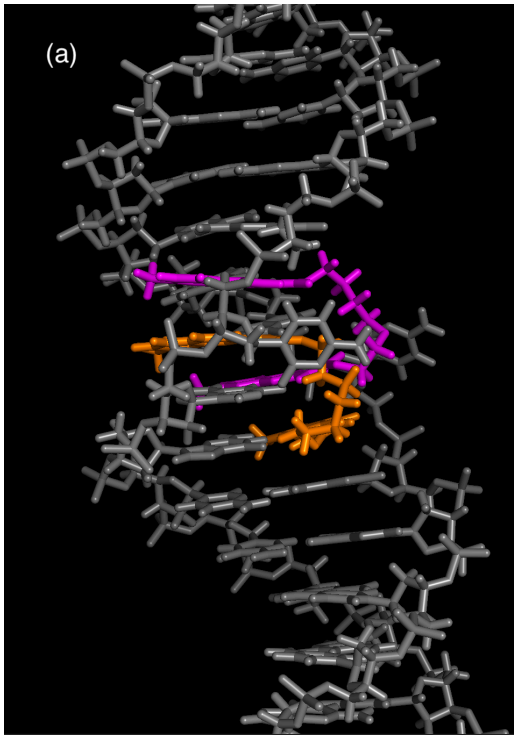
Supporting Information for

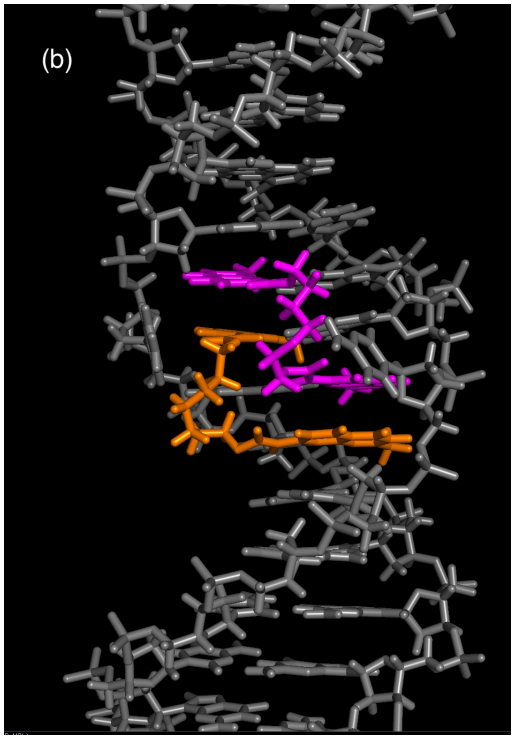
CGG repeat DNA assisted dimerization of CGG/CGG binding molecule through intermolecular disulfide formation

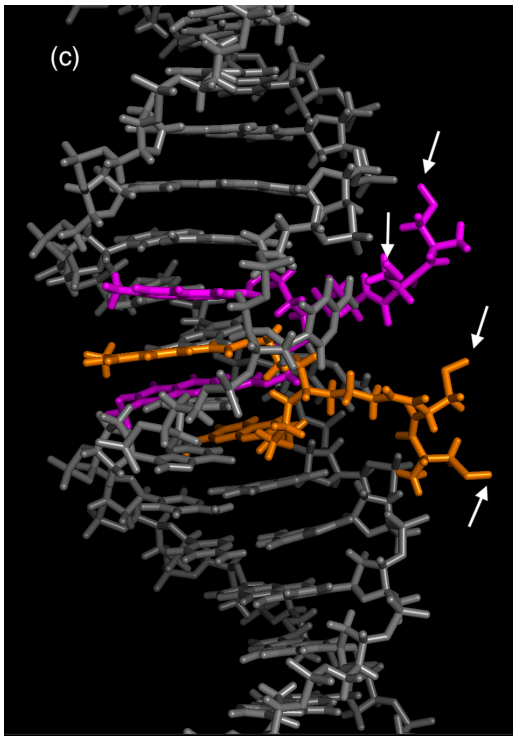
Takeshi Yamada,[a] Shouta Miki, Lu Ni, and Kazuhiko Nakatani[a]

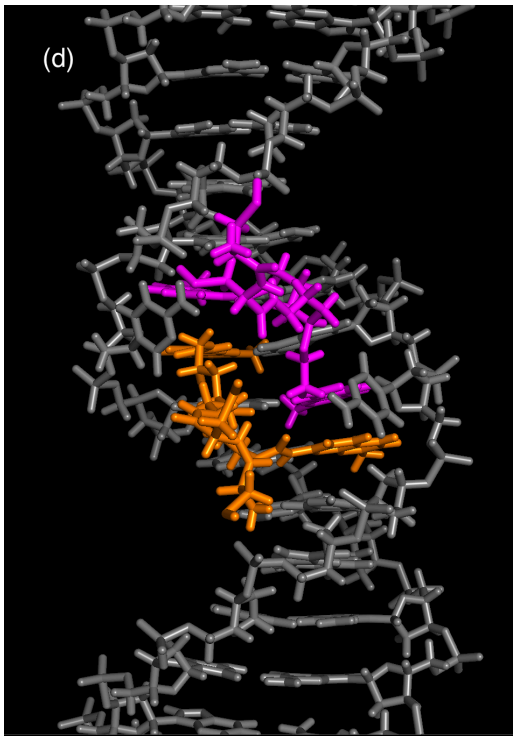
[a] *Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (ISIR), Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047. JAPAN*

