Acta Cryst. (1972). B28, 3352

The Crystal and Molecular Structure of L-*N*-Acetylhistidine Monohydrate: An Application of Direct Methods to Space Group *P*1*

BY THOMAS J. KISTENMACHER,† DAVID J. HUNT‡ AND RICHARD E. MARSH

Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91109, *U.S.A.*

(*Received* 16 *June* 1972)

L-*N*-Acetylhistidine crystallizes from aqueous solution as the monohydrate. The crystals are triclinic, space group *P*1, with cell constants $a=8{\cdot}865$ (2), $b=9{\cdot}097$ (2), $c=7{\cdot}346$ (2) Å, $\alpha=102{\cdot}24$ (1), $\beta=90{\cdot}30$ (1), $\gamma=117{\cdot}73$ (1)°. The unit cell contains two independent *N*-acetylhistidine molecules and two independent waters of crystallization. Intensities of 2152 reflections were collected on an automated diffractometer. The structure solution was initiated by a symbolic addition–tangent refinement–*E* map sequence. The *E* map indicated positions for 13 atoms – eight of which turned out to be correct. Continued application of tangent refinement and *E* maps eventually led to a model in which 18 atoms were positioned correctly and one incorrectly. Fourier and difference Fourier techniques successfully produced the complete structure (30 nonhydrogen atoms). Least-squares refinement gave a final *R* index of 0·029 and e.s.d.'s in the heavy-atom coordinates of about 0·003 Å. The molecule exists in the crystal as a zwitterion with protonation of the imidazole ring. The two independent molecules have very different molecular conformations. For this crystal structure, at least, it appears that packing and hydrogen bonding forces play a crucial role in determining the molecular conformation.

Introduction

The importance of the amino acid histidine in protein chemistry is associated with the frequency with which it occurs at active sites. In particular, histidine seems to be directly involved in the proteolysis mechanisms of the enzymes papain, subtilisin, chymotrypsin, trypsin, and other proteolytic enzymes, and in the coordination of zinc in zinc insulin (Tanford & Epstein, 1954) and zinc bacitracin (Craig, Philips & Burachik, 1969; Cornell & Gurney, 1970).

Several X-ray diffraction studies involving the histidine moiety have now been reported: L-histidine hydrochloride (Donohue, Lavine & Rollett, 1956; Donohue & Caron, 1964), DL-histidine hydrochloride (Bennett, Davidson, Harding & Morelle, 1970), monoclinic L-histidine (Madden, McGandy, Seeman, Harding & Hoy, 1972), orthorhombic L-histidine (Madden, McGandy & Seeman, 1972), and several metal–histidine coordination complexes (see Freeman, 1967). One notable feature of these studies is that the histidine moiety occurs in two conformations – one 'open' and one 'closed' (Kistenmacher & Marsh, 1971); the magnitude of the torsion angle about the C(2)–C(3) (C^{α}–C^{β}) bond differentiates between the two conformers. The 'closed' conformation has been observed in crystals of L-histidine hydrochloride and in the metal–histidine complexes; the 'open' conformation is found in two forms of L-histidine and in DL-histidine hydrochloride.

The crystal structure analysis of L-*N*-acetylhistidine is important for two reasons: (1) it is the first successful example of the application of direct methods to the most general phase problem, space group *P*1, and (2) the results show that there are two independent molecules of L-*N*-acetylhistidine in the unit cell, one having the 'open' conformation and one the 'closed' conformation.

A preliminary description of the structure of L-*N*-acetylhistidine has been published (Kistenmacher & Marsh, 1971).

Experimental

A sample of L-*N*-acetylhistidine was purchased from Nutritional Biochemicals Corporation. It was recrystallized from water, yielding colorless parallelepipeds somewhat elongated in the **b** direction. Rotation and Weissenberg photographs indicated a triclinic lattice, and the presumed optical purity of the compound requires the space group to be *P*1. Unit-cell dimensions were obtained from 2θ measurements both on Weissenberg films prepared in a Straumanis-type camera and on a diffractometer; the two sets of values agreed within experimental error. The density was measured by flotation. Crystal data are summarized in Table 1.

* Contribution No. 4475 from the Arthur Amos Noyes Laboratory of Chemical Physics. This investigation was supported in part by Public Health Service Research Grant No. GM 16966, from the National Institute of General Medical Sciences, National Institutes of Health, and in part by a National Science Foundation Postdoctoral Fellowship (T.J.K.).

† Present adress: Department of Chemistry, Remsen Hall, The Johns Hopkins University, Baltimore, Maryland 21218, U.S.A.

‡ Present address: I.B.M. (U.K.) Ltd., 126–150 Washway Road, Sale, Cheshire, England.

Table 1. *Crystal data for* L-*N-acetylhistidine monohydrate*

$a = 8{\cdot}865 \pm 0{\cdot}002$ Å	$C_8O_3N_3H_{11}\ H_2O$
$b = 9{\cdot}097 \pm 0{\cdot}002$	M.W. 215·2
$c = 7{\cdot}346 \pm 0{\cdot}002$	Space group: $P1$
$\alpha = 102{\cdot}24 \pm 0{\cdot}01°$	$Z = 2$
$\beta = 90{\cdot}30 \pm 0{\cdot}01$	$F(000) = 228$
$\gamma = 117{\cdot}73 \pm 0{\cdot}01$	$D_m = 1{\cdot}407$
$V = 508{\cdot}8$ Å^3	$D_x = 1{\cdot}404$
λ(Cu $K\alpha$) = 1·5418 Å	$\mu = 9{\cdot}8$ cm^{-1}

Intensities were measured on a modified Datex-automated General Electric diffractometer using Ni-filtered Cu $K\alpha$ radiation. The crystal was an approximate cube, about 0·4 mm on an edge, mounted along the b axis. A θ–2θ scan technique was used, with a scan speed of 2° (in 2θ) per minute and a scan range varying from 2° at $2\theta = 12°$ to 4° at $2\theta = 130°$; background was counted for 30 sec at each extremum. A single reflection, $\bar{1}04$, was monitored every 15 reflections; in addition, the intensities and alignments of several other reflections were checked by hand after every 400 measurements. There were no significant changes in intensities during the period of data collection. All reflections in a hemisphere of reciprocal space out to $2\theta = 154°$ were surveyed; these totaled 2152, of which four had net intensities less than zero. Observational variances, $\sigma^2(I)$, were assigned on the basis of counting statistics for the scan and background plus an additional term $(0{\cdot}02S)^2$, where S is the scan count. Intensities and their standard deviations were then corrected for Lorentz and polarization factors, but not for absorption ($\mu R = 0{\cdot}2$).

Solution of the structure

Our first attempts to derive the structure focused on the rotation function (Rossmann & Blow, 1962). In terms of the positions of the heavy atoms, the *N*-acetylhistidine molecule has but three degrees of conformational freedom. We hoped, then, that major fragments of the two molecules in the asymmetric unit would have the same conformation; in this case, a pronounced maximum in the rotation function would indicate the relative orientations of these two fragments. However, we were disappointed to find that no maximum in the rotation function was greatly predominant. Nevertheless, we spent considerable time in attempts to correlate the features of the rotation function with the three-dimensional Patterson function and with packing models of the molecules. These efforts were fruitless.

