112.2(4)°; C(4)-O(3)···O(4) angle 107.6(3)°; O(3)···O(4)-C(14)-C(8) dihedral angle −60.3(6)°; b) O(3)···O(4) distance 3.427(8) Å; O(3)···O(4)-C(14) angle 153.3(4)°; C(4)-O(3)···O(4) angle 123.3(4)°; O(3)···O(4)-C(14)-C(8) dihedral angle 25.5(1.3)°. Note that C(4)-O(3)-H is the hydroxy group of molecule A and O(4)−C(14) is in the carboxylic acid group of molecule B.

[21] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *J. Chem. Soc. Perkin Trans. 2* **1987**, S1.

[22] C. P. Brock, L. L. Duncan, *Chem. Mater.* **1994**, *6*, 1307.

[23] B. H. Meier, F. Graf, R. R. Ernst, *J. Chem. Phys.* **1982**, *76*, 767.

[24] K. Furic, *Chem. Phys. Lett*. **1984**, *108*, 518.

[25] B. H. Meier, R. Meyer, R. R. Ernst, A. Stöckli, A. Furrer, W. Hälg, I. Anderson, *Chem. Phys. Lett.* **1984**, *108*, 522.

[26] S. Nagaoka, T. Terao, F. Imashiro, N. Hirota, S. Hayashi, *Chem. Phys. Lett.* **1984**, *108*, 524.

[27] W. Scheubel, H. Zimmerman, U. Haeberlen, *J. Mag. Res.* **1988**, *80*, 401.

[28] C. C. Wilson, N. Shankland, A. J. Florence, *J. Chem. Soc. Faraday Trans.* **1996**, *92*, 5051.

[29] M. Neumann, D. F. Broughman, C. J. McGloin, M. R. Johnson, A. J. Horsewill, H. P. Trommsdorff, *J. Chem. Phys.* **1998**, *109*, 7300.

[30] M. A. Neumann, S. Craciun, A. Corval, M. R. Johnson, A. J. Horsewill, V. A. Benderskii, H. P. Trommsdorff, *Phys. Chem. Chem. Phys.* **1998**, *102*, 317.

[31] Form II of *p*-hydroxybenzoic acid represents a very rare example of a structure containing two crystallographically independent carboxylic acid dimer motifs in which one motif is ordered and the other is disordered. In support of this statement, we have carried out a search of the Cambridge Structural Database (CSD) to find structures containing *both* an unambiguously ordered carboxylic acid dimer (i.e. with one C−O bond in each carboxylic acid group in the range 1.21–1.25 Å representing C=O, and the other C−O bond in each carboxylic acid group in the range 1.31–1.35 Å representing C−OH) and an unambiguously disordered carboxylic acid dimer (i.e. with both C−O bonds in each carboxylic acid group in the range 1.26–1.30 Å). We note first that there is no overlap in the ranges used to distinguish the ordered and disordered carboxylic acid dimers, and second that for the structure reported here for form II of *p*-hydroxybenzoic acid, one carboxylic acid dimer is classed as ordered and the other is classed as disordered on the basis of the above distance criteria. In the CSD search, only one structure (octa-*cis*-2,*trans*-6-diene-1,8-dioic acid[32]) was found containing carboxylic acid dimers of both types (note that six structures were found containing individual carboxylic acid groups of both types, but it is only in the above structure that these carboxylic acid groups are present in ordered and disordered dimers). However, the estimated standard deviations in the C−O bond lengths in the structure of octa-*cis*-2,*trans*-6-diene-1,8-dioic acid (which was determined from Weissenberg photographic data) were in the range 0.02–0.04 Å, from which the reliability of the classification on the above distance criteria becomes significantly blurred. We note that these estimated standard deviations are substantially larger than those reported herein.

[32] E. Martuscelli, L. Mazzarella, *Acta Crystallogr. Sect. B* **1969**, *25*, 2022.

