Total Synthesis of the Bovine Pancreatic Trypsin Inhibitor (BPTI) and the Protein Diastereomer [Gly37D-Ala]BPTI using Boc Chemistry Solid Phase Peptide Synthesis

Alexander J. Donovan*‡, Joanna Dowle*§, Yanwu Yang[≠] , Michael A. Weiss[≠] , Stephen B.H. Kent

Department of Chemistry and Department of Biochemistry & Molecular Biology, University of Chicago, 929 E. 57th St., Chicago, Illinois 60637

[≠] Department of Biochemistry & Molecular Biology
635 Barnhill Dr
Indiana University School of Medicine
Indianapolis, IN 46202

‡Current address:
USG Corporate Innovation Center
700 N. U.S. Highway 45
Libertyville, IL 60048

§Current address:
Ferrier Institute, Victoria University of Wellington,
PO Box 33436,
Petone 5046, New Zealand

[*These two authors contributed equally to this work.]

Dedication

This paper is dedicated to Louis Carpino without whose innovations in amino group protection modern synthetic protein chemistry would not be possible.

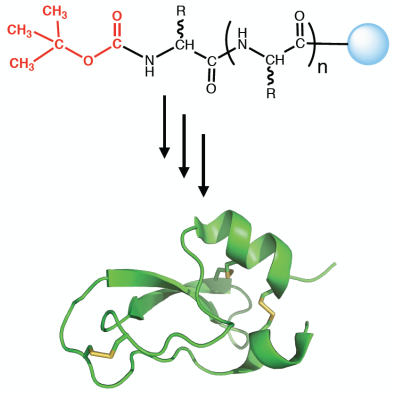
This is the author's manuscript of the article published in final edited form as:

Donovan, A. J., Dowle, J., Yang, Y., Weiss, M. A., & Kent, S. B. H. (2020). Total synthesis of bovine pancreatic trypsin inhibitor and the protein diastereomer [Gly37D-Ala]BPTI using Boc chemistry solid phase peptide synthesis. Peptide Science, e24166. https://doi.org/10.1002/pep2.24166

Abstract

Bovine pancreatic trypsin inhibitor (BPTI) is a well-studied model for investigation of protein folding and stability. Here we report the synthesis and characterization of wild-type BPTI and a diastereomeric protein analogue [Gly37D-Ala]BPTI. Each 58-residue polypeptide chain was made by native chemical ligation of two peptide segments, BPTI[1-29]-$^{\alpha}$thioester and the appropriate version of the [Cys^{30}-58]BPTI segment. Boc chemistry *in situ* neutralization solid phase synthesis was used to prepare the peptide segment reactants. The resulting full-length polypeptide chains were folded in a cysteine/cystine redox buffer to give synthetic protein molecules containing three disulfide bonds. Synthetic proteins were characterized by analytical LCMS and by natural-abundance ^{1}H-^{15}N HSQC NMR fingerprinting. These results illustrate the power of Boc chemistry peptide synthesis and its utility for the total chemical synthesis of protein molecules.

Graphical Contents



Introduction

Bovine pancreatic trypsin inhibitor (BPTI) is the prototypical Kunitz-type serine protease inhibitor.[1] BPTI has a 58 amino acid residue polypeptide chain in an alpha/beta fold constrained by three disulfide bonds. Dynamic features of the BPTI protein molecule have been widely studied, because it is a good model for NMR studies of tyrosine ring flipping and sub-global NH exchange in a globular protein. A [Gly37Ala]BPTI mutant has been described that destabilized the protein by ~5 kcal/mol.[2] Although the average structure of the BPTI protein was retained, the addition of just one side chain methyl group in an Ala residue at position 37 of the BPTI polypeptide chain caused dramatic perturbation in the dynamics, leading to a large decrease in global stability, and a significant increase in local flexibility of the loop domain. This was ascribed to disruption of the NH-aromatic-NH network of interactions (between the Gly37 backbone HN, the Tyr35 aromatic ring, and the side chain HN of Asn44), which is energetically costly and may involve other cooperatively linked interactions.[2] To further probe the role of the Gly37 residue in the NH-aromatic-NH network of interactions, we undertook the total chemical synthesis of a BPTI analogue with D-Ala substituted at position 37. The BPTI polypeptide chain backbone phi, psi angles for Gly37 [Φ +103°, Ψ -3°] suggest that a D-Ala residue could be accommodated at that position.[3] There are a number of published examples in which total or semi-synthesis of a protein molecule has been used to substitute a D-Ala residue for a Gly residue that has backbone phi, psi angles typical of a D-amino acid residue.[4-7]

