Pergamon

Tetrahedron Letters 42 (2001) 4691–4694

TETRAHEDRON LETTERS

Crystallographic characterisation of novel β-turn like folds in a model peptide: stabilisation by main-chain to side-chain interactions

A. K. Thakur and R. Kishore*

Institute of Microbial Technology, Sector 39-A, Chandigarh 160 036, India

Received 3 January 2001; revised 20 March 2001; accepted 11 May 2001

Abstract—X-Ray diffraction analysis of the model peptide Boc-Thr-Thr-OCH_3 reveals the existence of distinctive conformational characteristics: semi-extended ($\phi = -62.1°$; $\psi = 137.1°$) and semi-folded ($\phi = -130.3°$; $\psi_T = 5.7°$) of the N- and C-terminus Thr residues, respectively. Surprisingly, an overall significantly 'flat' conformation is stabilised by a number of novel main-chain to side-chain intramolecular hydrogen bonds, i.e. two non-conventional: C^{γ}_{i+1}–H···O=C_{i-1} and C^{γ}_{i}–H···O_{i+1} types and two conventional: N_i–H···O^{γ}_{i} and O^{γ}_{i}–H···N_i types of interactions.

Precise information about the intrinsic folding–unfolding propensities of an individual amino acid residue is of fundamental importance in the design and construction of diverse structural motifs since their distribution in proteins and polypeptides within the class of secondary structure is not random.[1a] Of the 20 naturally occurring proteinogenic residues, Thr is the only amino acid which has a C^{β}-branched amphiphilic stereogenic centre (Thr versus Ile, Fig. 1) and is expected to impose stronger constraints on the conformational preference.[1] In an attempt to resolve the relative β-sheet forming preferences of each of the 20 amino acids, as analysed by protein engineering experiments, Minor and Kim[2c] demonstrated that a Thr residue has the highest intrinsic propensity for a β-sheet conformation and assessed this to be due to a thermodynamic origin.[2a,b] Additionally, Zimm–Bragg parameters of 20 amino acids, determined experimentally by the host–guest technique and rationalised by conformational energy calculations, revealed that a Thr residue also exhibits a strong tendency to hydrogen bond to main-chain atoms in a non-helical conformation, i.e. which lowers the *s* value.[3a]

Our previous conformational analysis of a model peptide incorporating two C^{β}-stereogenic centres: hydrophobic and amphiphilic, resulted in characterisation of an uncommon intramolecularly H-bonded C_5-secondary structural feature, i.e. an N_i–H···O=C_i interaction across the Thr residue, which in turn favoured the formation of a unique main-chain to side-chain intramolecular H-bond, i.e. O^{γ}_{i}···H–N_{i+1} interaction.[4] In this Letter we extend the study and report the crystal molecular structure of a model peptide Boc-Thr-Thr-OCH_3 **1**, incorporating two amphiphilic C^{β}-stereogenic centres.[5] Unexpectedly, the analysis reveals the existence of an unusual significantly 'flat' backbone conformation stabilised by a number of novel main-chain to side-chain intramolecular H-bonding interactions.

One of the most attractive features of the crystal molecular structure of **1** is the characterisation of distinctive backbone conformational characteristics of the N- and C-terminus Thr residues (Fig. 2). While the torsional angles (Table 1) of the former Thr residue prefers a semi-extended conformation ($\phi = -62.1°$, $\psi = 137.1°$), the latter adopts a semi-folded structure ($\phi = -130.3°$, $\psi = 5.7°$). An overall significantly 'flat' architecture, resulting from semi-extended and semi-folded main-

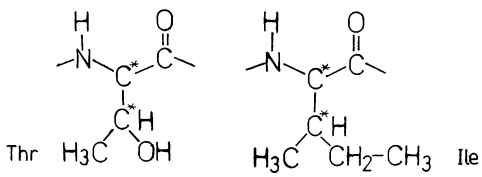

Figure 1. Schematic presentation of the side-chain functionalities: amphiphilic Thr (left) and hydrophobic Ile (right). Asterisks indicate C^{α} and C^{β} stereogenic centres.

* Corresponding author.