Supporting Figures









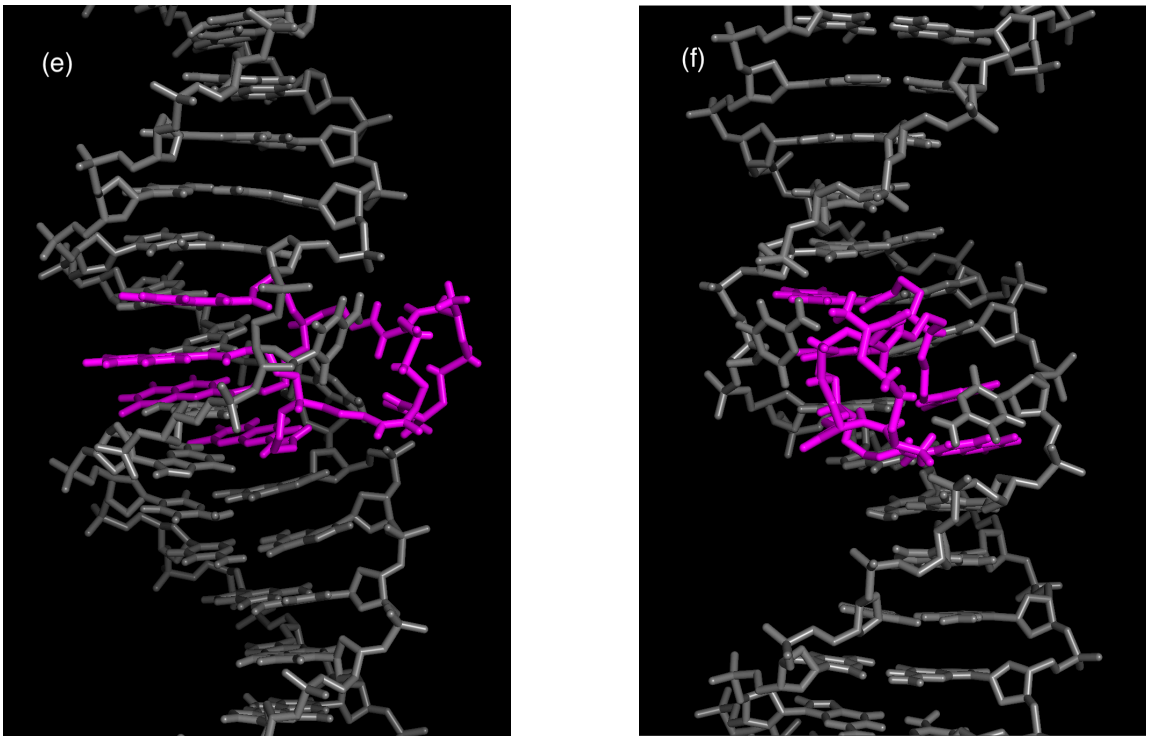


Figure S1. Molecular modeling simulation of the complex between a double strand DNA containing a CGG/CGG sequence with (a and b) two **NCD**s, (c and d) two **NCD-CC**s, and (e and f) one compound **5**. Each complex was shown in side views (a, c, and e) and view toward major groove (b, d, and f). Two ligand molecules in the complex were shown in magenta and orange colors. All structures were energy optimized by AMBER* force field in water using Maestro ver. 11.2. The geometry of **NCD-CC** in S1c and S1d indicated that binding of the **NCD** domain leave the cysteinylcystein moiety in the major groove, and thiol groups of two **NCD-CC**s exist in close vicinity (indicated by white arrows). The molecular modeling simulation of the complex of compound **5** and the CGG/CGG DNA (Figure S1e and S1f) indicated that the cyclopeptide moiety of compound **5** would restrict conformational flexibility of its naphthyridine moieties leading to strong affinity of compound **5** to G-G mismatch DNA.

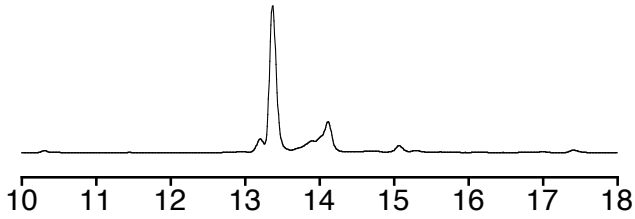


Figure S2. RP-HPLC profile of crude **NCD-CC**. The largest peak at 13.5 min was identified as **NCD-CC** by ESI-TOF-mass. HPLC condition: an endcappped C18 column, 0.1%TFA and MeCN were eluted 1 mL/min at room temperature. The percentage of MeCN was changed from 0 to 40% gradually for 20 min.

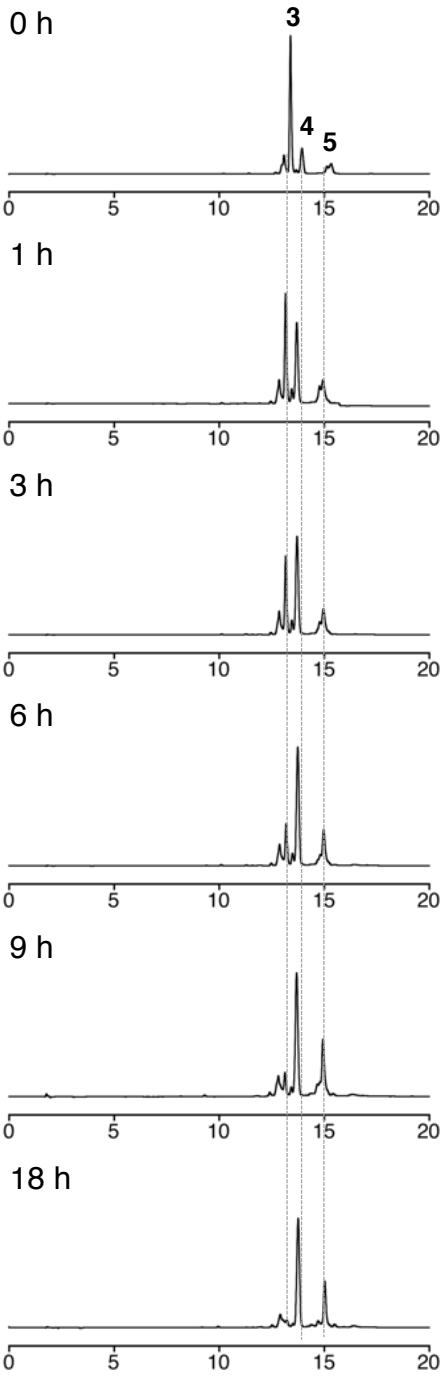


Figure S3. RP-HPLC profiles of NCD-CC (40 μM) in the presence of 5'-(TCAA CGG TTGA)-3'/3'-(AGTT GGC AACT)-5' (5 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM) after the indicated hours; RP-HPLC condition: an end-capped C18 column, 0.1% TFA and acetonitrile were eluted 1 mL/min at 40 °C. The percentage of acetonitrile was changed gradually from 0 to 20 % in 10 minutes, and then 20 to 40% in 20 minutes. Absorbance of 330 nm was monitored by a UV detector equipped to RP-HPLC.

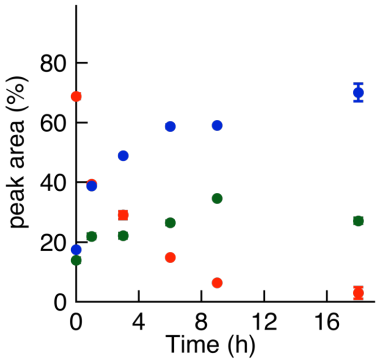


Figure S4. Time course of oxidation of **NCD-CC** in the presence of 5'-(TCAA CGG TTGA)-3'/3'-(AGTT GGC AACT)-5' (5 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM). Key: **NCD-CC**, (red); **4** (blue); and **5** (green).

