COMMUNICATION

ChemComm
www.rsc.org/chemcomm

A fully extended tetrapeptide consisting of natural amino acids†

Henrik Birkedal‡,*[a] Dieter Schwarzenbach[a] and Philip Pattison[ab]

[a] *Institute of Crystallography, University of Lausanne, BSP, Dorigny, CH-1015 Lausanne, Switzerland. E-mail: hbirkedal@chem.ucsb.edu; Fax: 01 805 893 4120; Tel: 01 805 893 8126*

[b] *Swiss-Norwegian Beam Line, ESRF, BP 220, F-38043 Grenoble Cedex, France*

Received (in Cambridge, UK) 28th August 2002, Accepted 8th October 2002
First published as an Advance Article on the web 25th October 2002

FGFG is the first example of a non-protected peptide consisting of natural amino acids that adopt a fully extended conformation in the crystalline state.

The backbone conformation of peptides and proteins is determined to a very large degree by the intra- and intermolecular interactions in which the molecule takes part. Examples of this are the α-helix and the β-sheet. In the β-sheet, the peptide backbone is pleated. Pauling and Corey[1] demonstrated that this pleating is necessary if the constituting amino acids have side chains larger than a single hydrogen, *i.e.* a flat β-sheet can only be envisaged in polyglycine. Flat peptides are therefore interesting since they exclude conventional intermolecular backbone hydrogen bonding as a stabilizing motif. Flat peptides adopt the fully extended peptide conformation, $\omega_i = \psi_i = \phi_i = 180°$. They are stabilized by consecutive intramolecular hydrogen bonds: $N_i–H_i\cdots O_i$. This interaction is referred to as the C_5 hydrogen bond as it forms a five-membered ring: $N_i–H_i\cdots O–C_i–C^{\alpha}{}_i$. Several fully extended peptide structures are known in the literature.[2] However, none of these consists exclusively of natural amino acids. Here we show that the tetrapeptide PheGlyPheGly (FGFG), consisting of natural amino acids only, is essentially flat in the crystalline state and that the non-terminal N–H and C=O groups do not participate in classical intermolecular hydrogen bonds. It is the first such peptide consisting only of natural amino acids and the first with unprotected end-groups.

We solved the crystal structure of FGFG using single crystal diffraction data collected at the Swiss Norwegian Beam Line at the ESRF, Grenoble, France. The sample was very small,§ 191 × 13 × 13 μm³. In spite of this minute sample size, data of high quality could be collected and the structure refined to R_1 = 0.0329 for all 1920 reflections.§ The total data collection time was less than 3 h. The molecular structure is shown in Fig. 1.

The peptide backbone is nearly fully extended: the absolute backbone torsion angles are all larger than 159°, the average being 168.7(2)°, see Table 1. There is a small buckling along the backbone. The angles between the least squares planes of the peptide units[3] vary between 5.69(10) and 32.06(12)°, the average being 19.8°. Despite these rather large interpeptidic angles, the backbone is essentially flat, Figs. 1 and 2. This is confirmed by the fact that none of the non-terminal N–H donors (N2, N3, N4) or C=O acceptors (O1, O2, O3) are involved in strong intermolecular interactions. Instead, they all participate in C_5 (Fig. 1) and in intermolecular C–H···O hydrogen bonds (Fig. 2).

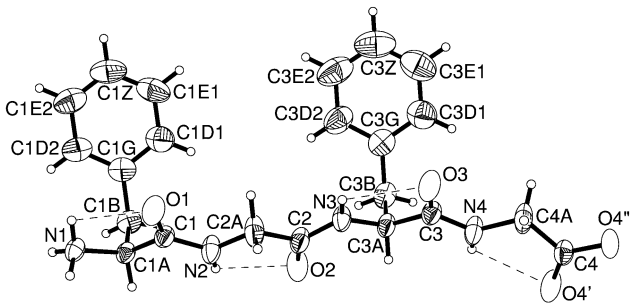

Fig. 1 Molecular structure of FGFG. Ellipsoids are drawn at the 50% probability level.

Table 1 Peptide backbone torsion angles[a] (°) for the experimental and two *ab initio* structures

Angle	Exptl.	RE	RP
ψ_1	162.9(2)	−176.73	155.49
ϕ_2	−165.3(2)	−168.44	96.04
ψ_2	170.6(2)	167.63	−118.89
ϕ_3	−159.5(2)	−173.51	−94.37
ψ_3	162.0(2)	171.91	68.47
ϕ_4	164.3(2)	−172.58	114.04
ψ_4	177.8(2)	178.87	−153.90

[a] $\psi_i = N_i–C_iA–C_i–N_{i+1}$, $\phi_i = C_{i-1}–N_i–C_iA–C_i$

The molecules pack in head-to-tail zigzag chains that run in an antiparallel fashion along the *c*-axis as shown in Fig. 2. The links between molecules along the chains are three-center charge-assisted hydrogen bonds between the carboxylate and ammonium groups. This packing leads to an effective shielding of the charged end-groups and to an overall cancellation of the

