Electronic Supplementary Material (ESI) for ChemComm.

Supplementary data

Nucleobase-involved native chemical ligation: a novel reaction between an oxanine nucleobase and *N*-terminal cysteine for oligonucleotide-peptide conjugation

Eui Kyoung Jang,[a] Yohei Koike,[b] Yuko Ide,[b] Kunihiko Tajima,[b] Kenji Kanaori,[*b] and Seung Pil Pack[*a]

[a] Department of Biotechnology and Bioinformatics, Korea University, Sejong-Ro 2511, Sejong, 30019 (Republic of Korea)
[b] Faculty of Molecular Chemistry and Engineering, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto, 606–8585 (Japan)

* These are co-corresponding authors;
Tel & Fax: +81-(75)-724-7825; Email: kanaori@kit.ac.jp (for K Kanaori)
Tel: +82-(44)-860-1419; Fax: +82-(44)-860-1598; Email: spack@korea.ac.kr (for SP Pack),

Table of Contents

1. EXPERIMENTAL PROCEDURES

1.1. Sample preparation.

Oxanosine (Oxo), 2'-deoxyoxanosine (dOxo), and oxanosine triphosphate (OTP) were synthesized and purified using reverse-phase HPLC (RP-HPLC).[1,3] The concentrations of dOxo, Oxo, and OTP were determined by measuring the sample light intensity with the molar lighting coefficient of Oxa (ε_{260} = 5100 $M^{-1}cm^{-1}$) using a UV-visible absorption apparatus Bio-630 manufactured by the JASCO Corporation.[2] The DNA oligomer, $d(T_5GT_6)$, was purchased from Hokaido System Science Co. Ltd., and modified by HNO_2 to $d(T_5OT_6)$ and purified as previously reported (Figure S11).[1] The final yield was 15-18%. The peptide, CRAEYS, used for the reaction analysis was solid-phase synthesized (Figure S4). For the DNA-peptide conjugation under aerobic conditions, *N*-terminal cysteinyl peptide (C.TAT; CGRKKRRQRRRPQ, PI 12.30) was purchased from Peptron Inc. (Daejeon, Republic of Korea). Other reagents were purchased from Wako Pure Chemical Industries, Ltd. and used without purification. For end-to-end conjugation (Section 1.4), ODN (/FAM/-CTCAGGTCGACAGTCTGCGG, 20 nt) was prepared by solid-phase chemical synthesis (Integrated DNA Technologies Co., Coralville, IA, USA). Oxa-labeling at the 3'-end of ODN was conducted using terminal deoxynucleotidyl transferase (TdT) that was purchased from New England Biolabs (M0315, NEB, Ipswich, MA, USA).

1.2. Secondary reaction rate constant analysis

Under second-order reaction rate conditions, the formation rates of Oxo and Cys are proportional to the instantaneous concentration of each reactant (Equation 1).

$$\frac{d[\mathrm{Oxo}-N-\mathrm{Cys}]}{dt} = k_{\mathrm{obs}}[\mathrm{Oxo}]_t[\mathrm{Cys}]_t \quad \text{(Eq. 1)}$$

By separating the variables of Equation 1 and integrating from t = 0 min to t = t min, Equation 2 is obtained. The initial concentrations of Oxo and Cys are $[Oxo]_0$ and $[Cys]_0$, respectively, while the concentrations of Oxo and Cys at time, t, are $[Oxo]_t$ and $[Cys]_t$, respectively.

$$k_{\mathrm{obs}}\mathbf{t} = \frac{1}{[\mathrm{Oxo}]_0-[\mathrm{Cys}]_0} ln \frac{[\mathrm{Cys}]_0[\mathrm{Oxo}]_t}{[\mathrm{Oxo}]_0[\mathrm{Cys}]_t} \quad \text{(Eq. 2)}$$

If $[Oxo]_0 = [Cys]_0$, the integrated rate equation is shown as follows:

$$\frac{1}{[\mathrm{Oxo}]_t} = k_{obs}t + \frac{1}{[\mathrm{Oxo}]_0} \quad \text{(Eq. 3)}$$

1.3. Reaction of Oxa with a peptide containing an N-terminal Cys residue

A peptide (sequence: CRAEYS) containing an N-terminal Cys residue and dOxo were reacted in 100 mM phosphate buffer (pH 7.4). The reaction was carried out in an NMR sample tube. For the polymer-to-polymer reaction, 5 mM d(T_5OT_6) and 5 mM CRAEYS (pH 7.4) were incubated at 320 K using a thermal cycler with nitrogen purging

1.4. Nucleobase-involved native chemical ligation (NbCL) under aerobic conditions

The optimized NbCL conditions were as follows: 4 μM Oxa-labeled DNA oligomer, 200 μM *N*-terminal cysteinyl peptide, and 4 mM TCEP were added to 50 mM MES (pH 6.4). The mixture was incubated at 310 K for 2 h in a PCR chamber. In this model, 21 mer of oxanine-labelled oligodeoxynucleotides (ODN-O1; FITC-CTCAGGTCGACAGTCTGCGG**O**, where **O** is Oxa) and *N*-terminal cysteinyl peptide (C.TAT; **C**GRKKRRQRRRPQ, PI 12.30) were used. The ODN-O1 were prepared by enzymatic synthesis.[3]

The final conjugation product could be purified by conventional PAGE extraction for further use. Briefly, the conjugation product was eluted from sliced and crushed gel by diffusion in ~10X volume of storage buffer (50 mM sodium phosphate, 150 mM NaCl). Next, the eluent was filtered using a syringe filter (06543-04, Nacalai Tesque, Inc, Kyoto, Japan), then it was concentrated and washed out with storage buffer using a 3-kDa centrifugal filter (Amicon UFC5003, Darmstadt, Germany).

