THE CRYSTAL STRUCTURE OF 5-ETHYL-6-METHYLURACIL

Thesis by

George N. Reeke

California Institute of Technology

Pasadena, California

1964

ACKNOWLEDGEMENTS

I should like to thank Dr. Richard Marsh for his patient instruction and guidance which has made this work possible.

I am also indebted to numerous postdoctoral fellows and graduate students at Caltech, particularly Noel Jones, Dave DuChamp, and B.D. Sharma, for their assistance and good fellowship.

I gratefully acknowledge the financial support I have received from the National Science Foundation.

ABSTRACT

The crystal structure of 5-ethyl-6-methyluracil has been solved in the (001) projection. The compound crystallizes in the triclinic space group $P\bar{1}$. The unit cell contains two molecules, and has dimensions $1/a^* = 7.100$ Å, $1/b^* = 11.597$ Å, $\gamma^* = 84.739°$. The structure was solved through a sharpened Patterson map, which revealed the locations of the major ring vectors. Data have been gathered for a complete three-dimensional analysis, which is nearing completion.

THE CRYSTAL STRUCTURE OF 5-ETHYL-6-METHYLURACIL

I. Introduction and Preliminary Data

Current interest in the structures of the nucleic acids led the author to select a pyrimidine derivative for his first attempt at an x-ray crystal analysis. A large number of organic compounds prepared in the laboratories of the late Prof. Treat B. Johnson have been made available to this laboratory by Prof. James English of the Department of Chemistry, Yale University. 5-Ethyl-6-methyluracil (formerly called 4-Methyl-5-ethyluracil) was selected from among these compounds for an x-ray diffraction analysis of its crystal structure.

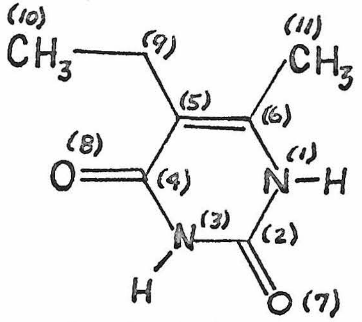

Fig. 1. 5-Ethyl-6-methyluracil showing the numbering scheme for heavy atoms used throughout this investigation.

The material was available in crystalline form. Typical crystals were white needles, averaging about 1 mm. in length and 0.1 mm. in diameter. Many crystals were poorly formed or bent. Recrystallization from water at room temperature yielded similar crystals.

A crystal was selected and mounted for rotation about the needle axis. An oscillation photograph indicated monoclinic symmetry with a needle period of 4.78 Å. A zero-layer Weissenberg photograph about this axis gave the other two shortest cell spacings: about 11.56 Å and 7.12 Å.

However, a first-layer Weissenberg photograph showed that the crystal must in fact be triclinic. It is likely that the false monoclinic symmetry of the original oscillation photograph was due to the use of a twinned crystal. It has subsequently been discovered that most of the crystals are actually mirror twins, and care has been exercised to select for mounting only single crystals.

It was then assumed that the space group is $P\bar{1}$, since a large proportion of triclinic crystals exhibit centric symmetry. Further observations and calculations have not disclosed any reason to suspect that this space group is incorrect, as the structure refines successfully in $P\bar{1}$.

The crystal density was measured roughly by flotation in dioxane-chloroform mixtures to be 1.25 gm/cc. The calculated value, assuming two molecules per unit cell, and assuming two right angles, was 1.29 gm/cc. We may conclude from this that the unit cell contains two molecules arranged about a center of symmetry.

An earlier investigation by Gerdil (1) reported that this compound crystallizes in two habits, according to the temperature. The crystals used here cerrespond to Gerdil's habit (a).

Gerdil's unit-cell parameters are given here for reference, relabelled to conform to the indexing I have chosen. Also given are parameters measured in this investigation (table I).

