Supporting Information

G-Quadruplex-Proximity Protein Labeling Based on Peroxidase Activity

Tatsuki Masuzawa,[a] Shinichi Sato,*[b,c] Tatsuya Niwa,[d] Hideki Taguchi,[d] Hiroyuki Nakamura,[c] Takanori Oyoshi*[e]

(a) Graduate School of Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8021, Japan
(b) Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai-shi, Miyagi, 980-8577, Japan
(c) Laboratory for Chemistry and Life Science, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, 226-8503, Japan
(d) Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, 226-8503, Japan
(e) Graduate School of Integrated Science and Technology, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka, 422-8529, Japan

List of contents

1. Supporting Data

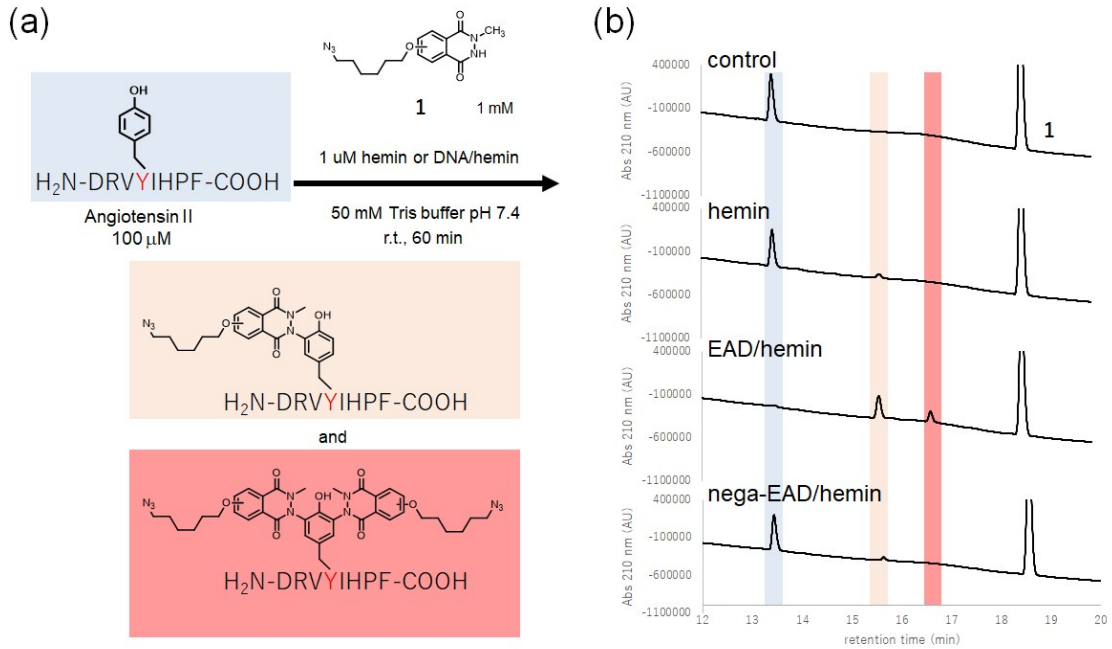


Figure S1. HPLC analysis of labeled peptide. (a) Scheme of labeling reaction using angiotensin II. (b) HPLC analysis of reaction mixtures. The modified peptide fragments were detected by HPLC (column: Inertsil ODS-4, 5 μm, 4.6 x 150 mm, GL Science) detecting absorbance at 210 nm. The micropump (1.0 mL/min) gradient method was used, as follows. Mobile phase A: 0.1% formic acid aq., mobile phase B: 100% methanol. 0−5 min: 5% B, 5−27 min: 5−100% B, 27−32 min: 100% B, 32−32.1 min: 100-5% B, 32.1−40 min: 5% B. EAD enhanced hemin-induced tyrosine labeling.

