Supporting Information

Development of an amino acid sequence-dependent analytical method for peptides using near-infrared spectroscopy

Mika Ishigaki[1,2]*, Atsushi Ito[3], Risa Hara[3], Shun-ichi Miyazaki[3]*, Kodai Murayama[3], Sana Tusji[1], Miho Inomata[1], Keisuke Yoshikiyo[1], Tatsuyuki Yamamoto[1,2], Yukihiro Ozaki[4]

[1]*Institute of Agricultural and Life Sciences, Academic Assembly, Shimane University, 1060 Nishikawatsu, Matsue, Shimane, 690-8504, Japan*
[2]*Raman Project Center for Medical and Biological Applications, Shimane University, 1060 Nishikawatsu, Matsue, Shimane 690-8504, Japan*
[3]*Research and Development Department, Yokogawa Electric Corporation, 2-9-32 Nakacho, Musashino, Tokyo 180-8750, Japan*
[4]*School of Biological and Environmental Sciences, Kwansei Gakuin University, 2-1 Gakuen-Uegahara, Sanda, Hyogo 669-1330, Japan*

*Authors to whom correspondence should be sent.
*E-mail: ishigaki@life.shimane-u.ac.jp (M.I.)
shun-ichi.miyazaki@yokogawa.com (S.M)

Table of Contents

SI 1. Information about amino acids and peptides
SI 2. Desalting operation
SI 3. Calculation of the background-subtracted second derivative spectra
SI 4. NIR spectra of amino acids and peptides

SI 1. Information about amino acids and peptides

Detailed information about amino acids and peptides is presented, and the chemical structures of (I) glycine and (II) Boc-protected triglycine is shown in Figure S1.

i. Amino acids

H-Gly-OH: G00018, Watanabe Chemical Ind. Ltd., Japan
H-Ala-OH: G00001, Watanabe Chemical Ind. Ltd., Japan
H-Ser-OH: 199-00402, FUJIFILM Wako Pure Chemical Co., Japan
H-Glu-OH: 070-00502, FUJIFILM Wako Pure Chemical Co., Japan
H-Lys-OH: L0129, Watanabe Chemical Ind. Ltd., Japan
H-Phe-OH: G00029, Watanabe Chemical Ind. Ltd., Japan
H-Tyr-OH: 202-03562, FUJIFILM Wako Pure Chemical Co., Japan
H-Pro-OH: G00030, Watanabe Chemical Ind. Ltd., Japan
H-Ser(*t*Bu)-OH: K01262, Watanabe Chemical Ind. Ltd., Japan
H-Glu(O*t*Bu)-OH: K01261, Watanabe Chemical Ind. Ltd., Japan
H-Lys(Boc)-OH: L00638, Watanabe Chemical Ind. Ltd., Japan
H-Tyr(*t*Bu)-OH: L00733, Watanabe Chemical Ind. Ltd., Japan
H-Gly-O*t*Bu·HCl: 27532-96-3, Tokyo Chemical Co., Ltd., Japan
H-Pro-O*t*Bu·HCl: 5497-76-7, Tokyo Chemical Co., Ltd., Japan

ii. *N-tert*-butoxycarbonyl (Boc)-protected amino acids and peptides

Boc-Gly-OH: K00172, Watanabe Chemical Ind. Ltd., Japan
Boc-GlyGly-OH: R00080, Watanabe Chemical Ind. Ltd., Japan
Boc-GlyPro-OH: R00089, Watanabe Chemical Ind. Ltd., Japan
Boc-ProGly-OH: R00434, Watanabe Chemical Ind. Ltd., Japan
Boc-ProPro-OH: Mw=303.39, >90%, PH Japan Co., Ltd, Japan
Boc-GlyGlyGly-OH: 28320-73-2, Watanabe Chemical Ind. Ltd., Japan
Boc-GlyGlyAla-OH: M_w=303.31, >90%, GL Biochem Ltd., China
Boc-GlyGlySer(*t*Bu)-OH: M_w=375.29, >80%, HiPep Laboratories, Japan
Boc-GlyGlyGlu(O*t*Bu)-OH: M_w=417.32, >80%, HiPep Laboratories, Japan
Boc-GlyGlyLys(Boc)-OH: M_w=460.39, >80%, HiPep Laboratories, Japan
Boc-GlyGlyPhe-OH: M_w=379.41, >90%, HiPep Laboratories, Japan
Boc-GlyGlyTyr(*t*Bu)-OH: M_w=451.38, >80%, HiPep Laboratories, Japan
Boc-GlyGlyPro-OH: M_W=329.35, >90%, GL Biochem Ltd., China
Boc-GlyProGly-OH: Mw=329.35, >90%, PH Japan Co., Ltd, Japan
Boc-ProGlyGly-OH: Mw=329.35, >90%, PH Japan Co., Ltd, Japan
Boc-ProProGly-OH: Mw=369.42, >90%, PH Japan Co., Ltd, Japan