1.5. Enzymatic synthesis of Oxa-labeled ODN

OTP preparation

OTP was prepared for supplying a substrate of TdT in the Oxa-labelling process for the target ODN. The conversion of GTP to OTP has been referred to in previous reports (3). Briefly, 10 mg of GTP was dissolved in 3.6 ml of 1 M acetate buffer (pH 3.7), then preheated for 5 min at 37°C. After the addition of 0.4 ml of 1 M $NaNO_2$ (final concentration: 100 mM) to the mixture, it was incubated for 4 h at 37°C. The mixture was neutralized by adding NaOH to stop the reaction. OTP was purified by HPLC (Young Lin Co., Korea, YL9100 System,

Column: ULTRON VX-ODS 150 × 6.0 mm, 5 m; gradient buffer system: 0% acetonitrile in 100 mM triethylammonium acetate (TEAA) at 0 min to 20% acetonitrile in 100 mM TEAA at 20 min, 1 ml/min flow rate) and then lyophilized. The final yield was 10-12%.

Oxa-labelling at 3'-end of ODN using TdT

We followed our previous report describing the TdT-in-TB method for enzymatic Oxa-labelling process (3). All components, including the initiator (ODN, 3 μM), $MgCl_2$ (final: 10 mM), $CoCl_2$ (final: 0.25 mM), TdT (final: 0.8 U/μL), and OTP (final: 0.2 mM) were mixed in the reaction buffer [100 mM Tris and 200 mM borate, pH 8.0]. The mixture was incubated at 37°C for 60 min. To stop the reaction, 4 μL of 0.3 M EDTA was added per 25 μL of a reaction mixture, then heated at 90°C for 10 min. The reaction product was concentrated by a 3-kDa MWCO centrifugal filter (Amicon™, Merck KGaA, Darmstadt, Germany) and was rid of residual substrates by a desalting column (Centri-Sep™, Thermo Fisher Scientific Inc., Waltham, MA, USA).

1.6. NMR and MS measurements.

The NMR experiments were carried out on a Bruker AV–600 spectrometer (600.19 MHz for ^{1}H). A set of two-dimensional (2D) NMR spectroscopy experiments were performed. The NMR spectrum in 90% H_2O/10% D_2O was measured using the AV-600 system (600.19 MHz for ^{1}H) manufactured by Bruker. The internal sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d4* was used for reference of ^{1}H chemical shifts. Signals were identified using ^{1}H-^{1}H TOCSY, ROESY, and ^{1}H-^{13}C HMBC. Electrospray ionization and time-of-flight mass (ESI-TOF-MS) spectra were measured using a Bruker micrOTOF device. The measurement solvent used was 50% H_2O/ 50% MeOH, and the measurement sample was directly injected into the MS system at a flow rate of 800 μL/min. Spectrometric data obtained from the analysis of reaction products is described in Supplementary Data, Section 1.

2. Spectrometric data of NMR and MS measurements

Spectrometric data of 1 (Oxo-*S*-Cys). ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 310 K): δ 8.07 (s, 1H, H-2). 5.78(d, 1H, H-1'). 4.55(t, 1H, H-2'). 4.38 (t, 1H, H-3'). 4.20 (dd, 1H, H-4'). 3.94 (dd, 1H, H-5'). 3.86 (dd, 1H, H-5'). 4.09(dd,1H, αH). 3.74(d, 1H, βH). 3.49(d, 1H, βH). ESI-TOF MS(negative): *m*/*z* 404.1 $[M-H]^-$

Spectrometric data of 2 (Oxo-*N*-Cys). ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 310 K): δ 8.03(s, 1H, H-2), 5.75(d, 1H, H-1'), 4.55(t, 1H, H-2'). 4.38 (t, 1H, H-3'). 4.20 (dd, 1H, H-4'). 3.94 (dd, 1H, H-5'). 3.86 (dd, 1H, H-5'). 4.62(dd,1H, αH). 3.18(d, βH). ESI-TOF MS(negative): *m*/*z* 404.1 $[M-H]^-$

Spectrometric data of 3 (Oxo-*N*-Cys *disulfide*). ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 310 K): δ 7.79 (s, 1H, H-2). 5.73 (d, 1H, H-1'). 4.53 (t, 1H, H-3'). 4.37 (t, 1H, H-4'). 4.28 (dd, 1H, H-4'), 3.94 (dd, 1H, H-5'). 3.83 (dd, 1H, H-5'). 4.66 (dd, 1H, αH). 3.40 (dd, 1H, βH). 3.18 (dd, 1H, βH). ESI-TOF MS(negative): *m*/*z* 403.1 $[M-2H]^{2-}$

Spectrometric data of 4 (Oxo-*N*-Cys *thioester*).[*1)] ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 310 K): δ 8.02(s, 1H, H-2). 7.97 (s, 1H, H-2). 5.74 (d, 1H, H-1'). 5.74 (d, 1H, H-1'). 4.53 (t, 1H, H-2'), 4.44 (t, 1H, H-3'). 4.37 (t, 1H, H-2'). 4.26 (t, 1H, H-3'). 4.18 (dd, 1H, H-4'). 3.94 (dd, 1H, H-5'). 3.83 (dd, 1H, H-5'). 3.79 (dd, 1H, βH). 3.48 (dd, 1H, βH).

dOxo –*S*-Cys. ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 283 K): δ 7.73(s, 1H, H-2), 5.83(t, 1H, H-1'), 4.30 (m, 2H, H-3', αH), 3.78(dd, 1H, H-4'), 3.53(dd, 1H, H-5'), 3.44(dd, 1H, H-5'), 2.78(d, 1H, βH), 2.35(m,1H, H-2'), 2.25(m, 1H, H-2').

dOxo –*N*-Cys. ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 283 K): δ 7.68(s, 1H, H-2), 5.80(t, 1H, H-1'), 4.30 (m, 2H, H-3', αH), 3.78(dd, 1H, H-4'), 3.53(dd, 1H, H-5'), 3.44(dd, 1H, H-5'), 2.78(d, 1H, βH), 2.35(m,1H, H-2'), 2.25(m, 1H, H-2'). ESI-TOF-MS(negative): *m*/*z* 404.1 $[M-H]^-$