Total chemical synthesis of proteins presents unique advantages: it provides total control over the covalent structure of the protein molecule and enables introduction of non-coded moieties such as the D-amino acids.[8] Chemical ligation methods, which are based on the chemoselective reaction between two mutually reactive groups in order to unambiguously condense unprotected peptide segments,[9] have dramatically increased the ease with which homogeneous preparations of protein molecules can be made by total synthesis. The most effective condensation method, native chemical ligation (NCL),[10] involves reaction between a peptide-$^{\alpha}$thioester and the amino-terminal cysteine of a second Cys-peptide segment: an initial thioester-linked intermediate undergoes spontaneous intramolecular

rearrangement to give a native peptide bond at the Xaa-Cys ligation site. Because the initial transthioesterification is reversible in the presence of added aromatic thiol under the reaction conditions used, internal Cys residues can be present in the reacting peptide segments.[8]

Stepwise Boc chemistry solid phase peptide synthesis (SPPS) can be used to make high purity, homogeneous peptides of up to 40 amino acids in good yield and defined covalent structure.[11] In particular, *in situ* neutralization Boc chemistry SPPS,[12] which was originally introduced in order to minimize so-called 'difficult' sequences in chain assembly, is ideally suited to the preparation of peptide thioesters for chemical protein synthesis by native chemical ligation.[13] Making peptide thioesters by conventional Boc chemistry SPPS is problematic, because a pre-existing thioester moiety is susceptible to attack by the neutralized peptide chain during every neutralization step.

In the work reported here, we describe the total synthesis of model protein BPTI and its disatereomeric analogue [Gly37D-Ala]BPTI, using peptides prepared by Boc chemistry SPPS. Full-length synthetic polypeptide chains were assembled by native chemical ligation, and folded with formation of three disulfides to give the two synthetic BPTI protein molecules; these were characterized by analytical LCMS and ^{1}H-^{15}N HSQC NMR spectroscopy.[14]

Materials and Methods

Boc-L-amino acids, Boc-Gly, Boc-D-Ala, S-trityl-β-mercaptopropionic acid, and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were obtained from the Peptide Institute Inc. (Osaka, Japan). Boc-Ala-OCH_2-Pam-resin and N,N-diisopropylethylamine (DIEA) were obtained from Applied Biosystems (Foster City, CA). N,N-dimethylformamide (DMF), dichloromethane (DCM), triisopropylsilane (TIPS), and acetonitrile were obtained from Fisher (Chicago, IL). Trifluoroacetic acid (TFA) was purchased from Halocarbon (New Jersey). 4-Mercaptophenylacetic acid (MPAA) and p-cresol were from Sigma-Aldrich (St. Louis, MO).

Analytical reversed-phase HPLC with online mass spectrometry (LCMS) was performed on an Agilent 1100 Series instrument equipped with an MSD ion trap, using a Vydac C18 column (2.1 mm X 150 mm). Chromatographic separations were performed using a linear gradient (5-65%) of buffer B in buffer A over 15 minutes at a flow rate of 0.5 mL/min. Buffer A = 0.1% TFA in water, buffer B = 0.08% TFA in acetonitrile. MS data were acquired across the entirety of each UV absorbing peak for which a mass is reported. The masses were derived from the m/z values of the different protonation states of a particular molecular species.

Peptides were made manually on Boc-Ala-OCH_2-Pam-(copoly)styrene-1%meta divinylbenzene resin using i*n situ* neutralization Boc chemistry solid phase peptide synthesis,[12] with HBTU activation. Side chain protection for Boc-amino acids was as follows: Arg(Tos), Asn(Xan), Asp(OcHex), Cys(4-CH_3-Bzl), Glu(OcHex), Lys(2-Cl-Z), Ser(Bzl), Tyr(Br-Z). All amino acids were of the L configuration, except for the Boc-D-Ala used in preparation of the diastereomer. Synthesis of BPTI(1-29)$^{\alpha}$thioester was performed on a 0.4 mmol scale. S-Trityl-β-mercaptopropionic acid was coupled to deprotected and neutralized Ala-OCH_2-Pam-Resin for 1 hr. The trityl group was removed by treatment with 2.5% TIPS/2.5% H_2O/95% TFA.[13] Boc-Leu was coupled to the β-mercaptopropionyl-Ala-OCH_2-Pam-Resin for 20 min. During SPPS chain elongations, when the N-terminal amino acid was Boc-Gln, DCM was used to wash before and after TFA deprotection.[12] Synthesis of the peptide (Cys^{30}-58)BPTI and the corresponding analogue (Cys^{30}-58) [Gly37D-Ala]BPTI peptide, was on a 0.2 mmol scale. After assembly of each peptide chain was complete, the peptide was deprotected and cleaved from the resin support using anhydrous HF/p-cresol (90:10, v/v) for 1 h at 0 ^{o}C. After evaporation of the HF under reduced pressure, the crude peptide product was precipitated and triturated with chilled diethyl ether, and then dissolved in 50% acetonitrile in H_2O containing 0.1% TFA. An aliquot from the solution containing the crude peptide product was analyzed by LCMS before lyophilization.