Boc-ProGlyPro-OH: Mw=369.42, >90%, PH Japan Co., Ltd, Japan
Boc-GlyProPro-OH: Mw=369.42, >90%, PH Japan Co., Ltd, Japan
Boc-GlyAlaGly-OH: M_w=303.32, >90%, GL Biochem Ltd., China
Boc-AlaGlyGly-OH: M_w=303.32, >90%, GL Biochem Ltd., China
Boc-AlaAlaGly-OH: M_w=317.34, >90%, GL Biochem Ltd., China
Boc-AlaGlyAla-OH: M_w=317.34, >90%, GL Biochem Ltd., China
Boc-GlyAlaAla-OH: M_w=317.34, >90%, GL Biochem Ltd., China

Amino acids were dissolved with ultrapure water or 0.5 mol/L NaOH depending on their solubility. The amino acids dissolved in NaOH aqueous solution were H-Glu-OH, H-Tyr-OH, H-Lys(Boc)-OH, and H-Tyr(*t*Bu)-OH, and the others were in ultrapure water. The maximum concentration of H-Phe-OH was set as 150 mM because it could not be solved up to 200 mM. Boc-protected amino acids and peptides were dissolved with DMSO.

Mixed DMSO solutions were prepared to reproduce the concentration gradients of the constituents that occurred in the micro flow reactor in the course of peptide synthesis. The detailed recipe for mixing DMSO solutions is shown in Table S1, and the changes in each solution over time are designated 1-9. In mixed DMSO solutions consisting of amino acids and peptides assumed to be raw materials and products in the course of peptide synthesis, the concentration of each substance was quantitatively estimated using PLSR analysis. Two types of raw materials and one product were assumed to be present in chemical reactions designed to synthesize dipeptides with one glycine and one proline (a, b) and tripeptides with two glycine and one proline (c-e) or one glycine and two prolines (f-h), as listed in Table S2. The PLSR results are shown in Tables S3 and S4.

SI 2. Desalting operation

Amino acids without protection groups are not dissolved in DMSO. Thus, glycine *t*Bu ester hydrochloride (H-Gly-O*t*Bu·HCl) and proline *t*Bu ester hydrochloride (H-Pro-O*t*Bu·HCl) were used after their desalting operation as model substances of raw materials for peptide synthesis.

The 2 grams of glycine *t*Bu ester hydrochloride and proline *t*Bu ester hydrochloride were dissolved in 50 mL of a sodium hydrogen carbonate aqueous solution (191-01305, FUJIFILM Wako Pure Chemical Co., Japan), and it was mixed with 30 mL of an ethyl acetate (051-00351, FUJIFILM Wako Pure Chemical Co., Japan) within a separatory funnel. After removing the upper ethyl acetate layer, the same operation where ethyl acetate was added to the lower solution and the upper solution was separated was performed twice. The ethyl acetate layer collected was mixed with 30 mL of an aqueous solution of a commercially available sodium chloride, and the same operation was performed twice for the upper layer of ethyl acetate. The ethyl acetate layer was dried with calcium sulfate (7778-18-9, Hammond Drierite Co. Ltd., U.S.A.), and the collected samples were used as naked amino acids without protection groups after removing ethyl acetate by an evaporator. After all operations, 1000 μL of an aqueous solution of silver nitrate (190-00834, FUJIFILM Wako Pure Chemical Co., Japan) was mixed with 0.1 μL of a desalted sample, and no white precipitate was confirmed to be generated.

Figures S2A and S2B show the background subtracted second derivative spectra in the 5000-4500 cm^{-1} region before and after desalting of H-Gly-O*t*Bu·HCl in DMSO with concentrations of 50, 100, 150, and 200 mM, respectively. The spectral patterns changed before and after desalting, and the bands at 4950 and 4735 cm^{-1} were clearly detected after desalting. They corresponded to the common bands observed in the second derivative spectra of NaOH aqueous solutions of glutamine, lysine, and tyrosine in Figure 3(A). That is, these bands were derived from an amino group changed from hydrochloride by desalting. The series of desalting operation was reconfirmed to be properly accomplished by NIR spectral changes.

The background subtracted second derivative spectra of Boc-Gly-OH in DMSO with concentrations of 50, 100, 150, and 200 mM are shown in Figure S2C. Since an amide bond was formed on the connecting part between a protecting group and the *N*-terminus of amino acids, the bands at

4838, 4711, and 4627 cm^{-1} were newly observed due to N-H stretching and amide II, the combination of the first overtone of amide II and amide I, and the combination of N-H stretching and amide III, respectively.[20,24,26] The NIR spectra were consistently understood without contradictions.

