

## The Crystal Structure of Thiamine Hydrochloride (Vitamin B<sub>1</sub>)

BY J. KRAUT AND H. J. REED

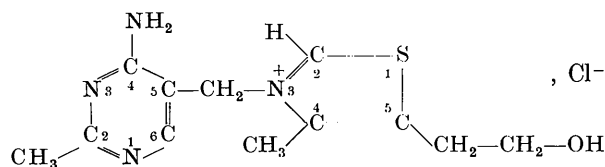
Department of Biochemistry, University of Washington, Seattle, Washington, U.S.A.

(Received 3 October 1961)

The crystal structure of thiamine hydrochloride has been determined by three-dimensional Patterson superposition methods and refined by full-matrix least-squares computations to a final  $R$  factor of 8.0% for 3039 reflections with measurable intensities. The unit cell is monoclinic, space group  $P2_1/c$ , with  $a = 6.99$ ,  $b = 20.59$ ,  $c = 12.73$  Å and  $\beta = 114.0^\circ$ . The locations of 17 out of 20 hydrogen atoms in the asymmetric unit have been determined from a difference-Fourier synthesis and an attempt has been made to refine the coordinates of these hydrogen atoms by least squares. The pyrimidine portion of the molecule is protonated at the ring nitrogen opposite the amino group. The planes of the thiazolium and pyrimidine rings are at a dihedral angle of  $76^\circ$ , and are turned so as to bring the amino group closer to the hydrogen-bearing carbon of the thiazolium ring. The structure contains two weak but distinct C-H...Cl<sup>-</sup> hydrogen bonds.

### Introduction

Thiamine is a required dietary constituent for all animals other than ruminants (White *et al.*, 1959). In man its deficiency is the cause of beriberi. Metabolically, thiamine (as the pyrophosphate) is involved in the decarboxylation of  $\alpha$ -keto acids. The usual structural formula and ring-numbering system given for thiamine is



Thiamine itself is the chloride salt of an organic cation, and may in turn have another proton added to produce the doubly charged acid form, as in thiamine hydrochloride. The pK of this proton addition is about 7 (Viscontini *et al.*, 1951). Although the acid form is perfectly stable, in neutral or alkaline solution thiamine spontaneously undergoes a complicated sequence of intramolecular reactions (Maier & Metzler, 1957).

A great deal has been written concerning possible mechanisms of thiamine catalyzed reactions. Present theories appear to favor a central role for the -CH= group at position 2 of the thiazolium ring. However, in spite of the fact that replacement of the pyrimidine amino group by an OH group causes the loss of catalytic activity, little attention is devoted in these theories to the part played by the pyrimidine portion of the molecule. A concise review of the status of the subject may be found in a recent paper by Pullman & Spanjaard (1961).

The purpose of this investigation was twofold. Primarily, we hoped to be able to obtain structural information which would be of use to biochemists in

understanding the function of this ubiquitous and highly important molecule. A secondary purpose, however, was to gain further insight into the vector-coincidence method for solving complex organic structures. Thiamine hydrochloride contains two chloride ions and a sulfur atom, all three of which should be locatable by examination of the sharpened Patterson function. Each of these heavy atoms may then serve as the basis of an independent Patterson superposition, and each superposition in turn ought to contain a reasonably complete structure, together with some spurious 'noise' peaks. We felt that it would be illuminating to compare the three separate superposition maps, particularly with regard to the locations of these spurious peaks. A similar noise distribution in all three maps would suggest interference, so to speak, from possible 'nearly homometric' structures. If each map were to contain a different noise distribution, on the other hand, one could assume that it originated in essentially random overlapping of Patterson peaks in the superposition. It would consequently be reasonable to expect that higher and higher order superpositions would produce progressively clearer images of the correct structure. Fortunately, as will be seen under *determination of the structure*, the random overlap hypothesis appears to be the correct one.

### Experimental

Thiamine hydrochloride (Merck) was crystallized by evaporation of a solution in a mixture of water and *N,N*-dimethylformamide. A lath-like habit was obtained, with the prominent faces of the form {010} and the long direction of the crystal parallel to the unit-cell  $a$  axis. Systematic absence of  $h0l$  reflections with  $l$  odd and of  $0k0$  reflections with  $k$  odd indicated space group  $P2_1/c$ . Unit-cell parameters were measured both by precession photography and on a General

Electric single-crystal diffractometer, using Cu  $K\alpha$  radiation. A wavelength of 1.5418 Å was assumed for the mean of  $K\alpha_1$  and  $K\alpha_2$ , or 1.54051 Å for  $K\alpha_1$  where the two components could be resolved in high-order reflections. Unit-cell parameters were found to be

$$a = 6.99, b = 20.59, c = 12.73 \text{ \AA}, \beta = 114^\circ 0'.$$

The experimental standard deviations of these parameters, as estimated from the results of several determinations, are 0.01, 0.02, 0.02 Å and 5'.

The crystal density measured by flotation was 1.401 g.cm.<sup>-3</sup>, corresponding to a molecular weight of 353.1 for one asymmetric unit. The molecular weight of a thiamine hydrochloride molecule plus one molecule of water is 355.3 and it was therefore concluded that the compound had crystallized as the monohydrate.

Integrated intensity data were collected on the diffractometer for all 3663 reflections within the Cu  $K\alpha$  limit. The  $2\theta$  scanning technique was used. The number of reflections having intensities measurably above background was 3039, or 83% of the total. A single crystal was used throughout, with dimensions approximately 0.3 × 0.05 × 0.2 mm. Absorption corrections were not made, although it was recognized that these would not be inappreciable. The rather high final  $R$  factor of 8.0% may well be due to this deficiency.

