THE CRYSTAL AND MOLECULAR STRUCTURES OF SOME INORGANIC, ORGANIC, AND BIOLOGICAL COMPOUNDS

by

NORMAN CAMERMAN

B.Sc.(Hon.), University of British Columbia, 1961

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the Department of

CHEMISTRY

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
September, 1964.
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Chemistry

The University of British Columbia,
Vancouver 8, Canada

Date July 23, 1964
The University of British Columbia

FACULTY OF GRADUATE STUDIES

PROGRAMME OF THE

FINAL ORAL EXAMINATION

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

of

NORMAN CAMERMAN

B.Sc., The University of British Columbia, 1961

WEDNESDAY, AUGUST 19th, 1964, AT 10:00 A.M.

IN ROOM 261, CHEMISTRY BUILDING

COMMITTEE IN CHARGE

Chairman: I. McT. Cowan

W.R. Cullen  S.A. Melzak
J.P. Kutney  E. Teghtsoonian
C.A. McDowell  J. Trotter

External Examiner: Dr. D.G. Watson

University of Washington
THE CRYSTAL AND MOLECULAR STRUCTURES OF SOME INORGANIC, ORGANIC, AND BIOLOGICAL COMPOUNDS

ABSTRACT

The crystal and molecular structures of cyanodimethylarsine (cacodyl cyanide), diiodomethylarsine, and "cacodyl disulphide" (dimethylarsino dimethylidithioarsinate) have been determined by x-ray diffraction of single crystals, in order to investigate the stereochemistry of arsenic.

The structure of cyanodimethylarsine was determined from projections along the three crystallographic axes. The central arsenic bond angle of 105° agrees well with that found in the halogendiphenylarsines, and is considerably larger than the value assumed in an electron-diffraction investigation of the corresponding dimethyl derivatives. There is an unusually short As...N intermolecular separation, which is indicative of charge-transfer bonding involving donation of nitrogen lone pair electrons to vacant arsenic 4d-orbitals, which probably accounts for the solidity of the compound whereas the halogendimethylarsines are all liquid at room temperature.

Diiodomethylarsine, one of the few simple arsenic derivatives which are solid at room temperature, was structurally determined from partial three-dimensional data in normal and generalized projections. Though severe absorption corrections had to be applied, values of the bond distances, valency angles, and intermolecular separations were obtained.

The molecular structure of "cacodyl disulphide" was determined precisely, utilizing the full three-dimensional data, and the analysis revealed the compound to be dimethylarsino dimethylidithioarsinate, an unusual structure, having one trivalent and one pentavalent arsenic atom in the molecule. The former atom has a trigonal-pyramidal configuration, while the latter is roughly tetrahedral. There is an unusually short intermolecular As\(^{III}\)...As\(^{III}\) separation which suggests charge-transfer bonding involving donation of lone pair electrons on each As\(^{III}\) to vacant 4d-orbitals on the other As\(^{III}\).

The structure of the alkaloid cleavamine, a product of the acidification of the Vinca rosea Linn alkaloid catharanthine, was determined from a three-dimensional x-ray analysis of cleavamine methiodide. The indole system is retained in cleavamine but the rest of the molecule
undergoes rearrangement via bond fission, the result being a tetracyclic alkaloid structurally resembling the known alkaloid quebrachamine. The presence of the iodide ion prevented accurate measurements of the bond distances and angles, but the measurements were sufficiently precise to differentiate between the different types of bonds and angles. The absolute configuration of cleavamine was also established.

The crystal and molecular structure of 5-iodo-2'-deoxyuridine, the first anti-viral agent to have proven clinical chemotherapeutic value, was elucidated by a three-dimensional x-ray investigation. The bond lengths and angles in the nucleoside were determined accurately, and indicated that the pyrimidine base is in the diketo form. The deoxyribose ring is puckered with C2' displaced 0.59 Å from the plane of the other four atoms; the distances and angles in the sugar ring are all normal. The most significant intermolecular distance is an I...O (carbonyl) separation of 2.96 Å; this is considerably shorter than the usual van der Waals contact (3.55 Å) and suggests charge-transfer bonding involving donation of oxygen lone pair electrons to vacant 5d-orbitals of the iodine atom. A strong intermolecular attraction of this type may be the cause of the increase in "melting temperature" observed when 5-iodo-2'-deoxyuridine is substituted for thymidine in DNA, and hence may be the molecular basis for the compound's anti-viral activity.

A preliminary x-ray investigation was carried out on crystals of sodium thymidylyl-(5'→3')-thymidylate-(5'). The cell constants and space group of the dinucleotide were determined, and the intensities of the three-dimensional diffraction maxima were recorded, but the absence of a relatively heavy atom made attempts at a complete structural elucidation unsuccessful.

A number of crystallographic computer programmes have been written for the IBM 1620 and 7040 computers; a list of these is given in an appendix.
GRADUATE STUDIES

Field of Study: Chemistry

Topics in Physical Chemistry
  J.A.R. Coope
  R.F. Snider
  A. Bree

Topics in Inorganic Chemistry
  N. Bartlett
  W.R. Cullen

Topics in Organic Chemistry
  D.E. McGreer
  J.P. Kutney
  R.E.I. Pincock

Crystal Structures
  J. Trotter

Related Studies:

  Calculus & Differential Equations
    W.H. Gage

  Digital Computer Programming
    C. Froese

  Linear Algebra
    R. Cleveland

  Quantum Mechanics
    W. Opechowski


8. N. Camerman and J. Trotter, "5-Iodo-2'-deoxyuridine: Relation of Structure to its Anti-viral Activity", Science, 144, 1348 (1964)
The crystal and molecular structures of cyanodimethylarsine (cacodyl cyanide), diiodomethylarsine, and "cacodyl disulphide" (dimethylarsino dimethyldithioarsinate) have been determined by x-ray diffraction of single crystals, in order to investigate the stereochemistry of arsenic.

The structure of cyanodimethylarsine was determined from projections along the three crystallographic axes. The central arsenic bond angle of 105° agrees well with that found in the halogendiphenylarsines, and is considerably larger than the value assumed in an electron-diffraction investigation of the corresponding dimethyl derivatives. There is an unusually short As-$\ldots$N intermolecular separation, which is indicative of charge-transfer bonding involving donation of nitrogen lone pair electrons to vacant arsenic 4d-orbitals, which probably accounts for the solidity of the compound whereas the halogendimethylarsines are all liquid at room temperature.

Diiodomethylarsine, one of the few simple arsenic derivatives which are solid at room temperature, was structurally determined from partial three-dimensional data in normal and generalized projections. Though severe absorption corrections had to be applied, values of the bond distances, valency angles, and intermolecular separations were obtained.

The molecular structure of "cacodyl disulphide" was determined precisely, utilizing the full three-dimensional data, and the analysis revealed the compound to be dimethylarsino dimethyldithioarsinate, an unusual structure, having one trivalent and one pentavalent arsenic atom in the molecule. The former atom has a trigonal-pyramidal configuration, while the latter is roughly tetrahedral. There is an unusually short intermolecular As$^{\text{III}}$...As$^{\text{III}}$ separation which suggests charge-transfer bonding involving donation of lone pair electrons on each As$^{\text{III}}$ to vacant 4d-orbitals on the other As$^{\text{III}}$. 
The structure of the alkaloid cleavamine, a product of the acidification of the *Vinca rosea* Linn alkaloid catharanthine, was determined from a three-dimensional X-ray analysis of cleavamine methiodide. The indole system is retained in cleavamine but the rest of the molecule undergoes rearrangement via bond fission, the result being a tetracyclic alkaloid structurally resembling the known alkaloid quebrachamine. The presence of the iodide ion prevented accurate measurements of the bond distances and angles, but the measurements were sufficiently precise to differentiate between the different types of bonds and angles. The absolute configuration of cleavamine was also established.

The crystal and molecular structure of 5-iodo-2'-deoxyuridine, the first anti-viral agent to have proven clinical chemotherapeutic value, was elucidated by a three-dimensional X-ray investigation. The bond lengths and angles in the nucleoside were determined accurately, and indicated that the pyrimidine base is in the di-keto form. The deoxyribose ring is puckered with C2' displaced 0.59 Å from the plane of the other four atoms; the distances and angles in the sugar ring are all normal. The most significant intermolecular distance is an I...O (carbonyl) separation of 2.96 Å; this is considerably shorter than the usual van der Waals contact (3.55 Å) and suggests charge transfer bonding involving donation of oxygen lone pair electrons to vacant 5d-orbitals of the iodine atom. A strong intermolecular attraction of this type may be the cause of the increase in "melting temperature" observed when 5-iodo-2'-deoxyuridine is substituted for thymidine in DNA, and hence may be the molecular basis for the compound's anti-viral activity.

A preliminary X-ray investigation was carried out on crystals of sodium thymidylyl-(5'→3')-thymidylate-(5'). The cell constants and space group of the dinucleotide were determined, and the intensities of the three-dimensional diffraction maxima were recorded, but the absence of a relatively heavy atom made attempts at a complete structural elucidation unsuccessful.

A number of crystallographic computer programs have been written for the IBM 1620 and 7040 computers; a list of these is given in an appendix.
To

MY MOTHER

for a lifetime of labor for her children
ACKNOWLEDGEMENTS

I am greatly indebted to Dr. J. Trotter for his invaluable assistance, advice, and friendship during the course of my graduate studies.

I would also like to express my thanks to Dr. W.R. Cullen for samples of the three arsenic-containing compounds, to Dr. J. Kutney for the sample of cleavamine methiodide, and to Dr. G.M. Tener for the dinucleotide sample and for much helpful discussion and advice on the dinucleotide and 5-ido-2'-deoxyuridine.

My thanks are also due to Dr. F.R. Ahmed for kindly making available his structure factor and Fourier programs, and to Dr. G.A. Mair for his anisotropic least squares program, for the IBM 1620 computer. I also thank Dr. J. Trotter for his assistance in writing several crystallographic computer programs.

For many of the illustrations in this thesis I am indebted to Mr. W. Griba.

Finally, I gratefully acknowledge the generous support of the National Research Council of Canada which awarded me a bursary for 1961-62, and a studentship for the period 1962-64.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>(i)</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>(ii)</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>(v)</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>(vi)</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>(viii)</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>(x)</td>
</tr>
<tr>
<td>GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I. SOME ASPECTS OF THE THEORY OF CRYSTAL STRUCTURE DETERMINATION</strong></td>
<td>4</td>
</tr>
<tr>
<td>I. Elementary Crystallography</td>
<td>5</td>
</tr>
<tr>
<td>A. External Geometry and Symmetry</td>
<td>5</td>
</tr>
<tr>
<td>B. Internal Structure - Lattice Theory</td>
<td>6</td>
</tr>
<tr>
<td>C. Space Groups</td>
<td>7</td>
</tr>
<tr>
<td>II. X-Ray Diffraction by a Crystal</td>
<td>9</td>
</tr>
<tr>
<td>A. Scattering by the Atoms</td>
<td>9</td>
</tr>
<tr>
<td>B. von Laue's Equations and Bragg's Law</td>
<td>10</td>
</tr>
<tr>
<td>C. Diffraction in Reciprocal Space</td>
<td>11</td>
</tr>
<tr>
<td>D. The Structure Factor</td>
<td>13</td>
</tr>
<tr>
<td>E. Intensities of Diffraction Maxima</td>
<td>14</td>
</tr>
<tr>
<td>F. Fourier Series and the Phase Problem</td>
<td>16</td>
</tr>
<tr>
<td>III. Structure Analysis Procedures</td>
<td>17</td>
</tr>
<tr>
<td>A. Determination of Structures</td>
<td>17</td>
</tr>
<tr>
<td>B. Refinement of Structures</td>
<td>19</td>
</tr>
<tr>
<td>C. Assessment of Accuracy</td>
<td>23</td>
</tr>
<tr>
<td><strong>PART II. THE STRUCTURES OF CYANODIMETHYLARSINE, DIIODOMETHYLARSINE, AND DIMETHYLARSINO DIMETHYLDITHIOARSENATE (&quot;CACODYL DISULPHIDE&quot;)</strong></td>
<td>24</td>
</tr>
<tr>
<td>I. CYANODIMETHYLARSINE</td>
<td>25</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>25</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>25</td>
</tr>
<tr>
<td>C. Structure Analysis</td>
<td>26</td>
</tr>
<tr>
<td>D. Discussion</td>
<td>29</td>
</tr>
<tr>
<td>II. DIIODOMETHYLARSINE</td>
<td>33</td>
</tr>
<tr>
<td>------------------------</td>
<td>----</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>33</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>33</td>
</tr>
<tr>
<td>C. Structure Analysis</td>
<td>34</td>
</tr>
<tr>
<td>D. Discussion</td>
<td>35</td>
</tr>
<tr>
<td>III. DIMETHYLARSINO DIMETHYLDITHIOARSDINATE (&quot;CACODYL DISULPHIDE&quot;)</td>
<td>40</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>40</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>41</td>
</tr>
<tr>
<td>C. Structure Analysis</td>
<td>42</td>
</tr>
<tr>
<td>D. Discussion</td>
<td>44</td>
</tr>
<tr>
<td>PART III. THE STRUCTURES OF CLEAVAMINE METHIODIDE AND 5-IODO-2'-DEOXYURIDINE, AND CRYSTAL DATA FOR SODIUM THYMIDYL-(5' → 3')-THYMIDYLATE-(5')</td>
<td>53</td>
</tr>
<tr>
<td>I. CLEAVAMINE METHIODIDE</td>
<td>54</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>54</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>55</td>
</tr>
<tr>
<td>C. Structure Analysis</td>
<td>56</td>
</tr>
<tr>
<td>D. Discussion</td>
<td>65</td>
</tr>
<tr>
<td>II. 5-IODO-2'-DEOXYURIDINE</td>
<td>69</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>69</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>70</td>
</tr>
<tr>
<td>C. Structure Analysis</td>
<td>70</td>
</tr>
<tr>
<td>D. Discussion</td>
<td>77</td>
</tr>
<tr>
<td>III. SODIUM THYMIDYL-(5' → 3')-THYMIDYLATE-(5')</td>
<td>84</td>
</tr>
<tr>
<td>A. Introduction</td>
<td>84</td>
</tr>
<tr>
<td>B. Experimental</td>
<td>86</td>
</tr>
<tr>
<td>C. Attempts at Structure Elucidation</td>
<td>87</td>
</tr>
<tr>
<td>APPENDIX. A BRIEF DESCRIPTION OF SOME CRYSTALLOGRAPHIC PROGRAMS FOR THE IBM 1620 AND 7040 COMPUTERS</td>
<td>90</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>94</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The 14 Bravais lattices</td>
<td>8</td>
</tr>
<tr>
<td>2.</td>
<td>Scattering from a row of lattice points</td>
<td>10</td>
</tr>
<tr>
<td>3.</td>
<td>Relationship between $\hat{s}$ and crystal plane</td>
<td>10</td>
</tr>
<tr>
<td>4.</td>
<td>Relationship between vectors $\hat{s}_0$, $\hat{s}$, and $\hat{S}$</td>
<td>11</td>
</tr>
<tr>
<td>5.</td>
<td>Construction of a reciprocal lattice</td>
<td>12</td>
</tr>
<tr>
<td>6.</td>
<td>Diffraction in reciprocal space</td>
<td>13</td>
</tr>
</tbody>
</table>

**CYANODIMETHYLARSINE**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Electron-density projection and view of the structure along c</td>
<td>27</td>
</tr>
<tr>
<td>8.</td>
<td>Electron-density projections along a and b</td>
<td>28</td>
</tr>
<tr>
<td>9.</td>
<td>Measured bond lengths and valency angles</td>
<td>30</td>
</tr>
<tr>
<td>10.</td>
<td>Packing of the molecules, projected along [001]</td>
<td>32</td>
</tr>
</tbody>
</table>

**DIIODOMETHYLARSINE**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.</td>
<td>Electron-density projection and view of the structure along b</td>
<td>36</td>
</tr>
<tr>
<td>12.</td>
<td>Cosine and sine hll generalized projections</td>
<td>37</td>
</tr>
<tr>
<td>13.</td>
<td>Measured bond lengths and valency angles</td>
<td>38</td>
</tr>
<tr>
<td>14.</td>
<td>Packing of the molecules, projected onto (010)</td>
<td>39</td>
</tr>
</tbody>
</table>

**DIMETHYLARSINO DIMETHYLTHIOARSINATE ("CACODYL DISULPHIDE")**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.</td>
<td>Superimposed sections of the three-dimensional electron-density distribution, through the atomic centres perpendicular to b, and perspective view of the molecule</td>
<td>45</td>
</tr>
<tr>
<td>16.</td>
<td>Packing of the molecules, projected along [100]</td>
<td>48</td>
</tr>
<tr>
<td>17.</td>
<td>View of the molecule along the $S_4$-$As_1$ bond</td>
<td>49</td>
</tr>
</tbody>
</table>
Figure  Page

CLEAVAMINE METHIODIDE

18. Superimposed sections of the three-dimensional electron-density distribution, through the atomic centres parallel to (010), and perspective view of the molecule .... 59
20. Packing of the molecules, projected onto (100) .... 63

5-iodo-2'-deoxyuridine

21. Superimposed sections of the three-dimensional electron-density distribution, through the atomic centres parallel to (100), and perspective view of the molecule .... 73
22. Bond lengths and valency angles .................... 76
23. Isometric projection showing the molecular packing .... 82
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. The 32 Point Groups or Crystal Symmetry Classes</td>
<td>6</td>
</tr>
<tr>
<td>CYANODIMETHYLARSINE</td>
<td></td>
</tr>
<tr>
<td>II. Positional Parameters</td>
<td>29</td>
</tr>
<tr>
<td>DIIODOMETHYLARSINE</td>
<td></td>
</tr>
<tr>
<td>III. Positional Parameters</td>
<td>35</td>
</tr>
<tr>
<td>DIMETHYLARSINO DIMETHYLTHIOARSINATE (&quot;CACODYL DISULPHIDE&quot;)</td>
<td></td>
</tr>
<tr>
<td>IV. Positional and Thermal Parameters with Standard Deviations</td>
<td>46</td>
</tr>
<tr>
<td>V. Bond Distances and Valency Angles with Standard Deviations</td>
<td>46</td>
</tr>
<tr>
<td>VI. Shorter Intermolecular Distances</td>
<td>47</td>
</tr>
<tr>
<td>VII. Progress of Refinement</td>
<td>47</td>
</tr>
<tr>
<td>CLEAVAMINE METHIODIDE</td>
<td></td>
</tr>
<tr>
<td>VIII. Positional and Thermal Parameters with Standard Deviations, and Peak Electron-Densities</td>
<td>60</td>
</tr>
<tr>
<td>IX. Shorter Intermolecular Distances</td>
<td>61</td>
</tr>
<tr>
<td>X. Deviations from Mean Planes</td>
<td>64</td>
</tr>
<tr>
<td>XI. Determination of the Absolute Configuration</td>
<td>66</td>
</tr>
<tr>
<td>5-IODO-2'-DEOXYURIDINE</td>
<td></td>
</tr>
<tr>
<td>XII. Crystal Data</td>
<td>71</td>
</tr>
<tr>
<td>XIII. Positional and Thermal Parameters with Standard Deviations</td>
<td>75</td>
</tr>
<tr>
<td>XIV. Deviations from Mean Planes</td>
<td>78</td>
</tr>
<tr>
<td>XV. Shorter Intermolecular Distances</td>
<td>79</td>
</tr>
</tbody>
</table>
GENERAL INTRODUCTION
This thesis is concerned with the structural elucidation of a number of inorganic, organic, and biologically-active compounds by single-crystal x-ray diffraction. The contents are divided into three parts:

Part I is a description of some of the theory of crystal symmetry, x-ray diffraction by single crystals, and methods presently employed in the elucidation of crystal structures. It is by no means a thorough or comprehensive exposition, and is intended only as an introduction for the reader to some of the methods employed in determining the structures which follow in parts II and III.