Structure Determination of an Oligopeptide Directly from Powder Diffraction Data**

Emilio Tedesco, Giles W. Turner, Kenneth D. M. Harris,* Roy L. Johnston, and Benson M. Kariuki

In recent decades, single-crystal X-ray diffraction techniques have underpinned many major breakthroughs in structural biology, although it is well recognized that an intrinsic limitation of these techniques is the requirement to prepare single crystals of sufficient size, quality, and stability to allow diffraction data of sufficiently high quality to be measured. In solid-state and materials sciences, similar limitations are encountered in the scope and applicability of single-crystal diffraction methods, but much progress has been made in recent years in the use of powder diffraction data as an alternative means of obtaining structural information. Herein, we apply a new methodology for structure determination from powder diffraction data to determine the crystal structure of the oligopeptide Phe-Gly-Gly-Phe (Figure 1), which demonstrates the opportunity for future extensions to more complex systems of interest in structural biology. In this regard, we note that knowledge of the structural properties and interactions in model oligopeptide systems can yield fundamental insights towards understanding structural properties of polypeptide sequences in proteins.[1–6]

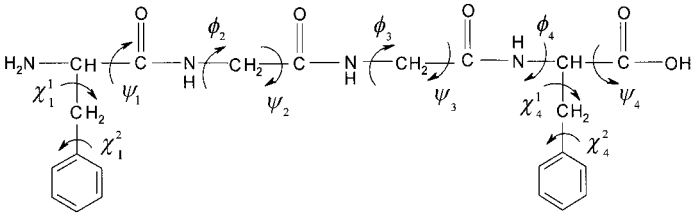

Figure 1. The molecular structure of Phe-Gly-Gly-Phe, showing the variable torsion angles in the genetic algorithm structure solution calculation.

In principle, powder diffraction data contains the same information as single-crystal diffraction data, but with the three-dimensional diffraction data compressed into one dimension. As a consequence, there is generally substantial overlap of peaks in the powder diffraction pattern, leading to severe difficulties in extracting the same quality of diffraction data (and hence structural information). Although the challenges in determining structures directly from powder diffraction data are considerable, new techniques and strat-

[*] Prof. K. D. M. Harris, Dr. E. Tedesco, G. W. Turner, Dr. R. L. Johnston, Dr. B. M. Kariuki
School of Chemistry, University of Birmingham
Edgbaston, Birmingham B15 2TT (UK)
Fax: (+44) 121-414-7473
E-mail: K.D.M.Harris@bham.ac.uk

[**] This work was supported by the EPSRC, the University of Birmingham, Wyeth-Ayerst plc, and Ciba Specialty Chemicals. We are grateful to Professor P. Balaram (Indian Institute of Science, Bangalore) for valuable discussions.

Supporting information for this article is available on the WWW under http://www.wiley-vch.de/home/angewandte/ or from the author.

  1433-7851/00/3924-4488 $ 17.50+.50/0

egies developed in recent years (reviewed in references[7–10]) have allowed substantial progress to be made, particularly with regard to the more challenging structure solution stage of the structure determination process. It is relevant here to recall that *structure solution* involves deriving an initial (approximate) structural model directly from the experimental diffraction data, starting from no prior knowledge of the arrangement of molecules in the unit cell. If this structural model is a sufficiently good representation of the true structure, *structure refinement* can then be carried out to obtain a good quality structure. Techniques for structure solution from powder diffraction data can be subdivided into "traditional" and "direct-space" approaches.

In the *traditional* approach, the aim is to extract the intensities $I(hkl)$ of individual reflections directly from the powder diffraction pattern, and to use these $I(hkl)$ data in the types of structure solution calculation that are used for single-crystal diffraction data. Although there have been many successful applications of this strategy,[7–9] organic molecular crystals represent a particularly challenging case. The main limitation of the traditional approach concerns the reliability of the intensities $I(hkl)$ that can be extracted from the experimental powder diffraction pattern, particularly in cases with substantial peak overlap.

