Crystal Structures

DOI: 10.1002/anie.201406886

The Polymorphs of L-Phenylalanine**

*Franziska Stefanie Ihlefeldt, Fredrik Bjarte Pettersen, Aidan von Bonin, Malgorzata Zawadzka, and Carl Henrik Görbitz**

Abstract: *The solid-state structure of the amino acid phenylalanine (Phe) offers a potential key to understanding the behavior of a large class of important aromatic compounds. Obtaining good single crystals is, however, notoriously difficult. The structure of the common polymorph of Phe, form I, was first reported by Weissbuch et al. (as* D*-Phe) in 1990, but the correctness of the published C2 unit cell with two disordered molecules in the asymmetric unit was later questioned and other space groups suggested. The identity of form I of* L*-Phe is here established to be $P2_1$ with $Z' = 4$, based on data from a well-diffracting single crystal grown from an acetic acid solution of the amino acid. A second new polymorph, form IV, together with the two recently described forms II and III provide unprecedented information on the structural complexity of this essential amino acid. It is furthermore documented that the racemate,* DL*-Phe, does not grow proper single crystals.*

Phenylalanine, Phe, is the biosynthetic precursor of tyrosine, signaling molecules such as dopamine, norepinephrine and epinephrine, and the skin pigment melanin. Inability to metabolize Phe is associated with the genetic disorder phenylketonuria (PKU). Phe accounts for about 4% of the amino acid residues in proteins, and invariably takes part in stabilizing aromatic–aromatic interactions, but has also been linked to amyloid formation.[1] Crystallization of Phe itself is not trivial; Khawas[2] noted that that it "could not be obtained as good single crystals by the ordinary methods of crystallization", an experience shared by numerous other researchers. Typically no diffraction is observed beyond 1.0 Å resolution, with poorly resolved low-angle reflections that are hard to index.

A single-crystal structure of Phe, as the D-enantiomer, first appeared in 1990 as part of a paper on oriented crystallization at the air–solution interface by Weissbuch et al.[3] [Cambridge Structural Database (CSD)[4] refcode SIMPEJ]. Refinement of this polymorph (form I, **F-I**) in space group $C2$ with $Z' = 2$, converged at an R-factor of 0.147, and the authors reported "too short a H···H distance of 1.8 Å between the hydrogen atoms of neighboring phenyl rings" and that "carbon atoms of the two independent phenyl rings show unusually large displacement parameters". They interpreted the results "in terms of orientational disorder of the phenyl rings". Later on, King et al.[5] carried out solid-state DFT optimizations and found that the phenyl rings rotated from their coplanar $C2$ geometry to energetically more favorable staggered orientations. The symmetry of the resulting unit cell corresponded to the space group $P2$ with four molecules in the asymmetric unit. Recently Williams et al.[6] used powder X-ray diffraction to elucidate the structure of a second polymorph, form II (**F-II**, QQQAUJ04), "stable only under rigorously dry conditions", which can be reversibly converted to the hemihydrate and monohydrate upon increased relative humidity. Mossou et al.[7] furthermore reported a new monoclinic structure ($P2_1$, $Z' = 4$, QQQAUJ03), here called form III (**F-III**). Single crystals were grown from a neutral aqueous solution containing 10 mg mL^{-1} L-Phe, 15% each of polyethylene glycol and propan-2-ol, and 0.05 M NaCl. Other claims for new polymorphs[8,9] have not been substantiated.

During the course of carrying out a series of co-crystallization experiments, we discovered that the quality of the usual, rhombus-shaped zwitterionic **F-I** crystals could be significantly improved, without a change of crystal habit, by the presence of either formic or acetic acid in the aqueous solution (see the Supporting Information). With excellent crystals at hand, we have taken the opportunity to reinvestigate the structure based on data collection from a specimen that diffracted to 0.69 Å.

