Structure Determination

International Edition: DOI: 10.1002/anie.201411520
German Edition: DOI: 10.1002/ange.201411520

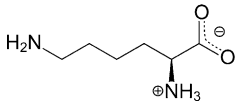
L-Lysine: Exploiting Powder X-ray Diffraction to Complete the Set of Crystal Structures of the 20 Directly Encoded Proteinogenic Amino Acids**

*P. Andrew Williams, Colan E. Hughes, and Kenneth D. M. Harris**

Abstract: *During the last 75 years, crystal structures have been reported for 19 of the 20 directly encoded proteinogenic amino acids in their natural (enantiomerically pure) form. The crystal structure is now reported for the final member of this set:* L*-lysine. As crystalline* L*-lysine has a strong propensity to incorporate water under ambient atmospheric conditions to form a hydrate phase, the pure (non-hydrate) crystalline phase can be obtained only by dehydration under rigorously anhydrous conditions, resulting in a microcrystalline powder sample. For this reason, modern powder X-ray diffraction methods have been exploited to determine the crystal structure in this final, elusive case.*

Over the past 75 years, progress has been made towards determination of the crystal structures of the complete set of the 20 directly encoded proteinogenic amino acids in their natural (enantiomerically pure) form,[1] a task begun in 1939 with the first determination of a crystal structure of glycine.[2] The 20 directly encoded proteinogenic amino acids are those that are encoded directly by the codons of the standard genetic code. Until recently, the crystal structures of 17 of these amino acids had been determined, the remaining cases being L-arginine, L-tryptophan, and L-lysine. Within the last two years, the crystal structures of L-arginine[3] and L-tryptophan[4] have now been reported. While the crystal structure of L-tryptophan was determined from single-crystal X-ray diffraction (XRD) data after significant efforts to produce single crystals of sufficient size and quality, the crystal structure of L-arginine was determined directly from powder XRD data, taking advantage of advances in methodology that have taken place during the last 20 years or so, which now allow the crystal structures of organic materials to be determined directly from powder XRD data.[5] Clearly, the availability of such techniques circumvents the requirement to prepare single crystals. Furthermore, several new polymorphs of L-phenylalanine (denoted polymorphs II,[6] III,[7] and IV[8]) have been discovered recently and their crystal structures reported.

Herein, we report the crystal structure of the final member of the set of 20 amino acids for which the crystal structure has not yet been reported: L-lysine (Scheme 1). Prior to this work, there were no known crystal structures of L-lysine (or D-lysine), DL-lysine, or any hydrate or solvate of either L-lysine (or D-lysine) or DL-lysine. However, the crystal structures of several cocrystals of L-lysine are known (the Cambridge Structural Database[9] contains 64 entries for cocrystals of L-lysine or DL-lysine), including simple salts such as L-lysine hydrofluoride[10] and L-lysine hydrochloride dihydrate,[11] and cocrystals formed with other amino acids, such as L-aspartic acid[12] and D-glutamic acid.[13]



Scheme 1. Molecular structure of L-lysine.

The fact that the crystal structure of L-lysine has remained undetermined (prior to the present work) is almost certainly due to the challenge of growing single crystals of sufficient size and quality for single-crystal XRD. Furthermore, solid L-lysine is very hygroscopic and readily incorporates atmospheric water. As a consequence, solid L-lysine is found to exist as a hydrate phase under ambient atmospheric conditions and deliquesces when exposed to an atmosphere of sufficiently high humidity. To obtain the anhydrous form of solid L-lysine, a rigorous procedure is required for dehydration of the hydrate phase. However, such solid-state dehydration procedures typically lead to a microcrystalline powder of the anhydrous phase, necessitating the use of powder XRD for structure determination. Furthermore, it is important to develop experimental procedures that ensure that the anhydrous material does not rehydrate prior to, or during, measurement of the powder XRD data.