#### Determination of the structure

A sharpened, origin-removed three-dimensional Patterson synthesis was prepared and the distribution of peaks noted. With three heavy atoms in the asymmetric unit, there are 21 different vectors between pairs of heavy atoms in the vector set of this space group. Thorough inspection of the Patterson map eventually yielded all the peaks required for correlation with a selfconsistent family for these 21 vectors, and from these the approximate coordinates of the three heavy atoms could be determined unambiguously.

Each of the heavy atom locations was then used to compute a separate four term minimum-function superposition. These were contoured and the 26–28 highest peaks in the asymmetric unit of each superposition map were selected as candidates for the 21 atomic sites. As an aid to interpretation, a simple three dimensional model of each superposition map was constructed, in which the possible atomic sites were represented by small clay balls. With the help of these models, it proved to be relatively easy to pick out the actual structure. Indeed, most or all of the atoms of the structure could be identified independently in all three models. Specifically, for the superposition based on S1 (see Figs. 1 and 2 for atom designations) all 21 atoms of the structure appeared in their approximately correct locations; for C120 only C8 was missing; for C121 only C5 and O9 did not appear in the model.

For the reasons set forth in the *Introduction*, a comparison was made among the noise peaks in

the three superposition maps, but no coincidences were found. This observation lends support to the notion that the spurious peaks always observed in such superposition maps probably arise through random overlapping of Patterson peaks.

When a final combined superposition map was prepared, consisting of the sum of the three separate minimum functions, the noise was satisfactorily suppressed and the correct structure stood out clearly. Comparison with the final set of refined coordinates showed that no atom was misplaced by more than 0.2 Å in this combined map. The initial  $R$  factor, based on the atomic coordinates derived from this map, was 35%.

#### Refinement

The following atomic scattering factors were used in the refinement: C, N, O and Cl<sup>-</sup> of Berghuis *et al.* (1955), H of McWeeny (1951) and S of Tomiie & Stam (1958).

The refinement was carried out entirely by the full-matrix least-squares method, using Busing & Levy's (1959) program ORXLS adapted to the IBM-709 with the aid of compatibility program IBIOC II. A weighting scheme similar to Hughes' (1941) was employed.

Three cycles of refinement of positional and isotropic temperature factor parameters reduced the  $R$  factor from 35% to 14%. Two further cycles of refinement were run on anisotropic temperature factors only, reducing  $R$  to 9.2%. The last cycle of both series of computations (cycles 3 and 5) indicated only small adjustments in the parameters. At this point the first difference-Fourier synthesis was prepared in order to locate the hydrogen atoms. There are a total of 20 hydrogen atoms in the structure; of these 17 were found readily, comprising as they did the most

Table 1. *Positional parameters of non-hydrogen atoms, in fractional coordinates × 10<sup>4</sup>, and their estimated standard deviations*

	$x/a$	$y/b$	$z/c$
S 1	7668 ± 2	2889 ± 1	3533 ± 1
C 2	7856 ± 7	3632 ± 2	3064 ± 4
N 3	6066 ± 5	3829 ± 2	2231 ± 3
C 4	4477 ± 7	3361 ± 2	1921 ± 4
C 5	5105 ± 7	2819 ± 2	2567 ± 4
C 6	2405 ± 8	3495 ± 2	986 ± 5
C 7	3820 ± 8	2237 ± 2	2571 ± 5
C 8	2307 ± 10	2399 ± 3	3121 ± 6
O 9	3315 ± 7	2673 ± 2	4233 ± 4
C10	5725 ± 7	4472 ± 2	1671 ± 4
N11	10288 ± 6	5555 ± 2	2198 ± 3
C12	10585 ± 7	5834 ± 2	3197 ± 4
N13	9420 ± 6	5707 ± 2	3755 ± 3
C14	7838 ± 7	5264 ± 2	3303 ± 4
C15	7489 ± 7	4937 ± 2	2246 ± 4
C16	8760 ± 7	5101 ± 2	1721 ± 4
C17	12332 ± 9	6312 ± 3	3675 ± 5
N18	6697 ± 7	5151 ± 2	3889 ± 4
O19	7410 ± 9	2091 ± 2	5698 ± 5
C120	2907 ± 2	4187 ± 1	3715 ± 1
C121	7140 ± 2	845 ± 1	4191 ± 1

Table 2. *Thermal parameters of non-hydrogen atoms and their estimated standard deviations* $\beta$  as given here is defined by:

$$T = \exp \{ -10^{-4} (\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + 2\beta_{12}hk + 2\beta_{13}hl + 2\beta_{23}kl) \}$$