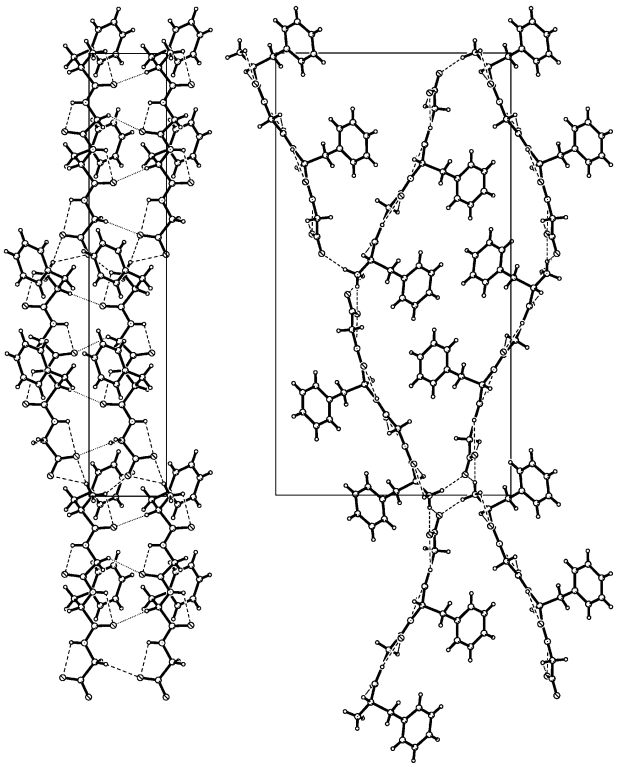

Fig. 2 Crystal packing of FGFG. Left: view onto the *a*,*c*-plane. Right: onto the *b*,*c*-plane. In both cases the *c*-axis is vertical. Dashed and dotted lines represent N–H···O and C–H···O hydrogen bonds, respectively.

† Electronic supplementary information (ESI) available: details of the crystallographic work, *ab initio* calculations (including coordinates) and data base searches. See http://www.rsc.org/suppdata/cc/b2/b208306j/

‡ Present address: Department of Chemistry and Biochemistry, University of California, Santa Barbara CA 93106-9510, USA.

DOI: 10.1039/b208306j

very large molecular dipole moment arising from the extended conformation. In addition, the packing leads to separation of the hydrophobic and hydrophilic parts of the peptide (Fig. 2).

The three carbonyl oxygens and O4′ are involved in the C_5 hydrogen bonds typical of the fully extended peptide conformation.[2] The dimensions are all very similar, the spread in N···O distances being only 0.023(3) Å with an average of 2.676(2) Å. It is the consecutive sequence of C_5 hydrogen bonds that leads to/supports the overall extended conformation of the peptide. The zigzag chains are linked parallel to the *a*-axis by C–H···O interactions, four per molecule (Fig. 2). The average H···O and C···O distances are 2.43 and 3.163(6) Å, respectively. In comparison, all the shortest intermolecular backbone (N–)-H···O distances are longer than 2.6 Å and do not represent hydrogen bonds.

The hydrophobic region is held together by van der Waals contacts and C–H···π interactions between the C1B and C3B protons and neigboring aromatic rings (H···ring-center ~3.1 Å). Interestingly, there are no short π···π contacts (ring-center to ring-center distances are all longer than 5.4 Å).

Ab initio calculations[4] confirm that the fully extended conformer is not the most stable geometry of the zwitterion *in vacuo*. Two geometries were investigated: (1) direct optimization of the X-ray structure (RE) and (2) optimization of a folded conformation found by a previous PM3 calculation (RP). RP is 248 kJ mol^{-1} more stable than RE predominantly due to the large charge separation in RE (~13.6 Å) compared to RP (~2.8 Å). RE is completely extended: the absolute values of the backbone torsion angles are in the range 167.6–178.9°, the average being 174.3°, see Table 1. RE is thus flatter than the conformation in the crystal. This difference is most likely caused by the crystal field, which arises from the presence of the charged end-groups and the separation of the hydrophobic from the hydrophilic parts of the peptide.

To gain further insight into the relation between the C_5 hydrogen bond and the flatness of the peptides, acyclic flat peptides where extracted from the Cambridge Structural Database[5] and combined with the recent literature (see ESI† for details). Two families of peptides were extracted: the fully extended ones (8 structures in total), and molecules that were close to planar but with one or more backbone N–H groups involved in intermolecular hydrogen bonding (3 structures). One example of the latter class is GGG,[6] for which the absolute values of the backbone torsion angles in the two independent molecules vary between 150 and 170°. The hydrogen bonding in GGG is complex and not all possible C_5 bonds are formed. The situation may best be described as a 'modified β-sheet': intermolecular backbone N–H···O hydrogen bonding occurs as in β-sheets but the $C_1^{\alpha}\cdots C_3^{\alpha}$ distances are 7.27 and 7.18 Å in the two molecules as opposed to 6.68 Å in the standard β-sheet. Similar bonding situations occur in [Met^5]enkephalin and in one of the crystal forms of [Leu^5]enkephalin, where backbone torsion angles cover much the same range as in GGG.[7] However, all three structures show much larger deviations from planarity than FGFG and conventional intermolecular backbone N–H···O hydrogen bonding occurs in all of them.

