N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION

SUPPORTING INFORMATION

General. Fmoc-PAL-PEG-PS supports (initial loading 0.17–0.24 mmol/g) for peptide synthesis were obtained from Applied Biosystems (Foster City, CA). Protected Fmoc-amino acid derivatives and coupling reagents were from Applied Biosystems or Advanced Chemtech (Louisville, KY), S-Trt-mercaptopropionic acid was from Peptide International (Louisville, KY), Boc-Aoa-OH, Fmoc-Dpr(Boc-Aoa)-OH, Fmoc-Dpr(Boc-Ser(*t*Bu))-OH, and Fmoc-Asn(Tmob)-OH were from Novabiochem (San Diego, CA), and Fmoc-Gln(Tmob)-OH was from Bachem Bioscience (Philadelphia, PA). TFA, piperidine, and DIEA were obtained from Fisher (Pittsburgh, PA). GdnHCl and urea 'ultrapure reagents' were from Gibco BRL (Rockville, MD), and DMSO, sodium periodate, ethylene glycol, and $CH_3ONH_2 \cdot HCl$ (25-30 wt % solution in water) 'ACS grade reagents' were from Aldrich (Milwaukee, WI). Et_2O was freshly distilled from Na.

Amino acid analyses were performed on a Beckman 6300 analyzer with a sulfated polystyrene cation-exchange column (0.4 cm x 21 cm). Peptide or protein samples were hydrolyzed in 6 N aqueous HCl plus one drop of liquefied phenol (to prevent degradation of Tyr), for 90 min at 155 °C. Cys was not quantitated and is therefore not reported. The Ile values were consistently low, due to incomplete hydrolysis of the peptide bond between Ile^{18} and Ile^{19}.

Analytical HPLC of the core module monomers, capped monomers, and of the model tetrapeptide and its derivatives, was performed using a Vydac C_{18} reversed-phase column (5 μm, 300 Å; 0.46 x 25 cm), and analytical HPLC of the core module dimers was performed using a Waters YMC C_4 reversed-phase column (5 μm, 200 Å; 0.46 x 25 cm), all on a Beckman instrument, configured with two 112 pumps and a 165 variable wavelength detector. UV detection was at 220 nm.

Condition A: linear gradients of 0.1% TFA in CH_3CN and 0.1% aqueous TFA were run, at 1.2 mL/min flow rate, from 1:9 to 2:3 over 30 min, then to 4:1 over the next 10 min.

Condition B: linear gradients of 0.1% TFA in CH_3CN and 0.1% aqueous TFA were run, at 1.2 mL/min flow rate, from 1:4 to 7:13 over 40 min, then to 4:1 over the next 10 min.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry SUPPORTING INFORMATION

Condition C: linear gradients of 0.1% TFA in CH_3CN–H_2O (3:2) and 0.1% aqueous TFA were run, at 1.2 mL/min flow rate, from 0:1 to 1:1 over 20 min, then to 4:1 over the next 10 min.

Condition D: linear gradients of 0.1% TFA in CH_3CN–H_2O (3:2) and 0.1% aqueous TFA were run, at 1.2 mL/min flow rate, at 1:3 for 5 min, then to 1:1 over 30 min, then to 4:1 over the next 10 min.

Semipreparative HPLC was carried out on a Waters DeltaPrep 3000 using a Vydac semipreparative C_{18} reversed-phase column (10 μm, 300 Å; 1 x 25 cm). UV detection was at 220 nm.

Condition E: linear gradients of CH_3CN and 0.1% aqueous TFA were run, at 5 mL/min flow rate, from 0:1 for 10 min, then to 19:81 over the next 5 min, then to 27:73 over the next 120 min, and finally to 4:1 over the next 10 min.

Condition F: linear gradients of CH_3CN and 0.1% aqueous TFA were run, at 5 mL/min flow rate, from 4:21 for 5 min, then to 9:16 over the next 45 min, and finally to 4:1 over the next 10 min.

Condition G: linear gradients of CH_3CN and 0.1% aqueous TFA were run, at 5 mL/min flow rate, from 0:1 for 10 min, then to 17:83 over the next 5 min, then to 1:3 over the next 120 min, and finally to 4:1 over the next 10 min.