In *direct-space* approaches, a large number of trial structures are generated/sampled in direct space by means of an appropriate search algorithm. The "quality" of each structure is assessed by comparing the powder diffraction pattern calculated for the structure and the experimental powder diffraction pattern, for example using the weighted profile R factor R_{wp}. As R_{wp} considers the whole digitized intensity profile, peak overlap is implicitly taken into account (and the need to extract integrated intensities $I(hkl)$ from the powder diffraction pattern is circumvented). In effect, the direct-space strategy involves searching the $R_{wp}(\Gamma)$ hypersurface to find the best structure solution (lowest R_{wp}), where Γ represents the set of variables that define the position, orientation, and intramolecular geometry of each molecule in the asymmetric unit (see below). Previous research has demonstrated the success of Monte Carlo[11, 12], simulated annealing,[13–17] and genetic algorithm (GA)[18–22] methods for searching R factor hypersurfaces in direct-space structure solution from powder diffraction data.

In general, the limitations of direct-space structure solution concern the number and nature of the variables in the set Γ (defining the dimensionality and nature of the $R(\Gamma)$ hypersurface) and clearly the number of variables is larger for flexible molecules defined by a significant number of variable torsion angles. In the case of organic molecular crystals, another factor is the absence of dominant X-ray scatterers in the structure. The oligopeptide Phe-Gly-Gly-Phe studied herein represents a case of substantial molecular flexibility with no dominant X-ray scattering atoms.

In our GA approach for structure solution[18, 19], a population of trial crystal structures is allowed to evolve subject to the types of rules and operations that govern evolutionary systems. Initially, the population comprises a set of N_P random structures. For the case with one molecule in the asymmetric unit, a given structure is specified by the coordinates $\{x, y, z\}$ of the center of mass or a predefined pivot atom, the orientation is specified by rotation angles $\{\theta, \phi, \psi\}$ around a set of orthogonal axes, and the intramolecular geometry is specified by a set of variable torsion angles $\{\tau_1, \tau_2, \ldots, \tau_n\}$. These variables represent the "genetic code" that uniquely characterizes each member of the population. The quality ("fitness") of each structure depends on its value of R_{wp} (lower R_{wp} represents better agreement between experimental and calculated powder diffraction patterns and therefore higher fitness). The population is allowed to evolve through several generations by using the operations of mating, mutation, and natural selection. In the *mating* procedure, a given number (N_M) of pairs of structures ("parents") are selected from the population, with the probability of selecting a given structure proportional to its fitness. New structures ("offspring") are then generated by combining genetic information from the two parents. In the *mutation* procedure, a given number (N_X) of structures are selected at random from the population and random changes are made to parts of their genetic code to create mutant structures (the original structures from which the mutants are derived are still retained within the population). In the *natural selection* procedure, only the best structures are allowed to pass from one generation to the next. After a sufficient number of generations, the fittest member of the population (the structure with lowest R_{wp}) should be close to the correct crystal structure. In our most recent implementation of the GA method,[23] each new structure generated during the GA calculation is subjected to local minimization of R_{wp} with respect to the structural variables in the set Γ, and only these minimized structures are used subsequently in the GA calculation. Introduction of local minimization in this way has been found to improve significantly the efficiency of finding the correct structure solution.

The powder X-ray diffraction pattern of Phe-Gly-Gly-Phe was recorded by using a conventional laboratory powder X-ray diffractometer.[24] Prior to the structure solution calculation, the unit cell and space group were determined directly from the powder diffraction pattern.[25] In the GA structure solution calculation, the molecule was defined by using standard bond lengths and angles. All peptide groups[27] R_1-CO-NH-R_2 were maintained as planar units with the O-C-N-H dihedral angle fixed at 180°, and each benzene ring was constrained to be planar. All other torsion angles were varied freely (Figure 1). Thus, each structure considered in the GA calculation was defined by 17 variables $\{x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \ldots, \tau_{11}\}$ (note that in Figure 1 and reference [29], a different notation for the torsion angles is used).

