Synthetic Circularized Analogues of Bovine Pancreatic Trypsin Inhibitor[‡]

Judit Tulla-Puche,[a] Irina V. Getun,[a] Jordi Alsina,[a] Fernando Albericio,[b] and George Barany*[a]

Dedicated to the memory of Professor Murray Goodman (1928−2004)[‡‡]

Keywords: Circular proteins / Native chemical ligation / Solid-phase synthesis / Orthogonal protecting groups / Thiolesters / Cyclization

Two cyclic analogues of the protein bovine pancreatic trypsin inhibitor (BPTI), c-$[R]_{Smc}$ and c-Cys5$[R]_{Abu}$, have been synthesized. For the first target, a semisynthetic approach featured a cyclization that took advantage of the constraint imposed by the three disulfides of the native protein. For the second target, an entirely unstructured 58-residue thiolester, prepared by total synthesis with mild Fmoc chemistry and a side-chain anchoring strategy, was cyclized by native chemical ligation.

Introduction

While cyclic peptides and peptidomimetics (typically having 2 to 20 amino acid residues), both naturally occurring and synthetically constructed, have been a staple topic of biochemical investigations for many years because of their presumed greater structural and metabolic stability,[1,2] it has been only within the past decade or so that examples of circular proteins (> 25 residues) have emerged.[3] Prominent among these is the plant-derived cyclotide family,[3c,3d] members of which show significant biological activities; all their covalent structures have three disulfide bridges in addition to the head-to-tail backbone connections of the termini. In contrast, the antibiotic protein enterocin AS-48 contains 70 amino acid residues, and the only constraint is an amide bond responsible for circularity.[4] Regarding the chemistry, the synthesis of cyclic peptides and, in particular, circular proteins[5] presents interesting challenges.[1,2] This is because successful outcomes of cyclization reactions are likely to depend on the capacity of a given linear sequence to adopt appropriate conformations in which the termini are sufficiently close to each other so that covalent coupling is feasible.

The artificially circularized protein prototype is cyclic bovine pancreatic trypsin inhibitor, c-[BPTI], described and characterized by Goldenberg and Creighton in an elegant series of papers.[6] Starting with the exceedingly stable 58-residue folded native protein that has three disulfide bridge constraints as well as its *N*- and *C*-termini in close three-dimensional proximity (Figure 1), a head-to-tail amide crosslink was created chemically.[6,7,8]

Our interest in circularized analogues of BPTI derives from the observation that pairwise replacement of Cys−Cys intramolecular bridges in BPTI by α-aminobutyric acid (Abu) isosteres results in partially folded or unfolded proteins.[9] Much of our work focuses on chemically synthesized variants where two of the three disulfides have been removed by substitution of Cys by Abu, and the *only* retained crosslink is the disulfide between Cys-14 and Cys-38; the resultant analogue $[14-38]_{Abu}$ is partially folded with an intact central core.[9b,9c] Interpreted through our lens, analogues − made by other investigators − that we refer to as $[30\text{-}51]_{Ala}$, $[30-51]_{Ser}$, $[5-55]_{Ala}$, and $[5-55]_{Ser}$ may similarly be considered as *partially folded* with the *same* central core.[10] In contrast, the "reduced" synthetic analogue referred to as $[R]_{Abu}$, with six Abu, i.e., no disulfides, is entirely *unfolded.* This is a prototype of what we term a "molten coil"; it does however retain some NMR-detectable *non-native* order.[9b,9d] Our generalized hypothesis is that *any* single crosslink in otherwise reduced BPTI destabilizes the unfolded state and thereby stabilizes more collapsed species (entropy effect); among collapsed

[‡] This contribution was taken in part from the February 2004 Ph.D. thesis of J. Tulla-Puche.