Part II consists of the determination of the crystal and molecular structures of three arsenic-containing compounds, undertaken as part of a project to investigate the stereochemistry of arsenic. At the commencement of my graduate studies there were no computing facilities available, and structure factor and Fourier calculations had to be done manually on a desk calculator; this severely limited the size of the molecules whose structural determination could be attempted, as well as limiting the methods which could be used. Thus the first two compounds discussed, cyanodimethylarsine and diiodomethylarsine, are relatively small molecules and were investigated by essentially two-dimensional projections. Toward the end of 1961 the UBC Computing Centre acquired an IBM 1620 computer and shortly thereafter Dr. F.R. Ahmed's structure-factor and Fourier programs were made available to us. Larger molecules could now be investigated, and the structure of "cacodyl disulphide" was determined, utilizing the full three-dimensional data.

Part III describes the complete structural determination by three-dimensional means of two large organic and biological molecules - the alkaloid cleavamine and a nucleoside, 5-ido-2'-deoxyuridine. In addition, preliminary data is given for a dinucleotide, sodium thymidylyl-(5'→3')-thymidylate-(5').
A brief appendix follows part III. In it are short descriptions of several crystallographic computer programs, written and repeatedly utilized in the course of the structural investigations which are described in this thesis.
PART I

SOME ASPECTS OF THE
THEORY OF CRYSTAL STRUCTURE
DETERMINATION
I. ELEMENTARY CRYSTALLOGRAPHY

A. External Geometry and Symmetry

It was recognized in the 17th century that while crystals of the same substance might appear in many shapes and forms, the angles between the principal faces remained constant. It is thus the orientation of the faces which is a property of the crystal.

If three non-coplanar vectors are chosen as a reference system and we arbitrarily choose a plane of the crystal (called the standard or parametral face), with axial intercepts of a, b, and c, then any face of the crystal has axial intercepts of a/h, b/k, c/l; h, k, l are termed the Miller indices of the plane. Hauy recognized in 1784 that the ratios of the indices of any face of any crystal are small integers (law of rational indices).

When we say a geometrical figure has symmetry, we mean that if some operation (e.g. rotation about an axis) is performed on the figure it will produce an orientation which is indistinguishable from the original figure. Though a geometrical figure can be constructed to display any desired degree of symmetry, the law of rational indices limits the symmetry elements possible in a crystal to 1-fold, 2-fold, 3-fold, 4-fold, or 6-fold rotation or rotatory inversion axes.

It was first shown by Hessel in 1830 that there were only 32 ways of combining the above symmetry operations into self-consistent sets, and these sets are known as the 32 crystal classes or point groups (Table I).
Table I. The 32 Crystal Classes
(The eleven classes of distinct Laue symmetry are separated by vertical lines)

<table>
<thead>
<tr>
<th>System</th>
<th>Axes and angles</th>
<th>Crystal classes (point groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>(a) (b) (c) (\alpha) (\beta) (\gamma)</td>
<td>1  (\bar{1})  (C_1)  (C_i)</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>(a) (b) (c) (90^\circ) (\beta) (90^\circ)</td>
<td>2  (m) or (\bar{2})  (2/m)  (C_2)  (C_s)  (C_{2h})</td>
</tr>
<tr>
<td>Orthorhombic</td>
<td>(a) (b) (c) (90^\circ) (90^\circ) (90^\circ)</td>
<td>222  (mm2)  (mmm)  (D_2)  (C_{2v})  (D_{2h})</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>(a) (a) (c) (90^\circ) (90^\circ) (90^\circ)</td>
<td>4  (\bar{4})  (\bar{4}m)  (4mm)  (\bar{4}m)  (4/mmm)  (D_4)  (C_{4v})  (D_{2d})  (D_{4h})</td>
</tr>
<tr>
<td>Trigonal (rhombohedral)</td>
<td>(a) (a) (a) (90^\circ) (90^\circ) (120^\circ)</td>
<td>3  (\bar{3})  (\bar{3}m)  (\bar{3}m)</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>(a) (a) (c) (90^\circ) (90^\circ) (120^\circ)</td>
<td>6  (\bar{6})  (\bar{6}m)  (6mm)  (\bar{6}m)  (6/mmm)  (D_6)  (C_{6v})  (D_{3h})  (D_{6h})</td>
</tr>
<tr>
<td>Cubic</td>
<td>(a) (a) (a) (90^\circ) (90^\circ) (90^\circ)</td>
<td>23  (m3)  (\bar{4}3m)  (m3m)</td>
</tr>
</tbody>
</table>

B. Internal Structure - Lattice Theory

The idea that the regular external geometry and symmetry of crystals reflected some ordered arrangement of the ultimate matter of which the crystal is composed, that is that crystals must be based on some sort of lattice structure, also arose in the 17th century. By lattice is meant some regular geometrical repetition in space of identical units, and it may be represented by an orderly three-dimensional array of points. Connecting these points results in an indefinitely extended regular series of parallelopipeds or unit cells, defined by the three non-coplanar edge vectors and the contained
angles. Since there are usually many ways of connecting the lattice points the choice of a unit cell is dictated by size, simplicity, and ease of visualizing the symmetry and carrying out mathematical calculations. There are seven possible unit cell shapes, characterized by the symmetry elements contained in their lattices, and these make up the seven crystal systems (Table 1). When the possibility of centering in the lattice is considered, only fourteen different Bravais (1) lattices belonging to the seven systems can be distinguished (Fig. 1).

C. Space Groups

Since lattices consist of regular arrays of identical units, typified by lattice points, the structure can be brought into self-coincidence in a new way, namely by translations along any of the lattice directions. This means that a new class of symmetry operations is applicable to lattice structures, obtained by combining reflections and rotations with translations, the former combination giving rise to glide planes, and the latter to screw axes. The self-consistent sets of symmetry elements, including those having translational properties, constitute groups of movements which are infinite but discontinuous, and these are called space groups. Federow (2), Schoenflies (3), and Barlow (4) showed independently that 230 different space groups are possible. These are tabulated extensively in the International Tables for X-Ray Crystallography, Vol. 1 (5).

The basis for space group determination by x-ray diffraction is that the presence of glide planes and screw axes produce systematic absences in the x-ray spectra. Because non-translational symmetry elements cause no such systematic absences and because Friedel's law asserts that x-ray diffraction patterns have an inherent center of symmetry, it is not always possible to pinpoint the space group from the diffraction pattern alone.
Fig. 1. The 14 Bravais lattices.
II. X-RAY DIFFRACTION BY A CRYSTAL

A. Scattering by the Atoms

If an electron is located in the path of an x-ray beam it is forced into oscillation by the electromagnetic field of the x-rays impinging upon it. Due to this acceleration the electron in turn becomes a source of radiation of the same frequency and wavelength, and in this way is said to scatter the impinging radiation.

The scattering power of an atom, designated \( f_0 \) (form factor), is expressed in terms of the scattering power of a single free electron, and therefore the maximum scattering by an atom is equal to its atomic number. Since the electrons are distributed throughout the volume of the atom, destructive interference of the wavelets scattered from different parts of the electron cloud sets in, and \( f_0 \) decreases with increasing Bragg angle \( \theta \).

Values of \( f_0 \) have been computed for most common atoms and ions by theoretical methods, and are tabulated as a function of \( \sin \theta / \lambda \) in the International Tables, Vol. III (5).

In calculating \( f_0 \) values the atoms are assumed to be at rest, but in actuality thermal oscillations of the atoms result in a spreading of the electron distribution, and so decrease the intensities of the spectra. The temperature corrected scattering factor, \( f \), is given by

\[
f = f_0 e^{-B \sin^2 \theta / \lambda^2}
\]  

(1)

\( B \), the Debye-Waller temperature factor (6,7), is related to the mean square displacement of the atom normal to the reflecting plane, \( \mu^2 \):

\[
B = 8 \pi^2 \mu^2
\]  

(2)

In crystal-structure analysis, \( B \) is usually treated as an empirical parameter.
B. von Laue's Equations and Bragg's Law

Consider a parallel beam of x-rays falling upon a space lattice in a direction defined by the unit vector \( \vec{s}_0 \) (Fig. 2). For constructive interference to occur in the direction of the unit vector \( \vec{s} \), the path difference between waves scattered from successive points separated by the basis vector \( \vec{a} \) must equal an integral number of wavelengths, \( n\lambda \). Thus

\[
a (\cos \alpha_0 - \cos \alpha) = n\lambda
\]  

or in three dimensions,

\[
a_i (\cos \alpha_{0i} - \cos \alpha_i) = n_i \lambda \quad i = 1, 2, 3
\]  

This set of equations was derived by von Laue as the conditions for diffraction from a crystal lattice, and may be put into vector form:

\[
\frac{\vec{a}_i}{n_i} \cdot \frac{\vec{s} - \vec{s}_0}{\lambda} = 1 \quad \text{or} \quad \frac{\vec{a}_i}{n_i} \cdot \frac{\vec{s}}{\lambda} = 1
\]

\( \vec{s} \) is known as the diffraction vector, and is in the direction of the bisector of the incident and diffracted beams, since the magnitudes of \( \vec{s} \) and \( \vec{s}_0 \) are equal. Subtraction of any two of equations (4) shows
\[
\left( \frac{\mathbf{a}_i}{n_i} - \frac{\mathbf{a}_j}{n_j} \right) \cdot \mathbf{S} = 0 \quad i \neq j
\]

which means that \( \mathbf{S} \) is perpendicular to the vectors \( \frac{\mathbf{a}_i}{n_i} - \frac{\mathbf{a}_j}{n_j} \); since these vectors are in the plane \( n_1n_2n_3 \), \( \mathbf{S} \) is perpendicular to this plane (Fig. 3). Thus the bisector of the incident and diffracted beams is identified with the normal to the \( n_1n_2n_3 \), \( \text{(or } hkl \text{)} \) plane - the first step in the proof of Bragg's law, and the justification for the concept of each diffraction as a reflection of the rays from lattice planes.

The magnitude of the vector \( \mathbf{S} \) is obviously \( 2 \sin \theta \) (Fig. 4) and the distance of the plane \( n_1n_2n_3 \) from the origin is

\[
d = \frac{\mathbf{a}_i \cdot \mathbf{S}}{n_i} = \frac{\lambda}{2 \sin \theta}
\]

or

\[
2d_{hkl} \sin \theta = \lambda
\]

which is Bragg's law, with the order \( n \) absorbed in the integers \( n_1n_2n_3 \).

Indices of diffraction, unlike Miller indices, can have common factors; if \( n \) is this common factor the spacing corresponding to indices \( nh, nk, nl \), is regarded as \( 1/n \)th the spacing corresponding to indices \( hkl \), and the \( n \)th order reflection from the \( hkl \) plane can be regarded as a 1st order reflection from this fictitious \( nh, nk, nl \) plane.

C. Diffraction in Reciprocal Space

To attempt to visualize a large number of planes of varying orientation in a crystal is very difficult; a great simplification occurs if we construct a reciprocal lattice. This is done by drawing the normal to each plane in direct space from a chosen origin point, and representing each plane by a
point on the normal at a distance \( d \) from the origin such that

\[
\rho d = K^2 \tag{7}
\]

where \( K \) is a constant. The infinite number of fictitious planes of sub-multiple spacings which correspond to second and higher order reflections from a given crystal plane have a common normal from the origin, but the corresponding reciprocal lattice points are regularly spaced out along each normal in accordance with (7). Extending Fig. 5 to three dimensions readily shows that the indices \((hkl)\) of any general plane in real space are simply the coordinates of the corresponding point in reciprocal space.

It is convenient to make the constant \( K^2 \) in (7) equal to the x-ray wave-length \( \lambda \), so that

\[
\rho = \frac{\lambda}{d} = 2 \sin \theta \tag{8}
\]

From (8) it is clear that a sphere described about the center of the reciprocal lattice of radius \( \rho = 2 \) contains all the reflections which may be observed with radiation of wave-length \( \lambda \). This is termed the limiting sphere (Fig. 6).

The reciprocal lattice affords a simple geometrical interpretation of diffraction on the basis of Bragg's law. (Fig. 6). From (8),

\[
\sin \theta = \frac{\rho(p)}{2} \tag{2}
\]
for any point $P$ at a distance $\rho$ from the origin, a condition satisfied for any point on the sphere OAP, as

$$\sin OAP = \frac{OP}{OA} = \frac{\rho(P)}{2}.$$  

Thus any point on the surface of this "sphere of reflection", and no other point, is capable of reflecting radiation incident in the direction of its diameter $AO$.

The basis of most single-crystal diffraction methods is to rotate or oscillate the crystals so that in effect, the direction of incidence of the x-rays is changed, and many reciprocal lattice points will be caused to intersect the sphere of reflection.

### D. The Structure Factor

In the derivation of Bragg's law we assumed the scattering centers (atoms) to lie on the reflecting crystal planes, so that each scattered wavelet was in phase with those scattered from successive parallel planes to build up a strong diffracted beam. In general, however, the atoms are not all situated on the various crystal planes, but are distributed through the cell, so that the waves scattered in any order $hk\ell$ by the atoms in a unit cell differ in phase and must be compounded vectorially.

For a set of planes $hk\ell$, the spacings $a/h$, $b/k$, $c/\ell$ correspond to positions from which scattering differs in phase by $2\pi$. The phase difference $P_x$ for displacement $x$ in the $a$ direction is given by

$$\frac{x}{a/h} = \frac{P_x}{2\pi} \quad \text{or} \quad P_x = 2\pi \frac{h(x/a)}{\alpha}$$

(11)
Extending to three dimensions, the total phase change that an atom at \((xyz)\) contributes to the plane \((hk\ell)\) is

\[ P = 2\pi (hx/a + ky/b + l\ell/c) \]  

(12)

Adding vectorially the contributions from all \(N\) atoms in the unit cell,

\[ F(hk\ell) = \sum_{j=1}^{N} f_j e^{i2\pi (hx_j/\alpha + ky_j/\beta + l\ell_j/\gamma)} \]  

(13)

The resultant \(F(hk\ell)\) is termed the structure factor. It is a complex quantity, with amplitude \(|F(hk\ell)|\) and phase \(\alpha(hk\ell)\). It can be evaluated by means of the expressions:

\[ |F(hk\ell)| = \sqrt{A^2 + B^2} \]

\[ \alpha(hk\ell) = \tan^{-1} \frac{B}{A} \]  

(14)

where

\[ A = \sum_{j=1}^{N} f_j \cos 2\pi (hx_j/\alpha + ky_j/\beta + l\ell_j/\gamma) \]

\[ B = \sum_{j=1}^{N} f_j \sin 2\pi (hx_j/\alpha + ky_j/\beta + l\ell_j/\gamma) \]

If the summations are carried over the coordinates of the equivalent positions of the space group, simplified expressions often result. In particular if a center of symmetry is present, and chosen as the origin, every vector of phase angle \(2\pi x_j/\alpha\) is accompanied by another of phase angle \(-2\pi x_j/\alpha\). The \(B\) part of the structure factor therefore cancels to zero, and the possible phase angles of \(F(hk\ell)\) are limited to 0 or \(\pi\) depending on whether the expression for \(A\) is positive or negative.

E. Intensities of Diffraction Maxima

Most single crystals behave as if they were not perfect crystals, but made up sub-microscopically of a conglomerate or mosaic of small blocks, each block a perfect crystal, but the adjacent blocks not accurately fitted together (8). The crystal thus reflects x-rays over a small range about the
Bragg angle. If the crystal is rotated through the reflecting position with angular velocity $w$, and $I_0$ is the intensity of the incident beam and $E$ the total reflected energy, the ratio $Ew/I_0$ is called the integrated reflection, and is characteristic of a given crystal plane. For a mosaic crystal block of volume $dV$, small enough that no absorption occurs, Darwin (8) has given the following relation

$$\frac{Ew}{I_0} = (N \frac{e^2}{mc^2})^2 \lambda^3 \frac{1+\cos^2\theta}{2\sin\theta} |F|^2 dV \quad (15)$$

In equation 15, $N$ is the number of unit cells per unit volume, and $e^2/mc^2$ arises from the classical formula for scattering by an electron. The polarization term, $(1+\cos^2\theta)/2$, arises because the reflected beam is partially polarized while the incident beam is unpolarized. The Lorentz factor, $1/\sin\theta$, takes into account the differing speeds at which different reciprocal lattice points move through the sphere of reflection. The expression for the Lorentz factor varies with the experimental conditions, and so equation 15 must be modified in use. If intensities are collected by the equi-inclination Weissenberg method, for example, (15) becomes

$$\frac{Ew}{I_0} \propto \frac{1+\cos^2\theta}{\xi \cos \theta} |F|^2 \quad (16)$$

where $\xi$ is the cylindrical coordinate of the reciprocal lattice point.

Other factors may also affect the intensities. X-rays are absorbed in crystals in accordance with the Beer-Lambert law, resulting in diminution of the various reflections in a complicated manner depending on the crystal size and shape. Absorption corrections are difficult to calculate, but values for crystal specimens of simple shapes are given in the International Tables (5).

Extinction effects also occur in crystals which are not ideal mosaics; primary extinction refers to the re-reflection of the x-rays from parallel
planes, while secondary extinction refers to the weakening of the incident beam as it progresses through the crystal due to reflections from planes in its path. Extinction effects are difficult to overcome, and are usually ignored.

F. Fourier Series and the Phase Problem

W.H. Bragg recognized in 1915 (9) that the electron density in a crystal is a triply periodic function, and may be represented in the form of a Fourier series

\[ \rho(xyz) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \sum_{r=0}^{\infty} A(pqr) e^{-2\pi i(px+qy+rz)} \]  

where \( x, y, z \) are fractions of the cell edges, \( x/a, y/b, z/c \).