Condition H: linear gradients of CH_3CN–H_2O (3:2) and 0.1% aqueous TFA were run, at 4 mL/min flow rate, from 0:1 to 3:17 over 20 min, then to 4:1 over the next 10 min.

Condition I: linear gradients of CH_3CN and 0.1% aqueous TFA were run, at 5 mL/min flow rate, from 0:1 for 10 min, then to 11:39 over the next 5 min, then to 3:7 over the next 120 min, and finally to 4:1 over the next 10 min.

Size exclusion chromatography (SEC) was performed using two Superdex Peptide HR 10/30 columns assembled in series, in an FPLC System (Pharmacia).

Condition J: injections of 500 μL were eluted with 0.1% TFA in CH_3CN–H_2O (3:7), at 0.2 mL/min.

Condition K: injections of 25 μL were eluted with CH_3CN–10 mM aqueous NaCl (1:19) at pH 2, at 0.2 mL/min.

MALDI-TOF mass spectrometric data were collected with a Bruker Biflex III instrument, equipped

with a N_2-laser (337 nm, 3 nanosecond pulse length) and a microchannel plate (MCP) in the reflectron mode, positive polarity, with an accelerating potential of 19 kV. Each spectrum was the accumulation of 200 laser shots. External calibration was performed using human angiotensin II (monoisotopic mass 1046.5 $[M+H]^+$; Sigma) and adrenocorticotropic hormone (ACTH) fragment 18-39 (monoisotopic mass 2465.2 $[M+H]^+$; Sigma). The matrix used for samples and standards was α-cyano-4-hydroxycinnamic acid (4-HCCA; Hewlett-Packard, obtained as solution in methanol), diluted 1:1 with 0.1% TFA in CH_3CN–H_2O (1:1). ESMS data were collected using a ThermoFinnigan LCQ Classic ion trap mass spectrometer equipped with an ESI source in full scan mode from 50-2000 *m/z* with a spray voltage setting of 4.0 kV and a heated capillary temperature of 200 °C. Samples were infused at a flow rate of 5 μL/min in 0.2% acetic acid in CH_3CN–H_2O (1:1), using the syringe pump on the instrument. Deconvolution of multiply charged protein peaks was performed using ThermoFinnigan's BioMASS add-on tool in the Xcalibur software. Molecular masses were calculated with the GPMAW Version 4.0 program.

Peptide synthesis of core module monomers: $I_{R17K(U)}$, $I_{R17K(UG)}$, and $I_{R17Z(U)}$ (Supporting Information Scheme 1), and $II_{L29K(J)}$, $II_{L29K(JG)}$, $II_{L29Z(J)}$, and $II_{N24K(J)}$ (Supporting Information Scheme 2 and Figure 4).

$I_{R17K(U)}$. The synthesis and characterization of this molecule is described in the main text.

$I_{R17K(UG)}$. The only difference between the procedure for this molecule, and for $I_{R17K(U)}$ described in the main text Experimental, was the extra incorporation of Fmoc-Gly-OH preceding Boc-Aoa-OH. As summarized in the main text, three repeat PyAOP/HOAt/DIEA (8:8:16 equiv with respect to peptide-resin)-mediated couplings in DMF, 1 h each, were required for the Fmoc-Gly-OH incorporation to go to completion. Pure peptide was characterized by analytical HPLC (t_R 22.8 min; condition A), amino acid analysis (Asp, 2.04; Thr, 0.93; Glu, 1.05; Gly, 4.07; Ala, 1.98; Val, 0.92; Ile, 1.17; Leu, 0.99; Tyr, 2.46; Phe 1.91; Lys, 2.87; Arg, 1.07), and MALDI-TOF; calcd monoisotopic mass: 2954.46, found: 2955.1 $[M+H]^+$.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry SUPPORTING INFORMATION

$I_{R17Z(U)}$. In the linear core module sequence, Arg^{17} was replaced by Fmoc-Dpr(Boc-Aoa)-OH. Cleavage, disulfide bond formation, and purification followed as already described for $I_{R17K(U)}$. Pure peptide was characterized by analytical HPLC (t_R 22.5 min; condition A), amino acid analysis (Asp, 1.83; Thr, 0.91; Glu, 1.04; Gly, 3.14; Ala, 1.71; Val, 1.02; Ile, 1.18; Leu, 1.04; Tyr, 2.53; Phe 1.85; Lys, 1.83; Arg, 0.93), and MALDI-TOF; calcd monoisotopic mass: 2855.39, found: 2856.03 $[M+H]^+$.