[a] Department of Chemistry, University of Minnesota
Minneapolis, MN 55455, USA
Fax: (internat.) + 1-612-626-7541
E-mail: barany@umn.edu
[b] Barcelona Biomedical Research Institute, Barcelona Science Park, University of Barcelona,
08028 Barcelona, Spain

[‡‡] Professor Murray Goodman was an international leader in the field of peptide science.[1]

Bovine Pancreatic Trypsin Inhibitor (BPTI)

Arg[1]-Pro-Asp-Phe-Xxx-
Leu-Glu-Pro-Pro-Tyr-
Thr-Gly-Pro-Xxx-Lys-
Ala-Arg-Ile-Ile-Arg-
Tyr-Phe-Tyr-Asn-Ala-
Lys-Ala-Gly-Leu-Xxx-
Gln-Thr-Phe-Val-Tyr-
Gly-Gly-Xxx-Arg-Ala-
Lys-Arg-Asn-Asn-Phe-
Lys-Ser-Ala-Glu-Asp-
Xxx-Met-Arg-Thr-Xxx-
Gly-Gly-Ala[58]

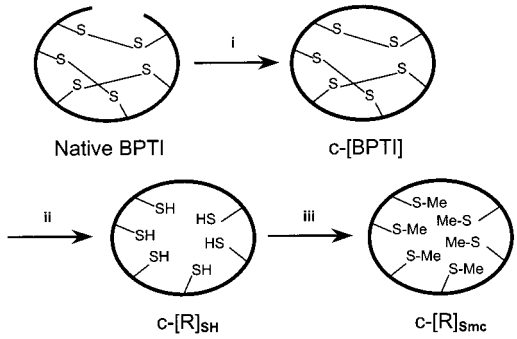
Figure 1. Left: Ribbon diagram (PyMOL software) structure of BPTI (PDB accession no. 6PTI); the wild-type cysteine positions are labelled, but the three naturally occurring disulfides, 5−55, 14−38, and 30−51, are not shown as crosslinks; note the central antiparallel β-sheet core, the 14−38 disulfide near the trypsin specificity pocket, and the proximity of the termini; right: target amino acid sequence, in which the six replaced cysteines are indicated by Xxx; in c-[R]$_{Smc}$, the termini are joined covalently, and all Xxx are *S*-methylcysteine (Smc); in c-Cys5[R]$_{Abu}$, the termini are also joined covalently, Xxx at position 5 remains as Cys, and the remaining Xxx are α-aminobutyric acid (Abu)

species, those with a nativelike central core are favored. The goal of the present investigation is to prepare representatives of the c-[R]$_{Abu}$ family in order to find out whether adding a *non-disulfide crosslink* leads to a partially folded BPTI.

Results and Discussion

Initially, a semisynthetic approach was followed, based on the previously established similarity[9d,9e] of [R]$_{Abu}$, the prototype linear molten coil, to the analogue in which each half-cystine residue is replaced by *S*-methylcysteine (Smc). The thought was that the same set of reactions used to convert native BPTI to [R]$_{Smc}$ could be carried out instead on c-[BPTI]. Following the literature,[6a,6b] BPTI was closed covalently by water-soluble carbodiimide treatment at pH 4 [Scheme 1, step (i)]. The resultant c-[BPTI], obtained reproducibly in approximately 50% yield, was subjected to comprehensive reduction/permethylation, followed by careful HPLC purification, to provide pure c-[R]$_{Smc}$ in 15% overall isolated yield (Scheme 1).

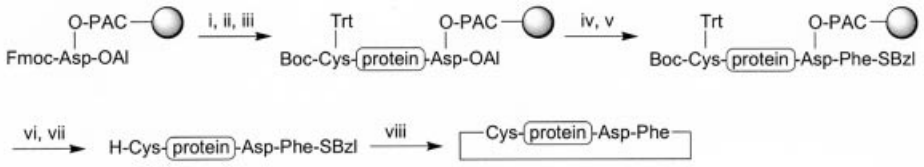
As the research progressed, it became apparent that, because of the extra steric bulk of the *S*-methyl groups, c-[R]$_{Smc}$ is not an ideal model for the desired c-[R]$_{Abu}$. However, the goal to make a cyclic analogue containing only isosteric Abu residues introduces a new challenge, because appropriate linear precursors lack the constraints to facilitate the proximity of the termini to be joined. In our hands, multiple and varied efforts involving strategies and methodologies, including *B*ackbone *A*mide *L*inker (BAL) approaches,[11] that are often successful for smaller structures[1,2] were problematic when applied to the c-[R]$_{Abu}$ target. We therefore sought related structures amenable to alternative chemistries and were encouraged by several reports that native chemical ligation reactions[12] can be harnessed to achieve cyclization in peptides and proteins.[7,8a,8g,13] Such latter strategies require at the *N*-terminus a cysteine (or equivalent) residue with both the N^{α}-amine and the β-thiol free, along with a thiolester at the *C*-terminus. To date, the peptide thiolester intermediates required for cyclization have almost invariably been assembled by *tert*-butyloxycarbonyl (Boc) chemistry,[1,12] rather than the milder fluorenylmethoxycarbonyl (Fmoc) chemistry.[1,8g,14]