The sample of L-lysine used here was purchased from Sigma–Aldrich (purity $\geq 98\,\%$) and was identified as a hydrate phase. Anhydrous L-lysine was prepared by subjecting this sample to a drying procedure in which it was stored for an extended period of time (three weeks) in the presence of P_2O_5 in a sealed desiccator.[14]

To prepare the sample of anhydrous L-lysine for the powder XRD data collection, the purchased (hydrated) sample of L-lysine was mixed with amorphous starch in a 2:1 ratio (to reduce the effects of preferred orientation in the powder sample) and packed into three glass capillaries which were left unsealed and placed in a sealed sample vial containing P_2O_5 for three weeks. After this period of dehydration, the capillaries were rapidly flame-sealed to prevent rehydration and transferred to the sample stage of the powder XRD instrument for measurement of the powder XRD data. We note that the powder XRD pattern of the dehydrated material is significantly different from that of the original purchased sample of L-lysine (that is, the hydrated phase before dehydration). In separate experiments, powder XRD confirmed that the dehydrated material transforms

[*] Dr. P. A. Williams, Dr. C. E. Hughes, Prof. K. D. M. Harris
School of Chemistry, Cardiff University
Cardiff CF10 3AT, Wales (UK)
E-mail: HarrisKDM@cardiff.ac.uk

[**] We are grateful to Cardiff University for financial support.

back to the same phase as the original sample on exposure to the atmosphere, consistent with the assignment that the original sample is a hydrate phase.

The powder XRD pattern[15] of L-lysine was indexed using the KOHL[16] algorithm in the CRYSFIRE[17] indexing suite, giving the following unit cell with monoclinic metric symmetry: $a = 9.54$ Å, $b = 5.14$ Å, $c = 17.06$ Å, $\beta = 97.7°$ ($V = 828.9$ Å^3). Profile-fitting and unit-cell refinement, carried out using the Le Bail procedure[18] in the program GSAS,[19] gave a good quality of fit (Figure 1, $R_{wp} = 3.13\,\%$, $R_p = 2.34\,\%$). On the basis of systematic absences, the space group was assigned as $P2_1$. From density considerations, the crystal structure has four molecules of L-lysine in the unit cell and hence two independent molecules in the asymmetric unit for space group $P2_1$.

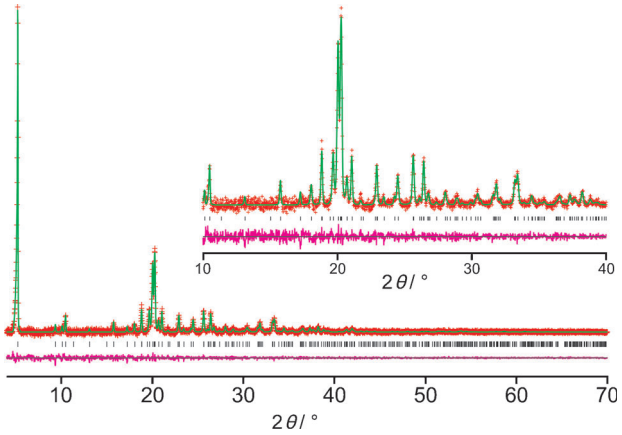


Figure 1. Le Bail fit for the powder XRD pattern (with background subtracted) of L-lysine, showing the experimental data (red + marks), calculated data (green line), difference plot (magenta line), and predicted peak positions (black tick marks). The inset shows an expanded view of the region $10° \le 2\theta \le 40°$.

Structure solution was carried out directly from the powder XRD data using the direct-space genetic algorithm technique[20] in the program EAGER.[21] In the direct-space structure-solution calculations, each L-lysine molecule was in the zwitterionic form shown in Scheme 1 and the crystal structure was defined by 21 variables. One molecule was defined by two positional variables (for space group $P2_1$, the origin can be fixed arbitrarily along the b-axis, allowing the positional variable along this axis to be fixed for one molecule), three orientational variables, and five torsional variables (total variables = 10). The second molecule was defined in the same way but with three positional variables (total variables = 11).

In total, 16 independent structure-solution calculations were carried out. In each case, the population comprised 400 trial structures, with 40 mating operations and 200 mutation operations per generation. Each calculation was run for 500 generations. The same structure solution of highest quality was obtained in 10 cases. This structure was then used as the starting model for Rietveld refinement,[22] carried out using the GSAS[19] program.