0040-4039/01/$ - see front matter
PII: S0040-4039(01)00788-2

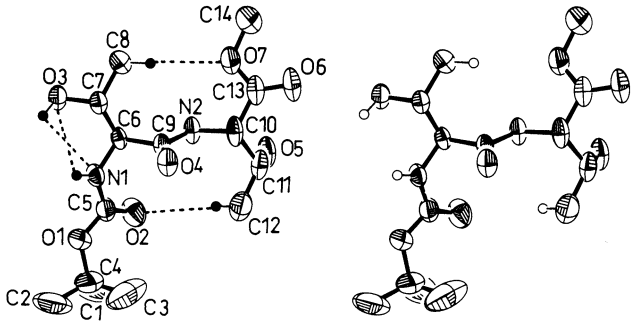

Figure 2. An ORTEP presentation of the crystal molecular structure of Boc-Thr-Thr-OCH_3 **1**. The thermal ellipsoids are shown to the 50% probability level. Dotted lines indicate H-bonds. (·) H-atoms involved in interactions.

chain conformations of the two Thr residues in a short linear peptide is indeed surprising since detailed analysis of protein three-dimensional structures, coupled with the conclusions of protein engineering experiments, revealed that the residue strongly prefers an extended β-sheet conformation.[1–4] Consistent with this proposal[3] the preferred non-helical backbone conformation of **1** may be the consequence of four specific intramolecular main-chain to side-chain H-bonding interactions.

There exist two non-conventional[6] (C12–H···O2=C5 and C8–H···O7-ethereal) and two conventional (N1–H···O3 and O3–H···N1) short-range interactions in **1**. The geometric parameters (Table 2) are well accommodated within the accepted definition of H-bonding interactions. Of the two non-conventional H-bonds the C12–H···O2=C5 interaction encompasses a novel ten-membered ring motif: reminiscent of a β-turn like-fold, topologically similar to the well-known Asx turn (the side-chain C=O of the Asp/Asn residue *i*, H-bonds to main-chain NH of residue *i*+2), whereas the C8–H···O7 interaction constitutes an unusual nine-membered ring motif (Fig. 2). Each of the conventional H-bonding interactions encompasses a five-membered ring motif. The simultaneous existence of a variety of short-range main-chain to side-chain interactions in a short linear peptide of the size in question, is indeed interesting. Energetic contributions of non-conventional weak C-H···O interactions in proteins and polypeptides[7] are accepted as potential stabilising forces.[8] The energy of such interactions is suggested to range between ~0.5–5.0 kcal/mol.[6b]

Since, the OH group can function both as a H-bond donor and an acceptor, the O^{γ}–H functionality of the N-terminus Thr residue precisely participates in the fabrication of two intraresidue H-bonds i.e. N_i–H···O_i^{γ} and O_i^{γ}–H···N_i interactions (Table 2). As expected, in these pentagonal ring motifs, the D–H···A H-bond angles are notably non-linear.[4] While the existence of an N_i–H···O_i^{γ} H-bond is the most frequently observed main-chain to side–chain interaction in proteins and polypeptides, the characterisation of the O_i^{γ}–H···N_i interaction in **1**, at atomic resolution, represents another attractive feature of the crystal molecular structure. The stabilisation of an intraresidue O^{γ}–H···N interaction across the Thr residue, so far, is supported by theoretical calculations.[3b,c] However, proof of its existence in crystallised proteins would not have been possible because of the low accuracy of the H-atoms atomic coordinates.[8] In our study the H-atoms participating in closing the ring motifs could be unambiguously located in a difference Fourier map.[5] The statistically observed distribution of Thr side-chains in proteins however, indicates that the O^{γ}H functionalities occur predominantly as either *gauche*$^+$ or *gauche*$^-$, whereas *trans* orientations are rare.[1,3] Moreover, a survey of the side-chain distributions of the Thr residues, in 258 crystal structures of oligopeptides, further validate the stereochemical and the energetic preference for O^{γ}H as *gauche*.[1c,9] Interestingly, in **1** the preferred *trans* orientation of the $C^{\gamma}H_3$ group of the N-terminus Thr, the torsion angle being ≈−176°, clearly favours its participation in the formation of an unusual non-conventional intramolecular interaction that stabilises a β-turn like-fold.

We propose that the preferred structural motifs in **1** may be unique in nature, at least in Thr containing peptides. Compelling evidence for this is provided from

Table 2. The geometric parameters of the intramolecular H-bonding interactions for Boc-Thr-Thr-OCH_3 **1**

Donor	Acceptor	Distances (Å)		Angles (°)
D	A	D···A	H···A	D–H···A
Conventional				
N1–H	O3	2.810	2.617	96.5
O3–H	N1	2.810	2.663	90.0
Non-conventional				
C12–H	O2	3.648	2.770	158.1
C8–H	O7	3.702	2.907	132.6