Our unit cell parameters were essentially the same as those reported in Ref. [3] (Table 1), but the lattice was clearly primitive, and structure solution in space group $P2_1$ yielded four molecules A–D in the asymmetric unit (Figure 1a). Refinement proceeded without restraints to an R-factor of 0.0417. No positional disorder was indicated. To confirm that no phase transition had taken place upon cooling, a second data set was recorded for the same crystal at 293 K. Apart

Table 1: Crystallographic unit cells of L-Phe.

Form	I[a]	I[b]	II[c]	III[d]	IV[b]
space group	$C2$	$P2_1$	$P2_1$	$P2_1$	$C2$
a [Å]	8.804	8.7829/8.7955	12.063	6.0010	9.6806
b [Å]	6.041	5.9985/6.0363	5.412	30.8020	5.2362
c [Å]	31.509	31.0175/31.5233	13.676	8.7980	15.8474
β [°]	96.60	96.9220/96.6441	99.5976	90.120	96.291
V [Å^3]	1667.6	1622.12/1662.40	880.3	1626.24	798.46
Z'	2	4	2	4	1
T [K]	295[e]	105/293	294	100	100

[a] Previous interpretation.[3] [b] This work. [c] Powder data, Williams et al.[6] [d] Mossou et al.[7] [e] Room temperature.

[*] F. S. Ihlefeldt, F. B. Pettersen, M. Zawadzka, Prof. C. H. Görbitz
Department of Chemistry, University of Oslo
P.O. Box 1033 Blindern, N-0315 Oslo (Norway)
E-mail: c.h.gorbitz@kjemi.uio.no
A. von Bonin
Department of Physics, University of Oslo
P.O. Box 1048 Blindern, N-0316 Oslo (Norway)

[**] This work was supported by the Department of Chemistry,University of Oslo

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201406886.

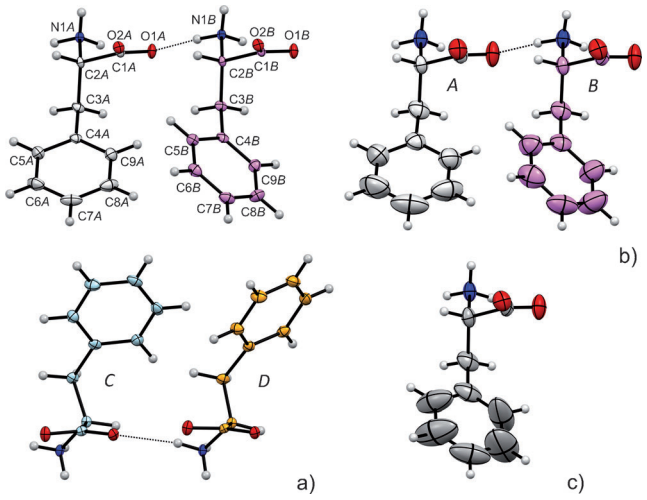

Figure 1. a) The asymmetric unit of **F-I** at 105 K with four independent molecules *A* (gray), *B* (violet), *C* (light blue), and *D* (orange) (50% probability displacement ellipsoids). Atomic numbering is indicated for *A* and *B*. b) Molecules *A* and *B* at 293 K. c) Average of molecule *A* and *B* at 293 K resulting from a *C*2 space group assignment.

from an increase in thermal vibrations (Figure 1 b), there are no significant changes.

The crystal structure in Figure 2 a is divided into hydrophobic and hydrophilic layers. Pseudo-*C*-centering cause reflections with $h+k=2n+1$ to be systematically weak; their average intensity is about 10% of that of the reflections with $h+k=2n$ at 293 K. Since they used a classic diffractometer and a Polaroid film for indexing, it is highly conceivable that the authors of Ref. [3] in 1990 overlooked the weaker reflections and thus ended up with a *C*-centered rather than a primitive unit cell. This removes the difference between *A* and *B* and vice versa for *C* and *D*. The averaged room-temperature *C*2 structure, refined here to an *R*-factor of 0.0708, has large displacement ellipsoids (Figure 1 c) with orientations that give an unphysical interpretation of the thermal vibrations in the molecule.

