THE CRYSTAL STRUCTURE OF PURINE

Robert M. Sweet

May 24, 1965

Abstract

The crystal structure of Purine, $C_5N_4H_4$, has been determined and refined by Fourier and least-squares analyses of complete three-dimensional diffraction data from Copper X-radiation. The crystals are Orthorhombic with space group $Pna2_1$. The unit cell dimensions are $\underline{a}$ = 15.551, $\underline{b}$ = 9.3751, and $\underline{c}$ = 3.6626 Å. Least-squares refinement of the positional parameters of all 13 atoms and of the individual anisotropic temperature factors of the nine heavy atoms, using the 570 observed reflections of non-zero weight, yeilded a final R factor of 0.070.

The molecule containing two fused rings is planar. There is a Hydrogen on N(7) and none on N(9). N(7) participates in a Hydrogen bond of length 2.85 Å. with N(9).

Experimental

Purine was crystallized from water, alcohol, and water-alcohol solutions by slow evaporation at room temperature to form orthorhombic needles with the needle axis parallel to c. Intensity data was obtained from multi-film, equi-inclination Weisenberg photographs taken with CuK_α radiation. Layers 0-3 about c and 0-5 about b were photographed. The crystal used in the c-axis photographs was a thin needle from a methanol solution mounted lengthwise. That in the b-axis photographs was a larger needle from a water-methanol solution, mounted crosswise to the needle axis and shaped to approximately equal dimensions in the a and c directions by means of water-soaked tissue paper. The intensities were estimated visually by means of a standard intensity strip and corrected for Lorentz and polarization factors. Of the 730 reflections within the Copper sphere, 640 were recorded, 570 of these being strong enough to be observed. The largest portion of the reflections not recorded were those of the form hk4, which were too weak to be of any use on the photographs about the b-axis. The data were put on a single scale using the 280 reflections common to both axes by use of a program written for the CRYRM Crystallographic Computing System by C.M. Gramaccioli. This CRYRM System was written for Caltech's IBM 7094 by several authors, notably David DuChamp, Carlo Gramaccioli, Albert Kendig, and Ivars Ambats. The final values of F_o^2 were an average between the values from each axis, less weight

being given to the reflections about the b-axis. The only reflection coming solely from b-axis data is the 002 reflection. This scaling program gave values for the estimated standard deviation of any one reflection, these sigmas being used as basis for a weighting scheme in later least squares refinement: $\text{sigma}(F^2) = 2.4446 - 0.6398F + 0.2127F^2$, and $\text{sigma}(F) = 0.4121 + 0.0002F + 0.0033F^2$.

The absence of reflections $k+l$ odd in $0kl$ and h odd in $h0l$ indicated the Orthorhombic space group to be $Pna2_1$. Values for a, b, and c were found using data from precision Weisenberg photographs about b and c, taken with CuK_α radiation; $\lambda CuK_{\alpha_1} = 1.54050$Å. and $\lambda CuK_{\alpha_2} = 1.54434$Å. Calculation was done by a least squares fit unit cell parameter program written for the Burroughs 220 computer by Dr. Noel Jones. Eccentricity and absorbtion corrections were made in the calculations of a and b. Absorbtion corrections only were made in the calculation of c. These cell dimensions and their estimated sigmas are:

$$a = 15.551 \pm 0.0015\text{Å}.$$

$$b = 9.3751 \pm 0.0005\text{Å}.$$

$$c = 3.6626 \pm 0.0003\text{Å}.$$

The calculated density of the crystal, assuming four molecules per unit cell was 1.493_5 g./cc., while that obtained from suspension in mixed liquids was 1.491 g./cc.

Determination of the Structure

A trial structure for the unit cell was devised, taking into consideration the shape of the unit cell, which did not allow superposition of the molecules along c, approximate dimensions of the molecule, and opportunities for Hydrogen bonding, especially between the 7 and 9 Nitrogens of adjacent molecules. Using the heavy atom x and y coordinates derived from this trial, the structure factors of 15 reflections were calculated by hand. When these were found to compare favorably with observed structure factors, a Fourier electron density map in the 001 plane, calculated using these 15 reflections, reasonably well confirmed this approximate structure. Structure factors were computed on the Burroughs 220, using the structure factor - least squares program for the Orthorhombic system written by Dr. A. Hybl, and a new Fourier map was made, again using the Burroughs 220 and the Fourier electron density map program written by Dr. K. Hoogsteen. The x and y coordinates of the heavy atoms indicated by this map were then used to begin least squares refinement of the Carbon and Nitrogen x and y coordinates.

Least squares of hk0 data was begun by refining first Carbon and Nitrogen positions, then Carbon and Nitrogen positions with isotropic temperature factors refining, and then heavy atom coordinates and anisotropic temperature factors with the three Hydrogens on C(2), C(6), and C(8) present but not refining. Soon after this point the

weighting function of $\sqrt{w} = 1/(1+F_o)$ which had been used up to then was dropped and $\sqrt{w} = 1/(1+F_o^2)$ was substituted. After several cycles like this, the positions of the Hydrogens were also refined. At this point, after the R factor had dropped from 0.380 to 0.110, a difference Fourier map was calculated which rather clearly placed the final Hydrogen on N(7). Several more cycles of least squares dropped the R factor to 0.102 and completed the two-dimensional refinement. After estimating the z coordinates of the atoms by considering the original trial structure, least squares on first the positions and then positions and isotropic temperature factors of the heavy atoms was begun in three-dimensions with the Hydrogens present but fixed. At this point all further calculations involved in the refinement were entrusted to the CRYRM computing system. The weighting function adopted was $\sqrt{w}$ equals the reciprocal of the sigma(F_o) given by Carlo Gramaccioli's data scaling program. Subsequently, anisotropic temperature factors for the Carbons and Nitrogens were allowed to refine for a few cycles. The weighting function was changed to $\sqrt{w} = 1/\text{sigma}(F_o^2)$ and after several more least squares cycles, the positions of the Hydrogen atoms were allowed to refine. Refinement was considered completed when the R factor for the 568 observed reflections of non-zero weight had been reduced to 0.070 and no parameter shift exceeded 35% of its standard deviation; indeed most shifts were considerably smaller than this. Unobserved reflections were entered into the least

squares calculations only when their calculated value exceeded the minimum observable value. During the final stages of this refinement a difference Fourier map in the plane of the nine heavy atoms was calculated both with and without the Hydrogens present. The first of these, shown in Fig. 1, again indicates the positions of the Hydrogen atoms to be where we had put them. The other showed merely a minimal amount of noise with no systematic peaks which might indicate an occasional flipping of the molecule in the crystal lattice or disorder with respect to H(13). A final Fourier map in the plane of the molecule is shown in Fig. 2, and a Fourier map in projection down c is shown in Fig. 3.