Native chemical ligations were performed in parallel in 3 mL aqueous solution containing 6 M Gn.HCl, 0.2 M Na_2HPO_4, 50 mM MPAA, 20 mM TCEP-HCl, pH 6.7, at a concentration of ~2 mg/mL (~0.6 mM) of each reacting peptide. Progress of the reactions was analyzed by

analytical LCMS. Ligations were essentially complete at 6 h, but for reasons of convenience, the reactions were allowed to continue overnight. After each native chemical ligation reaction was complete, solid-phase extraction was performed to remove low molecular weight compounds using Alltech C18-LP cartridges. The cartridges were first equilibrated with 10 mL 100% methanol, 10 mL 50%/50% A/B, and 10 mL 100% A. The reaction mixtures were slowly loaded onto the cartridges dropwise and subsequently washed with 5 mL 5% B in A. The crude peptide was eluted using 50%/50% A/B and the eluate analyzed by LCMS.

The wild-type BPTI polypeptide chain and the analogue [Gly37D-Ala]BPTI polypeptide chain were separately folded, with concomitant formation of disulfide bonds, at a polypeptide concentration of ~0.3 mg/mL in aqueous solution containing 2 M Gn.HCl, 6.13 mM L-cysteine, 0.8 mM L-cystine, 67 mM Tris, pH 7.7-7.9. During the folding reactions, no stirring was performed. Progress of the reactions was monitored by analytical LCMS. Folding was complete within 6-7 h. Folded BPTI and folded analogue [Gly37D-Ala]BPTI were purified by preparative HPLC performed on a Vydac C18 column (10 mm X 250 mm) at a flow rate of 5 mL/min, with a gradient of 5-25% B over 20 minutes, then 25-50% B over 25 minutes. Fractions were pooled based on LCMS analysis. Purified folded wild-type BPTI and folded [Gly37D-Ala]BPTI proteins were characterized by analytical LCMS.

^{1}H-^{15}N HSQC spectra of wild-type BPTI and [Gly37D-Ala]BPTI were acquired at natural abundance (proton frequency 700 MHz) in H_2O/D_2O (10% D_2O) at pH 4.6 and 25 ^{o}C using a Bruker Avance NMR spectrometer equipped with [x,y,z]-gradient triple-resonance cryoprobe. Protein concentrations were in each case ~1 mM.

Results and Discussion

The amino acid sequence of the BPTI polypeptide chain is shown in Figure 1.[15] The BPTI protein molecule contains three disulfide bonds, connecting residues Cys[5]-Cys[55], Cys[14]-Cys[38], and Cys[30]-Cys[51].

R·[1]PDFCLEPPY T[11]GPCKARIIR Y[21]FYNAKAG**LC** Q[31]TFVYGGCRA
K[41]RNNFKSAED C[51]MRTCGGA[58]

Figure 1. Amino acid sequence of BPTI.

Synthetic Scheme. The 58 amino acid residue polypeptide chain has a Leu29-Cys30 ligation site at the exact center of its sequence, ideally placed for synthesis by native chemical ligation.[8] A scheme for the convergent synthesis of BPTI from two 29 amino acid residue peptides is shown in Figure 2.