The structure was eventually solved by the direct method of Karle & Karle (1966) for acentric space groups. Since as far as we are aware this was the first structure in space group $P1$ to be solved by this method, and since in any event the size of the structure (30 heavy atoms) made the solution a rather formidable undertaking, we describe our procedure in somewhat more detail than usual.

Three linearly independent reflections ($0\bar{7}1$, $\bar{1}63$, $4\bar{2}5$) were assigned arbitrary phases to fix the origin. Careful examination of the interactions among reflections with $|E| \geq 2{\cdot}0$ and partial phase expansion by hand led to the assignment of three symbolic phases (Zachariasen, 1952; Karle & Karle, 1966, equation 4.3). With these six phases (Table 2) as a starting set, the symbolic addition method was applied to the 110 reflections with $|E| \geq 1{\cdot}85$, using a program written by Duchamp (1969). During this process one symbolic phase ($37\bar{4}$) was rejected because it was involved in too small a number of high-reliability interactions (Germain, Main & Woolfson, 1970). The phase of the $6\bar{1}4$ reflection was restricted to values between 0 and 180° to specify the enantiomorph.

Table 2. *Initial phase data*

$h\ k\ l$	Phase (°)	$\|E\|$	Parity
$0\ \bar{7}\ 1$	45	3·66	*eoo*
$\bar{1}\ 6\ 3$	90	2·51	*oeo*
$4\ \bar{2}\ 5$	135	2·63	*eeo*
$6\ \bar{1}\ 4$	*A*	2·75	*eoe*
$0\ 6\ 1$	*B*	3·14	*eeo*
$3\ 7\ \bar{4}$	*C*	2·45	*ooe*

A total of 97 reflections were assigned phases during the symbolic addition process. The list of phases was gradually expanded and refined using the familiar tangent formula (Karle & Hauptman, 1956; Karle & Karle, 1966, equation 4.5), resulting in 218 phased reflections with $|E| \geq 1{\cdot}40$ which were used as input to the initial E map. From this map, 13 peaks were selected as representing atomic sites; in retrospect, 8 of these peaks corresponded to correct atom postitions while five were spurious [Fig. 1(*a*)]. Structure factors were calculated ($R = \sum||F_o| - |F_c||/\sum|F_o| = 0{\cdot}56$) and phases were accepted for a second cycle of tangent refinement if $|F_c|/|F_o| \geq 0{\cdot}40$ and $|E| \geq 1{\cdot}60$ (Karle, 1968). Expansion to include all reflections with $|E| \geq 1{\cdot}45$ led to assignment of 278 phases which were the basis of a second E map. From it 16 atomic sites were selected of which 12 turned out to be correct [Fig. 1(*b*)]. A second structure-factor ($R = 0{\cdot}53$) tangent-refinement cycle led to 270 phases ($|E| \geq 1{\cdot}45$); on the basis of the resulting E map, 7 new atoms were added [five correctly; Fig. 1(*c*)] and three correct atoms which were represented by anomalously large peaks [C(1), O(1), and O(2) of molecule *A*] were removed. Structure factors were again calculated ($R = 0{\cdot}49$) and phases accepted for tangent refinement if $|F_c|/|F_o| \geq 0{\cdot}60$ and $|E| \geq 1{\cdot}60$. A total of 225 phased E's formed the final E map, which indicated 19 atom positions [18 correct; Fig. 1(*d*)]. The R index was now 0·47.

At this point an electron density map was calculated and compared with the last E map. Five new and correct atoms were positioned, including the three that had been deleted previously, and the resulting R index was 0·40. At this stage 24 atoms had been located, of which only one was incorrect. Further electron density and difference maps readily located the remaining 7

atoms and indicated the removal of the incorrect one. With all 30 major atoms located and assigned form factors of carbon, the R index was 0·30.

The course of the phase refinement is summarized in Fig. 2, which shows the relative phase errors at three stages of the analysis. In hindsight, it appears as though we allowed too many phases (97) to be determined by the relatively inaccurate symbolic addition method; we now feel that we should have been more selective in assigning starting phases for tangent refinement. The removal from the phasing procedure of the three atoms having excessively large peaks on the E map seems to have quickened the pace of the convergence [compare Fig. 1(*c*) and (*d*)].

Refinement

Refinement of the structure was by least squares, the quantity minimized being $\sum w(F_o^2-s^2F_c^2)^2$ where $1/s$ is the scale factor and weights were taken equal to $1/\sigma^2(F_o^2)$; the four reflections with net intensities less than zero were given zero weight. Atomic form factors for C, N and O were taken from *International Tables for X-ray Crystallography* (1962) and for H from Stewart, Davidson & Simpson (1965). Calculations were carried out on an IBM 360-75 computer using the CRYM system. The coordinates of atom O(1) of molecule B were held fixed to prevent a singular matrix.

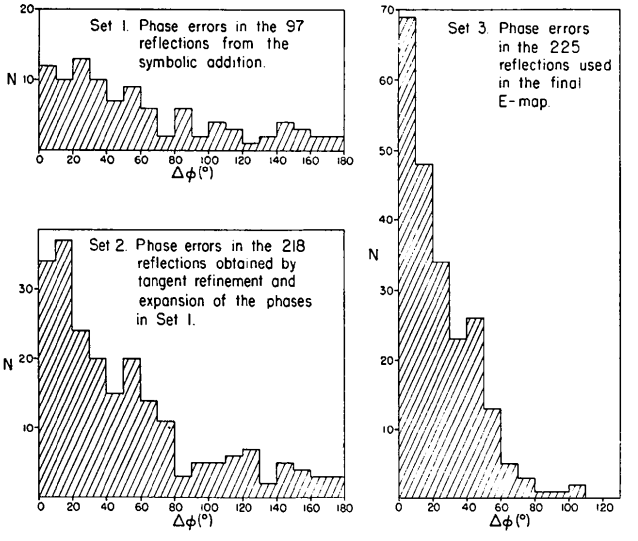

Fig. 2. Distribution of phase errors at three stages of the phase refinement.