SI 3. Calculation of the background subtracted second derivative spectra

Figure S3A depicts the second derivative spectra in the 5000-4500 cm^{-1} region of DMSO and Boc-Gly-OH in DMSO with concentrations of 50, 100, 150, and 200 mM. The spectral features showed almost those of DMSO, and slight variations depending on the Boc-Gly-OH concentration were observed. To remove the DMSO contributions, the subtracted second derivative spectra were calculated by defining DMSO second derivatives as background (Figure S3B). The peaks derived from amino acids should have stronger intensities in a concentration dependent manner. The peaks at approximately 4837, 4712, and 4627 cm^{-1} were confirmed to be changed depending on amino acid concentration, and they were assigned to the combination of N-H stretching and amide II, the combination of the first overtone of amide II and amide I, and the combination of N-H stretching and amide III, respectively. [20,24,26]

SI 4. NIR spectra of amino acids and peptides

Figure S4A depicts an NIR absorbance spectrum in the 10000-4000 cm^{-1} region for glycine aqueous solution with 200 mM concentration. Figures S4B and S4C show the second derivative spectra in the 5000-4200 and 6200-5750 cm^{-1} regions, respectively, for four kinds of amino acids aqueous solutions with *t*Bu group with 200 mM concentration. Figure S4D expresses the second derivative spectra in the 6200-5700 cm^{-1} region for 200 mM solutions of eight amino acids in DMSO. Figure S5 exhibits the background subtracted second derivative spectra in the 5000-4500 cm^{-1} region for six kinds of tripeptides composed of glycine and alanine. Figure S6 shows the PLSR loading plots of Factor 1 obtained for six tripeptides (c-h) in Table S2.

Table S1: Recipe for mixing organic solutions.

No.	Boc-GlyGly-OH (μL) 800 mM	H-Pro-OH (μL) 800 mM	Boc-GlyGlyPro-OH (μL) 400 mM	DMSO (μL)	Total volume (μL)
1	250	250	0	500	1000
2	225	225	50	500	1000
3	187.5	187.5	125	500	1000
4	150	150	200	500	1000
5	125	125	250	500	1000
6	100	100	300	500	1000
7	62.5	62.5	375	500	1000
8	25	25	450	500	1000
9	0	0	500	500	1000

Table S2: The list of substances assumed to be raw materials and products in peptide synthesis.

	Raw material 1	Raw material 2	Product
(a)	Boc-Gly-OH	H-Pro-OH	Boc-GlyPro-OH
(b)	Boc-Pro-OH	H-Gly-OH	Boc-ProGly-OH
(c)	Boc-GlyGly-OH	H-Pro-OH	Boc-GlyGlyPro-OH
(d)	Boc-GlyPro-OH	H-Gly-OH	Boc-GlyProGly-OH
(e)	Boc-ProGly-OH	H-Gly-OH	Boc-ProGlyGly-OH
(f)	Boc-GlyPro-OH	H-Pro-OH	Boc-GlyProPro-OH
(g)	Boc-ProGly-OH	H-Pro-OH	Boc-ProGlyPro-OH
(h)	Boc-ProPro-OH	H-Gly-OH	Boc-ProProGly-OH

Table S3: Results of the coefficient of determination (R^2) and root mean square error (RMSE) for both calibration and validation evaluated using a PLSR model to determine the concentration of each constituent within mixed DMSO solutions reproducing dipeptide synthesis reactions (a) and (b) in Table S2.

		Boc-Gly-OH	**H-Pro-OH**	**Boc-GlyPro-OH**
(a)	**Calibration**	R^2=0.999 RMSE=1.82	R^2=0.999 RMSE=1.82	R^2=0.999 RMSE=1.82
	Validation	R^2=0.999 RMSE=2.53	R^2=0.999 RMSE=2.53	R^2=0.999 RMSE=2.53
		Boc-Pro-OH	**H-Gly-OH**	**Boc-ProGly-OH**
(b)	**Calibration**	R^2=0.999 RMSE=1.24	R^2=0.999 RMSE=1.24	R^2=0.999 RMSE=1.24
	Validation	R^2=0.999 RMSE=1.88	R^2=0.999 RMSE=1.88	R^2=0.999 RMSE=1.88

Table S4: Results for the coefficient of determination (R^2) and root mean square error (RMSE) for both calibration and validation evaluated using the PLSR model to determine the concentration of raw material 1 within mixed DMSO solutions reproducing dipeptide synthesis reactions (c)-(h) in Table S2.