Spectrometric data of dOxo-CRAEYS. ^{1}H-NMR (600 MHz, 90% H_2O/10% D_2O at 288 K): δ 7.92 (s, 1H, H-2). 6.03 (t, 1H, H-1'), 4.51 (dt, 1H, H-3'). 4.51 (dd, 1H, H-4'). 3.75 (dd, 1H, H-5'). 3.67 (dd, 1H, H-5'). 2.54 (dd, 1H, H-2'). 2.47 (dd, 1H, H-2'). 8.54 (d, 1H, NH in Arg). 8.38 (d, 2H, NH in Ala and Glu). 8.15 (d, 1H, NH in Cys). 8.09 (d, 1H, NH in Tyr). 7.81 (d, 1H, NH in Ser), 7.15 (t, 1H, $NH_2C(=NH)NH$- in Arg). 4.58 (t, 1H, αH in Cys). 4.56 (t, 1H, αH in Tyr). 4.24 (t, 1H, αH in Arg). 4.15 (t, 1H, αH in Ser). 4.08 (t, 1H, αH in Ala and Glu). 3.74 (d, 2H, βH in Ser). 3.06 (m, 2H, δH in Arg). 3.05(dd, 1H, βH in Tyr). 2.93(d, 2H, βH in Cys). 2.80(dd, 1H, βH in Tyr). 2.01(m, 2H, γH in Glu). 1.75(m, 1H, γH in Arg). 1.72 (m, 2H, βH in Glu). 1.64(m, 1H, γH in Arg). 1.52(m, 2H, βH in Arg). 1.24(d, 3H, βH in Ala). ESI-TOF-MS (negative): *m*/*z* 994.4 $[M-1]^-$

CRAEYS. ESI-TOF-MS (negative): *m/z* 724.5 $[M-H]^-$

d(T_5OT_6). ESI-TOF-MS (negative): *m/z* 3614.0 $[M-4H]^{4-}$

d(T_5OT_6)-Cys. ESI-TOF-MS (negative): *m/z* 932.8 $[M-4H]^{4-}$

d(T_5OT_6)-Cys-S-S-Cys. ESI-TOF-MS (negative): *m/z* 962.6 $[M-4H]^{4-}$

d(T_5OT_6)-CRAEYS. ESI-TOF-MS (negative): *m/z* 1084.5 $[M-4H]^{4-}$

d(T_5OT_6)-CRAEYS-S-S-CRAEYS. ESI-TOF-MS (negative): *m/z* 1265.9 $[M-4H]^{4-}$

[*1)] As the peak of the αH of Oxo-*N*-Cys thioester overlaps with the peak of the measuring solvent, it was impossible to determine an accurate chemical shift.

3. Supporting Figures

3.1. Identification of the products from the reaction of Oxo and Cys

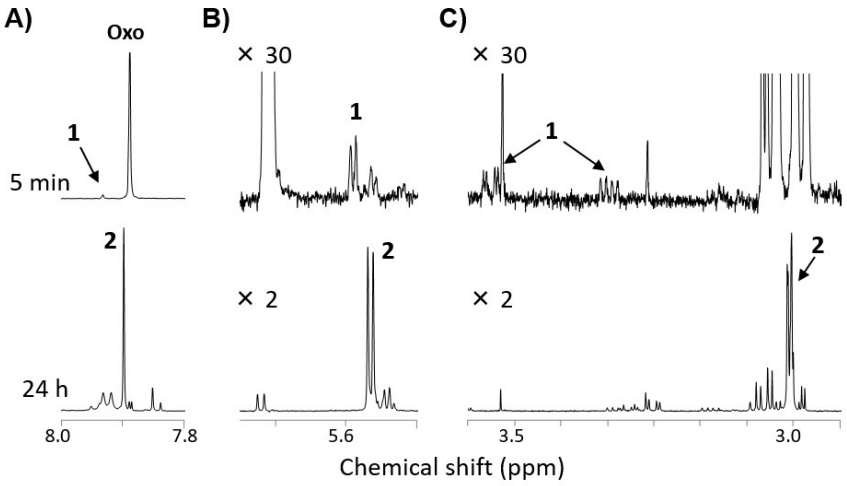


Figure S1. Time courses of a reaction of 10 mM Oxo with 10 mM Cys at 298 K and pH 7.4 at 1 atm N_2 pressure. The spectral regions are A) Oxo H-2, B) Oxo H-1', and C) Cys βH.