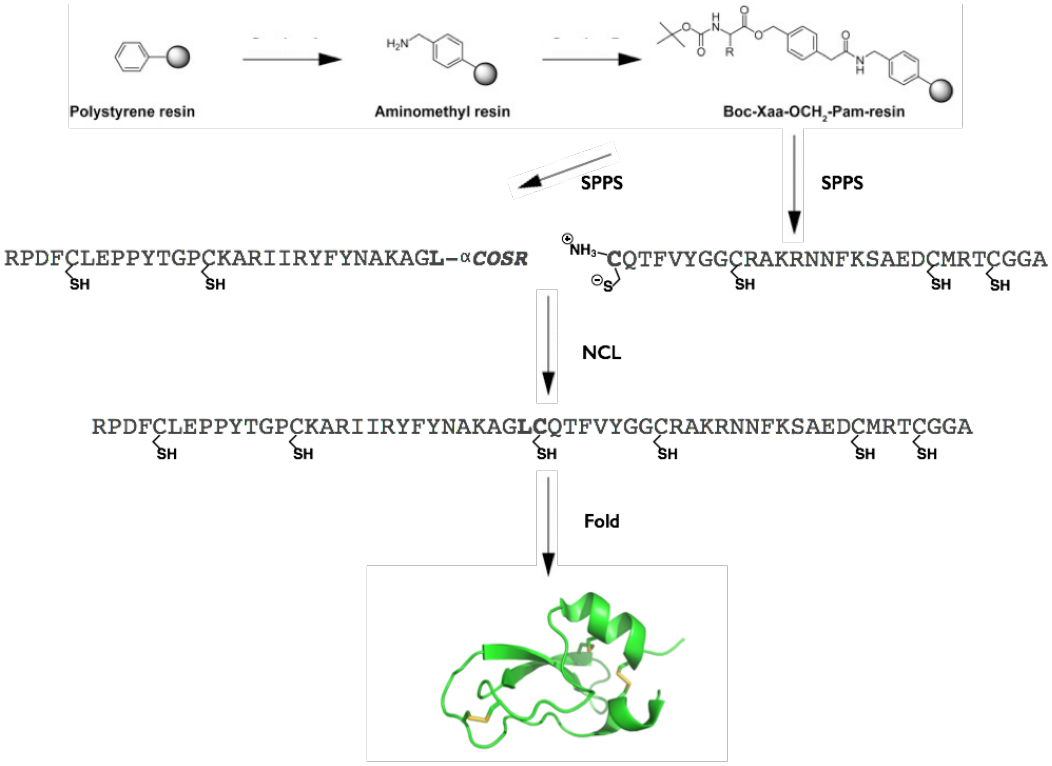


Figure 2. Convergent total synthesis of BPTI. Peptide segment reactants were made by Boc chemistry SPPS and covalently condensed by native chemical ligation (NCL).[8] The resulting 58 amino acid residue full-length polypeptide chain was folded with concomitant formation of disulfides under standard redox conditions. (Image of BPTI generated from PDB code: 1BPI)

Peptide Synthesis. Peptide segments were made by *in situ* neutralization Boc chemistry stepwise SPPS.[12] For synthesis of the thioester segment BPTI(1-29)-*COSR*, where R= SCH_2CH_2-Ala, the C-terminal Boc-Leu residue of the target sequence was coupled to $HSCH_2CH_2$-Ala-OCH_2Pam-Resin. After the peptide chain had been assembled in stepwise fashion, the resin-bound peptide was simultaneously deprotected and cleaved from the resin by treatment with HF, to give the peptide thioester BPTI(1-30)-*$COSCH_2CH_2$-Ala-OH*. The wild-type BPTI sequence Cys-peptide segment Cys^{30}-Ala^{58} was synthesized directly on Boc-Ala-OCH_2Pam-Resin, and the analogue segment was made in the same fashion, substituting D-Ala for Gly at residue position 37. Crude peptides were purified by preparative HPLC, and characterized by LCMS. Examples of the synthetic peptide segments are shown in **Figure 3.**

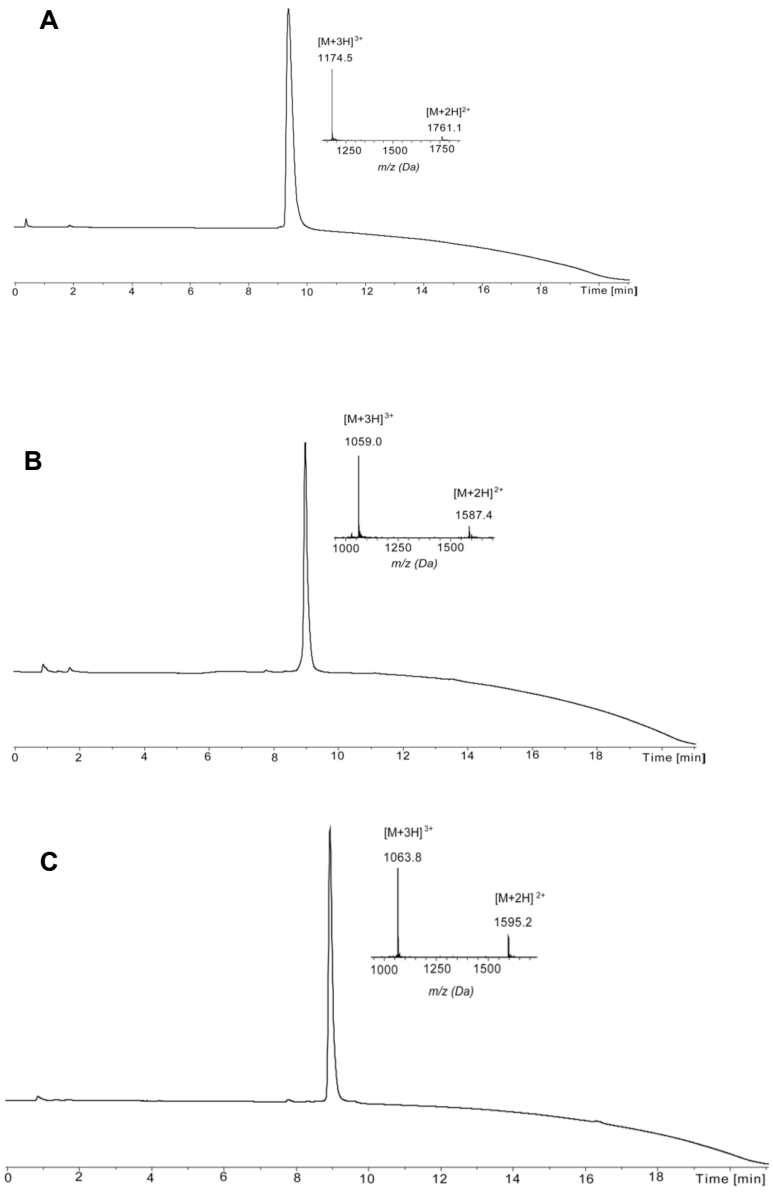


Figure 3. LCMS characterization of purified peptide segments made by Boc chemistry SPPS. **Panel A.** Purified BPTI(1-29)-thioester. Observed mass 3520.3 ± 0.2 Da [Calculated mass: 3520.0 Da (average isotope composition)] **Panel B.** Purified (Cys^{30}-58)BPTI. Observed mass 3174.4 ± 0.4 Da [Calculated mass: 3174.6 Da (average isotope composition)] **Panel C.** Purified (Cys^{30}-58) [Gly37D-Ala]BPTI. Observed mass 3188.4 ± 0.2 Da [Calculated mass: 3188.6 Da (average isotope composition)]