	Calibration	**Validation**
(c)	R^2=0.999, RMSE=1.60	R^2=0.998, RMSE=2.78
(d)	R^2=0.998, RMSE=2.90	R^2=0.997, RMSE=3.89
(e)	R^2=0.998, RMSE=2.66	R^2=0.995, RMSE=5.36
(f)	R^2=0.999, RMSE=2.39	R^2=0.997, RMSE=4.00
(g)	R^2=0.999, RMSE=0.76	R^2=0.999, RMSE=1.94
(h)	R^2=0.999, RMSE=2.39	R^2=0.998, RMSE=3.35

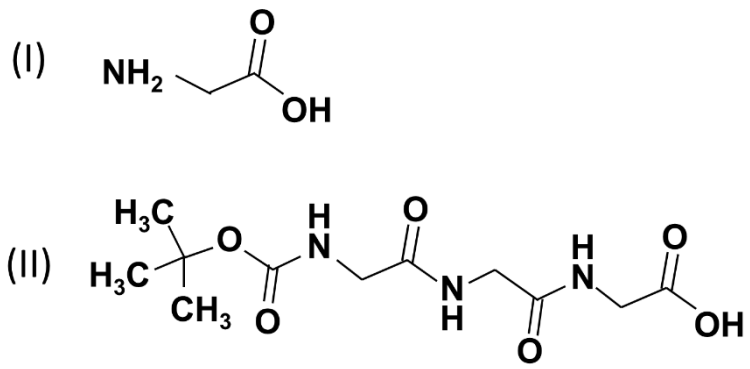
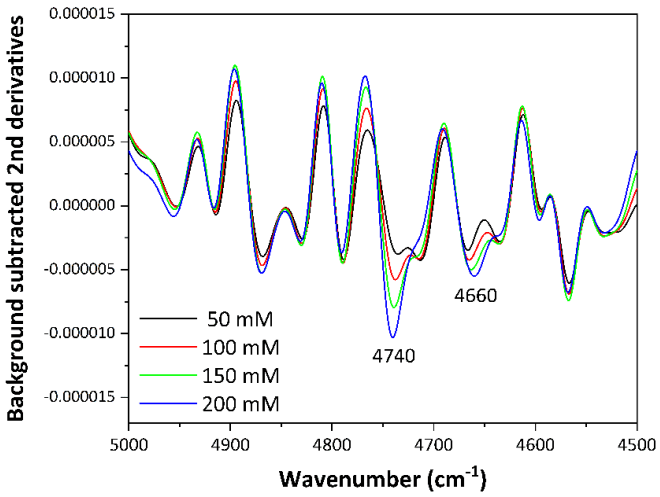
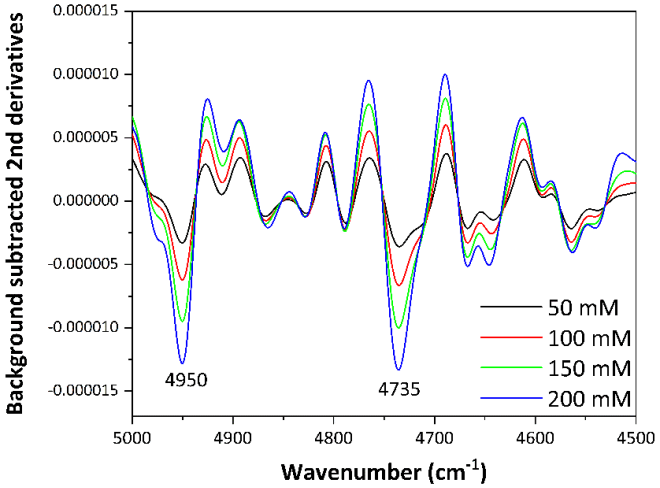


Figure S1: Chemical structures of (I) glycine and (II) Boc-protected triglycine.

(A)



(B)



(C)

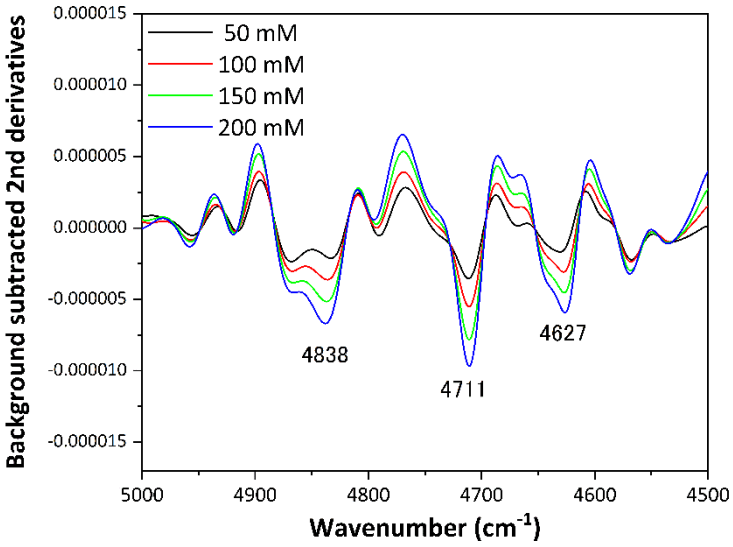
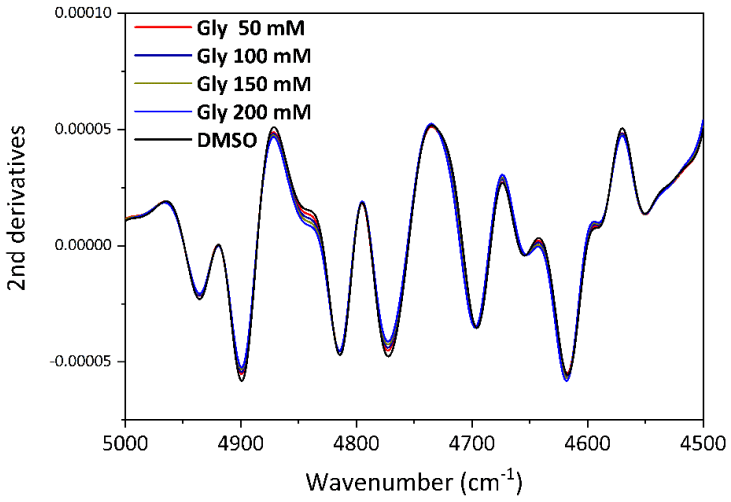