It can easily be shown that the Fourier coefficients, \( A(pqr) \), are equal to \( F(hkl)/V \), so

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

Letting \( F(hkl) = A + iB \), \( F(hkl') = A - iB \), and \( 2\pi(hx+ky+lz) = \Theta \), then,

\[ \rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} (A \frac{e^{i\Theta} + e^{-i\Theta}}{2} - iB \frac{e^{i\Theta} - e^{-i\Theta}}{2}) \]  

Hence the electron density is everywhere real, and

\[ \rho(-x,-y,-z) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} (A \cos \Theta + B \sin \Theta) \]  

which means that the electron densities at \((X,Y,Z)\) and \((-X,-Y,-Z)\) will differ unless \( B = 0 \), which, recalling section D, occurs when the origin of the space group is at a center of symmetry.
Writing

\[
A = |F(hkl)| \cos \alpha (hkl) \\
B = |F(hkl)| \sin \alpha (hkl),
\]
equation 19 becomes

\[
\rho(xyz) = \frac{1}{V} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} |F(hkl)| \cos [\theta - \alpha(hkl)]
\]

(20)

Equation 20 illustrates the phase problem in X-ray crystallography. We can find the values of $|F(hkl)|$ from the reflection intensities, but the relative phases are lost and must be recovered by some means or other before the electron-density distribution in the crystal can be calculated.

III. STRUCTURE ANALYSIS PROCEDURES

A. Determination of Structures

(a) Trial and Error Methods

An approximately correct structure (i.e. correct positioning of the atoms in the unit cell) is sought from chemical considerations (chemical formula, study of models of the molecules) combined with crystallographic data (limitations on possible positioning in the cell due to packing considerations). An indication of an approximately correct structure is reasonable agreement between observed and calculated structure amplitudes, generally indicated by a low ($\leq \sim 0.4$) value of the discrepancy factor

\[
R = \sum |F_o| - |F_c| / \sum |F_o|
\]

When this is attained, a Fourier series with the measured structure amplitudes and calculated phases as coefficients can be summed, and from the resulting electron-density plot revised and more accurate atomic coordinates can be chosen, from which to re-calculate structure factors. This iterative procedure is repeated until a minimum value of the discrepancy is attained.

(b) Heavy Atom Methods - the Patterson Synthesis

By far the most widely-used method, and the one which has yielded the most results, is the heavy atom method. A heavy atom is one which contains a
much larger number of electrons than do the other atoms of the molecule and is therefore of higher scattering power to x-rays. Hence its contribution to the structure factors tends to dominate the contributions of the other atoms for a high percentage of the various orders \((hkl)\). If the position of the heavy atom is known, structure factors may be calculated, based on the heavy atom alone, whose phases are near enough correct to show up the other atoms in an electron density distribution plotted with the observed structure amplitudes and heavy atom phases as Fourier coefficients.

The location of the heavy atom can usually be ascertained by use of a Patterson synthesis. In 1934 Patterson (10) showed that the function

\[
P(xyz) = \frac{1}{V^2} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F(hkl)|^2 \cos 2\pi (hx+ky+lz) \tag{21}
\]

which can be computed from the observed data alone, has maxima whose vector distances from the origin correspond physically to all possible interatomic vector distances in the crystal. The height of a Patterson peak is dependent on the atomic numbers of the two atoms involved, so vectors between heavy atoms will give rise to Patterson peaks of relatively great prominence. Systematic procedures have been developed to derive the atomic coordinates from the vector distribution (11).

Though it may now appear that the structure of any crystal may be determined simply by interpreting its Patterson function, it must be pointed out that a crystal with \(N\) atoms in the unit cell has \(N^2\) vectors between those atoms, and thus for a structure of moderate complexity the resultant Patterson map contains a large number of broad unresolved peaks due to overlap, severely limiting the amount of information which can be derived from it.

(c) Direct Methods

The phase values are sought directly from a knowledge of the structure amplitudes alone. Harker and Kasper (12), applied Cauchy's and Schwartz's
inequalities to modified structure factors and derived inequality relationships between some structure factors which impose limitations on phases due to crystal symmetry. More general systems of inequalities have been derived by Karle and Hauptman (13) who took into account the positiveness of the electron density. Relationships between the signs of some structure amplitudes in centrosymmetric space groups have been derived by Sayre (14), Cochran (15), and Zachariasen (16). The most recent and perhaps most promising approach has been the statistical one of Hauptman and Karle (17).

Although some structures have been completely elucidated by all of the above methods, these direct methods have not met with widespread use as of yet, perhaps partly because of their complexity.

B. Refinement of Structures

(a) Fourier Methods

Refinement of coordinates by successive Fourier syntheses is most commonly carried out in the first stages of refinement, until all atoms are fairly accurately located. The major source of error in Fourier syntheses is due to series termination effects, i.e. terminating the Fourier series while the remaining coefficients are still appreciable.

The summation of a triple Fourier series is extremely laborious, and when adequate computational facilities are not available it is customary to plot the electron-density projected onto two or more planes by use of a double Fourier series, which can be computed from zonal data alone.

A normal projection is mapped on a plane at right angles to the direction of projection (usually a crystallographic axis), and corresponds to an integration of the electron-density over the period of the direction of projection. For example,

$$\rho(\text{XYZ}) = \int_0^c \rho(\text{xyz}) \, dz$$

or in fractional coordinates,

$$\rho(\text{xy}) = \int_0^1 \rho(\text{xyz}) \, dz$$
Substituting from (18) for $\rho(xyz)$ and integrating,

$$\rho(xy) = \frac{1}{S} \sum_{h} \sum_{k} F_{hko} e^{-2\pi i(hx+ky)}$$

where $S$ is the cross-sectional area normal to the axis of projection.

In a generalized projection, the electron density is modulated by a factor $e^{2\pi iLz}$, where $L$ is a constant value of the index $l$. Thus

$$\rho_L(xy) = \int_0^1 \rho(xyz) e^{2\pi iLz} dz$$

$$\rho_L(xy) = \frac{1}{S} \sum_{h} \sum_{k} F_{hko} e^{-i2\pi(hx+ky)} = c_L(xy) + is_L(xy)$$

by the same mathematical procedure as above.

Generalized projections can also yield approximate third coordinates of the atoms. The expression for the contributions to a structure factor belonging to the $L$th layer from an atom $j$ is

$$F_{j}(hkL) = (f_j(hkL)e^{2\pi iLz}) e^{2\pi i(hx+ky)}$$

If the zero and $L$th layers are close together, $f_j(hkL) \sim f_j(hk0)$ and substituting into the expressions for the electron densities, we get

$$\rho(xyz) \sim \rho_L(xy) e^{2\pi iLz_j}$$

which means that the generalized projection of the $j$th atom is the same as its normal projection except that it is modulated by the factor $e^{2\pi iLz_j}$.

Thus $z_j$ can be evaluated from the expression $2\pi Lz_j = \tan^{-1} S_L/C_L$.

Series termination errors can be eliminated from Fourier methods by refinement by difference syntheses, utilizing $(|F_o| - |F_c|)$ as coefficients in the Fourier series (19). Atoms which deviate slightly from their true positions lie on steep slopes of the difference map and should be shifted up the gradients. The temperature factor of the atom can also be refined by noting if the atom lies in a region of positive or negative density on the
difference map. Difference syntheses are also commonly employed to locate hydrogen atoms.

(b) Differential Synthesis

First suggested by Booth (20), the differential synthesis seeks the electron-density maxima by evaluating the first differentials of the electron-density at the present atomic centers. Writing \( \rho = \frac{1}{V} \sum q \rho_0 \cos(\theta - \alpha) \), the slope of the electron density in the x-direction at the point \((x_n, y_n, z_n)\) is

\[
\left( \frac{\partial \rho}{\partial x} \right)_n = \frac{-2\pi}{aV} \sum q \rho_0 \sin(\theta_n - \alpha) \] (27)

The amounts by which the point of maximum electron density deviates from this point are then the solutions of the three equations

\[
\left( \frac{\partial^2 \rho}{\partial x^2} \right)_n \Delta x_n + \left( \frac{\partial^2 \rho}{\partial x \partial y} \right)_n \Delta y_n + \left( \frac{\partial^2 \rho}{\partial x \partial z} \right)_n \Delta z_n + \left( \frac{\partial \rho}{\partial x} \right)_n = 0 \] (28)

et cetera.

When the crystallographic axes are orthogonal, and the atoms spherically symmetric, these equations reduce to

\[
\Delta x_n = - \frac{\left( \frac{\partial \rho}{\partial x} \right)_n}{\left( \frac{\partial^2 \rho}{\partial x^2} \right)_n} \] etc. (29)

Series termination errors also affect the differential synthesis, but may be allowed for by computing an \( F_2 \) synthesis separately and applying back-shift corrections to the atomic coordinates (21). The advantage of the differential synthesis over the usual Fourier methods is that the point of maximum electron-density is precisely located without recourse to interpolation, and the tedious labor of plotting Fourier maps and determining coordinates therefrom is obviated.
(c) Least Squares

The method of least squares was first applied to crystal structure refinement by Hughes (22), and with the increased use of digital computers, is becoming the most common method of refinement today.

The theory of errors predicts that if errors in the measured $F_0$'s follow the normal or Gaussian law, then the best atomic parameters are those which result in a minimization of the quantity

$$\sum w(hk\ell) \left( |F_0(hk\ell)| - |F_c(hk\ell)| \right)^2$$

where $w(hk\ell)$ is a weight inversely proportional to the square of the probable error in $F_0(hk\ell)$. A small change $\Delta x_n$ in the x-coordinate of the nth atom of a structure changes $F_c$ by an amount

$$\frac{\partial F_c}{\partial x_n} \Delta x_n$$

Changes to all coordinates simultaneously result in a change in $F_c$ of amount

$$\Delta F_c = \sum_{n=1}^{N} \left( \frac{\partial F_c}{\partial x_n} \Delta x_n + \frac{\partial F_c}{\partial y_n} \Delta y_n + \frac{\partial F_c}{\partial z_n} \Delta z_n \right)$$

The correct values of $\Delta x_n$, etc., are those which most nearly equate $\Delta F_c$ to $F_0 - F_c$ for all the equations of this type possible (one for each $F_0$ value).

These linear "observational equations" must be reduced to a set of "normal equations", the nth of which is formed by multiplying both sides of (30) by $w \frac{\partial F_c}{\partial x_n}$, and then adding the q left-hand sides and the q right-hand sides to get equations of the type:

$$\sum_q w \left( \frac{\partial F_c}{\partial x_n} \Delta x_n \right) + \sum_m \frac{\partial F_c}{\partial x_m} \left( \frac{\partial F_c}{\partial y_m} \Delta y_m + \frac{\partial F_c}{\partial z_m} \Delta z_m \right) = \sum_q w (F_0 - F_c) \frac{\partial F_c}{\partial x_n}$$

(31)
The final result is a set of \(3N\) simultaneous equations which have to be solved for \(3N\) unknowns, \(\Delta x_n, \Delta y_n, \Delta z_n\).

Refinement by least squares has many advantages: it is free from series termination errors, it allows varying weights according to the reliability of the data, and it is possible to also refine a scale factor, and individual isotropic or anisotropic temperature factors for each atom.

\[ \text{(c) Assessment of Accuracy} \]

The standard deviations of electron density and atomic coordinates have been given by Cruickshank (23) as

\[ \sigma(p) = \frac{1}{V} \left\{ \sum (F_0 - F_C)^2 \right\}^{\frac{1}{2}} \]

and

\[ \sigma(x_n) = \frac{2\pi}{aV} \left\{ \sum F_0 (F_0 - F_C)^2 \right\}^{\frac{1}{2}} \]

where \(C_n\) is the central curvature, \(\frac{\partial^2 \rho}{\partial x^2}\), at the center of the \(n\)th atom. Similar formulas apply for the \(y\) and \(z\) coordinates. If the structure does not have a center of symmetry, each of (32) must be doubled.

The standard deviation of a bond length between two atoms is given by

\[ \sigma^2(d_{12}) = \sigma_1^2 + \sigma_2^2 \]

where \(\sigma_1\) and \(\sigma_2\) are the standard deviations of the atomic positions in the direction of the bond. Differences in equivalent bond distances under different measurements are not significant unless they differ by at least about 2 to 3 times the standard deviation.

The standard deviation of an angle \(\theta\) between two bonds \(d_{12}\) and \(d_{23}\) is given by

\[ \sigma^2(\theta) = \frac{\sigma_1^2}{d_{12}} + \frac{\sigma_3^2}{d_{23}^2} + \sigma_2^2 \left( \frac{1}{d_{12}^2} - \frac{2\cos\theta}{d_{12}d_{23}} + \frac{1}{d_{23}^2} \right) \]
PART II

THE CRYSTAL AND MOLECULAR STRUCTURES OF
CYANODIMETHYLARSINE,
DIIDOMETHYLARSINE, AND
"CACODYL DISULPHIDE", DIMETHYLARSINO DIMETHYLTDITHIOARSINATE
I. CYANODIMETHYLARSINE

A. Introduction

Previous works in this laboratory have resulted in the determination of structures of some halogenodiphenylarsines. The C-As-C angles are 105°, considerably larger than the value assumed in an electron-diffraction investigation of the corresponding dimethyl derivatives, so that it was felt that it would be useful to investigate the structure of a compound of this latter type. The halogenodimethylarsines are all liquids at room temperature, but cyanodimethylarsine (cacodyl cyanide), whose structure is described here, is a solid, m.p. 30°. This is in itself an interesting fact, and suggests that there might be some type of intermolecular interaction.

B. Experimental

Colorless needle-shaped crystals of cyanodimethylarsine were prepared by the reaction of iododimethylarsine with silver cyanide in vacuo, the desired product being purified by sublimation. The compound is very volatile and extremely toxic, and crystals were sealed in thin-walled Lindemann-glass capillaries. The unit cell dimensions and space group were determined from rotation and oscillation photographs about the a-axis and 0kL and lkL Weissenberg and h0L and hko precession films.

Crystal Data

Cyanodimethylarsine, (CH₃)₂AsCN; M, 131.0; m.p. 30°C

Triclinic, a = 6.31, b = 8.02, c = 6.27 Å, \( \alpha = 110°00', \beta = 119°45', \gamma = 81°47' \)

Volume of the unit cell = 258.7 Å³

\( D_{\text{calc}} \) (with \( Z = 2 \)) = 1.68 g cm⁻³

Absorption coefficients for X rays, \( \lambda = 1.542 \) Å, \( \mu = 76.8 \) cm⁻¹;
\( \lambda = 0.7107 \) Å, \( \mu = 67.2 \) cm⁻¹.
F(000) = 128

Space group is P1 or P1. Analysis has proceeded satisfactorily on the basis of P1.

Due to the extreme volatility and toxicity of the crystals, no effort was made to measure the density, but since it seemed unlikely that it was as low as 0.8 or as high as 3.4 g cm\(^{-3}\), Z = 2 was assumed. This was confirmed by the subsequent analysis of the structure.

The intensities of the 0kl reflections were recorded on Weissenberg films (Cu Kα) and of the h0l and hk0 reflections on precession films (Mo Kα). The estimates were made visually, and the structure amplitudes derived as usual, the absolute scale being established later by correlation with the calculated structure factors. The crystal used had cross section 0.1 \(\times\) 0.1 mm and no absorption corrections were considered necessary. Eighty-four independent 0kl, forty-eight h0l, and seventy-seven hk0 reflections were observed.

C. Structure Analysis

The position of the arsenic atom was determined from the three Patterson projections, and the other atoms were then located on electron-density maps computed with signs based on the arsenic contributions alone. Structure factors were then calculated for all three zones, using the scattering factors of Berghuis et al. (24) with an overall temperature factor B = 5.1 \(\AA^2\), determined from plots of ln (|Fc|/|Fo|) against \(\sin^2 \theta/\lambda^2\). Refinement was completed by computing successive Fourier and difference syntheses. The final electron-density projections are shown in Figs. 7 and 8, and measured and calculated structure factors are listed in (25). The discrepancy factors for the observed planes are: R(0kl) = 14.7%; R(h0l) = 17.3%; R(hk0) = 14.0%.
Fig. 7. Electron-density projection along the c-axis, with contours at intervals of 2 electrons about C and N, and at intervals of 5, 10, 20, 30 about As, and projection of the structure along c.
Fig. 8. Electron-density projections along the a- and b-axes.
Coordinates and Molecular Dimensions

The final positional parameters of the As, C, and N atoms are listed in Table II, \( x, y, \) and \( z \) being referred to the triclinic crystal axes and expressed as fractions of the unit cell edges, and \( x', y', z' \) coordinates in Å units referred to orthogonal axes \( a' = a \sin \gamma, b, \) and \( c' \) (normal to \( a' \) and \( b \)).

The bond distances and valency angles in the molecule are shown in Fig. 9.

The standard deviations of the atomic positions, calculated from Cruickshank's (23) formulae, are \( \sigma(x) = \sigma(y) = \sigma(z) = 0.005 \) Å for As and 0.047 Å for C and N. These values are probably optimistic, especially for the C and N atoms, because two of the projections are poorly resolved.

D. Discussion

On the basis of the estimated standard deviations, the observed differences between As-CH\(_3\) and As-CN bond lengths are not significant. The mean value of the bond lengths and angles, with standard deviations, are:

- \( \text{As-C} = 1.96 \pm 0.03 \) Å
- \( \text{C-N} = 1.16 \pm 0.07 \) Å
- \( \angle \text{Me-As-Me} = 105 \pm 2^\circ \)
- \( \angle \text{Me-As-CN} = 91 \pm 1.5^\circ \)
- \( \angle \text{As-C-N} = 180 \pm 4^\circ \)
Fig. 9. Measured bond lengths and valency angles.
The valency angles differ somewhat from those reported for the halogenodimethylarsines, which were determined by electron diffraction (26). The Me-As-Me angle in these halides was assumed to be 96°, the same as in trimethylarsine (27), and this is significantly smaller than the Me-As-Me angle of 105° now measured for the cyano derivative, which is similar to the angle in bromodiphenylarsine, 105° (28), and in chlorodiphenylarsine, 105° (29). The Me-As-CN angles are slightly less than those reported for Me₂AsBr (96°). The As-C and C=N distances are normal (30).