$II_{L29K(J)}$. The synthesis and characterization of this molecule is described in the main text.

$II_{L29K(JG)}$. The only difference between the procedure for this molecule, and for $II_{L29K(J)}$ (Figure 4b) described in the main text Experimental, was the extra incorporation of Fmoc-Gly-OH preceding Fmoc-Ser(*t*Bu)-OH. As summarized in the main text, three repeat PyAOP/HOAt/DIEA (8:8:16 equiv with respect to peptide-resin)-mediated couplings in DMF, 1 h each, were required for the Fmoc-Gly-OH incorporation to go to completion. Semipreparative HPLC (condition G) provide *separately* analytical-HPLC pure P1 and P2 [t_R 18.4 min (P1) and t_R 18.8 min (P2); condition A]. Purified P1 and P2 were indistinguishable by amino acid analysis (Asp, 1.98; Thr, 0.91; Glu, 1.00; Gly, 3.99; Ala, 2.00; Val, 1.01; Ile, 1.37; Tyr, 2.86; Phe, 1.92; Lys, 2.79; Arg, 2.02) and MALDI-TOF; calcd monoisotopic mass: 2980.45, found: 2981.3 $[M+H]^+$.

$II_{L29Z(J)}$. In the linear core module sequence, Leu^{29} was replaced by Fmoc-Dpr(Boc-Ser(*t*Bu))-OH. Cleavage, disulfide bond formation, and purification followed as already described for $II_{L29K(J)}$ (Figure 4b). Semipreparative HPLC (condition G) provide *separately* analytical-HPLC pure P1 and P2 [t_R 18.8 min (P1) and 19.2 (P2) min; condition A]. Purified P1 and P2 were indistinguishable by amino acid analysis (Asp, 2.01; Thr, 0.93; Glu, 1.00; Gly, 3.12; Ala, 1.99; Val, 0.94; Ile, 1.24; Tyr, 2.67; Phe 1.93; Lys, 1.94; Arg, 2.07) and MALDI-TOF; calcd monoisotopic mass: 2881.38, found: 2882.1 $[M+H]^+$.

$II_{N24K(J)}$. The procedure for this molecule was essentially as described for $II_{L29K(J)}$ (Figure 4b), except that Asn^{24} was replaced by Fmoc-Lys(Alloc)-OH and Leu^{29} was added normally. Semipreparative HPLC (condition G) provide *separately* analytical-HPLC pure P1 and P2 [t_R 20.8 min (P1) and 21.8 min (P2); condition A]. Purified P1 and P2 were indistinguishable by amino acid analysis (Asp, 1.01; Thr,

0.86; Glu, 1.02; Gly, 3.09; Ala, 1.98; Val, 0.96; Ile, 1.58; Leu, 1.00; Tyr, 2.57; Phe, 1.97; Lys, 2.97; Arg, 1.97) and MALDI-TOF; calcd monoisotopic mass: 2922.47, found: 2923.12 $[M+H]^+$.

Oxime-forming ligation reactions

Capping of monomers

$\mathbf{I}_{R17K(U^*)}$. The synthesis and characterization of this molecule is described in the main text.

$\mathbf{I}_{R17Z(U^*)}$. Pure peptide was characterized by analytical HPLC (t_R 27 min; condition A), amino acid analysis (Asp, 2.01; Thr, 0.92; Glu, 0.99; Gly, 3.01; Ala, 2.02; Val, 0.95; Ile, 1.11; Leu, 1.00; Tyr, 2.85; Phe, 1.96; Lys, 2.05; Arg, 1.05), and MALDI-TOF; calcd monoisotopic mass: 2895.42, found: 2896.09 $[M+H]^+$.