		$\beta_{11}$	$\beta_{22}$	$\beta_{33}$	$\beta_{12}$	$\beta_{13}$	$\beta_{23}$
S 1		165 ± 3	16 ± 0	74 ± 1	7 ± 1	32 ± 1	9 ± 0
C 2		152 ± 10	18 ± 1	65 ± 4	0 ± 3	32 ± 5	-1 ± 2
N 3		145 ± 8	13 ± 1	50 ± 3	0 ± 2	32 ± 4	-3 ± 1
C 4		176 ± 10	12 ± 1	62 ± 3	-4 ± 2	38 ± 5	-5 ± 1
C 5		177 ± 11	14 ± 1	71 ± 4	1 ± 3	48 ± 5	-2 ± 2
C 6		198 ± 12	18 ± 1	75 ± 4	-6 ± 3	14 ± 6	-4 ± 2
C 7		236 ± 13	13 ± 1	96 ± 5	-11 ± 3	51 ± 7	-2 ± 2
C 8		263 ± 15	23 ± 1	104 ± 5	-22 ± 4	71 ± 8	-1 ± 2
O 9		377 ± 15	23 ± 1	105 ± 4	-3 ± 3	100 ± 6	5 ± 2
C 10		193 ± 11	12 ± 1	54 ± 3	-2 ± 3	23 ± 5	2 ± 1
N 11		213 ± 10	12 ± 1	60 ± 3	-5 ± 2	54 ± 5	2 ± 1
C 12		182 ± 10	10 ± 1	57 ± 3	1 ± 2	38 ± 5	1 ± 1
N 13		199 ± 10	14 ± 1	56 ± 3	-6 ± 2	47 ± 4	-5 ± 1
C 14		179 ± 10	13 ± 1	53 ± 3	0 ± 2	42 ± 5	-1 ± 1
C 15		188 ± 11	11 ± 1	58 ± 3	-3 ± 2	44 ± 5	1 ± 1
C 16		225 ± 12	13 ± 1	64 ± 3	-4 ± 3	59 ± 5	0 ± 1
C 17		259 ± 14	20 ± 1	83 ± 5	-27 ± 4	47 ± 7	-6 ± 2
N 18		244 ± 11	19 ± 1	69 ± 3	-16 ± 3	75 ± 5	-7 ± 1
O 19		449 ± 18	26 ± 1	121 ± 5	-1 ± 4	95 ± 6	-7 ± 2
Cl 20		171 ± 3	17 ± 0	64 ± 1	-5 ± 1	45 ± 1	-6 ± 0
Cl 21		239 ± 3	16 ± 0	62 ± 1	7 ± 1	58 ± 1	-1 ± 0

prominent positive regions of the difference map, with an average peak height of 0.6 e.Å<sup>-3</sup>. A rather elongated positive region, of height 0.7 e.Å<sup>-3</sup>, and located about 1.5 Å from O9, was originally thought to represent the hydrogen atom attached to O9, and a hydrogen atom was consequently carried along at this position in the subsequent refinement. In the end, however, it was abandoned on general stereochemical grounds. No satisfactory indications of the two hydrogen atoms attached to the water oxygen, O19, were found at this stage.

The hydrogen atoms were next introduced into the structure with positional parameters derived from the difference map, given isotropic temperature factors of  $B = 3.0 \text{ \AA}^2$ , and a further round (cycle 6) of refinement computed on the positional parameters of the non-hydrogen atoms. The calculated corrections were small and the  $R$  factor dropped only slightly, to 8.5%. At this point we were originally prepared to terminate the refinement. However, the following question arose. Since the anisotropic temperature factor parameters were last adjusted before the hydrogen atoms had been introduced, and since the thermal parameters presumably tended to compensate for the missing hydrogen atoms, would not further refinement result in appreciable changes in both thermal parameters and hydrogen atom positions?

Largely in order to determine the extent of this effect, further refinement was carried out as follows. The hydrogen atoms were assigned the anisotropic temperature factors of the atoms to which they were covalently bonded, and six more cycles (7 through 12) of least-squares refinement computed, in which the parameters adjusted were, in order, non-hydrogen temperature factors, non-hydrogen positions, non-hydrogen temperature factors again, and three cycles of hydrogen atom positions. The latter tended to

oscillate somewhat, but with diminishing amplitude in succeeding cycles of refinement. The mean final adjustment in each type of parameter in terms of their standard deviations was 0.35 $\sigma$  for non-hydrogen atom positional parameters (cycle 8), 0.16 $\sigma$  for non-hydrogen atom thermal parameters (cycle 9) and 0.31 $\sigma$  for hydrogen atom positional parameters (cycle 12).

The final anisotropic thermal parameters were not appreciably different from the same parameters as they stood at the end of cycle 6; the average change was 0.4 $\sigma$  and the largest change (for  $\beta_{33}$  of C14) was 1.7 $\sigma$ .

In order to get some idea for the efficacy of the least-squares refinement of the hydrogen atom positional parameters in cycles 10–12, C–H bond lengths were computed for the two methyl groups and for the three methylene groups using the original parameters taken directly from the difference map. These

Table 3. *Positional parameters of hydrogen atoms, in fractional coordinates  $\times 10^3$ , and their estimated standard deviations*