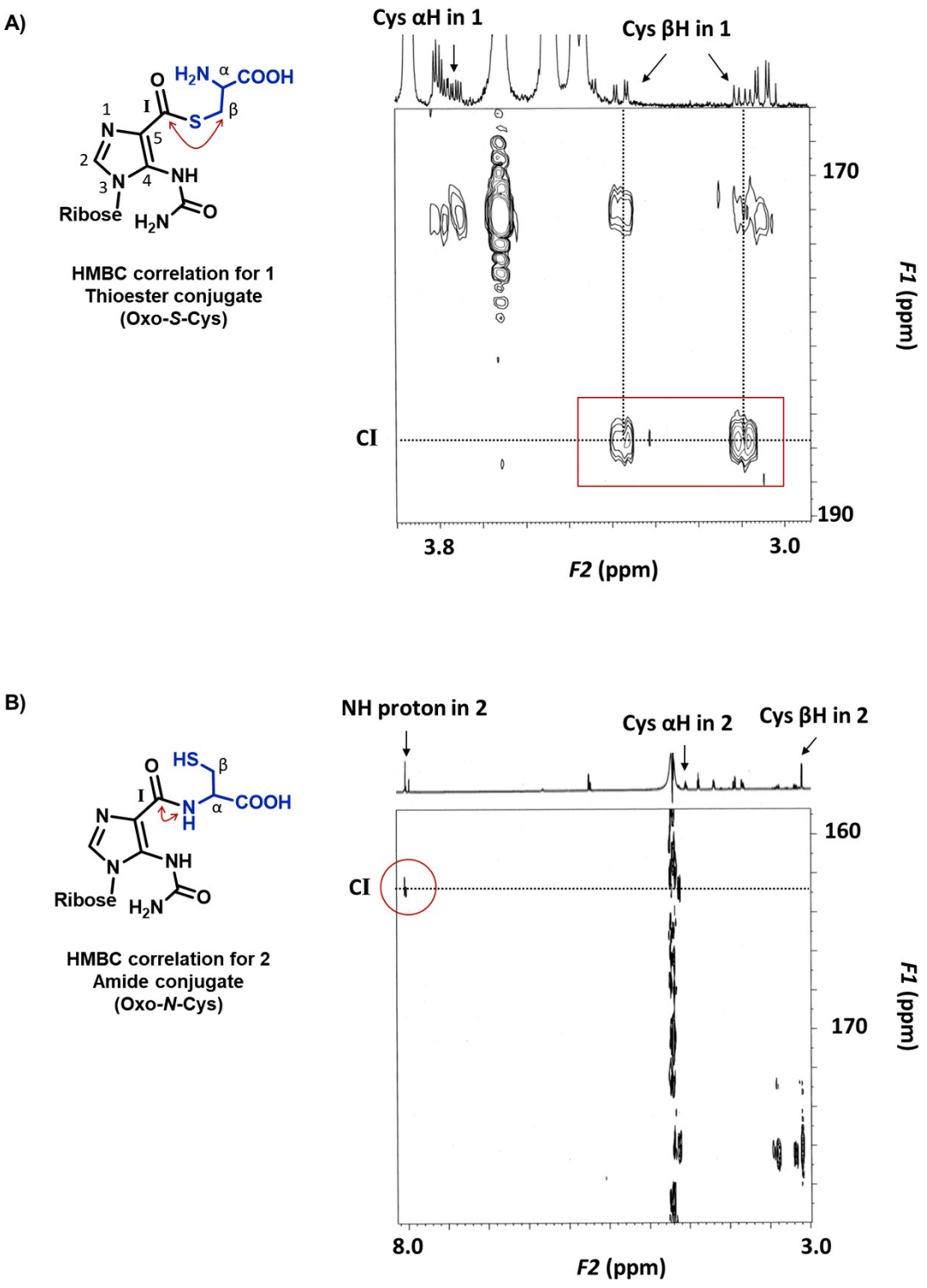
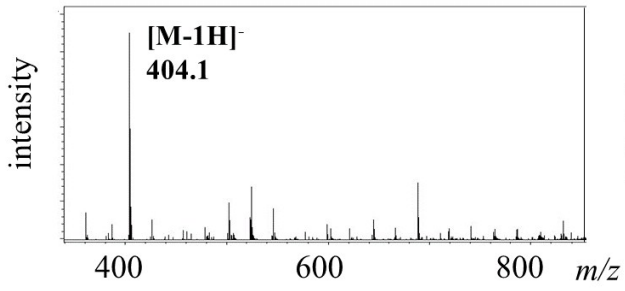


Figure S2. A) ^{1}H-^{13}C HMBC spectrum of compound **1**. Owing to the difficulty in isolating **1**, the result was obtained with the reaction proceeding at a low temperature (278 K) in 90% H_2O/10% D_2O; B) HMBC spectrum of **2** formed by 20 mM Oxo and 50 mM Cys at 310 K and pH 7.4.



	Theoretical MW
Oxo	284.2
Cysteine	121.0
Oxo-Cys Conjugate ($C_{13}H_{19}N_5O_8S$)	405.2

Figure S3. ESI-TOF MS spectrum (negative mode) of compound **2**. The molecular weight of 405.1 was presumed to be deprotonations from a compound (molecular weight: 405.2) that was formed after Oxo (molecular weight: 284.2) and Cys (molecular weight: 121.0) reacted at a ratio of 1:1.

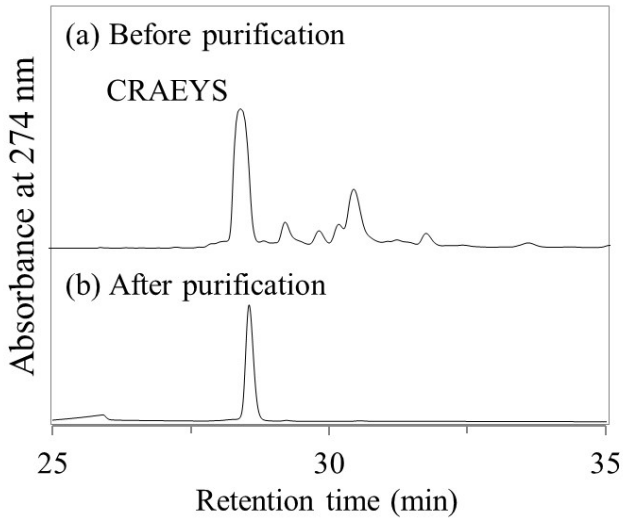


Figure S4. HPLC chromatogram of the peptide (CRAEYS) at 274 nm (a) before purification and (b) after purification. Peptide synthesis was performed using standard F-moc chemistry. The peptide concentration of CRAEYS was determined both by weight and absorption of tyrosine at 280 nm using the molar absorption coefficient 1490 $M^{-1}cm^{-1}$.