Table 1. Selected torsion angles (°) for Boc-Thr-Thr-OCH_3 **1**

Torsion angles			Torsion angles		
C4–O1–C5–N1	(θ1)	−165.5(8)	C6–C9–N2–C10	(ω1)	173.5(3)
O1–C5–N1–C6	(ω0)	179.5(3)	C9–N2–C10–C13	(ϕ2)	−130.3(3)
C5–N1–C6–C9	(ϕ1)	−62.1(3)	N2–C10–C13–O7	(ψT)	5.7(4)
N1–C6–C9–N2	(ψ1)	137.1(2)	C10–C13–O7–C14	(ωT)	−177.8(4)
N1–C6–C7–O3[a]	(χ1)	−56.4(3)	N2–C10–C11–O5[a]	(χ2)	73.5(3)
N1–C6–C7–C8	(χ1)	−176.1(3)	N2–C10–C11–C12	(χ2)	−52.9(4)

[a] In the text, the side-chain torsion angle χ of the Thr residue is defined as N^{α}–C^{α}–C^{β}–O^{γ}.

the X-ray crystal structure analysis of Boc-Phe-D-Leu-Thr-OCH_3, which also has the Thr-OCH_3 moiety;[10] the torsion angles: ($\phi_{Leu}=66.8°$, $\psi_{Leu}=41.4°$, $\phi_{Thr}=-95.6°$, $\chi_{Thr}=49.9°$) and ($\phi_{Thr}=-95.6°$, $\psi_{Thr}=14.3°$ and $\chi_{Thr}=49.9°$), are fully compatible with two main-chain to side-chain intramolecular H-bonds, i.e. Thr C^{γ}_{i+2}–H···O=C_i Phe (non-conventional) and N_{i+2}–H···O^{γ}_{i+2} (conventional) interactions, encompassing ten- and five-membered ring motifs, respectively. The geometric parameters of these two weak interactions compare well with those determined for the corresponding H-bonding interactions in **1** (data not shown).

In conclusion, the conformational analysis of our short linear model peptide unambiguously demonstrates that short-range main-chain to side-chain intramolecular interactions are of primary importance in dictating an overall unusual folding behaviour. An observation of two distinct: quasi-extended and folded, conformations of the Thr residues in **1** are at variance from those reported from protein engineering experiments.[2] The present analysis is valuable in understanding the backbone dependent distributions of the side-chain dihedral angles, which is the first step towards the rational prediction of secondary and complex tertiary structures. We believe that the manifestation of C–H···O intramolecular interactions in short linear peptides, particularly in the absence of main-chain to main-chain amide–amide interactions, may dramatically influence and stabilise unusual preferred folded–unfolded conformations.[8] Finally, the information can also be exploited for the design and construction of biologically active peptide analogues, particularly the 'peptide T' (H-Ala-Ser-Thr-Thr-Thr-Asn-Tyr-Thr-OH), which is shown to be a potent antagonist for the human immunodeficiency virus (HIV) attachment to its receptor(s) on human T-cells.[11]

Acknowledgements

This investigation was financially supported in part by DBT and CSIR, Government of India. R. Kishore gratefully acknowledges the Director for his consistent encouragement. This is IMTech Communication No. 42/2000.