Scheme 1. Semisynthesis of c-[R]$_{Smc}$; the steps are: (i) 10 mM of 1-[3-(dimethylamino)propyl]-3-ethyl-carbodiimide hydrochloride (EDC) in the presence of *N*-hydroxysuccinimide (3 equiv.) in H_2O at pH 4; (ii) 10 mM of DTT in 0.1 M Tris·HCl, 6 M guanidinium hydrochloride, 1 mM EDTA, pH 8.7; (iii) 16 mM of CH_3I in 0.1 M Tris·HCl, 1 mM EDTA, pH 8.7; the final product was characterized by MALDI-TOF (calcd. 6584.7, found 6584.4)

With the aforementioned in mind, we dissected the linearly permuted BPTI sequence between Phe-4 and Cys-5, and replaced each of the five other Cys residues (i.e., 14, 30, 38, 51, and 55) with Abu. The chosen cyclization point presents several advantages: (i) adjacent to the chosen *C*-terminus Phe-4, an Asp-3 residue is available as a useful site for side-chain anchoring[15] to the solid-phase synthesis support; (ii) a robust "backing-off" procedure to introduce the necessary *C*-terminal thiolester has already been developed;[11b,14] (iii) combining the previous two insights, completion of main chain assembly by Fmoc chemistry could be followed by a selective deprotection step and then direct introduction of thiolester-activated Phe-4; (iv) compared to the original goal of c-[R]$_{Abu}$, substitution of Cys-5 for an Abu is considered an insignificant perturbation, since Abu-5 in other single-crosslink analogues is relatively disordered.

Based on the overall outline (Scheme 2 and immediately preceding text), the synthesis started with Fmoc-Asp(O-PAC-PEG-PS)-OAllyl;[14,15] and further chain elongation was carried out by standard Fmoc solid-phase peptide synthesis, using HATU/1-hydroxy-7-azabenzotriazole (HOAt)/DIEA couplings in DMF. N^{α}-Boc, *S*-tritylcysteine was used as precursor to the key *N*-terminal Cys residue. Removal of the allyl protecting group was mediated by Pd^0 in the presence of $PhSiH_3$,[16] and attachment of phenylalanine *S*-benzyl ester[11b,14] completed establishment of the overall



Scheme 2. Synthesis of c-Cys5[R]$_{Abu}$ featuring side-chain anchoring to a polyethylene glycol−polystyrene (PEG-PS) support; the steps are: (i) standard peptide chain elongation by Fmoc chemistry; (ii) piperidine − *N*,*N*-dimethylformamide (DMF) (1:4); (iii) Boc-Cys(Trt)-OH, *N*,*N*′-diisopropylcarbodiimide (DIPCDI)/1-hydroxybenzotriazole (HOBt) in DMF; (iv) Pd(PPh$_3$)$_4$ (0.1 equiv.)/PhSiH$_3$ (24 equiv.) in CH_2Cl_2; (v) H-Phe-SBzl, 1-[bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium hexafluorophosphate 3-oxide (HATU)/*N*,*N*-diisopropylethylamine (DIEA) in CH_2Cl_2; (vi) Reagent R [trifluoroacetic acid (TFA) − thioanisole − 1,2-ethanedithiol − anisole (90:5:3:2)]; (vii) preparative HPLC; (viii) 0.25 M sodium phosphate, 6 M guanidinium hydrochloride, 1% thiophenol, at pH 7.5; throughout chain elongation, there was *partial* oxidation at Met 52, and this was exacerbated at step (v); the final purified product showed, upon high resolution ESMS, a mass value characteristic of complete oxidation to Met(O) (calcd. 6423.6627, found 6424.3094); there are ample literature methods[1,9a] to reduce inadvertently formed Met(O) to Met; an alternative strategy would be to make an analogue with Nle instead of Met