Table I. Certain Unit Cell Parameters

	a	b	c	alpha	beta	gamma	1/a*	1/b*	gamma*
A.	7.45	11.7	4.83	99°55'	108°	92°30'			
B.			4.78				7.12	11.56	
C.							7.100	11.597	84.74°

A Values given by Gerdil (1)
B Preliminary measurements, this investigation
C More precise measurements, described below

All lengths in Angstroms, all angles in degrees.

II. Two-Dimensional Refinement

A preliminary effort was made to arrive at approximate structures suitable for least-squares refinement by drawing models of the molecule based on standard bond lengths and angles, and moving these models about a drawing of the unit cell. Possible modes of hydrogen bonding were considered. A few structure factors calculated for several such structures failed to correlate with observed spot intensities.

Therefore, a Patterson map was calculated in the (001) projection, based on intensities of spots observed on zero-layer c-axis Weissenberg photographs. The intensities were estimated visually, using a standard intensity strip of graduated spots. The Lorentz and polarization corrections were applied to the intensities, and the Patterson function was calculated by hand, using Beevers-Lipson strips. The Patterson transform for the space group $P\bar{1}$ is given by:

$$P(U,V)=\sum_{h=0}^{\infty}\sum_{k=0}\left[\left(F_{hk0}^{2}+F_{h\bar{k}0}^{2}\right)\cos 2\pi hU\cos 2\pi kV-\left(F_{hk0}^{2}-F_{h\bar{k}0}^{2}\right)\sin 2\pi hU\sin 2\pi kV\right]$$

The calculation was done at an interval of 1/60 along V and 1/30 along U. Peaks near the origin on this Patterson map were not well resolved, and trial structures derived from it failed to give calculated structure factors which agreed with the observations.

It was therefore decided to construct a sharpened Patterson map. The crystal used by Gerdil had been discovered, mounted for rotation about the needle (c) axis. An oscillation photograph revealed that this crystal was not a twin. This crystal was therefore used to obtain complete c-axis intensity data for the zero through fourth Weissenberg layers, using the equi-inclination method. Two packs of three films each were exposed for the zero-layer, with exposure times such as to give a graduated series of six exposures (21 hrs. and 1 hr.). The intensities were estimated visually, using a standard comparison strip. The upper layers were similarly observed, except that each required two series of exposures, 180° apart in rotation, in order to observe all the spots. Corresponding packs 180° apart were always exposed for the same length of time in order to avoid a laborious scaling calculation.

The zero-layer intensities were used for the sharpened

Patterson, after application of Lorentz and polarization corrections by a program written for the Burroughs 220 computer by Albert Hybl (2). The corrected data were analyzed for scale and temperature factors by Wilson's statistical method (3), using atomic scattering factors for carbon and oxygen taken from a Burroughs 220 program written by B.D. Sharma (4). Nitrogen atoms were treated as oxygen, with the thought that the excess scattering factor thereby introduced would be qualitatively compensated by the presence of the neglected hydrogen atoms. The resulting calculated average temperature coefficient was 3.00.

The intensity data were then corrected to the absolute scale, and corrected for temperature and scattering factors. The origin peak was subtracted out, and finally, the modifying function was introduced. This function was taken from tabulated values in the B. D. Sharma program referred to above (5). Finally, the Patterson synthesis was summed using a Fourier summation program written for the Burroughs 220 by Hoogsteen (6) and modified by the author.

A reproduction of this map, with arbitrary contours, is included here (Fig. 2). The origin, placed at the center of the cell, is marked by a large dot. (The cell is drawn roughly to scale, based on the limited unit cell data then available.) Marked by lines are the three peaks nearest the origin, which were taken to be the three major ring vectors.