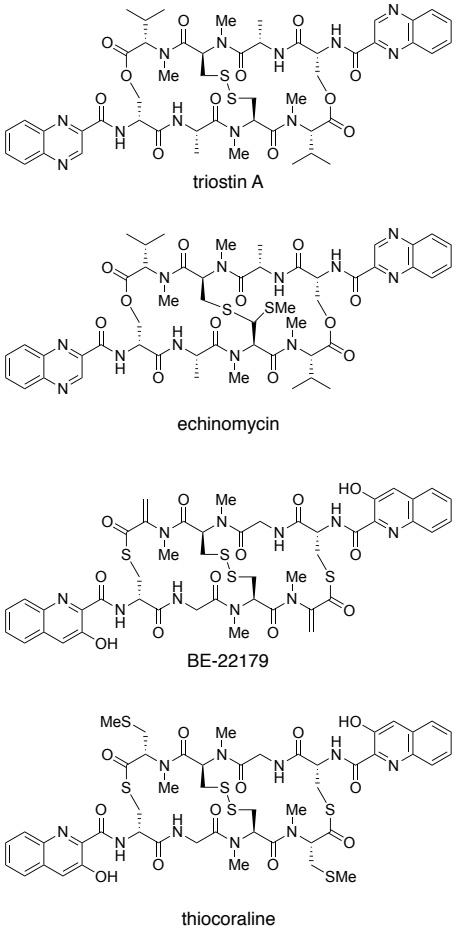


Figure S5. Chemical Structure of natural cyclic bisintercalator.

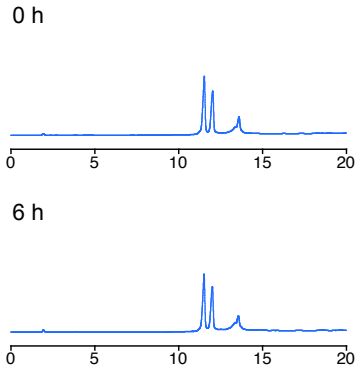


Figure S6. RP-HPLC profiles of **NCD-CC** (40 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing NaCl (100 mM), 2'-deoxyadenosine (internal standard, 20 μM) and dithioerythritol (160 μM) after the indicated hours; RP-HPLC condition: an endcappped C18 column, 0.1% TFA and MeCN were eluted 1 mL/min at 80 °C. The percentage of MeCN was changed from 0 to 40 gradually for 20 min. Absorbance of 330 nm was monitored by PDA.

Experimental section

General. Reagents and solvents were purchased from standard suppliers and used without further purification. Reactions were monitored with TLC plate silica gel 60 F_{254}. Spots on TLC were monitored with UV, phosphomolybdic acid, ninhydrin, or anisaldehyde. A C-200 Silica gel was used for silica gel flash chromatography. ^{1}H-NMR and ^{13}C-NMR spectra were measured with 600 MHz and 150 MHz NMR. The multiplicity was expressed as follows; s = singlet, d = doublet, t = triplet, m = multiplet, br = broad. The chemical shifts are expressed in ppm relative to residual solvent as an internal standard, and coupling constants (*J* values) were represented in hertz.

Molecular modeling of binding structure of compounds to the CGG/CGG triad.

Molecular modeling studies were carried out using Macromodel 11.2 (Schrodinger, LLC, NY, 2017). Initial structures were constructed manually from the NMR–structure that we determined previously. The resulting compound–DNA complexes were subjected to energy minimization using the AMBER* force field.

Preparation of NCD–CC solution

NCD–CC (1 mM) was dissolved in 0.1 M HCl to avoid oxidation during storage. The solution was neutralized with 0.1 M NaOH right before its use.

Kinetic studies of NCD–CC's oxidative dimerization.

NCD–CC (40 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing 2'-deoxyadenosine (20 μM), and NaCl (100 mM) was kept in air at 37 °C in an incubator. A 25 μL portion was took from the solution at the indicated time. Each of them was analyzed by RP–HPLC with a C18 endcapped column (4.6 mm * 150 mm) at 80 °C in a column oven. 0.1% TFA in H_2O and acetonitrile were used as eluting solvents. The percentage of acetonitrile was increased from 0 to 40% gradually in 20 minutes. The absorbance of the eluted compounds was measured by a photo diode array. The profiles of 330 nm were shown in Figure 2A because absorbance of 330 nm is characteristic for 2-amino-1,8-

naphthyridine moieties. This experiment was done twice, and the average values were plotted on Figure 3A.

Kinetic study of NCD–CC's oxidative dimerization in the presence of the CGG repeat DNA.

NCD–CC (40 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing 5'-d(CGG)$_9$-3' (5 μM), 2'-deoxyadenosine (20 μM), and NaCl (100 mM) was kept in air at 37 °C in an incubator. A 25 μL portion was took from the solution at the indicated time. Each of them was analyzed by RP–HPLC with a C18 endcapped column (4.6 mm * 150 mm) at 80 °C in a column oven. 0.1% TFA in H_2O and acetonitrile were used as eluting solvents. The percentage of acetonitrile was increased from 0 to 40% gradually in 20 minutes. The absorbance of the eluted compounds was measured by a photo diode array. The profiles of 330 nm were shown in the Figure 2B. This experiment was done twice, and average values were plotted on Figure 3B.

Kinetic study of NCD–CC's oxidative dimerization in the presence of the G-G mismatch DNA.

NCD–CC (40 μM) in sodium phosphate buffer (10 mM, pH 7.0) containing 5'-(TCAA CGG TTGA)-3'/3'-(AGTT GGC AACT)-5' (5 μM), and NaCl (100 mM) was kept in air at 37 °C in an incubator. A 25 μL portion was took from the solution at the indicated time. Each of them was analyzed by RP–HPLC with a C18 endcapped column (4.6 mm * 150 mm) at 40 °C in a column oven. 0.1% TFA in H_2O and acetonitrile were used as eluting solvents. The percentage of acetonitrile was increased from 0 to 20% in 10 minutes, then 20 to 40% gradually in 20 minutes. The absorbance of the eluted compounds was measured by a photo diode array. The profiles of 330 nm were shown in the Figure S3. This experiment was done twice, and average values were plotted on Figure S4.