Accuracy of the Results

The value of the quantity known as goodness of fit,

$$\left(\frac{\Sigma w(\Delta F^2)^2}{m-s}\right)^{1/2},$$

which is the standard deviation of an observation of unit weight, is 0.76. Although this quantity is usually slightly larger than unity at the end of refinement, this unusually small value can be explained by the fact that as least squares refinement progressed, several large reflections were calculating considerably larger than they had been observed. As this indicated that extinction had occurred, these reflections were given an external weight equal to zero and refinement was continued without them. Unfortunately the data scaling program, in calculating the weighting functions, had taken the probable inaccuracies of these values into account, hence their absence left the goodness of fit too small. Indeed, adding in the missing $w(\Delta F^2)^2$'s brought this value up to almost exactly unity.

The standard deviations in the positional parameters were calculated from the diagonal terms of the inverse matrix of the normal equations. The average standard deviation of the heavy atom positions is 0.004Å., leading to 0.0057Å. for the standard deviation of interatomic distances involving only heavy atoms. The limit of error, chosen to be three times the standard deviation, is then 0.017Å. for heavy atom bond lengths. The limit of error in bond angles involving only heavy atoms is 0.9°. For bonds involving Hydrogen atoms, the

limits of error in bond length and bond angle are 0.14Å. and 4.7° respectively.

The standard deviations of the temperature factor parameters of the heavy atoms were calculated from the diagonal coefficients of the inverse matrices of the normal equations.

Discussion of the Structure

The Intermolecular Environment:

The structure of the Purine unit cell viewed along its c-axis is shown in Fig. 4. As may be seen, Hydrogen bonding takes place between N(7) of one molecule and N(9) of the adjacent molecule. These Hydrogen bonded strips of molecules running parallel to the 100 plane crisscross with adjacent strips to form an interwoven lattice. The Hydrogen bond length is 2.85Å. The other close contacts between molecules are the distances of 3.50Å. and 3.48Å. between C(2) and N(3) of one molecule and N(3) and C(2) respectively of an adjacent molecule. The molecules are tilted 22.8° from parallel to the 001 plane, the interplanar distance being 3.38Å.

Geometry of the Molecule:

The equation for the best least squares plane through the nine heavy atoms is

$$0.1861X - 0.3390Y + 0.9221Z = -0.4837\text{Å}.,$$

where the coefficients are direction cosines relative to the crystallographic axes. The planarity of the heavy atoms is, in general, pretty good. The average deviation from the plane is 0.006Å. and the greatest deviation is 0.015Å., this being N(7) which is bent down to form a Hydrogen bond. These deviations from the plane give the molecule a slightly curved shape as if its edges are bent away from a line running nearly through C(6) and N(9).

The lengths of the bonds and their respective angles are shown in both Fig. 5 and Table 1. These observed bond lengths can almost be explained by considering the canonical resonance structures of Purine shown in Fig. 6, and assuming that each structure contributes the indicated percentage to the actual structure. The bond lengths calculated from this combination of these structures are shown in the final drawing in Fig. 6. The calculated bond lengths of the two C-C bonds with bond numbers 1.62 and 1.38 were gotten by extrapolation from Table 7-9 of Pauling (1960). The lengths of the C-N bonds with bond numbers 1.50, 1.38, 1.24, 1.54, 1.46, and 1.0 were obtained from a curve drawn through points representing observed C-N bond distances for bond numbers 1.0, 1.5, and 2.0. This curve comes from Acta Cryst. (1962) 15, 310, Marsh et al. Cursory examination shows that bonds N(1)-N(2) and C(5)-N(7) calculate too short and too long respectively. This may be evidence that several resonance forms with negative charges on Carbon atoms are also present among the more likely ones. These forms are normally considered to be too high energy to compete with those shown; they would, however, give more single bond character to N(1)-N(2) and more double bond character to C(5)-N(7), making them calculate nearer the observed values.

The Temperature Factors

The Semi-major axes of the atomic thermal vibration ellipsoids with their respective direction cosines were calculated from the anisotropic temperature factors of each of the heavy atoms. These values, along with the values of root mean square displacements are recorded in Table 3. The direction cosines are relative to a set of orthogonal axes x_1, x_2, and x_3. Axis x_1 lies in the plane of the molecule, approximately along the line N(1)-C(5); x_2, also in the plane of the molecule, runs approximately along C(6)-C(3); and x_3 is the molecular plane normal.

A more or less quantitative drawing of these thermal ellipsoids viewing perpendicular to the plane of the molecule is shown in Fig. 7. The relative magnitudes of different directions of vibration are much what one might suspect. There is a definite, small libratory motion, on the order of 3° r.m.s. amplitude. This is small enough that the changes in bond length due to this libration would be on the order of only 0.1%, hence we do not modify our final values of bond lengths. As might be predicted, the smallest vibrations are those parallel to the line of the hydrogen bonds. The vibrations along the length of the molecule calculate the largest and those perpendicular to the plane are nearly as large. These are indeed seen to be the directions in which the molecule has the greatest degree of freedom.

Conclusion

The determination of the crystal structure of purine was undertaken as part of a continuing program of research on the structures of compounds related to the nucleic acids. As many of the purines have already been studied, elucidation of the parent compound will help in interpreting the results of these earlier studies and in verifying theoretical work done on purine. The most interesting discovery made in this study is that purine exists in its crystalline state as 7-purine, rather than 9-purine as was previously assumed.

I would like to thank my research advisor, Dr. Richard E. Marsh, for his untiring help and advice, and the National Science Foundation, whose financial help made this work possible.

Bibliography

Marsh, DuChamp, Gramaccioli, Kendig, Ambats. The CRYRM Crystallographic Computing system for the IBM 7094.

