Coexistence of both histidine tautomers in the solid state and stabilisation of the unfavourable Nδ–H form by intramolecular hydrogen bonding: crystalline L-His-Gly hemihydrate

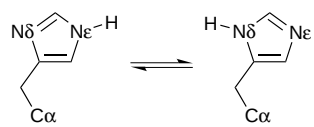
Thomas Steiner* and Gertraud Koellner

Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany

The crystal structure of L-His-Gly hemihydrate contains two independent peptide molecules which exhibit different histidine tautomers; the less favourable N(δ)–H form is stabilized by an intramolecular Nδ(*n*)–H···O(*n* + 1) hydrogen bond facilitated by a suitable molecular conformation.

It is well known that the imidazole moiety of the amino acid histidine (His), if unchanged, may exist in two tautomers (Scheme 1). At pH 7, the Nε–H form is dominant. The microscopic p*K* values are 6.7 for the Nδ–H tautomer, and 6.1 for the Nε–H tautomer (determined by titration experiments on a series of suitable derivates).[1] In consequence, histidine peptide crystals grown from water are normally found with the Nε–H form of His, and if grown at low p*H*, His is positively charged with Nδ and Nε both carrying hydrogen atoms.[2,3] The present communication reports for the first time on the stable coexistence of both tautomers in one peptide crystal structure.

From solution in pure water, the dipeptide L-histidyl-glycine crystallizes as a hemihydrate (L-His-Gly·0.5H_2O, **1**), with two peptide molecules and one water molecule per asymmetric crystal unit† (crystal structures of the chloride and dichloride salts have been published previously[4]). Very surprisingly, the two symmetry-independent molecules are found in completely different conformations, and with different tautomers of the histidine imidazole ring (Fig. 1, selection of torsion angles given in the figure legend). The tautomeric identity has been unambiguously determined by location of the hydrogen atoms in difference Fourier calculations and subsequent isotropic refinement, but could in principle also be deduced from the crystal contacts of the non-hydrogen atoms.‡ A section of the crystal packing is shown in Fig. 2, with some geometrical hydrogen bond data given in the figure legend. Molecule **A** adopts a standard elongated conformation and the imidazole moiety is in the expected Nε-H form, very similar to the closely related molecule L-His-L-Ala crystallized from water as a dihydrate.[5] Molecule **B**, on the contrary, is found in a folded conformation as the Nδ–H histidine tautomer [Fig. 1(*b*)]. This is associated with a short and linear intramolecular side-chain to main-chain hydrogen bond Nδ(1)–H··O(2) [for N–H normalized to 1.04 Å: N··O = 2.742(5) Å, H··O = 1.71 Å, N–H··O = 170°]. This situation is exceptional not only from the chemical, but also from the crystallographic point of view: typically in crystal structures with more than one moleule per asymmetric crystal unit, the symmetry-independent molecules may have different conformations, but have the same chemical identities. It is obvious that crystal growth of **1** requires unusually complex mechanisms of molecular recognition at the crystal–solvent interface to incorporate molecules of both identities at the right places. Once incorporated in the crystal, the individual molecules are trapped in their tautomeric form.



Scheme 1

Previous surveys have reported that in crystal structures with uncharged His side-chains, *only* the more stable Nε–H tautomer is found[2,3] (conflicting with the present results). Since several relevant crystal structures have been published after the surveys, a new search through the Cambridge Structural Database[6] (CSD) was performed (spring 1996 version, R < 0.08, all hydrogen atoms given, no metal atoms present). Of 39 His residues in 36 crystal structures, 20 are charged. Of the remaining 19 His residues, 18 are found in the Nε–H, and one in the Nδ–H form. The latter occurs in unsolvated L-His-L-Leu, **2**.[7] No previous example of the solid-state coexistence of *both* tautomers is found.

