Threonine Phosphorylation is a Bioreversible Surrogate of Proline Hydroxylation in a Collagen Triple Helix

Isabella M. B. Pearl and Ronald T. Raines*
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

*E-mail: rtraines@mit.edu

Table of Contents

1. Abbreviations Used

Ac	acetyl
ACN	acetonitrile
CD	circular dichroism
CHCA	α-cyano-4-hydroxycinnamic acid
CIP	calf intestinal phosphatase (EC 3.1.3.1; UniProtKB P19111)
DCM	dichloromethane
DIC	diisopropylcarbodiimide
DIPEA	*N*,*N*′-diisopropylethylamine
DMF	dimethylformamide
DODT	2,2-(ethylenedioxy)diethanethiol
Fmoc	9-fluorenylmethoxycarbonyl
Gly	glycine
Hyp or O	(2*S*,4*R*)-4-hydroxyproline
HPLC	high-performance liquid chromatography
MALDI–TOF	matrix-assisted laser desorption/ionization–time-of-flight
Oxyma	ethyl cyano(hydroxyimino)acetate
Pase	phosphatase
PPII	polyproline type II
Pro or P	(2*S*)-proline
^{p}Ser or ^{p}S	*O*-phospho-(2*S*)-serine
^{p}Thr or ^{p}T	*O*-phospho-(2*S*,3*R*)-threonine
RT	room temperature (~23 °C)
SD	standard deviation
Ser or S	(2*S*)-serine
*t*Bu	*tert*-butyl
TEA	triethylamine
TFA	trifluoroacetic acid
THD	triple-helical domain
Thr or T	(2*S*,3*R*)-threonine
TIS	triisopropylsilane

2. General Experimental Procedures

Reagents. Commercial chemicals were of reagent grade or better, and were used without further purification. FmocGlyOH, FmocSer(*t*Bu)OH, FmocThr(*t*Bu)OH, FmocSer(PO(OBzl)OH)OH, FmocThr(PO(OBzl)OH)OH, Oxyma, and Rink Amide ProTide Resin (low-loading) were from CEM Corporation (Matthews, NC). FmocHyp(*t*Bu)OH was from AK Scientific (Union City, CA). DIC and 4-methylpiperidine were from Oakwood Chemical (Tampa, FL). Acetic anhydride, DIPEA, TIS, CHCA, and DODT were from Sigma–Aldrich (St. Louis, MO). NEBuffer 2 (product #B7002S) and calf intestinal phosphatase (CIP) (product #M0525S) were from New England Biolabs (Ipswich, MA). TFA was from Beantown Chemical (Hudson, NH). All other reagents were from Fisher Scientific (Hampton, NH).

Conditions. All procedures were performed at ambient temperature (~23 °C) and pressure (1.0 atm) unless indicated otherwise.

3. Peptide Synthesis—General

This section was adapted from Dones et al. (2019).[1] Peptides were synthesized with a Liberty Blue Automated Microwave Peptide Synthesizer from CEM. All peptides were synthesized following CEM standard methods for both microwave and coupling cycles. Standard solutions of an activator, DIC (0.5 M in DMF), an activator base, Oxyma (1 M in DMF), deprotection solution, 4-methylpiperidine (20% v/v in DMF) + 0.1 M Oxyma, Fmoc-protected amino acids (0.2 M in DMF), and a capping solution (10% v/v acetic anhydride in DMF) were prepared for each synthesis.

Microwave-Assisted Deprotection. The microwave was set to 155 W at 75 °C for 15 s, followed by 30 W at 90 °C for 50 s.

Microwave-Assisted Coupling. The microwave was set to 170 W at 75 °C for 15 s, followed by 30 W at 90 °C for 225 s.

Conventional coupling. The reaction was carried out in the instrument reaction vessel for 15 min in the absence of microwave heating.

Standard Coupling Cycle. Rink Amide ProTide Resin (1 equiv) was added to the CEM reaction vessel, and the resin was allowed to swell for 5 min in DMF. The Fmoc group was removed using the standard deprotection solution and the microwave-assisted deprotection methods described above. The resin was then washed (4×), and Fmoc-AA-OH (5 equiv) was added, followed by DIC (20 equiv) and Oxyma (40 equiv). Standard microwave-assisted coupling was performed with additional Fmoc-protected amino acids, and the resin was washed (2×) and drained. When double-coupling was required, the cycle was repeated without the deprotection step.

4. Peptide Synthesis—Specific

Nonphosphorylated peptides: Ac-(POG)$_6$PO-NH$_2$, Ac-(POG)$_3$(SOG)(POG)$_2$PO-NH$_2$, Ac-(POG)$_3$(TOG)(POG)$_2$PO-NH2, Ac-(POG)$_3$(PSG)(POG)$_2$PO-NH$_2$, and Ac-(POG)$_3$(PTG)(POG)$_2$PO-NH$_2$. Nonphosphorylated peptides were synthesized by a non-interrupted continuous method. After the deprotection of Fmoc-protected Rink Amide ProTide Resin, Fmoc-protected amino acids were coupled by either a single or a double standard coupling cycle until completion. The low nucleophilicity of Hyp required a double standard coupling cycle. Following final deprotection, the resin was capped with acetic anhydride. Next, the resin was moved from the reaction vessel to a cleavage filter, washed with DCM (3×), and air-dried crude peptides were then cleaved from the resin using a cleavage cocktail composed of 95:2.5:2.5 TFA/H_2O/TIS for 2 h. Peptide mixtures were filtered and precipitated in ice-cold diethyl ether. The peptides were collected by centrifugation, the supernatants were decanted, and the solid peptide was dissolved in 5 mL of 70:30 H_2O/ACN. The solutions were frozen and lyophilized using a FreeZone benchtop instrument from Labconco (Kansas City, MO). The crude peptide mixture was then subjected to purification.