N.D. Jones. Least Squares Fit Unit-Cell Parameter Program for the Burroughs 220.

A. Hybl. Orthorhombic Structure Factor- Least Squares Program for the Burroughs 220.

K. Hoogsteen. Fourier Electron Density Map Program for the Burroughs 220.

Marsh, et al. Acta Crystallographica. 15, 310.

Table 1.

Bond distances and angles. The average estimated standard deviations for bonds involving only heavy atoms are 0.006Å. and 0.3°, and for bonds involving Hydrogen atoms are 0.05Å. and 1.9°.

N(1)-C(2)	1.349Å.	C(2)-N(1)-C(6)	117.5°
N(1)-C(6)	1.330	N(3)-C(2)-H(10)	117.
C(2)-N(3)	1.324	N(3)-C(2)-N(1)	128.5
C(2)-H(10)	0.97	H(10)-C(2)-N(1)	114
N(3)-C(4)	1.336	C(4)-N(3)-C(2)	113.4
C(4)-C(5)	1.398	C(5)-C(4)-N(9)	109.6
C(4)-N(9)	1.369	C(5)-C(4)-N(3)	123.1
C(5)-C(6)	1.385	N(9)-C(4)-N(3)	127.3
C(5)-N(7)	1.375	C(6)-C(5)-N(7)	136.0
C(6)-H(11)	0.93	C(6)-C(5)-C(4)	118.5
N(7)-C(8)	1.337	N(7)-C(5)-C(4)	105.5
N(7)-H(13)	0.84	H(11)-C(6)-N(1)	114
C(8)-N(9)	1.313	H(11)-C(6)-C(5)	126
C(8)-H(12)	1.06	N(1)-C(6)-C(5)	119.1
Intermolecular Hydrogen bond length:		C(8)-N(7)-H(13)	127
		C(8)-N(7)-C(5)	106.3
N(7)-N(9)	2.849	H(13)-N(7)-C(5)	127
		N(9)-C(8)-H(12)	125
		N(9)-C(8)-N(7)	114.1
		H(12)-C(8)-N(7)	120
		C(4)-N(9)-C(8)	104.6

Table 2.

Atomic parameters. Coordinates of the atoms in percentages of the unit cell. Standard deviations of the z coordinates were not calculated because of the indeterminacy of these coordinates. Anisotropic temperature factors for each heavy atom.

Atom	Scale	Parameter	Stnd. Dev. x 10^4
N(1)	x	.002738	2.60
	y	.195915	3.76
	z	.039016	-
	B_{11}	.0061201	1.82
	B_{22}	.0128030	4.22
	B_{33}	.1056910	46.3
	B_{12}	-.0000659	4.89
	B_{13}	-.0004846	15.3
	B_{23}	-.0011645	30.0
C(2)	x	.016001	3.27
	y	.331078	4.79
	z	.152285	-
	B_{11}	.0057729	2.28
	B_{22}	.0132719	5.82
	B_{33}	.1172022	69.1
	B_{12}	.0024600	6.06
	B_{13}	.0015858	18.9
	B_{23}	-.0041249	31.7

Atom	Scale	Parameter	Stnd. Dev. x 10^4
N(3)	x	.088917	2.37
	y	.404015	3.57
	z	.159494	-
	B_{11}	.0066173	1.90
	B_{22}	.0102790	3.56
	B_{33}	.0840424	36.3
	B_{12}	.0026863	4.51
	B_{13}	.0011045	14.8
	B_{23}	-.0088429	21.6
C(4)	x	.156301	2.71
	y	.330046	3.75
	z	.035624	-
	B_{11}	.0060340	1.97
	B_{22}	.0080471	4.00
	B_{33}	.0825194	40.4
	B_{12}	.0001287	4.63
	B_{13}	-.0022710	16.5
	B_{23}	-.0072599	22.8
C(5)	x	.150053	2.74
	y	.189737	3.99
	z	-.090755	-
	B_{11}	.0058738	1.90
	B_{22}	.0082287	3.61
	B_{33}	.0732691	42.3

Atom	Scale	Parameter	Stnd. Dev. x 10^4
C(5)	B_{12}	.0003087	4.42
	B_{13}	-.0047297	15.3
	B_{23}	-.0050497	21.4
C(6)	x	.070375	2.98
	y	.123850	4.34
	z	-.085361	-
	B_{11}	.0067937	2.38
	B_{22}	.0097345	4.21
	B_{33}	.1064597	55.5
	B_{12}	-.0021017	5.59
	B_{13}	-.0045461	19.5
	B_{23}	-.0062114	26.3
N(7)	x	.231624	2.33
	y	.153457	3.64
	z	-.201776	-
	B_{11}	.0059913	1.76
	B_{22}	.0099230	3.79
	B_{33}	.0921380	41.6
	B_{12}	.0020228	4.57
	B_{13}	-.0018060	15.2
	B_{23}	-.0156405	22.0
C(8)	x	.281239	3.15
	y	.266800	4.56
	z	-.133142	-

Atom	Scale	Parameter	Stnd. Dev. x 10^4
C(8)	B_{11}	.0054449	2.09
	B_{22}	.0132057	5.27
	B_{33}	.0895567	44.6
	B_{12}	.0002542	5.04
	B_{13}	.0002782	19.8
	B_{23}	-.0034262	29.7
N(9)	x	.239636	2.27
	y	.375865	3.56
	z	.006230	-
	B_{11}	.0059475	1.67
	B_{22}	.0105696	3.68
	B_{33}	.0946416	40.5
	B_{12}	-.0016609	4.46
	B_{13}	-.0030279	15.8
	B_{23}	-.0106179	22.9
H(10)	x	-.003535	28.7
	y	.375388	49.6
	z	.264890	-
	B	3.50	-
H(11)	x	.056483	25.1
	y	.036276	50.1
	z	-.190809	-
	B	3.50	-

Atom	Scale	Parameter	Stnd. Dev. x 10^4
H(12)	x	.346376	35.2
	y	.267733	42.2
	z	-.216435	-
	B	3.50	-
H(13)	x	.245937	27.5
	y	.079750	51.9
	z	-.316148	-
	B	3.50	-

Table 3.