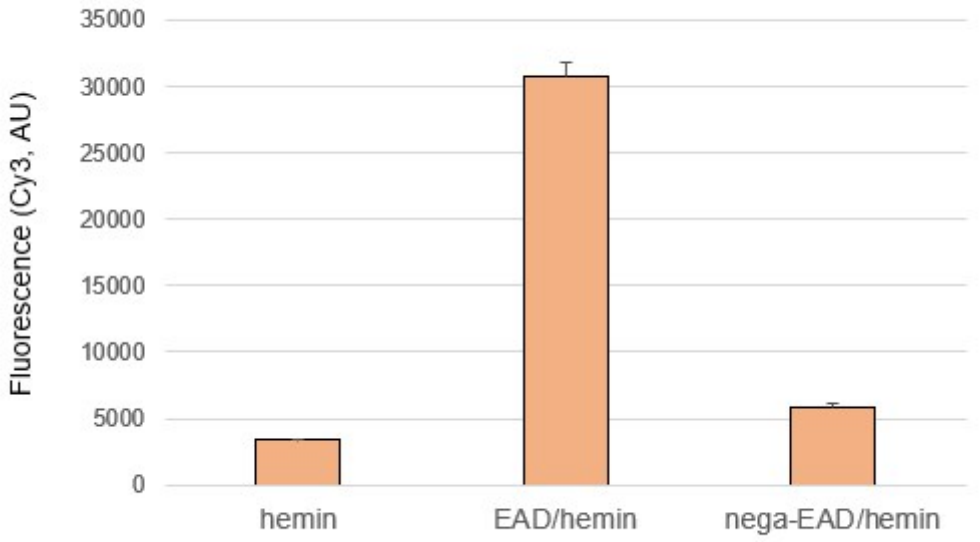


Figure S2. Fluorescence intensities of Cy3-labeled BSAs. Reaction condition: BSA (10 μM), **1** (1 mM), hemin/DNA (1 μM), H_2O_2 (1 mM), in 50 mM Tris-HCl pH 7.4, room temperature, 60 min; DBCO-Cy3 (10 μM) 37 °C, 60 min. Excess DBCO reagent was removed using a Sephadex G-25 gel (GE Healthcare) filtration column (2000×g, 4 min).