Hydrophilic layers in Figure 2 are composed of two sheets. An individual L1 sheet, one of five different types for hydrophobic amino acids,[10] is illustrated in Figure 3 a. In many racemates L1 sheets are paired with mirror-image D1 sheets to form L1–D1 layers. The L1–L1 layer of **F-I** and **F-III** is much less common with only two other observations: 4-fluorophenylalanine (**4F-F**, EXAXEG) with $Z'=2$[11] (Figure 2 d) and the *C*2 form of L-Leu (LEUCIN03) with $Z'=1$.[12]

Most amino acids with hydrophobic side chains crystallize with $Z'=2$ (L2–L2 hydrogen bonding), a preference that has been explained as a way to achieve a better hydrogen-bonding arrangement where the two molecules participate in slightly different interactions.[10] Molecules *A*, *B*, *C*, and *D* of **F-I** are indistinguishable from this point of view, meaning that the high Z'-value must instead be dictated by the need to get a proper stacking of the side chains.

While the isobutyl group of L-Leu is fully compatible with the *C*2 space group, an aromatic side chain inevitable leads to short contacts. The authors of Ref. [3] pointed out that these

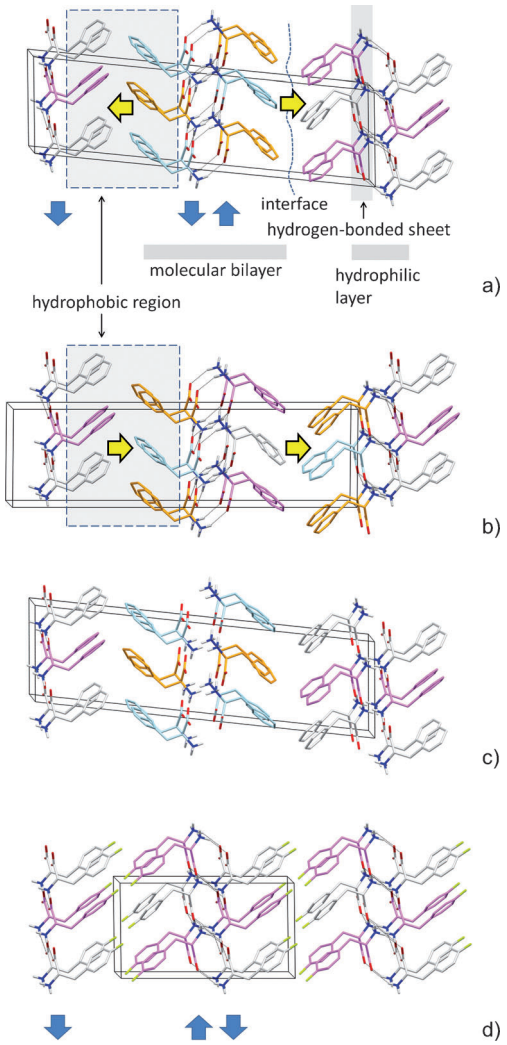

Figure 2. a) Unit cell and crystal packing of **F-I** viewed approximately along the *b*-axis with various terms used to describe the structures. Blue arrows give the directions of the N1–H2···O1 hydrogen bonds within each hydrogen-bonded sheet, yellow arrows the directions of each interface as defined in the text. b) The crystal structure of **F-III**.[7] c) Structural model in space group *P*2.[5] d) Crystal packing of **4F-F**.[11] Color coding in (b–d) has been adapted from (a) to indicate molecular conformation. Gray and orange molecules have the same conformation, as have violet and light blue ones.

could be made acceptable "by rotating the phenyl rings by 15° about the C(4)–C(7) (or C(42)–C(72)) axis. The rotation may be clockwise or counterclockwise, but the direction of rotation will be the same for a row of molecules related by translation along the *b* axis". We find rotations about C2-C3-C4-C5 of $\pm 18°$ (*A* and *D* $\approx 96°$, *B* and *C* $\approx 60°$, see the Supporting Information) to give ordered, independent stacks of *A*, *B*, *C*, and *D* molecules. Accordingly, the Z'-value of 2 for **4F-F** and the initial increase from 1 to 2 for **F-I** and **F-III** comes from the relief of steric repulsion.