Melting temperature (T_m) measurements of the G–G mismatch DNA in the presence of NCD.

Thermal denaturation profiles of the solutions of the G–G mismatch DNA (5'-(TCAA CGG TTGA)-3'/3'-(AGTT GGC AACT)-5': 5 μM), in the absence or the presence of **NCD** (40 μM), sodium phosphate buffer (10 mM, pH 7.0), NaCl (100 mM), and 0.1% Tween 20 were analyzed by a UV-Vis spectrophotometer equipped with a temperature controller. The absorbance of the samples was monitored at 260 nm. All T_m experiments were carried out three times to obtain the mean values and the standard deviations.

Melting temperature (t_m) measurements of the G–G mismatch DNA under condition A

The solutions of **NCD–CC** (40 μM) in 10 mM sodium phosphate buffer (pH 7.0) and 100 mM NaCl containing was kept at 37 °C in an incubator for 24 h to make compound **4** *in situ*. Then the G–G mismatch DNA (5 μM) was added to the solution right before t_m measurement. The samples were analyzed in the same manner shown above.

Melting temperature (t_m) measurements of the G–G mismatch DNA under condition B

The solutions of the G–G mismatch DNA (5 μM) and **NCD–CC** (40 μM) in 10 mM sodium phosphate buffer (pH 7.0) and 100 mM NaCl was kept at 37 °C in an incubator for 24 h to make compound **5** *in situ*. The samples were analyzed in the same manner shown above.

Calculation of TA (tangent angle) of a inflection point of a Tm profile

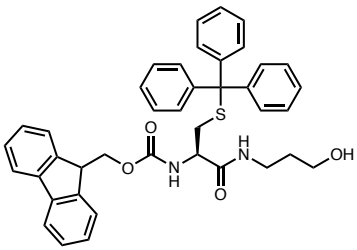
The tangent line of a T_m profile (TA) could be defined as below.

$$TA = \lim_{x \to 0} \frac{Abs(x+\Delta x) - \mathrm{Abs}(\Delta x)}{\Delta x} \ldots (1)$$

Therefore, the approximate values of derivatives were calculated by applying calculation formula (1) to two adjacent points on the T_m profiles. Thus, the value (2) shown below was calculated for each of the adjacent plots (X1, Y1), (X2, Y2) near the T_m.

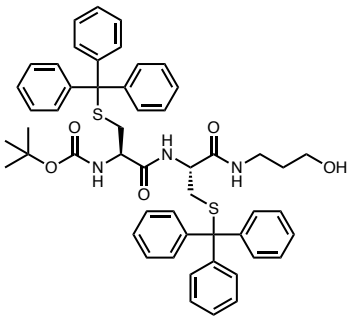
$$(Y2-Y1)/(X2-X1) \cdots (2)$$

The largest values (2) were shown in Table 1 as those tangent angles.



Chemical Formula: $C_{40}H_{38}N_2O_4S$
Exact Mass: 642.25523
Molecular Weight: 642.81400

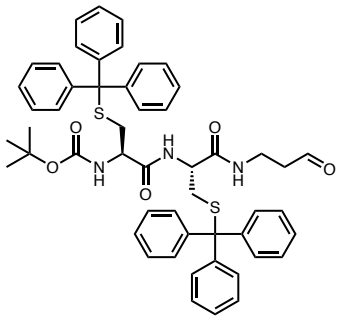
(*R*)-(9*H*-fluoren-9-yl)methyl (1-((3-hydroxypropyl)amino)-1-oxo-3-(tritylthio)propan-2-yl)carbamate (**6**). To the solution of ***N*-Fmoc-*S*-trityl-*L*-cysteine** (5.25 g, 8.96 mmol) in DMF (25 mL) was added 3-amino-1-propanol (696 mg, 9.27 mmol), EDC·HCl (1.80 g, 9.39 mmol), and HOBt·H_2O (1.37 g, 8.95 mmol). The solution was stirred at room temperature for 2 h. Then the solution was diluted with ethyl acetate (20 mL), and washed with 5% $NaHCO_3$ aqueous solution and brine. The organic layer was dried over anhydrous $MgSO_4$, and concentrated under reduced pressure. The obtained pale yellow oil was purified by column chromatography on silica. Eluting solvent of ethyl acetate/hexane = 2/1 (v/v) gave compound **6** (5.69 g, 8.85 mmol, 99%) as a white foam. R_f (ethyl acetate/hexane = 2/1, v/v) 0.3; HR-ESIMS calcd for $C_{40}H_{38}NaN_2O_4S^+$ [$(M+Na)^+$] 665.2450, found 665.2446.



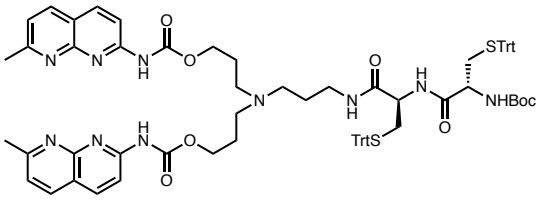
Chemical Formula: $C_{52}H_{55}N_3O_5S_2$
Exact Mass: 865.35831
Molecular Weight: 866.14800

***tert*-butyl ((*R*)-1-(((*R*)-1-((3-hydroxypropyl)amino)-1-oxo-3-(tritylthio)propan-2-yl)amino)-1-oxo-3-(tritylthio)propan-2-yl)carbamate** (**7**). To the solution of compound **6** (647 mg, 1.01 mmol) in MeOH (5 mL) was added diethylamine (0.52 mL, 367 mg, 5.03 mmol). The solution was vigorously stirred at room temperature for 1 h. TLC analysis showed the Fmoc group of compound **6** was completely cleaved under this condition. Then the solution was diluted with MeOH (25 mL) and hexane (25 mL) and vigorously stirred for 1 min, and then the solution became two phases. TLC analysis showed the residual Fmoc groups were almost extracted to hexane layer, and the deprotected cysteine residue was extracted to MeOH layer. The MeOH layer was collected, and concentrated under

reduced pressure. To the obtained colorless oil were added **N-Boc-S-trityl-L-cysteine** (465 mg, 8.96 mmol) in DMF (5 mL), EDC·HCl (194 mg, 1.02 mmol), and HOBt·H_2O (155 mg, 1.01 mmol). The pale yellow solution was stirred at room temperature for 2 h. Then the solution was diluted with ethyl acetate (50 mL), and washed with 5% $NaHCO_3$ aqueous solution and brine. The organic layer was dried over anhydrous $MgSO_4$, and concentrated under reduced pressure. The obtained residue was purified by column chromatography on silica. Eluting solvent of ethyl acetate/hexane = 3/1 (v/v) gave compound **7** (508 mg, 0.586 mmol, 58%, 2 steps) as a colorless oil. R_f (ethyl acetate/hexane = 4/1, v/v) 0.5; HR-ESIMS calcd for $C_{52}H_{55}NaN_3O_5S_2^+$ [$(M+Na)^+$] 888.3481, found 888.3483.