In the Rietveld refinement, standard restraints were applied to bond lengths and angles, and planar restraints were applied to the carboxylate groups. Two common isotropic displacement parameters were refined for the non-hydrogen atoms (one value for each molecule). The isotropic displacement parameter of the hydrogen atoms was set to 1.2 times the refined value for the non-hydrogen atoms in the same molecule. After Rietveld refinement, the non-hydrogen atoms were found to occupy structurally sensible positions, although the hydrogen-bonding arrangement (particularly for the NH_2 groups) was geometrically non-optimal (this situation is not uncommon in structure solution of organic materials from powder XRD data, reflecting the fact that XRD is rather insensitive to the locations of hydrogen atoms). The structure was optimized through use of DFT energy minimization using CASTEP[23] (Academic Release version 6.1). The calculation was performed with a basis set cut-off energy of 700 eV, ultrasoft pseudopotentials, PBE functional, semiempirical dispersion corrections (TS correction scheme), fixed unit cell, preserved space group symmetry ($P2_1$), and periodic boundary conditions. The energy minimization produced a geometrically more favorable hydrogen-bonding arrangement but with no significant change in the rest of the structure.

The crystal structure with optimized hydrogen bonding was used as the initial structural model for further Rietveld refinement, carried out as described above but with intermolecular distance restraints applied to retain the hydrogen-bond geometry. As discussed below, disorder was assessed for the positions of the hydrogen atoms of the side-chain NH_2 groups. The final Rietveld refinement gave a good fit to the experimental powder XRD data (Figure 2; $R_{wp} = 3.38\,\%$, $R_p = 2.50\,\%$) with final refined parameters: $P2_1$, $a = 9.5060(7)$ Å, $b = 5.1289(4)$ Å, $c = 16.9941(13)$ Å, $\beta = 97.697(9)°$; $V = 821.09(12)$ Å^3 (2θ range, 4–70°; 3867 profile points; 177 refined variables).[24] In the following discussion, the two crystallographically independent L-lysine molecules are denoted A and B (where A is the first molecule listed in the crystal information file).[24]

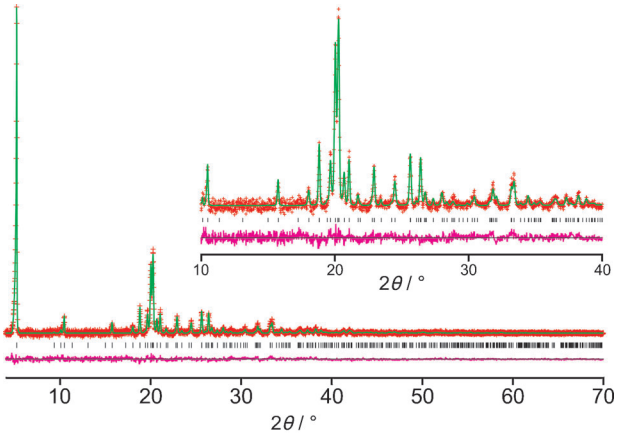


Figure 2. Final Rietveld fit for the background-subtracted powder XRD pattern of L-lysine. The inset shows an expanded view of the region $10° \le 2\theta \le 40°$.

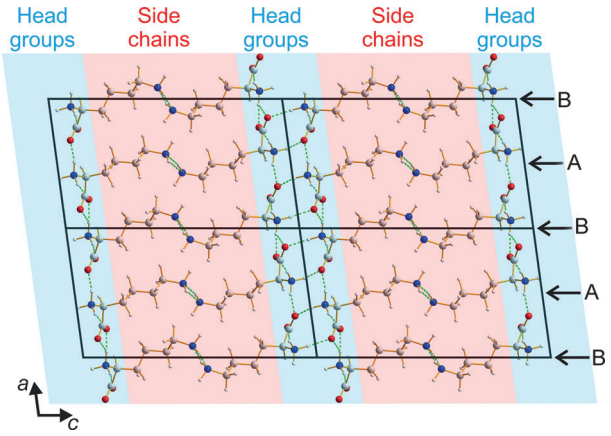


Figure 3. Crystal structure of L-lysine viewed along the *b*-axis showing the layered nature of the structure (the plane of the layers is parallel to the *ab*-plane). Regions containing the head groups are shaded in cyan and regions containing the side chains are shaded in red. Dashed green lines represent hydrogen bonds. The two crystallographically independent molecules are labeled A and B.

The crystal structure of L-lysine (Figure 3) may be described as a layered arrangement, in common with the crystal structures of many other amino acids (L-isoleucine,[25] L-leucine,[26] L-methionine,[27] polymorphs I,[8] II,[6] III,[7] and IV[8] of L-phenylalanine, L-tryptophan,[4] and L-valine[28]). However, in contrast to these other amino acids, the side chain of L-lysine has an NH_2 group at the chain terminus, which is able to engage in hydrogen-bonding interactions.