The further increase to $Z'=4$ is the result of the way neighboring molecular bilayers are arranged along the *c*-axis. The pattern for **4F-F** in Figure 2 d is typical in that not only the directions of the two hydrogen-bonded sheets within a hydrophilic layer are antiparallel, but also sheets on

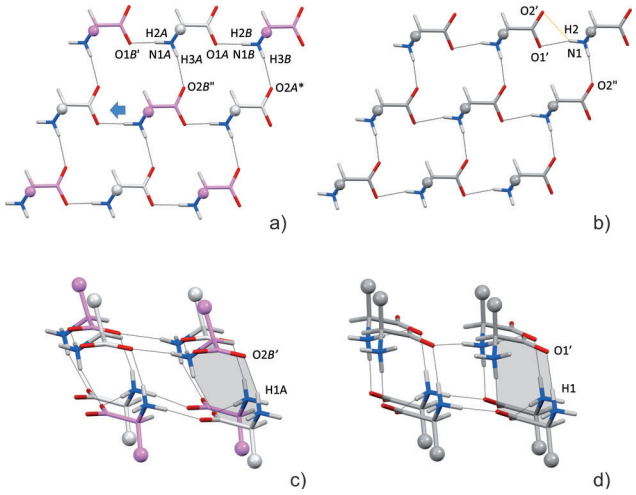

Figure 3. Hydrogen-bonded sheets in a) **F-I**, generated by *A* and *B* molecules, and b) in **F-IV**. Side chains are shown as small spheres. Symmetry-generated atoms are O1B′(−1+*x*, 1+*y*, *z*), O2B"(−1+*x*, *y*, *z*) and O2A*(*x*, −1+*y*, *z*) in (a) and O1′($^1/_2$+*x*, $^1/_2$+*y*, *z*) and O2"(*x*, 1+*y*, *z*) in (b). The single, long contact to O2′ shown in orange in (b) is 2.55 Å, corresponding distances in (a) are 2.74–2.80 Å. c) Formation of a hydrogen-bonded layer from two adjacent sheets in **F-I**. O2B′ is at (1−*x*, $^1/_2$+*y*, −*z*). d) Equivalent for **F-IV**, O1′ is at (1−*x*, *y*, 1−*z*). Dimers in both polymorphs have been highlighted by gray shading.

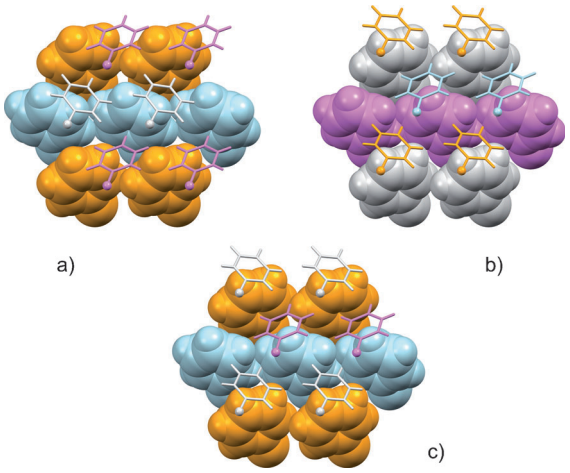

Figure 4. a) Detail of the "knobs-into-holes" interactions between phenyl rings inside the box highlighted in Figure 2a with *C* and *D* shown in a space-filling representation. b) Same interface as in (a), but with *A* and *B* in a space-filling representation. c) Interaction at the interface of the hydrophobic region of **F-III**[7] in Figure 2b. Color coding as in Figure 2.