Phosphorylated peptides: Ac-(POG)$_3$(^{p}SOG)(POG)$_2$PO-NH$_2$, Ac-(POG)$_3$(^{p}TOG)(POG)$_2$PO-NH$_2$, Ac-(POG)$_3$(P^{p}SG)(POG)$_2$PO-NH$_2$, and Ac-(POG)$_3$(P^{p}TG)(POG)$_2$PO-NH$_2$. Phosphorylated peptides were synthesized by a non-interrupted continuous method. After deprotection of Fmoc-protected Rink Amide ProTide Resin, Fmoc-protected amino acids were coupled by either a single or a double standard coupling cycle until completion. The low nucleophilicity of Hyp required a double standard coupling cycle. For coupling of Fmoc-Ser(PO(OBzl)OH)-OH, Fmoc-Thr(PO(OBzl)OH)-OH, and the amino acid N-terminal to these couplings, a conventional coupling cycle was performed to minimize elimination of the phosphate group. After the last coupling, the N-terminal amino acid (proline), was deprotected and capped by treating with acetic anhydride.

The resin was then moved from the reaction vessel to a cleavage filter, washed with DCM (3×), and air-dried crude peptides were then cleaved from the resin using a cleavage cocktail composed of 92.5:2.5:2.5:2.5 TFA/H_2O/TIS/DODT for 2 h. Peptide mixtures were then filtered and precipitated in ice-cold diethyl ether. The peptides were collected by centrifugation, the supernatants were decanted, and the solid peptide was dissolved in 5 mL of 70:30 H_2O/ACN. The solutions were frozen and lyophilized using a FreeZone benchtop lyophilizer from Labconco. The crude peptide mixture was then subjected to purification.

Phosphorylated peptides were also obtained from ABI Scientific (Sterling, VA). These peptides were indistinguishable from those synthesized in our laboratory.

5. Peptide Purification and Characterization

The crude peptide products were purified by preparative reversed-phase HPLC using an XSelect Peptide CSH C18 OBD Prep Column 130 Å, 5 μm, 19 mm × 250 mm from Waters (Milford, MA) and a 1260 Infinity II instrument from Agilent Technologies (Santa Clara, CA). Crude products were dissolved in the minimum amount of ACN and eluted with a linear gradient of 5–80% v/v ACN in H_2O containing TFA (0.1% v/v). After reviewing the initial chromatogram, the method was updated, if necessary. Fractions containing purified peptide were pooled, lyophilized, and analyzed with analytical reversed-phase HPLC and MALDI–TOF mass spectrometry.

Analytical HPLC. Analytical HPLC was performed with an Agilent 1260 Infinity II instrument equipped with an XSelect CSH C18 column from Waters. Solvent A was 95:5 H_2O/ACN containing TFA (0.1% v/v), and solvent B was 95:5 ACN/H_2O containing TFA (0.1% v/v).

MALDI–TOF Mass Spectrometry. MALDI–TOF mass spectrometry analyses were carried out with a microflex LRF mass spectrometer from Bruker (Billerica, MA). Peptides were analyzed by using a saturated solution of α-cyano-4-hydroxycinnamic acid in 30:70 ACN/water containing TFA (0.1% v/v).

6. Circular Dichroism Spectroscopy

Peptide secondary structure was analyzed via circular dichroism (CD) spectroscopy. Peptides were dried under vacuum for ≥48 h before weighing and dissolving to 0.4 mM in 10 mM sodium phosphate buffer, pH 7.4. To ensure consistency across samples, peptide concentrations were verified by their absorbance at 205 nm (ε = 31 mL mg^{-1} cm^{-1}) using a DS-11 Integrated UV–vis spectrophotometer from DeNovix (New Castle County, DE). The resulting solutions were heated to 65 °C and then cooled to 4 °C at 0.2 °C/min. After ≥48 h at 4 °C, CD spectra were recorded with a Model J-1500 CD spectrometer from JASCO (Easton, MD) and a bandpass of 1 nm at the MIT Biophysics Instrumentation Facility. The signal was averaged for 3 s during the wavelength scan. Peptide solutions were heated at 0.2 °C/min while monitoring molar ellipticity at 225 nm. Values of T_m were determined by fitting the molar ellipticity at 225 nm to a four-parameter Hill equation. All CD spectroscopy experiments were performed in triplicate.

7. Peptide Dephosphorylation

Peptides with sequences XaaYaa = ProThr and XaaYaa = Pro^{p}Thr (Table S1) were dissolved separately to 0.5 mM in NEBuffer 2, which is 10 mM Tris–HCl buffer, pH 7.9, containing NaCl (50 mM), $MgCl_2$ (10 mM), and DTT (1 mM). The peptides were annealed into homotrimeric triple helices by heating and cooling as described in Section 6. CIP was added to the solution of the XaaYaa = Pro^{p}Thr peptide at a concentration of 1 unit per pmol of peptide. The reaction mixture was incubated at 37 °C for 24 h. Dephosphorylation was assessed with MALDI–TOF mass

spectrometry. Values of T_m were determined before and after CIP addition by CD spectroscopy as described in Section 6, except in NEBuffer 2.

8. Bioinformatic Analyses

The UniProt database (https://www.uniprot.org) was searched for collagen amino acid sequences from selected mammalian species. To quantify Ser and Thr in these long amino acid sequences, we wrote a Python script to count Ser or Thr residues in the Xaa or Yaa position. This script identified 4-residue phrases "GS*G" or "GT*G" to count instances at the Xaa position, or "G*SG" or "G*TG" to count instances at the Yaa position, where * can be any amino acid.

Table S1. Amino Acid Sequences of CMPs Used in this Work.