The shortest intermolecular contact is an As--N separation of 3.18 Å (Fig.10); this is considerably less than the sum of the van der Waals radii (3.5 Å) and suggests a charge-transfer bond involving donation of nitrogen lone pair electrons to the vacant arsenic 4d orbitals. The C=N...As angle is 168°, so that the nitrogen lone pair, which is expected to be collinear with the C=N bond, is directed towards the arsenic atom. Figure 10 shows that, in addition, the arsenic lone pair (assuming that it makes approximately equal angles with the As-C bonds) is directed away from the nitrogen atom, so that the intermolecular bonding would be little hindered by lone-pair repulsions. Cyanodimethylarsine is a solid at room temperature, probably as a result of this intermolecular bonding, while the corresponding halogen derivatives are liquids. The cyanide melting point is 65° higher than that of the iodide, while in other series such as methyl iodide and cyanide and the corresponding ethyl, propyl, and phenyl derivatives this difference is never greater than 20°. Comparison of these differences with the increases in melting point which usually result from hydrogen bonding suggests that the energy of the charge-transfer bonds is only of the order of one or two kilocalories per mole.

The donor properties of the tripositive state of the Group Vb elements diminish with increase in atomic number, and are pronounced only for nitrogen;
Fig. 10. Projection of the structure along [001], showing the shorter intermolecular contacts.
the other members of the group can act as electron acceptors. For arsenic
this is illustrated by the formation of complexes between AsCl$_3$ and various
aromatic amines, which, although their structures have not been determined,
probably involve intermolecular bonds similar to those in cyanodimethylarsine;
and by the formation of complex ions such as AsCl$_4^-$. All the other intermolecular distances in cyanodimethylarsine correspond
to normal van der Waals interactions.

II. DIIODOMETHYLARSINE

A. Introduction

As part of a series of investigations of compounds containing arsenic,
the crystal and molecular structure of diiodomethylarsine has been determined;
it is one of the few simple arsenic derivatives which are solid at room
temperature.

B. Experimental

Crystals of diiodomethylarsine, which are yellow-orange, are volatile and
melt at about room temperature. For recording the X-ray data, crystals were
sealed in capillaries and cooled by a stream of nitrogen which was first
passed through a coil immersed in an ice-bath. The unit-cell dimensions and
space group were determined from various rotation, oscillation, Weissenberg
(Cu Kα) and precession (Mo Kα) films.

Crystal data (at 5-10°C; λ(Cu Kα) = 1.5418 Å, λ(Mo Kα) = 0.7107 Å).

Diiodomethylarsine, CH$_3$AsI$_2$; M, 343.85; m.p. 26°C.

Monoclinic, a = 14.45, b = 4.60, c = 19.97 Å, β = 114°20' .

U = 1209.5 Å$^3$.

D$_x$ (with Z=8) = 3.8 g.cm$^{-3}$.

Absorption coefficient, μ(Cu Kα) = 939 cm$^{-1}$.

F(000) = 1184.
Absent reflections: \( hkl \) when \((h+k)\) is odd, \( h0l \) when \( l \) is odd.

Space group is \( \text{Cc} \) or \( \text{C}2/c \). Analysis has proceeded satisfactorily in \( \text{C}2/c \).

No suitable flotation medium was available for measuring the density; the density of the liquid, measured at room temperature by means of a density bottle, was \( 3.1 \, \text{g.cm}^{-3} \), and, since it seemed likely that the solid at slightly reduced temperatures would have a higher density, \( Z=8 \) was assumed. This was confirmed by the structure analysis.

Intensity data for the \( h0l \) and \( hll \) reflections were recorded on Weissenberg films and estimated visually, and the structure amplitudes were derived as usual, the absolute scale being established later by correlation with the calculated structure factors. The crystal used was a needle, elongated along \( b \), with a rectangular cross-section \( 0.4 \times 0.13 \, \text{mm} \), the (001) face being developed. The films were textbook examples (31) of severe absorption effects, and corrections were applied (32). These absorption correction factors applied to the intensities varied from 1 to about 150, and since they are approximate, the accuracy of the measured structure amplitudes is probably rather limited. 140 independent \( h0l \) reflections (77\% of the possible) and 213 \( hll \) reflections (62\%) were observed.

C. Structure Analysis

The \( x \)- and \( z \)-parameters were determined from \( h0l \) Patterson and electron-density projections, and the \( y \)-coordinates by some trials with \( 0k0 \) and \( hll \) reflections. Structure factors were calculated by means of the scattering factors of Berghuis, Haanappel, Potters, Loopstra, MacGillavry & Veenendaal (24) for \( \text{As} \) and \( \text{C} \), and of Thomas & Umeda (33) for \( \text{I} \), and with an overall temperature factor \( B=5.1 \, \text{Å}^2 \) determined from a plot of \( \ln (|F_C|/|F_0|) \) against \( \sin^2 \theta/\lambda^2 \). Refinement proceeded by \( h0l \) \( (F_0-F_C) \) syntheses and \( hll \) cosine and sine difference generalized projections, refining all parameters simultaneously (34). Refinement was complete after four cycles; measured and calculated structure
factors are listed in (35), the final R values for the observed reflections being 0.18 for h0l and 0.22 for h1l. The final electron-density projection is shown in Fig.11, and the h1l cosine and sine generalized projections in Fig.12.

The final positional parameters are listed in Table III, the standard deviations (23) are $\sigma(x)=\sigma(y)=\sigma(z)=0.006 \, \text{Å}$ for I, 0.010 \, Å for As, and 0.07 \, Å for C. The bond lengths and valency angles are shown in Fig.13, and the shorter intermolecular contacts, all of which correspond to van der Waals interactions, in Fig.14.

Table III.

<table>
<thead>
<tr>
<th>Atom</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>0.2045</td>
<td>0.3804</td>
<td>0.1501</td>
</tr>
<tr>
<td>I1</td>
<td>0.0590</td>
<td>0.0436</td>
<td>0.1384</td>
</tr>
<tr>
<td>I2</td>
<td>0.3475</td>
<td>0.0349</td>
<td>0.1615</td>
</tr>
<tr>
<td>C</td>
<td>0.156</td>
<td>0.380</td>
<td>0.037</td>
</tr>
</tbody>
</table>

D. Discussion

The structure analysis has utilized a large part of the observable three-dimensional data; the rest of the available reflections were not used since any inaccuracies in the final results (as reflected in the rather high R values) are due not to a lack of data (353 reflections have been used to determine twelve positional parameters), but to the poor quality of the intensity measurements as a result of the absorption effects. The standard deviations of the atomic positions indicate that the carbon atom has been located rather imprecisely in the presence of the heavier I and As atoms.

The arsenic atom has the usual pyramidal configuration. The As-I distance (2.54±0.01 Å) is similar to corresponding lengths in As I$_3$ (2.52 Å), Me$_2$AsI (2.54 Å), and Ph$_2$AsI (2.53 Å) (30,36). The I-As-I angle (104°±0.4°)
Fig. 11. Electron-density projection along the b-axis; contours at intervals of 5, 10, 20, 30... eÅ⁻² at As and I atoms, and 3, 4, 5 eÅ⁻² at the C atom.
Fig. 12. Cosine and sine hkl generalized projections.
Fig. 13. Bond lengths and valency angles in Me As I₂.
Fig. 14. Projection of the structure onto (010), showing the shorter intermolecular contacts.
is a little larger than the angles in $\text{AsI}_3$. Bonds and angles involving the carbon atom have been determined less precisely, and do not differ significantly from normal values.

III. "CACODYL DISULPHIDE", DIMETHYLARSINO DIMETHYLDITHIOARSINATE

A. Introduction

"CACODYL DISULPHIDE," $\left[(\text{CH}_3\right)_2\text{As}\right]_2\text{S}_2$, was first prepared by Bunsen (37) by passing hydrogen sulphide through a concentrated alcoholic solution of cacodylic acid:

$$2(\text{CH}_3)_2\text{As}(\text{OH})+3\text{H}_2\text{S} \rightarrow \left[(\text{CH}_3)_2\text{As}\right]_2\text{S}_2+\text{S}+4\text{H}_2\text{O}$$

and by reaction of cacodyl sulphide with sulphur. Dehn and Wilcox (38) obtained cacodyl sulphide from dimethylarsine and sulphur, but found that the disulphide was formed when the arsine reacted with a greater amount of sulphur:

$$2(\text{CH}_3)_2\text{AsH}+3\text{S} \rightarrow \left[(\text{CH}_3)_2\text{As}\right]_2\text{S}_2+\text{H}_2\text{S}$$

It appears to have been generally considered that the compound is a true disulphide, and it is listed as a cacodyl derivative (39) so that structure (I) has been assumed. However, its preparation from cacodyl sulphide and an excess of sulphur suggests that the disulphide contains pentavalent arsenic, since trivalent arsines usually react with sulphur to give pentavalent derivatives. This is substantiated by Bunsen's finding that the sulphide reacts with metal halides to yield salts of dimethyldithioarsinic acid, $(\text{CH}_3)_2\text{As(S).SH}$. Formulations such as (II) or (III) are therefore possible.

![Diagram](image_url)
The present analysis was undertaken to determine the structure of "cacodyl disulphide" by X-ray diffraction methods, and the results indicate that the compound is dimethylarsino dimethyldithioarsinate (III).

B. Experimental

"Cacodyl disulphide" was prepared from cacodylic acid and hydrogen sulphide (37) and by reaction of dimethylarsine with an excess of sulphur (38). X-ray powder photographs indicated that the two samples were identical (40). Crystals from ether are needles elongated along the a-axis. The unit-cell dimensions and space group were determined from rotation, Weissenberg, and precession photographs, and on the G.E. Spectrogoniometer. Efforts to measure the density by flotation were unsuccessful, since the crystal reacted with aqueous solutions and dissolved in organic liquids of suitable density; however, it appeared to be between about 1.5 and 2.0 g.cm\(^{-3}\) and this sufficed to establish the number of molecules in the unit cell as two.

Crystal Data (Cu-K\(\alpha\)=1.5418 Å, Mo-K\(\alpha\)=0.7107 Å).

Dimethylarsino dimethyldithioarsinate, \(\text{C}_4\text{H}_{12}\text{As}_2\text{S}_2\); \(\text{M}, 274.1\). Triclinic, \(a = 6.34\pm0.01\) Å, \(b = 7.11\pm0.01\) Å, \(c = 11.35\pm0.02\) Å, \(\alpha = 100^\circ 14'\pm 5'\), \(\beta = 95^\circ 46'\pm 5'\), \(\gamma = 89^\circ 55'\pm 5'\). Volume of the unit cell = 500.4 Å\(^3\). \(D_x\) (with \(Z=2\)) = 1.820 g.cm\(^{-3}\). Absorption coefficient for X-rays, \(\lambda = 0.7107\) Å, \(\mu = 74\) cm\(^{-1}\). \(F(000) = 268\). No absent reflections; space group \(\text{P}1\) or \(\text{P}1\); \(\text{P}1\) from structure analysis.

The intensities of the reflections were measured on a General Electric XRD-5 spectrogoniometer with single crystal orieneter, with a scintillation counter and Mo-K\(\alpha\) radiation, an approximately monochromatic beam being obtained by use of a zirconium filter and a pulse-height analyser. The moving crystal-moving counter technique (41) was used. All the reflections in the range \(0 \leq 2\theta \leq 40^\circ\) (corresponding to a minimum interplanar spacing \(d = 1.04\) Å) were examined and 805 were measurable, 87% of the total number of reflections
in this range. All the intensities were corrected for background, which was found to be a function of θ only. The crystal used (mounted with a* parallel to the φ-axis of the goniostat) was small and had cross-section 0.1 × 0.1 mm. perpendicular to a and length 0.5 mm. parallel to a; absorption was fairly low and no correction was applied. Lorentz and polarization corrections were made and the structure amplitudes derived.

C. Structure Analysis

The largest peaks in the $0kl$ and $h0l$ Patterson projections could be interpreted in terms of two pairs of arsenic atoms related by a centre of symmetry. The co-ordinates derived for the arsenic atoms immediately eliminated structure (II), since no two arsenic atoms were close enough together to be directly bonded. Four-fold vector convergence (minimum) functions were then derived by placing the origins of the Patterson functions in turn at the four arsenic positions (II). The $0kl$ map showed only three peaks; two of these were situated at arsenic, one being considerably higher than the other and probably corresponding to overlapping arsenic and sulphur atoms; the third peak could be ascribed to a single sulphur atom. The $h0l$ map also had three significant peaks, and a smaller fourth peak adjacent to one of the arsenic atoms, but in such a position that it could not be involved in bonding with the second arsenic; the peaks at the arsenic positions were of equal height, and the third peak was about half this height. An effort was made to interpret both projections in terms of structure (I), and positions were derived for 2 arsenic and 2 sulphur atoms. Structure factors were calculated for the $0kl$ and $h0l$ reflections by using standard scattering factors (5), with $B = 4.5 \, \AA^2$ for all the atoms; the discrepancy factors were $R(0kl) = 33.4\%$, $R(h0l) = 38.9\%$. A similar calculation with arsenic atoms only gave $R(0kl) = 31.5\%$, $R(h0l) = 41.5\%$. 
Fourier series were summed for both projections, by using as coefficients the measured structure amplitudes with signs based on the arsenic atoms only. The resulting electron-density maps showed the same features as the vector-convergence functions, including the fourth peak in the b-axis projection noted above, with much improved resolution, and there were in addition other smaller peaks. It was clear at this stage that neither structure (I) nor structure (II) could be fitted to these maps, but that the electron-density distributions corresponded to structure (III). Co-ordinates could be derived with confidence for 2 arsenic, 2 sulphur, and 2 carbon atoms; there were peaks which apparently corresponded to two other carbon atoms, but as there was some ambiguity they were omitted at this stage. Structure factors were recalculated, and the R values were reduced to 19.3% for Okl and 25.8% for h0l reflections.

A second set of electron-density projections was then computed, and these maps showed good resolution of all the atoms. Inclusion of the two other carbon atoms in the structure factor calculations gave \( R(Ok\ell) = 13.2\% \), \( R(h0\ell) = 23.4\% \). Examination of the structure-factor agreement suggested that the temperature factor was too high. \( B \) was therefore reduced to 3.0 \( \AA^2 \), and structure factors were calculated for the complete three-dimensional data; \( R(hk\ell) \), for the observed reflections, was 18.2%.

Refinement

The positional and isotropic thermal parameters of the arsenic, sulphur, and carbon atoms, together with an overall scale factor were then refined by least squares, with a programme described in part III of this thesis. The function minimized was \( \sum w(F_O - F_c)^2 \), with \( \sqrt{w} = F_O/35 \) when \( F_O < 35 \), and \( \sqrt{w} = 35/F_O \) when \( F_O \geq 35 \). Refinement proceeded smoothly and was complete in four cycles (Table VII).

The final measured and calculated structure factors, calculated from the parameters output from the fourth least-squares cycle \( (R = 9.1\% \) for the 805 observed reflections), are listed in (42). A three-dimensional Fourier series
was summed, and superimposed sections of the resulting electron-density
distribution taken through the atomic centres are shown in Fig.15.

Co-ordinates and Molecular Dimensions

The final positional and thermal parameters are listed, with their
standard deviations, in Table IV: x, y, and z are fractional co-ordinates
referring to the triclinic cell axes. No effort was made to determine positions
of the hydrogen atoms. The bond distances and valency angles are given, with
their standard deviations, in Table V. The shorter intermolecular contacts
are listed in Table VI, and the packing of the molecules is shown in Fig.16.

D. Discussion

The analysis has established that the compound called "cacodyl disulphide"
is dimethylarsino dimethyldithioarsinate (III) (Figs. 15 and 17). It therefore
contains one trivalent and one pentavalent arsenic atom. Other compounds are
known with two arsenic atoms in the same molecule in different valency states,
and in fact once one of the arsenic atoms in compounds of the type \( R_2\text{As}_2 \cdot \text{X} \cdot \text{AsR}_2 \)
has been oxidized (say, by forming a methiodide), it is often quite difficult
to oxidize the other (43). Formulation (III) is in accord with the preparation
of the compound from cacodyl sulphide and sulphur (38) and with the formation
of salts of dimethyldithioarsinic acid by reaction with metal halides (37).

In dimethylarsino dimethyldithioarsinate the configuration around the
trivalent arsenic (As\(_2\)) is trigonal pyramidal and that around the pentavalent
arsenic (As\(_1\)) is tetrahedral. The angles at the pentavalent arsenic atom
differ significantly from the regular tetrahedral value; those angles
involving the doubly bonded sulphur, \( S_3 \) (S=As\(^V\)-S\(_4\), S=As\(^V\)-C\(_5\), S=As\(^V\)-C\(_6\)) are
all larger than 109° 28' (Table V), while the angles not involving \( S_3 \) are
smaller than the regular tetrahedral angle and have values which are not too
much larger than the angles at the trigonal As\(^{III}\) atom. There are, in addition,
deviations from the symmetry which might pertain in the two parts of the molecule.
Fig. 15. Superimposed sections of the final three-dimensional electron-density distribution, through the atomic centres perpendicular to $b$; contours at intervals of 2, 3, 4, 5 eÅ$^{-3}$ for C; 2, 4, 6... for S; and 5, 10, 15... for As. A perspective drawing of the molecule is also shown.
Table IV

Final positional parameters (fractional) with standard deviations (Å), and temperature factors and standard deviations (Å²).

<table>
<thead>
<tr>
<th>Atom</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>σ(x)</th>
<th>σ(y)</th>
<th>σ(z)</th>
<th>B(Å²)</th>
<th>σ(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>As₁</td>
<td>0.3161</td>
<td>0.2240</td>
<td>0.3668</td>
<td>0.0024</td>
<td>0.0024</td>
<td>0.0025</td>
<td>3.22</td>
<td>0.05</td>
</tr>
<tr>
<td>As₂</td>
<td>0.3700</td>
<td>-0.0619</td>
<td>0.1033</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0026</td>
<td>3.76</td>
<td>0.05</td>
</tr>
<tr>
<td>S₃</td>
<td>0.6419</td>
<td>0.2461</td>
<td>0.3645</td>
<td>0.0074</td>
<td>0.0073</td>
<td>0.0074</td>
<td>5.09</td>
<td>0.16</td>
</tr>
<tr>
<td>S₄</td>
<td>0.1789</td>
<td>-0.0498</td>
<td>0.2641</td>
<td>0.0069</td>
<td>0.0065</td>
<td>0.0068</td>
<td>4.18</td>
<td>0.14</td>
</tr>
<tr>
<td>C₅</td>
<td>0.2180</td>
<td>0.2229</td>
<td>0.5238</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>4.63</td>
<td>0.56</td>
</tr>
<tr>
<td>C₆</td>
<td>0.1664</td>
<td>0.4265</td>
<td>0.2969</td>
<td>0.026</td>
<td>0.026</td>
<td>0.026</td>
<td>4.29</td>
<td>0.54</td>
</tr>
<tr>
<td>C₇</td>
<td>0.5950</td>
<td>-0.2441</td>
<td>0.1467</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
<td>5.37</td>
<td>0.63</td>
</tr>
<tr>
<td>C₈</td>
<td>0.1786</td>
<td>-0.2432</td>
<td>-0.0100</td>
<td>0.026</td>
<td>0.026</td>
<td>0.027</td>
<td>5.09</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table V

Bond distances (Å) and valency angles, and standard deviations

<table>
<thead>
<tr>
<th>Bond</th>
<th>As₁-S₃</th>
<th>As₁-C₅</th>
<th>As₁-S₄</th>
<th>As₁-C₆</th>
<th>As₁-C₇</th>
<th>S₄-As₁-S₃</th>
<th>S₄-As₁-C₅</th>
<th>S₄-As₁-C₆</th>
<th>S₄-As₁-C₇</th>
<th>S₄-As₁-C₈</th>
<th>S₄-As₁-C₉</th>
<th>S₄-As₁-C₁₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>As₁-S₃</td>
<td>2.075±0.007</td>
<td>1.947±0.025</td>
<td>113.3±0.3°</td>
<td>101.0±0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As₁-C₅</td>
<td>2.214±0.007</td>
<td>1.963±0.025</td>
<td>106.2±0.7°</td>
<td>99.1±0.7°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As₁-C₆</td>
<td>2.279±0.007</td>
<td>2.009±0.027</td>
<td>115.7±0.7°</td>
<td>96.2±0.7°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As₁-C₇</td>
<td>1.972±0.026</td>
<td>1.972±0.026</td>
<td>111.9±0.7°</td>
<td>99.2±1.1°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

 Separately: S₄-As₃-110.1° ≠ S₄-As₃-110.0°, ≠ S₄-As₅-111.9°. These differences are probably a result of minor steric interferences in the molecule; Figs. 15 and 17 show that C₆ is much closer to As₂-110.0° than is C₅, and C₇ is much closer to S₃ than is C₈. Slight repulsions due to these closer approaches would tend to increase S₄-As₅-110.0° and S₄-As₃-110.1°.
Table VI

Shorter intermolecular distances

(All distances ≤ 5 Å between molecule 1, at x, y, z, and neighbouring molecules were calculated; only the crystallographically independent distances ≤ 4 Å are listed.)