linear sequence. The required peptide thiolester intermediate was cleaved from the resin by treatment with Reagent R, with concomitant removal of all side-chain protecting groups as well as the *N*-terminal Boc group. This crude material was purified by preparative reversed-phase HPLC under acidic conditions, and native chemical ligation was effected after the purified peptide thiolester was taken up in pH 7.5 buffer in the presence of guanidinium hydrochloride as a denaturant and thiophenol as an exchange reagent.[13f] After 2 h, intramolecular solution ligation was essentially complete, and preparative reversed-phase HPLC purification gave the circular protein in 11% overall yield (Scheme 2, Figure 2). After tryptic digestion (pH 7.2, 37 °C, 1 h), seven fragments were separated by HPLC. One of these, Thr-Abu-Gly-Gly-Ala-Arg-Pro-Asp-Phe-Cys-Leu-Glu-Pro-Pro-Tyr-Thr-Gly-Pro-Abu-Lys,[17] which spans the key Phe−Cys amide bond formed during ligation/cyclization, supports assignment of the cyclic structure.

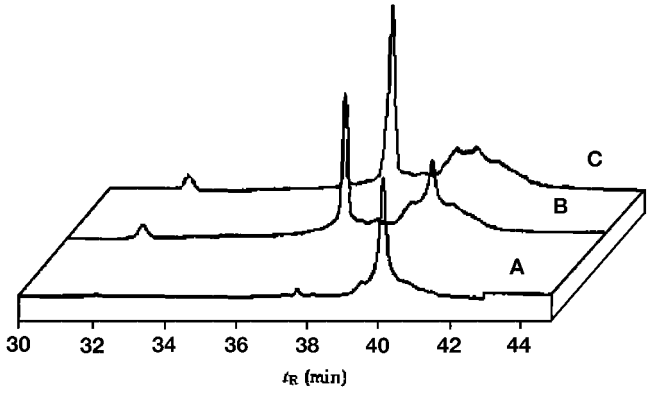


Figure 2. Circularization by native chemical ligation, under conditions of Scheme 2, step (viii), as followed by HPLC; panel A is at start of reaction (note linear peptide thiolester at t_R = 39.4 min), panel B is after 1 h (note appearance of cyclized protein product at t_R = 37.5 min), and panel C is after 2 h (considered to be the endpoint); reversed-phase analytical HPLC was carried out on Vydac C-4 (210TP1010; 10 μm particle size, 0.46×15 cm), and the components were eluted with a linear gradient of 0.1% aqueous TFA−0.1% TFA in CH_3CN, from 9:1 to 1:1 over 40 min, run at 1.2 mL/min flow rate

Conclusions

Two methods have been demonstrated to achieve circularization of BPTI analogues. Semisynthesis of c-[R]$_{Smc}$ was accomplished by treatment of native BPTI with a water-soluble carbodiimide, followed by comprehensive reduction of all three disulfide bridges, and final permethylation. The analogue c-Cys5[R]$_{Abu}$ was obtained in respectable yield through construction of the appropriate 58-residue peptide linear precursor by convenient, mild Fmoc solid-phase synthesis chemistry with side-chain anchoring, careful on-resin installation of the needed *C*-terminal (of the permuted sequence) thiolester, and then use of native chemical ligation *in solution* to effect circularization. This is the largest peptide thiolester prepared to date by Fmoc chemistry, and among the largest ligation/cyclizations yet achieved − moreover, without any evident conformational driving force. There is every reason to believe that by varying the locus of side-chain anchoring (Asp in this work, but routes involving Asn, Glu, Gln, Lys, or Tyr are readily envisaged), the method can be generalized to allow synthesis of a wide range of circular proteins.

Acknowledgments

We thank Prof. C. K. Woodward, Dr. Y. M. Angell, and Mr. L. R. Masterson for rewarding collaborations and critical readings of the manuscript, and gratefully acknowledge financial support from NIH (GM 42722 and 51628).

[1] M. Goodman (Ed.), *Methods of Organic Chemistry*: *Synthesis of Peptides and Peptidomimetics (Houben-Weyl)*, Georg Thieme Verlag, Stuttgart, New York, **2002** and **2003**, vols. E 22a−e.