Ac-POGPOGPOG**PO**GPOGPOGPO-NH_2
Ac-POGPOGPOG**SO**GPOGPOGPO-NH_2
Ac-POGPOGPOG**TO**GPOGPOGPO-NH_2
Ac-POGPOGPOG**PS**GPOGPOGPO-NH_2
Ac-POGPOGPOG**PT**GPOGPOGPO-NH_2
Ac-POGPOGPOG^{p}**SO**GPOGPOGPO-NH_2
Ac-POGPOGPOG^{p}**TO**GPOGPOGPO-NH_2
Ac-POGPOGPOG**P**^{p}**S**GPOGPOGPO-NH_2
Ac-POGPOGPOG**P**^{p}**T**GPOGPOGPO-NH_2

Table S2. Values of T_m for Triple-Helical CMPs with the Sequence Ac-(ProHypGly)$_3$-(XaaYaaGly)-(ProHypGly)$_2$(ProHyp)-NH_2.

Condition	XaaYaa	T_m (°C)[a]	XaaYaa	T_m (°C)[a]	ΔT_m
1	ProHyp	46.6	—	—	—
1	SerHyp	35.4	^{p}SerHyp	41.3	+5.9
1	ProSer	33.1	Pro^{p}Ser	31.0	−2.1
1	ThrHyp	32.9	^{p}ThrHyp	39.8	+6.9
1	ProThr	35.0	Pro^{p}Thr	45.5	+10.5
2	Pro^{p}Thr[b]	45.6	Pro^{p}Thr + Pase	37.3	−8.3
2	ProThr	36.6			+0.7

[a]Values were determined with CD spectroscopy and are the mean of 3 experiments. [b]Value is from 1 experiment. Thermal denaturation curves are shown in Figures S11 and S12B. Condition 1: CMP (0.4 mM) in 10 mM sodium phosphate buffer, pH 7.4. Condition 2: CMP (0.5 mM) in 10 mM Tris–HCl buffer, pH 7.9, containing NaCl (50 mM), $MgCl_2$ (10 mM), and DTT (1 mM).

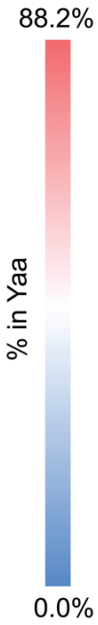
Table S3. Ser and Thr Residues in the Xaa and Yaa Positions of 30 Mammalian Collagens.

Fibrillar Collagens

Collagen		Ser				Thr			
Subtype	Uniprot ID	% overall	Xaa	Yaa	% in Yaa	% overall	Xaa	Yaa	% in Yaa
Bovine α1(I)	P02453	3.23	17	18	51.4	1.57	2	15	88.2
Human α1(I)	P02452	3.35	17	17	50.0	1.97	4	16	80.0
Mouse α1(I)	P11087	3.65	18	19	51.4	1.48	3	12	80.0
Bovine α2(I)	P02465	3.42	16	19	54.3	1.76	7	11	61.1
Human α2(I)	P08123	3.16	16	16	50.0	1.87	6	13	68.4
Mouse α2(I)	Q01149	4.72	25	23	47.9	2.26	8	15	65.2
Bovine α1(II)	P02459	2.56	9	17	65.4	2.17	6	16	72.7
Human α1(II)	P02458	3.16	16	16	50.0	2.07	5	16	76.2
Mouse α1(II)	P28481	3.16	11	21	65.6	2.17	4	18	81.8
Rat α1(II)	P05539	2.96	12	18	60.0	2.37	5	19	79.2
Bovine α1(III)	P04258	4.39	27	18	40.0	1.37	6	8	57.1
Human α1(III)	P02461	4.09	27	15	35.7	1.46	6	9	60.0
Mouse α1(III)	P08121	4.59	29	18	38.3	1.56	9	7	43.8
Human α1(V)	P20908	2.08	10	11	52.4	1.98	3	17	85.0
Mouse α1(V)	O88207	2.18	12	10	45.5	2.18	3	19	86.4
Human α2(V)	P05997	3.22	29	7	19.4	2.60	12	17	58.6

Nonfibrillar Collagens

Collagen		Ser				Thr			
Subtype	Uniprot ID	% overall	Xaa	Yaa	% in Yaa	% overall	Xaa	Yaa	% in Yaa
Human α1(IV)	P02462	3.00	32	6	15.8	1.42	11	7	38.9
Mouse α1(IV)	P02463	3.39	35	8	18.6	1.74	13	9	40.9
Human α2(IV)	P08572	1.85	19	5	20.8	1.85	11	13	54.2
Mouse α2(IV)	P08122	2.47	19	13	40.6	2.08	13	14	51.9
Human α3(IV)	Q01955	2.79	32	7	17.9	2.01	16	12	42.9
Mouse α3(IV)	Q9QZS0	1.79	20	5	20.0	1.79	15	10	40.0
Human α4(IV)	P53420	2.44	20	14	41.2	0.86	5	7	58.3
Mouse α4(IV)	Q9QZR9	1.94	21	6	22.2	0.65	4	5	55.6
Human α5(IV)	P29400	1.70	14	10	41.7	1.06	9	6	40.0
Human α6(IV)	Q14031	3.81	31	23	42.6	1.83	15	11	42.3
Human α1(VI)	P12109	2.09	5	2	28.6	0.60	2	0	0.0
Mouse α1(VI)	Q04857	1.79	4	2	33.3	0.60	1	1	50.0
Human α2(VI)	P12110	2.99	6	4	40.0	0.90	3	0	0.0
Mouse α2(VI)	Q02788	3.29	8	3	27.3	0.90	3	0	0.0