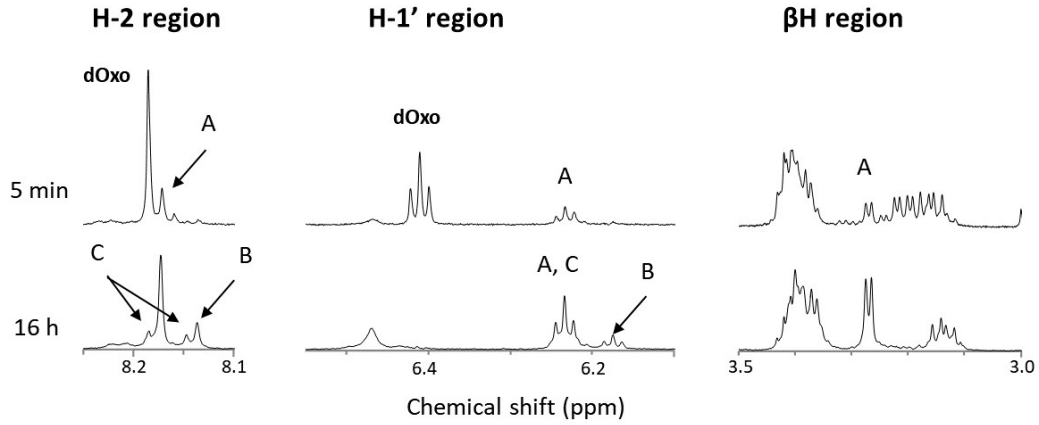


Figure S5. Time courses of the ^{1}H-NMR spectra of a mixture of 5 mM dOxo and 5 mM CRAEYS at pH 7.4, 310 K, where A represents the amide conjugate of the dOxo-*N*-CRAEYS with a free SH group of Cys, B represents the disulfide compounds composed by two dOxo-*N*-CRAEYS, and C represents the dOxo-*N*-Cys(-*S*-Oxo)RAEYS that is formed by the reaction of the SH group of dOxo-*N*-CRAEYS with dOxo. Two dOxo molecules are bound to the *N*-terminal Cys residue with amide and thioester bonds.

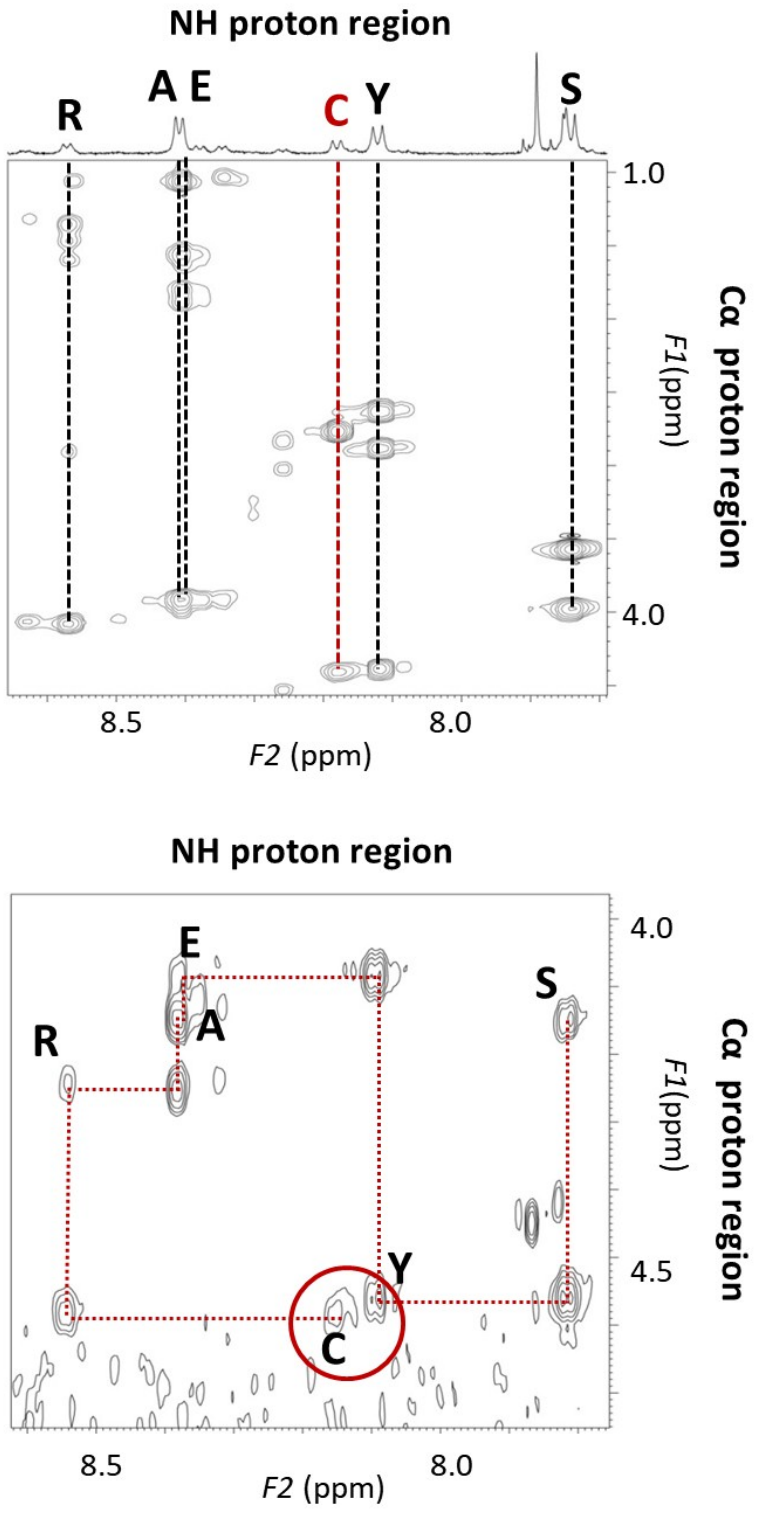
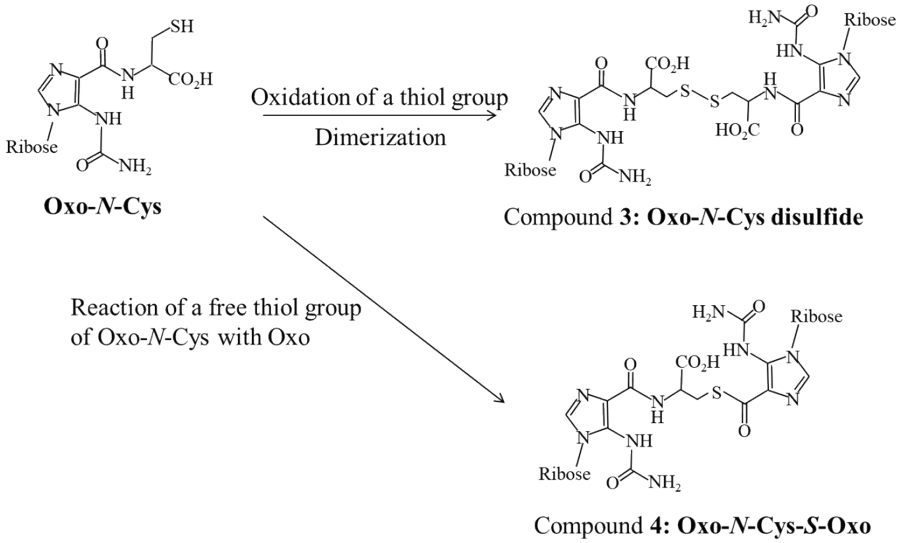


