CHEMCOMM Communication
www.rsc.org/chemcomm

Ab initio structure determination of a peptide β-turn from powder X-ray diffraction data†

Emilio Tedesco,[a] Kenneth D. M. Harris,*[a] Roy L. Johnston,[a] Giles W. Turner,[a] K. Muruga Poopathi Raja[b] and Padmanabhan Balaram[b]

[a] *School of Chemistry, University of Birmingham, Edgbaston, Birmingham, UK B15 2TT. E-mail: K.D.M.Harris@bham.ac.uk; Fax: +44-121-414-7473; Tel: +44-121-414-7474*
[b] *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India*

Received (in Cambridge, UK) 30th April 2001, Accepted 26th June 2001
First published as an Advance Article on the web 20th July 2001

***Ab initio* crystal structure determination of the peptide Piv-Pro-Gly-NHMe directly from powder X-ray diffraction data, using the genetic algorithm technique for structure solution, has allowed the complete structural characterization of the Type II β-turn conformation and the intermolecular interactions in this structure, and highlights the opportunities that now exist for structure determination of peptide systems when single crystals appropriate for single crystal X-ray diffraction experiments cannot be prepared.**

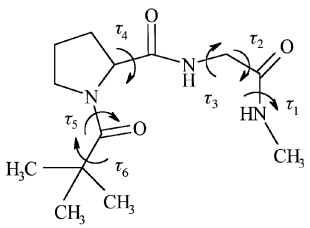
Knowledge of the conformational properties and interactions in model peptide systems can yield important insights concerning the structural properties of polypeptide sequences in proteins. An example concerns β-turns, which are structural elements that permit polypeptide chain reversals in proteins.[1] Tight turns in proteins and peptides, involving two residues as folding nuclei, have been very widely investigated,[2–5] and the understanding of β-turn stereochemistry that now exists has been facilitated by extensive high resolution structure determination of peptides[4] and proteins[2,3] from *single crystal* X-ray diffraction data. However, an intrinsic limitation of this technique is the requirement to prepare a crystal of sufficient size, quality and stability to allow single crystal diffraction data of appropriate quality to be measured. To circumvent this problem, progress has been made recently[6–10] in the development of new techniques for determining crystal structures directly from *powder* diffraction data, particularly with regard to the structure solution stage of the structure determination process. In the field of molecular crystals, the direct-space strategy[10,11] for powder structure solution has been a particularly important development. This paper reports the application of a genetic algorithm (GA) technique for *ab initio* structure determination of the peptide *N*-pivaloyl-L-prolyl-glycyl-*N*′-methylamide (Piv-Pro-Gly-NHMe) from powder X-ray diffraction data.

In the direct-space strategy for powder structure solution, trial structures are generated in direct space, with the quality of each trial structure assessed by comparing the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern (this comparison is made here using the powder profile R-factor R_{wp}, which implicitly takes peak overlap into consideration). In this work, direct-space structure solution was carried out using our GA method[12–19] to locate the trial structure corresponding to the global minimum in R_{wp}. In the GA method,[12–21] a population of trial structures is allowed to evolve subject to rules and operations (mating, mutation and natural selection) analogous to those that govern evolution in biological systems. Each structure is specified by its 'genetic code', which represents, for each molecule in the asymmetric unit, the position $\{x, y, z\}$ and orientation $\{\theta, \phi, \psi\}$ of the molecule, and the molecular conformation (defined by variable torsion angles $\{\tau_1, \tau_2, ..., \tau_n\}$). New structures are generated by the mating and mutation operations, and in our implementation used here,[16] each new structure is subjected to local minimization of R_{wp}. In natural selection, only the structures of highest 'fitness' (*i.e.* lowest R_{wp}) are allowed to pass from one generation to the next generation. Details of our GA methodology for powder structure solution[12–16] and examples of its application[17–19] are given in the cited refs.

The peptide Piv-Pro-Gly-NHMe was prepared by standard procedures,[22] purified by reverse phase HPLC on a C_{18} column and fully characterized by ^{1}H NMR (500 MHz) and electrospray mass spectrometry. The powder X-ray diffraction pattern of a lightly ground sample was recorded at ambient temperature on a Siemens D5000 diffractometer [transmission; $CuK_{\alpha 1}$ (Ge-monochromated); linear position-sensitive detector covering 8° in 2θ]. The 2θ range was 5 to 60°, measured in steps of 0.02° over 12 hours. The powder diffraction pattern was indexed by the program ITO,[23] giving a triclinic unit cell [final refined unit cell parameters following Rietveld refinement: a = 7.9747(3); b = 9.1814(3); c = 5.8456(2) Å; α = 97.020(2); β = 99.429(2); γ = 114.801(2)°]. Density considerations suggested that there is one molecule in the unit cell, and the space group was assigned as $P\bar{1}$. Lineshape and linewidth parameters were determined using the POWDERFIT program,[24] which uses a modified Pawley fitting procedure.[25]

The GA structure solution calculation was carried out using the program EAGER.[26] The structural fragment (Scheme 1) comprised all non-hydrogen atoms of the molecule (with standard bond lengths and angles). For each structure, the genetic code comprised 9 variables $\{\theta, \phi, \psi, \tau_1, \tau_2, ..., \tau_6\}$, with the variable torsion angles allowed to take any value, except τ_5 which was allowed to take only the values 0 or 180°. We note that the position of the structural fragment $\{x, y, z\}$ is fixed arbitrarily in space group $P\bar{1}$. The O–C–N–H torsion angle between τ_3 and τ_4 was fixed[27] at 180°. The population comprised 100 structures, and in each generation 100 offspring (50 pairs of parents) and 20 mutants were produced. The calculation was carried out for 50 generations.