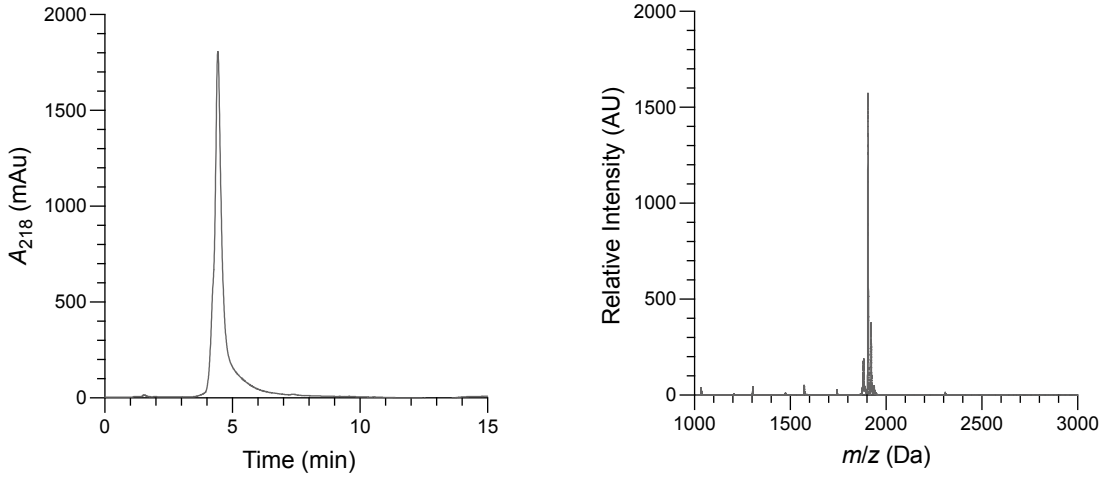


Figure S1. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_6PO\text{-}NH_2$. RT, 4.46 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + H]^+$ expected, 1873.01; found, 1873.55.

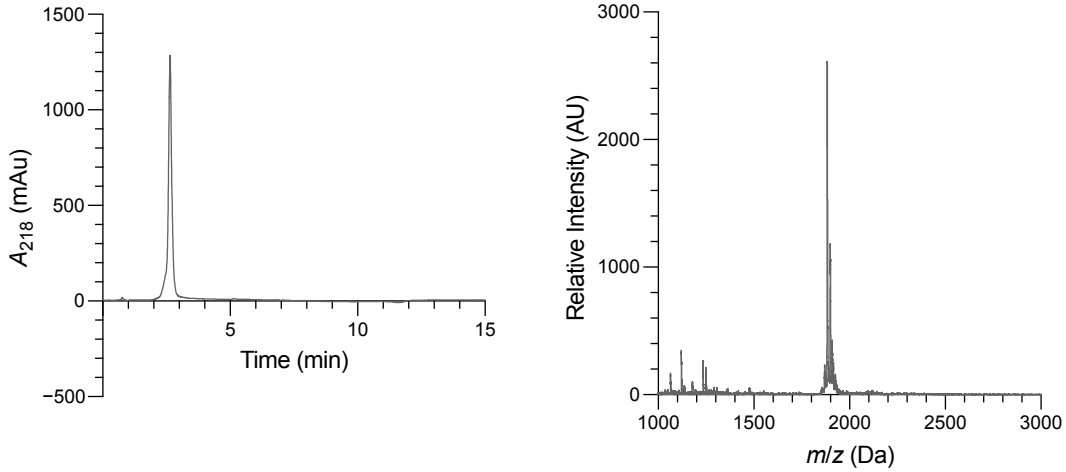


Figure S2. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_3$(**SO**G)$(POG)_2PO\text{-}NH_2$. RT, 2.65 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + Na]^+$ expected, 1884.84; found, 1883.44.

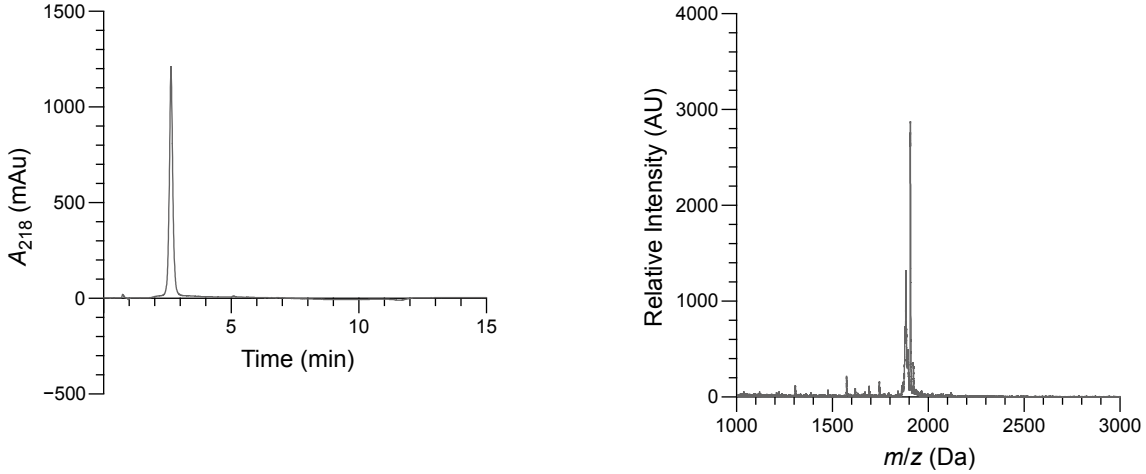


Figure S3. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_3$(**TO**G)$(POG)_2$PO-NH_2. RT, 2.66 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + H]^+$ expected, 1875.86; found, 1875.47.