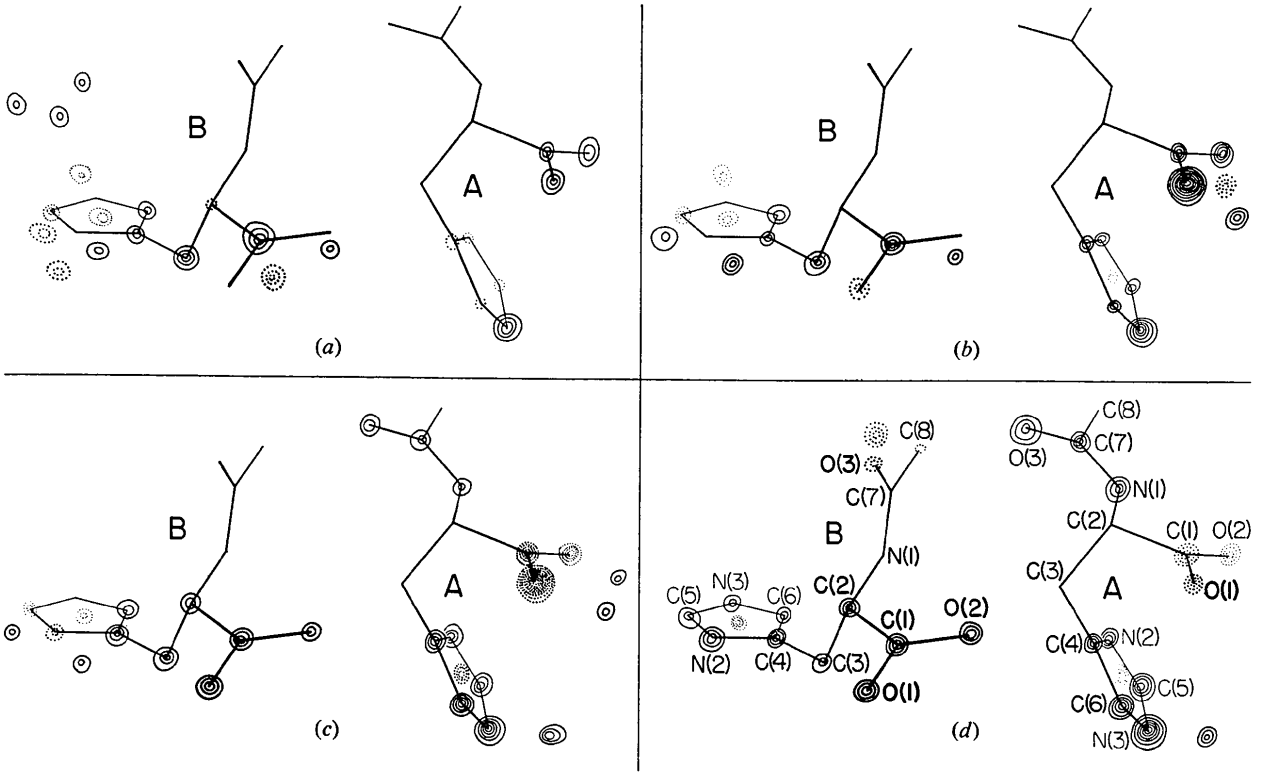

Fig. 1. Portions of E maps calculated at four stages of the structure solution. Solid contours represent sites chosen as atomic positions; dashed contours indicate additional features. Final atomic positions are indicated by the molecular skeletons. The relative orientation of molecules A and B is not their relative orientation in the crystals. (*a*) Initial E map, calculated after the symbolic-additional tangent-refinement cycle (218 phases); (*b*) E map calculated after first structure-factor tangent-refinement cycle (278 phases); (*c*) E map calculated after second structure-factor tangent-refinement cycle (270 phases); (*d*) final E map, calculated after third structure-factor tangent-refinement cycle (225 phases). Contours are on a uniform, arbitrary scale.

Assignment of the correct form factor to each heavy atom and refinement of their positional and isotropic temperature parameters led to an R index of 0·11. A difference map allowed the positions of the 26 hydrogen atoms to be assigned, although there was some doubt as to the orientations of the methyl groups because of residual electron density representing anisotropic thermal motions of the carbon atoms.

Refinement was continued, adding successively to the parameter list: (1) anisotropic thermal factors for the heavy atoms; (2) an isotropic extinction parameter (Zachariasen, 1963; Larson, 1967, equation 3); (3) coordinates of the hydrogen atoms; (4) isotropic temperature factors of the hydrogen atoms. Between stages (2) and (3), difference maps in the planes of the methyl hydrogen atoms were recalculated; slight changes in orientations were indicated. Difference maps in the planes of the imidazole rings were also calculated (Fig. 3); they confirmed the protonation of all the ring nitrogen atoms. Because of the limited size of core storage, the parameters were divided into three matrices: (1) scale factor, extinction parameter, and anisotropic temperature parameters of the C, N and O atoms; (2) coordinates of the C, N and O atoms; (3) coordinates and isotropic temperature parameters of the H atoms. Two of these three sets of parameters were adjusted in each cycle. During the final cycles, no parameter shifted by as much as 0·3 e.s.d. A final difference map showed no unusual features.

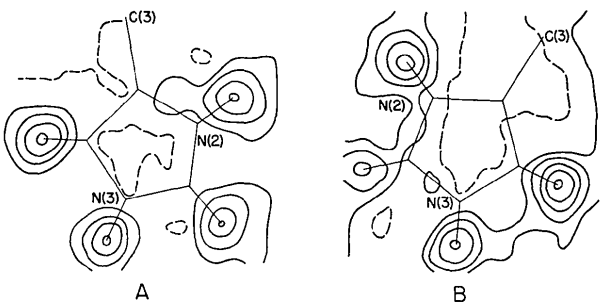

Fig. 3. Difference Fourier maps in the planes of the imidazole rings. Contours are at −0·05 (dashed), 0·05, 0·15 ... e.Å^{-3}.

The final value of the R index was 0·029 and the goodness-of-fit, $\sum w(F_o^2-F_c^2)^2/(n-p)$ for $n=2148$ observations of non-zero weight and $p=373$ parameters, was 1·5. Final parameters of the heavy atoms are given in Table 3 and of the hydrogen atoms in Table 4. Ob-

Table 3. *Coordinates and anisotropic temperature parameters of the nonhydrogen atoms* ($\times 10^4$)

Estimated standard deviations in parentheses. Temperature parameters in the form exp $(-b_{11}h^2 \ldots -b_{23}kl)$.