[2] Reviews: [2a] H. Kessler, *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 512−523. [2b] J. Rizo, L. M. Gierasch, *Annu. Rev. Biochem.* **1992**, *61*, 387−418. [2c] S. A. Kates, N. A. Solé, F. Albericio, G. Barany, in *Peptides: Design, Synthesis, and Biological Activity* (Eds.: C. Basava, G. M. Anantharamaiah), Birkhauser, Berlin, **1994**, pp. 39−58. [2d] J. N. Lambert, J. P. Mitchell, K. D. Roberts, *J. Chem. Soc., Perkin Trans. 1* **2001**, *5*, 471−484. [2e] J. S. Davies, *J. Pept. Sci.* **2003**, *9*, 471−501.

[3] [3a] O. Saether, D. J. Craik, I. D. Campbell, K. Sletten, J. Juul, D. G. Norman, *Biochemistry* **1995**, *34*, 4147−4158. [3b] Y.-Q. Tang, J. Yuan, G. Ösapay, K. Ösapay, D. Tran, C. J. Miller, A.

J. Ouellette, M. E. Selsted, *Science* **1999**, *286*, 498−502. [3c] D. J. Craik, N. L. Daly, T. Bond, C. Waine, *J. Mol. Biol.* **1999**, *294*, 1327−1336. [3d] Review: M. Trabi, D. J. Craik, *Trends Biochem. Sci.* **2002**, *27*, 132−138.

[4] M. Martinez-Bueno, M. Maqueda, A. Galvez, B. Samyn, J. Van Beeumen, J. Coyette, E. Valdivia, *J. Bacteriol.* **1994**, *176*, 6334−6339.

[5] Beyond the extensions/generalizations of traditional synthetic approaches directed at cyclic peptides, two ingenious types of approaches to obtain circular proteins merit mention (see ref.[5a,b]). [5a] Action of the engineered enzyme subtligase on chemically synthesized precursors (to date, up to 21 residues), as described by D. Y. Jackson, J. P. Burnier, J. A. Wells, *J. Am. Chem. Soc.* **1995**, *117*, 819−820. [5b] Intein-based methods in which the cyclization substrate is assembled entirely by molecular biology techniques, as reviewed by F. B. Perler, E. Adam, *Curr. Opin. Biotech.* **2000**, *11*, 377−383.

[6] [6a] D. P. Goldenberg, T. E. Creighton, *J. Mol. Biol.* **1983**, *165*, 407−413. [6b] D. P. Goldenberg, T. E. Creighton, *J. Mol. Biol.* **1984**, *179*, 527−545. [6c] Review: D. P. Goldenberg, *J. Cell. Biochem.* **1985**, *29*, 321−335.

[7] In contrast to the semisynthetic Goldenberg−Creighton route, a completely synthetic route that featured as the final step an intramolecular ligation was used to obtain a circularized analogue of BPTI that was amenable to X-ray crystallographic structural analysis. See: I. Botos, Z. Wu, W. Lu, A. Wlodawer, *FEBS Letters* **2001**, *509*, 90−94.

[8] For representative further examples of designed cyclic proteins (including design rationales), see: [8a] J. A. Camarero, J. Pavel, T. W. Muir, *Angew. Chem. Int. Ed.* **1998**, *37*, 347−349. [8b] J. A. Camarero, T. W. Muir, *J. Am. Chem. Soc.* **1999**, *121*, 5597−5598. [8c] H. Iwai, A. Plückthun, *FEBS Letters* **1999**, *459*, 166−172. [8d] Q. Yu, R. I. Lehrer, J. P. Tam, *J. Biol. Chem.* **2000**, *275*, 3943−3949. [8e] J. A. Camarero, D. Fushman, S. Sato, I. Giriat, D. Cowburn, D. P. Raleigh, T. W. Muir, *J. Mol. Biol.* **2001**, *308*, 1045−1062. [8f] H. Iwai, A. Lingel, A. Pluckthun, *J. Biol. Chem.* **2001**, *276*, 16548−16554. [8g] S. Deechongkit, J. W. Kelly, *J. Am. Chem. Soc.* **2002**, *124*, 4980−4986. [8h] Review: H.-X. Zhou, *J. Mol. Biol.* **2003**, *332*, 257−264.