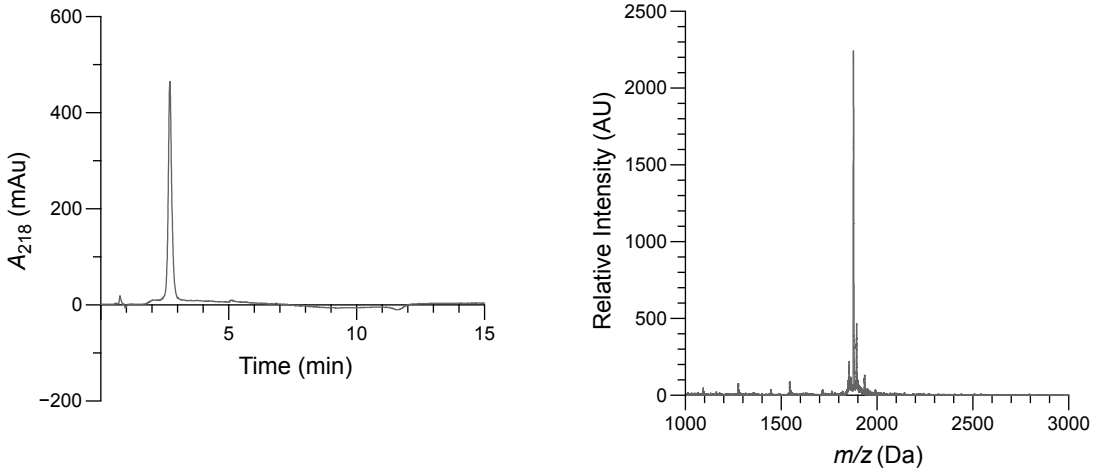


Figure S4. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_3$(**PS**G)$(POG)_2$PO-NH_2. RT, 2.72 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + H]^+$ expected, 1845.85; found, 1845.41.

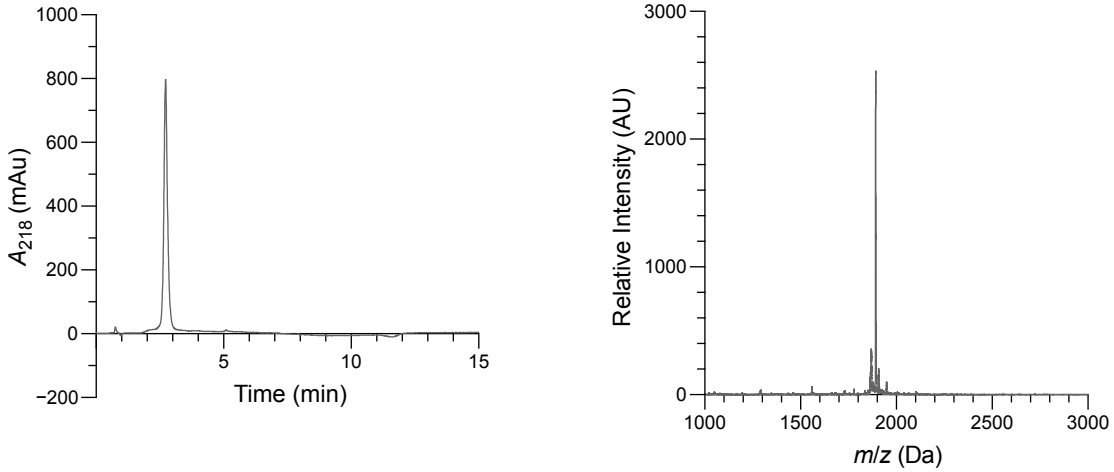


Figure S5. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_3$(**PT**G)$(POG)_2$PO-NH_2. RT, 2.73 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + H]^+$ expected, 1860.88; found, 1860.85.

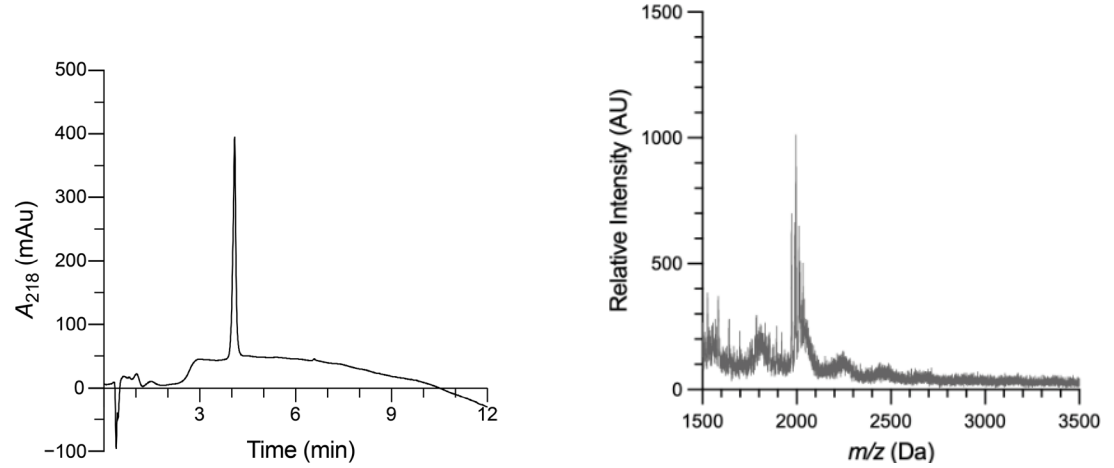


Figure S6. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-$(POG)_3$(ᴰ**SO**G)$(POG)_2$PO-NH_2. RT, 4.1 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + Na]^+$ expected, 1963.85; found, 1963.88.

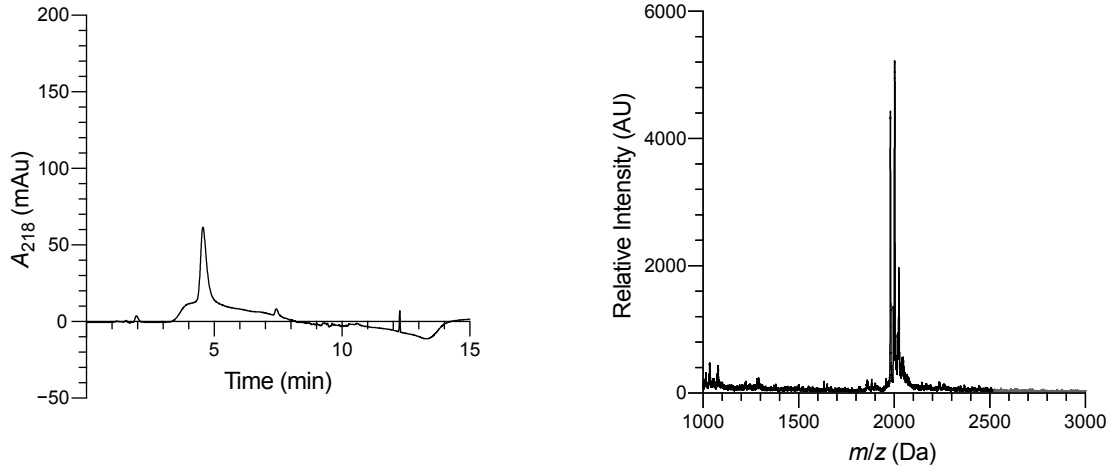


Figure S7. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-(POG)$_3$(**^{p}TO**G)(POG)$_2$PO-NH$_2$. RT, 4.61 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + Na]^+$ expected, 1978.99; found, 1980.88; $[M + K]^+$ expected, 1995.09; found, 1996.82.