<table>
<thead>
<tr>
<th>Atom (in molecule 1) to Atom in Molecule</th>
<th>d(Å)</th>
<th>Atom (in molecule 1) to Atom in Molecule</th>
<th>d(Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>As₂</td>
<td>4</td>
<td>S₄</td>
<td>5</td>
</tr>
<tr>
<td>As₂</td>
<td>4</td>
<td>C₅</td>
<td>5</td>
</tr>
<tr>
<td>As₂</td>
<td>4</td>
<td>C₂</td>
<td>2</td>
</tr>
<tr>
<td>S₃</td>
<td>2</td>
<td>C₆</td>
<td>7</td>
</tr>
<tr>
<td>S₃</td>
<td>2</td>
<td>C₇</td>
<td>3</td>
</tr>
<tr>
<td>S₃</td>
<td>6</td>
<td>C₇</td>
<td>3</td>
</tr>
<tr>
<td>S₃</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Molecule 1 at x, y, z
2 at 1 + x, y, z
3 at -x, -y, -z
4 at 1 - x, -y, -z
5 at -x, -y, 1 - z

Molecule 6 at 1 - x, -y, 1 - z
7 at 1 - x, 1 + y, z
8 at 1 + x, 1 + y, z
9 at - x, 1 - y, 1 - z
10 at 1 - x, 1 - y, 1 - z

Table VII

Progress of refinement

Coordinates from

<table>
<thead>
<tr>
<th>Patterson projections (2As + 2S)</th>
<th>R₀kl</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.4</td>
<td>38.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patterson projections (2As)</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.5</td>
<td>41.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st Fourier projections (2As + 2S + 2C)</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.3</td>
<td>25.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd Fourier projections (2As + 2S + 4C)</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.2</td>
<td>23.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd Fourier projections + adjusted B's</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.2</td>
<td>2818</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1st least squares</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1</td>
<td>1062</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd least squares</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.3</td>
<td>824</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3rd least squares</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.2</td>
<td>714</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4th least squares</th>
<th>R₀₀₀l</th>
<th>R₀₀₀l</th>
<th>ΣwΔF²</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>700</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 16. Projection of the structure along [100], illustrating the packing of the molecules. The short intermolecular contact is shown as a broken line, and the lone pairs are shown schematically.
Fig. 17. View of the molecule along the $S_4$-$As_1$ bond.
The angle \( \text{As}^{\text{III}} \text{-S} \text{-As}^{\text{V}} \) is 96.5°, only a little larger than the angle in \( \text{H}_2\text{S} \) (93°) (30). The corresponding angles in arsenic trisulphide and arsenic sulphide (realgar) are 100° ± 2° and about 102°, respectively (30), and the S-As\(^{\text{III}}\)-S angle in arsenic trisulphide is 114° ± 2°, very similar to the S-As\(^{\text{V}}\)=S angle in dimethylarsino-dimethyldithioarsinate (113.5° ± 0.3°). However, this agreement is probably fortuitous since the molecules are so different. There appear to be no results available with which the other angles can be strictly compared, data for pentavalent arsenic in particular being very sparse. For trivalent arsenic in compounds of the types \( \text{R}_2\text{AsX} \) (\( \text{R} = \text{alkyl or aryl}, \text{X} = \text{halogen} \)) and \( \text{R}_3\text{As} \) (\( \text{R} = \text{alkyl or aryl} \)) some measurements are available. The C-As-X and C-As-C angles are each about 96-98° (within rather wide limits of error) in the halogenodimethylarsines (30); 95-96° ± 1° and 105° ± 2°, respectively, in the halogenodiphenylarsines (29); and 91° ± 1.2° and 105° ± 2°, respectively, in cyanodimethylarsine (cacodyl cyanide) (25). The C-As-C angles are 96° ± 5° in trimethylarsine (30), 100° ± 3.5° in perfluorotrimethylarsine (30), and 102° ± 2° in both tri-p-tolylarsine (44) and tri-p-xylylarsine (45). The angles at As\(^{\text{III}}_2\) in the present analysis (96.2°, 99.1°, and 99.2°) are in the same range as the corresponding angles in these other molecules; no stricter comparison is useful in any case since the bond angles are so readily distorted by slight steric strains in the various molecules.

The As-S single bond distances are 2.28 ± 0.007 Å for As\(^{\text{III}}\)-S and 2.21 ± 0.007 Å for As\(^{\text{V}}\)-S. The difference between these values is highly significant, and suggests a slightly smaller covalent radius (by 0.07 Å) for As\(^{\text{V}}\). The As\(^{\text{V}}_1\)-S\(_3\) distance (2.07 Å) indicates that this is a double bond, the difference between As\(^{\text{V}}\)-S\(_4\) and As\(^{\text{V}}\)=S\(_3\) being 0.14 Å, about the same as the difference between single- and double-bond covalent radii for sulphur (0.10 Å) (46). The only As-S distances given in the literature are those in realgar (2.21-2.23 ± 0.02 Å) and in arsenic trisulphide (2.25 ± 0.02 Å) (30).
The arsenic-carbon distances average $1.990 \pm 0.019 \, \text{Å}$ for $\text{As}^{\text{III}} - \text{C}$ and $1.955 \pm 0.018 \, \text{Å}$ for $\text{As}^{\text{V}} - \text{C}$; as with the arsenic-sulphur distances, the bonds involving pentavalent arsenic are slightly shorter ($0.035 \, \text{Å}$), but on the basis of the estimated standard deviations the difference cannot be claimed to be significant in this case. The $\text{As} - \text{C}$ distances are very similar to the lengths of corresponding bonds in other arsenic derivatives (mean value about $1.98 \, \text{Å}$).

The thermal parameters of the atoms vary in a manner (Table IV) which might be expected from the molecular structure. $\text{As}_2$, which is bonded to only three other atoms, is able to execute slightly larger vibrations ($B = 3.76 \, \text{Å}^2$) than $\text{As}_4$ ($B = 3.22 \, \text{Å}^2$), which is bonded to four atoms. Similarly, $\text{S}_3$ vibrates more than $\text{S}_4$, and $\text{C}_7$ and $\text{C}_8$, which are bonded to $\text{As}_2$, vibrate more than $\text{C}_5$ and $\text{C}_6$ (we might, however, have expected that $\text{S}_3$ would vibrate less than $\text{C}_5$ and $\text{C}_6$).

The shortest intermolecular contact is an $\text{As}^{\text{III}} - \text{As}^{\text{III}}$ separation of $3.24 \, \text{Å}$ (Table VI). This is considerably less than the sum of the van der Waals radii ($4.0 \, \text{Å}$) (46), and suggests charge-transfer bonding involving donation of lone-pair electrons (probably in $\text{sp}^3$-hybrid orbitals) on each $\text{As}^{\text{III}}$ to vacant $4d$-orbitals on the other arsenic atom. The direction of the short intermolecular contact is shown in Fig. 16, which indicates that the molecules are associated in pairs and that the lone pairs (assumed to make approximately equal angles with the bonds at each arsenic atom) are so directed that there is little steric interference between them, and each can therefore overlap with a $4d$-orbital of the other atom. Similar interactions have been observed in other structures, but usually with the arsenic acting only as an electron-acceptor, the donor atom being nitrogen. In cyanodimethylarsine (25) an intermolecular $\text{As} - \text{N}$ separation of $3.18 \, \text{Å}$ is observed (sum of van der Waals radii (46) $3.5 \, \text{Å}$), and in arsenic tricyanide (47) a similar but even shorter contact ($2.85 \, \text{Å}$). Both these compounds have relatively high melting points and low volatilities in comparison with the corresponding halides, and
these facts suggest that the short intermolecular distances represent charge-transfer bonds with energies of 1 - 2 kcal. mole\(^{-1}\) in excess of the usual van der Waals forces. The situation in the present case is somewhat similar, except that each trivalent arsenic atom acts as both an electron-donor and an electron-acceptor.

All the other intermolecular distances (Table VI) correspond to normal van der Waals interactions.
PART III

THE CRYSTAL AND MOLECULAR STRUCTURES OF
CLEAVAMINE METHIODIDE
AND 5-IODO-2' -DEOXYURIDINE,
AND CRYSTAL DATA FOR
SODIUM THYMIDYL-(5' → 3')-THYMIDYLATE-(5')
I. CLEAVAMINE METHIODIDE

A. Introduction

The study of the action of acidic reagents on some of the simpler alkaloids isolated from Vinca rosea Linn provides important information about the effect of such reagents on the biologically important alkaloids vincaleukoblastine and leurosine. The reaction of catharanthine (IV) with concentrated hydrochloric acid yields two products, one of which is simply desmethoxy-carbonylcatharanthine. The second product, cleavamine, C_{19}H_{24}N_{2}, involves a more drastic rearrangement; analytical, ultraviolet, infrared and nuclear magnetic resonance data suggest a tetracyclic ring skeleton for cleavamine, with retention of the indole chromophore and ethylenic double bond, and loss of the ester function of catharanthine (48). Consideration of the evidence thus far suggested a structural relationship between cleavamine and the known alkaloid quebrachamine (VI); this relationship was strengthened by a mass spectral comparison of the two (49).

The following describes an X-ray analysis of cleavamine methiodide, which establishes the structure of the alkaloid, including the absolute configuration,

![IV](image)

![V](image)

as (V). This formulation is in accord with all the chemical evidence.
B. Experimental

Crystals of cleavamine methiodide are plates elongated along c with (010) developed and smaller (100) faces. The density was measured by flotation in aqueous potassium iodide, and the unit-cell dimensions and space group were determined from various rotation, Weissenberg and precession photographs and on the G.E. spectrogoniometer.

**Crystal data** \( (\lambda, \text{Cu K}_\text{K}=1.5418 \text{ Å}, \lambda, \text{Mo K}_\text{K}=0.7107 \text{ Å}) \).

Cleavamine methiodide, \( C_{20}H_{27}N_2I \), mol. wt. 422.3, m.p. 244-245°C (dec.).

Orthorhombic, \( a=7.86 \pm 0.02, b=14.86 \pm 0.03, c=16.32 \pm 0.04 \text{ Å} \).

\( U = 1906 \text{ Å}^3 \).

\( D_m = 1.46 \text{ g.cm}^{-3}, \ Z=4, \ D_x = 1.47 \text{ g.cm}^{-3} \).

Absorption coefficient for X-rays, \( \mu = 0.7107 \text{ Å}, \ \mu = 16.9 \text{ cm}^{-1} \).

\( F(000) = 856 \).

Absent spectra: \( h00 \) when \( h \) is odd, \( 0k0 \) when \( k \) is odd, \( 00l \) when \( l \) is odd.

Space group \( P2_12_12_1 \) (\( D_{d}^{2} \)).

A preliminary photographic survey revealed a rapid fall-off in intensity with increasing Bragg angle. The intensities of all reflections with \( 2\theta(Mo K_\alpha) \leq 35^\circ \) (corresponding to an interplanar spacing \( d = 1.18 \text{ Å} \)) were measured on a General Electric XRD-5 Spectrogoniometer with Single-Crystal Orienter, a scintillation counter and Mo K_\alpha radiation being used, an approximately monochromatic beam being obtained by use of a zirconium filter and a pulse height analyser. The moving crystal-moving counter technique \((41)\) was used. 582 reflections in the range \( 0 < 2\theta < 35^\circ \) were observed, representing about 80% of the total number of reflections in this range. For reflections at higher Bragg angles to the limit of the Cu K_\alpha sphere, only those few visible on the preliminary films were examined in the spectrogoniometer, and the intensities (all very weak) recorded. All the intensities were corrected
for background, which was found to be a function of $\theta$ only. Lorentz and polarization factors were applied and the structure amplitudes derived. The crystal used in recording the intensities was mounted with $c$ parallel to the $\phi$ axis, and had dimensions 0.4 mm, 0.12 mm, 0.03 mm parallel to $c$, $a$ and $b$ respectively; absorption was fairly low and no corrections were applied.

C. Structure Analysis

The position of the iodine atom was determined from the [100] and [010] Patterson projections as (0.250, 0.350, 0.140), and structure factors were calculated for all the data for the iodine only, the scattering factor listed by Sagel (50) for uncharged I being used, with $B=4.0$ Å$^2$. The value of $R$, the usual discrepancy factor, was 33.0% for the observed reflections. A three-dimensional Fourier series was summed with phases based on the iodine atom, since $x_I$ was $\frac{1}{4}$, the resulting electron-density distribution had a false mirror plane at $x=\frac{1}{4}$. However it was possible to pick out fused five- and six-membered rings which obviously corresponded to the indole nucleus. Many other atoms were also clearly resolved, but the ambiguities introduced by the false symmetry prevented the deduction of the whole molecular structure at this stage. The nine atoms of the indole group and the two carbon atoms bonded to the five-membered ring were introduced into the structure factor calculations, all with the carbon scattering factor (50) and $B=4.0$ Å$^2$; $R$ was reduced to 23.2%. A second three-dimensional electron-density distribution revealed ten more atoms, introduction of these reduced $R$ to 21.8%, and the final atom was located on a third three-dimensional electron-density distribution. At this stage all the gross features of the molecular structure had been established and $R$ was 21.5%.

Refinement of the structure

Further refinement proceeded by computation of difference electron-density projections on (010) and (001). It was felt that these projections, being centrosymmetrical, were more likely to reveal any small deviations of $x_I$ from $\frac{1}{4}$,
than a synthesis using the three-dimensional data.* The maps did suggest a reduction of $x_I$ to about 0.248, small shifts in the $y$ and $z$ coordinates of the iodine, and an increase in the temperature factor to 5.5 $\text{Å}^2$. At this stage also more realistic scattering factors were introduced. That for $I^-$ was obtained from the curve for uncharged $I$ (5) by comparison with the differences in the values of $X^-$ and $X$ ($X=F, Cl, Br$); it was corrected for anomalous dispersion, according to the expression

$$f(\text{corrected}) = \sqrt{(f'_I + \Delta f'_I)^2 + (\Delta f''_I)^2}$$

using the values $\Delta f'$, $\Delta f''$ given in (5). The nitrogen scattering factor was introduced for $N(2)$ and the carbon curve was used for all the carbon atoms and for $N^+(3)$. These changes reduced $R$ for the three-dimensional data to 15.4%.

All the positional and isotropic thermal parameters, together with an overall scale factor, were then refined by least squares, using a program for the IBM 1620 computer (Addendum). The function minimized was $\sum w(|F_o| - |F_c|)^2$, with $w=40/F_o$ when $F_o < 40$, and $w=40/F_o$ when $F_o > 40$. Five cycles reduced $R$ to 8.4%. The shifts in the fifth cycle were generally less than about one-third of a standard deviation, except for one or two as large as one standard deviation.

An attempt was then made to refine the structure independently by the differential synthesis method, using calculated syntheses to apply 'backshift' corrections to the atomic coordinates (51), and corrections to the thermal parameters. Starting with the final least-squares positions the first differential syntheses cycle increased $R$ to 9.7%, two more cycles reduced this to 8.9%, and a fourth cycle caused a slight increase to 9.1%. The changes in positional parameters from those of the least-squares refinement were generally small, but there were several fairly large differences in bond lengths and valency angles (to be described later). The temperature factors

* I am indebted to Dr. A.W. Hanson for suggesting this procedure.
were generally about the same as or a little larger than those of the least-squares analysis. A final least-squares cycle, using as input the parameters from the fourth differential syntheses, shifted the atoms towards the positions given by the fifth least-squares cycle.

The final measured and calculated structure factors, calculated from the parameters output from the fifth least-squares cycle (R=8.4% for the 591 observed reflections), are listed in (52). A final three-dimensional Fourier series was summed using as coefficients the measured structure amplitudes with calculated phase angles. Superimposed sections of the resulting electron-density distribution taken through the atomic centres are shown in Fig.18. All the peak heights are quite low (Fig.18 and Table VIII), a consequence of the rapid fall-off in intensity of the reflections. This is particularly so for atom C(23) (Fig.18), and this was also the atom which was last to appear in the electron-density distributions, and for which the coordinate differences between least-squares and differential syntheses were largest; however in no other region was there any electron density which might indicate another position for or any disordering of atom C(23).