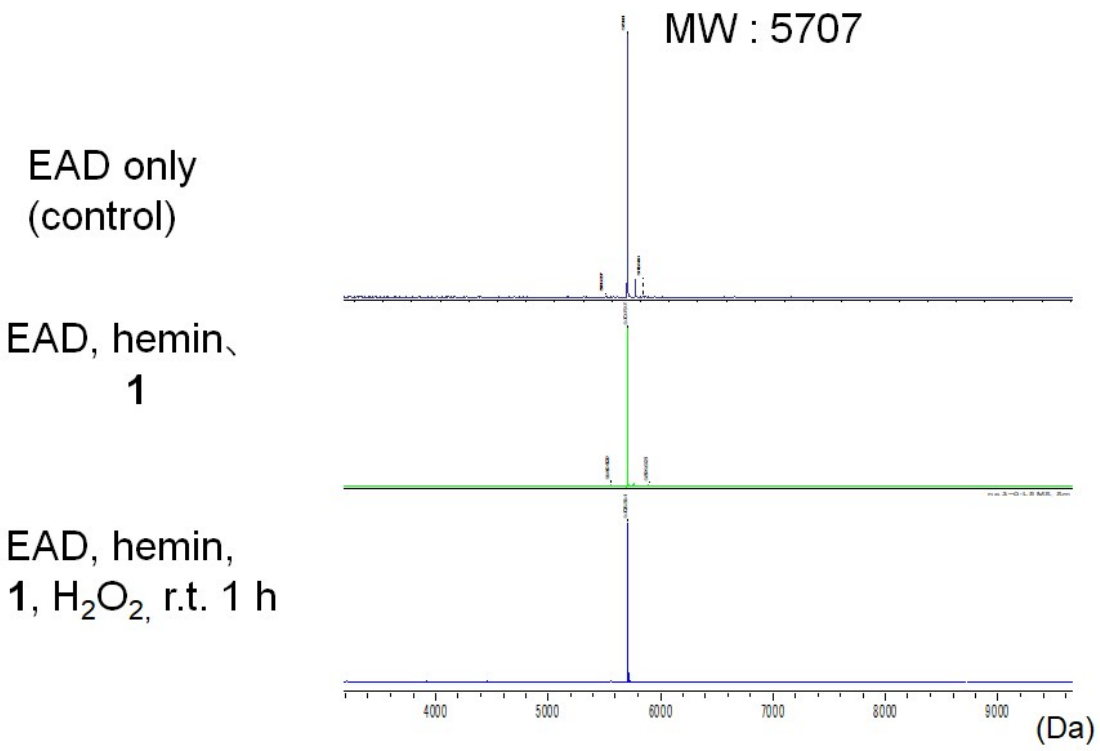


Figure S3. G4 DNA was not oxidized or labeled under the labeling reaction condition.