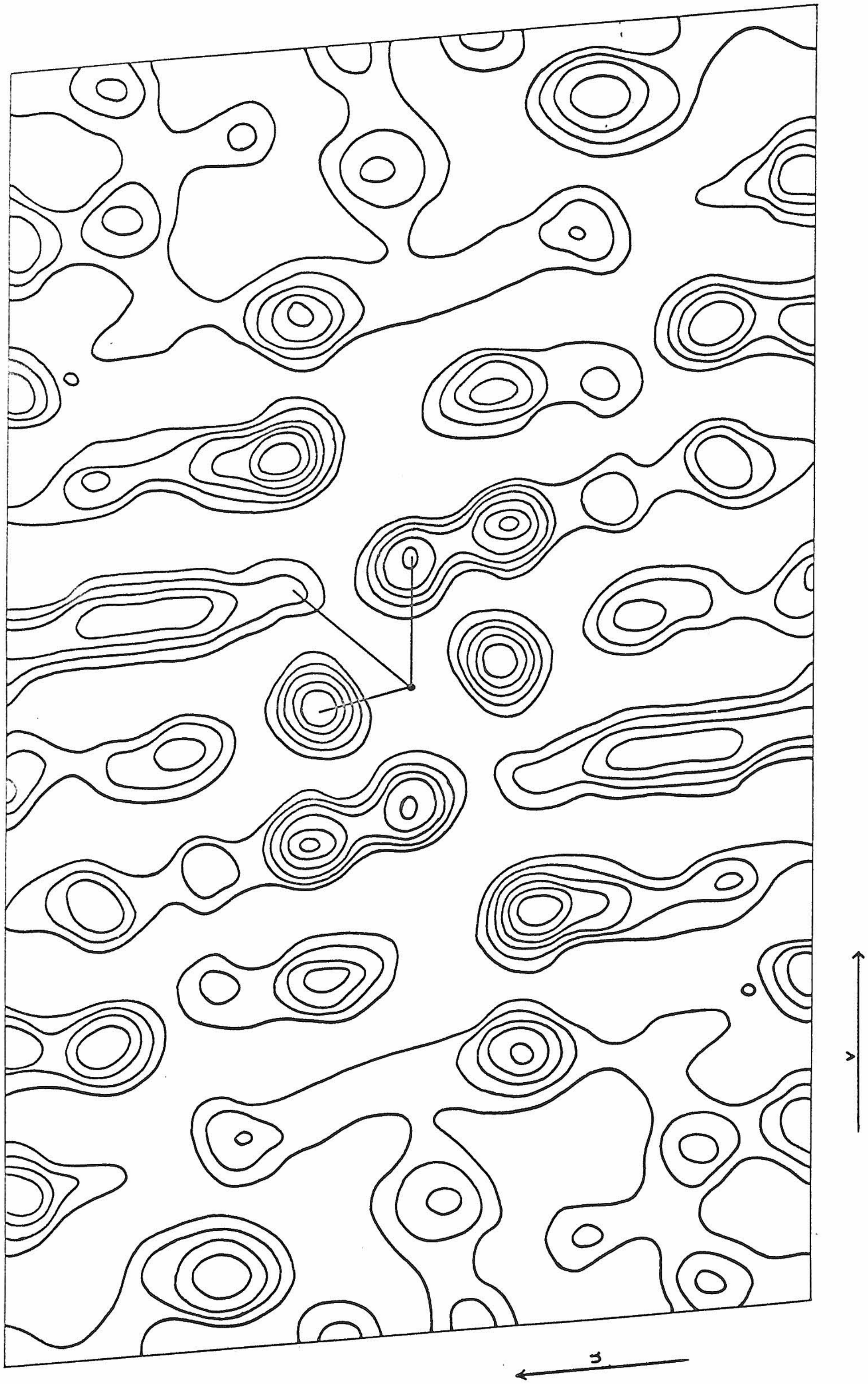

Fig. 2. Sharpened Patterson Map. (001) Projection

Five trial structures were drawn, based on the ring size and orientation defined by the sharpened Patterson. Structure factors were computed on the 220 computer, using a program written by R. Marsh (7). (This program was used for all structure factor and least-squares calculations in this investigation.) Comparison of calculated and observed structure factors gave good agreement for certain strong reflections in the case of the fourth of these trials. Consequently, an electron density Fourier map was calculated, using 30 strong reflections for which the sign of F_{obs} could be determined with reasonable certainty. This Fourier revealed the positions of all the heavy atoms. These new atomic positions were used to calculate new structure factors, and one round of least-squares refinement was performed. At this point, the R factor, defined by

$$R = \frac{\sum \left| |KF_{obs}| - |F_{calc}| \right|}{\sum |KF_{obs}|}$$

(K is a scaling factor)

was 47.1%.