Chemical Ligation. To prepare the full-length synthetic polypeptide chains, the peptide-thioester and Cys-peptide segments were condensed by native chemical ligation in aqueous 6 M Gn.HCl in phosphate buffer at pH 6.7 in the presence of 50 mM MPAA as an aromatic thiol catalyst. The NCL reactions were essentially complete after six hours, but for convenience were allowed to proceed for 24 hours. Data for synthesis of the wild-type BPTI and [Gly37D-Ala]BPTI polypeptide chains are shown in **Figures 4 & 5**, respectively.

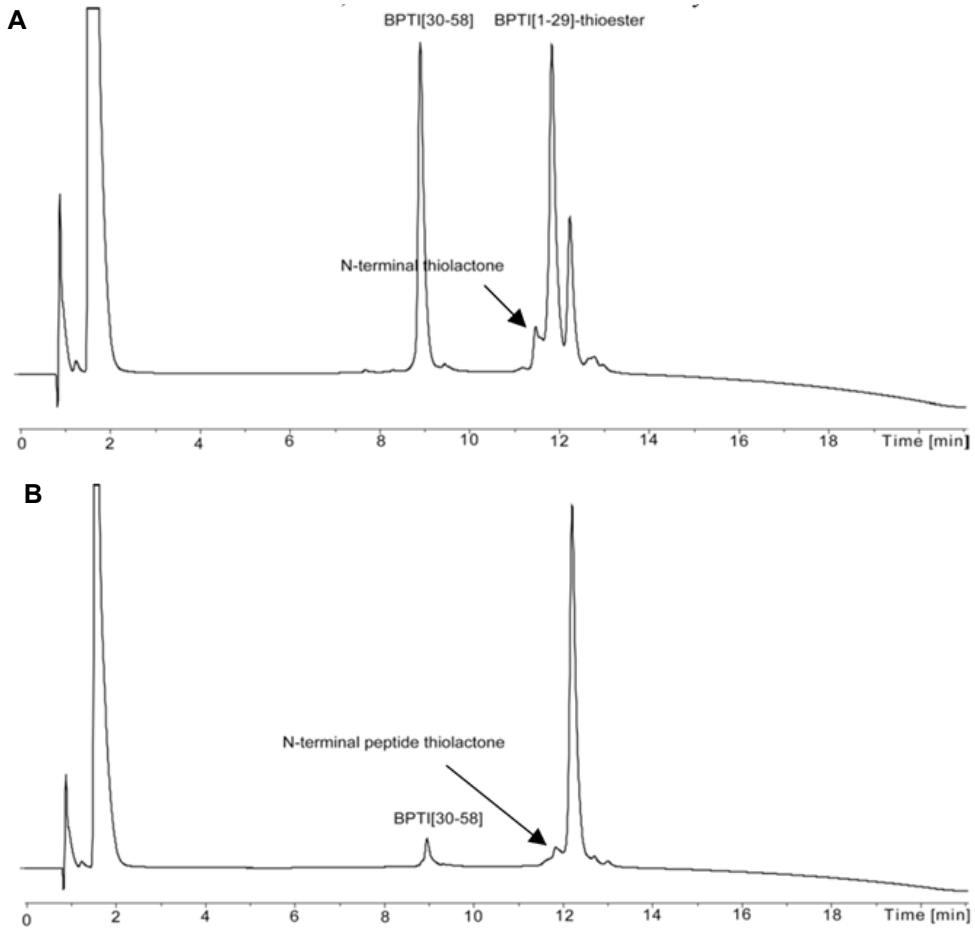


Figure 4. Synthesis of the wild-type BPTI(1-58) polypeptide chain. **Panels A & B**: Analytical LCMS monitoring of the NCL reaction at T=0; t= 24h, respectively. Observed mass of the ligation product: 6516.0 ± 0.5 Da [Calculated masses: 6513.1 Da (monoisotopic); 6517.5 Da (average isotope composition)] MS data were acquired across the entirety of each UV absorbing peak for which a mass is reported.