Coordinates and molecular dimensions

The final positional and thermal parameters are listed in Table VIII. Two sets of parameters are listed, those from the fifth least-squares cycle and those from the third differential syntheses; the least-squares coordinates give a slightly lower discrepancy factor (8.4 as against 8.9%), but it is felt that it is not really possible to decide between them. Despite the reliable counter intensities, it appears that the difficulty of obtaining accurate carbon positions in the presence of the heavier iodine is the limiting factor in the analysis. Also given in Table VIII are the standard deviations of the atomic parameters, calculated from the least-squares residuals (Addendum), and the peak electron densities from the third observed differential synthesis.
Fig. 18. Superimposed sections of the final three-dimensional electron-density distribution, through the atomic centres parallel to (010); contours for the carbon and nitrogen atoms are at intervals of $\frac{1}{2}e\text{Å}^{-3}$ starting at $2e\text{Å}^{-3}$, and for the iodine at intervals of $5e\text{Å}^{-3}$ starting at $5e\text{Å}^{-3}$. Also shown is a perspective drawing of the molecule. Both drawings show the correct absolute configuration, the positive direction of the b-axis being towards the viewer.
Table VIII. Positional parameters (fractional) with standard deviations (Å), thermal parameters and standard deviations (Å²), and peak electron-densities (e.Å⁻³)

<table>
<thead>
<tr>
<th>Atom</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>σ(x)</th>
<th>σ(y)</th>
<th>σ(z)</th>
<th>B(Å²)</th>
<th>σ(B)</th>
<th>5th Least-squares cycle</th>
<th>3rd Differential cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (1)</td>
<td>0.247</td>
<td>0.353</td>
<td>0.137</td>
<td>0.004</td>
<td>0.002</td>
<td>0.002</td>
<td>5.78</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (2)</td>
<td>0.064</td>
<td>0.443</td>
<td>0.309</td>
<td>0.031</td>
<td>0.029</td>
<td>0.029</td>
<td>4.26</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N+3</td>
<td>-0.456</td>
<td>0.532</td>
<td>0.416</td>
<td>0.034</td>
<td>0.034</td>
<td>0.032</td>
<td>3.91</td>
<td>0.77</td>
<td>-0.452</td>
<td>0.533</td>
</tr>
<tr>
<td>C (4)</td>
<td>-0.030</td>
<td>0.509</td>
<td>0.336</td>
<td>0.041</td>
<td>0.040</td>
<td>0.036</td>
<td>3.48</td>
<td>0.92</td>
<td>-0.041</td>
<td>0.508</td>
</tr>
<tr>
<td>C (5)</td>
<td>-0.048</td>
<td>0.523</td>
<td>0.429</td>
<td>0.034</td>
<td>0.031</td>
<td>0.030</td>
<td>3.29</td>
<td>0.75</td>
<td>-0.053</td>
<td>0.523</td>
</tr>
<tr>
<td>C (6)</td>
<td>0.031</td>
<td>0.435</td>
<td>0.457</td>
<td>0.033</td>
<td>0.034</td>
<td>0.032</td>
<td>2.72</td>
<td>0.77</td>
<td>0.035</td>
<td>0.432</td>
</tr>
<tr>
<td>C (7)</td>
<td>0.054</td>
<td>0.400</td>
<td>0.532</td>
<td>0.040</td>
<td>0.039</td>
<td>0.038</td>
<td>4.02</td>
<td>0.93</td>
<td>0.063</td>
<td>0.398</td>
</tr>
<tr>
<td>C (8)</td>
<td>0.158</td>
<td>0.317</td>
<td>0.536</td>
<td>0.037</td>
<td>0.036</td>
<td>0.036</td>
<td>3.20</td>
<td>0.91</td>
<td>0.150</td>
<td>0.319</td>
</tr>
<tr>
<td>C (9)</td>
<td>0.209</td>
<td>0.278</td>
<td>0.467</td>
<td>0.036</td>
<td>0.033</td>
<td>0.033</td>
<td>4.70</td>
<td>0.90</td>
<td>0.210</td>
<td>0.280</td>
</tr>
<tr>
<td>C (10)</td>
<td>0.195</td>
<td>0.308</td>
<td>0.385</td>
<td>0.034</td>
<td>0.037</td>
<td>0.034</td>
<td>3.81</td>
<td>0.98</td>
<td>0.193</td>
<td>0.312</td>
</tr>
<tr>
<td>C (11)</td>
<td>0.092</td>
<td>0.395</td>
<td>0.385</td>
<td>0.039</td>
<td>0.037</td>
<td>0.036</td>
<td>3.80</td>
<td>0.95</td>
<td>0.090</td>
<td>0.394</td>
</tr>
<tr>
<td>C (12)</td>
<td>-0.111</td>
<td>0.587</td>
<td>0.477</td>
<td>0.040</td>
<td>0.039</td>
<td>0.038</td>
<td>3.65</td>
<td>0.92</td>
<td>-0.115</td>
<td>0.588</td>
</tr>
<tr>
<td>C (13)</td>
<td>-0.104</td>
<td>0.581</td>
<td>0.286</td>
<td>0.040</td>
<td>0.038</td>
<td>0.036</td>
<td>2.69</td>
<td>0.95</td>
<td>-0.103</td>
<td>0.584</td>
</tr>
<tr>
<td>C (14)</td>
<td>-0.244</td>
<td>0.514</td>
<td>0.223</td>
<td>0.065</td>
<td>0.034</td>
<td>0.032</td>
<td>4.47</td>
<td>0.86</td>
<td>-0.237</td>
<td>0.548</td>
</tr>
<tr>
<td>C (15)</td>
<td>-0.428</td>
<td>0.532</td>
<td>0.254</td>
<td>0.036</td>
<td>0.034</td>
<td>0.033</td>
<td>3.38</td>
<td>0.84</td>
<td>-0.431</td>
<td>0.533</td>
</tr>
<tr>
<td>C (16)</td>
<td>-0.519</td>
<td>0.620</td>
<td>0.265</td>
<td>0.047</td>
<td>0.041</td>
<td>0.041</td>
<td>5.13</td>
<td>1.11</td>
<td>-0.515</td>
<td>0.624</td>
</tr>
<tr>
<td>C (17)</td>
<td>-0.624</td>
<td>0.643</td>
<td>0.337</td>
<td>0.041</td>
<td>0.042</td>
<td>0.037</td>
<td>3.79</td>
<td>0.95</td>
<td>-0.623</td>
<td>0.644</td>
</tr>
<tr>
<td>C (18)</td>
<td>-0.618</td>
<td>0.581</td>
<td>0.416</td>
<td>0.037</td>
<td>0.034</td>
<td>0.036</td>
<td>2.30</td>
<td>0.81</td>
<td>-0.622</td>
<td>0.579</td>
</tr>
<tr>
<td>C (19)</td>
<td>-0.323</td>
<td>0.608</td>
<td>0.445</td>
<td>0.039</td>
<td>0.043</td>
<td>0.040</td>
<td>5.27</td>
<td>1.17</td>
<td>-0.326</td>
<td>0.605</td>
</tr>
<tr>
<td>C (20)</td>
<td>-0.750</td>
<td>0.727</td>
<td>0.351</td>
<td>0.072</td>
<td>0.035</td>
<td>0.037</td>
<td>7.31</td>
<td>0.90</td>
<td>-0.747</td>
<td>0.730</td>
</tr>
<tr>
<td>C (21)</td>
<td>-0.740</td>
<td>0.778</td>
<td>0.265</td>
<td>0.076</td>
<td>0.043</td>
<td>0.043</td>
<td>8.65</td>
<td>1.19</td>
<td>-0.739</td>
<td>0.776</td>
</tr>
<tr>
<td>C (22)</td>
<td>-0.472</td>
<td>0.468</td>
<td>0.499</td>
<td>0.046</td>
<td>0.042</td>
<td>0.040</td>
<td>5.38</td>
<td>1.14</td>
<td>-0.470</td>
<td>0.468</td>
</tr>
<tr>
<td>C (23)</td>
<td>-0.418</td>
<td>0.485</td>
<td>0.352</td>
<td>0.051</td>
<td>0.046</td>
<td>0.045</td>
<td>7.62</td>
<td>1.22</td>
<td>-0.413</td>
<td>0.481</td>
</tr>
</tbody>
</table>

Table VIII. Positional parameters (fractional) with standard deviations (Å), thermal parameters and standard deviations (Å²), and peak electron-densities (e.Å⁻³)
The bond distances in the molecule, calculated from both sets of coordinates of Table VIII, are shown in Fig. 19(a), and the valency angles (least-squares parameters) in Fig. 19(b). The standard deviations of these are about 0.05 Å for bond distances and 3° for angles. The most significant intermolecular contacts are given in Table IX (least-squares parameters), and the molecular packing is illustrated in Fig. 20.

Table IX. Shorter intermolecular distances
(For C-C, C-N, all distances \( \leq 3.8 \) Å are listed; for C-I, N-I, all distances \( \leq 5.0 \) Å)

<table>
<thead>
<tr>
<th>Atom (molecule 1)</th>
<th>Atom in Molecule</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3.41 Å</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>4.22 Å</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>4.84 Å</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>4.80 Å</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3.54 Å</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>4.66 Å</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>4.35 Å</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4.16 Å</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4.38 Å</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>4.37 Å</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>4.32 Å</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>3.93 Å</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>4.38 Å</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>4.33 Å</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>3.68 Å</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>3.77 Å</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>3.66 Å</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>3.50 Å</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>3.58 Å</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>3.73 Å</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3.78 Å</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3.79 Å</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>3.62 Å</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>3.69 Å</td>
</tr>
</tbody>
</table>

Molecule General coordinates
1 \( x,y,z \)
2 \( 1+x,y,z \)
3 \(-x,\frac{1}{2}+y,\frac{1}{2}-z \)
4 \( \frac{1}{2}+x,\frac{1}{2}-y,1-z \)
5 \( \frac{1}{2}-x,1-y,\frac{1}{2}+z \)
Fig. 19. (a) Bond distances from least squares, and (in parentheses) differential syntheses refinements.
(b) Valency angles (least squares parameters).
Fig. 20. Projection of the structure onto (100), illustrating the packing of the molecules.
There are only two parts of the molecule where it is useful to examine planarity of the atoms; one of these is the indole group and the other the neighbourhood of the ethylenic double bond. The best plane through the nine atoms of the indole rings has equation (least-squares parameters):

\[0.8546X+0.5114Y+0.0902Z=4.2135\]  (A)

where \(X\), \(Y\) and \(Z\) are coordinates in \(\text{Å}\). The plane through these nine atoms plus the two carbon atoms bonded to the five-membered ring is:

\[0.8532X+0.5150Y+0.0831Z=4.1804\]  (B)

The best plane through the two atoms of the ethylenic double bond plus the three atoms bonded to them is:

\[0.7802X+0.5198Y+0.3480Z=3.0449\]  (C)

The deviations of the atoms from these various planes (Table X) indicate that, within experimental error, the indole nucleus and the two attached carbon atoms are planar, and that the two ethylenic carbons and the three carbon atoms bonded to them also lie in one plane.

<table>
<thead>
<tr>
<th>Atom</th>
<th>A</th>
<th>B</th>
<th>Atom</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.038 Å</td>
<td>-0.058 Å</td>
<td>15</td>
<td>0.059 Å</td>
</tr>
<tr>
<td>4</td>
<td>0.055</td>
<td>0.033</td>
<td>16</td>
<td>-0.067</td>
</tr>
<tr>
<td>5</td>
<td>-0.069</td>
<td>-0.081</td>
<td>17</td>
<td>-0.009</td>
</tr>
<tr>
<td>6</td>
<td>0.030</td>
<td>0.028</td>
<td>18</td>
<td>-0.018</td>
</tr>
<tr>
<td>7</td>
<td>0.030</td>
<td>0.038</td>
<td>20</td>
<td>0.038</td>
</tr>
<tr>
<td>8</td>
<td>-0.040</td>
<td>-0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.013</td>
<td>0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-0.004</td>
<td>-0.006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.027</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-0.005</td>
<td>-0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.082</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Absolute configuration**

The final step in the analysis was the determination of the absolute configuration of the molecule by the anomalous dispersion method (53). Structure factors were calculated for all the \(hkl\) and \(\bar{h}\bar{k}\bar{l}\) reflections, using a scattering factor for I- of the form
\[ f = (f_I + \Delta f_I') + i \Delta f_I'' \]

With Mo K\(\alpha\) radiation the differences between \(F_c(hkl)\) and \(F_c(\bar{h}\bar{k}\bar{l})\) were small, and for about only a dozen reflections did the differences seem large enough to be detectable even with the counter equipment. These differences could have been increased by using Cu K\(\alpha\) radiation, but probably only at the expense of introducing absorption corrections (and of realigning the instrument). In an effort to make the determination as objective as possible the author calculated the structure factors and suggested which pairs of reflections were to be measured, carefully omitting any mention of the indicated directions of the differences. Dr. J. Trotter then measured the intensities. The results are given in Table XI. Of the fourteen pairs of reflections for which the differences were expected to be measurable, four showed no significant difference; the other ten pairs unambiguously indicated that the parameters used to calculate the structure factors (those of Table VIII referred to a conventional right-handed set of axes) represent the true absolute configuration. (V) and Fig. 18 therefore also depict the correct absolute configuration.

D. Discussion

The present analysis has established the structure of cleavamine methiodide, including the absolute configuration, as that shown in Figs. 18, 19, and 20, and the structure of the parent alkaloid as (V). The general shape of the molecule is clear from Figs. 18 and 20. The atoms of the indole rings and the two attached carbon atoms all lie in one plane. In the other six-membered ring four of the atoms (those of the C=C bond and the two attached atoms) are coplanar, C(23) is only slightly displaced (-0.14 \(\text{Å}\)), but N(3) lies significantly off the plane (-0.63 \(\text{Å}\)); C(20) and C(21) do not deviate significantly from the plane (displacements +0.04 \(\text{Å}\) and +0.07 \(\text{Å}\) respectively). The general boat conformation of the nine-membered ring is also evident from Figs. 18 and 20.
Table XI. Determination of the absolute configuration

(Mo Kα radiation)

| h  | k  | l  | \(|F_0|\) | \(|F_c(hkl)|\) | \(|F_c(hkl)|^2\) | \(|F_o(hkl)|^2\) | \(I_o(hkl)\) | \(I_o(hkl)\) |
|----|----|----|----------|----------------|----------------|----------------|-------------|-------------|
| 1  | 2  | 2  | 12.4     | 3.8            | 8.2            | 0.21           | 34          | 59          | 0.58        |
| 2  | 3  | 3  | 23.0     | 18.6           | 13.9           | 1.79           | 39          | 30          | 1.30        |
| 1  | 4  | 6  | 61.5     | 61.1           | 65.5           | 0.87           | 189         | 224         | 0.84        |
| 3  | 4  | 1  | 8.5      | 17.5           | 14.4           | 1.48           | 6           | 8           | 2.00        |
| 2  | 9  | 6  | 17.7     | 17.5           | 14.5           | 1.46           | 8           | 8           | 1.00        |
| 1  | 12 | 3  | 41.4     | 41.5           | 47.4           | 0.77           | 64          | 64          | 1.00        |
| 1  | 3  | 3  | 29.7     | 32.9           | 36.0           | 0.84           | 87          | 104         | 0.84        |
| 4  | 6  | 6  | 15.9     | 17.4           | 19.1           | 0.83           | 5           | 6           | 0.83        |
| 1  | 11 | 1  | 36.8     | 27.3           | 29.4           | 0.86           | 31          | 31          | 1.00        |
| 2  | 2  | 3  | 57.3     | 63.7           | 66.6           | 0.91           | 250         | 291         | 0.86        |
| 1  | 3  | 2  | 35.7     | 40.8           | 38.7           | 1.11           | 141         | 134         | 1.05        |
| 2  | 1  | 5  | 37.8     | 38.1           | 36.1           | 1.11           | 82          | 82          | 1.00        |
| 1  | 2  | 6  | 35.2     | 30.4           | 32.6           | 0.87           | 65          | 81          | 0.80        |
| 3  | 3  | 10 | 35.0     | 32.5           | 35.1           | 0.86           | 34          | 45          | 0.76        |

The structure is very markedly changed from that of catharanthine (W), the atoms C(13) and C(18), which are directly bonded in catharanthine, being separated in cleavamine by 4.6 Å.

The differences between the bond distances determined by the least-squares and differential syntheses refinements (Fig. 19(a)) suggest that the measurements are of limited accuracy; only for the bonds involving C(23) do the differences exceed 2σ however. The C=C bond, C(16)-C(17) (whose position in the molecule is well established chemically, and in the present analysis by the planar arrangement of the atoms around it) has a length of 1.39 Å (differential syntheses parameters) or of 1.48 Å (least-squares parameters); even the latter value is less than the mean* C-C single bond distance in the molecule, which is 1.55 Å, not significantly different from the normal single bond distance.

The mean C-N+ length is 1.52 Å, the mean C-C bond length in the aromatic six-membered ring is 1.39 Å, and the mean C-N(2) distance 1.38 Å. The two C-C bonds in the five-membered ring appear rather long (average value 1.53 Å); the presence of an indole nucleus, rather than dihydroindole, is again well established chemically and by the planar arrangement of atoms, and the length

* The mean values are averages of the least-squares and differential syntheses results.
of these bonds might indicate a slight inaccuracy in the position of atom 5, the valency angles at this atom being somewhat anomalous (Fig. 19(b)).

The valency angles (Fig. 19(b)) are all quite reasonable; the mean tetrahedral angle is 111°, mean angle in the aromatic six-membered ring 120°, mean in the five-membered ring 108°, and mean at C=C double bond 119°.

The most significant packing distance is an I⁻-indole nitrogen separation of 3.41 Å (least squares; 3.47 Å differential syntheses) as shown in Figs. 18 and 20. This probably represents an I...H-N hydrogen bond, although the hydrogen, if it is situated on the bisector of the C(4)-N(2)-C(11) angle, is displaced from the I...N line (N-H bond displaced 14°, Fig. 19(b)). All the other distances involving the iodide ion correspond to normal separations, the next shortest being 3.93 Å (Table IX). All the C-C and C-N intermolecular contacts involve normal van der Waals interactions, the shortest being 3.50 Å.

This work has provided evidence for a novel and remarkable acid rearrangement in the Iboga alkaloid series which may have interesting mechanistic and biogenetic implications.

It is pertinent to note that cleavamine has been obtained as one of the products in the acidic hydrolysis of vincaleukoblastine and leurosine (49). This work now provides strong support for the presence of a catharanthine-like moiety as one of the units in the structure of these dimeric alkaloids.
ADDENDUM

The least-square program was written in FORTRAN-2 for the IBM 1620 computer with 40K memory and card input unit.

The general method followed was that given by Cruickshank (54). The block-diagonal approximation was used, 3X3 and 1X1 matrices for the coordinates and isotropic temperature parameters, and a 2X2 matrix for correlation of the scale factor and the average temperature factor (the memory size prevented the use of anisotropic temperature factors). The standard deviations of the positional and temperature parameters were computed according to Cruickshank's equation (2.5).

The input cards (one per plane) contained the indices, \( \sin^2 \theta \), and the scattering-factors for that plane (this was chosen as input because these cards were available for the structure factor program on hand).