Comparison of calculated and observed structure factors established the signs of 135 reflections. A Fourier map based on these reflections was made, and new atomic positions read from the resulting plot. Six rounds of least-squares refinement were performed on these positions, including refinement of isotropic temperature factors, and using one-half of the calculated shifts in each round.

The resulting atomic parameters are listed in table II, along with the calculated temperature factors and standard deviations. At this stage, the standard deviations should not be assigned a very high significance. After the last stage of least squares refinement, the R factor was 16.8 %. 187 reflections were included in the least squares, ten of which were "less than" readings for which F_{calc} was greater than the minimum F_{obs}. These ten reflections were not included in the calculation of R.

A final electron density Fourier map was calculated, including 183 reflections, and using F_{calc} for the strong (010) reflection, which was not observed. This Fourier map is reproduced here (Fig. 3.). It is contoured at levels of 1.5 electrons per square Angstrom from 1.5 to 10.5 $e/Å^2$, based on a unit cell area of 81.86 $Å^2$ derived from the rough unit cell dimensions. Two unit cells are shown, so that a complete molecule is visible. It should be kept in mind that this is the (001) projection.

One molecule has been drawn in, and the atoms labelled. Hydrogen bonding between amine and carbonyl groups of adjacent molecules is indicated by dotted lines. It is apparent that these hydrogen bonds make use of all available -NH and -CO groups to hold the molecules in long chains involving pairs of symmetry-related molecules. The normal to the molecular ring is roughly parallel to the c axis. The ethyl group is roughly in the same plane as the rest

Table II. Refined Atomic Parameters and Standard Deviations

Atom	Kind	x	y	B	sigma (x)	sigma (y)
1	N	0.45091	0.10744	3.478	.00258	.00163
2	C	0.30489	0.02240	2.426	.00274	.00169
3	N	0.18257	0.04023	3.302	.00250	.00159
4	C	0.17586	0.13824	2.685	.00277	.00170
5	C	0.32981	0.23365	2.207	.00267	.00169
6	C	0.45773	0.21156	2.519	.00270	.00170
7	O	0.29990	-0.06978	4.133	.00226	.00145
8	O	0.06105	0.14714	3.766	.00216	.00137
9	C	0.31858	0.34801	3.592	.00317	.00198
10	C	0.17271	0.42119	4.975	.00369	.00242
11	C	0.63511	0.28909	6.133	.00400	.00281

x is the fractional distance of the atom along the a axis from the center of symmetry.
y is the fractional distance of the atom along the b axis from the center of symmetry.
B is the isotropic temperature factor. The atoms in the ring (atoms 1-6) show smaller vibrations than the relatively free side group atoms, as expected.