Notably, the backbone and histidine conformation of **2** is almost identical to that of molecule **1B**, and the intramolecular Nδ(1)–H··O(2) hydrogen bond is formed with very similar geometry (N··O = 2.71 Å, H··O = 1.70 Å, N–H··O = 163° for normalized hydrogen position). Since the intermolecular surroundings are very different for **1B** and **2**, it is assumed that

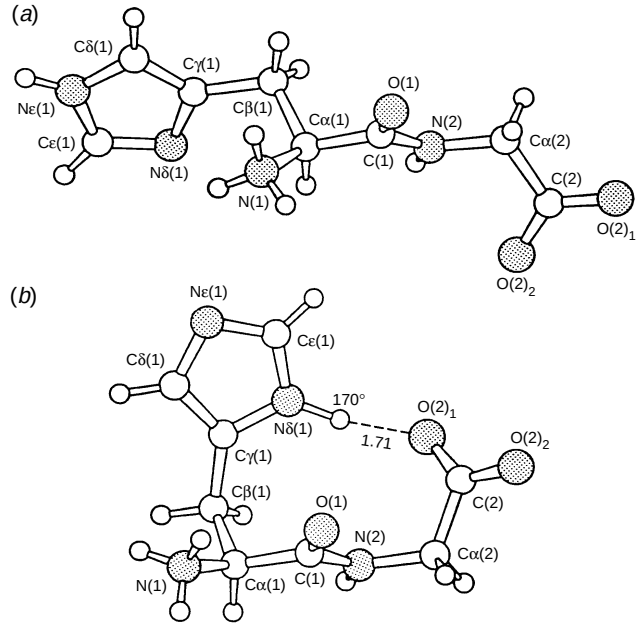


Fig. 1 Molecular structure and atomic numbering scheme for both peptide molecules in L-His-Gly hemihydrate. (*a*) molecule **A**, (*b*) molecule **B**. The molecules are drawn in the same projection with respect to the peptide bonds. Oxygen and nitrogen atoms are drawn shaded. Selection of torsion angles defining the molecular conformation: Molecule **A**: ψ_1 = 147.0(4)°; ω_1 = −178.1(4)°; φ_2 = 130.4(5)°; N(2)–Cα(2)–C(2)–$O(2)_1$ = −179.8(5)° (corresponds to ψ_2); N(2)–Cα(2)–C(2)–$O(2)_2$ = −1.1(7)°; χ_1^1 = −52.4(5)°; χ_1^2 = −58.2(6)°. Molecule **B**: ψ_1 = 166.2(4)°; ω_1 = 179.1(4)°; φ_2 = −78.4(6)°; N(2)–Cα(2)–C(2)–$O(2)_1$ = −27.1(7)°; N(2)–Cα(2)–C(2)–$O(2)_2$ = 154.5(4)° (corresponds to ψ_2); χ_1^1 = 69.5(6)°; χ_1^2 = 78.4(6)°.

the occurrence of the less favourable tautomer is not induced by the crystal surrounding, but is closely associated with the conformation, in particular with the formation of a short intramolecular Nδ(n)–H··O(n + 1) hydrogen bond. This interaction appears to stabilize the Nδ–H tautomer to a sufficient degree to allow crystallisation in this form. In a polypeptide chain, this kind of hydrogen bond represents an interaction of His Nδ–H with the carbonyl oxygen atom of the next residue, *i.e.* Nδ(n)–H···O(n + 1). The peptide conformation is required to be close to that of **1B** (ψ_n = 166°, φ_{n+1} = −78°, ψ_{n+1} = 155°, χ_n^1 = 69°, χ_n^2 = 78°). This is not an uncommon conformation, and several examples are quoted in ref. 7; a closer examination of these examples, however, shows that all except **1B** and **2** are formed by peptides with charged His residues, where Nδ is protonated irrespective of the conformation.