$\mathbf{I}_{R17K(U^*G)}$. Pure peptide was characterized by analytical HPLC (t_R 25.5 min; condition A), amino acid analysis (Asp, 1.99; Thr, 0.86; Glu, 1.03; Gly, 3.95; Ala, 2.02; Val, 0.99; Ile, 1.63; Leu, 1.03; Tyr, 2.78; Phe, 1.95; Lys, 2.91; Arg, 1.05), and MALDI-TOF; calcd monoisotopic mass: 2994.49, found: 2996.3 $[M+H]^+$.

$\mathbf{II}_{L29K(J^*)}$. The synthesis and characterization of this molecule is described in the main text.

$\mathbf{II}_{L29K(J^*G)}$. Pure peptide was characterized by analytical HPLC (t_R 21.3 min; condition A), amino acid analysis (Asp, 2.00; Thr, 1.03; Glu, 1.13; Gly, 3.89; Ala, 2.05; Val, 1.07; Ile, 1.40; Tyr, 2.89; Phe 2.01; Lys, 2.76; Arg, 1.87), and MALDI-TOF; calcd monoisotopic mass: 3009.47, found: 3010.9 $[M+H]^+$.

$\mathbf{II}_{L29Z(J^*)}$. Pure peptide was characterized by analytical HPLC (t_R 20.1 min; condition A), amino acid analysis (Asp, 1.96; Thr, 0.83; Glu, 1.06; Gly, 2.96; Ala, 1.97; Val, 0.98; Ile, 1.59; Tyr, 2.56; Phe, 1.87; Lys, 1.96; Arg, 2.06), and ESMS; calcd average mass: 2912.35, found: 2912.6 ± 1.4.

$\mathbf{II}_{N24K(J^*)}$. Pure peptide was characterized by analytical HPLC (t_R 24.6 min; condition A), amino acid analysis (Asp, 1.14; Thr, 1.01; Glu, 1.25; Gly, 3.00; Ala, 1.78; Val, 1.13; Ile, 1.15; Leu, 1.15; Tyr, 2.45; Phe, 1.96; Lys, 2.56; Arg, 1.95), and MALDI-TOF; calcd monoisotopic mass: 2951.49, found: 2953.2 $[M+H]^+$.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry SUPPORTING INFORMATION

Dimer formation

$\mathbf{I}_{\mathbf{R17K(U)}}$–$\mathbf{II}_{\mathbf{L29K(J)}}$. The synthesis and characterization of this molecule is described in the main text.

$\mathbf{I}_{\mathbf{R17Z(U)}}$–$\mathbf{II}_{\mathbf{L29K(J)}}$. Pure peptide was characterized by analytical HPLC (t_R 29.1 min; condition B), amino acid analysis (Asp, 3.95; Thr, 1.82; Glu, 2.03; Gly, 5.95; Ala, 3.96; Val, 1.93; Ile, 2.56; Leu, 1.06; Tyr, 5.55; Phe. 3.90; Lys, 4.89; Arg, 2.83), and ESMS; calcd average mass: 5764.68, found: 5764.1 ± 1.8.

$\mathbf{I}_{\mathbf{R17Z(U)}}$–$\mathbf{II}_{\mathbf{L29Z(J)}}$. Pure peptide was characterized by analytical HPLC (t_R 29.0 min; condition B), amino acid analysis (Asp, 3.96; Thr, 1.87; Glu, 2.01; Gly, 5.97; Ala, 3.94; Val, 1.92; Ile, 2.46; Leu, 1.06; Tyr, 5.57; Phe, 3.91; Lys, 3.91; Arg, 2.87), and ESMS; calcd average mass: 5722.66, found: 5719.8 ± 1.5.

$\mathbf{I}_{\mathbf{R17K(U)}}$–$\mathbf{II}_{\mathbf{L29K(JG)}}$. Pure peptide was characterized by analytical HPLC (t_R 28.6 min; condition B), amino acid analysis (Asp, 3.91; Thr, 1.63; Glu, 2.01; Gly, 7.26; Ala, 3.91; Val, 1.88; Ile, 2.91; Leu, 0.93; Tyr, 4.30; Phe, 3.78; Lys, 5.47; Arg, 2.81), and ESMS; calcd average mass: 5862.87, found: 5863.56 ± 0.64.