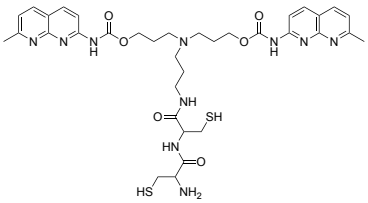


***tert*-butyl ((*R*)-1-oxo-1-(((*R*)-1-oxo-1-((3-oxopropyl)amino)-3-(tritylthio)propan-2-yl)amino)-3-(tritylthio)propan-2-yl)carbamate** (**8**). To the solution of compound **7** (58 mg, 0.063 mmol) in DMSO (0.6 mL) was added Et_3N (78 μL, 0.564 mmol) and pydine – sulfur trioxide complex (Py·SO_3, 45 mg, 0.282 mmol). The solution was stirred at room temperature for 18 h. Then the reaction was quenched with H_2O (30 mL). Then to the solution was added ethyl acetate (30 mL). The organic layer was collected, washed with brine several times, collected, dried over anhydrous $MgSO_4$, and concentrated under reduced pressure. The obtained residue was was purified by silica-gel column chromatography. Elution with ethyl acetate/hexane = 1/1 (v/v), gave compound **8** (41 mg, 0.047 mmol, 75%) as colorless oil. R_f (ethyl acetate/hexane = 1/1, v/v) 0.4; 1H-NMR (600 MHz, CHLOROFORM-D) δ 9.62 (s, 1H), 7.45-7.11 (30H), 6.88 (d, *J* = 5.2 Hz, 1H), 6.08 (d, *J* = 7.2 Hz, 1H), 4.88 (d, *J* = 5.0 Hz, 1H), 4.15 (m, 1H), 3.62 (m, 1H), 3.37 (m, 1H), 3.31 (m, 1H), 3.06 (dd, *J* = 12.6, 5.2 Hz, 1H), 2.65 (dd, *J* = 12.9, 4.6 Hz, 1H), 2.50 (t, *J* = 6.6 Hz, 2H), 2.44 (dd, *J* = 12.9, 7.1 Hz, 1H), 2.26 (dd, *J* = 12.6, 4.6 Hz, 1H), 1.36 (s, 9H); HR-ESIMS calcd for $C_{52}H_{53}NaN_3O_5S_2^+$ [$(M+Na)^+$] 886.3324, found 886.3322.



Chemical Formula: $C_{78}H_{82}N_{10}O_8S_2$
Exact Mass: 1350.57585
Molecular Weight: 1351.69600

NCD-L-Cys(Tr)-L-Cys(Tr)-NHBoc (**9**). Hydrochloric salt of **NCD** (26 mg, 0.048 mmol) was dissolved in H_2O (1 mL). To the solution was added 28% ammonia aqueous solution (1 mL) to give white precipitate. The solution was vigorously stirred with $CHCl_3$ (1 mL). Then the $CHCl_3$ layer was collected, dried over Na_2SO_4, concentrated under reduced pressure, and pumped to give free **NCD**. It was dissolved in MeOH (0.25 mL). To the solution were added compound **8** (41 mg, 0.047 mmol) in DCM (0.2 mL) and $NaBH_3CN$ (3 mg, 0.050 mmol). The solution was vigorously stirred at room temperature for 18 h. Then the solution was concentrated under reduced pressure. The obtained residue was dissolved in $CHCl_3$, and washed with brine. The organic layer was collected, dried over Na_2SO_4, and concentrated under reduced pressure. The residue was purified by silica gel column chromatography. The eluting solvent of 8% methanol in $CHCl_3$ gave compound **9** (32 mg, 0.024 mmol, 50%); 1H-NMR (600 MHz, CHLOROFORM-D) δ 9.07 (s, 1H), 8.56 (s, 2H), 8.27 (d, J = 8.9 Hz, 2H), 8.10 (d, J = 8.9 Hz, 2H), 7.97 (d, J = 8.2 Hz, 2H), 7.37-6.98 (32H), 6.66 (d, J = 6.2 Hz, 1H), 5.84 (d, J = 7.6 Hz, 1H), 4.40 (d, J = 4.8 Hz, 1H), 4.28-4.25 (3H), 4.20 (m, 2H), 4.11-3.99 (m, 1H), 3.18 (m, 1H), 3.10 (m, 1H), 2.89 (m, 1H), 2.68 (s, 6H), 2.62-2.20 (8H), 1.82 (m, 4H), 1.60 (m, 2H), 1.23 (s, 9H); HR-ESIMS calcd for $C_{78}H_{83}N_{10}O_8S_2^+$ [$(M+H)^+$] 1351.5831, found 1351.5842.