	x	y	z	b_{11}	b_{22}	b_{33}	b_{12}	b_{13}	b_{23}
Molecule *A*									
C(1)	7456 (3)	7086 (3)	8191 (4)	87 (3)	81 (3)	161 (4)	41 (5)	19 (6)	79 (6)
C(2)	6321 (3)	5450 (3)	6690 (4)	83 (3)	71 (3)	180 (5)	44 (5)	−11 (6)	75 (6)
C(3)	7384 (3)	5281 (3)	5074 (4)	107 (3)	75 (3)	199 (5)	69 (5)	19 (6)	30 (6)
C(4)	8429 (3)	6937 (3)	4530 (4)	90 (3)	92 (3)	144 (4)	81 (5)	26 (5)	33 (6)
C(5)	8988 (3)	9169 (3)	3409 (4)	102 (3)	107 (3)	175 (5)	69 (5)	0 (6)	90 (6)
C(6)	10107 (3)	8104 (3)	5052 (4)	88 (3)	123 (4)	164 (5)	86 (5)	11 (6)	68 (6)
C(7)	3313 (3)	3979 (3)	5406 (4)	87 (3)	107 (4)	168 (5)	33 (5)	−5 (6)	68 (6)
C(8)	1737 (3)	4134 (4)	5042 (5)	92 (4)	150 (4)	260 (7)	66 (6)	−35 (7)	59 (8)
N(1)	4762 (3)	5439 (3)	6063 (4)	83 (3)	83 (3)	177 (4)	48 (4)	−14 (5)	59 (5)
N(2)	7764 (3)	7641 (3)	3501 (4)	77 (3)	100 (3)	153 (4)	53 (4)	−14 (5)	47 (5)
N(3)	10415 (3)	9482 (3)	4343 (4)	85 (3)	99 (3)	163 (4)	29 (5)	5 (5)	63 (5)
O(1)	8632 (3)	7055 (3)	9107 (4)	96 (2)	90 (2)	218 (4)	34 (4)	−72 (5)	87 (5)
O(2)	7199 (3)	8329 (3)	8309 (4)	145 (3)	93 (3)	215 (4)	115 (4)	−42 (5)	12 (5)
O(3)	3286 (3)	2571 (3)	5137 (4)	105 (3)	88 (3)	296 (5)	15 (4)	−47 (6)	46 (5)
Molecule *B*									
C(1)	3384 (3)	6120 (3)	11023 (4)	92 (3)	61 (3)	181 (5)	58 (5)	−14 (6)	67 (6)
C(2)	3443 (3)	7877 (3)	11493 (4)	87 (3)	67 (3)	134 (4)	74 (5)	9 (5)	51 (5)
C(3)	3395 (3)	8400 (3)	9659 (4)	115 (3)	69 (3)	131 (4)	103 (5)	−5 (5)	27 (5)
C(4)	3208 (3)	9981 (3)	9888 (4)	90 (3)	73 (3)	120 (4)	80 (5)	−4 (5)	40 (5)
C(5)	1857 (3)	11502 (3)	10091 (4)	100 (3)	87 (3)	221 (5)	107 (5)	8 (6)	81 (6)
C(6)	4389 (3)	11657 (3)	10184 (4)	91 (3)	82 (3)	184 (5)	85 (5)	15 (6)	65 (6)
C(7)	5005 (3)	9363 (3)	14626 (4)	119 (3)	93 (3)	137 (4)	117 (5)	12 (6)	52 (6)
C(8)	6561 (4)	10895 (4)	15790 (4)	182 (5)	120 (4)	151 (5)	99 (7)	−36 (7)	3 (7)
N(1)	4930 (3)	9208 (3)	12777 (4)	101 (3)	63 (3)	123 (3)	52 (4)	8 (5)	45 (4)
N(2)	1631 (3)	9922 (3)	9845 (4)	83 (3)	70 (3)	177 (4)	63 (4)	−3 (5)	59 (5)
N(3)	3512 (3)	12571 (3)	10302 (4)	100 (3)	63 (2)	198 (4)	77 (4)	10 (5)	66 (5)
O(1)*	2000	4880	10200	112 (3)	67 (2)	399 (6)	65 (4)	−174 (6)	21 (6)
O(2)	4694 (3)	5993 (2)	11394 (4)	90 (2)	70 (2)	242 (4)	71 (4)	−81 (5)	24 (5)
O(3)	3880 (3)	8325 (3)	15339 (4)	148 (3)	170 (3)	158 (4)	104 (5)	50 (5)	137 (5)
Water									
O(4)	7994 (3)	1229 (3)	1036 (4)	143 (3)	108 (3)	194 (4)	99 (4)	30 (5)	103 (5)
O(5)	8698 (3)	4241 (3)	10016 (5)	137 (3)	143 (4)	672 (10)	81 (6)	−43 (9)	367 (9)

* Coordinates of O(1) of molecule *B* were held fixed to prevent a singular matrix.

served and calculated structure factors are collected in Table 5.

Standard deviations in the atomic parameters (Tables 3 and 4) may be slightly underestimated because of the partitioning of the matrices, but we doubt if the effect is as large as 10%.

Discussion

(i) *Bond distances and angles*

Drawings of the two molecules, showing the conformational features and the bond distances between heavy atoms, are shown in Fig. 4; the corresponding bond angles are listed in Table 6. Bond distances and angles involving the hydrogen atoms are listed in Table 7. Estimated standard deviations are in the range 0·004–0·005 Å and 0·1–0·2° for bonds involving only the heavy atoms and 0·03–0·05 Å and 1·5–3·0° for bonds involving hydrogen atoms. Comparison of corresponding values in the two molecules – and particularly within the five-membered rings, where nonbonding forces in the two molecules should be very nearly equivalent – suggests that these e.s.d.'s are reasonable.

Table 4. *Coordinates* ($\times 10^3$) *and isotropic temperature parameters of the hydrogen atoms*

Estimated standard deviations are in parentheses

	x	*y*	*z*	*B*
Molecule *A*				
H(1)	595 (3)	452 (3)	725 (4)	3·3 (5) Å²
H(2)	464 (3)	643 (3)	623 (4)	3·5 (5)
H(3)	667 (3)	444 (3)	392 (4)	3·5 (5)
H(4)	820 (3)	493 (3)	545 (4)	3·4 (5)
H(5)	663 (4)	719 (4)	292 (4)	5·9 (8)
H(6)	882 (4)	988 (4)	276 (4)	4·8 (7)
H(7)	1148 (4)	1044 (4)	458 (4)	5·4 (8)
H(8)	1097 (3)	801 (3)	576 (4)	3·8 (6)
H(9)	105 (4)	337 (4)	407 (5)	6·5 (8)
H(10)	190 (4)	530 (4)	507 (4)	5·9 (8)
H(11)	97 (5)	380 (5)	590 (5)	6·8 (8)
Molecule *B*				
H(1)	242 (3)	773 (3)	1207 (3)	2·0 (4)
H(2)	579 (3)	991 (3)	1234 (3)	3·0 (5)
H(3)	437 (3)	859 (3)	899 (3)	2·7 (4)
H(4)	238 (3)	741 (3)	884 (3)	3·0 (5)
H(5)	46 (4)	887 (4)	960 (4)	4·5 (6)
H(6)	94 (4)	1178 (4)	1004 (4)	4·4 (6)
H(7)	402 (4)	1379 (4)	1060 (4)	5·2 (7)
H(8)	565 (3)	1225 (3)	1029 (4)	3·5 (6)
H(9)	705 (5)	1042 (5)	1654 (5)	7·8 (10)
H(10)	739 (4)	1158 (4)	1507 (4)	5·6 (7)
H(11)	616 (4)	1153 (4)	1663 (4)	5·4 (7)
Water molecules				
H(12)	778 (4)	39 (4)	8 (5)	6·3 (8)
H(13)	816 (3)	203 (3)	70 (4)	3·7 (6)
H(14)	993 (4)	471 (4)	1028 (5)	5·9 (8)
H(15)	857 (5)	499 (5)	980 (5)	7·8 (10)