$\mathbf{I}_{\mathbf{R17K(UG)}}$–$\mathbf{II}_{\mathbf{L29K(JG)}}$. Pure peptide was characterized by analytical HPLC (t_R 28.8 min; condition B), amino acid analysis (Asp, 4.15; Thr, 1.98; Glu, 2.21; Gly, 7.46; Ala, 4.00; Val, 2.02; Ile, 2.90; Leu, 1.15; Tyr, 5.83; Phe, 3.93; Lys, 5.79; Arg, 2.85), and ESMS; calcd average mass: 5919.92, found: 5920.31 ± 0.48.

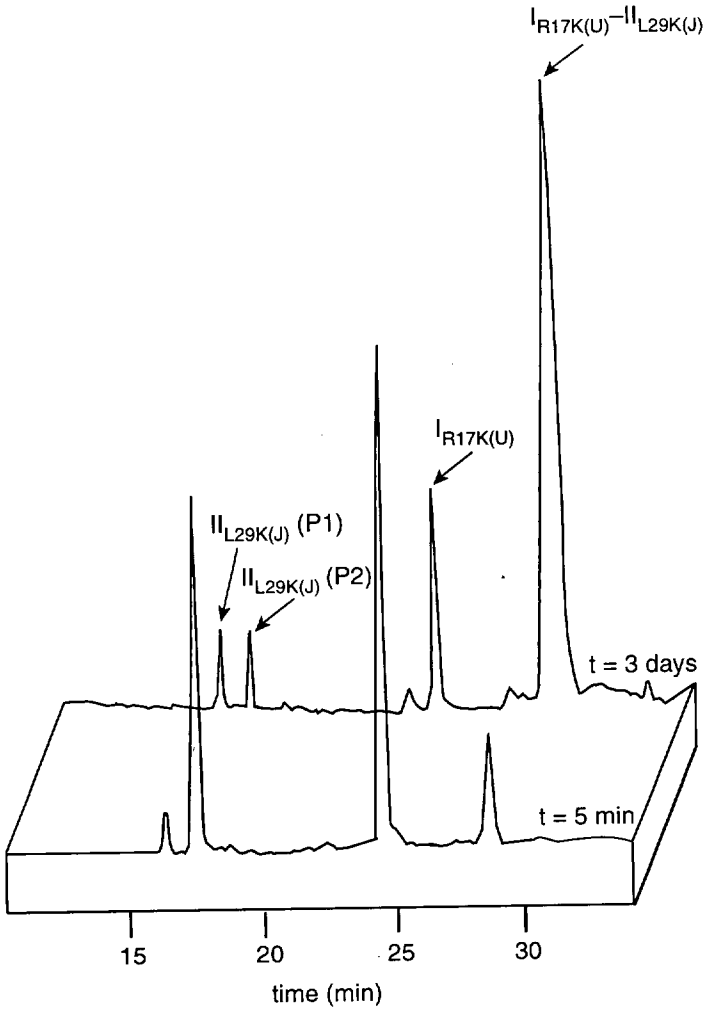
$\mathbf{I}_{\mathbf{R17K(U)}}$–$\mathbf{II}_{\mathbf{N24K(J)}}$. Pure peptide was characterized by HPLC (t_R 29.2 min; condition B), amino acid analysis (Asp, 2.99; Thr, 1.61; Glu, 2.01; Gly, 6.15; Ala, 4.02; Val, 1.89; Ile, 3.35; Leu, 1.99; Tyr, 5.08; Phe, 3.91; Lys, 5.91; Arg, 2.87), and ESMS; calcd average mass: 5805.82, found: 5805.00 ± 0.78.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION



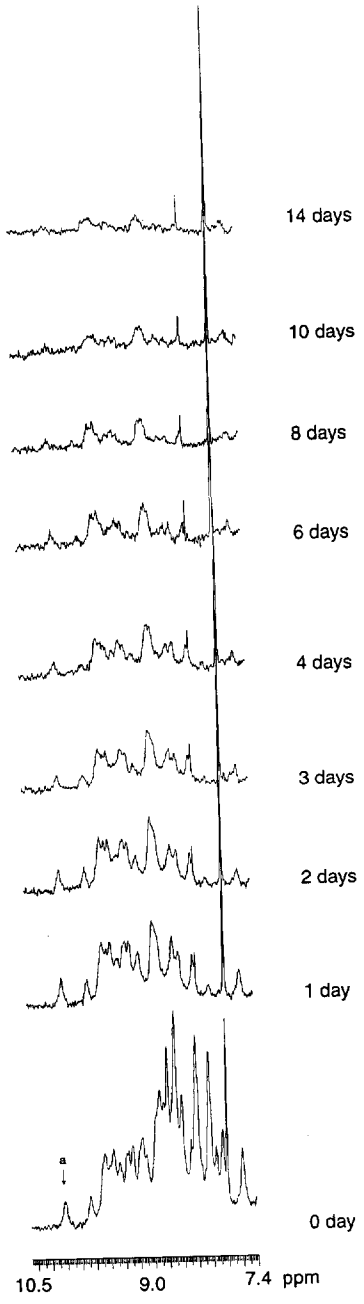
Supporting Figure 1. RP-HPLC analysis (C_4 column; condition B) to document oxime-forming reaction to produce OxCM dimer, carried out in 6 M GdnHCl in 0.1 M NaOAc buffer, pH 4.6. Front panel, at start of reaction (5 min), and rear panel, at point where, no appreciable further product was formed (3 days). Peaks in the rear panel are labeled by the system introduced in Figure 2 of the main text; the peaks in the front panel are not explicitly labeled but correspond. Also, see Figure 4 for definitions of P1 and P2.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION



Supporting Figure 2. $^{1}H/^{2}H$ isotope exchange at pH 2 and 5 °C, of $I_{R17K(U)}$–$II_{L29K(J)}$, monitored by a series of 1D ^{1}H NMR spectra at 600 MHz. Amide resonance 'a', indicated by arrow, was sufficiently resolved from the remaining envelope to allow calculation of hydrogen isotope exchange rate.