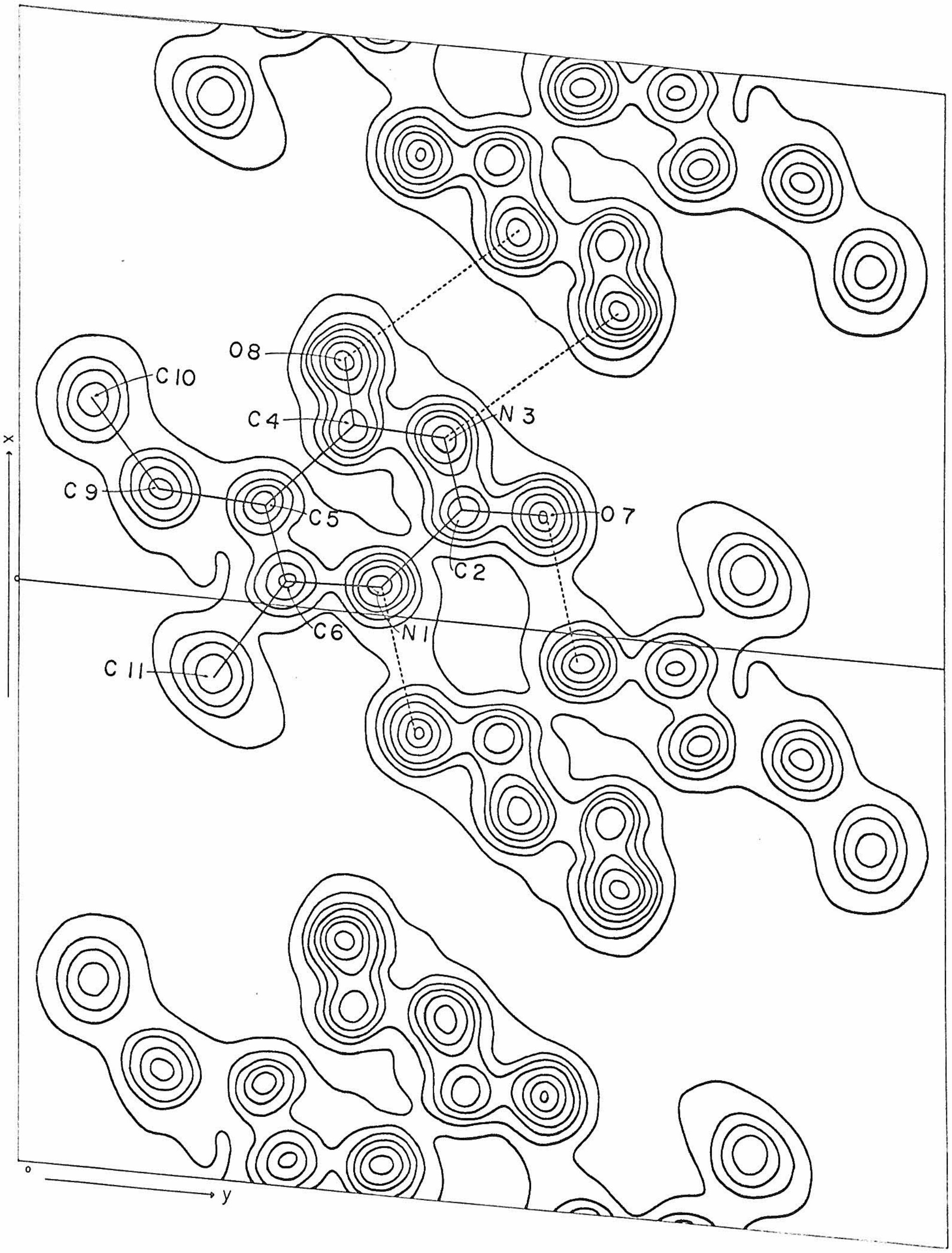

Fig. 3. Electron Density Fourier Map.
(001) Projection

of the molecule. The structure contains no unusual features. Detailed information as to the degree of tilt of the molecule, the exact shape of the unit cell, etc., shall have to await the completion of the three-dimensional analysis.

III. Three-Dimensional Analysis

A complete three-dimensional structure for this compound has not yet been determined. However, intensity data are available for all observable layers about the c axis. These data are nearly complete; therefore, gathering of intensity data about other axes, except for scaling purposes, is not contemplated.

A major difficulty in the triclinic case is the choice of a unit cell and the indexing of upper-layer photographs consistent with the set of axes chosen. The method employed here for indexing was as follows:

The needle axis had already been chosen as the c axis, although this choice is not consistent with the convention recommended by Buerger (8). It was therefore desirable that the scheme chosen would preserve the existing indexing of zero-layer c axis spots.

The reciprocal lattice planes corresponding to the zero- and first-layer Weissenberg c axis photographs were reconstructed on polar graph paper using a method described by Buerger. (9) When the plots of the two layers were superimposed, it was clear that one of four points on the first layer nearest to the (000) below must be chosen as (001). Calculation of direct cell parameters based on each of these four choices failed to establish unequivocally which choice satisfied crystallographic convention (shortest axes and obtuse angles).