opposite sides of a hydrophobic region. **F-I** and **F-III** in contrast have parallel arrangements; this was seen previously only for α-Gly,[13] L-2-aminobutyric acid,[14] and a L-Val:D-Met complex.[15] This is the result of a 180° rotation of the center molecular bilayer around the horizontal axis compared to **4F-F**, which has produced, instead of the close-to-parallel arrangement of aromatic rings of **4F-F**, a nice herringbone-like pattern with favorable edge-to-face interactions at the central interface of the hydrophobic region, such as C7B–H71B···C7C(*x*+1,*y*,*z*) with H···C = 2.28 Å (at 105 K) for **F-I**. These can clearly not be present for the 4-fluorine analogue.

Judged by the unit cell parameters in Table 1 and the packing diagrams in Figure 2a and b, the **F-I** and **F-III** forms are closely related, so an analysis of the actual differences between them is needed. A key element is that, in distinction to, for example, **4F-F** (Figure 2d), no symmetry operator relates the two halves of a hydrophobic region. This means that there are two distinct ways of fitting phenyl H-atoms into "holes" on the opposing surface (Figure 4a and b).

Consequently, we can define a direction for each interface, here chosen arbitrarily to go into the plane in Figure 4b (or out of the plane in Figure 4a). Using the symbol ‖ to denote a hydrogen-bonded layer, this gives ←‖→‖←‖→ as a schematic representation for **F-I** in Figure 2a.

Compared to **F-I** the central molecular bilayer of **F-III** in Figure 2b has been shifted half an axis length along the shortest (vertical) axis. This would suggest that the interfaces of **F-I** and **F-III** are different, but a comparison of Figure 4b and c reveals that they are in fact identical, only the directions have changed. The resulting schematic description for **F-III** is then →‖→‖→‖→. As detailed in the Supporting Infor-maation, this sequence produces crystallographic screw axes parallel to the longest unit cell axis (*b*). The angle between the 6.0 and 8.8 Å axes, confined by symmetry to be 90.0° for **F-I** (γ angle), relaxes to 90.120(4)° for **F-III** (Table 1). The relationship between the two forms, with identical hydrogen-bonding patterns as well as aromatic–aromatic interactions, is unprecedented among amino acid structures and provides a remarkable example of intricate polymorphism.

The *P*2 structure model presented by King et al.[5] as "the true crystal structure of L-phenylalanine" is shown in Figure 2c. It is close to the **F-I** structure in Figure 2a, but is not supported by experimental data.

In addition to the rhombus-shaped specimens of **F-I**, most test tubes also contained variable amounts of extremely thin fibers, which were also described by several other authors.[5,8] A single-crystal structure determination to 1.0 Å resolution (see the Supporting Information) reproduced the monohydrate (**F-w**, GOFWOP) described previously by Williams et al.[6] A single test tube furthermore yielded visibly different crystals in the shape of rather large elongated wedges. A piece was cut from a larger specimen and used for data collection of what proved to be the new form IV, **F-IV**. The unit cell was indisputably *C*2, and structure solution yielded a single L-Phe molecule as the asymmetric unit. Subsequent refinement revealed that the side chain was severely disordered, and, after a series of alternatives were tested, a model with three disorder parts was adopted. (Restraints were employed during refinement, details are in the SHELXL.res file which is part of the submitted .cif file; see the Supporting Information.) The molecular structure is shown in Figure 5a. All disorder parts have the side chain in the N-C2-C3-C4 *trans* orientation, as for **F-I**, and together describe a wagging motion of the side chain. The aromatic rings are stacked in a slightly different way than in form I, giving an even more compact structure (d = 1.374 vs. 1.353 g cm^{-1}). It is thus possible that **F-IV** is in fact the thermodynamically most favorable form. Considered alone, the disorder part 1 makes close, but not prohibitively close contacts with its neighbors;