Since the program was in FORTRAN, and the calculations were in floating-point mode, the time taken was necessarily quite long, being for cleavamine methiodide (23 atoms and 4 equivalent positions, with 23 different isotropic temperature factors) approximately one minute per reflection. The advantages of FORTRAN with floating-point are ease of programming and of modification of the program, and automatic handling of the decimal point.
II. 5-iodo-2'-deoxyuridine

A. Introduction

5-iodo-2'-deoxyuridine (IDU) (Fig. 22) has been used in the treatment and cure of herpes simplex keratitis (55) and is the first anti-viral agent to have proven clinical chemotherapeutic value, giving the first clear-cut demonstration that true viral disease can be effectively treated without obvious harm to the host. The blocking of the metabolic pathways of viral synthesis by IDU has been attributed to either the selective action of IDU on a virus-specific enzyme system which may be involved in the synthesis of viral DNA, or to the incorporation of IDU itself in place of thymidine into an aberrant DNA which presumably cannot be utilized to form infective virus particles (56). IDU and the bromine analogue (BrDU) can certainly be incorporated in place of thymidine into DNA of bacteria, bacteriophages, and even human and other mammalian cells (57); heavily substituted DNA can sometimes still undergo full biological functioning but with retarded growth rate, increased genetic instability, and increased radiation sensitivity. Incorporation of IDU and BrDU into DNA results in significantly higher melting temperatures (temperature at which the two DNA helices separate) than that exhibited by normal DNA. Of the several explanations put forward to explain this enhanced stability, the most probable has been that the substitution of I or Br for the CH\(_3\) in thymidine alters the electron density in the pyrimidine ring with concomitant strengthening of the hydrogen bonding to the purine base in the complementary DNA chain (58). It would thus appear that the incorporation of IDU and BrDU instead of thymidine into DNA hinders further DNA synthesis by increasing the inter-coil attraction and thereby impairing the ability of the complementary chains to separate and serve as templates for additional DNA.

The crystal and molecular structures of 5-iodo-2'-deoxyuridine and 5-bromo-2'-deoxyuridine have been investigated in the hope that the detailed structures might yield some direct information on the possible role of these compounds in combatting viral diseases.
B. Experimental

Crystals of 5-iodo-2'-deoxyuridine and 5-bromo-2'-deoxyuridine are plates elongated along a with (001) developed. Most of the crystals were twinned, but some single ones were obtained from alcohol-water. The density of IDU was measured by flotation in a CCl₄-CHBr₃ mixture, and the unit cell dimensions of both compounds were determined from rotation, Weissenberg (CuKα, λ = 1.5418 Å) and precession (MoKα, λ = 0.7107 Å) films. The two are isomorphous, and crystal data are given in Table XII.

The iodo-compound was chosen for detailed analysis, and the intensities of all reflections with 2θ_MoKα ≤ 57.3° (corresponding to a minimum interplanar spacing d = 0.74 Å) were measured on a General Electric XRD-5 Spectrogoniometer, with Single Crystal Orienter, using a scintillation counter, approximately monochromatic MoKα radiation (zirconium filter and pulse height analyser), and the moving crystal-moving counter technique (41). All the intensities were corrected for background, Lorentz and polarization factors were applied, and the structure amplitudes were derived. The crystal used was mounted with a* parallel to the φ axis of the goniostat, and had length 0.50 mm parallel to a and cross-section 0.16 x 0.07 mm perpendicular to a; with MoKα radiation absorption is fairly low and no corrections were considered necessary. 1434 reflections were observed, 99.2% of the total number (1445) in the range 0 ≤ 2θ_MoKα ≤ 57.3°.

C. Structure Analysis

Since the compound is optically active the space group is P1, and with only one molecule in the unit cell the origin may be taken at the iodine atom. A three-dimensional Fourier series was summed with all phase angles taken as 0°. The resulting electron-density distribution necessarily contained a false centre of symmetry, but the sugar ring and all its substituent groups were very clearly resolved. The base was poorly resolved since it was
<table>
<thead>
<tr>
<th></th>
<th>5-iodo-2'-deoxyuridine</th>
<th>5-bromo-2'-deoxyuridine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>C₉H₁₁N₂O₂I</td>
<td>C₉H₁₁N₂O₂Br</td>
</tr>
<tr>
<td>M.W.</td>
<td>354.1</td>
<td>307.1</td>
</tr>
<tr>
<td>Crystal system</td>
<td>Triclinic</td>
<td>Triclinic</td>
</tr>
<tr>
<td>a (Å)</td>
<td>4.98 ± 0.01</td>
<td>4.87 ± 0.01</td>
</tr>
<tr>
<td>b (Å)</td>
<td>6.83 ± 0.01</td>
<td>6.72 ± 0.01</td>
</tr>
<tr>
<td>c (Å)</td>
<td>9.60 ± 0.02</td>
<td>9.56 ± 0.02</td>
</tr>
<tr>
<td>α</td>
<td>101°40’ ± 5’</td>
<td>100°10’ ± 5’</td>
</tr>
<tr>
<td>β</td>
<td>109°18’ ± 5’</td>
<td>107°24’ ± 5’</td>
</tr>
<tr>
<td>γ</td>
<td>98°20’ ± 5’</td>
<td>98°31’ ± 5’</td>
</tr>
<tr>
<td>U (Å³)</td>
<td>293</td>
<td>285</td>
</tr>
<tr>
<td>Dm (gcm⁻³)</td>
<td>2.014</td>
<td>-</td>
</tr>
<tr>
<td>Z</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dx (gcm⁻³)</td>
<td>2.008</td>
<td>1.789</td>
</tr>
<tr>
<td>F(000)</td>
<td>172</td>
<td>154</td>
</tr>
<tr>
<td>μ (for CuKα) (cm⁻¹)</td>
<td>222</td>
<td>-</td>
</tr>
<tr>
<td>μ (for MoKα) (cm⁻¹)</td>
<td>28</td>
<td>-</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
</tr>
</tbody>
</table>
situated close to a false centre of symmetry, but it was possible, by making use of the atoms which were resolved and by assuming conventional dimensions, to obtain coordinates for all the atoms (excluding hydrogen) in the molecule. Structure factors were calculated using the scattering factors of the International Tables (5), with \( B = 4.5 \text{ Å}^2 \) for all atoms; the iodine scattering factor was corrected for anomalous dispersion according to the relation:

\[
\tilde{f}_{\text{corrected}} = \sqrt{\left( f_0 + \Delta f' \right)^2 + \left( \Delta f'' \right)^2},
\]

and \( \Delta f' \), \( \Delta f'' \) were taken from the International Tables (5). \( R \), the usual discrepancy factor, was 0.249. A second Fourier series was summed with phases based on all the atoms, and the resulting three-dimensional electron-density distribution showed very good resolution of all the atoms, with no spurious detail.

Refinement of the positional and isotropic thermal parameters and a scale factor was then carried out by (block-diagonal) least-squares. The function minimized was \( \sum w \left( |F_0| - |F_c| \right)^2 \), with \( \sqrt{w} = |F_0| / 20 \) when \( |F_0| < 20 \) and \( \sqrt{w} = 20 / |F_0| \) when \( |F_0| \geq 20 \). Refinement was complete in four cycles, during which \( R \) was reduced from 0.249 to 0.142, and \( \sum w \Delta^2 \) from \( 12 \times 10^3 \) to \( 5.6 \times 10^3 \). At this stage facilities for computing anisotropic least squares became available, and two more cycles of refinement reduced \( R \) to 0.054 and \( \sum w \Delta^2 \) to \( 1.3 \times 10^3 \); in the anisotropic refinement no significant changes in positional parameters were indicated, so that the improvement in structure factor agreement is entirely a result of introducing anistropic thermal parameters.

The final calculated structure amplitudes are compared with the measured values in (59) \( (R = 0.054 \text{ for all } 1445 \text{ planes with } d \geq 0.74 \text{ Å}) \). A final three-dimensional Fourier series was summed, and sections of the resulting electron-density distribution, taken through the atomic centres parallel to \((100)\), are shown in Fig.21.
Fig. 21. Superimposed sections of the electron-density distribution, taken through the atomic centres parallel to (100). Contours start at 0 eÅ⁻³ and are at intervals of 20 eÅ⁻³ about the iodine, and 2 eÅ⁻³ about the light atoms. A perspective drawing of the molecule is also shown.
Atomic parameters and molecular dimensions

The final positional and anisotropic thermal parameters are given in Table XIII, \(x\), \(y\), and \(z\) being fractional coordinates referred to the triclinic crystal axes. Since the compound is derived from d-ribose the absolute configuration is established; the coordinates in Table XIII referred to a right-hand set of axes give the correct absolute configuration, and all the diagrams also show the true configuration. \(B_{ij}\) in Table XIII are the coefficients in the expression:

\[
\exp \left\{- \left\{ -B_{11}h^2 + B_{22}k^2 + B_{33}l^2 + B_{23}kl + B_{13}hl + B_{12}hk \right\} \right\}
\]

The standard deviations of the atomic coordinates (in Å) are included in Table XIII; these were estimated from the inverses of the diagonal elements of the matrix of the least squares normal equations.

The bond distances and valency angles in the molecule are shown in Fig. 22; the standard deviations of the bond lengths vary from 0.015 - 0.028 Å, and of the angles from 1.4° - 1.9°.

The best plane through the pyrimidine nucleus and the I, O (2), O (6), and C (1') atoms (all atoms, including I, given equal weight) has equation

\[
0.82253 X' + 0.56395 Y' - 0.07341 Z' + 0.80200 = 0 \quad \ldots \quad (D)
\]

where \(X', Y',\) and \(Z'\) are coordinates in Å units referred to orthogonal axes \(a', b',\) and \(c'.\) A least-squares plane was calculated through all five atoms of the deoxyribose ring, but the displacements indicated that the ring was significantly non-planar. Planes through all possible combinations of four atoms of the ring were calculated, and one of these was much more satisfactory than the others, atoms C (1'), C (3'), C(4'), and O (1') being essentially coplanar, the plane having equation

\[
0.68910 X' - 0.72450 Y' + 0.01496 Z' + 0.98961 = 0 \quad \ldots \quad (E)
\]

with C(2') displaced 0.59 Å from this plane. The deviations of the atoms
Table XIII. Final positional parameters (fractional), standard deviations (Å), and anisotropic thermal parameters (x 10^5).

<table>
<thead>
<tr>
<th>Atom</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>σ(x)</th>
<th>σ(y)</th>
<th>σ(z)</th>
<th>B_{11}</th>
<th>B_{22}</th>
<th>B_{33}</th>
<th>B_{23}</th>
<th>B_{13}</th>
<th>B_{12}</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.0035</td>
<td>0.0034</td>
<td>0.0035</td>
<td>3422</td>
<td>1185</td>
<td>1099</td>
<td>687</td>
<td>1748</td>
<td>2322</td>
</tr>
<tr>
<td>N(1)</td>
<td>0.5809</td>
<td>0.6132</td>
<td>1.1563</td>
<td>0.0138</td>
<td>0.0137</td>
<td>0.0142</td>
<td>4215</td>
<td>1285</td>
<td>1323</td>
<td>508</td>
<td>2396</td>
<td>3338</td>
</tr>
<tr>
<td>C(2)</td>
<td>0.5719</td>
<td>0.4787</td>
<td>1.0257</td>
<td>0.0186</td>
<td>0.0179</td>
<td>0.0189</td>
<td>3479</td>
<td>927</td>
<td>1298</td>
<td>1516</td>
<td>1767</td>
<td>3720</td>
</tr>
<tr>
<td>O(2)</td>
<td>0.7339</td>
<td>0.3577</td>
<td>1.0370</td>
<td>0.0125</td>
<td>0.0125</td>
<td>0.0127</td>
<td>4978</td>
<td>2282</td>
<td>1366</td>
<td>1410</td>
<td>1994</td>
<td>5416</td>
</tr>
<tr>
<td>N(3)</td>
<td>0.3761</td>
<td>0.4918</td>
<td>0.8902</td>
<td>0.0143</td>
<td>0.0142</td>
<td>0.0145</td>
<td>3583</td>
<td>1340</td>
<td>709</td>
<td>395</td>
<td>1301</td>
<td>3387</td>
</tr>
<tr>
<td>C(4)</td>
<td>0.2192</td>
<td>0.6401</td>
<td>0.8804</td>
<td>0.0176</td>
<td>0.0173</td>
<td>0.0182</td>
<td>4162</td>
<td>976</td>
<td>579</td>
<td>-309</td>
<td>1507</td>
<td>958</td>
</tr>
<tr>
<td>C(5)</td>
<td>0.2357</td>
<td>0.7770</td>
<td>1.0065</td>
<td>0.0131</td>
<td>0.0128</td>
<td>0.0140</td>
<td>2070</td>
<td>1350</td>
<td>320</td>
<td>111</td>
<td>-246</td>
<td>2265</td>
</tr>
<tr>
<td>C(6)</td>
<td>0.4290</td>
<td>0.7627</td>
<td>1.1581</td>
<td>0.0147</td>
<td>0.0145</td>
<td>0.0153</td>
<td>2503</td>
<td>1060</td>
<td>887</td>
<td>736</td>
<td>1390</td>
<td>1493</td>
</tr>
<tr>
<td>O(6)</td>
<td>0.4554</td>
<td>0.8796</td>
<td>1.2769</td>
<td>0.0119</td>
<td>0.0119</td>
<td>0.0120</td>
<td>4824</td>
<td>1525</td>
<td>761</td>
<td>192</td>
<td>1324</td>
<td>1579</td>
</tr>
<tr>
<td>C(1')</td>
<td>0.3567</td>
<td>0.3480</td>
<td>0.7458</td>
<td>0.0164</td>
<td>0.0161</td>
<td>0.0169</td>
<td>2844</td>
<td>1215</td>
<td>900</td>
<td>474</td>
<td>1729</td>
<td>1643</td>
</tr>
<tr>
<td>O(1')</td>
<td>0.4699</td>
<td>0.4619</td>
<td>0.6613</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0118</td>
<td>3101</td>
<td>1863</td>
<td>675</td>
<td>641</td>
<td>1162</td>
<td>1254</td>
</tr>
<tr>
<td>C(2')</td>
<td>0.0386</td>
<td>0.2374</td>
<td>0.6368</td>
<td>0.0166</td>
<td>0.0163</td>
<td>0.0170</td>
<td>3121</td>
<td>1352</td>
<td>1104</td>
<td>604</td>
<td>2038</td>
<td>1533</td>
</tr>
<tr>
<td>C(3')</td>
<td>0.0782</td>
<td>0.1796</td>
<td>0.4804</td>
<td>0.0169</td>
<td>0.0167</td>
<td>0.0174</td>
<td>3095</td>
<td>1101</td>
<td>793</td>
<td>614</td>
<td>1378</td>
<td>1748</td>
</tr>
<tr>
<td>O(3')</td>
<td>0.2026</td>
<td>0.0019</td>
<td>0.4789</td>
<td>0.0130</td>
<td>0.0130</td>
<td>0.0131</td>
<td>6380</td>
<td>893</td>
<td>1665</td>
<td>866</td>
<td>3541</td>
<td>3108</td>
</tr>
<tr>
<td>C(4')</td>
<td>0.3312</td>
<td>0.3672</td>
<td>0.5011</td>
<td>0.0159</td>
<td>0.0154</td>
<td>0.0161</td>
<td>3206</td>
<td>1070</td>
<td>926</td>
<td>414</td>
<td>2095</td>
<td>2316</td>
</tr>
<tr>
<td>C(5')</td>
<td>0.2177</td>
<td>0.5277</td>
<td>0.4163</td>
<td>0.0192</td>
<td>0.0190</td>
<td>0.0195</td>
<td>3088</td>
<td>1836</td>
<td>1230</td>
<td>1778</td>
<td>1418</td>
<td>2058</td>
</tr>
<tr>
<td>O(5')</td>
<td>-0.0101</td>
<td>0.5904</td>
<td>0.4526</td>
<td>0.0123</td>
<td>0.0124</td>
<td>0.0126</td>
<td>2891</td>
<td>1550</td>
<td>1517</td>
<td>1189</td>
<td>1562</td>
<td>2423</td>
</tr>
</tbody>
</table>
Fig. 22. Bond lengths and angles in 5-iodo-2'-deoxyuridine.
from the various planes are given in Table XIV.

All the intermolecular separations $\leq 4.5$ Å were calculated, and the more significant contacts are listed in Table XV.

D. Discussion

The pyrimidine base

The base is in the di-keto form, the two C-O bonds having an average length of 1.22 Å (1.21 and 1.23 Å). The four C-N bond distances are virtually indentical (1.36, 1.37, 1.37, and 1.38 Å), averaging 1.37 Å. The value of 1.34 Å for the C (4) - C (5) distance confirms its true double-bond character, in better agreement with the results of 1.35 Å obtained for thymine monohydrate (60) and 1.31 Å for calcium thymidylate (61), than with the value of 1.41 Å obtained in a structural determination of uracil (62). The C (5) - C (6) bond is a single bond, measuring 1.49 Å, somewhat larger than the value of 1.45 Å found in thymine and calcium thymidylate, and very significantly longer than the 1.41 Å determined for uracil. The C (5) - I length of 2.05 Å is normal.

All of the bond angles are almost identical with those found in thymine, the difference exceeding 1° only in the exo-cyclic C-O angles. The values of 121 and 122° about C (6) - O (6) in IDU are more symmetrical than the corresponding 118 and 126° in thymine, but the distortions may be due to intermolecular packing in thymine; similarly the angles of 119° and 125° about C (2) - O (2) in IDU (122 and 123° in thymine) reflect the participation of O (2) in a short intermolecular contact.

The general agreement of the structure of the base in IDU with thymine (the structural determination of which was the most precise of the three cited above) demonstrates that the substitution of I for CH$_3$ does not significantly alter the structure of the base, and suggests that perhaps the differences found between thymine and uracil are not real.
Table XIV. Deviations from molecular planes (Å)

<table>
<thead>
<tr>
<th>Plane D</th>
<th>Deviation</th>
<th>Plane E</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atom</td>
<td></td>
<td>Atom</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.046</td>
<td>C (1')</td>
<td>0.015</td>
</tr>
<tr>
<td>N (1)</td>
<td>-0.046</td>
<td>C (3')</td>
<td>-0.013</td>
</tr>
<tr>
<td>C (2)</td>
<td>0.005</td>
<td>C (4')</td>
<td>0.032</td>
</tr>
<tr>
<td>O (2)</td>
<td>0.077</td>
<td>O (1')</td>
<td>-0.032</td>
</tr>
<tr>
<td>N (3)</td>
<td>-0.034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C (4)</td>
<td>0.010</td>
<td>C (2')</td>
<td>-0.590</td>
</tr>
<tr>
<td>C (5)</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C (6)</td>
<td>-0.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O (6)</td>
<td>-0.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C (1')</td>
<td>0.007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table XV. Shorter intermolecular contacts (Å).

All distances < 4.5 Å between a standard molecule (1) and neighbouring molecules were calculated. Only those < 3.9 Å involving iodine, and < 3.5 Å not involving iodine are listed.