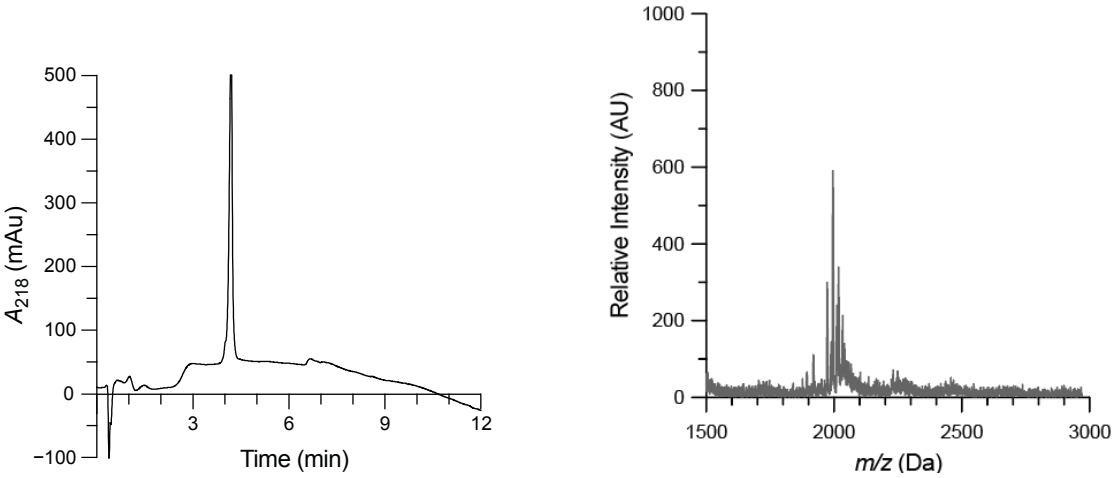


Figure S8. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-(POG)$_3$(**P^{p}S**G)(POG)$_2$PO-NH$_2$. RT, 4.1 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + K]^+$ expected, 1964.78; found, 1963.23.

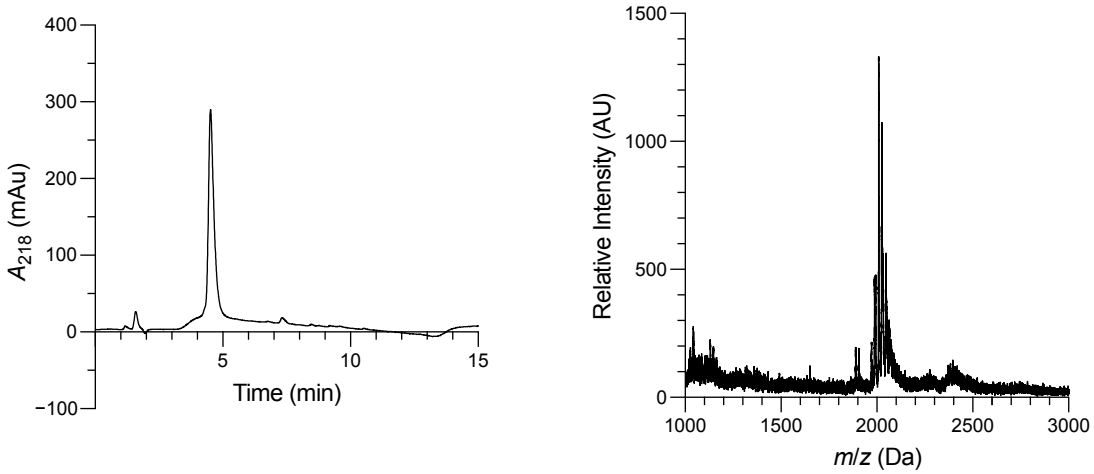


Figure S9. Analytical HPLC trace (left) and MALDI–TOF mass spectrum (right) of purified peptide Ac-(POG)$_3$(**P^{p}T**G)(POG)$_2$PO-NH$_2$. RT, 4.54 min (0–100% v/v B over 10 min, then 5 min 10% v/v B). MALDI–TOF *m*/*z* $[M + Na]^+$ expected, 1962.98; found, 1962.62; $[M + K]^+$ expected, 1978.95, found, 1978.60.