At this point another crystal was mounted about a direction roughly perpendicular to the needle axis and parallel to a crystal face, and lined up about an axis which appeared to be densely populated with the aid of oscillation photographs. A zero-layer Weissenberg photograph was then taken. By comparing intensities with the earlier photographs, it was established that this photograph contained spots previously indexed as the (hh0) series, as well as spots from the as yet unindexed first-layer c axis.

A rather arbitrary choice was now made for the (001) spot, which allowed this new Weissenberg photograph to be consistently indexed as rotation about the $(1\bar{1}1)$ axis. The upper layer c axis photographs could now be indexed by plotting a few reciprocal lattice points on polar graph paper as before and superimposing the plots on these of the already indexed layers. In this way, it was possible to consistently index all of the photographs. It is felt that if the choice of axes taken fails to condorm to convention when precise direct cell parameters are finally calculated, then the entire indexing system can easily be transformed to a more suitable one.

A program was written for the Burroughs 220 computer for the purpose of computing film factors and scaling observed intensity data for the various layers. However, this program was used for the first layer only, as completion of appropriate portions of the new CRYRM system for the IBM 7090 has made it convenient to perform all further calculations using this system. (CRYRM is a unified program system for performing all common crystallographic calculations.)

Finally, the problem of obtaining precise unit cell parameters was attacked. Zero-layer Weissenberg photographs about the (001) and $(1\bar{1}1)$ axes were made with a special Straumanis type camera (10), and the Bragg angle measured for all visible reflections. A crystal has been successfully mounted about the (010) axis, so that a Straumanis photograph with this crystal will give the remaining data needed to calculate all the lattice parameters, using a least-squares fit to the observed Bragg angles (11).

For the present, 1/a*, 1/b*, and gamma* were calculated for the data about the (001) axis. The results of this calculation are reported in Table I.

Completion of the calculation of the lattice parameters, three-dimensional atomic positions, and Fourier electron density maps will conclude the investigation of the structure of 5-ethyl-6-methyluracil.

References:

1. Raymond Gerdil, "A preliminary crystallographic investigation of 1-phenylhydantoin and three derivatives of uracil." Acta Crystallographica, 13, 165 (1960).
2. Albert Hybl, Ph. D. Dissertation, California Institute of Technology, Pasadena, (1961). p. 206.
3. H. Lipson and W. Cochran, "The Determination of Crystal Structures," G. Bell & Sons Ltd., London, 1957, p. 61.
4. These are essentially the scattering factors of Berghuis et. al., Acta Crystallographica, 8, 478 (1955).
5. This modifying function was essentially:

$$M=\left(\frac{2\sin\theta}{\lambda}\right)^4 \exp\left[-\left(4.4\frac{\sin\theta}{\lambda}\right)^2\right] \qquad \theta = \text{Bragg angle}, \quad \lambda = \text{wavelength } (\text{Å}).$$

   The original source is no longer available.
6. K. Hoogsteen, "Crystallographic Fourier Summation Program for the Burroughs 220 Computer," California Institute of Technology, Pasadena (1960).
7. R. Marsh, "Structure Factor Least Squares Program for Monoclinic Space Groups," California Institute of Technology, Pasadena, (1960). This program minimizes the function

$$\sum_{h,k,l} \omega\left[|F_{obs}|^2 - |F_{calc}|^2\right]^2$$

   Each observation in the least squares was weighted according to the function

$$\sqrt{\omega} = \frac{1}{F_{obs}^2} \quad (F_{obs} \geq 4F_{min}) \qquad \sqrt{\omega} = \frac{1}{F_{obs} \cdot 4F_{min}} \quad (F_{obs} \leq 4F_{min})$$

8. M. J. Buerger, "X-ray Crystallography," John Wiley & Sons, New York, 1942, pp. 364-369.

9. *Ibid.* p.244

10. *Ibid.* p.394

11. N. D. Jones, Ph. D. Dissertation, California Institute of Technology, Pasadena, (1964). p. 116.