The C_5 N···O distance distribution of the modified β-sheet structures is centered around a significantly longer distance than that of the fully extended peptides, Fig. 3. This reflects that the C_5's are in competition with the intermolecular, strong, N–H···O hydrogen bonds in the former, while they in the latter are the only hydrogen bonds present. The C_5 N···O distances in the FGFG crystal structure are in between those of the previously known fully extended peptides and the modified β-sheet structures, see Fig. 3. The RE structure, on the other hand, is lying within the traditional fully extended distribution. This indicates that the slight buckling of the FGFG backbone and the lengthening of the C_5 hydrogen bonds in the crystal are caused by the influence of the strong crystal field.

It is unlikely that the fully extended conformation is stable in solution and the *ab initio* calculations clearly show that it is not the stable form *in vacuo*. This is easily explained by the long distance between the charged end groups that are neutralized by neighboring molecules in the crystal. In a protic solvent like water, some charge neutralization would take place by interaction with the solvent, but it is more likely that a folded conformation akin to RP would form.

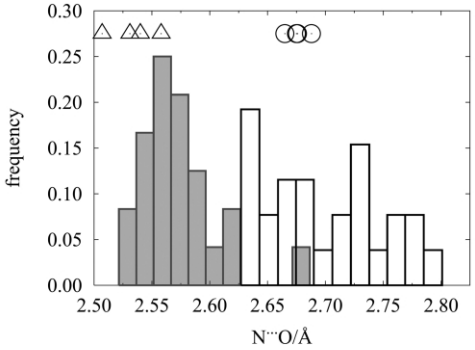

Fig. 3 Distribution of the C_5 intramolecular peptide N–H···O hydrogen bond for fully extended peptides (shaded) and modified β-sheet structures (unshaded). The N···O distances in the FGFG crystal structure (circles) and the RE *ab initio* structure (triangles) are also given.

We thank the staff of the Swiss–Norwegian Beam Line for their kind support, the Swiss National Science Foundation for funding and Mrs Snigireva (ESRF) for access to and assistance with the electron microscope. H. B. thanks the Danish Research Training Council for additional financial support.

Notes and references

§ The single crystal sample size was determined from an electron micrograph. *Crystal data*: $C_{22}H_{26}N_4O_5$, M = 426.47, crystals are transparent and orthorhombic, space group $P2_12_12_1$ (no. 19), Z = 4. a = 5.063(1), b = 15.379(3), c = 28.912(6) Å, U = 2251.2(8) Å^3. 12283 reflections measured, 1920 unique (R_{int} = 0.041). R_1 = 0.0298 for 1795 reflections with $F_o^2 > 2\sigma(F_o^2)$ and R_1 = 0.0329 for all reflections. The absolute conformation was fixed by the known configuration of the peptide. MAR345 imaging plate. λ = 0.8008 Å, T = 293 K. Data reduction was performed with the HKL-package (Y. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307–326). Structure solution and refinement were done with SHELXS-97 (Sheldrick, G. M. 1997. SHELXS-97, program for the solution of crystal structures, University of Göttingen, Germany) and SHELXL-97 (G. M. Sheldrick, SHELXL-97, program for the refinement of crystal structures, University of Göttingen, Germany, 1997). CCDC 192545. See http://www.rsc.org/suppdata/cc/b2/b208306j/ for crystallographic data in CIF or other electronic format.

1 L. Pauling and R. B. Corey, *Proc. Natl. Acad. Sci. USA*, 1951, **37**, 729.

2 (*a*) C. Toniolo and E. Benedetti, in, *Molecular Conformation and Biological Interactions: G. N. Ramachandran Festschrift*, ed. P. Balaram and S. Ramaseshan, Indian Academy of Sciences, Bangalore, India, 1991, pp. 511–521 and references therein; (*b*) M. Crisma, F. Formaggio, C. Toniolo, T. Yoshikawa and T. Wakamiya, *J. Am. Chem. Soc.*, 1999, **121**, 3272; (*c*) N. Imakawa, M. Tanaka and H. Suemune, *Helv. Chim. Acta*, 2000, **83**, 2823.

3 Defined by C_i, C_i^{α}, O_i, N_{i+1} except for residue four, where the plane is defined by C_4, C_4^{α}, O_4' and O_4''.

4 Calculations were done at the RHF/6-31G(d) level using Gaussian 98 (Revision A.7), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

5 Cambridge Structural Database version 5.17, April 1999. F. Allen and O. Kennard, *Chem. Des.. Autom. News*, 1993, **8**, 31.

6 T. Srikrishnan, N. Winiewicz and R. Parthasarathy, *Int. J. Pept. Protein Res.*, 1982, **19**, 103.

7 J. F. Griffin, D. A. Langs, G. D. Smith, T. L. Blundell, I. J. Tickle and S. Bedarkar, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 3272.