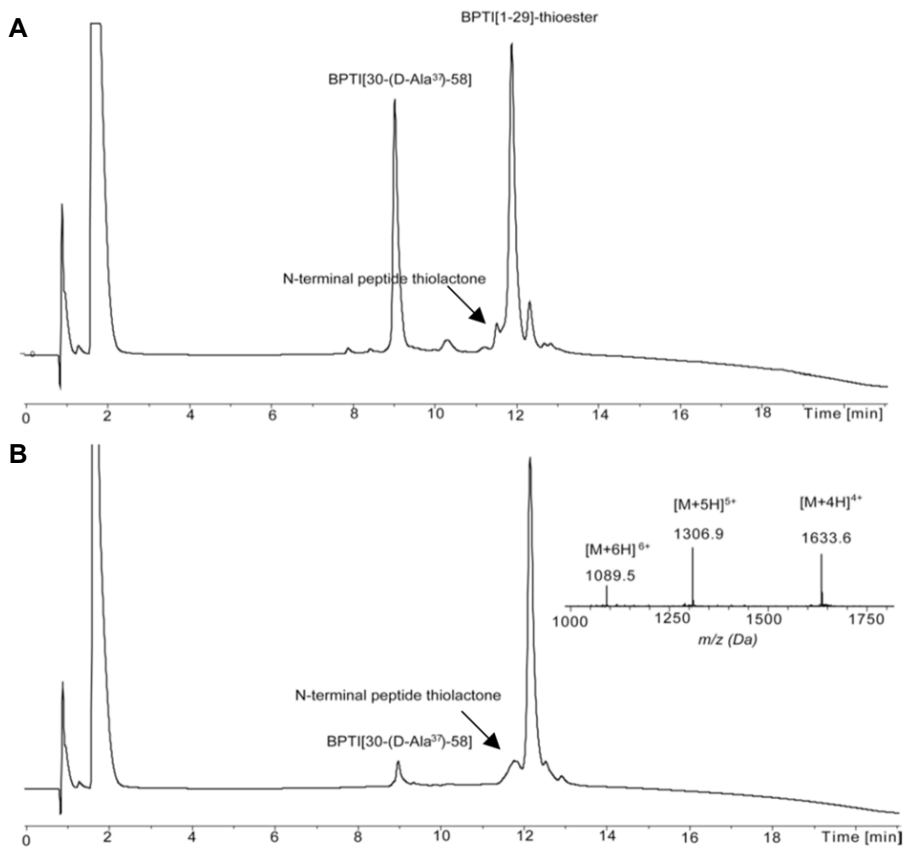


Figure 5. Synthesis of the [Gly37D-Ala]BPTI(1-58) polypeptide chain. **Panels A & B**: Analytical LCMS monitoring of the NCL reaction at T=0; t= 24h, respectively. The full-length polypeptide chain [Gly37D-Ala]BPTI(1-58) had observed mass: 6530.0 ± 0.5Da [Calculated masses: 6527.1 Da (monoisotopic); 6531.5 Da (average isotope composition)]

Folding. Synthetic protein molecules were obtained by separately folding the two full-length polypeptide chains, with concomitant formation of disulfides, in a redox buffer containing a moderate concentration of guanidine.HCl as a solubilizing agent in order to keep misfolded forms of the polypeptide chains in solution, according to the principles developed by Rudolph and coworkers for refolding proteins expressed as inclusion bodies in recDNA-engineered *E. coli.*[16] Here we used a redox buffer consisting of 6.13 mM cysteine and 0.8 mM cystine in 67 mM Tris at pH 7.8, with 2 M Gn.HCl. The synthetic protein molecules were purified by reversed phase HPLC.

Characterization. The purified synthetic proteins were characterized by analytical LCMS. Data for the folded wild-type BPTI synthetic protein is shown in Figure 6.

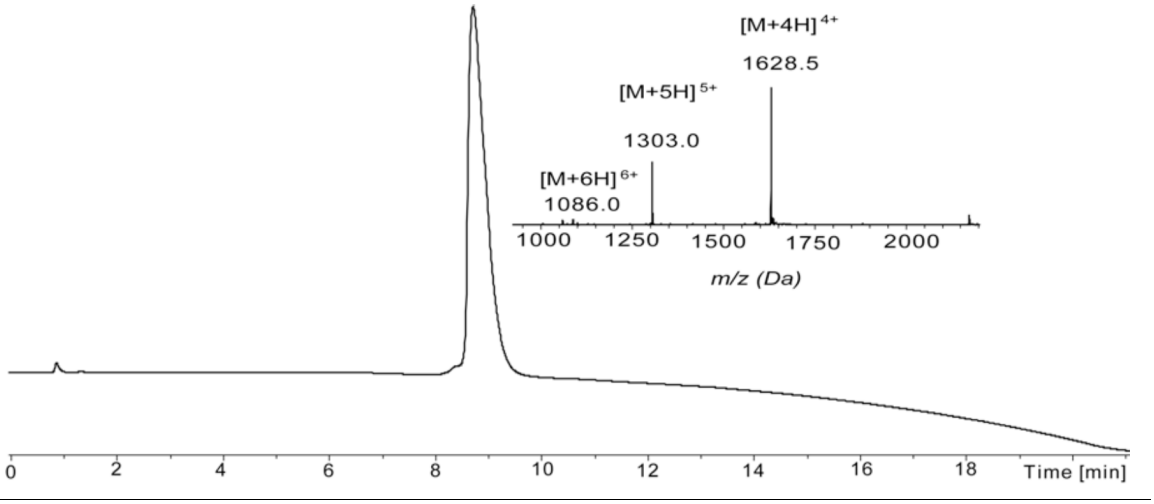


Figure 6. Analytical LCMS of purified wild-type BPTI synthetic protein. The folded synthetic protein had observed mass 6510.0 ± 0.5 Da [Calculated masses: 6507.1 Da (monoisotopic); 6511.5 Da (average isotope composition)]