The magnitudes B, root mean square displacements μ, and direction cosines q relative to abc of the principle axes of the thermal vibration ellipsoids.

Atom	Axis i	B_i ($Å^2$)	q_i^1	q_i^2	q_i^3	μ_i (Å)
N(1)	1	1.48	1.000	0.008	-0.011	0.274
	2	1.42	0.008	0.272	0.962	0.268
	3	1.12	-0.011	0.962	-0.272	0.239
C(2)	1	1.59	-0.071	0.231	0.970	0.284
	2	1.49	0.858	0.511	-0.058	0.275
	3	1.05	-0.509	0.828	-0.234	0.321
N(3)	1	1.65	0.946	0.313	0.078	0.289
	2	1.20	0.063	0.059	-0.996	0.246
	3	0.78	-0.317	0.948	0.036	0.199
C(4)	1	1.47	0.998	0.057	-0.006	0.273
	2	1.13	0.002	0.069	0.998	0.239
	3	0.67	-0.057	0.996	-0.069	0.184
C(5)	1	1.46	0.994	0.043	-0.102	0.272
	2	0.97	0.100	0.045	0.994	0.222
	3	0.70	-0.047	0.998	-0.041	0.188
C(6)	1	1.71	0.978	-0.160	-0.132	0.295
	2	1.41	0.145	0.076	0.986	0.267
	3	0.80	0.147	0.984	-0.097	0.201

Atom	Axis i	B_i (Å^2)	a_i^1	a_i^2	a_i^3	μ_i (Å)
N(7)	1	1.54	0.885	0.256	-0.388	0.280
	2	1.30	0.395	0.025	0.918	0.257
	3	0.71	-0.244	0.966	0.079	0.190
C(8)	1	1.32	0.969	0.189	0.157	0.258
	2	1.24	0.104	0.263	-0.959	0.251
	3	1.12	-0.222	0.946	0.235	0.243
N(9)	1	1.48	0.990	-0.125	-0.057	0.273
	2	1.34	-0.042	0.116	-0.992	0.261
	3	0.81	0.131	0.985	0.110	0.203

Table 4.

Crystal data for Purine.

(λ_{CuK_α} = 1.5418 Å.)

Orthorhombic; Space Group $Pna2_1$

$\underline{a}$ = 15.551 ± 0.0015 Å.

$\underline{b}$ = 9.3751 ± 0.0005 Å.

$\underline{c}$ = 3.6626 ± 0.0003 Å.

V = 534.0 Å^3

ρ_o = 1.493_5 g.cm.$^{-3}$

ρ_c = 1.491 g.cm.$^{-3}$

Illustrations

Fig. 1: Fourier electron density difference map in the plane of the molecule. Dotted line is at 0.5 electrons per Å^3. Succeding contours are at 0.5 $e/\text{Å}^3$ intervals.

Fig. 2: Fourier electron density map in the plane of the molecule. Dotted line is at 0.5 $e/\text{Å}^3$, solid contours are at 1 $e/\text{Å}^3$ intervals, starting at 1 $e/\text{Å}^3$.

Fig. 3: Fourier electron density map in projection down the c-axis. Dotted contour is at 0.5 $e/\text{Å}^2$, and solid contours are at 1 $e/\text{Å}^2$, starting at 1 $e/\text{Å}^2$.

Fig. 4: Packing arrangement of purine molecules in the unit cell. Shown in projection down the c-axis.

Fig. 5: Bond lengths and angles in the purine molecule. Approximate errors of these values listed on page 7 of paper.

Fig. 6: Canonical resonance forms contributing to the structure of purine.

Fig. 7: Temperature factor ellipsoids showing the principal modes of vibration of the purine molecule in the crystal.