Figure S6. TOCSY and ROESY spectra of the product of dOxo and CRAEYS at 288 K. The dOxo-CRAEYS conjugate is formed with the amino group of Cys changing to the amide group, and then the amide NH proton of the Cys residue can be observed in the 1H-NMR spectra, including TOCSY and ROESY. The cross-peak related to the Cys NH proton is indicated by a red circle.

3.2. Identification of by-products from the successive reaction of Oxo-N-Cys

To investigate the side reaction of Oxa with the N-terminal cysteine in detail, the reaction products under the conditions of excess Oxo and in the presence of air were analyzed (Scheme S1). Excess Oxo (10 mM Oxo and 3 mM Cys) was mixed with a 50 mM phosphate buffer solution (pH 7.4) in an air atmosphere, reacted at 310 K, and the ^{1}H-NMR spectra were measured (Figure S7, Table S3). When the reaction time exceeded 1 h, the intensity of the signal derived from Oxo-*N*-Cys gradually decreased, and signals for new compounds (compounds **3** and **4**) appeared. Signals of the Cys $C_{\beta}H$ of **3** were observed at 3.40 and 3.18 ppm, while signals of Cys $C_{\beta}H$ of **4** were observed at 3.79 and 3.48 ppm. Signals of all Cys and Oxo-*N*-Cys disappeared completely after 40 h, and those of **3** and **4** were observed. It was suggested that **3** or **4** are structures in which multiple Oxo bonds are bonded per molecule of Cys because the ^{1}H-NMR spectra after 40 h show that the ratio of Oxo signals to the signals of **3** and **4** is almost one. The ^{13}C-NMR chemical shift of **3** almost corresponds with that of Oxo-*N*-Cys, suggesting that **3** contains the same structure as Oxo-*N*-Cys. The $C_{\beta}H$ signal of Oxo-*N*-Cys was observed at 3.18 ppm, whereas that of $C_{\beta}H$ of **3** was observed at two locations, 3.40 ppm and 3.18 ppm. The ESI-TOF MS spectra (negative mode) of **3** showed a signal at *m/z* 403.1 (Figure S8). The molecular weight of 403.1 indicated by this signal was presumed to be two deprotonations from the dimer of Oxo-*N*-Cys (molecular weight: 808.7).

As for compound **4**, according to the ^{1}H-^{1}H TOCSY measurement, **4** contains an amide proton in its structure. Additionally, as the chemical shifts of the methylene protons (3.79 and 3.48 ppm) of the cysteine residues of **4** were observed in a lower magnetic field than the chemical shifts (3.40 and 3.18 ppm) of the methylene protons of **3**, **4** is free of Oxo-*N*-Cys. There is a high possibility that Oxa is bonded to the thiol group of (Oxo-*N*-Cys-*S*-Oxo).