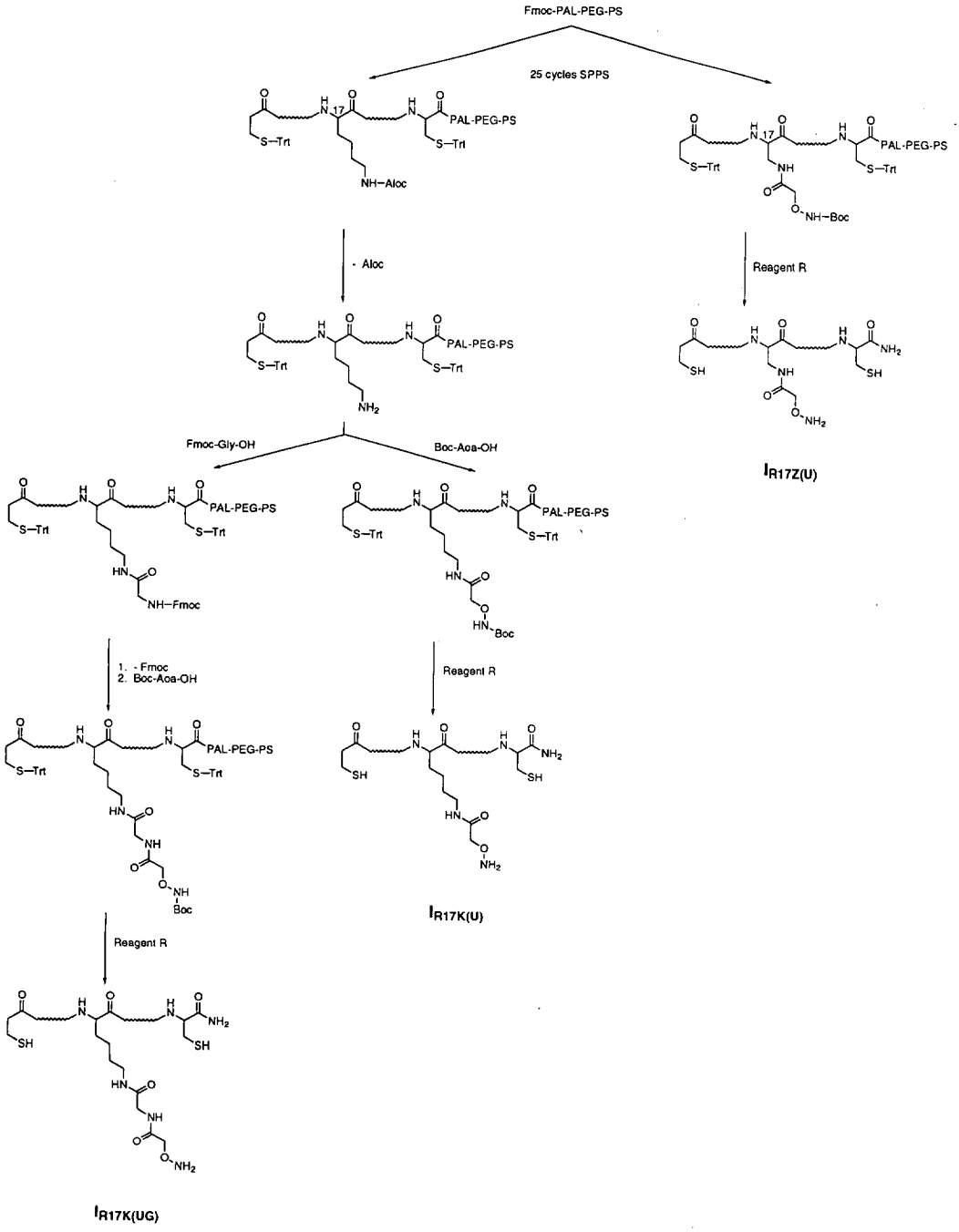
N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION

Supporting Scheme 1. Synthesis and cleavage of monomers of type I containing the (aminooxy)acetyl functionality.



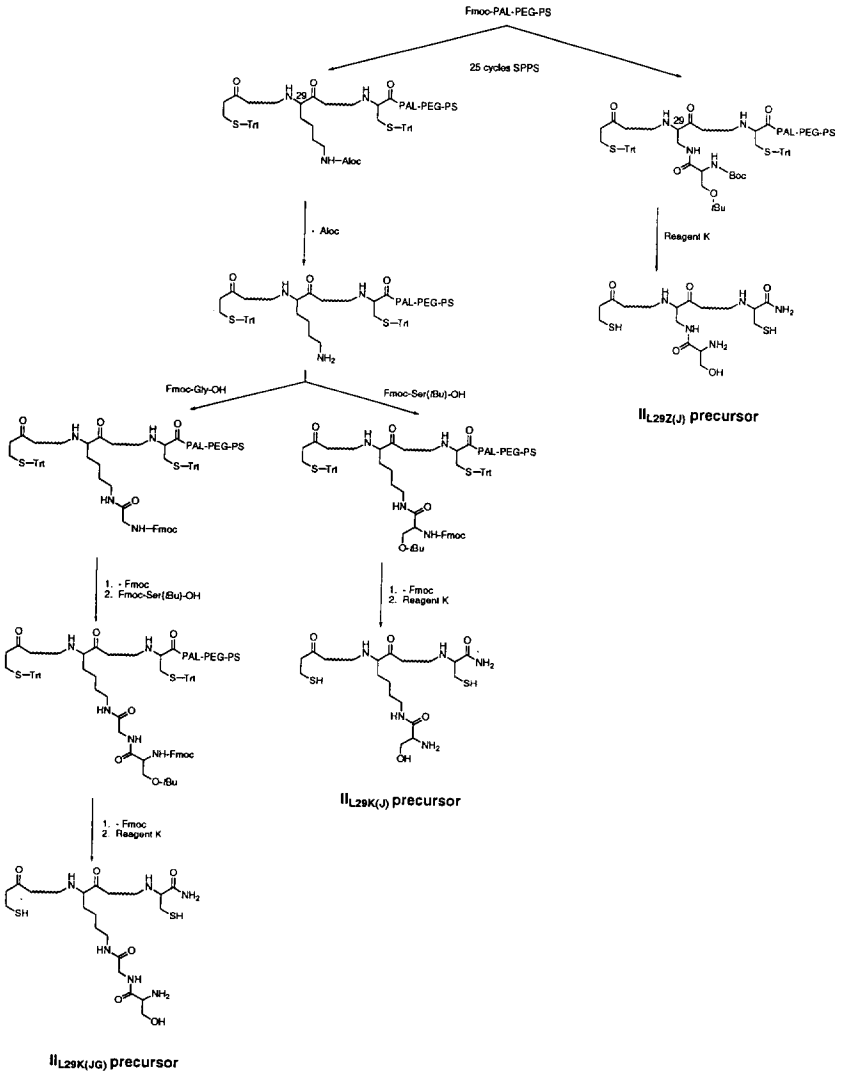
N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION

Supporting Scheme 2. Synthesis and cleavage of monomers of type II containing the glyoxylyl functionality. [a],[b]



[a] The schemes are drawn up to the point of reduced precursors with a free Ser residue at the *N*-terminus of the branch. Further manipulations to form the disulfide and create the Gxy moiety are discussed in the main text.

[b] $II_{N24K(J)}$ (not shown in the Scheme) was prepared as $II_{L29K(J)}$, with the only difference that Fmoc-Lys(Alloc)-OH was incorporated at position 24.

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION

Supporting Table 1. ^{1}H chemical shifts for H-Tyr-Dpr(Gxy)-Ala-Lys-NH_2, recorded at 500 MHz in $^{2}H_2O$ at pH 4.6 in the absence and presence of denaturants

Residue	Position	$^{2}H_2O$	6 M GdnDCl-d_6	8 M urea-$^{15}N_2$,d_4
Tyr	CαH	4.47	4.42	4.41
	CβH	3.33, 3.40	3.33, 3.37	3.30, 3.36
	CδH	7.10	7.03	7.07
	CεH	7.37	7.31	7.33
Dpr (Gxy)	CαH	4.77	4.81	4.79
	CβH	3.70, 3.83	3.71, 3.79	3.70, 3.83
	CεH[(a)]	5.52	5.52, 5.71, 5.72	5.57, 5.61
Ala	CαH	4.50	4.54	4.50
	CβH	1.66	1.61	1.66
Lys-amide	CαH	4.50	4.52	4.48
	CβH	2.05, 2.08	1.97, 2.01	2.00, 2.07
	CγH	1.69, 1.72	1.65	1.66, 1.69
	CδH	1.94	1.89	1.92
	CεH	3.25	3.20	3.22

[(a)] Main text, Figures 5, 6, and Table 2 refer to this as H-C(2).

N. Carulla, C. Woodward, and G. Barany

Towards New Designed Proteins Derived from Bovine Pancreatic Trypsin Inhibitor (BPTI): Covalent Cross-linking of Two 'Core Modules' by Oxime-forming Ligation.

Bioconjugate Chemistry

SUPPORTING INFORMATION

Supporting Table 2. SEC chromatography of OxCM dimers[a]

	Molecular Weight	Retention time (mL)
BPTI[b]	6512.7	20.86
$I_{R17K(UG)}$–$II_{L29K(JG)}$	5919.9	22.64
$I_{R17K(U)}$–$II_{L29K(JG)}$	5862.9	22.90
$I_{R17K(U)}$–$II_{L29K(J)}$	5806.8	24.06
$I_{R17K(U)}$–$II_{N24K(J)}$	5805.8	24.23
$I_{R17Z(U)}$–$II_{L29K(J)}$	5764.7	24.30
$I_{R17Z(U)}$–$II_{L29Z(J)}$	5722.7	24.56
Gly[b]	75.1	37.84

[a] The proteins were eluted in CH_3CN–10 mM aqueous NaCl (1:19) at pH 2 and 3 °C.

[b] BPTI and Gly are used as standards.