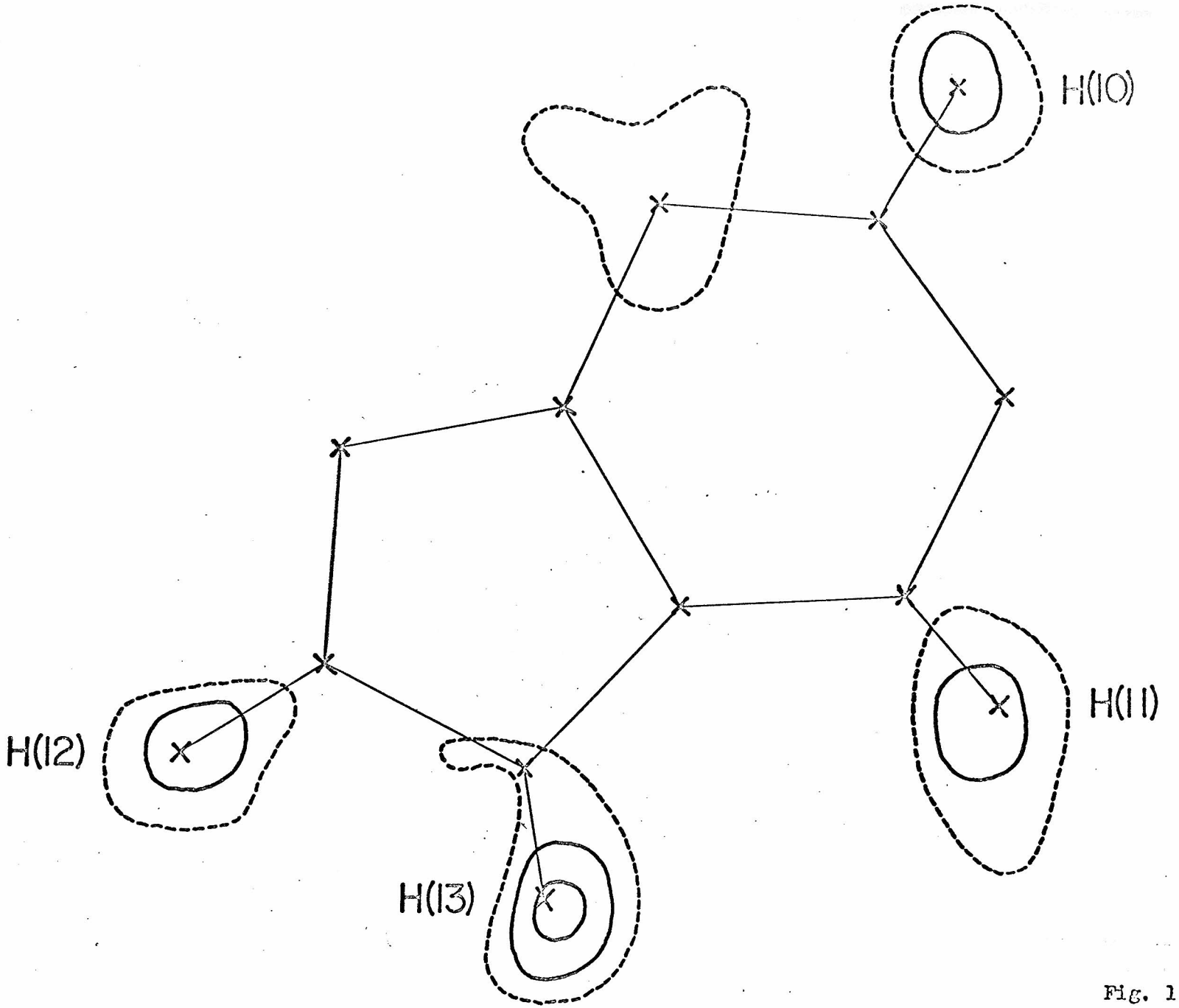


Fig. 1

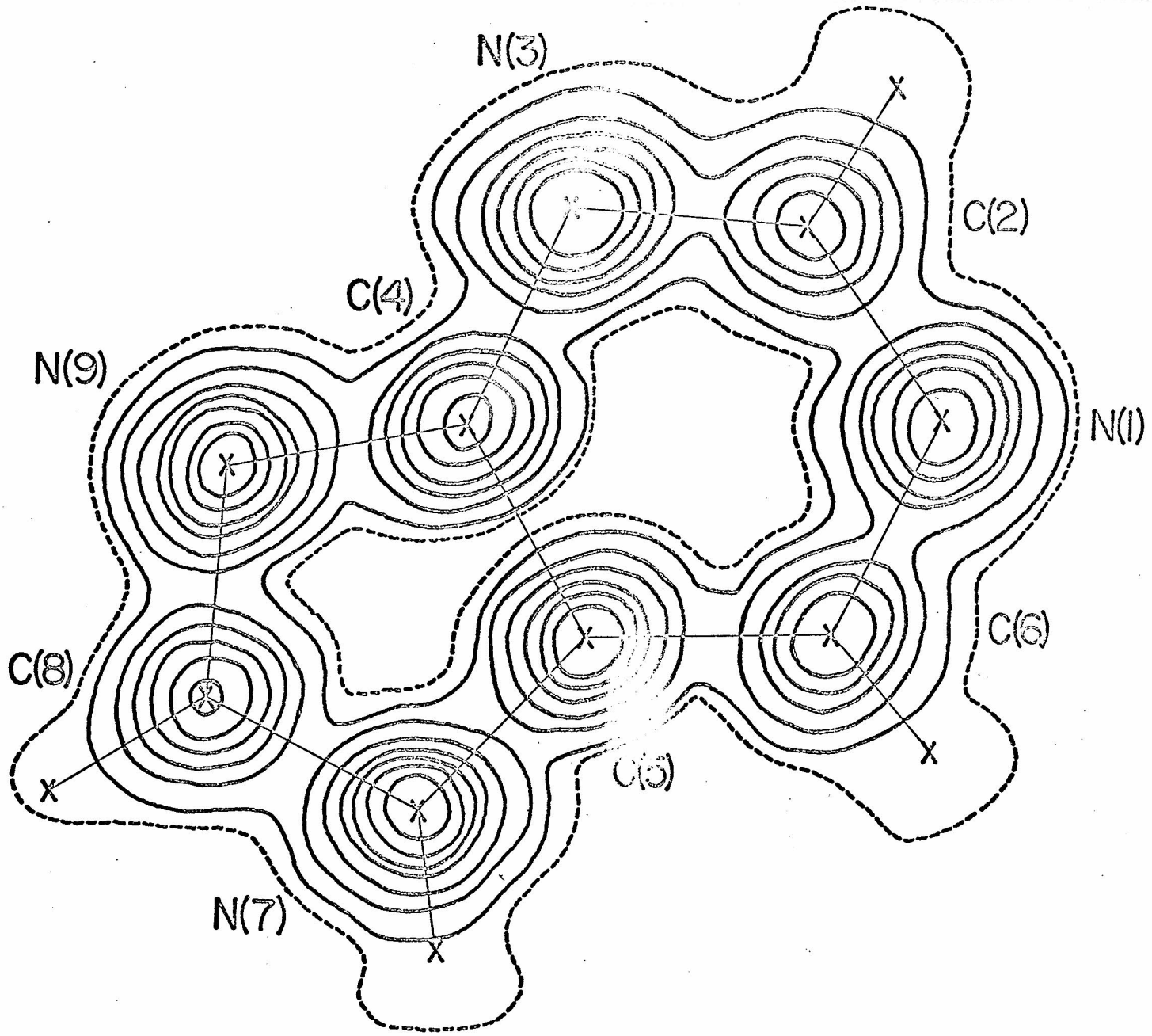


Fig. 2

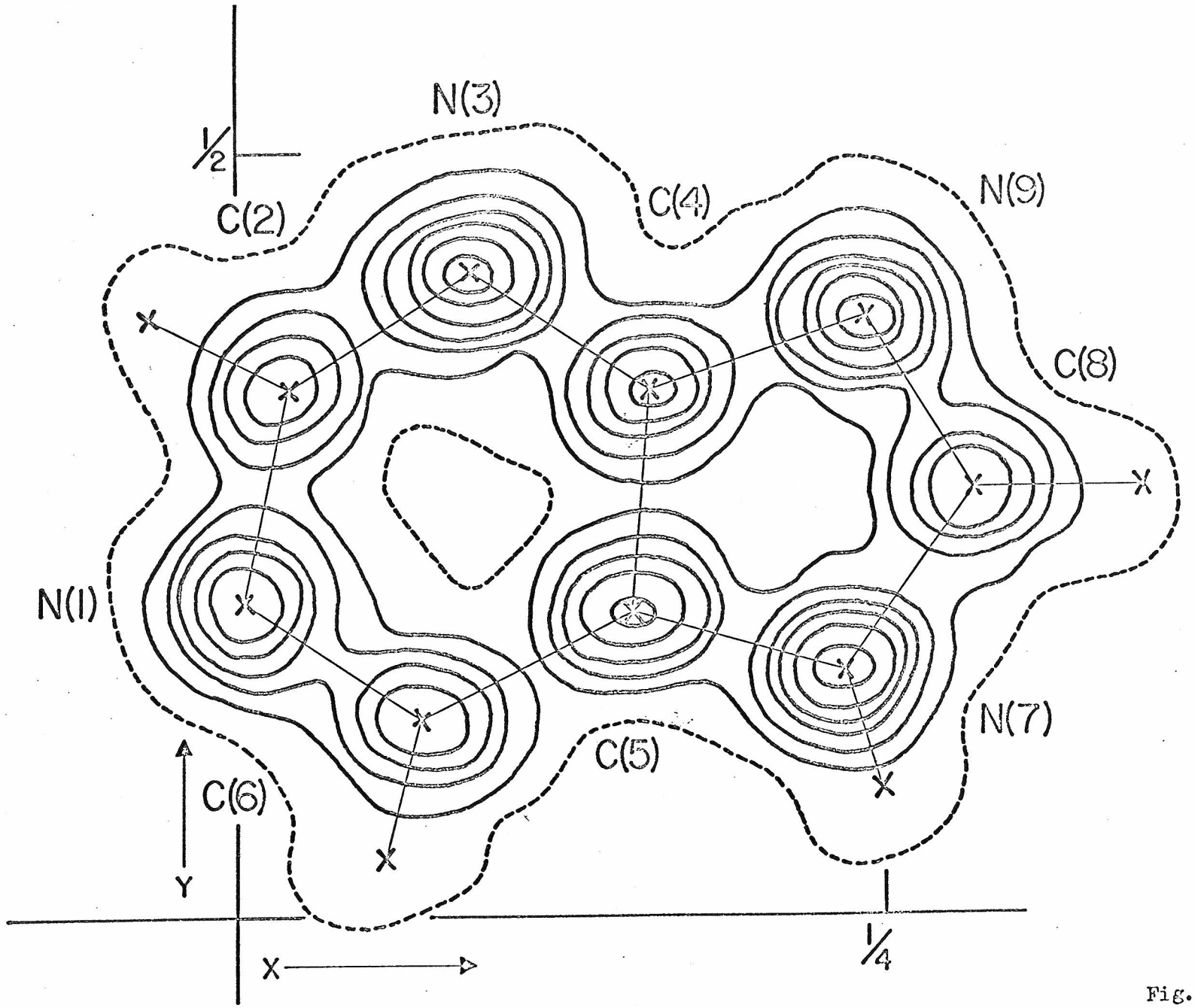


Fig. 3

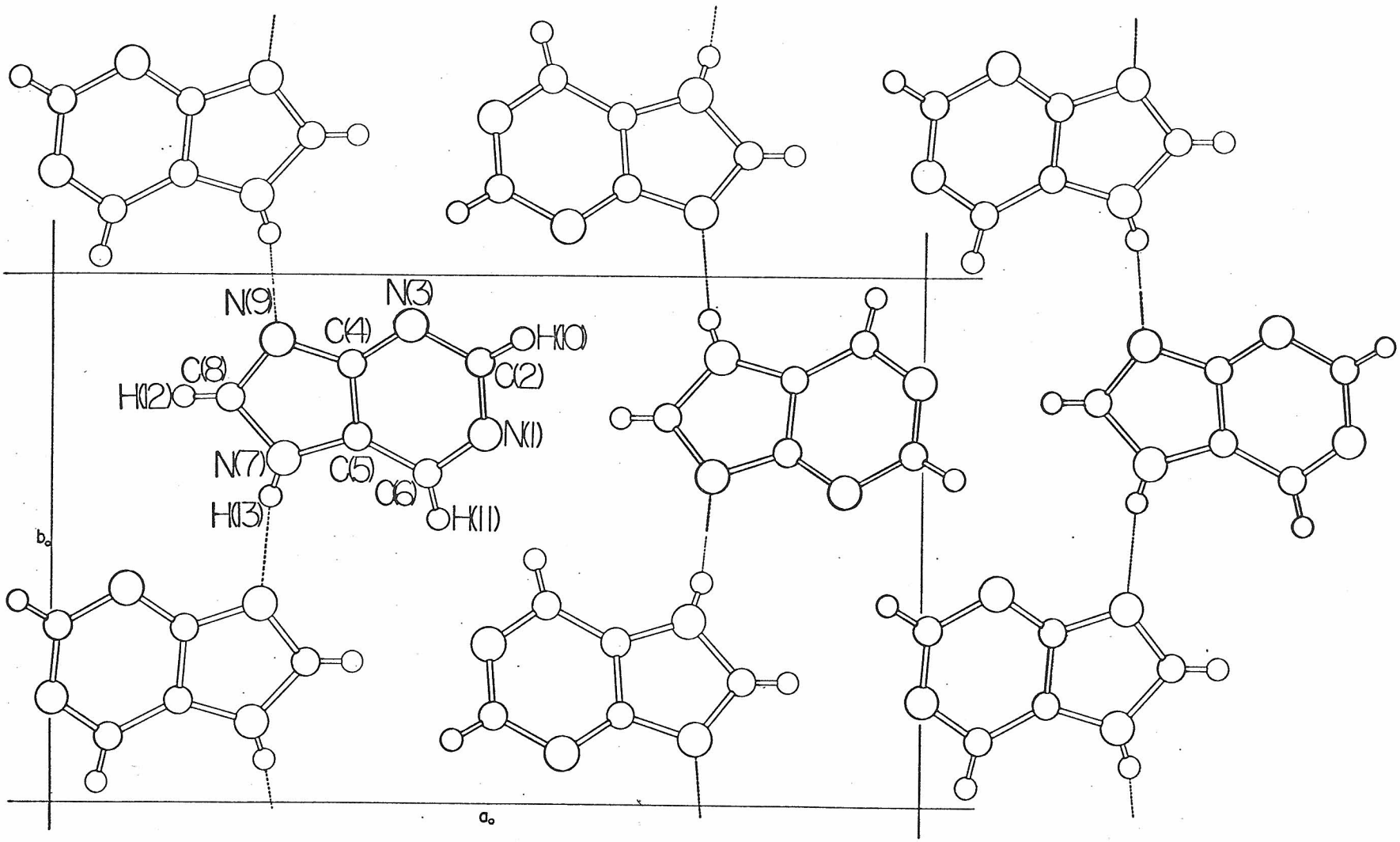


Fig. 4