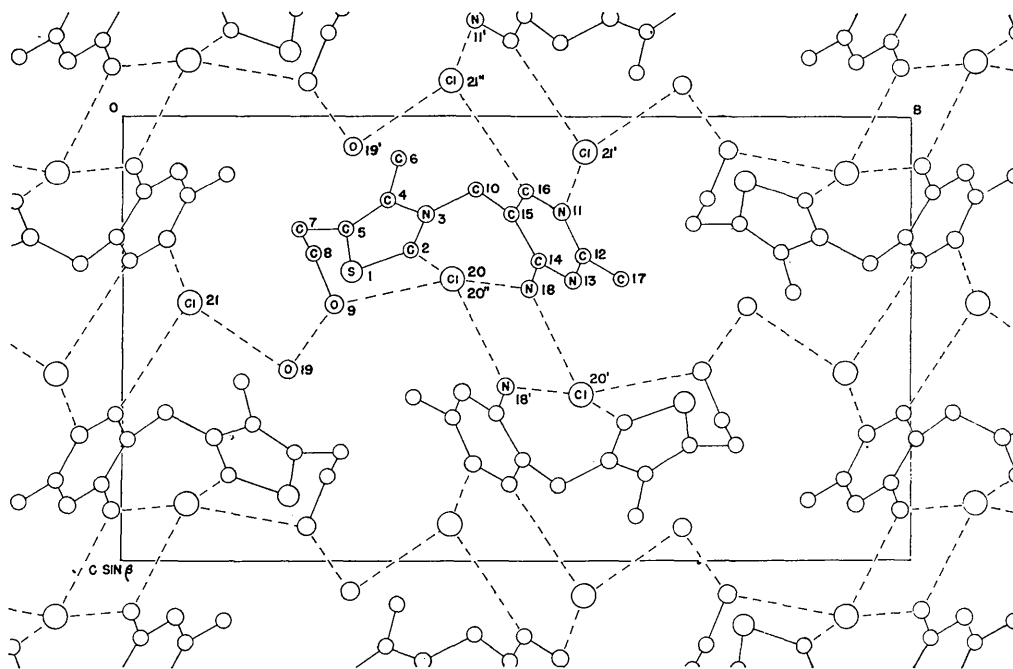
	$x/a$	$y/b$	$z/c$
H23	108 ± 12	275 ± 4	253 ± 7
H24	158 ± 12	182 ± 4	326 ± 7
H25	300 ± 12	210 ± 3	170 ± 7
H26	473 ± 11	192 ± 3	302 ± 6
H27	164 ± 11	385 ± 4	121 ± 6
H28	258 ± 11	349 ± 4	35 ± 6
H29	149 ± 11	312 ± 4	72 ± 6
H30	916 ± 10	381 ± 3	335 ± 6
H31	432 ± 10	458 ± 3	169 ± 5
H32	555 ± 11	435 ± 3	85 ± 6
H33	868 ± 10	488 ± 3	92 ± 6
H34	1096 ± 11	563 ± 3	189 ± 6
H35	1293 ± 11	642 ± 4	313 ± 6
H36	1162 ± 12	677 ± 4	379 ± 6
H37	1332 ± 12	619 ± 4	442 ± 7
H38	698 ± 11	533 ± 3	455 ± 6
H39	561 ± 11	489 ± 4	364 ± 6



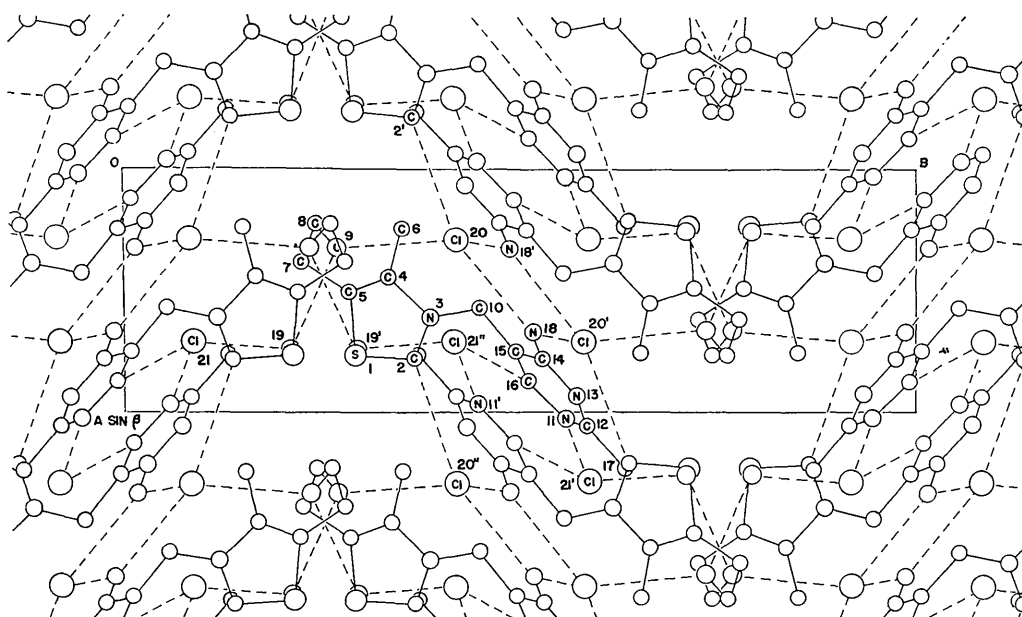
Table 4 (cont.)

Table with multiple columns of numerical data, including integers and floating-point numbers, organized in a grid-like structure with various sub-sections and labels.





(a)



(b)

Fig. 2. Packing of molecules in the crystal structure of thiamine hydrochloride. (a) Orthographic projection along *a* axis. (b) Orthographic projection along *c* axis.

the *b* axis, and approximately  $1 \text{ e.}\text{\AA}^{-3}$  deep. Correspondingly, the least-squares  $F_o$  scale constant, as recomputed separately for each level of *k*, showed a distinct trend, reaching a minimum of 0.93 at  $k=10$  and rising on either side to 1.13 for  $k=0$  and 1.17 for

$k=25$ . We believe that this behavior is due to neglect of the absorption correction, which, owing to the crystal shape and method of data collection, should have been a strong function of the *k* index but only weakly dependent on *h* and *l*.

### Results and discussion

Fig. 1 shows a single molecule of thiamine hydrochloride in orthographic projection along the direction [0·5725, -0·0528, 0·1671]. The 17 hydrogen atoms located in the first difference-Fourier synthesis are shown with their final positional parameters as computed by the least-squares refinement. Hydrogen atoms not actually found in the difference map are omitted, although their approximate locations can be inferred from the distribution of hydrogen bonds, indicated by dashed lines. The *b* axis in Fig. 1 is horizontal, and both the *a* and *c* axes are pointing obliquely away from the observer.