The GA calculation comprised 45 generations of a population of $N_P = 50$ structures. In each generation, 50 offspring ($N_M = 25$ pairs of parents) and $N_X = 10$ mutants were generated. The fitness of each structure is defined as a function of R_{wp}, and in the present case the linear fitness function $F(R_{wp}) = 1 - ((R_{wp} - R_{min})/(R_{max} - R_{min}))$ was used, where R_{min} and R_{max} represent the minimum and maximum values of R_{wp} in the current population (R_{min} and R_{max} are updated as the population evolves under the GA). For the mating procedure between two selected parent structures, the 17 variables for a given parent were subdivided into three groups $\{x, y, z\}$, $\{\theta, \phi,$

  1433-7851/00/3924-4489 $ 17.50+.50/0

ψ}, and {τ_1, τ_2, ..., τ_{11}}. The three groups from each parent were distributed between the two offspring with no restriction on which combination of groups may come from each parent. In carrying out the mutation procedure on a selected structure, eight variables were selected at random, and a new random value was assigned to each of the selected variables (i.e. static mutation[18, 19]). The evolution of R_{wp} during the GA calculation is shown in Figure 2, and demonstrates clearly that the quality of the population improves as the population evolves.

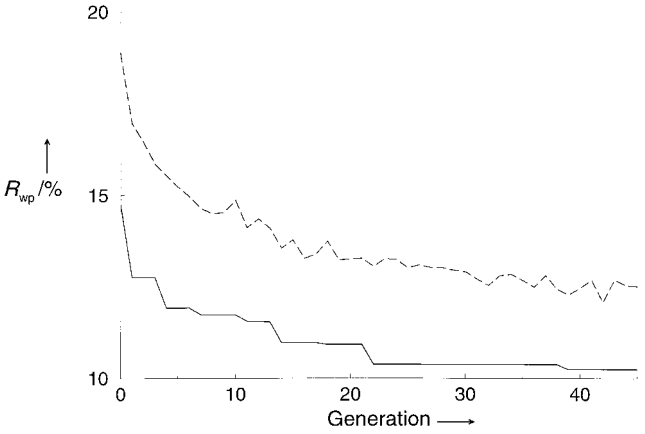

Figure 2. Evolutionary progress plot showing the evolution of R_{wp} for the best structure in the population (solid line) and the average R_{wp} for the structures in the population (dashed line) as a function of generation number in the GA structure solution calculation of Phe-Gly-Gly-Phe. Note that the values of R_{wp} refer to structures following local minimization.

The best structure solution (with lowest R_{wp} in the final generation) was taken as the starting structural model for Rietveld refinement using the GSAS program package.[28] All atoms (with hydrogen atoms introduced into the structural model in calculated positions) were included in the refinement, with standard geometric restraints applied to the bond lengths and bond angles. Two isotropic displacement parameters were refined—a common parameter for the non-hydrogen atoms of the CH_2Ph groups, and a common parameter for all non-hydrogen atoms of the oligopeptide backbone. The isotropic displacement parameter for the hydrogen atoms was fixed at a standard value. In the final stages, a preferred orientation parameter was refined. The final Rietveld refinement (Figure 3) gave $R_{wp} = 0.053$ and $R_p = 0.034$ (data range, $5^\circ \leq 2\theta \leq 60^\circ$; number of reflections: 403; number of profile points: 2857; number of refined variables: 105). The fact that our successful structure determination of Phe-Gly-Gly-Phe was based on the use of experimental data collected on a conventional laboratory powder X-ray diffractometer emphasizes that it is not essential to use synchrotron X-ray powder diffraction data in this field.