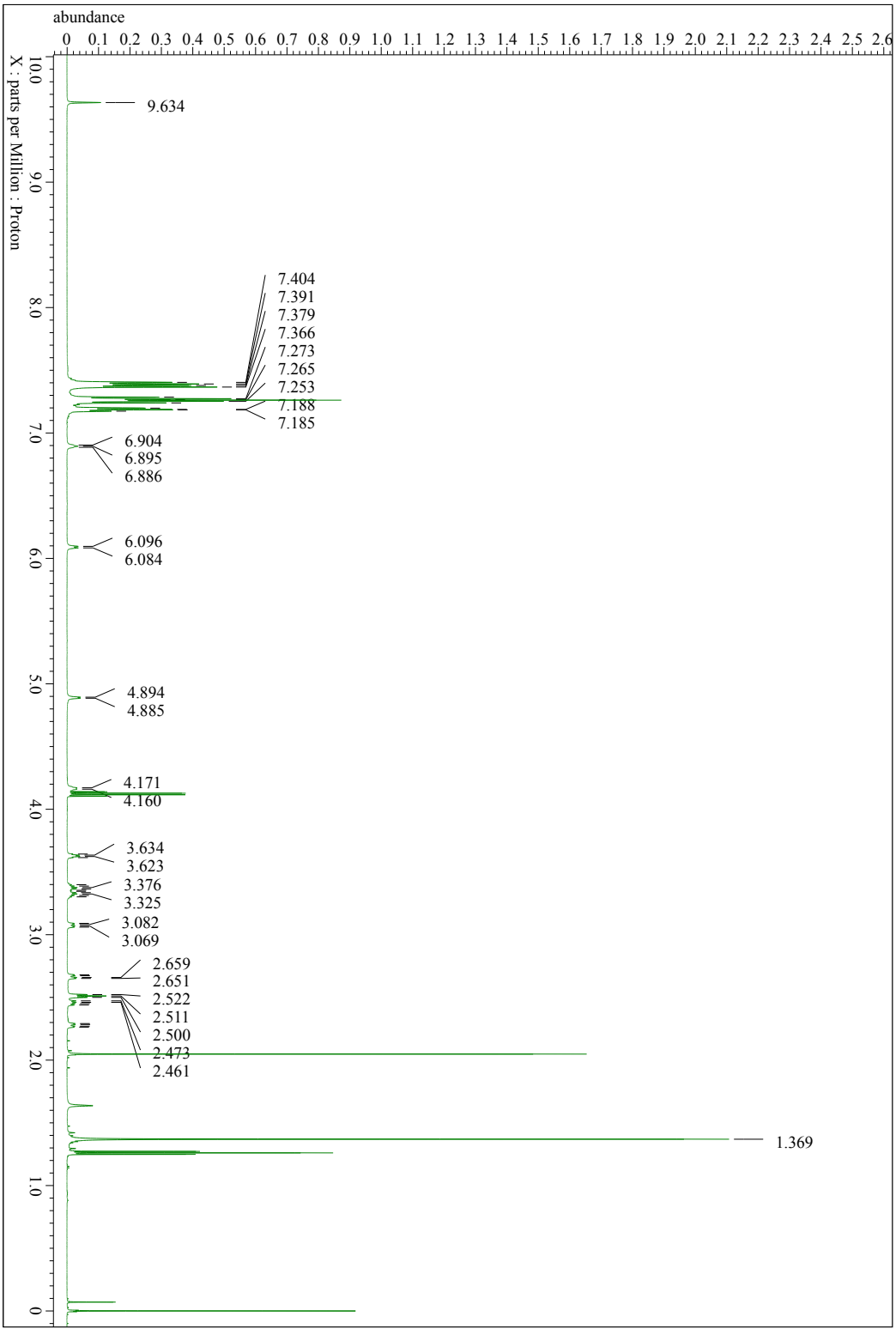


Chemical Formula: $C_{35}H_{46}N_{10}O_6S_2$
Exact Mass: 766.30432
Molecular Weight: 766.93700

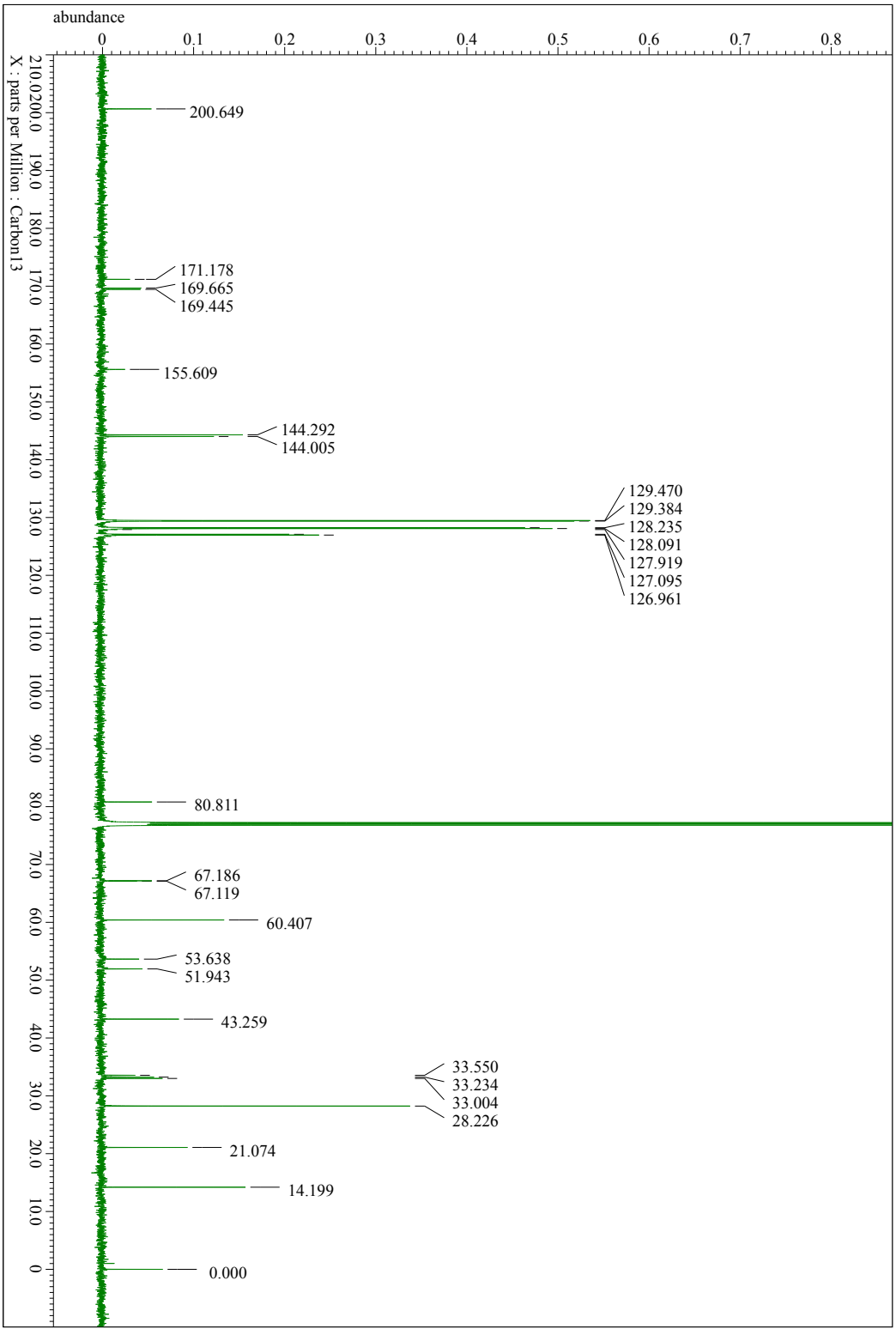
NCD-Cys-Cys (**3**). To the solution of **9** (19 mg, 0.014 mmol) in 0.5 mL of DCM was added TFA (0.5 mL), and then triethylsilane (20 μL) on an ice bath. The colorless solution was stirred at 0 °C to room temperature for 30 min. Then the solution was concentrated under reduced pressure. To the residue was added 4 M HCl in ethyl acetate (10 mL) at 0 °C for 2 h. Then the solution was concentrated under reduced pressure. The obtained residue was dissolved in 10 mL of mixed solution of MeOH/MeCN (1/1, v/v). The solution was washed with hexane (10 mL) several times. The MeOH/MeCN layer was concentrated under reduced pressure again to give crude compound **3** (11 mg, 0.014 mmol, quant) with the intramolecular cyclized product compound **4**. The residue was further purified by RP-HPLC. (HPLC condition: an C18 endcapped column was used at room