The best structure solution (*i.e.* with lowest R_{wp} in the final generation) was taken as the starting structure for Rietveld refinement,[28] which was carried out using the GSAS program.[29] Standard restraints were applied to bond lengths and angles, and hydrogen atoms were inserted in calculated positions. Three common isotropic displacement parameters were refined, for



Scheme 1 Structural fragment used in the GA structure solution calculation for Piv-Pro-Gly-NHMe, with variable torsion angles indicated by arrows.

† Electronic supplementary information (ESI) available: fractional coordinates and isotropic displacement parameters for atoms in the final refined crystal structure. See http://www.rsc.org/suppdata/cc/b1/b103876c/

 DOI: 10.1039/b103876c

atoms of the main peptide chain, atoms of the ring system and oxygen atoms, respectively. For hydrogen atoms, a fixed common isotropic displacement parameter was used. In the final stages of refinement, a preferred orientation parameter was refined. The final Rietveld refinement (Fig. 1; supplementary information) gave R_{wp} = 4.14, R_p = 5.78 and R_F = 8.97% (75 variables; 269 reflections; 2857 data points).

Fig. 2 shows the final refined structure of Piv-Pro-Gly-NHMe. As shown in Fig. 2a, the molecule adopts a Type II β-turn conformation[30] (ϕ_{Pro} = −58.3; ψ_{Pro} = 127.2; ϕ_{Gly} = 71.4; ψ_{Gly} = 26.1°) stabilized by an intramolecular 4 → 1 hydrogen bond between the C=O group of the Piv residue and the methylamide N–H group (N···O, 2.99 Å; N···O–C, 140.6°). The conformation found here for Piv-Pro-Gly-NHMe is similar to those in other structures containing Type II β-turns, determined previously from single crystal X-ray diffraction data.[31] We note that Pro-Gly sequences in proteins and peptides

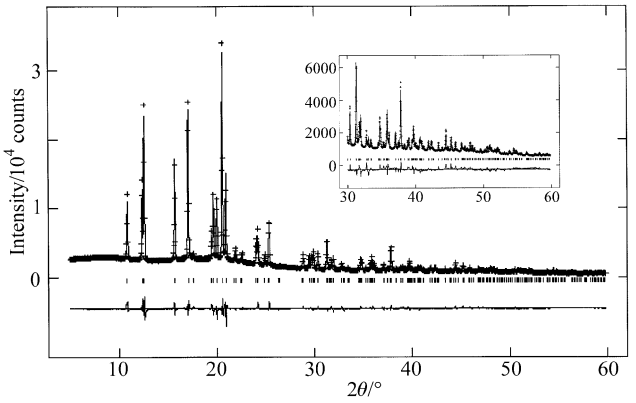


Fig. 1 Experimental (+ marks), calculated (solid line) and difference (lower line) powder X-ray diffraction profiles for the final Rietveld refinement of Piv-Pro-Gly-NHMe. The inset shows an expanded plot for 2θ = 30–60°.

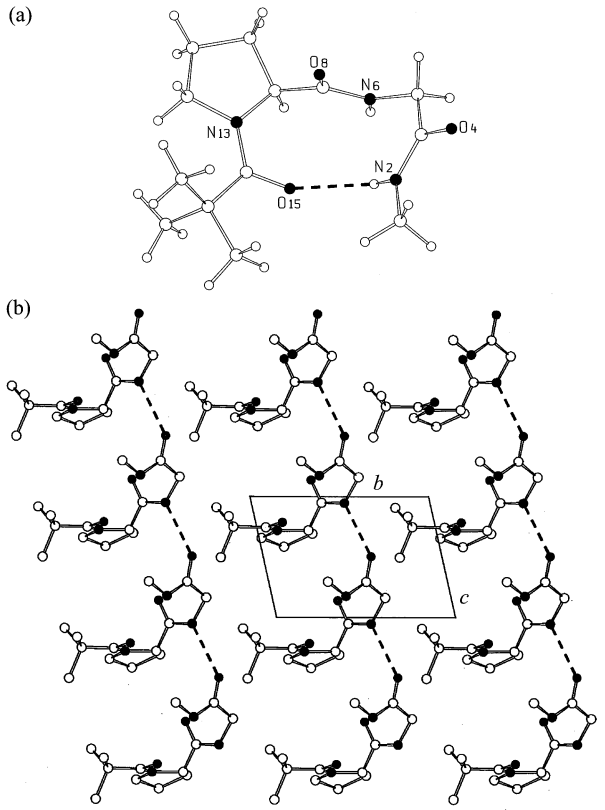


Fig. 2 (a) Molecular geometry of Piv-Pro-Gly-NHMe in the crystal structure, with the intramolecular hydrogen bond shown as a dashed line. (b) Crystal structure of Piv-Pro-Gly-NHMe viewed along the *a*-axis, with N–H···O hydrogen bonds shown as a dashed lines. Hydrogen atoms are omitted for clarity.