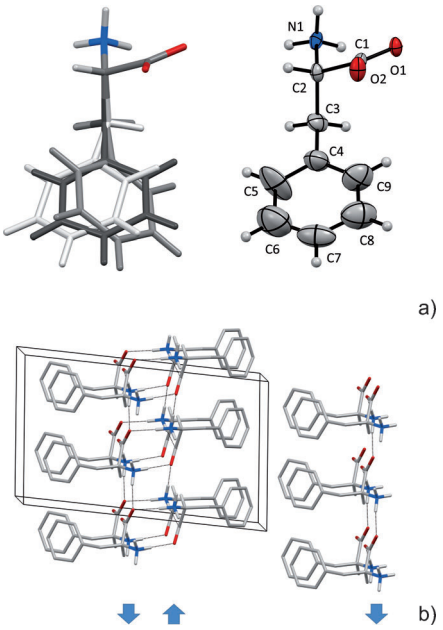

Figure 5. a) Molecular structure of **F-IV** at 100 K. The stick drawing shows the side-chain disorder: part 1 (dark gray) has occupancy 0.471(15), part 2 (gray) 0.260(10), and part 3 (light gray) 0.267(12). The ellipsoid drawing shows part 1 with atom numbers indicated (parts 2 and 3 were refined isotropically). b) Crystal packing viewed approximately along the *b*-axis. Only the major side chain orientation is shown, blue arrows used as in Figure 2.

C8–H8···H5–C5($1/2+x,1/2+Y,z$) has H···H = 2.14 Å and H8···C5/C8···H5 = 2.50 Å. Nevertheless, a complex stacking disorder is observed, and the large displacement ellipsoids in Figure 5a should be seen to reflect this disorder rather than normal thermal vibrations.

A hydrogen-bonded sheet of type Lx[10] for **F-IV** is shown in Figure 3b. The subtle differences compared to the L1 sheets of **F-I** and **F-III** are merely the fingerprint of the distinct ways sheets generate layers (Figure 3c and d). H1 always serves as the connecting atom, but while it is accepted by O2 in **F-I**, O1 is the acceptor in **F-IV**, a shift brought about by a 125° rotation of one sheet relative to the other. Transitions between the two polymorphs are thus not expected. **F-IV** (Figure 5b) and isostructural L-norleucine[16] and (2*S*,4*S*)-5-fluoroleucine[17] have the normal antiparallel arrangement of sheets on opposite sides of a hydrophobic region. L-2-Aminobutyric acid[14] and the hexagonal form of L-cystine[18] are the only other amino acids with Lx–Lx hydrogen bonding.

In 1970[2] Khawas reported a $P2_1$ unit cell with $b = 6.59$ Å and volume 861.2 Å^3. In his 1984 paper[19] no reference is made to the previous work, but two new orthorhombic polymorphs α and β are introduced with unit cell volumes of 901.9 ($c = 6.48$ Å, $Z = 4$) and 1803.9 Å^3 ($Z = 8$), respectively. In view of the unit cell volumes in Table 1, it appears unlikely that these could correspond to yet other solvent-free L-Phe modifications. Rather, Khawas may have obtained unknown solvates or made erroneous unit cell assignments. Notably, it is very uncommon to find amino acid crystal structures with the shortest unit cell axis between 6.2 and 7.0 Å, as this is normally incompatible with formation of direct hydrogen bonds between the amino and carboxylate groups.[20]

After completing the investigation of L-Phe, we continued to test a number of rhombus-shaped crystals grown in a similar manner, but from solutions of DL-Phe rather than L-Phe. Diffraction was variable, but overall reasonable, and several limited and one complete data set were collected. Subsequent displays of the reciprocal lattices indicated that the specimens mounted were not proper single crystals, as views along the *c**-axis invariably showed multiple stacks of reflections. Extensive twinning analysis with the program CELL_NOW[21] eventually identified the unit cell already known for **F-I**. We thus conclude that DL-Phe does not form a true racemate, but rather a racemic twin or conglomerate with additional, as yet unidentified types of non-merohedral twinning. Further efforts to resolve this matter were considered futile.