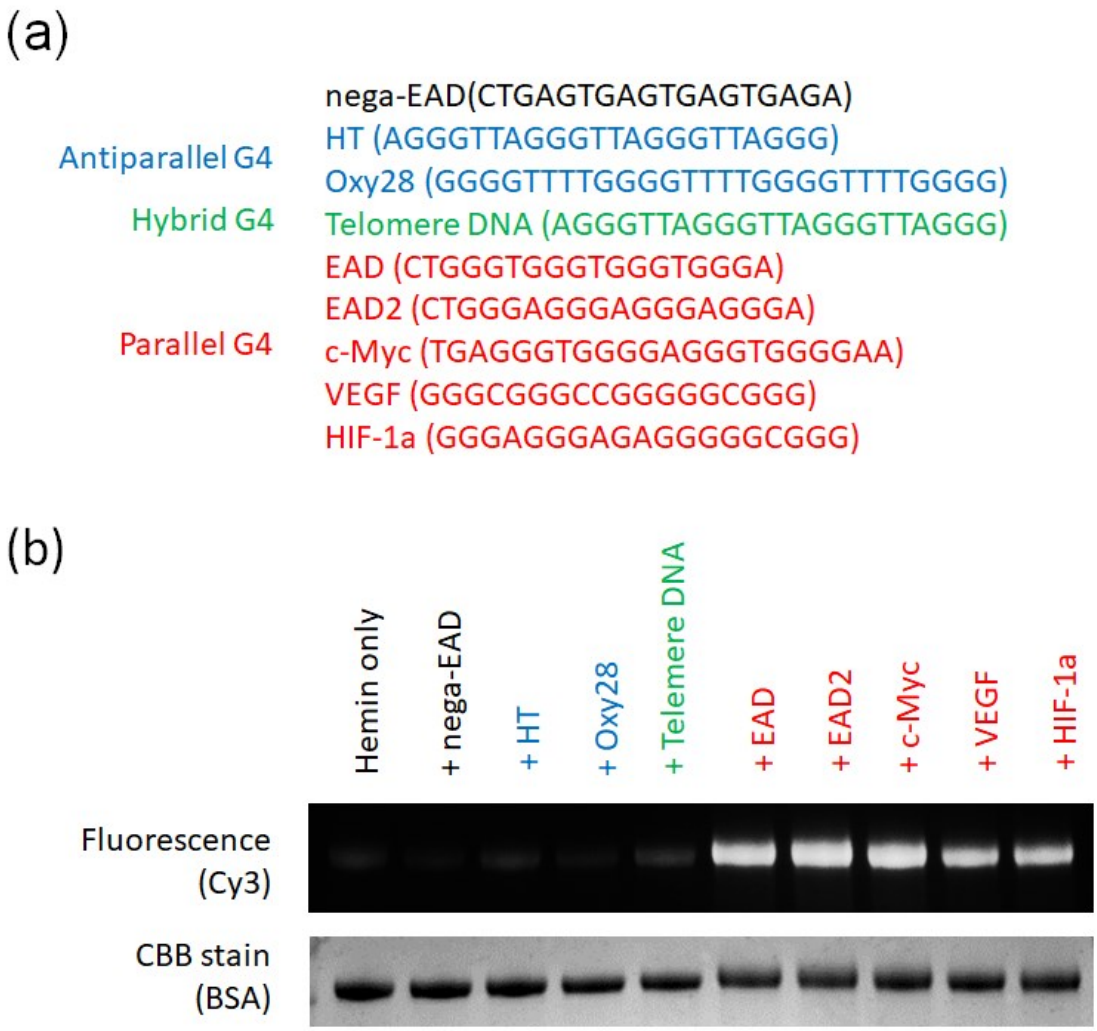


Figure S4. Importance of parallel G4 structure in the promoting effect of hemin-catalyzed protein labeling. (a) Sequences of DNA used in the experiments. (b) Promoting effect of the protein labeling with various sequences of DNA (G4/hemin 1 μM).

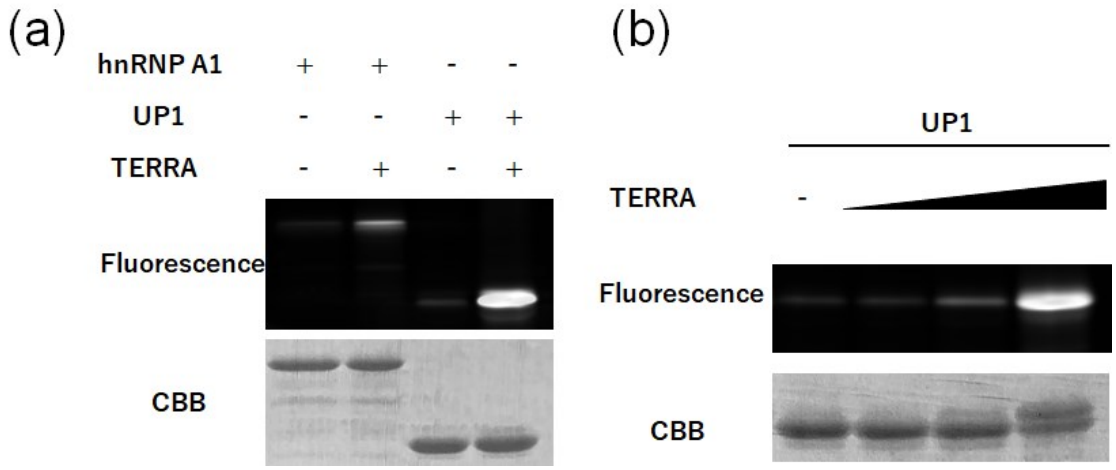


Figure S5. Modification of hnRNP A1 and UP1 with **1**. (a) hnRNP A1 or UP1 were treated with **1**, hemin and H_2O_2 in the presence or absence of TERRA. (b) UP1 was treated with **1**, hemin, H_2O_2, and TERRA (0, 0.2, 0.6 or 1.0 μM). After modification, each protein was visualized with DBCO-Cy3.