Fig. 2 shows the packing of molecules in the crystal

Table 5. Bond lengths of covalent bonds not involving hydrogen atoms, and their estimated standard deviations

Bond	Length
S 1-C 2	1·667 ± 0·006 Å
S 1-C 5	1·718 ± 0·006
C 2-N 3	1·332 ± 0·007
N 3-C 4	1·401 ± 0·007
C 4-C 5	1·349 ± 0·007
C 4-C 6	1·479 ± 0·008
C 5-C 7	1·499 ± 0·008
C 7-C 8	1·524 ± 0·010
C 8-O 9	1·417 ± 0·009
N 3-C 10	1·476 ± 0·007
C 10-C 15	1·496 ± 0·008
N 11-C 12	1·333 ± 0·007
C 12-N 13	1·306 ± 0·007
N 13-C 14	1·367 ± 0·007
C 14-C 15	1·434 ± 0·007
C 15-C 16	1·354 ± 0·007
N 11-C 16	1·362 ± 0·007
C 12-C 17	1·492 ± 0·008
C 14-N 18	1·316 ± 0·007

Table 6. Bond angles for covalent bonds not involving hydrogen atoms, and their estimated standard deviations

Angle	Degrees
C 2-S 1-C 5	91·8 ± 0·3
S 1-C 2-N 3	112·3 ± 0·4
C 2-N 3-C 4	113·3 ± 0·4
C 2-N 3-C 10	124·9 ± 0·4
C 4-N 3-C 10	121·8 ± 0·4
N 3-C 4-C 5	111·9 ± 0·4
N 3-C 4-C 6	120·2 ± 0·4
C 5-C 4-C 6	127·9 ± 0·5
S 1-C 5-C 4	110·7 ± 0·4
S 1-C 5-C 7	121·1 ± 0·4
C 4-C 5-C 7	128·1 ± 0·5
C 5-C 7-C 8	110·8 ± 0·5
C 7-C 8-O 9	113·0 ± 0·5
N 3-C 10-C 15	113·5 ± 0·4
C 12-N 11-C 16	120·7 ± 0·4
N 11-C 12-N 13	122·9 ± 0·4
N 11-C 12-C 17	117·2 ± 0·4
N 13-C 12-C 17	119·9 ± 0·4
C 12-N 13-C 14	118·8 ± 0·4
N 13-C 14-C 15	120·5 ± 0·4
N 13-C 14-N 18	116·9 ± 0·4
C 15-C 14-N 18	122·5 ± 0·4
C 14-C 15-C 16	116·8 ± 0·4
C 10-C 15-C 16	119·6 ± 0·4
C 10-C 15-C 14	123·5 ± 0·4
N 11-C 16-C 15	120·3 ± 0·4

structure. Hydrogen atoms are omitted for clarity. Hydrogen bonds are indicated by dashed lines. Fig. 2(a) represents an orthographic projection along the direction of the *a* axis; Fig. 2(b) is an orthographic projection along the direction of the *c* axis. The labeled molecule is the same one shown in Fig. 1.

Bond lengths and angles for covalent bonds not involving hydrogen atoms are given in Tables 5 and 6. Hydrogen-bond lengths and angles are listed in Tables 7 and 8. Table 9 contains the bond lengths of covalent bonds involving hydrogen atoms.

### General features

Somewhat unexpectedly, the extra proton has been found on N11 opposite the amino group of the pyrimidine ring. This is in contrast to adenine hydrochloride, where the extra proton appears on the ring nitrogen adjacent to the amino group of the pyrimidine

Table 7. Hydrogen-bond lengths and their estimated standard deviations

Bond	Length
N 18-C120	3·245 ± 0·008 Å
H 39-C120	2·4 ± 0·1
N 18-C120'	3·248 ± 0·008
H 38-C120'	2·4 ± 0·1
N 11-C121'	3·050 ± 0·007
H 34-C121'	2·3 ± 0·1
C 16-C121''	3·533 ± 0·009
H 33-C121''	2·5 ± 0·1
C 2-C120''	3·475 ± 0·009
H 30-C120''	2·6 ± 0·1
O 9-C120	3·175 ± 0·009
O 9-O 19	2·961 ± 0·010
O 19-C121	3·163 ± 0·009

Table 8. Hydrogen-bond angles and their estimated standard deviations

Angle	Degrees
C 14 -N 18 -C120	137·4 ± 0·3
C 14 -N 18 -C120'	128·2 ± 0·3
C120 -N 18 -C120'	94·4 ± 0·1
H 39 -N 18 -C120	16 ± 5
H 38 -N 18 -C120'	8 ± 5
C 12 -N 11 -C121'	124·7 ± 0·3
C 16 -N 11 -C121'	114·6 ± 0·3
H 34 -N 11 -C121'	2 ± 5
N 11 -C 16 -C121''	133·4 ± 0·3
C 15 -C 16 -C121''	106·3 ± 0·3
H 33 -C 16 -C121''	18 ± 5
S 1 -C 2 -C120''	115·9 ± 0·2
N 3 -C 2 -C120''	128·6 ± 0·3
H 30 -C 2 -C120''	10 ± 5
C 8 -O 9 -C120	102·5 ± 0·3
C 8 -O 9 -O 19	115·4 ± 0·4
O 19 -O 9 -C120	120·7 ± 0·3
O 9 -O 19 -C121	98·2 ± 0·3
O 9 -C120 -N 18	125·8 ± 0·2
O 9 -C120 -N 18'	104·5 ± 0·2
O 9 -C120 -C 2'	73·9 ± 0·2
N 18 -C120 -N 18'	85·6 ± 0·1
N 18 -C120 -C 2'	160·0 ± 0·2
N 18' -C120 -C 2'	93·0 ± 0·2
C 16 -C121'' -O 19'	88·6 ± 0·2
C 16 -C121'' -N 11'	112·0 ± 0·2
O 19' -C121'' -N 11'	127·3 ± 0·2