The crystal structure of L-lysine contains two different hydrogen-bonded regions. One hydrogen-bonded region is formed by the head groups of the two independent L-lysine molecules, with the ammonium groups acting as hydrogen-bond donors and the carboxylate groups acting as hydrogen-bond acceptors; as shown in Figure 3, this region is composed of two hydrogen-bonded sheets that are connected to each other by further hydrogen bonds. The second hydrogen-bonded region involves the NH_2 groups at the termini of the side chains, with hydrogen bonding occurring solely between NH_2 groups that are related to each other by the 2_1 screw axis. As a consequence, there are two independent chains of N–H···N–H···N–H hydrogen bonds along the *b*-axis (Figure 4), with one chain involving only molecules of type A and the other chain involving only molecules of type B. It is conceivable that these hydrogen-bonded chains could run either along the positive direction of the *b*-axis (that is, N–H···N–H···N–H) or along the negative direction of the *b*-axis (H–N···H–N···H–N), and disorder models incorporating fractional populations of both of these hydrogen-bonding arrangements were considered in the Rietveld refinement. However, the quality of fit to the powder XRD data for these disordered models was indistinguishable from the quality of fit for ordered hydrogen-bonding arrangements. For this reason, in the final Rietveld refinement, the positions of the hydrogen atoms of the NH_2 groups within these hydrogen-bonded chains were selected on the basis of the positions that are considered to correspond to the most favorable hydrogen-bonding geometry.[29]

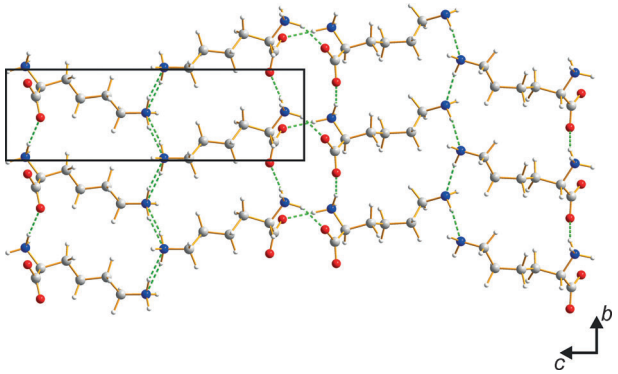


Figure 4. Part of the crystal structure of L-lysine viewed along the *a*-axis, showing the two independent N–H···N–H···N–H hydrogen bonded chains that run along the *b*-axis (vertical). The chain in the left half of the figure involves only molecules of type A. The chain in the right half of the figure involves only molecules of type B.

The two crystallographically independent molecules in the crystal structure have distinctly different conformations[30] as evident from the overlays in Figure 5.

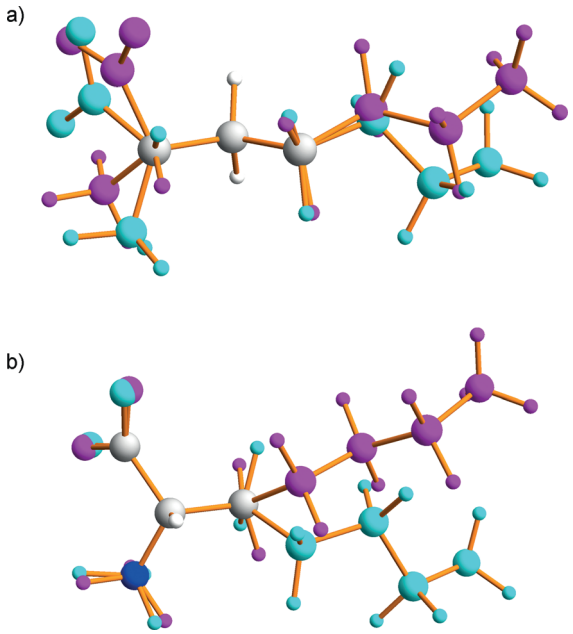


Figure 5. Overlays showing that the two crystallographically independent molecules in the crystal structure of L-lysine have different conformations: a) with Cα, Cβ, and Cγ superimposed, b) with the carboxylate C, ammonium N, Cα, and Cβ superimposed. The non-superimposed atoms of the two molecules are shown in magenta (molecule A) and cyan (molecule B).