Figure S2: The background subtracted second derivative spectra in the 5000-4500 cm^{-1} region of (A) glycine *t*Bu ester hydrochloride, (B) glycine *t*Bu ester, and (C) Boc-protected glycine in DMSO with concentrations of 50, 100, 150, and 200 mM.

(A)



(B)

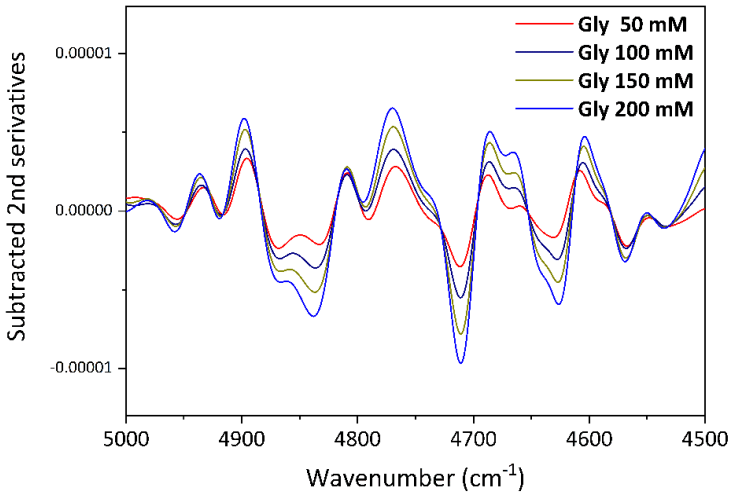
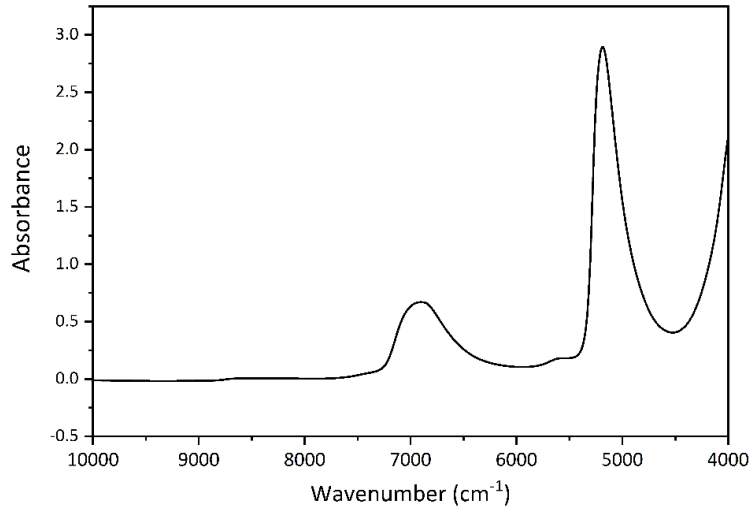
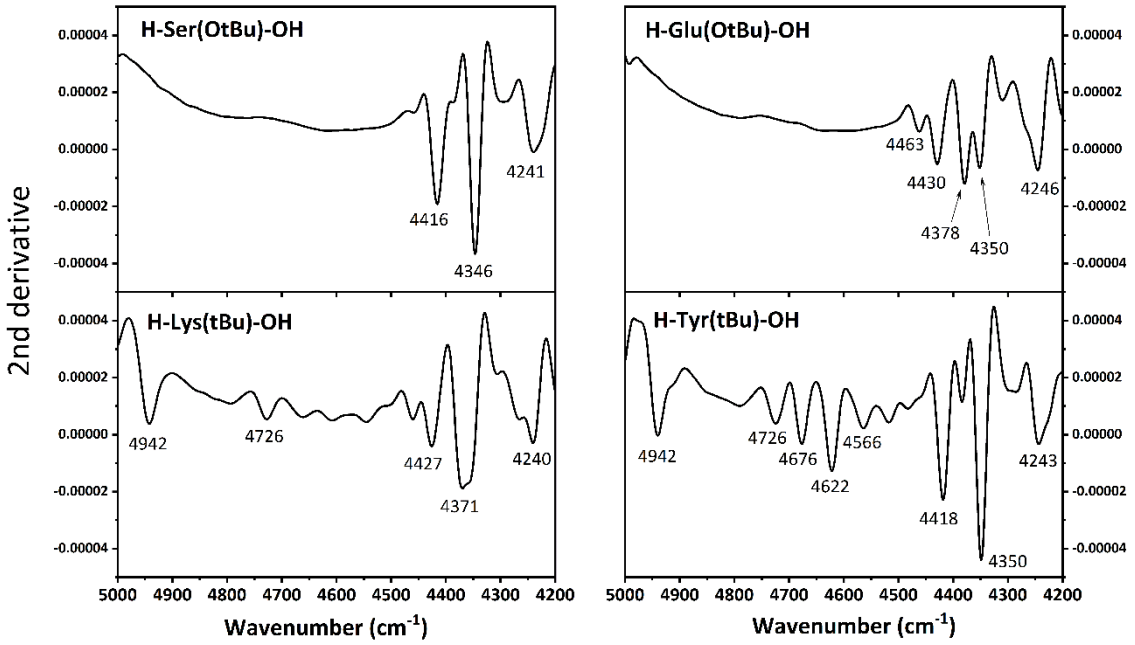