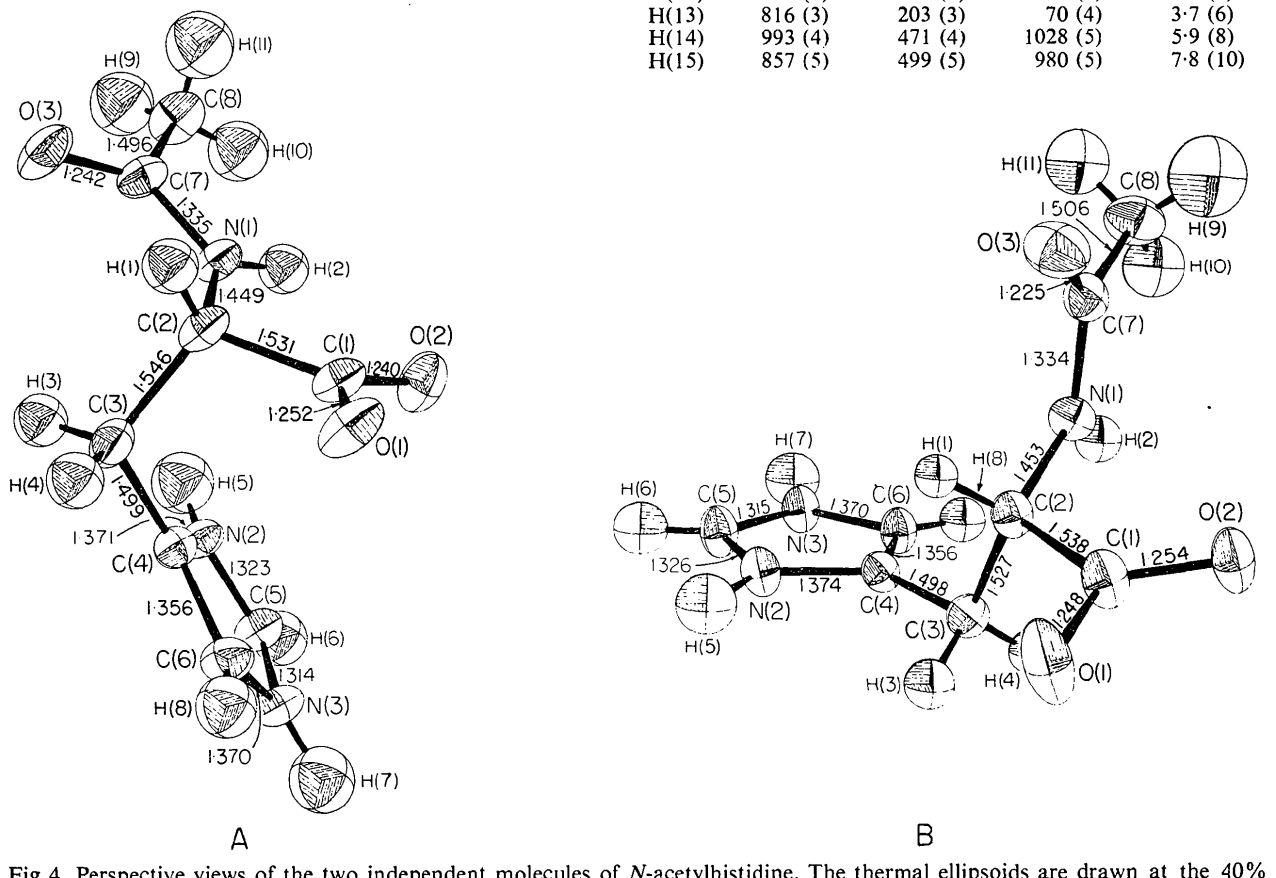

Fig. 4. Perspective views of the two independent molecules of *N*-acetylhistidine. The thermal ellipsoids are drawn at the 40% probability level. Standard deviations in the bond lengths are 0·004–0·005 Å.

Table 5. *Observed and calculated structure amplitudes*

The three columns contain values, reading from left to right, of h, $10F(\text{obs})$, and $10|F(\text{calc})|$. The four reflections designated with an asterisk (*) had negative values for their net intensities, and were assigned zero weight.

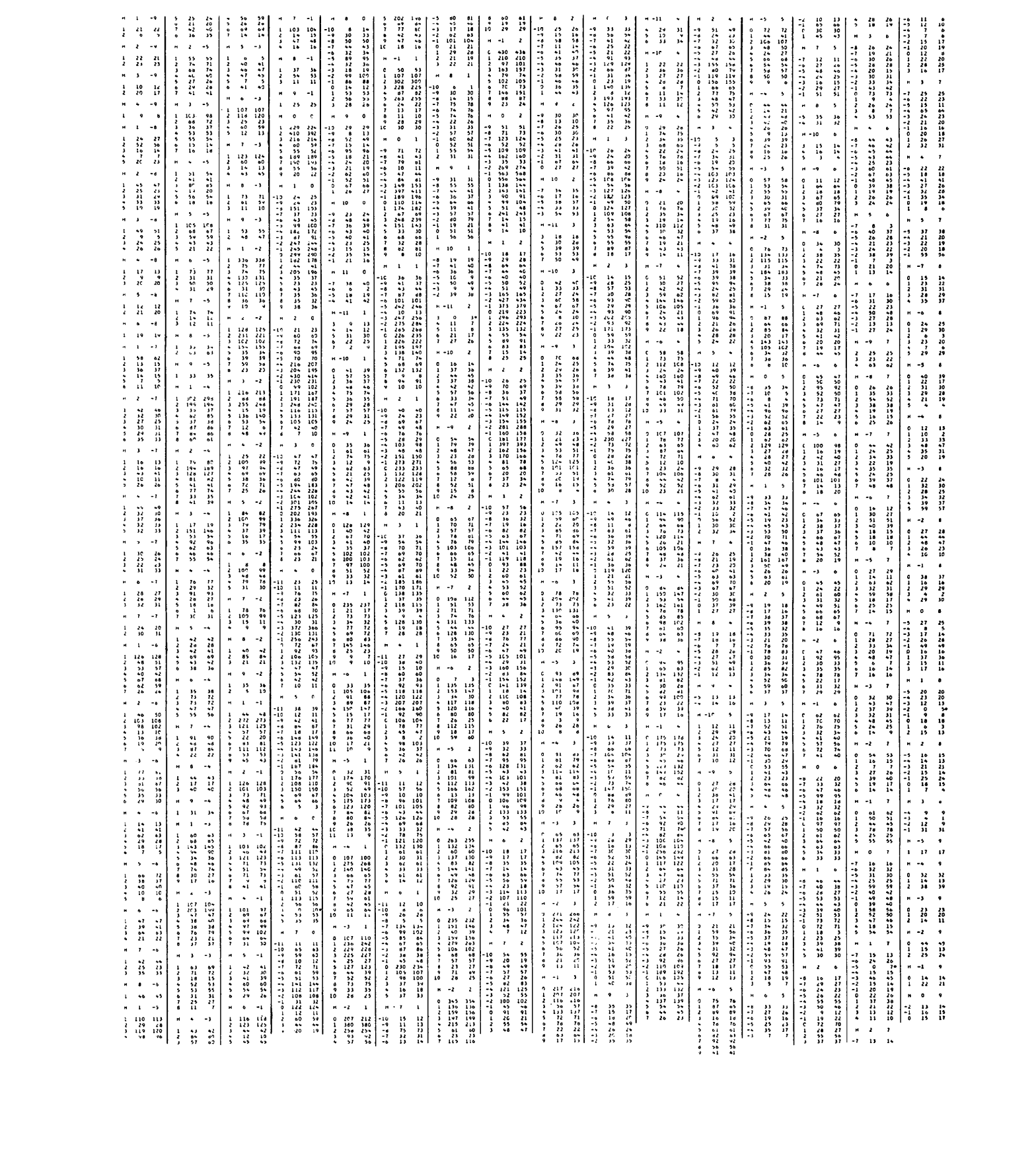

Table 6. *Bond angles involving the heavy atoms*

Standard deviations are in the range 0·1–0·2°.