[9] Key papers stressing total chemical synthesis, semisynthesis, and/or biophysical characterization include: [9a] M. Ferrer, C. Woodward, G. Barany, *Int. J. Pept. Prot. Res.* **1992**, *40*, 194−207. [9b] M. Ferrer, G. Barany, C. Woodward, *Nat. Struct. Biol.* **1995**, *2*, 211−217. [9c] E. Barbar, G. Barany, C. Woodward, *Biochemistry* **1995**, *34*, 11423−11434. [9d] H. Pan, E. Barbar, G. Barany, C. Woodward, *Biochemistry* **1995**, *34*, 13974−13981. [9e] G. Barany, C. M. Gross, M. Ferrer, E. Barbar, H. Pan, C. Woodward, in *Techniques in Protein Chemistry VII* (Ed.: D. Marshak), Academic Press, San Diego, **1996**, pp. 503-514.

[10] Most of this literature, from the laboratories of T. E. Creighton and of P. S. Kim, was cited and thoroughly discussed in ref.[9b] Of further note, a variety of heterologous pairwise substitutions of [30−51] have been prepared and characterized, offering further details beyond the usual strategy of using the same residue to replace both paired Cys. See: Y. Liu, K. Breslauer, S. Anderson, *Biochemistry* **1997**, *36*, 5323−5335.

[11] [11a] K. J. Jensen, J. Alsina, M. F. Songster, J. Vágner, F. Albericio, G. Barany, *J. Am. Chem. Soc.* **1998**, *120*, 5441−5452. [11b] J. Alsina, T. S. Yokum, F. Albericio, G. Barany, *J. Org. Chem.* **1999**, *64*, 8761−8769. [11c] Review: J. Alsina, K. J. Jensen, F. Albericio, G. Barany, *Chem. Eur. J.* **1999**, *5*, 2787−2795.

[12] [12a] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776−779. [12b] Review: J. P. Tam, J. Xu, K. D. Eom, *Biopolymers (Peptide Science)* **2001**, *60*, 194−205. [12c] Review: S. B. H. Kent, *J. Pept. Sci.* **2003**, *9*, 574−593.

[13] Selected examples include: [13a] J. A. Camarero, T. W. Muir, *Chem. Commun.* **1997**, 1369−1370. [13b] J. A. Camarero, G. J. Cotton, A. Adeva, T. W. Muir, *J. Peptide. Res.* **1998**, *51*, 303−316. [13c] L. Zhang, J. P. Tam, *J. Am. Chem. Soc.* **1997**, *119*, 2363−2370. [13d] Y. Shao, W. Lu, S. B. H. Kent, *Tetrahedron Lett.* **1998**, *39*, 3911−3914. [13e] J. P. Tam, Y.-A. Lu, Q. Yu, *J. Am. Chem. Soc.* **1999**, *121*, 4316−4324. [13f] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526−533, and references cited in all of these publications.

[14] Our recent report demonstrates *on-resin* native chemical ligation/cyclization compatible with Fmoc chemistry, and summarizes relevant literature for Fmoc solid-phase synthesis of peptide thiolesters: J. Tulla-Puche, G. Barany, *J. Org. Chem.* **2004**, *69*, 4104−4107.

[15] [15a] A. Trzeciak, W. Bannwarth, *Tetrahedron Lett.* **1992**, *33*, 4557−4560. [15b] S. A. Kates, N. A. Solé, C. R. Johnson, D. Hudson, G. Barany, F. Albericio, *Tetrahedron Lett.* **1993**, *34*, 1549−1552.

[16] N. Thieriet, J. Alsina, E. Giralt, F. Guibé, F. Albericio, *Tetrahedron Lett.* **1997**, *38*, 7275−7278.

[17] The aforementioned peptide had the correct mass by high resolution ESMS (calcd. 2075.1204, found 2074.9854). Moreover, MS/MS gave nested sequencing ions, starting from the *C*-terminus from Lys through Pro-Pro-Tyr-Thr-Gly-Pro-Abu-Lys, and starting from the *N*-terminus from Thr-Abu through Thr-Abu-Gly-Gly-Ala-Arg-Pro-Asp-Phe-Cys-Leu-Glu-Pro, while Edman sequential degradation correctly identified amino acids starting from Thr up to and including Leu, but not Abu (for which there was no standard).

Received July 20, 2004