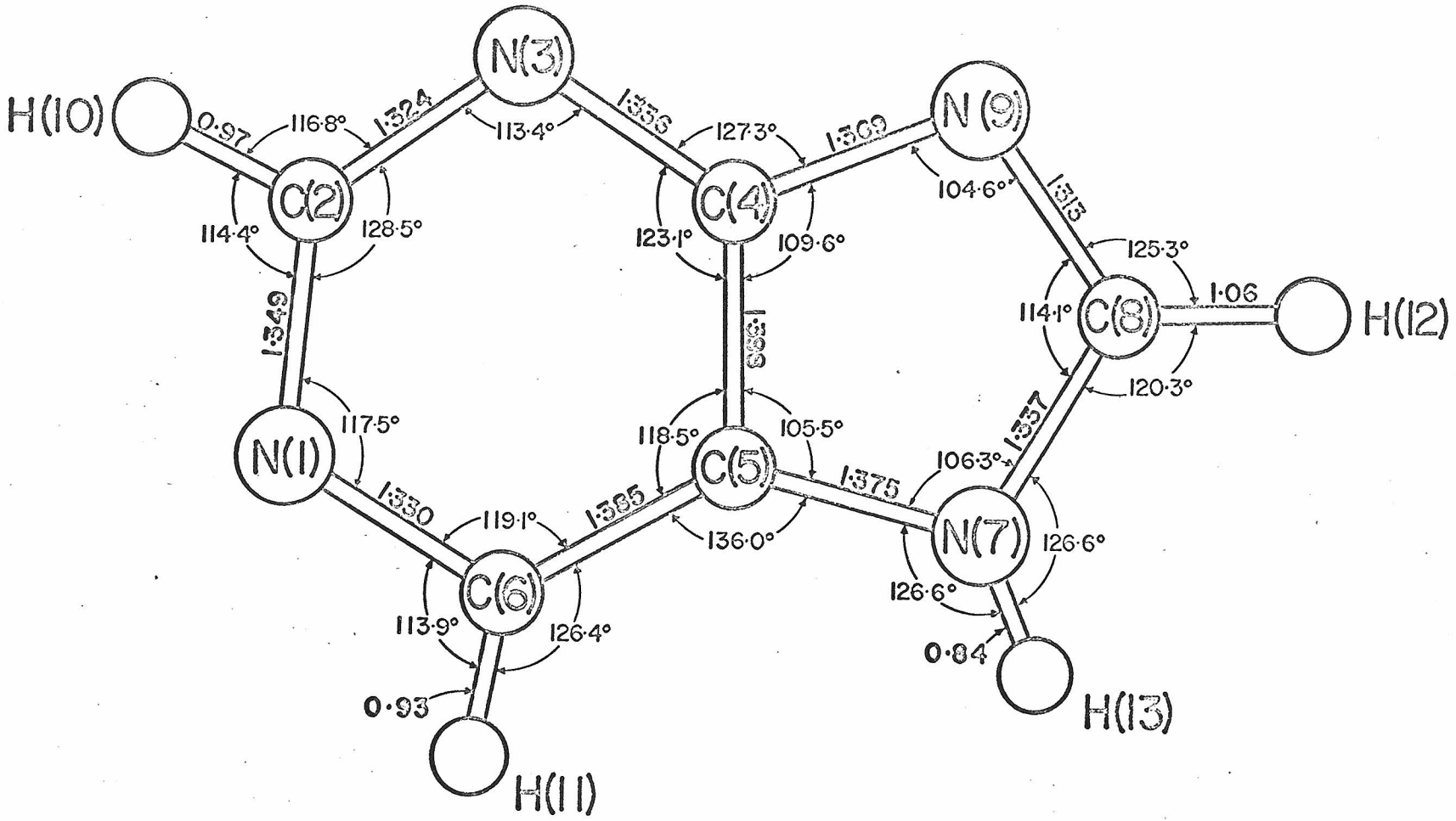
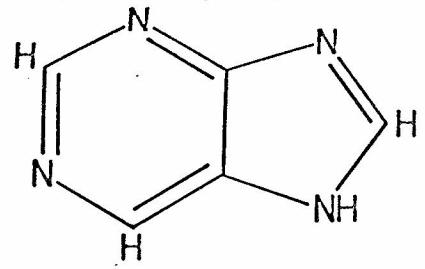
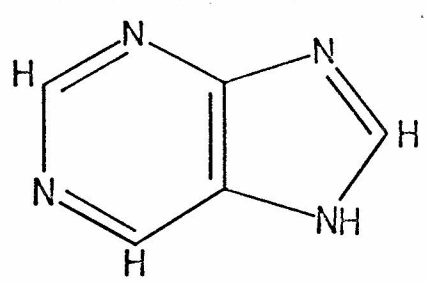

H(10)
0·97
116·8°
1·324
N(3)
113·4°
1·356
127·3°
1·369
N(9)
104·6°
1·313
C(2)
C(4)
114·4°
128·5°
123·1°
109·6°
125·3°
1·06
H(12)
114·1°
C(8)
120·3°
1·343
1·386
1·337
117·5°
118·5°
105·5°
N(1)
C(5)
106·3°
1·330
119·1°
1·365
1·375
N(7)
126·6°
136·0°
C(6)
126·6°
113·9°
126·4°
0·84
0·93
H(13)
H(11)