The observed mass of the folded wild-type BPTI protein was 6510.0 ± 0.5 Da compared with 6516.0 ± 0.5 Da for the linear polypeptide chain, a mass difference of 6 Da, consistent with the formation of three disulfide bonds in the folded protein molecule. A similar reduction in mass of 6.5 Da was observed, from 6530.0 ± 0.5 Da for the linear polypeptide chain to 6523.5 ± 0.5 Da, for the folded [Gly37D-Ala]BPTI analogue protein, which was thus also shown to contain three disulfides.

Synthetic wild-type BPTI and [Gly37D-Ala]BPTI proteins were characterized by ^{1}H-^{15}N natural-abundance HSQC NMR 'fingerprinting' (Figure 7).[13'] Both proteins exhibited marked dispersion of 2D NMR cross-peaks as is characteristic of well-folded protein molecules that have a single, well-defined tertiary structure. The synthetic proteins exhibited similar NMR fingerprint spectra, with only ~7 residues showing significantly perturbed chemical shifts.

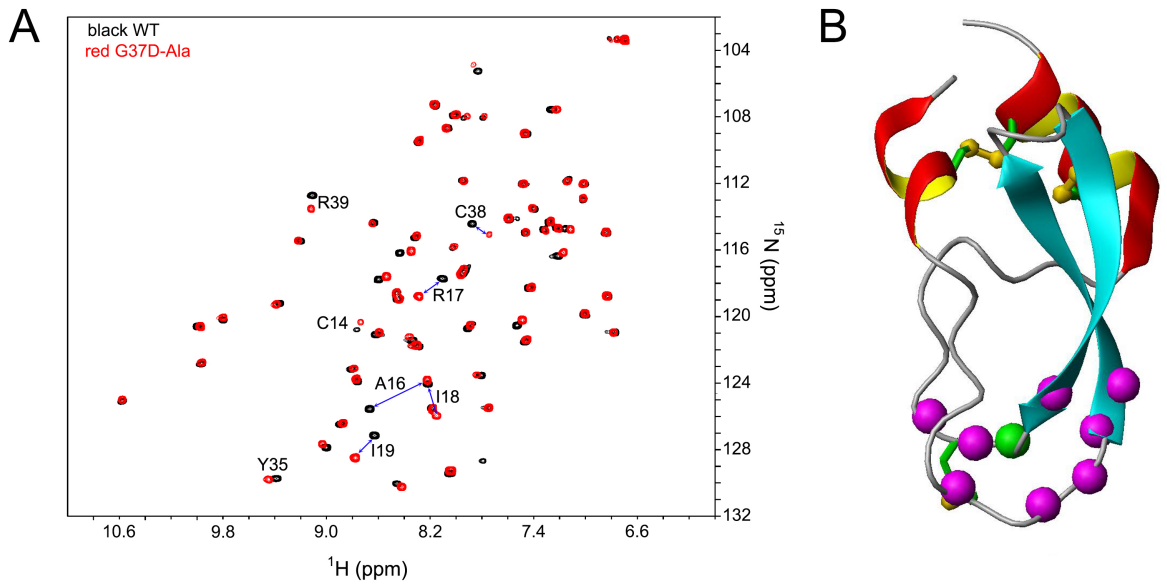


Figure 7. ^{1}H-^{15}N HSQC fingerprints of wild-type BPTI and [Gly37D-Ala]BPTI. (A) Overlay of ^{1}H-^{15}N HSQC spectra of wild-type BPTI (Gly37; black cross-peaks) and the [Gly37D-Ala]BPTI analogue (red cross-peaks). (B) Ribbon presentation of BPTI crystal structure (PDB entry 1K6U). The green sphere indicates position of the Gly37 main-chain nitrogen, and magenta spheres indicate main-chain nitrogen atoms of residues that exhibit perturbed chemical shifts in the [Gly37D-Ala]BPTI analogue's HSQC spectrum. Disulfide bridges are highlighted in gold, with smaller spheres indicating the sulfur atoms.