It is tempting to speculate on the relevance of such observations for proteins. Side-chain to main-chain His Nδ(n)··O(n + 1) hydrogen bonds do occur in some proteins[8] (normally, however, the charge situation is not unambiguously determined), and might there stabilize the less favourable tautomeric form. More generally, it can be assumed that the Nδ–H form can be effectively stabilized also by other hydrogen bond configurations;§ this assumption is not yet backed up by peptide structure analysis.

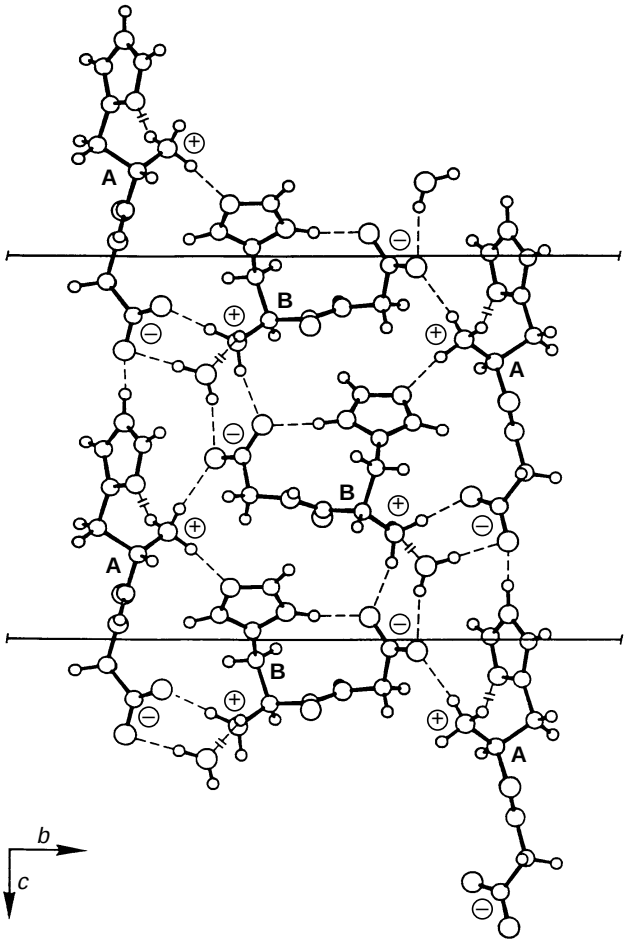


Fig. 2 Section of the crystal packing, O/N–H··O/N hydrogen bonds are shown as dashed lines. Molecules **A** and **B** are labelled near the Cα(1) atoms. The charge situation is indicated by symbols ⊕ and ⊖. The imidazole rings are involved in the following N–H··O/N hydrogen bond interactions (for normalized hydrogen atom positions with N–H = 1.04 Å): Molecule **A**: Nδ(1) (**A**)··H$_2$–N(1) (**A**) (x − 1, y, z), H··N = 1.81 Å, N··N = 2.815(6) Å, N–H··N = 160°. Nε(1) (**A**)–H··O(2)$_1$ (**A**) (x, y, z − 1), H··O = 1.73 Å, N··O = 2.715(6) Å, N–H··O = 156°. Molecule **B**: Nδ(1) (**B**)–H··O(2)$_1$ (**B**) (x, y, z), H··O = 1.71 Å, N··O = 2.742(5) Å, N–H··O = 170°. Nε(1) (**B**)··H$_3$–N(1) (**A**) (x, y, 1 + z), H··N = 1.82 Å, N··N = 2.859(6) Å, N–H··N = 177°.

Footnotes

* E-mail: steiner@chemie.fu-berlin.de

† A commercial sample of L-His-Gly (Sigma) was dissolved in water; slow evaporation yielded very thin, needle shaped colourless crystals.