may alternatively adopt a Type I β-turn conformation, which is related to the Type II β-turn by a flip of the central peptide unit without disrupting the 4 → 1 hydrogen bond.[5] As shown in Fig. 2b, adjacent molecules in the crystal structure of Piv-Pro-Gly-NHMe are linked along the *c*-axis by intermolecular N–H···O hydrogen bonds (N···O, 2.87 Å; N···O–C, 135.3°).

Our *ab initio* determination of the crystal structure of Piv-Pro-Gly-NHMe from powder X-ray diffraction data has allowed the complete structural characterization of the Type II β-turn conformation as well as the intermolecular interactions in this structure, demonstrating that the direct-space approach for powder structure solution, with the molecular conformation treated as almost completely flexible, is well suited to applications in the field of polypeptides. The potential for applying powder diffraction techniques in this field is clearly important in cases that do not yield single crystals of suitable size and quality for single crystal diffraction studies.

We thank EPSRC, University of Birmingham, Wyeth-Ayerst plc and Ciba Specialty Chemicals for financial support.

Notes and references

1 J. S. Richardson, *Adv. Protein Chem.*, 1981, **34**, 167.
2 G. D. Rose, L. M. Gierasch and J. A. Smith, *Adv. Protein Chem.*, 1985, **37**, 1.
3 C. M. Wilmot and J. M. Thornton, *J. Mol. Biol.*, 1988, **203**, 221.
4 C. Toniolo, *CRC Crit. Rev. Biochem.*, 1980, **9**, 1.
5 K. Gunasekaran, L. Gomathi, C. Ramakrishnan, J. Chandrasekhar and P. Balaram, *J. Mol. Biol.*, 1998, **284**, 1505.
6 A. K. Cheetham and A. P. Wilkinson, *Angew. Chem., Int. Ed. Engl.*, 1992, **31**, 1557.
7 K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554.
8 D. M. Poojary and A. Clearfield, *Acc. Chem. Res.*, 1997, **30**, 414.
9 A. Meden, *Croat. Chim. Acta*, 1998, **71**, 615.
10 K. D. M. Harris, M. Tremayne and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2001, **40**, 1626.
11 K. D. M. Harris, M. Tremayne, P. Lightfoot and P. G. Bruce, *J. Am. Chem. Soc.*, 1994, **116**, 3543.
12 B. M. Kariuki, H. Serrano-González, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189.
13 K. D. M. Harris, R. L. Johnston, B. M. Kariuki and M. Tremayne, *J. Chem. Res. (S)*, 1998, 390.
14 K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Acta Crystallogr. Sect. A*, 1998, **54**, 632.
15 K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Anal. Quím., Int. Ed.*, 1998, **94**, 410.
16 G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Chem. Phys. Lett.*, 2000, **321**, 183.
17 B. M. Kariuki, P. Calcagno, K. D. M. Harris, D. Philp and R. L. Johnston, *Angew. Chem., Int. Ed.*, 1999, **38**, 831.
18 B. M. Kariuki, K. Psallidas, K. D. M. Harris, R. L. Johnston, R. W. Lancaster, S. E. Staniforth and S. M. Cooper, *Chem. Commun.*, 1999, 1677.
19 E. Tedesco, G. W. Turner, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2000, **39**, 4488.
20 K. Shankland, W. I. F. David and T. Csoka, *Z. Kristallogr.*, 1997, **212**, 550.
21 K. Shankland, W. I. F. David, T. Csoka and L. McBride, *Int. J. Pharm.*, 1998, **165**, 117.
22 B. N. N. Rao, A. Kumar, H. Balaram, A. Ravi and P. Balaram, *J. Am. Chem. Soc.*, 1983, **105**, 7423.
23 J. W. Visser, *J. Appl. Crystallogr.*, 1969, **2**, 89.
24 G. E. Engel, S. Wilke, O. König, K. D. M. Harris and F. J. J. Leusen, *J. Appl. Crystallogr.*, 1999, **32**, 1169.
25 G. S. Pawley, *J. Appl. Crystallogr.*, 1981, **14**, 357.
26 S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, O. J. Lanning, E. Tedesco and G. W. Turner, *EAGER*, University of Birmingham, 2000 (an extended version of the program *GAPSS*, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, University of Birmingham, 1997).
27 G. Fischer, *Chem. Soc. Rev.*, 2000, **29**, 119.
28 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65.
29 A. C. Larson and R. B. Von Dreele, Los Alamos Lab. Report No. LA-UR-86-748, 1987.
30 Standard notation is defined in: IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, 1970, **9**, 3471.
31 P. A. Aubry, J. Protas, G. Boussard and M. Marraud, *Acta Crystallogr. Sect. B*, 1980, **36**, 2822.