Conclusions

Wild-type BPTI and a diastereomeric analogue BPTI protein substituted with D-Ala at position 37 were synthesized by Boc chemistry SPPS combined with native chemical ligation. The D-Ala substituted BPTI variant [G37D-Ala]BPTI should provide additional insights into the effect of restrictions of main chain dihedral angles on local and larger-distance conformational fluctuations in the BPTI protein molecule.[2] Studies of the effects of the 'D-Gly^{37}'→D-Ala^{37} substitution on folding efficiency, thermodynamic stability and dynamic features will be reported separately. Of particular interest will be potential stereospecific damping of Tyr^{35} side-chain 'flipping' and subglobal native-state amide proton exchange in D_2O solution relative to the markedly enhanced flexibility described in the L-Ala^{37} analogue.[2]

The utility of Boc chemistry SPPS is vividly illustrated by the work reported here. In 1957 Carpino introduced the acid labile tert-butyloxycarbonyl (Boc) group for amine protection.[17] It was swiftly adopted for use in peptide chemistry,[18] and N$^{\alpha}$Boc-protected amino acid chemistry dominated the burgeoning field of peptide synthesis for more than four decades, and is still used in the present day. In 1972 Carpino introduced the base labile fluorenylmethyloxycaronyl (Fmoc) group for amine protection.[19] Embodied in Fmoc chemistry SPPS,[20] it is now almost universally used for the chemical synthesis of peptides and, in combination with modern chemical ligation methods, is becoming widely used for the total chemical synthesis of protein molecules.[21-23]

Acknowledgements

NMR studies were supported in part by NIH R01 DK040949 (MAW). We thank Q.-X. Hua and N.B. Phillips for discussions.

Authors' Contributions

Alexander Donovan: synthesized peptide segments and protein molecules.

Joanna Dowle: synthesized peptide segments and protein molecules, and assisted in preparation of the manuscript.

Yanwu Yang: carried out the HSQC NMR studies on the synthetic proteins.

Michael Weiss: suggested the synthesis of BPTI and the diastereomeric BPTI analogue, and supervised the HSQC NMR studies.

Stephen Kent: designed and supervised the synthesis of BPTI and the diastereomeric BPTI analogue, and wrote the manuscript with input from other authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest

References

1. P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. De Cristofaro, E. Menegatti, Curr. Protein & Peptide Sci. **2003**, *4,* 231.
2. J. L. Battiste, R. Li, C. Woodward, Biochemistry **2002**, *41*, 2237.
3. S. Parkin, B. Rupp, H. Hope, Acta Crystallogr. Sect. D **1996**, *52*, 18.
4. D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, S. B. Kent, Angew. Chem. Int. Ed. Eng. **2005**, *44*, 3852.
5. V. Yu. Torbeev, H. Raghuraman, D. Hamelberg, M. Tonelli, W. M. Westler, E. Perozo & S. B. H. Kent, Proc. Natl. Acad. Sci.USA **2011**, *108*, 20982.
6. F. I. Valiyaveetil, M. Leonetti, T. W. Muir, R. MacKinnon, Science **2006**, *314*, 1004.
7. Y. Sohma, B. L. Pentelute, J. Whittaker, Q.-x. Hua, L. J. Whittaker, M. A. Weiss, S. B. H. Kent, Angew. Chem. Int. Ed. Engl. **2008**, *47*, 1102.
8. S. B.H. Kent, Chem. Soc. Reviews **2009**, *38*, 338.
9. M. Schnölzer, S.B.H. Kent Science **1992**, *256*, 221.
10. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B.H. Kent, Science **1994**, *266*, 776.
11. S. B. H. Kent, Ann. Rev. Biochem. **1988**, *57*, 957.
12. M. Schnölzer, P. Alewood, D. Alewood, S. B. H. Kent, Int. J. Pept. Prot. Res. **1992**, *40*, 180. Republished: Int. J. Pept. Res. Therapeutics **2007**, *13*, 31.
13. T. M. Hackeng, J. H. Griffin, P. E. Dawson, Proc. Natl. Acad. Sci. U S A **1999**, *96*, 1006.
14. T. Rehm, R. Huber, T, A. Holak, Structure **2002**, *10*, 1613.
15. B. Kassell, M. Radicevic, M. J. Ansfield, M. Laskowski Sr., Biochem. Biophys. Res. Com. **1965**, *18*, 255.
16. C.E. De Bernardez, E. Schwarz, R. Rudolph, Methods Enzymol. **1999**, *309*, 217.
17. L.A. Carpino, J. Am. Chem. Soc. **1957**, *79*, 98.
18. F. C. McKay, N. F. Albertson, J. Am. Chem. Soc. **1957**, *79*, 4686.
19. L.A. Carpino, G.Y. Han, J. Org. Chem. **1972**, *37*, 3404.
20. W. C. Chan, P.D. White, Fmoc Solid Phase Peptide Synthesis, Oxford University Press, Oxford 2000.
21. G.-M. Fang, J.-X. Wang, L. Liu, Angew. Chem. Int. Ed. **2012**, *51*, 10347.
22. J. B. Blanco-Canosa, B. Nardone, F. Albericio, P. E. Dawson, J. Am. Chem. Soc. **2015**, *137*, 7197.

23. B. Dhayalan, K. Mandal, N. Rege, M. A. Weiss, S. H. Eitel, T. Meier, R. O. Schoenleber, S. B.H. Kent, Chem. Eur. J. **2017**, *23*, 1709.