References

1. (a) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211–222; (b) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357–386; (c) McGregor, M.; Islam, S. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1987**, *198*, 295–310; (d) Richardson, J. S.; Richardson, D. V. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G. D., Ed.; Plenum Press: New York, 1990; pp. 1–98; (e) Chakrabarti, P.; Pal, D. *Prot. Eng.* **1998**, *11*, 631–647.
2. (a) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267–270; (b) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510–5517; (c) Minor, Jr., D. L.; Kim, P. S. *Nature* **1994**, *367*, 660–663.
3. (a) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535. The '*s*' is an equilibrium constant (statistical weight) which defines helical units; (b) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Israel J. Chem.* **1973**, *11*, 121–152; (c) Scheraga, H. A. In *Perspectives in Structural Biology*; Vijayan, M.; Yathindra, N.; Kolaskar, A. S., Eds.; Universities Press: India, 1999; pp. 275–292.
4. (a) Ashish; Kishore, R. *Tetrahedron Lett.* **1997**, *37*, 5747–5750; (b) Ashish; Kishore, R. *FEBS Lett.* **1997**, *417*, 97–100; (c) Thakur, A. K.; Ashish; Kishore, R. *Chem. Commun.* **1999**, *1*, 1643–1644 and references therein.
5. The peptide Boc-Thr-Thr-OCH_3 **1**, was synthesised using standard solution phase procedures and purified to homogeneity employing silica-gel (60–120 mesh) column chromatography (≈2–3% MeOH–$CHCl_3$ mixtures). White solid; mp = 140°C; $[\alpha]_D^{25}=-28.6$ ($c=0.5$, MeOH); $R_f=0.52$ (10% MeOH–$CHCl_3$ mixture); ν_{max} (FT-IR, KBr, crystal) 3447, 3346, 3279, 1728, 1714, 1660, 1570, 1529 cm^{-1}; δ_H (300 MHz ^1H NMR, $CDCl_3$, 25°C, 8 mg/ml, TMS) 1.22 (6H, d, $^3J=6.3$ Hz, Thr^{1,2} $C^{\gamma}H_3$), 1.46 (9H, s, $(CH_3)_3$), 2.78/3.46 (2H, s/s, $O^{\gamma}H$), 3.78 (3H, s, OCH_3), 4.13/4.58 (2H, d/dd, $^3J=7.6/8.9$ Hz, Thr^{1,2} $C^{\alpha}H$), 4.35 (2H, m, Thr^{1,2} $C^{\beta}H$), 5.56 (1H, d, $^3J=7.8$ Hz, Thr^1 NH), 8.8 (1H, d, $^3J=7.0$ Hz, Thr^2 NH). Single crystals suitable for X-ray diffraction study were obtained from EtOAc:pet. ether (2:3 v/v) solutions by slow evaporation at room temperature. The diffraction data were collected on Siemens P4 single-crystal diffractometer equipped with MoKα radiation and highly oriented graphite monochromator ($\lambda=0.71073$ Å). Crystal data **1**: molecular formula $C_{14}H_{26}N_2O_7$, molecular weight = 334.37, monoclinic, $P2_1$, cell constants: $a=11.190$, $b=7.382$, $c=11.905$ Å, $\alpha=\gamma=90$, $\beta=107.14°$, $V=939.14$ Å^3, $Z=2$, $D_c=1.182$ mg m^{-3}, $T=293$ K, final $R=0.0343$, final $R_w=0.0967$. Details of the crystal structure solution and refinement by application of direct method programme SHELX-97, shall be presented elsewhere.
6. (a) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 5063–5070; (b) Desiraju, G. R. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2311–2327; (c) Desiraju, G. R. *Acc. Chem. Res.* **1996**, *29*, 441–449; (d) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond in Structural Chemistry and Biology*; Oxford University Press: Oxford, 1999; (e) Gu, Y.; Kar, T.; Scheiner, S. *J. Am. Chem. Soc.* **1999**, *121*, 9411–9422; (f) Vargas, R.; Garza, J.; Dixon, D. A.; Hay, B. P. *J. Am. Chem. Soc.* **2000**, *122*, 4750–4755 and references cited therein.
7. (a) Derewenda, Z. S.; Derewenda, U.; Kobos, P. M. *J. Mol. Biol.* **1994**, *241*, 83–93; (b) Derewenda, Z. S.; Lee, L.; Derewenda, U. *J. Mol. Biol.* **1995**, *252*, 248–262; (c) Bella, J.; Berman, H. M. *J. Mol. Biol.* **1996**, *264*, 734–742; (d) Musah, R. A.; Jensen, G. M.; Rosenfield, R. A.; McRee, D. E.; Goodin, D. B.; Bunte, S. W. *J. Am. Chem. Soc.* **1997**, *119*, 9083–9084; (e) Chakrabarti, P.; Chakrabarti, S. *J. Mol. Biol.* **1998**, *284*, 867–873.
8. (a) Crisma, M.; Formaggio, F.; Toniolo, C.; Yoshikawa, T.; Wakamiya, T. *J. Am. Chem. Soc.* **1999**, *121*, 3272–3278; (b) Thakur, A. K.; Kishore, R. *Biopolymers* **2000**, *53*, 447–454; (c) Thakur, A. K.; Venugopalan, P.; Kishore, R. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 492–498; (d) Arvinda, S.; Shamala, N.; Pramanik, A.; Das, C.; Balaram, P. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 933–936; (e) Thakur, A. K.; Venugopalan, P.; Kishore, R. *J. Peptide Res.* **2000**, *56*, 55–58.

9. (a) Benadetti, E.; Morelli, G.; Nemethy, G.; Scheraga, H. A. *Int. J. Peptide Protein Res.* **1983**, *22*, 1–15; (b) Aubry, A.; Marraud, M. *Acta. Crystallogr., Sect. C* **1985**, *41*, 65–67; (c) Ref. 3 and references therein.
10. Doi, M.; In, Y.; Tanaka, M.; Inoue, M.; Ishida, T. *Acta. Crystallogr., Sect. C* **1993**, *49*, 1530–1532.
11. (a) Pert, C. B.; Hill, J. M.; Ruff, M. R.; Berman, R. M.; Robey, W. G.; Arthur, L. O.; Ruscetti, F. W.; Farrar, W. L. *Proc. Natl. Acad. Sci. USA* **1986**, *23*, 9254–9258; (b) Picone, D.; Temussi, P. A.; Marastoni, M.; Tomatis, R.; Motta, A. *FEBS Lett.* **1988**, *231*, 159–163.