Scheme S1. By-products from the successive reaction of Oxo-N-Cys

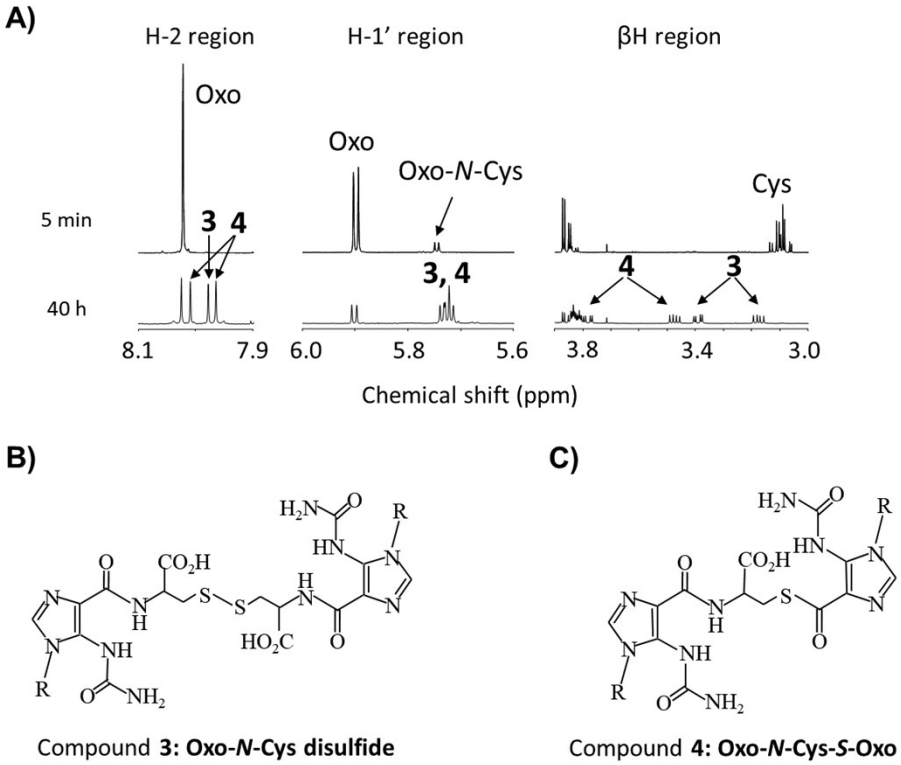


Figure S7. A) Time courses of a reaction of 10 mM Oxo with 3 mM Cys at 298 K and pH 7.4 under atmospheric pressure conditions. B) and C) possible structures of the by-products.

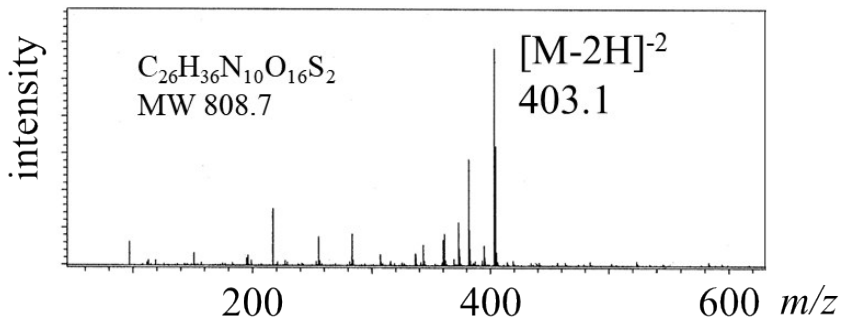


Figure S8. ESI-TOF MS spectrum (negative mode) of compound **3.**

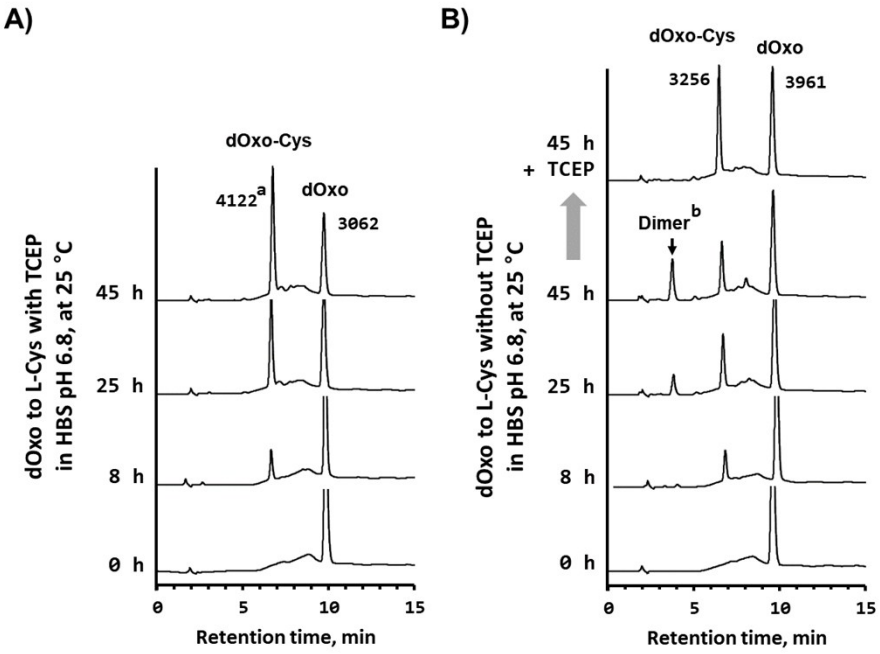


Figure S9. Chromatograph of dOxo-Cys reaction under aerobic conditions: A) with or B) without TCEP. Reaction of 0.5 mM dOxo and 1 mM Cys with 5 mM TCEP in 200 mM Hepes (pH 6.8). **a** Integrated area of the specific peak (mV*min); **b** Dimer inferred as dOxo-Cys-Cys-dOxo.

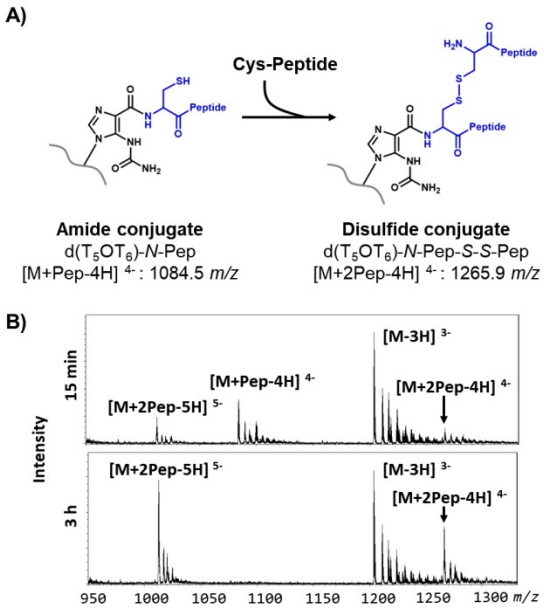


Figure S10. ESI-TOF (negative mode) spectra of the reaction products of $d(T_5OT_6)$ with CRAEYS after A) 15 min and B) 3 h. M: $d(T_5OT_6)$, Pep: CRAEYS