<table>
<thead>
<tr>
<th>Atom to atom (molecule 1)</th>
<th>atom in molecule</th>
<th>d (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O (2) .................... I</td>
<td>6</td>
<td>2.96</td>
</tr>
<tr>
<td>O (5')---------------------H—N (1)</td>
<td>7</td>
<td>2.95</td>
</tr>
<tr>
<td>O (3')---------------------H—N (6)</td>
<td>4</td>
<td>2.70</td>
</tr>
<tr>
<td>O (3')---------------------H—N (5')</td>
<td>2</td>
<td>2.79</td>
</tr>
<tr>
<td>C (3')</td>
<td>O (6)</td>
<td>8</td>
</tr>
<tr>
<td>C (5')</td>
<td>O (6)</td>
<td>3</td>
</tr>
<tr>
<td>O (1')</td>
<td>C (2')</td>
<td>5</td>
</tr>
<tr>
<td>O (3')</td>
<td>C (5')</td>
<td>2</td>
</tr>
<tr>
<td>O (3')</td>
<td>O (6)</td>
<td>8</td>
</tr>
</tbody>
</table>

Molecule

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x</td>
<td>-1 + y</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>-1 + y</td>
</tr>
<tr>
<td>5</td>
<td>1 + x</td>
<td>y</td>
</tr>
<tr>
<td>6</td>
<td>1 + x</td>
<td>-1 + y</td>
</tr>
<tr>
<td>7</td>
<td>-1 + x</td>
<td>y</td>
</tr>
<tr>
<td>8</td>
<td>-1 + x</td>
<td>-1 + y</td>
</tr>
</tbody>
</table>
Least-squares planes calculated through all of the base and attached atoms, and through all but C (1'), showed that C (1') lies on the plane of the others. The maximum deviation from the plane (plane D of Table XIV), is that of atom O (2) (0.077 Å) and probably is due to the short intermolecular I---O (2) distance. The significant displacements of the I and N (1) (0.046, -0.046 Å) from the plane are probably also caused by intermolecular contacts, the deviation of N (1) possibly to facilitate the hydrogen-bonding scheme.

The sugar

The deoxyribose ring is puckered with atom C (2') lying 0.59 Å out of the plane (plane E of Table XIV) of the other four atoms (maximum deviation of the other four from the plane is 0.03 Å). This contrasts with calcium thymidylate, in which C (3') was the atom which was out of the plane of the others, and suggests that which of C (2') or C (3') deviates from planarity is a function of the packing in the individual crystal structures.

The bond distances in the ring are normal, the average C-C length being 1.56 Å, and C-O, 1.42 Å. The internal bond angles average 104° at C, and the angle is 111° at O (1'); the external angles average 112° at carbon. These values are similar to those (105°, 108°, 112°) predicted by Spencer (63) from a survey of similar molecules.

The glycosidic N (3) - C (1') bond measures 1.49 Å, similar to that in adenosine-5'-phosphate (64) and close to the value of 1.47 Å in calcium thymidylate.

Conformation of the molecule

The plane of the sugar ring is at a dihedral angle of 81° to the base, a little larger than the values of 76° in adenosine-5'-phosphate and cytidine (65), and 75° in calcium thymidylate.
In discussing the conformation about the glycosidic C-N bond, it is convenient to use the $\phi_{CN}$ torsion angle defined by Donohue and Trueblood (66). This is defined as the angle formed by the trace of the plane of the base with the projection of the C (1') - O bond when the projection is taken along the glycosidic bond itself. As for all of the nucleosides and nucleotides so far studied, IDU is in the anti conformation, with a value of $\phi_{CN}$ of -22° (as contrasted with $\phi_{CN}$ of about 150° expected in the syn conformation), similar to the value in adenosine-5'-phosphate of -18°.

**Intermolecular distances**

An isometric projection of the intermolecular packing is shown in Fig. 23. The most significant approach is an I---O (2) distance of 2.96 Å, which is considerably shorter than the normal iodine - oxygen van der Waals contact (3.55 Å), and suggests a charge-transfer bond involving donation of oxygen lone-pair electrons to vacant 5d orbitals of the iodine. The carbonyl group-iodine arrangement is not too far from linear, the C (2) - O (2) - I angle being 166°. Such charge-transfer bonds involving halogens as acceptors have been previously reported (67) with oxygen-, nitrogen-, and even sulfur- and selenium-containing compounds acting as donors. The stability of the bond increases with heavier halogens, and I---O distances considerably shorter than even the one reported here have been found. The ability of the iodine to form such charge-transfer bonds may be the molecular basis for the antiviral activity of 5-iododeoxyuridine. DNA which has had some thymidine substituted by IDU exhibits a higher melting temperature than does normal DNA; this may reflect extra inter-coil attraction caused by the formation of a charge-transfer bond by the iodine with an oxygen or nitrogen on the complementary purine base (or possibly on the complementary helical skeleton itself), either in addition to the normal hydrogen-bonding scheme, or by replacing it with an alternate one. The result would be hindrance of strand
Fig. 23. Isometric projection showing the packing of the molecules. The short I...O (2) contact is shown by a dotted line, and the hydrogen bonds by dashed lines.
separation, thus impairing the ability of the DNA to reproduce itself. In the case of herpes simplex keratitis treatment with IDU, this may kill the virus DNA, or merely retard its growth sufficiently to allow the body to overcome it (68).

The other short distances correspond to normal hydrogen bonds, and are shown by dashed lines in Fig. 23. Each hydroxyl group on the sugar ring (O (3') and (5')) is involved in two hydrogen bonds; O (3') uses its hydrogen in an H-bond to a carbonyl O (6) and shares an O (5') hydrogen in a second attraction, while the second O (5') bond is through an N (1) hydrogen. The O...H-N distance is 2.95 Å, while the O...H-O distances average 2.75 Å. Together with the iodine - O (2) contact, this system holds the crystal together in a three-dimensional network. Of the possible hydrogen-bonding atoms, only O (1') does not appear to take part in the intermolecular bonding scheme.

All other approaches correspond to normal van der Waals interactions.
III. CRYSTAL DATA FOR
SODIUM THYMIDYL-(5'→3')-THYMIDYLATE-(5')

A. Introduction

The key role of deoxyribonucleic acid (DNA) in transporting genetic information has resulted in a great deal of attention being drawn, in recent years, to the problem of elucidating its molecular structure. Because of the complexity of DNA, and the fact that it gives only fibre diffraction patterns, a single-crystal x-ray investigation of its structure is not feasible. Several early attempts at constructing molecular models which would agree with the known chemical, physical and biological properties of DNA, resulted in the double-helix model of Watson and Crick (69) which fits remarkably well with all of the experimental evidence gathered before and since. However, as Crick (70) and others have stated, further elaboration and understanding of the structure must await the accumulation of additional basic physical chemical data, including precise data on the intra- and intermolecular geometrical relationships in polynucleotides.

The structures of several nucleosides and nucleotides have been accurately determined, but of only one nucleotide of the deoxyribose series (61). One phosphate-linked dinucleoside (71) has been structurally elucidated, but the linkage was 2':5' rather than 3':5' as is the case in DNA; to date there have been no structure determinations of true dinucleotides reported.

The investigation of the structure of the dinucleotide, sodium thymidylyl-(5'→3')-thymidylate-(5'), was undertaken to provide the first structural information on the mutual arrangement of the two mononucleotides, and the geometry of the internucleotide linkage. Such information would be of great interest and importance in the establishment of the geometrical relations to be expected in polynucleotides and hence in DNA.
Thymidylyl-(5'→3')-thymidlyc acid-(5') was first prepared (calcium salt) by Michelson and Todd (72) by the condensation of 3'-O-acetylthymidine with thymidine 3'-(benzyl phosphorochoridate) 5'-(dibenzy1 phosphate) and subsequent removal of the protecting groups. This represented the first preparation of a dinucleotide by chemical means, and since the synthetic material behaved toward enzymes exactly as the dinucleotidic fragments obtained by degrading deoxyribonucleic acids, the postulate of a 3':5'-internucleotidic linkage in the latter was further confirmed.

A more general method for synthesizing polynucleotides with 3':5' internucleotidic linkages was reported by Gilham and Khorana (73), who achieved the 3':5' linkage by reacting a suitably protected deoxynucleotide with a second protected deoxy-nucleoside or -nucleotide in the presence of dicyclohexylcarbodiimide or p-toluenesulfonyl chloride.

The thymidine-dinucleotide sample used in this analysis was prepared by this latter method and very kindly provided to me as the sodium salt by Dr. G.M. Tener.
B. Experimental

Crystals of sodium thymidylyl-(5'→3')-thymidylate-(5')(VII) are colorless plates elongated along a with (001) developed. Most of the crystals gave diffraction patterns with extended streaky reflections, but suitable single crystals were ultimately recrystallized from 50% ethanol solutions. The density was measured by flotation in CHCl₃-CHBr₃ solution, and the unit cell dimensions and space group were determined from various rotation, oscillation, Weissenberg, and precession photographs, and on the G.E. spectrogoniometer.

Crystal data

\[ \lambda, \text{CuK} \alpha = 1.5418 \text{ Å}, \quad \lambda, \text{MoK} \alpha = 0.7107 \text{ Å} \]

Sodium thymidylyl-(5'→3')-thymidylate-(5'), \( \text{C}_{20}\text{H}_{25}\text{N}_4\text{O}_{15}\text{P}_2\text{Na}_3 \), mol. wt. 692.4

Orthorhombic, \( a = 16.06 \pm .04 \), \( b = 15.13 \pm .04 \), \( c = 15.65 \pm .04 \) Å

\( U = 3803 \text{ Å}^3 \)

\( \text{D}_m = 1.588 \text{ g.cm}^{-3}, \quad Z = 4, \quad \text{D}_x = 1.210 \text{ g.cm}^{-3} \) With 12 H₂O of hydration per molecule of dinucleotide

Absorption coefficient for x-rays, \( \lambda = 1.5418 \text{ Å}, \quad \mu = 23.0 \text{ cm}^{-1} \)

\( F(000) = 1904 \)

Absent spectra: h00 when h is odd, 0k0 when k is odd.

Space group \( P2_12_12 \quad (D_2^3) \)

Although Gilham and Khorana (73) reported 13 H₂O of hydration per molecule of dinucleotide (sodium salt) on the basis of chemical analysis, the density measured here corresponds closely with a value of 12 H₂O per dinucleotide molecule (\( \text{d}_x \) for 11, 12, and 13 H₂O = 1.555, 1.587, 1.618 g.cm⁻³ respectively, while \( \text{d}_m = 1.588 \)).

A preliminary photographic survey revealed a rapid fall-off in intensity with increasing Bragg angle. The intensities of all reflections with
2θ (CuKα) ≤ 80° (corresponding to an interplanar spacing d = 1.20 Å) were measured on a General Electric XRD 5 Spectrogoniometer with Single Crystal Orienter, a scintillation counter and CuKα radiation being used, approximately monochromatic radiation being obtained by use of a nickel filter and a pulse height analyser. The moving crystal-moving counter technique was employed (41). 953 reflections with 2θ ≤ 80° were observed, representing about 70% of the total in this range. For those reflections at higher Bragg angles, only those few visible on the films were examined and their intensities (all weak) recorded. The intensities were corrected for background, Lorentz and polarization corrections applied, and the structure amplitudes derived. The crystal used in recording the intensities was mounted with a parallel to the φ axis of the goniostat, and had dimensions 0.85 mm parallel to b and 0.013 mm parallel to c; no absorption corrections were applied.

C. Attempts at Structure Elucidation

The scattering power of an atom declines rapidly with sin θ/λ because of destructive interference from different parts of the electron cloud as the scattering direction departs from that of the direct beam. But if all of the scattering matter of the atom were concentrated at a point, this collection of electrons would scatter in phase in all directions, and the scattering power of the atom would be a constant value, the atomic number, Z.

We may denote the structure factor which would result if the crystal were composed of point atoms, all of type 1, as S^F_1, and it follows that

\[ S^F_1(hk\ell) = \frac{Z_1}{f_1(hk\ell)} F_1(hk\ell). \]

Crystals are not usually composed of only one type of atom, but the f-curves of most atoms have similar shapes, and differ mainly by a scale-factor, the ratio of the atomic numbers. This suggests that for a particular crystal one can find an average shape of the scattering curves of the several atoms present, scaled to unit scattering power, \( \hat{F} \). The scattering power of any atom is then approximately Z\( \hat{F} \).
We can then "sharpen" the F's by the relation \( f(hkL) = \frac{sF(hkL)}{F(hkL)} \).

The usual use for sharpened structure amplitudes is as coefficients in a Patterson synthesis, resulting in a sharpening and better resolution of the Patterson peaks.

A three-dimensional sharpened Patterson series was computed for the dinucleotide. Since the intensities of most of the planes at the highest values of sin \( \Theta \) were very weak, and hence had the highest probable error, the curve used as the sharpening function was made to peak in the middle sin \( \Theta \) range and tail off at the maximum recorded sin \( \Theta \) values. The Patterson plot was intensively studied for the locations of the peaks due to phosphorus, but because of the complexity of the molecular structure and the resultant high degree of overlap of peaks, no positions for the phosphorus could be confidently and unambiguously derived. The results that could be drawn from the three-dimensional Patterson synthesis were:

(a) the concentration of peaks on the \( Z = 0 \) section confirms the space group as \( P2_12_12 \).

(b) choosing P-P vectors to lie on the stronger regions, the only positions which fit the Patterson map are \( P_1(\frac{13}{60}, -\frac{3\cdot5}{60}, 0) \) and \( P_2(\frac{15}{60}, \frac{11}{60}, 0) \). These are not too convincing, as firstly they both have \( z=0 \), and secondly they seem to be too close together for best packing (when a different equivalent position is chosen for one of them they then seem too far apart for best packing).

A comparison of the origin peak height with \( \sum\overline{Z_j^2(11)} \) (j summed over all atoms of the cell), further indicated that the heights expected for the P-P peaks could be very small, so that the P-P vectors need not be chosen on the larger Patterson peaks. However, even the consideration of the smaller peaks led to no other reasonable positions for the phosphorus atoms.
An eightfold vector convergence (minimum) function was then computed, superimposing the Patterson function at the eight phosphorus positions in the unit cell (the two given above and their equivalent positions). The resulting three-dimensional Fourier map had regions of density scattered through it, but unfortunately it could not be interpreted in terms of any recognizable structure. (The superposition program could not interpolate, but could only superpose at the nearest lattice point ($\frac{1}{120}$ ths of the axial lengths), and perhaps this contributed to the lack of resolution of the Fourier map).

Several attempts were then made by the author and Dr. G.M. Tener to prepare crystalline heavier metal salts of the dinucleotide, with caesium, rubidium, silver, and thallium as the positive ion, but in no case could a crystalline sample be obtained.

After extensive and intensive re-consideration of the Patterson synthesis of the sodium salt failed to yield any further results, the project was shelved.

Recently the intensity data were given to M. Sundaralingam at the University of Washington, where they have a superposition program which can interpolate between lattice points. At the time of this writing no information has been received from them.

The intensity data for sodium thymidylyl-(5'→3')-thymidylate-(5') have been recorded by the author and are not given here.
APPENDIX

A BRIEF DESCRIPTION OF SOME CRYSTALLOGRAPHIC PROGRAMS

WRITTEN FOR THE IBM 1620 AND 7040 COMPUTERS
The programs were originally written for the IBM 1620 computer, with 40,000 digits core memory, card input/output, and printer output. They have been written in Fortran-2, and are easily adaptable to many other IBM computers; they have already been adapted here for the IBM 7040 computer.

A brief listing and summary of the programs is given below. A detailed description of their operation and a complete listing of the actual program instructions is obtainable from the Chemistry Department, University of British Columbia, c/o Dr. J. Trotter.

(1) **Goniostat Settings**

The program generates indices $hk\ell$ for all planes with $\sin \theta$ below a specified maximum value, and computes the three setting angles $(2\theta, \chi, \phi)$ for the General Electric Spectrogoniometer and Single Crystal Orienter. It is required that the crystal be mounted with a reciprocal axis parallel to the $\phi$ axis of the goniostat.

The input is only two cards, one specifying the cell constants, and the other the initial values and intervals desired for the indices $hk\ell$. The output is on cards (1620 computer) or magnetic tape (7040) in order to be conveniently sorted before listing.

The computations may be started at any value of the $H_3$ index, and zones of data may be obtained if desired.

(2) **Intensity Reduction** *(Equi-inclination Weissenberg films or G.E. Spectrogoniometer data)*.

Lorentz and polarization corrections are calculated and applied to each intensity and the structure amplitudes derived. A scale factor or an absorption correction may be applied to each reflection. Values of $F_0^2$ are also output for a Patterson synthesis, and in the 7040 version, scattering factor values are also output.
(3) **Bond Lengths and Valency Angles**

The program requires as input the unit cell constants, fractional atomic coordinates, and bond angles required. It computes, and outputs on the printer, orthogonal coordinates in Å referred to axes a' (=a.sinγ), b, and c'(perpendicular to a' and b), and for three atoms 1, 2, and 3, the distances 1-2, 2-3 (Å), and angle 1-2-3 (degrees).

(4) **Mean Molecular Plane**

The program requires as input the unit cell constants and fractional atomic coordinates. It outputs on the printer the values of A, B, C, D in the equation of the least squares best plane:

\[ AX' + BY' + CZ' + D = 0 \]

\((X', Y', Z')\) are coordinates in Å with respect to the same orthogonal axes as in program (3). The deviation of each atom from the plane is also output. Atoms may be given zero weight in the calculation of the plane, and their deviations computed.

(5) **Intermolecular Distances**

The program requires fractional atomic coordinates for a standard molecule and the general coordinates of up to 14 or more equivalent molecules. It then generates the atomic coordinates for the equivalent molecules and computes and prints out all the distances between all atoms of the standard molecule and all atoms of the generated molecules which are less than a specified maximum distance.

(6) **Standard Deviations of Atomic Coordinates**

The program computes and prints the standard deviations of the atomic coordinates and of the electron density, according to Cruickshank's formulae (23). The input required is the structure factor data and the atomic curvatures.
(7) **Least Squares Refinement**

This program has been already described, in Part III of this thesis (p. 68). It uses a block-diagonal approximation to refine the atomic coordinates. Refinement of a scale factor and of individual isotropic temperature factors is also performed, and the standard deviations of the positional and thermal parameters computed.

In the IBM 7040 adaptation of the program, provision has been made to compute and refine individual anisotropic temperature factors as well, lxl matrices being presently employed for each bij.
REFERENCES
2. E. Fedorow, Trans. Russian Min. Soc. 21, 1 (1888); 25, 1 (1890).
8. C.G. Darwin, Phil. Mag. 27, 315, 675 (1914).
40. W.R. Cullen, personal communication.
49. J.P. Kutney, personal communication.
58. S. Kit, in ibid., p.133.