temperature. 0.1% TFA and MeCN were used for eluents. The percentage of MeCN was changed from 20% to 60% gradually for 20 min; HR-ESIMS calcd for $C_{35}H_{46}N_{10}O_6S_2^+$ [$(M+H)^+$] 767.3116, found 767.3117.

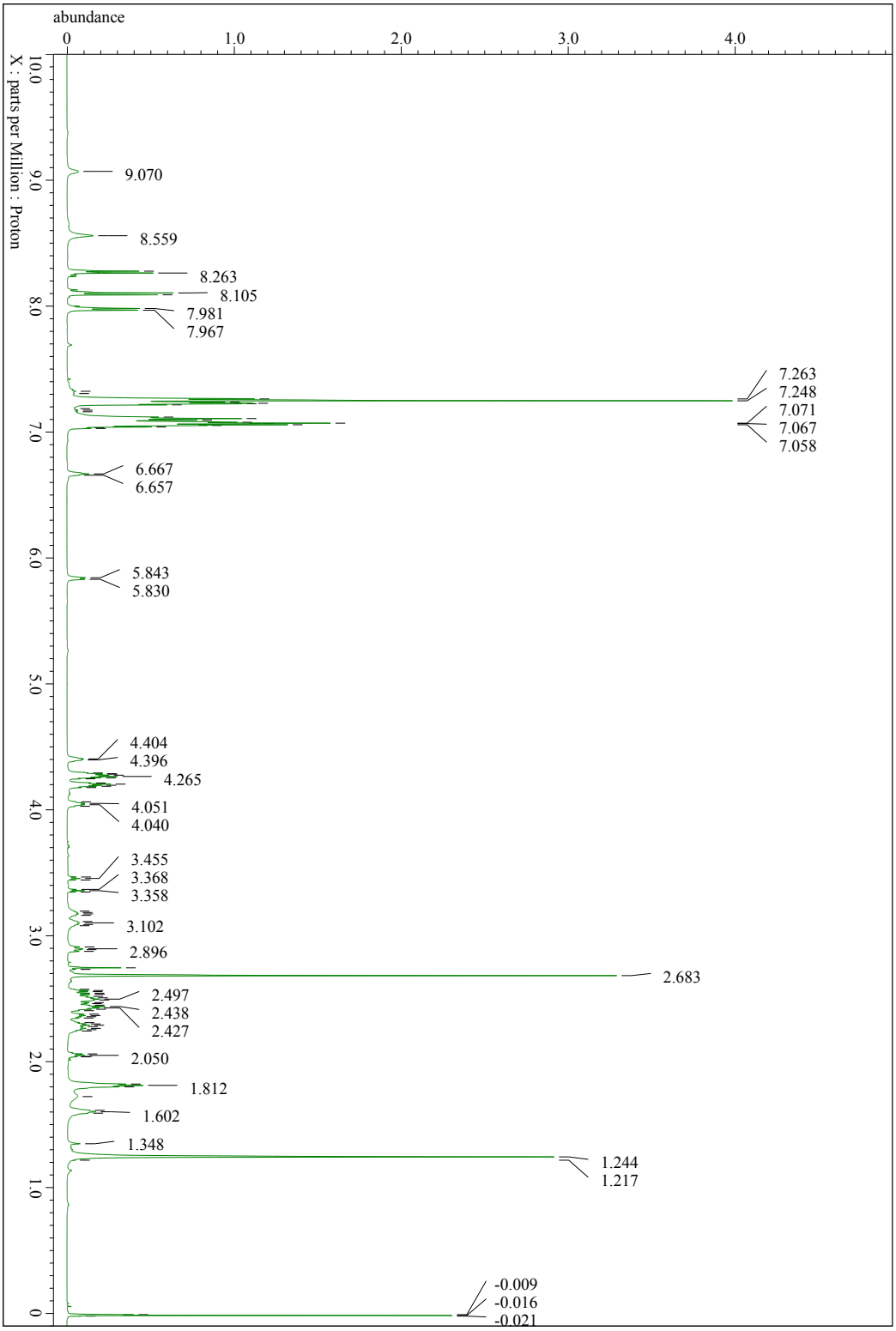
1H-NMR of aldehyde 8