Table 9. Bond lengths of covalent bonds involving hydrogen atoms

Standard deviations are approximately 0.1 Å

Bond	Length	Bond	Length
C 8-H 23	1.1 Å	C 10-H 31	1.0 Å
C 8-H 24	1.3	C 10-H 32	1.0
C 7-H 25	1.1	C 16-H 33	1.1
C 7-H 26	0.9	N 11-H 34	0.7
C 6-H 27	1.0	C 17-H 35	1.0
C 6-H 28	0.9	C 17-H 36	1.1
C 6-H 29	1.0	C 17-H 37	1.0
C 2-H 30	0.9	N 18-H 38	0.9
		N 18-H 39	0.9

portion (Cochran, 1951). A possible explanation for this behavior in terms of qualitative resonance theory is given below.

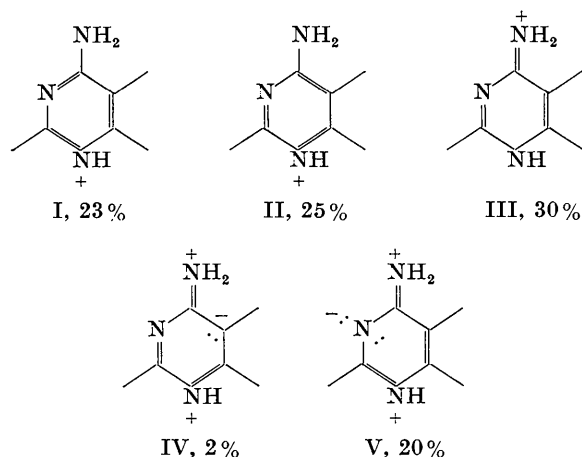
Both the thiazolium ring and the pyrimidine ring are planar; for both the r.m.s. deviation of ring atoms from the least-squares plane is 0.007 Å, and the maximum deviation is 0.011 Å. In addition, the pyrimidine ring, attached atoms C10, C17 and N18, and the hydrogen-bonded chloride ions Cl20, Cl20', Cl21' and Cl21'' are approximately coplanar, with no atom displaced by more than 0.09 Å from the least-squares plane. The normal to this plane is approximately parallel to the reciprocal lattice direction [1, 4.5, 1] and consequently the reflections 1, 4, 1 and 1, 5, 1 are two of the strongest, comparable only to 2, 0, 0.

The planes of the thiazolium and pyrimidine rings are at a dihedral angle of 76° and are oriented so that the amino group of the pyrimidine could readily interact with C2 of the thiazolium group. Experiments with commercial models suggest that rotation of the two rings with respect to one another is not completely unhindered, although other conformations are certainly possible. In these, however, C2 and N18 are less favorably disposed for interaction. Assuming that the conformation found in the crystal is also the most stable in solution, it is tempting to conjecture that this explains the rapidity of the reaction investigated by Maier & Metzler (1957). These authors have reported the immediate conversion of thiamine to a yellow form at pH 11.6. Spectroscopic evidence indicates the occurrence of a sequence of reactions involving removal of two protons and formation of a bond between N18 and C2 to produce a tricyclic compound. One can also imagine that the spatial relationship of these two groups may be of significance for the biological activity of thiamine. It is known (White *et al.*, 1959) that oxythiamine, in which a hydroxyl group is substituted for the amino group on the pyrimidine ring, is actually a thiamine antimetabolite.

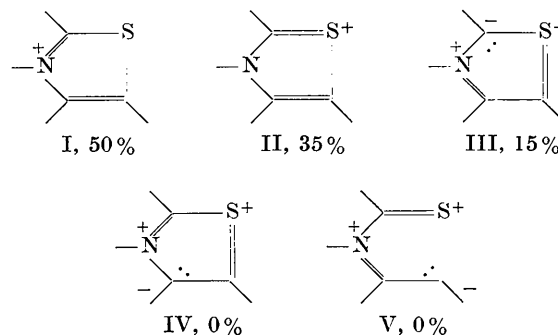
#### Bond lengths and angles

The four C-C bonds projecting from the rings are somewhat short, with an average length of 1.492 Å. Other bond lengths and angles external to the two rings are within the normal range.

Within the rings, bond lengths are satisfactorily explained by simple resonance theory, which gives the following partial contributions from canonical structures for the pyrimidine group:



and for the thiazolium group:



The above resonance contributions were calculated by applying the well known formula of Pauling (1960, equation 7-3) relating bond number and interatomic distance. Standard single-bond distances  $D_1$  were shortened by 0.04 Å for C-C and C-N, and by 0.05 Å for C-S bonds to allow for the bent-bond effect. The numerical values given must of course be regarded as rough approximations only, owing to the large errors inherent in such calculations. Nevertheless the fractional contributions arrived at are quite reasonable in view of the charge distributions of the various canonical structures.

As mentioned above, the extra proton of the pyrimidine portion of the molecule has been found attached to N11, opposite the amino group, whereas in adenine hydrochloride the ring nitrogen adjacent to the amino group is protonated instead. The following reasoning shows this behavior to be consistent with the large contribution of pyrimidine canonical structure III.