	Molecule *A*	Molecule *B*
C(2)–C(1)–O(1)	115·5	116·2
C(2)–C(1)–O(2)	117·8	120·4
O(1)–C(1)–O(2)	126·5	123·3
C(3)–C(2)–C(1)	108·7	108·2
N(1)–C(2)–C(1)	111·0	114·0
C(3)–C(2)–N(1)	112·9	110·0
C(4)–C(3)–C(2)	112·7	113·8
N(2)–C(4)–C(3)	124·2	122·1
C(6)–C(4)–C(3)	129·5	131·7
C(6)–C(4)–N(2)	106·0	106·2
C(5)–N(2)–C(4)	109·2	108·9
N(3)–C(5)–N(2)	108·7	108·8
C(6)–N(3)–C(5)	108·7	108·8
C(4)–C(6)–N(3)	107·4	107·3
C(7)–N(1)–C(2)	121·6	122·1
C(8)–C(7)–N(1)	116·5	116·0
O(3)–C(7)–N(1)	121·2	122·2
C(8)–C(7)–O(3)	122·3	121·8

The molecules are zwitterions, the imidazole rings being protonated. In histidine itself (Madden, McGandy & Seeman, 1972; Madden, McGandy, Seeman, Harding & Hoy, 1972) the α-amino nitrogen atom, N(1), is protonated rather than the imidazole ring; the acetyl group, which forms an amide linkage N(1)–C(7) with appreciable double-bond character, inhibits transfer of the proton to N(1) in the present compound. Both the imidazole ring and the α-amino groups are protonated in histidine hydrochloride (Bennett, Davidson, Harding & Morelle, 1969; Donohue & Caron, 1964; Donohue, Lavine & Rollett, 1956).

The dimensions of the protonated imidazole rings in the present structure agree satisfactorily with those found in DL-histidine hydrochloride (Bennett *et al.*, 1969) and in L-histidine hydrochloride (Donohue & Caron, 1964). As pointed out by Bennett *et al.* (1969), the difference between the external angles at C(4) reflects the relatively large double-bond character of the C(4)–C(6) bond.

The atoms of each imidazole ring are coplanar within experimental error (Table 8). The relatively large deviation of C(3) from the plane of the imidazole ring in molecule *A* reflects the strain in this molecule (see later discussion).

Table 8. *Least-squares planes and the deviations of individual atoms from these planes*

Direction cosines q are relative to the unit-cell axes; D is the origin-to-plane distance. Atoms indicated by an asterisk (*) were given zero weight in calculating the planes; other atoms were weighted equally.

Imidazole ring

(*A*) $q(\mathbf{a}) = -0{\cdot}505$, $q(\mathbf{b}) = 0{\cdot}430$, $q(\mathbf{c}) = 0{\cdot}759$; $D = 1{\cdot}464$ Å

C(4)	0·002 Å	N(3)	0·003 Å
C(6)	−0·003	N(2)	0·000
C(5)	−0·002	C(3)	0·125*

(*B*) $q(\mathbf{a}) = -0{\cdot}021$, $q(\mathbf{b}) = -0{\cdot}125$, $q(\mathbf{c}) = 0{\cdot}996$; $D = 6{\cdot}039$ Å

C(4)	−0·003	N(3)	0·000
C(6)	0·002	N(2)	0·003
C(5)	−0·002	C(3)	0·007*

Carboxylate group

(*A*) $q(\mathbf{a}) = -0{\cdot}538$, $q(\mathbf{b}) = -0{\cdot}195$, $q(\mathbf{c}) = 0{\cdot}781$; $D = -0{\cdot}142$ Å

C(2)	−0·006	O(1)	−0·008
C(1)	0·022	O(2)	−0·008
	N(1)	0·380*	

(*B*) $q(\mathbf{a}) = -0{\cdot}262$, $q(\mathbf{b}) = -0{\cdot}209$, $q(\mathbf{c}) = 0{\cdot}955$; $D = 5{\cdot}768$ Å

C(2)	−0·004	O(1)	−0·005
C(1)	0·015	O(2)	−0·005
	N(1)	0·298*	

Amide group

(*A*) $q(\mathbf{a}) = -0{\cdot}241$, $q(\mathbf{b}) = -0{\cdot}152$, $q(\mathbf{c}) = 0{\cdot}969$; $D = 2{\cdot}608$ Å

C(8)	0·039	N(1)	−0·062
C(7)	−0·017	C(2)	0·050
O(3)	−0·008	H(2)	−0·06*

(*B*) $q(\mathbf{a}) = -0{\cdot}836$, $q(\mathbf{b}) = 0{\cdot}866$, $q(\mathbf{c}) = -0{\cdot}024$; $D = 3{\cdot}418$ Å

C(8)	0·025	N(1)	−0·043
C(7)	−0·009	C(2)	0·033
O(3)	−0·006	H(2)	−0·12*

Table 7. *Distances and angles involving the hydrogen atoms*

Standard deviations are about 0·03–0·05 Å and 1·5–3·0°.

	A	*B*		*A*	*B*		*A*	*B*
C(2)–H(1)	0·94 Å	0·97 Å	C(1)–C(2)–H(1)	108°	107°	N(1)—C(2)–H(1)	105°	108°
			C(3)–C(2)–H(1)	111	109			
N(1)–H(2)	0·94	0·86	C(2)–N(1)–H(2)	123	119	C(7)—N(1)–H(2)	115	119
C(3)–H(3)	0·99	0·96	C(2)–C(3)–H(3)	113	115	C(4)—C(3)–H(3)	106	106
C(3)–H(4)	0·98	0·99	C(2)–C(3)–H(4)	110	105	C(4)—C(3)–H(4)	106	110
			H(3)–C(3)–H(4)	108	107			
N(2)–H(5)	0·95	1·01	C(4)–N(2)–H(5)	129	128	C(5)—N(2)–H(5)	122	123
C(5)–H(6)	0·95	0·96	N(2)–C(5)–H(6)	124	124	N(3)—C(5)–H(6)	128	127
N(3)–H(7)	0·92	0·96	C(5)–N(3)–H(7)	130	125	C(6)—N(3)–H(7)	121	126
C(6)–H(8)	0·97	0·98	N(3)–C(6)–H(8)	124	120	C(4)—C(6)–H(8)	128	133
C(8)–H(9)	0·87	0·98	C(7)–C(8)–H(9)	113	105	H(9)—C(8)–H(10)	112	111
C(8)–H(10)	1·00	0·96	C(7)–C(8)–H(10)	117	114	H(10)–C(8)–H(11)	102	113
C(8)–H(11)	0·92	0·94	C(7)–C(8)–H(11)	113	107	H(9)—C(8)–H(11)	96	107