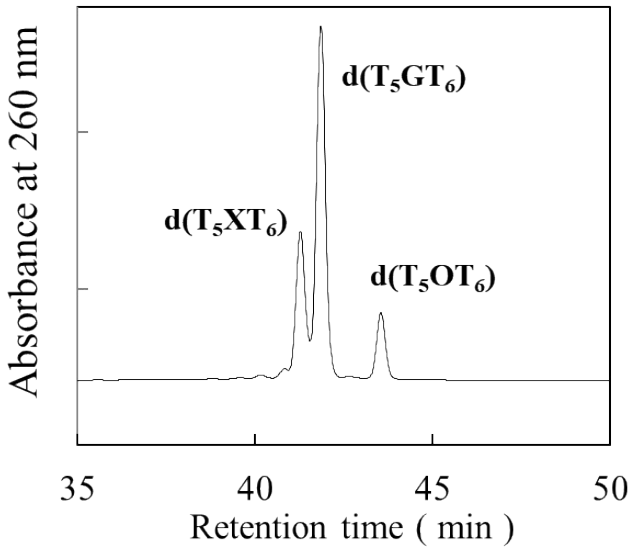
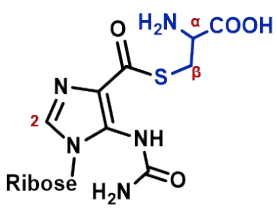
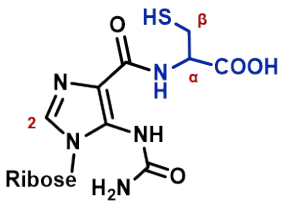


Figure S11. RP-HPLC chromatogram of $d(T_5OT_6)$. Gradient, 10%-20% acetonitrile (linear, 0-60 min) in 100 mM triethylammonium acetate (TEAA) buffer (pH 7.4); Temperature, 40°C. The peak of $d(T_5OT_6)$ was collected and lyophilized. The coefficient value of 9.4×10^4 was used for the calculation of concentration of the $d(T_5OT_6)$ oligomer.[5]

4. Supporting Tables



Compound **1**
Oxo-*S*-Cys



Compound **2**
Oxo-*N*-Cys

Table S1. ^{1}H-NMR chemical shifts of **1** and **2** at pH 7.4 and 310 K in 90% H_2O/10% D_2O

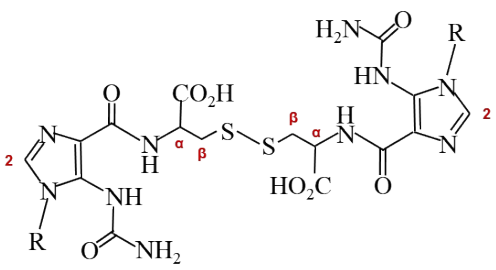
Compound	Chemical shifts (ppm)							
	Base		Ribose				Cysteine	
	H2	H1'	H2'	H3'	H4'	H5'	αH	βH
Compound **1** (Oxo-*S*-Cys)	8.07	5.78	4.55	4.38	4.20	3.94, 3.86	4.09	3.74, 3.49
Compound **2** (Oxo-*N*-Cys)	8.03	5.75					4.62	3.18
Oxa-*S*-GSH[*1)]	7.66						4.71	3.60, 3.33
Oxa-*N*-GSH[*2)]	8.62						4.58	2.93

[*1)] Chemical shifts of Oxa-*S*-GSH were obtained in D_2O[4]

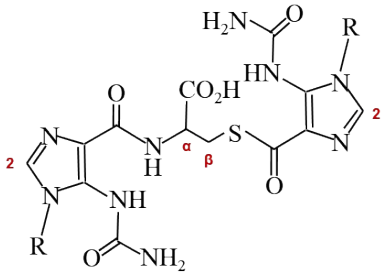
[*2)] Chemical shifts of Oxa-*N*-GSH were obtained in 1% TFA/D_2O solution[4]

Table S2. Second-order product rate constants of Oxa and amino acids at pH 7.4 and 310 K, as well as the p*K*a values of nucleophilic side chains

Compound	k_{obs} / $mM^{-1}min^{-1}$	p*K*a
Oxo-*N*-Cys	3.8×10^{-3}	8.15
dOxo-*N*-Cys	3.2×10^{-3}	8.15
dOxo-*N*-CRAEYS	2.0×10^{-3}	-
dOxo-*S*-GSH	3.1×10^{-3}	8.56
Oxo-*S*-(*N*-acetyl Cys)	4.2×10^{-4}	9.52
Oxo-*N*-Ala	2.0×10^{-5}	10.06



Compound **3**
Oxo-*N*-Cys disulfide



Compound **4**
Oxo-*N*-Cys-*S*-Oxo

Table S3. ^{1}H-NMR chemical shifts of **3** and **4** at pH 7.4 and 310 K in 90% H_2O/10% D_2O

Compound	Chemical shifts (ppm)							
	Base	Ribose					Cysteine	
	H2	H1'	H2'	H3'	H4'	H5'	αH	βH
Compound **3** (Oxo-*N*-Cys disulfide)	7.79	5.73	4.53	4.37	4.28	3.94, 3.83	4.66	3.40, 3.18
Compound **4** (Oxo-*N*-Cys thioester)	8.02, 7.97	5.74, 5.72		4.44, 4.37	4.26, 4.18		-	3.79, 3.48

References

(1) S. P. Pack, M. Nonogawa, T. Kodaki, K. Makino, *Nucleic acids research,* 2005, **33**, 5771.
(2) T. Suzuki, M. Yamada, H. Ide, K. Kanaori, K. Tajima, T. Morii, K. Makino, *Chemical research in toxicology,* 2000, **13**, 227.
(3) E. K. Jang, R. G. Son, S. P. Pack, *Nucleic acids research,* 2019, **47**, e102.
(4) H. J. Chen, C. J. Hsieh, L. C. Shen, C. M. Chang, *Biochemistry,* 2007, **46**, 3952.
(5) T. Suzuki, Y. Matsumura, H. Ide, K. Kanaori, K. Tajima, & K. Makino, *Biochemistry,* 1997. **36**, 8013-8019.