The crystal structure of Phe-Gly-Gly-Phe (Figure 4a; see also Table 1 in Supporting Information) comprises ribbons of Phe-Gly-Gly-Phe molecules[29] extending along the c axis. Adjacent molecules in these ribbons interact through three N–H···O hydrogen bonds (N···O distances 2.68–3.37 Å) involving the three peptide groups within the oligopeptide backbone (Figure 4b). Importantly, the structure and interactions between molecules in these ribbons are analogous to an antiparallel β-sheet arrangement. In addition to hydrogen bonding, edge-to-face interactions between the phenyl rings of adjacent molecules in these ribbons may also be identified. As shown in Figure 4a, the end-groups of the oligopeptide chains are located around the 4_1 screw axis, and also engage in intermolecular N–H···O hydrogen bonding (N···O distances 3.05–3.22 Å). Around the 4_1 screw axis in a given unit cell there are four NH_3^+ end-groups (related to each other by the 4_1 screw) and four CO_2^- end-groups (related by the 4_1 screw). The hydrogen bonding pattern comprises two intertwined helical chains (Figure 4c), with N–H···O hydrogen bonding interactions existing within each chain but not between the chains. Each individual hydrogen-bonded chain comprises an alternating arrangement of NH_3^+ end-groups and CO_2^- end-groups, and has 2_1 screw symmetry. The overall 4_1 screw symmetry arises when the two individual chains are intertwined.

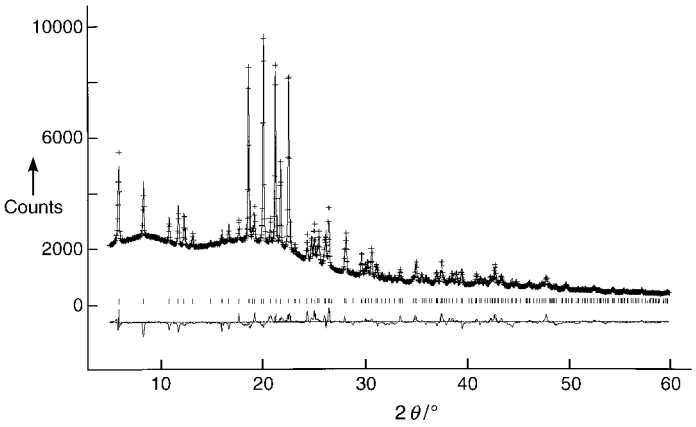

Figure 3. Experimental (+ marks), calculated (solid line), and difference (lower line) powder X-ray diffraction profiles for the Rietveld refinement of Phe-Gly-Gly-Phe. Reflection positions are marked. The calculated powder diffraction profile is for the final refined crystal structure. The refined lattice parameters are: $a = 15.084(1)$, $c = 9.7369(8)$ Å.

In conclusion, the successful determination of the crystal structure of the oligopeptide Phe-Gly-Gly-Phe demonstrates the scope and potential of current techniques for the complete determination of molecular crystal structures from powder diffraction data in the case of flexible molecules containing no dominant X-ray scattering atoms, and extends the application of techniques for powder structure solution into the arena of structural biology. The opportunities demonstrated here provide considerable promise for future successful applications of these techniques to solve crystal structures of increasing complexity in this field.

Received: July 27, 2000 [Z 15539]

[1] W. F. DeGrado, *Adv. Protein Chem.* **1988**, *39*, 51.
[2] L. Regan, *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 257.
[3] R. Kaul, P. Balaram, *Bioorg. Med. Chem.* **1999**, *7*, 105.
[4] C. A. Orenga, D. T. Jones, J. M. Thornton, *Nature* **1994**, *372*, 631.
[5] I. L. Karle, J. L. Flippen-Anderson, K. Uma, P. Balaram, *Int. J. Pept. Protein Res.* **1993**, *42*, 401.
[6] I. L. Karle, S. K. Awasthi, P. Balaram, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8189.