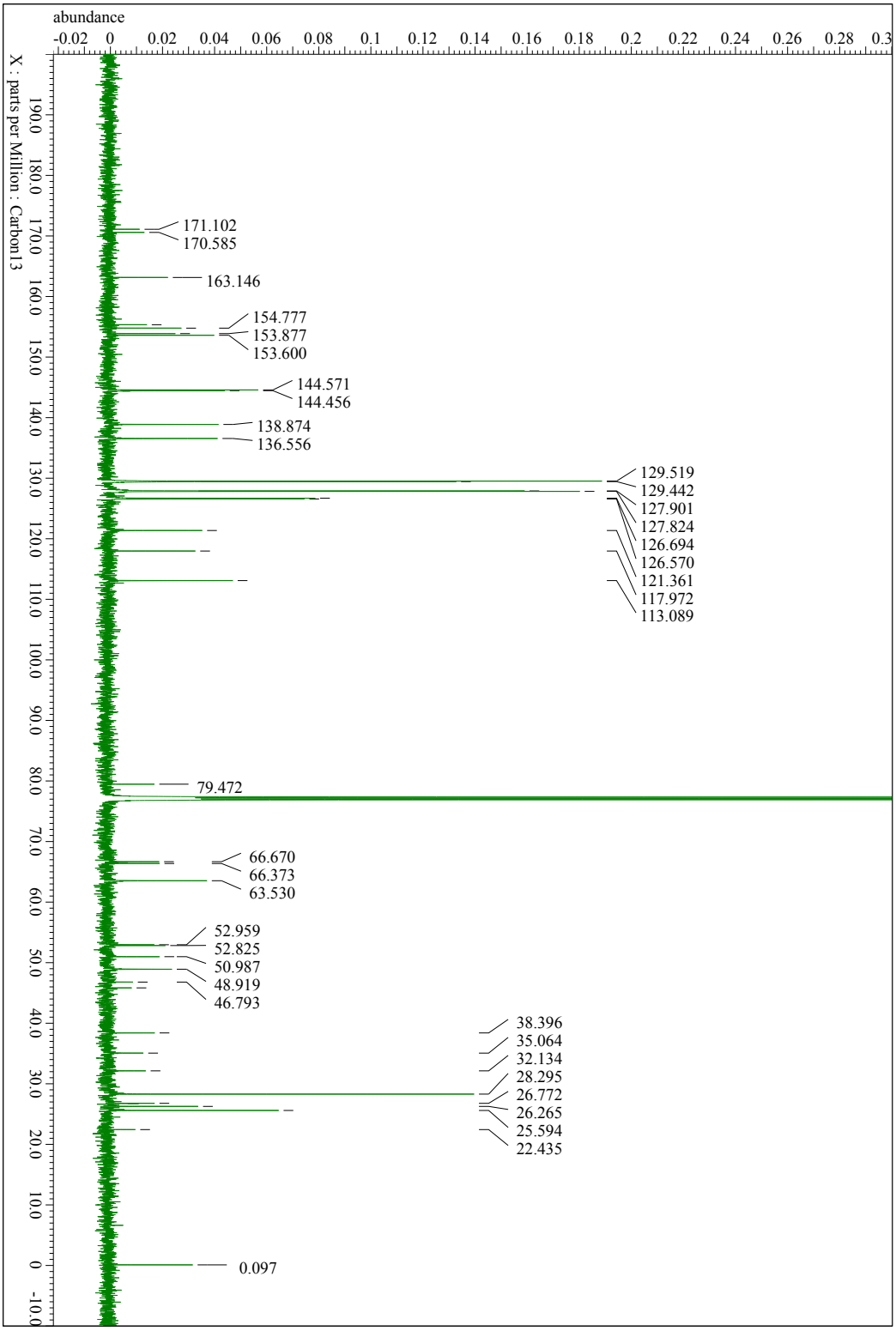
13C-NMR of aldehyde 8



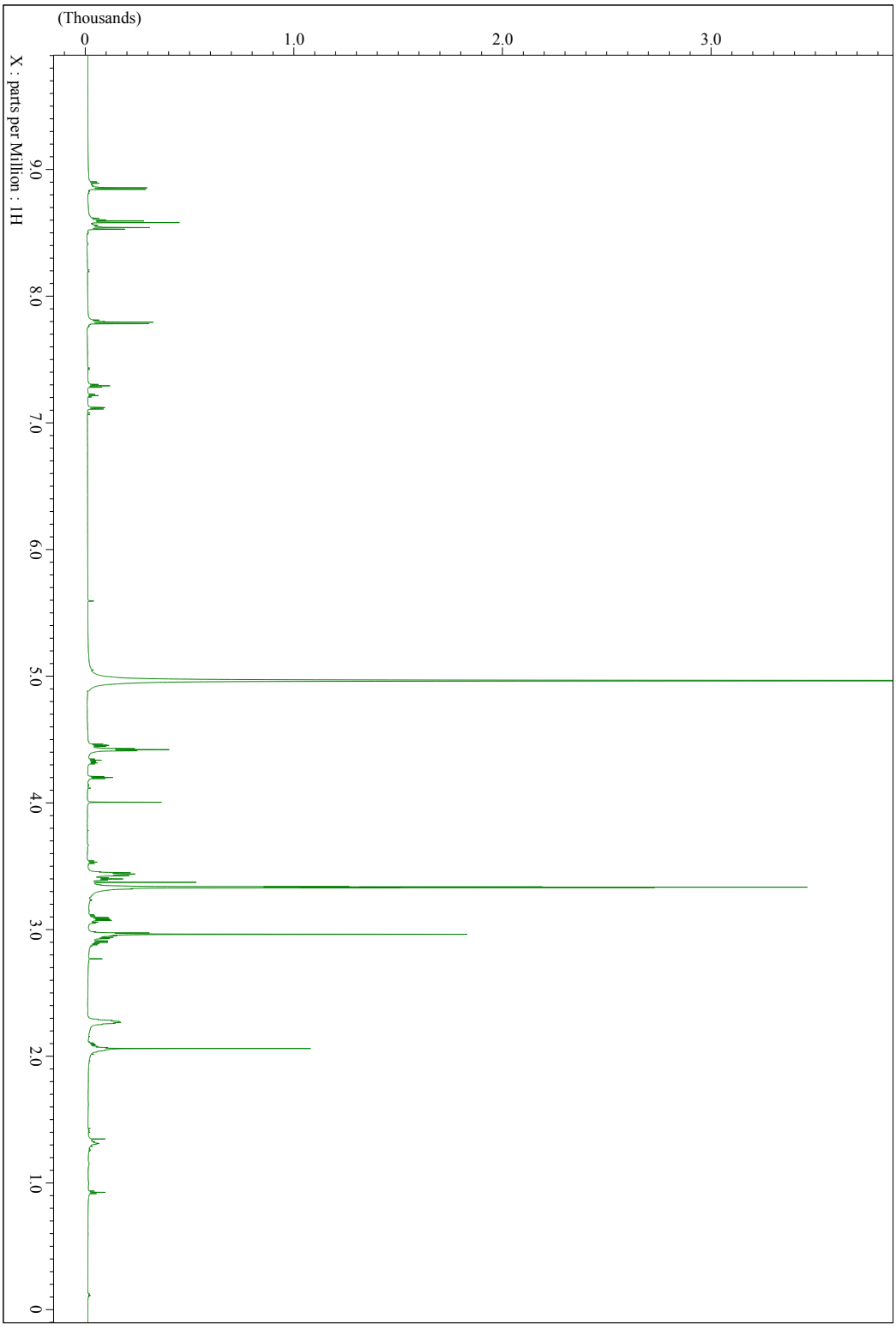
1H-NMR of compound 9 (NCD-C(Trt)-C(Trt)(Boc)



13C-NMR of compound 9 (NCD-C(Trt)-C(Trt)(Boc)



1H-NMR of crude compound 3



13C-NMR of crude compound 3