Consider the possible canonical structures which would contribute to the *unprotonated* form of the pyrimidine: five structures analogous to those shown

Table 10. *Magnitudes direction cosines of the principal axes of thermal vibration ellipsoids*

	Axis <i>i</i>	<i>B<sub>i</sub></i>	<i>c<sub>ia</sub></i>	<i>c<sub>ib</sub></i>	<i>c<sub>ic</sub></i>		Axis <i>i</i>	<i>B<sub>i</sub></i>	<i>c<sub>ia</sub></i>	<i>c<sub>ib</sub></i>	<i>c<sub>ic</sub></i>
S 1	1	4.73	-0.379	0.368	0.930	C12	1	3.21	-0.685	0.007	0.944
	2	2.98	0.821	0.556	-0.220		2	2.94	-0.728	-0.052	-0.329
	3	2.16	0.426	-0.745	0.295		3	1.69	0.033	-0.999	0.023
C 2	1	3.80	-0.443	-0.104	0.994	N13	1	3.44	0.583	-0.382	0.418
	2	2.99	0.075	-0.994	-0.102		2	2.94	0.812	0.255	-0.810
	3	2.47	0.893	0.032	0.046		3	2.04	0.017	0.889	0.412
N 3	1	2.90	-0.387	-0.377	0.926	C14	1	2.95	0.721	-0.063	0.337
	2	2.37	-0.909	0.006	-0.011		2	2.78	0.693	0.095	-0.935
	3	2.03	-0.153	0.926	0.378		3	2.20	-0.021	0.994	0.111
C 4	1	3.61	-0.375	-0.271	0.962	C15	1	3.12	0.275	0.034	0.766
	2	2.92	0.925	0.177	-0.068		2	3.06	0.947	-0.181	-0.627
	3	1.88	0.066	0.946	0.263		3	1.84	-0.165	-0.983	0.142
C 5	1	3.92	-0.163	-0.141	0.958	C16	1	3.82	0.683	-0.094	0.384
	2	2.83	0.961	0.202	-0.218		2	3.03	0.713	-0.128	-0.920
	3	2.29	-0.224	0.969	0.185		3	2.11	-0.158	-0.987	0.082
C 6	1	5.53	-0.734	-0.039	0.918	C17	1	5.42	-0.848	0.522	0.423
	2	3.30	-0.444	0.783	-0.217		2	4.72	0.104	0.323	-0.902
	3	2.48	0.514	0.621	0.332		3	2.34	0.519	0.789	0.089
C 7	1	5.49	-0.445	0.033	0.999	N18	1	4.87	0.542	-0.494	0.401
	2	4.04	-0.849	0.303	-0.050		2	2.77	0.395	-0.565	-0.822
	3	2.00	-0.286	-0.952	0.018		3	2.67	-0.742	-0.661	0.403
C 8	1	5.67	-0.239	0.124	0.977	O19	1	7.35	-0.980	-0.043	0.224
	2	5.28	0.709	-0.655	-0.050		2	6.72	0.170	0.307	-0.924
	3	2.78	-0.663	-0.745	0.207		3	4.19	-0.099	0.951	0.309
O 9	1	6.36	0.666	0.048	0.409	C120	1	3.85	-0.040	-0.562	0.771
	2	5.22	0.695	-0.403	-0.826		2	2.74	0.952	-0.272	-0.514
	3	3.61	-0.272	-0.914	0.387		3	2.54	0.303	0.781	0.376
C10	1	4.01	-0.878	0.143	0.775	C121	1	4.01	0.834	0.258	0.106
	2	2.62	-0.473	-0.098	-0.608		2	3.21	0.374	0.386	-0.922
	3	2.01	-0.080	-0.985	0.173		3	2.46	-0.406	0.886	0.372
N11	1	3.57	0.736	-0.110	0.310						
	2	3.06	0.606	-0.330	-0.907						
	3	1.96	-0.300	-0.937	0.284						

above, but lacking an H<sup>+</sup> at N11 (position 1 in the conventional ring numbering system). Among these, structures I, II, IV and V would probably contribute in about the same proportion as above. But structure III, with a negative charge at N11 in the unprotonated form, would be expected to make a much smaller contribution than above because of its large separation of opposite charges. It is evident upon examining the distribution of charges in these canonical structures that more resonance stabilization is to be gained by adding a proton at N11, and thereby allowing structure III to go from no contribution to a large one, than would be gained by adding a proton at N13 (position 3 in the conventional ring numbering system) and so somewhat further increasing the relative contribution of structure V. In brief, adding the proton at N11 creates another stable canonical structure (III) whereas this does not occur upon addition of a proton at any other position.

A similar discussion for adenine is somewhat more complicated, since it involves the relative contributions of 15 canonical structures. Nevertheless, examination shows that the affinities of the two corresponding ring nitrogens for a proton should be much more nearly equal. Furthermore, the stabiliza-

tion achieved upon addition of the proton to either would be less important because of the larger number of possible canonical structures. This probably explains in part the fact that the corresponding pK in adenine is about 4 (Albert, 1959) instead of 7 as in thiamine.

In the thiazolium ring canonical structure III makes a sizable contribution since it places a negative charge immediately between two positive charges. Structures IV and V appear to contribute nothing, within the rather large limits of error involved. This is in accord with the results of the molecular orbital computations of Pullman & Spanjaard (1961) who found a formal charge of 0.134e at C2. This was considered by them to support the catalytic mechanism proposed by Breslau (1958) and by Breslau & McNellis (1959).

Bond angles within the pyrimidine ring are all close to 120°. The small difference between the angle of 120.7° at N11 and 118.8° at N13 is somewhat surprising in view of the fact that N11 is externally bonded to a hydrogen atom whereas N13 is not. Pauling & Corey (1956) have pointed out that one would expect the angle at the ring nitrogens in pyrimidines to approach 125° for the externally bonded case or 106° for the unbonded case. Support for this view has been given by Spencer (1959) who

noted that the average of the corresponding angles found in all pertinent structure determinations was  $125^\circ$  and  $114^\circ$ .