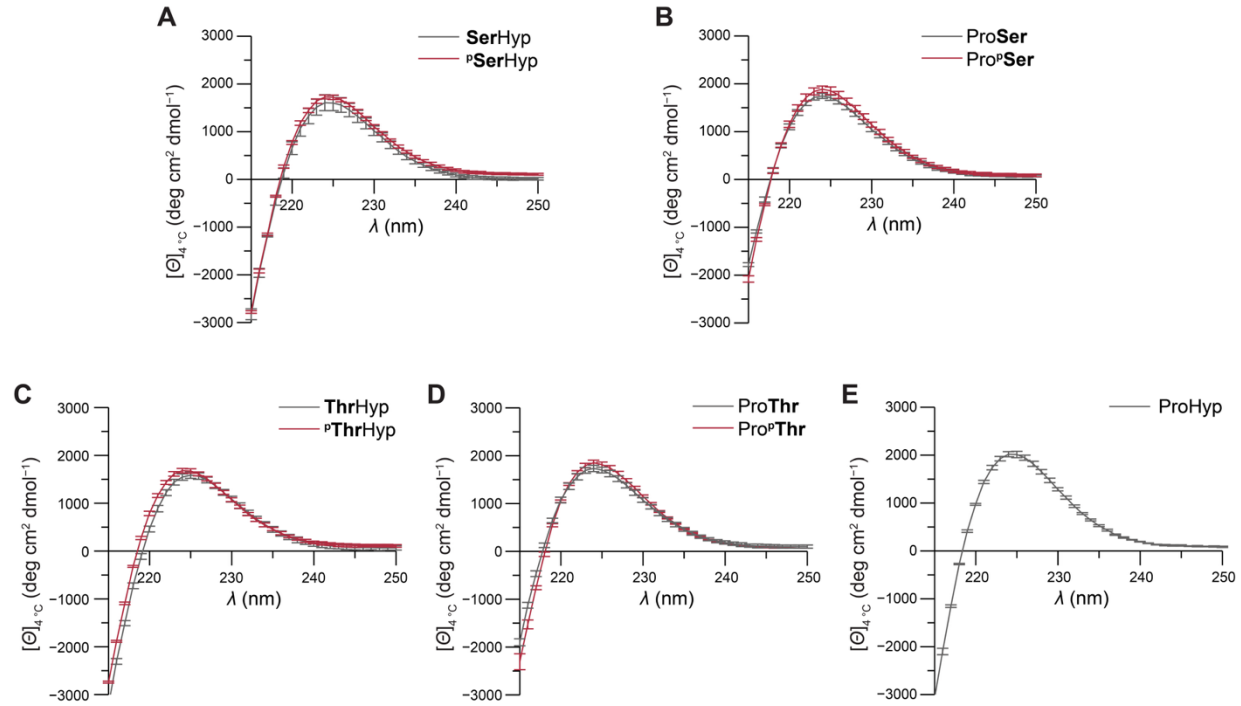


Figure S10. Circular dichroism spectra of annealed collagen-mimetic peptides (0.4 mM in 10 mM sodium phosphate buffer, pH 7.4) with the sequence Ac-(POG)$_3$(XYG)(POG)$_2$PO-NH$_2$ equilibrated at 4 °C. XY are the following residues: (A) SerHyp or ^{p}SerHyp, (B) ProSer or Pro^{p}Ser, (C) ThrHyp or ^{p}ThrHyp, (D) ProThr or Pro^{p}Thr, and (E) ProHyp. Values are the mean ± SD from 3 experiments. The data in Panel D are also shown in Figure 2A.

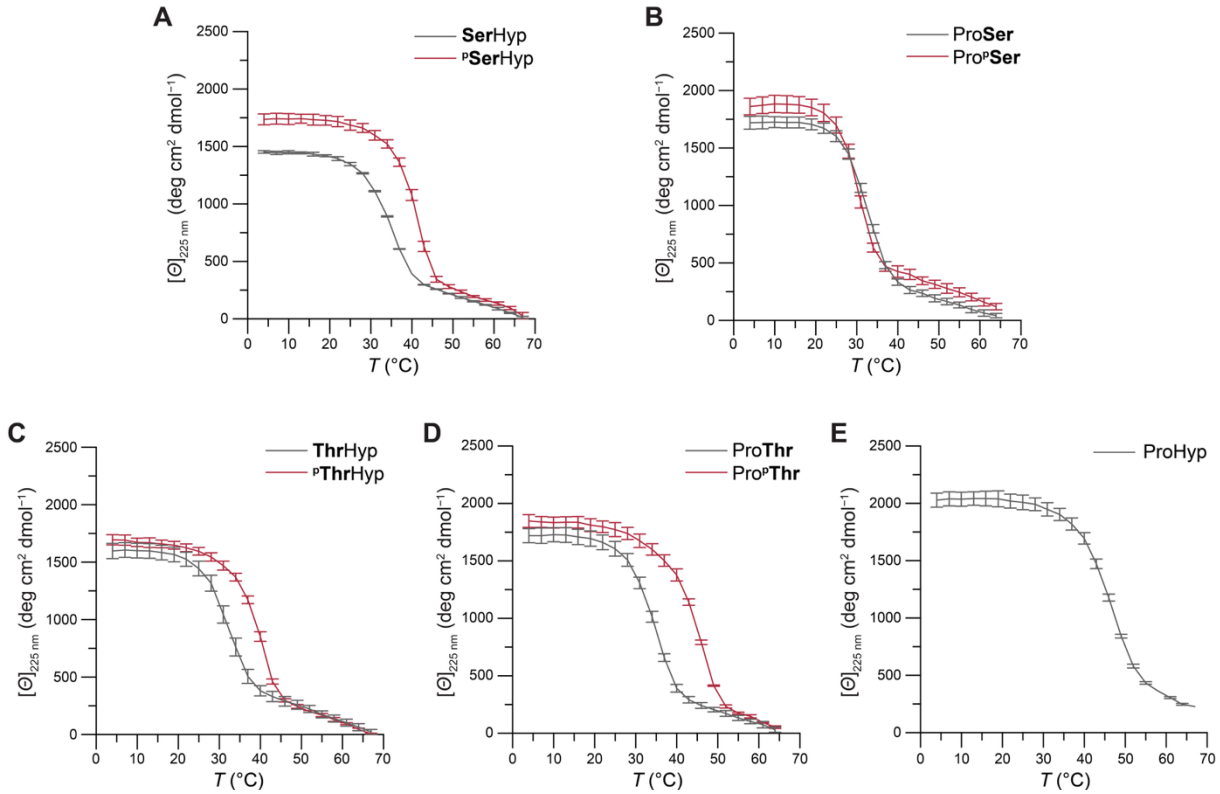


Figure S11. Effect of temperature on the molar ellipticity at 225 nm of annealed collagen-mimetic peptides (0.4 mM in 10 mM sodium phosphate buffer, pH 7.4). CMPs had the sequence Ac-(POG)$_3$(XYG)(POG)$_2$PO-NH$_2$. XY are the following residues: (A) SerHyp or ^{P}SerHyp, (B) ProSer or Pro^{P}Ser, (C) ThrHyp or ^{P}ThrHyp, (D) ProThr or Pro^{P}Thr, and (E) ProHyp. Values are the mean ± SD from 3 experiments. The data in Panel D are also shown in Figure 2B. Values of T_m are listed in Table S2.