The crystal structures of amino acids can be classified by the hydrogen-bonding arrangement of the head groups using the classification system proposed by Görbitz et al.[31] According to this system, the hydrogen-bonded sheets (Figure 6) in the crystal structure of L-lysine are assigned to the L2 structure type, in common with L-isoleucine, L-leucine, L-methionine and L-valine (and also found in the co-crystal of

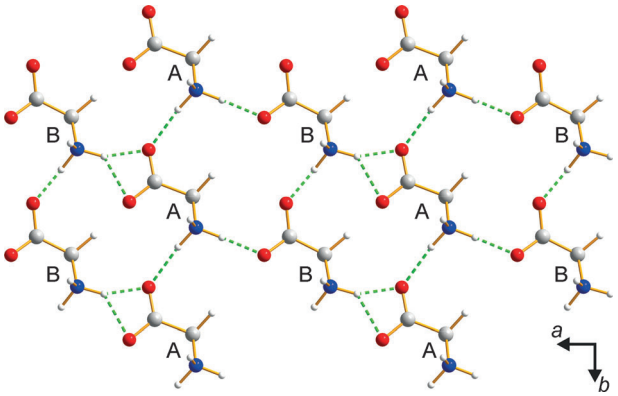


Figure 6. A single sheet within the hydrogen-bonded region of the crystal structure that involves the head groups of the L-lysine molecules. The two crystallographically independent molecules are marked A and B.

L-lysine with D-glutamic acid). The prevalence of the L2 structure type among this set of amino acids is in contrast to the formation of L1 sheets in polymorphs I and III of L-phenylalanine and L*x* sheets in polymorph IV of L-phenylalanine. The hydrogen-bonding arrangements in the crystal structures of L-tryptophan[4] and polymorph II of L-phenylalanine,[6] on the other hand, do not fit any of the classifications proposed by Görbitz et al.[31]

In conclusion, we emphasize that the crystal structure of L-lysine reported herein was determined directly from powder XRD data, necessitated by the fact that single crystals of this material of suitable size and quality for single-crystal XRD studies cannot be prepared. Furthermore, we highlight the fact that another essential component of the success of this work was the development of procedures for the preparation and handling of a rigorously anhydrous sample of L-lysine. In addition to completing the set of crystal structures of the 20 directly encoded proteinogenic amino acids, thus giving the opportunity to understand the solid-state structural properties of all members of this biologically important class of compounds, this work serves to demonstrate the significant opportunities that now exist for determining the crystal structures of organic materials by exploiting state-of-the-art techniques for the analysis of powder XRD data.

Keywords: amino acids · L-lysine · powder x-ray diffraction · structure determination

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 3973–3977
Angew. Chem. **2015**, *127*, 4045–4049

[1] One member (glycine) of this set of 20 amino acids is non-chiral. The other 19 amino acids in this set are chiral and are found to exist in nature only as the L-enantiomer.