Figure S3: (A) The second derivative spectra in the 5000-4500 cm^{-1} region for DMSO and Boc-Gly-OH in DMSO with concentrations of 50, 100, 150, and 200 mM. (B) The background subtracted second derivative spectra calculated by defining DMSO second derivatives as background.

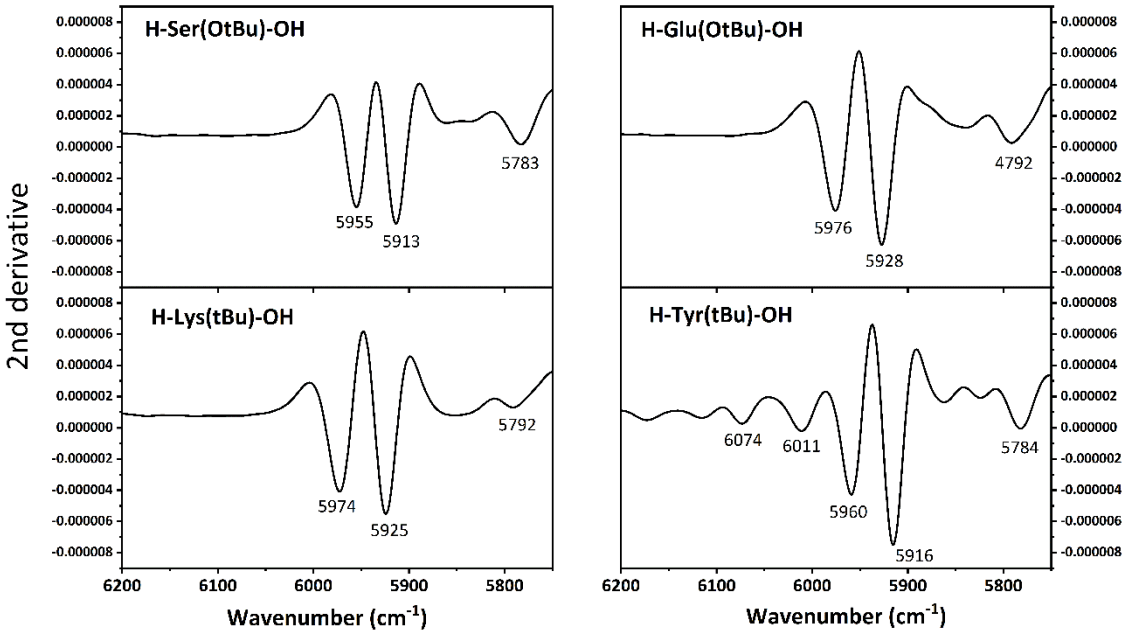
(A)


Absorbance
Wavenumber (cm-1)

(B)


H-Ser(OtBu)-OH
4416
4346
4241
H-Glu(OtBu)-OH
4463
4430
4378
4350
4246
H-Lys(tBu)-OH
4942
4726
4427
4371
4240
H-Tyr(tBu)-OH
4942
4726
4676
4622
4566
4418
4350
4243
2nd derivative
Wavenumber (cm-1)

(C)


H-Ser(OtBu)-OH
5955
5913
5783
H-Glu(OtBu)-OH
5976
5928
4792
H-Lys(tBu)-OH
5974
5925
5792
H-Tyr(tBu)-OH
6074
6011
5960
5916
5784
2nd derivative
Wavenumber (cm-1)

(D)