Water molecules

O(4)–H(12)	0·87 Å	O(4)–H(13)	0·77 Å	H(12)–O(4)–H(13)	109 Å
O(5)–H(14)	0·97	O(5)–H(15)	0·79	H(14)–O(5)–H(15)	106

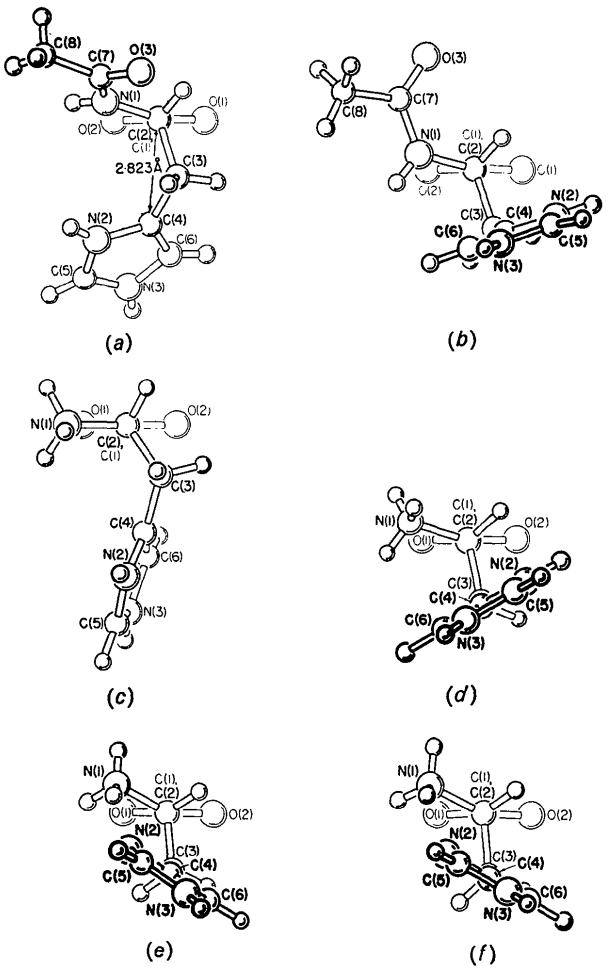

Fig. 5. Views along the C(2)–C(1) bond, showing the conformations of the histidine moiety in various crystal structures. (*a*) *N*-Acetylhistidine, molecule *A*. (*b*) *N*-Acetylhistidine, molecule *B*. (*c*) L-Histidine hydrochloride (Donohue, Lavine & Rollett, 1956; Donohue & Caron, 1964). (*d*) DL-Histidine hydrochloride (Bennett *et al.*, 1970). (*e*) L-Histidine, monoclinic modification (Madden *et al.*, 1972). (*f*) L-Histidine, orthorhombic modification (Madden, McGandy & Seeman, 1972).

The bond distances and angles in the amide groups are in good agreement with the average values in a peptide link as tabulated by Marsh & Donohue (1967). In both molecules the atoms of the amide group are significantly non-planar (Table 8); the non-planarity reflects primarily a twist about the N(1)–C(7) bond, by about 7° in molecule *A* and 5° in molecule *B*. The small but highly significant difference in the two C(7)–O(3) distances is undoubtedly related to the difference in hydrogen bonding; atom O(3) in molecule *A* accepts a much shorter and, hence, stronger hydrogen bond than does atom O(3) in molecule *B*.

The four atoms of the carboxylate groups are satisfactorily coplanar (Table 8). The C–C and C–O distances are typical of amino acids, as are the small deviations of atoms N(1) from the planes of the carboxylate groups (Marsh & Donohue, 1967). The bond angles in the amino-acid grouping show significant differences in the two molecules (Table 6); these differences presumably reflect the different hydrogen-bonding environments.

(ii) *Molecular conformation*

The most striking result of this investigation is the large difference in conformation of the two molecules. Views of the two molecules and of the histidine moiety in four other crystal structures are shown in Fig. 5. As pointed out previously (Kistenmacher & Marsh, 1971), molecule *A* adopts a 'closed' conformation similar to that found in L-histidine hydrochloride (Donohue, Lavine & Rollett, 1956; Donohue & Caron, 1964) while molecule *B* adopts an 'open' conformation as in crystals of DL-histidine hydrochloride (Bennett *et al.*, 1970) and in two modifications of L-histidine (Madden, McGandy & Seeman, 1972; Madden *et al.*, 1972). The difference between the two conformations is reflected primarily in the torsion angle about C(2)–C(3); a small value of this torsion angle results in a closed conformation in which the imidazole ring is folded back on top of the carboxylate group, while a larger value (near 180°) results in an extended or open conformation.

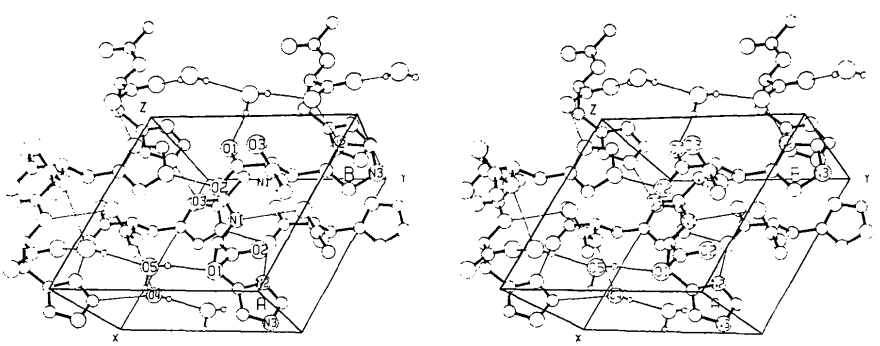

Fig. 6. A stereoscopic view of the crystal structure of *N*-acetylhistidine. The labeled molecules correspond to the coordinates given in Table 3.

Values of this and other torsion angles are listed in Table 9.

Table 9. *Torsion angles*

A positive angle corresponds to a right-handed screw. Standard deviations are about 0·3°. Values for L-histidine and DL-histidine crystals are given by Madden *et al.* (1972).

Angle	Molecule *A*	Molecule *B*
N(1)–C(2)–C(1)–O(2)	−19·1°	−14·9°
C(4)–C(3)–C(2)–C(1)	−44·0	172·2
N(2)–C(4)–C(3)–C(2)	−75·6	−86·9
C(7)–N(1)–C(2)–C(1)	−152·0	−80·2

The torsion angle about C(2)–C(3) in molecule *A* is the smallest yet found for histidine compounds. The resulting conformation is extremely tight, leading to a very short C(1)···C(4) non-bonded distance of 2·82 Å (Fig. 5) and other evidence of molecular strain. This evidence includes: (1) a lengthening of the C(2)–C(3) bond by nearly 0·02 Å relative to the value found in molecule *B*; (2) a bending of the C(4)–C(3) bond out of the plane of the imidazole ring, by about 5° (Table 8). There seems to be little doubt that the crystal forces – hydrogen bonds, or whatever – have been sufficient to cause molecule *A* to adopt a strained conformation; they must also be responsible for the large difference in the torsion angles about the N(1)–C(2) bonds (Table 9). To quote from Bennett *et al.* (1970): 'These results emphasize that predictions about conformation made on the basis of minimum energy or any other calculations for an isolated molecule, may not always be correct for the molecule in a crystal – or in any other environment.'

(iii) *Crystal packing and hydrogen bonding*

A stereoscopic view (Johnson, 1965) of the structure is shown in Fig. 6; details of the hydrogen bonding are given in Table 10. All protons attached to nitrogen or oxygen atoms are involved in the hydrogen bonding. We also include in the list the $C(5)_A$–H···O(4) interaction; although the C···O separation is rather large (3·20 Å), the angle at the hydrogen atom is close to 180° and the H···O separation is somewhat shorter than the sum of the van der Waals radii.

The environments of the water molecules are shown in Fig. 7. Atom O(4) has four close neighbors [including the C(5)–H group] in a roughly tetrahedral arrangement; the thermal motion of O(4) is moderately small and isotropic. Atom O(5), on the other hand, is involved in but three, closely coplanar, hydrogen bonds. It shows a highly anisotropic thermal motion with the major axis directed only 14° from the normal to the mean plane of O(5), O(4), $O(1)_A$, and $O(1)_B$.

Additional C–H···O interactions are indicated in Fig. 7. While the H···O distances are too long to correspond to significant hydrogen bonds, the geometries of these interactions are of some interest. Other intermolecular contacts include $C(5)_B$–H···$O(1)_B$, with H···O=2·50 Å and C···O=3·000 Å, and $N(3)_B$–H···$O(1)_B$, with H···O=2·45 Å and N···O=2·983 Å. The tendency of the C(5)–H(6) groups in both molecules to participate in relatively close interactions with neighboring electronegative groups presumably reflects a partial positive charge on C(5), as one might expect.