[2] G. Albrecht, R. B. Corey, *J. Am. Chem. Soc.* **1939**, *61*, 1087–1103. (This crystal structure was subsequently named as the α polymorph of glycine, following the discovery of further polymorphs.).
[3] a) E. Courvoisier, P. A. Williams, G. K. Lim, C. E. Hughes, K. D. M. Harris, *Chem. Commun.* **2012**, *48*, 2761–2763; b) J. M. Thomas, *ChemPhysChem* **2012**, *13*, 2637–2638.
[4] C. H. Görbitz, K. W. Törnroos, G. M. Day, *Acta Crystallogr. Sect. B* **2012**, *68*, 549–557.
[5] a) K. D. M. Harris, M. Tremayne, P. Lightfoot, P. G. Bruce, *J. Am. Chem. Soc.* **1994**, *116*, 3543–3547; b) B. M. Kariuki, D. M. S. Zin, M. Tremayne, K. D. M. Harris, *Chem. Mater.* **1996**, *8*, 565–569; c) B. M. Kariuki, P. Calcagno, K. D. M. Harris, D. Philp, R. L. Johnston, *Angew. Chem. Int. Ed.* **1999**, *38*, 831–835; *Angew. Chem.* **1999**, *111*, 860–864; d) B. M. Kariuki, K. Psallidas, K. D. M. Harris, R. L. Johnston, R. W. Lancaster, S. E. Staniforth, S. M. Cooper, *Chem. Commun.* **1999**, 1677–1678; e) K. D. M. Harris, M. Tremayne, B. M. Kariuki, *Angew. Chem. Int. Ed.* **2001**, *40*, 1626–1651; *Angew. Chem.* **2001**, *113*, 1674–1700; f) V. V. Chernyshev, *Russ. Chem. Bull.* **2001**, *50*, 2273–2292; g) W. I. F. David, K. Shankland, L. B. McCusker, C. Baerlocher, IUCr/OUP, **2002**; h) K. D. M. Harris, *Cryst. Growth Des.* **2003**, *3*, 887–895; i) M. Tremayne, *Philos. Trans. R. Soc. London Ser. A* **2004**, *362*, 2691–2707; j) V. Favre-Nicolin, R. Černý, *Z. Kristallogr.* **2004**, *219*, 847–856; k) V. Brodski, R. Peschar, H. Schenk, *J. Appl. Crystallogr.* **2005**, *38*, 688–693; l) H. Tsue, M. Horiguchi, R. Tamura, K. Fujii, H. Uekusa, *J. Synth. Org. Chem. Jpn.* **2007**, *65*, 1203–1212; m) W. I. F. David, K. Shankland, *Acta Crystallogr. Sect. A* **2008**, *64*, 52–64; n) K. D. M. Harris, *Top. Curr. Chem.* **2012**, *315*, 133–178.
[6] P. A. Williams, C. E. Hughes, A. B. M. Buanz, S. Gaisford, K. D. M. Harris, *J. Phys. Chem. C* **2013**, *117*, 12136–12145.
[7] E. Mossou, S. C. M. Teixeira, E. P. Mitchell, S. A. Mason, L. Adler-Abramovich, E. Gazit, V. T. Forsyth, *Acta Crystallogr. Sect. C* **2014**, *70*, 326–331.
[8] F. S. Ihlefeldt, F. B. Pettersen, A. von Bonin, M. Zawadzka, C. H. Görbitz, *Angew. Chem. Int. Ed.* **2014**, *53*, 13600–13604; *Angew. Chem.* **2014**, *126*, 13818–13822.
[9] CSD Version 5.35 (November **2013**).
[10] R. P. Sukiasyan, H. A. Karapetyan, A. K. Atanesyan, R. A. Apreyan, *Spectrochim. Acta Part A* **2014**, *123*, 206–210.
[11] D. A. Wright, R. E. Marsh, *Acta Crystallogr.* **1962**, *15*, 54–64.
[12] T. N. Bhat, M. Vijayan, *Acta Crystallogr. Sect. B* **1976**, *32*, 891–895.
[13] J. Soman, C. G. Suresh, M. Vijayan, *Int. J. Pept. Protein Res.* **1988**, *32*, 352–360.
[14] Further studies focused on the hydration and dehydration behavior of L-lysine will be published in due course.
[15] The powder XRD pattern of L-lysine was recorded on a Bruker D8 instrument operating in transmission mode using Ge-monochromated Cu Kα_1 radiation (step size, 0.017°; data collection time, 15 s per step).
[16] F. Kohlbeck, E. M. Hörl, *J. Appl. Crystallogr.* **1978**, *11*, 60–61.
[17] R. Shirley, *The CRYSFIRE System for Automatic Powder Indexing: User's Manual*, The Lattice Press, Guildford, UK, **1999**.
[18] A. Le Bail, H. Duroy, J. L. Fourquet, *Mater. Res. Bull.* **1988**, *23*, 447–452.
[19] A. C. Larson, R. B. Von Dreele, *Los Alamos National Laboratory Report* **2004**, LAUR 86-748.
[20] a) B. M. Kariuki, H. Serrano-González, R. L. Johnston, K. D. M. Harris, *Chem. Phys. Lett.* **1997**, *280*, 189–195; b) K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Acta Crystallogr. Sect. A* **1998**, *54*, 632–645; c) G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Chem. Phys. Lett.* **2000**, *321*, 183–190; d) S. Habershon, K. D. M. Harris, R. L. Johnston, *J. Comput. Chem.* **2003**, *24*, 1766–1774; e) K. D. M. Harris, S. Habershon, E. Y. Cheung, R. L. Johnston, *Z. Kristallogr.* **2004**, *219*, 838–846.
[21] a) E. Tedesco, G. W. Turner, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Angew. Chem. Int. Ed.* **2000**, *39*, 4488–4491;