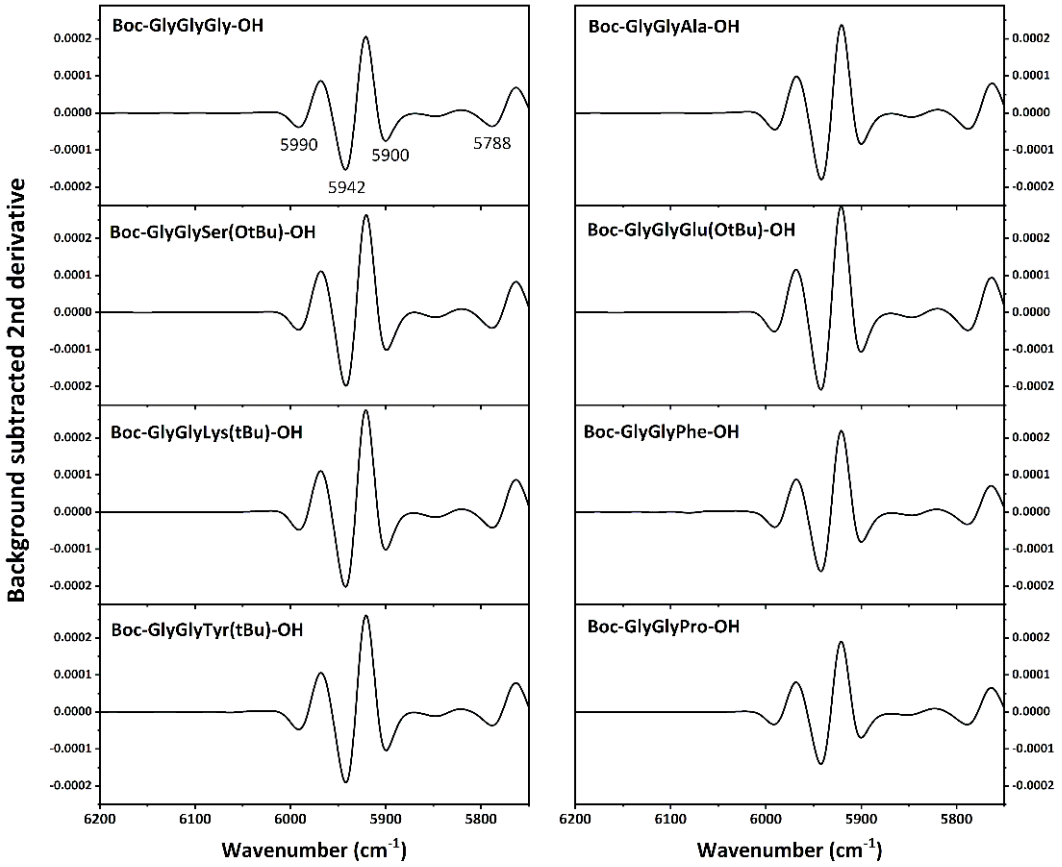

Boc-GlyGlyGly-OH
5990
5942
5900
5788
Boc-GlyGlyAla-OH
Boc-GlyGlySer(OtBu)-OH
Boc-GlyGlyGlu(OtBu)-OH
Boc-GlyGlyLys(tBu)-OH
Boc-GlyGlyPhe-OH
Boc-GlyGlyTyr(tBu)-OH
Boc-GlyGlyPro-OH
Background subtracted 2nd derivative
Wavenumber (cm-1)

Figure S4: (A) NIR absorbance spectra in the 10000-4000 cm^{-1} region for 200 mM glycine aqueous solution. The second derivative spectra of four kinds of amino acids aqueous solutions with *t*Bu in the (B) 5000-4200 and (C) 6200-5750 cm^{-1} regions with 200 mM concentrations. (D) The second derivative spectra in the 6200-5700 cm^{-1} region for 200 mM solutions of eight amino acids in DMSO.