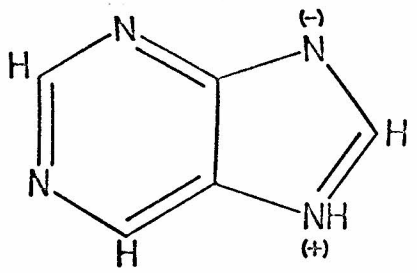
Fig. 5



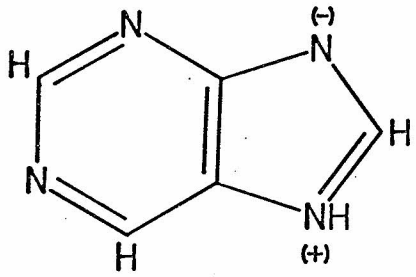
27%



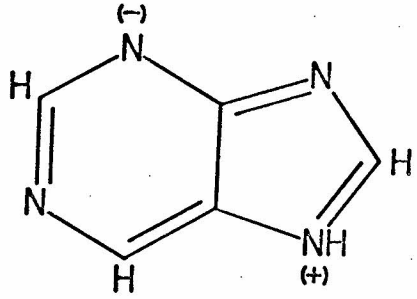
27%



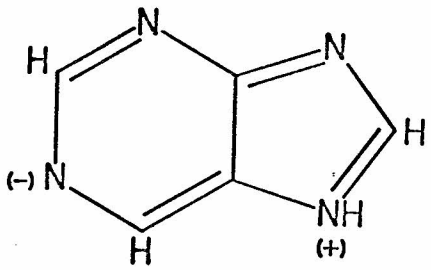
11%



11%



12%



12%

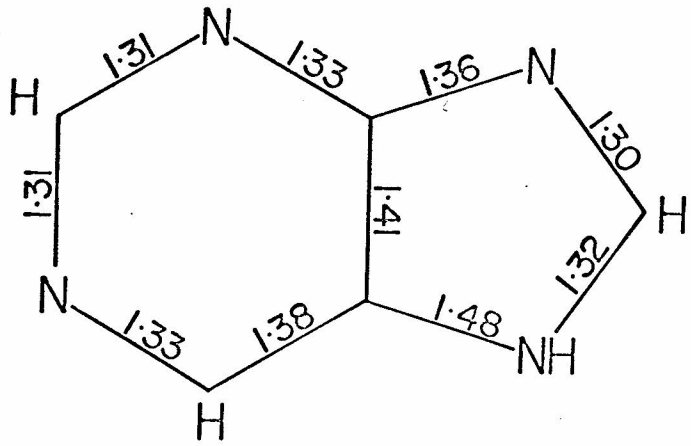


Fig. 6

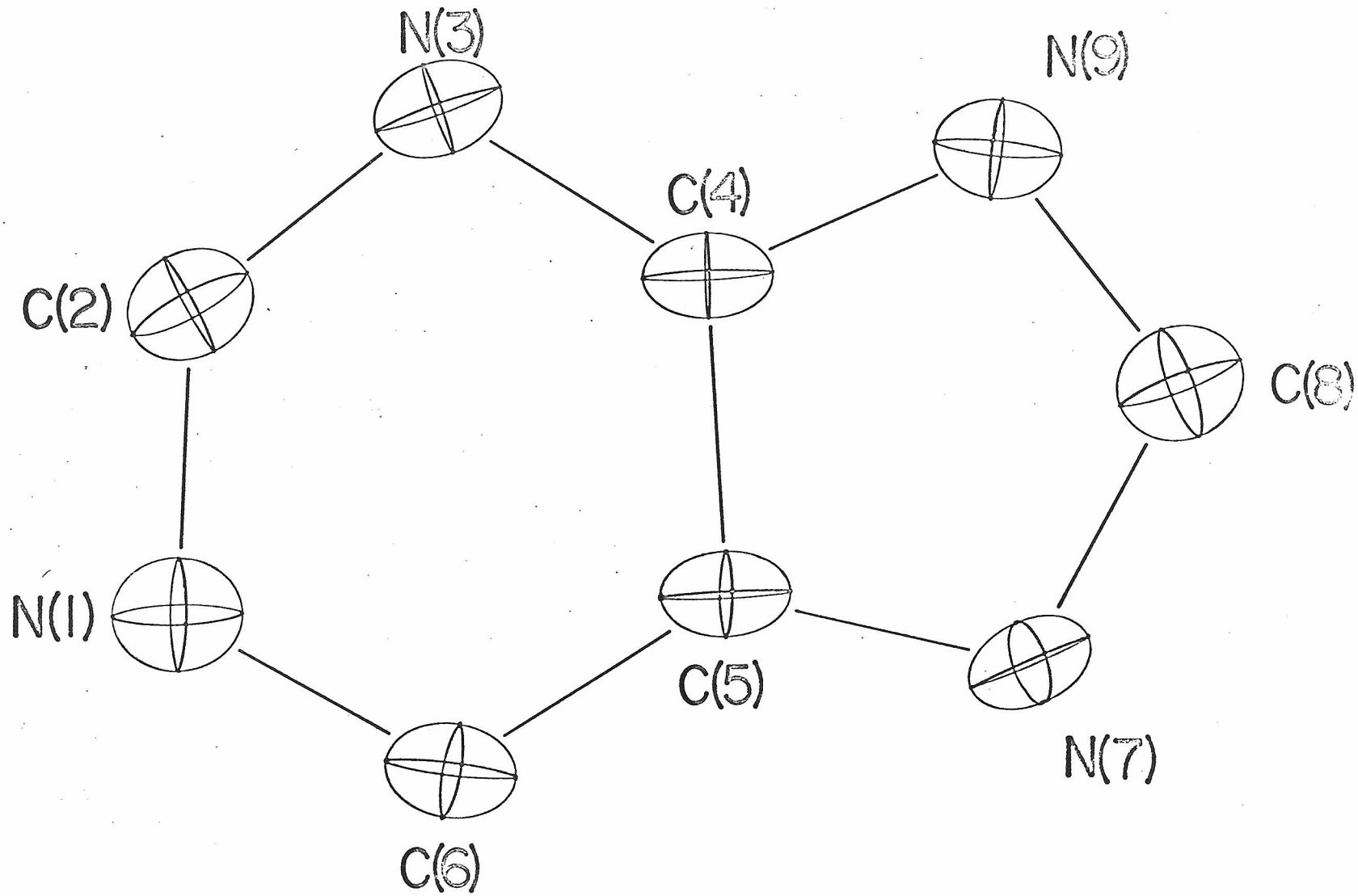


Fig. 7