There are no other unusual intermolecular contacts, nor is there any appreciable stacking of the imidazole rings. It seems clear that the strong intermolecular hydrogen-bonding forces have the predominant influence

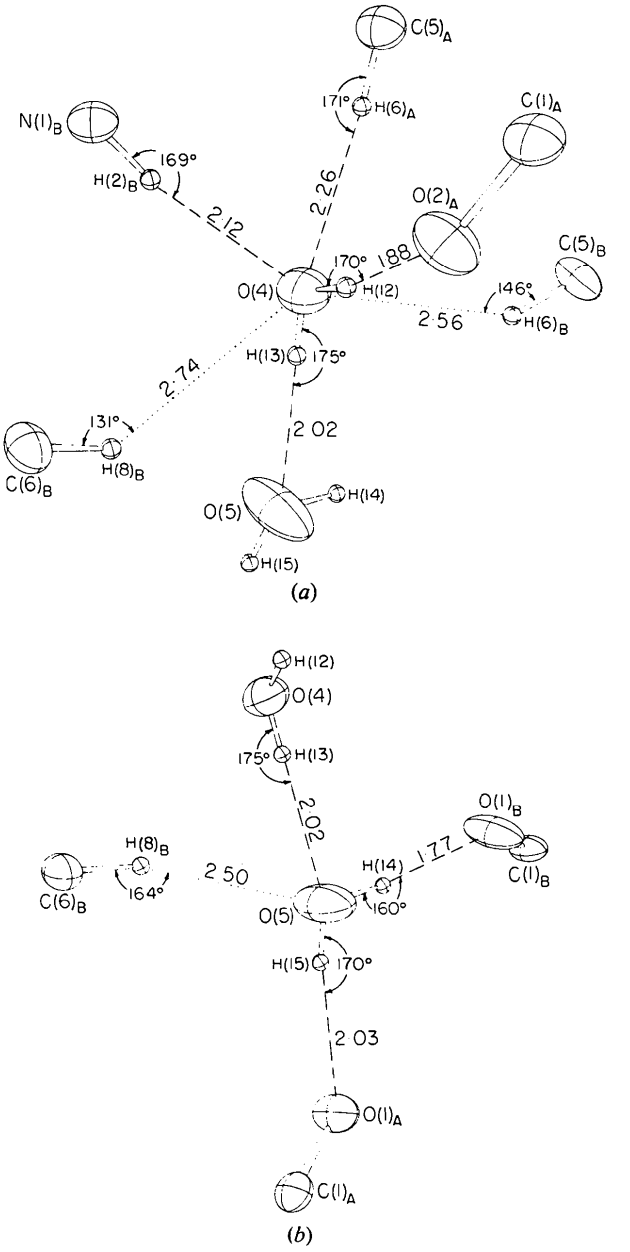

Fig. 7. Perspective views showing the environments of the water molecules. Hydrogen bonds are indicated by dashed lines. The thermal ellipsoids of the heavy atoms are drawn at the 40%-probability level; the hydrogen atoms are represented as artificially small spheres. (*a*) The surroundings of O(4). (*b*) The surroundings of O(5).

Table 10. *Distances and angles in the hydrogen bonds*, A–H···B

A	H	B	A···B	H···B	∠H–A···B
$N(1)_A$	$H(2)_A$	$O(3)_B(a)$	3·207 Å	2·33 Å	156°
$N(1)_B$	$H(2)_B$	O(4) (*b*)	2·963	2·12	169
$N(2)_A$	$H(5)_A$	$O(2)_B(a)$	2·688	1·77	164
$N(2)_B$	$H(5)_B$	$O(1)_A(c)$	2·664	1·65	174
$N(3)_A$	$H(7)_A$	$O(3)_A(d)$	2·711	1·80	167
$N(3)_B$	$H(7)_B$	$O(2)_B(e)$	2·707	1·76	171
O(4)	H(12)	$O(2)_A(f)$	2·736	1·88	170
O(4)	H(13)	O(5) (*a*)	2·783	2·02	175
O(5)	H(14)	$O(1)_B(g)$	2·703	1·77	160
O(5)	H(15)	$O(1)_A$	2·805	2·03	170
$C(5)_A$	$H(6)_A$	O(4) (*e*)	3·199	2·26	171

(*a*) $x, y, z-1$ (*b*) $x, y+1, z+1$ (*c*) $x-1, y, z$ (*d*) $x+1, y+1, z$
(*e*) $x, y+1, z$ (*f*) $x, y-1, z-1$ (*g*) $x+1, y, z$

in determining the conformations of the molecules. There is no intramolecular hydrogen bond. On the other hand, both crystal forms of L-histidine (Madden *et al.*, 1972) show strong internal hydrogen bonds from the protonated ammonium group N(1) to atom N(2) of the imidazole ring. This hydrogen bond presumably stabilizes the 'open' conformation of the molecules. Such stabilization would probably not be available to a histidine residue within the peptide chain of a protein, as the donor nitrogen atom would be part of a peptide group and the hydrogen bond would be less favorable. Without much doubt, then, the conformation of the histidine residue in a particular polypeptide chain in a protein will be dictated by its local hydrogen-bonding environment.

References

BENNETT, I., DAVIDSON, A. G. H., HARDING, M. M. & MORELLE, I. (1970). *Acta Cryst.* B**26**, 1722.

CORNELL, N.W. & GURNEY, D. G. (1970). *Biochem. Biophys. Res. Commun.* **40**, 530.

CRAIG, L. C., PHILIPS, W. F. & BURACHIK, M. (1969). *Biochem. Wash.* **8**, 2348.

DONOHUE, J. & CARON, A. (1964). *Acta Cryst.* **17**, 1178.

DONOHUE, J., LAVINE, L. R. & ROLLETT, J. S. (1956). *Acta Cryst.* **9**, 655.

DUCHAMP, D. J. (1969). Private communication.

FREEMAN, H. C. (1967). *Advanc. Protein Chem.* **22**, 235.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274.

International Tables for X-ray Crystallography (1962). Vol. III, p. 202. Birmingham: Kynoch Press.

JOHNSON, C. K. (1965). *ORTEP*, Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

KARLE, J. (1968). *Acta Cryst.* B**24**, 182.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.

KISTENMACHER, T. J. & MARSH, R. E. (1971). *Science*, **172**, 945.

LARSON, A. C. (1968). *Acta Cryst.* **23**, 664.

MADDEN, J. J., MCGANDY, E. L. & SEEMAN, N. C. (1972). Private communication and *Acta Cryst.* B**28**, 2377.

MADDEN, J. J., MCGANDY, E. L., SEEMAN, N. C., HARDING, M. M. & HOY, A. (1972). Private communication and *Acta Cryst.* B**28**, 2382.

MARSH, R. E. & DONOHUE, J. (1967). *Advanc. Protein Chem.* **22**, 235.

ROSSMAN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24.

STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175.

TANFORD, C. & EPSTEIN, J. (1954). *J. Amer. Chem. Soc.* **76**, 2170.

ZACHARIASEN, W. H. (1952). *Acta Cryst.* **5**, 68.

ZACHARIASEN, W. H. (1963). *Acta Cryst.* **16**, 1139.