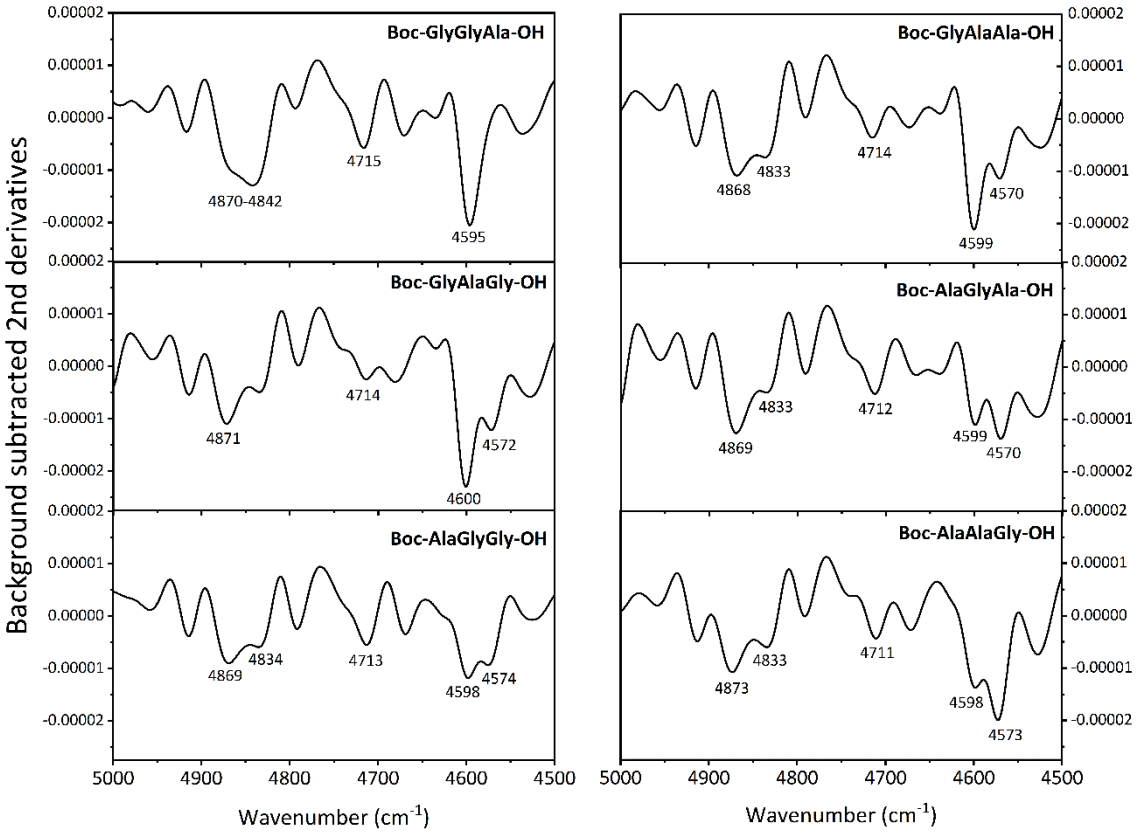


Figure S5: The background subtracted second derivative spectra in the 5000-4500 cm^{-1} region for six kinds of tripeptides composed of glycine and alanine in DMSO with 200 mM concentration.

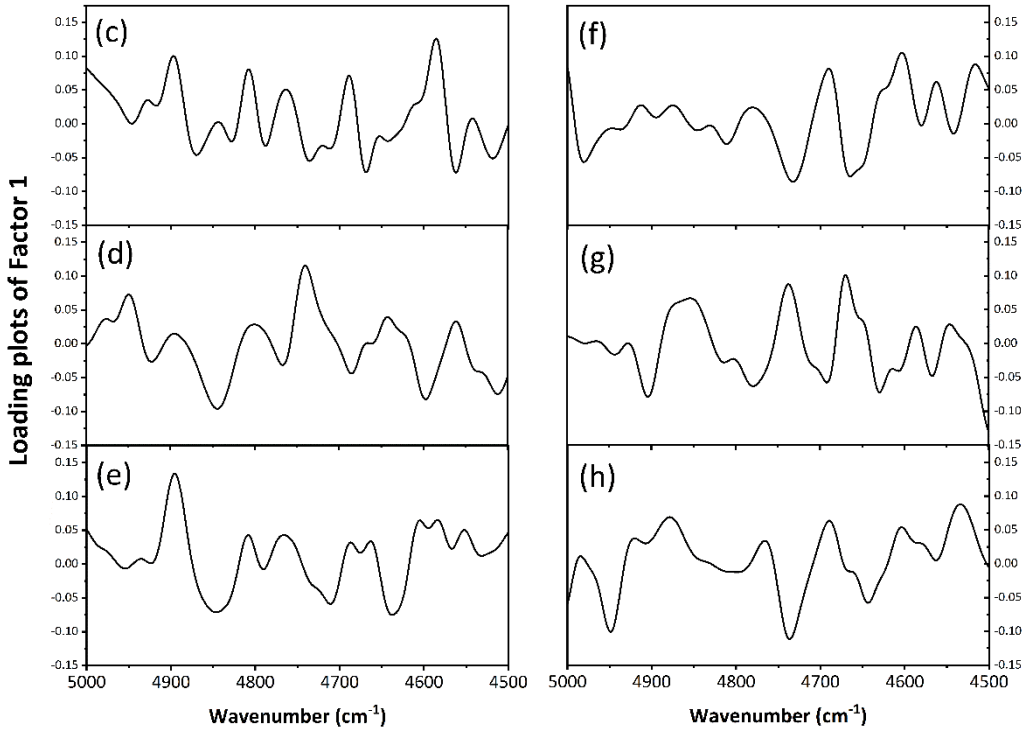


Figure S6: The PLSR loading plots of Factor 1 obtained for six tripeptides (c-h) in Table S2.