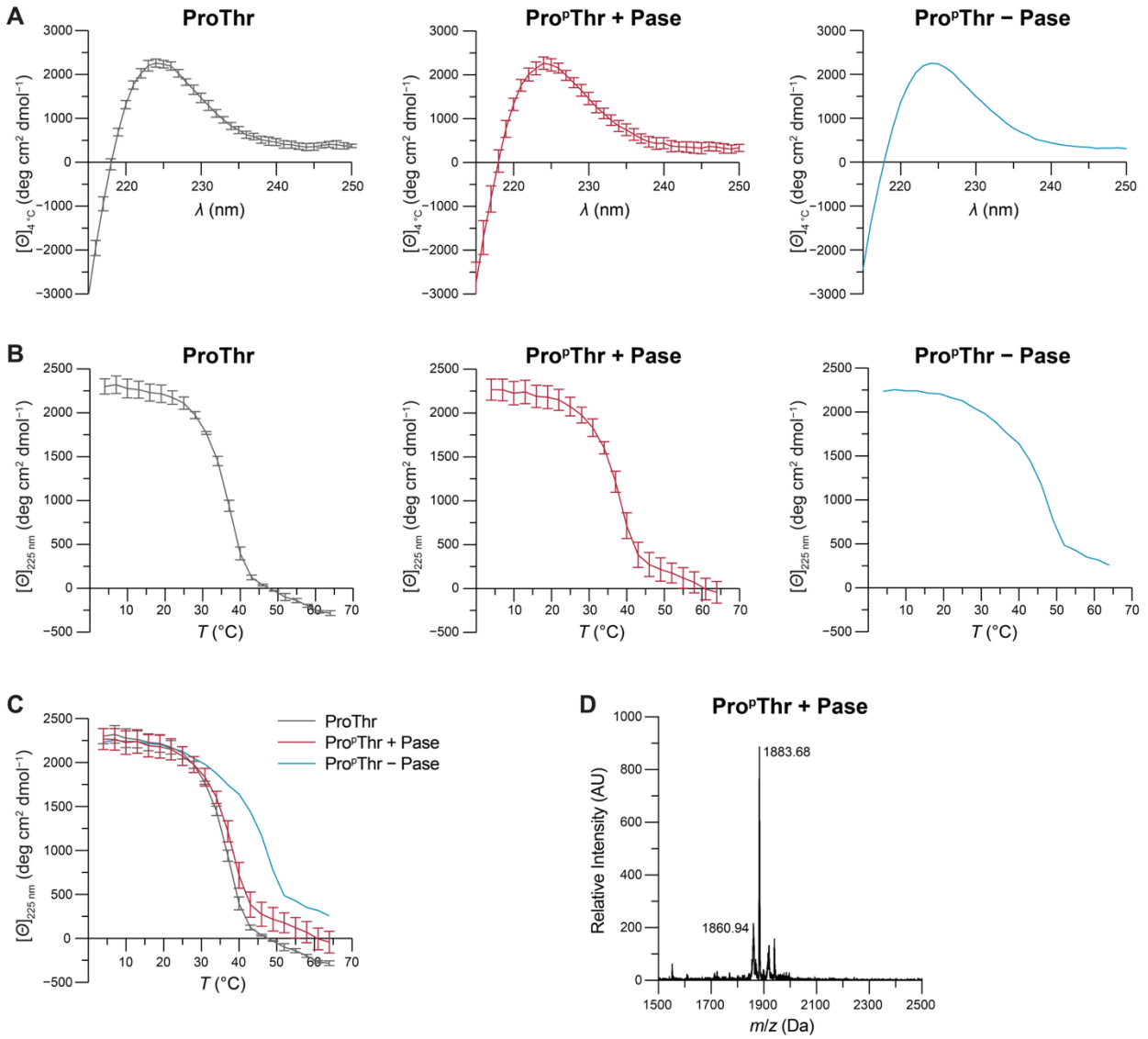


Figure S12. Effect of a calf intestinal phosphatase on annealed collagen-mimetic peptides (0.5 mM in 10 mM Tris–HCl buffer, pH 7.9, containing 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT). (A) Circular dichroism spectra of the annealed CMPs with the sequence Ac-$(POG)_3$(PYG)$(POG)_2$PO-NH_2 equilibrated at 4 °C. Y is Thr or ᵖThr. (B) Effect of temperature on the molar ellipticity at 225 nm of the annealed CMPs. Values of T_m are listed in Table S2. (C) Compilation of the data in panel B. Values for ProThr and ProᵖThr + Pase are the mean ± SD from 3 experiments. (D) MALDI–TOF mass spectrum of Ac-$(POG)_3$(P^pTG)$(POG)_2$PO-NH_2 treated with calf intestinal phosphatase (Pase). *m/z* $[M + H]^+$ expected for dephosphorylated peptide, 1860.88; found, 1860.94; $[M + Na]^+$ expected for dephosphorylated peptide, 1882.86; found, 1883.68.

Reference

(1) Dones, J. M.; Tanrikulu, I. C.; Chacko, J. V.; Schroeder, A. B.; Hoang, T. T.; Gibson, A. L. F.; Eliceiri, K. W.; Raines, R. T. Optimization of interstrand interactions enables burn detection with a collagen-mimetic peptide. *Org. Biomol. Chem.* **2019**, *17*, 9906–9912.