In summary, with the high-resolution structures of form I and form IV of anhydrous L-phenylalanine presented here,[22] the number of well-characterized polymorphs has in one year seen a dramatic increase from zero to four.[6,7] Together with two additional hydrates,[6] they provide detailed information on the solid-state structures of this prototypical aromatic amino acid. For the common form I, distinguished from form III only by the mode of layer stacking, we find that the space group, contrary to previous results, is monoclinic $P2_1$ with four molecules in the asymmetric unit. The disordered high-density form IV with $Z' = 1$ has a different hydrogen-bonding pattern. It is probably not possible to grow proper racemic crystals for DL-Phe.

Received: July 4, 2014
Revised: August 26, 2014
Published online: October 21, 2014

Keywords: amino acids · hydrogen bonds · polymorphism · pseudosymmetry · racemic twins

[1] R. Azriel, E. Gazit, *J. Biol. Chem.* **2001**, *276*, 1919–1922.
[2] B. Khawas, *Acta Crystallogr. Sect. B* **1970**, *26*, 1919–1922.
[3] I. Weissbuch, F. Frolow, L. Addadi, M. Lahav, L. Leiserowitz, *J. Am. Chem. Soc.* **1990**, *112*, 7718–7724.
[4] F. Allen, *Acta Crystallogr. Sect. B* **2002**, *58*, 380–388.
[5] M. D. King, T. N. Blanton, T. M. Korter, *Phys. Chem. Chem. Phys.* **2012**, *14*, 1113–1116.
[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] Z. Liu, C. Li, *Biophys. Chem.* **2008**, *138*, 115–119.
[9] J. Yano, H. Füredi-Milhofer, E. Wachtel, N. Garti, *Langmuir* **2000**, *16*, 10005–10014.
[10] C. H. Görbitz, K. Vestli, R. Orlando, *Acta Crystallogr. Sect. B* **2009**, *65*, 393–400.
[11] Y. In, S. Kishima, K. Minoura, T. Nose, Y. Shimohigashi, T. Ishida, *Chem. Pharm. Bull.* **2003**, *51*, 1258–1263.
[12] M. Yamashita, S. Inomata, K. Ishikawa, T. Kashiwagi, H. Matsuo, S. Sawamura, M. Kato, *Acta Crystallogr. Sect. E* **2007**, *63*, o2762–o2764.

[13] S. Mondal, S. J. Prathapa, S. van Smaalen, *Acta Crystallogr. Sect. A* **2012**, *68*, 568 – 581.
[14] C. H. Görbitz, *Acta Crystallogr. Sect. B* **2010**, *66*, 253 – 259.
[15] B. Dalhus, C. H. Görbitz, *Acta Crystallogr. Sect. C* **1999**, *55*, 1547 – 1555.
[16] K. Torii, Y. Iitaka, *Acta Crystallogr. Sect. B* **1973**, *29*, 2799 – 2807.
[17] J.-D. Charrier, D. S. Hadfield, P. B. Hitchcock, D. W. Young, *Org. Biomol. Chem.* **2004**, *2*, 474 – 482.
[18] S. Dahaoui, V. Pichon-Pesme, J. A. K. Howard, C. Lecomte, *J. Phys. Chem. A* **1999**, *103*, 6240 – 6250.
[19] B. Khawas, *Indian J. Phys. A* **1984**, *59*, 219 – 226.
[20] C. H. Görbitz, *Crystallogr. Rev.* **2015**, DOI: 10.1080/0889311X.2014.964229.
[21] http://xray.tamu.edu/pdf/manuals/cell_now.pdf.
[22] CCDC 1012154 (**F-I**, 105 K), 1012155 (**F-I**, 293 K), 1012156 (**F-III**) and 1021197 (**F-w**) contain 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.