  1433-7851/00/3924-4490 $ 17.50+.50/0

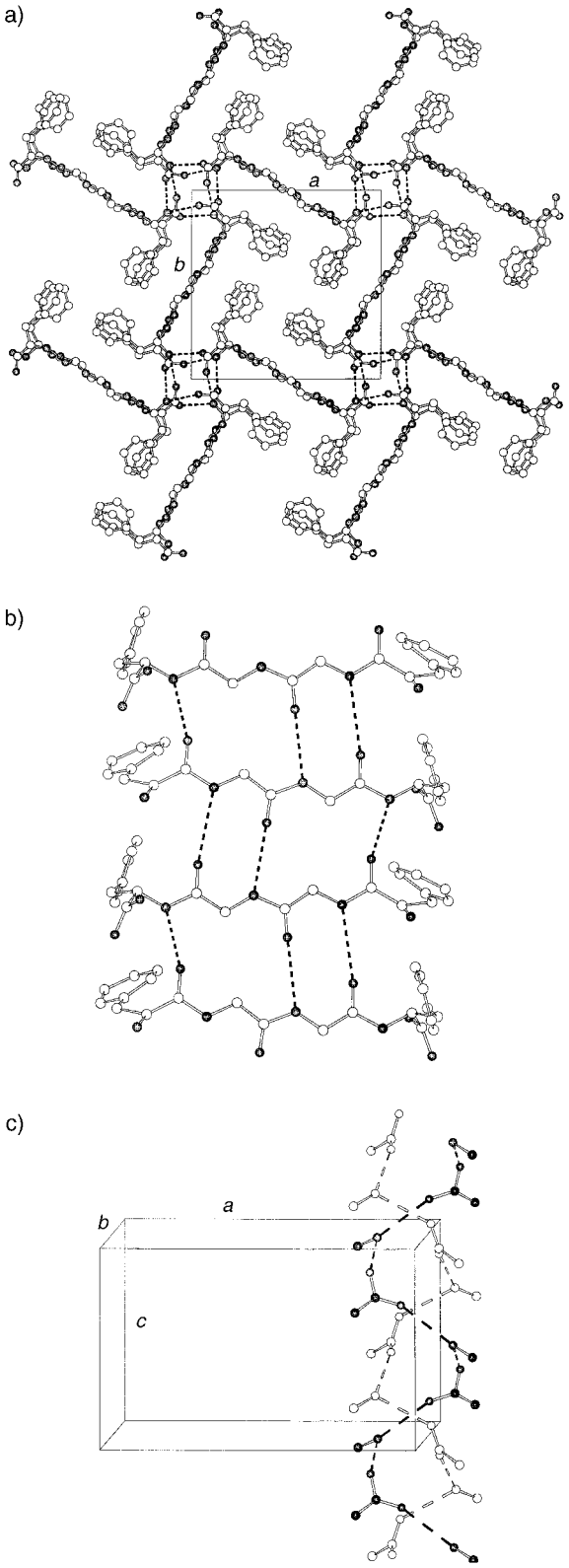

Figure 4. a) Final refined crystal structure of Phe-Gly-Gly-Phe viewed along the *c* axis. b) Interactions between adjacent molecules in the crystal structure of Phe-Gly-Gly-Phe (viewed perpendicular to the *c* axis) illustrating the formation of an antiparallel β-sheet arrangement. For clarity, hydrogen atoms are not shown. c) Illustration of the two intertwined helical hydrogen-bonded chains (open circles and filled circles, respectively) running along the *c* axis (vertical). Fragments comprising four atoms represent $C{-}CO_2^-$ end-groups; fragments comprising two atoms represent the C–N component of $C{-}NH_3^+$ end-groups.