Angew. Chem. **2000**, *112*, 4662 – 4665; b) D. Albesa-Jové, B. M. Kariuki, S. J. Kitchin, L. Grice, E. Y. Cheung, K. D. M. Harris, *ChemPhysChem* **2004**, *5*, 414 – 418; c) F. Guo, K. D. M. Harris, *J. Am. Chem. Soc.* **2005**, *127*, 7314 – 7315; d) Z. Pan, M. Xu, E. Y. Cheung, K. D. M. Harris, E. C. Constable, C. E. Housecroft, *J. Phys. Chem. B* **2006**, *110*, 11620 – 11623; e) F. Guo, J. Martí-Rujas, Z. Pan, C. E. Hughes, K. D. M. Harris, *J. Phys. Chem. C* **2008**, *112*, 19793 – 19796; f) K. Fujii, M. T. Young, K. D. M. Harris, *J. Struct. Biol.* **2011**, *174*, 461 – 467; g) J. Martí-Rujas, L. Meazza, G. K. Lim, G. Terraneo, T. Pilati, K. D. M. Harris, P. Metrangolo, G. Resnati, *Angew. Chem. Int. Ed.* **2013**, *52*, 13444 – 13448; *Angew. Chem.* **2013**, *125*, 13686 – 13690.

[22] H. M. Rietveld, *J. Appl. Crystallogr.* **1969**, *2*, 65 – 71.

[23] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, M. C. Payne, *Z. Kristallogr.* **2005**, *220*, 567 – 570.

[24] CCDC 1042478 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

[25] C. H. Görbitz, B. Dalhus, *Acta Crystallogr. Sect. C* **1996**, *52*, 1464 – 1466.

[26] C. H. Görbitz, B. Dalhus, *Acta Crystallogr. Sect. C* **1996**, *52*, 1754 – 1756.

[27] K. Torii, Y. Iitaka, *Acta Crystallogr. Sect. B* **1973**, *29*, 2799 – 2807.

[28] B. Dalhus, C. Henrik Görbitz, *Acta Chem. Scand.* **1996**, *50*, 544 – 548.

[29] For the chains involving molecules of type B, the C–N···N angle for the C–N–H···N hydrogen bond is 105.8° for one direction of hydrogen bonding (that is, N–H···N–H···N–H) and 131.1° for the other direction of hydrogen bonding (that is, H–N···H–N···H–N). As 105.8° is significantly closer to the tetrahedral angle (corresponding to a more linear C–N–H···N hydrogen bond), this direction of the hydrogen-bonded chain is considered geometrically more favorable. For this reason, in the final refinement, the hydrogen bonding in this chain was ordered, with the H atom located (with occupancy of 1) only in the C–N–H···N hydrogen bond for which the C–N···N angle is 105.8°. For the chains involving molecules of type A, the C–N···N angle for the C–N–H···N hydrogen bond is almost identical (116.8° and 117.6°) for the two directions of hydrogen bonding and thus the N–H···N–H···N–H and H–N···H–N···H–N directions of hydrogen bonding are considered to be essentially equally probable on the basis of geometric considerations. For this reason, in the final refinement, the hydrogen bonding in this chain was disordered, with the H atom in each direction of the hydrogen-bonded chain having occupancy of 0.5.

[30] Torsion angles in molecule A: C(O)-Cα-Cβ-Cγ 100.0°, Cα-Cβ-Cγ-Cδ 159.4°, Cβ-Cγ-Cδ-Cε 174.1°, Cγ-Cδ-Cε-N 171.2°. Torsion angles in molecule B: C(O)-Cα-Cβ-Cγ 149.6°, Cα-Cβ-Cγ-Cδ 166.9°, Cβ-Cγ-Cδ-Cε 125.2°, Cγ-Cδ-Cε-N 161.3°.

[31] C. H. Görbitz, K. Vestli, R. Orlando, *Acta Crystallogr. Sect. B* **2009**, *65*, 393 – 400.

Received: November 28, 2014
Published online: February 4, 2015