In the thiazolium ring the bond angle at S1 is  $91.8^\circ$ , comparable to  $91^\circ$  for the corresponding angle in thiophen as determined by electron diffraction (Schomaker & Pauling, 1939). The remaining angles within the thiazolium ring are very similar to one another and to those in thiophen.

#### Hydrogen atoms and hydrogen bonds

All hydrogen atoms of the structure have been located except those attached to the alcohol group O9 and to the water oxygen O19. Possibly the extreme thermal motion of O9 and O19 is the cause of our failure to find the attached hydrogen atoms. Bond distances for C-H and N-H bonds are normal; bond angles are not tabulated but they are reasonable in view of the large errors involved.

An unusual feature of the structure is the apparent formation of two C-H...Cl<sup>-</sup> hydrogen bonds, one between C16 and Cl21'', the other between C2 and Cl20''. That these are very probably true hydrogen bonds is indicated by shortening of the interatomic distances H33...Cl21'' and H30...Cl20'' to 2.5 Å and 2.6 Å. The normal van der Waals distance would be 3.0 Å. Furthermore, the angles H33-C16-Cl21'' =  $18^\circ$  and H30-C2-Cl20'' =  $10^\circ$  are within the range commonly found in more orthodox hydrogen bonds. These bonds must, however, be somewhat weaker than N-H...Cl<sup>-</sup> bonds, for which the hydrogen to chloride distances are 2.4, 2.4 and 2.3 Å in this structure.

The chloride ions Cl20 and Cl21 are acting as hydrogen bond acceptors toward four and three donor groups respectively. Viewed from the chloride ions, these bonds are not symmetrically disposed, as will be seen from the last nine hydrogen-bond angles in Table 8.

The hydrogen bond distribution at the alcohol group and the water molecule is less well established than elsewhere in the structure since the hydrogen atoms involved have not been found. However the interatomic distances and angles make the assignment of hydrogen bonds shown in Figs. 1 and 2 appear quite reasonable; e.g. the water angle O9-O19-Cl21 =  $98.2^\circ$ , the distance O19-Cl21 = 3.163 Å (van der Waals distance would be 3.20 Å).

#### Thermal parameters

The parameters of the ellipsoids of thermal vibration have been reduced to principal axes, and the resulting magnitudes and direction cosines are given in Table 10. The quantities in columns headed  $c_{ia}$ ,  $c_{ib}$  and  $c_{ic}$  are cosines of the angles between the  $i$ th principal axis

and the  $a$ ,  $b$  and  $c$  axes of the unit cell. The  $B_i$  are given in Å<sup>2</sup>. As would be expected, the water molecule, side chain and projecting groups show the largest amplitudes of thermal motion. It is likely that the thermal parameters are strongly influenced by the neglected absorption corrections.

This investigation was supported by Research Grant RG-4859 from the U.S. Public Health Service, by a grant from the Initiative 171 Section of the Research Committee of the Graduate School, University of Washington, by a grant from the University of Washington Institutional Cancer Grant Committee, and by a grant of free computer time from the Research Computer Laboratory of the University of Washington. We are also grateful to the IBM corporation and the Boeing Airplane Company for three hours of IBM-7090 time. Finally, we are indebted to the referee of the original version of this paper for valuable criticisms and suggestions.

#### References

- ALBERT, A. (1959). *Heterocyclic Chemistry*. New Jersey: Fairlawn, N. J.: Essential Books.
- BERGHUIS, J., HAANAPPEL, IJ. M., POTTERS, M., LOOPSTRA, B. O., MACGILLAVRY, C. H. & VEENENDAAL, A. L. (1955). *Acta Cryst.* **8**, 478.
- BRESLAU, R. (1958). *J. Amer. Chem. Soc.* **80**, 3719.
- BRESLAU, R. & McNELIS, E. (1959). *J. Amer. Chem. Soc.* **81**, 3080.
- BUSING, W. R. & LEVY, H. A. (1959). *A Crystallographic Least Squares Refinement Program for the IBM 704*, U.S. Atomic Energy Commission Publication ORNL 59-4-37.
- COCHRAN, W. (1951). *Acta Cryst.* **4**, 81.
- HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
- MAIER, G. D. & METZLER, D. E. (1957). *J. Amer. Chem. Soc.* **79**, 4386.
- MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.
- PAULING, L. & COREY, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164.
- PAULING, L. (1960). *The Nature of the Chemical Bond*, 3rd edition. Ithaca: Cornell University Press.
- PULLMAN, B. & SPANJAARD, C. (1961). *Biochim. Biophys. Acta*, **46**, 576.
- SCHOMAKER, V. & PAULING, L. (1939). *J. Amer. Chem. Soc.* **61**, 1769.
- SPENCER, M. (1959). *Acta Cryst.* **12**, 59.
- SUTTON, L. E. (1958). *Tables of Interatomic Distances and Configuration in Molecules and Ions*. Chemical Society Special Publication No. 11. London.
- TOMIE, Y. & STAM, C. H. (1958). *Acta Cryst.* **11**, 126.
- VISCONTINI, M., BONETTI, G., EBNÖTHER, C. & KARRER, P. (1951). *Helv. Chim. Acta*, **34**, 1388.
- WHITE, A., HANDLER, P., SMITH, E. L. & STETTEN, D. (1959). *Principles of Biochemistry*, 2nd ed. New York: McGraw-Hill.