Crystal data: L-histidyl-glycine hemihydrate, $C_8H_{12}N_4O_3 \cdot 0.5H_2O$; FW = 221.2, space group orthorhombic $P2_12_12_1$, a = 4.9547(6), b = 32.356(6), c = 12.587(4) Å, V = 2017.9(8) Å^3, Z = 8, Z' = 2, D_c = 1.46 g cm^{-3}. Enraf-Nonius Turbo-CAD diffractometer, room temperature, Ni-filtered Cu-Kα X-rays, ω-scan mode, $\lambda/2 \sin\theta_{max}$ = 0.89 Å, 0.65 × 0.04 × 0.02 mm^3 crystal glued to a glass pin, 3436 reflections measured, 2999 unique, 1889 with $I > 2\sigma(I)$, no absorption correction. Structure solution and refinement with standard methods (SHELXS86[10] and SHELXL93[11]). All hydrogen atom positions were located from difference Fourier calculations; to obtain the geometrically most realistic model, the hydrogen atoms were then treated in the riding model,[11] with exception of the hisidine (N)H atoms which were refined unconstrained isotropic, the water H-atoms, which were refined with O–H bond lengths constrained to 0.90(2) Å, and the ammonium groups which were allowed to rotate. Final R = 0.059 (for observed reflections), $\omega R(F^2)$ = 0.148 (for all reflections). Atomic coordinates, bond lengths and angles, and thermal parameters have been deposited at the Cambridge Crystallographic Data Centre (CCDC). See Information for Authors, Issue No. 1. Any request to the CCDC for this material should quote the full literature citation and the reference number 182/467.

‡ The tautomeric forms of molecules **A** and **B** can be deduced merely from the configuration of non-hydrogen atoms: Nε(1) (**B**) is in hydrogen bond contact with an ammonium group and can therefore only be non-protonated, and Nδ(1) (**B**) forms a hydrogen bond with a carboxylate group and must therefore carry a hydrogen atom (Fig. 2). For the 'conventional' molecule **A**, the situation is reversed: Nε(1) (**A**) is in hydrogen bond contact with a carboxylate group, and Nδ(1) (**A**) with an ammonium group, inferring the Nε–H tautomer for His(**A**).

§ In the well-defined example of a glutamic acid specific serine protease (Glu-SGP), such a case seems actually to occur: a charge-relay system involving a remarkable triad of presumably uncharged histidine residues was proposed (ref. 9). Based on the configuration of the non-hydrogen atoms, it can be inferred that in this triad, initially two residues are in the Nε–H and one in the Nδ–H form. Upon action of the charge-relay mechanism, the situation is switched to two Nδ–H and one Nε–H tautomer.

References

1 M. Tanokura, *Biochim. Biophys. Acta*, 1983, **742**, 576.
2 R. Ramani and R. J. Boyd, *Can. J. Chem*., 1981, **59**, 3232.
3 C. H. Görbitz, *Acta Crystallogr., Sect. B*, 1989, **45**, 390.
4 T. Steiner, *Acta Crystallogr., Sect. C*, 1997, **53**, 255; *ibid*, in the press.
5 T. Steiner, *Acta Crystallogr., Sect. C*, 1996, **52**, 2554.
6 F. H. Allen, J. E. Davies, J. E. Galloy, J. J. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith and D. G. Watson, *J. Chem. Inf. Comput. Sci*., 1991, **31**, 187.
7 J. A. Krause, P. W. Baures and D. S. Eggleston, *Acta Crystallogr., Sect. B*, 1993, **49**, 123.
8 E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol*., 1984, **44**, 97.
9 V. L. Nienaber, K. Breddam and J. J. Birktoft, *Biochemistry*, 1993, **32**, 11469.
10 G. A. Sheldrick, SHELXS86, *Program for the solution of crystal structures*, 1986, University of Göttingen, Germany.
11 G. A. Sheldrick, SHELXL93, *Program for the refinement of crystal structures*, 1993, University of Göttingen, Germany.

Received in Cambridge, UK, 21st November 1996; Revised manuscript received, 14th March 1997; Com. 7/01807J