[7] A. K. Cheetham, A. P. Wilkinson, *Angew. Chem.* **1992**, *104*, 1594; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1557.
[8] K. D. M. Harris, M. Tremayne, *Chem. Mater.* **1996**, *8*, 2554.
[9] D. M. Poojary, A. Clearfield, *Acc. Chem. Res.* **1997**, *30*, 414.
[10] K. D. M. Harris, M. Tremayne, B. M. Kariuki, *Angew. Chem.* **2001**, *113*, in press; *Angew. Chem. Int. Ed.* **2001**, *40*, in press.
[11] K. D. M. Harris, M. Tremayne, P. Lightfoot, P. G. Bruce, *J. Am. Chem. Soc.* **1994**, *116*, 3543.
[12] M. Tremayne, B. M. Kariuki, K. D. M. Harris, *Angew. Chem.* **1997**, *109*, 788; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 770.
[13] J. M. Newsam, M. W. Deem, C. M. Freeman, *Accuracy in Powder Diffraction II*, *NIST Special Publ. No. 846*, **1992**, p. 80.
[14] D. Ramprasad, G. P. Pez, B. H. Toby, T. J. Markley, R. M. Pearlstein, *J. Am. Chem. Soc.* **1995**, *117*, 10694.
[15] Y. G. Andreev, P. Lightfoot, P. G. Bruce, *Chem. Commun.* **1996**, 2169.
[16] W. I. F. David, K. Shankland, N. Shankland, *Chem. Commun.* **1998**, 931.
[17] G. E. Engel, S. Wilke, O. König, K. D. M. Harris, F. J. J. Leusen, *J. Appl. Crystallogr.* **1999**, *32*, 1169.
[18] B. M. Kariuki, H. Serrano-González, R. L. Johnston, K. D. M. Harris, *Chem. Phys. Lett.* **1997**, *280*, 189.
[19] K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Acta Crystallogr. Sect. A* **1998**, *54*, 632.
[20] K. Shankland, W. I. F. David, T. Csoka, *Z. Kristallogr.* **1997**, *212*, 550.
[21] B. M. Kariuki, P. Calcagno, K. D. M. Harris, D. Philp, R. L. Johnston, *Angew. Chem.* **1999**, *111*, 860; *Angew. Chem. Int. Ed.* **1999**, *38*, 831.
[22] 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.
[23] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Chem. Phys. Lett.* **2000**, *321*, 183.
[24] A sample of Phe-Gly-Gly-Phe was obtained from Sigma-Aldrich and used directly as received. The powder X-ray diffraction pattern was recorded at 22 °C in transmission mode on a Siemens D5000 diffractometer, using Ge-monochromated $Cu_{K\alpha 1}$ radiation and a linear position-sensitive detector covering 8° in 2θ. The total 2θ range was 5° to 60°, measured in steps of 0.02° over 12 h.
[25] The powder X-ray diffraction pattern was indexed (using the first 20 observable peaks) by the program TREOR,[26] giving the unit cell: $a = b = 15.06$, $c = 9.72$ Å, $\alpha = \beta = \gamma = 90°$. Systematic absences are consistent with the tetragonal space group $P4_1$, and density considerations suggest that there is one molecule of Phe-Gly-Gly-Phe in the asymmetric unit. Following unit cell determination and prior to structure solution, parameters describing the experimental lineshape were determined by fitting the powder diffraction profile with arbitrary intensities using the program PowderFit.[17]
[26] P.-E. Werner, L. Eriksson, M. Westdahl, *J. Appl. Crystallogr.* **1985**, *18*, 367.
[27] G. Fischer, *Chem. Soc. Rev.* **2000**, *29*, 119.
[28] A. C. Larson, R. B. Von Dreele, *Los Alamos Lab. Report No. LA-UR-86-748*, **1987**.
[29] The molecular conformation in the crystal structure is specified by the following values of the torsion angles (in standard notation[30]) defined in Figure 1: $\psi_1 = -103(1)°$; $\chi_1^1 = -175(1)°$; $\chi_1^2 = -33(1)°$; $\phi_2 = +164(1)°$; $\psi_2 = +35(1)°$; $\phi_3 = +176(1)°$; $\psi_3 = +35(1)°$; $\phi_4 = -107(1)°$; $\psi_4 = -70(1)°$; $\chi_4^1 = -50(1)°$; $\chi_4^2 = -67(1)°$.
[30] IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry* **1970**, *9*, 3471.

  1433-7851/00/3924-4491 $ 17.50+.50/0