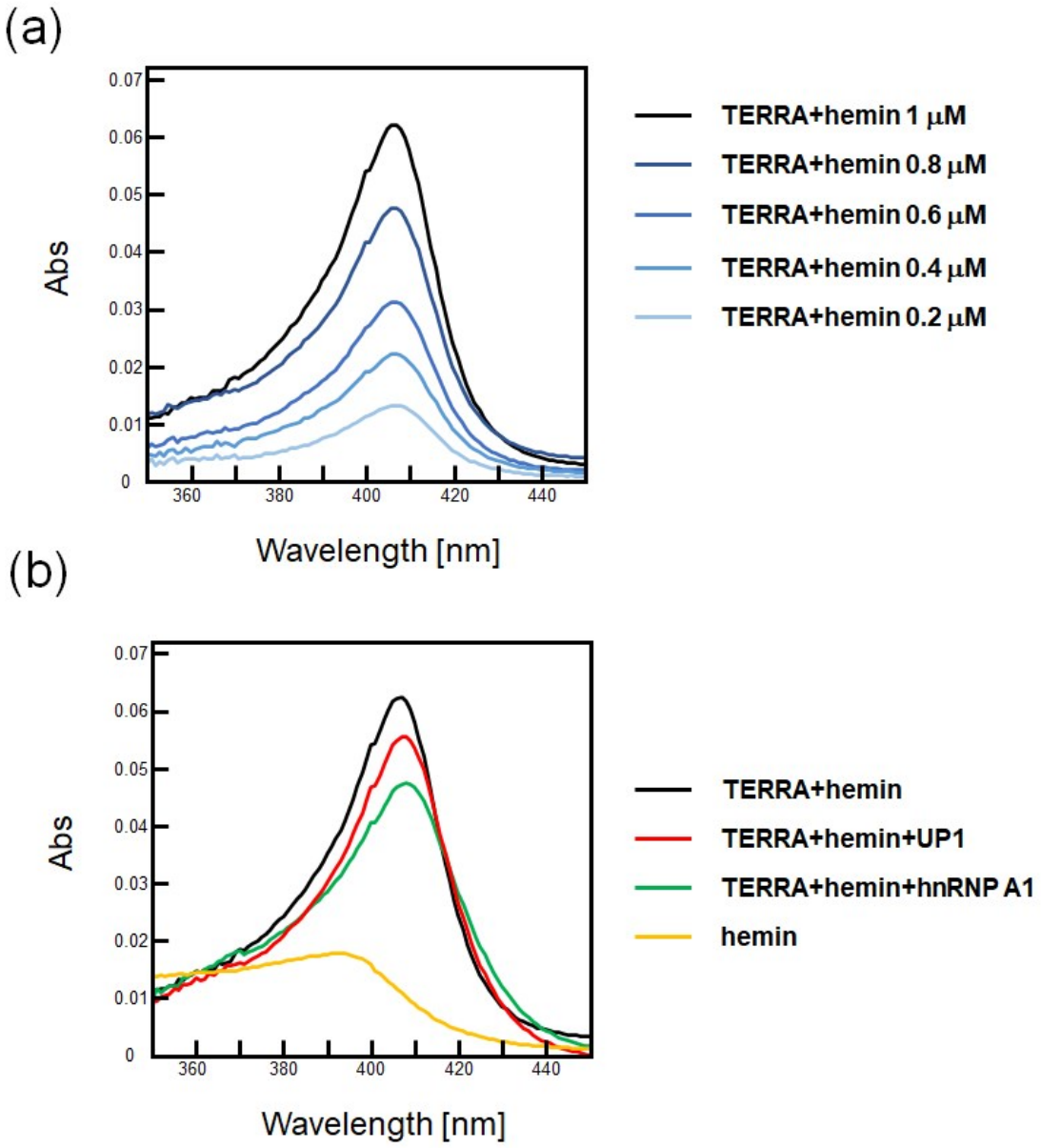


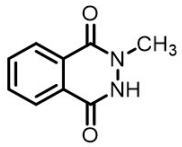
Figure S6. (a) Soret absorption of various amounts of hemin (0.2, 0.4 0.6, 0,8, and 1.0 μM) in Tris-HCl buffer (pH 7.5) containing 100 mM KCl in the presence of 1.0 μM TERRA at room temperature. Line colors: black = TERRA + 1.0 μM hemin; dark blue = TERRA + 0.8 μM hemin; blue = TERRA + 0.6 μM hemin; cyan = TERRA + 0.4 μM hemin; light blue = TERRA + 0.2 μM hemin. (b) Comparison of soret absorption of 1.0 μM hemin in Tris-HCl buffer (pH 7.5) containing 100 mM KCl in the presence of 3.0 μM UP1 or hnRNP A1 with 1.0 μM TERRA at room temperature. Line colors: black = TERRA + hemin; red = TERRA + hemin + UP1; green = TERRA + hemin + hnRNP A1; yellow = hemin.

(a)

Covered sequence by nanoLC-MS/MS

MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPN
TKRSRGFGFVTYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRP
GAHLTVKKIFVGGIKEDTEEHHLRDYFEQYGKIEVIEIMTDRGSGKKRGFA
FVTFDDHDSVDKIVIQKYHTVNGHNCEVRKALSKQEMASASSSQRG

(b)



N-Me-Lumi

(c)

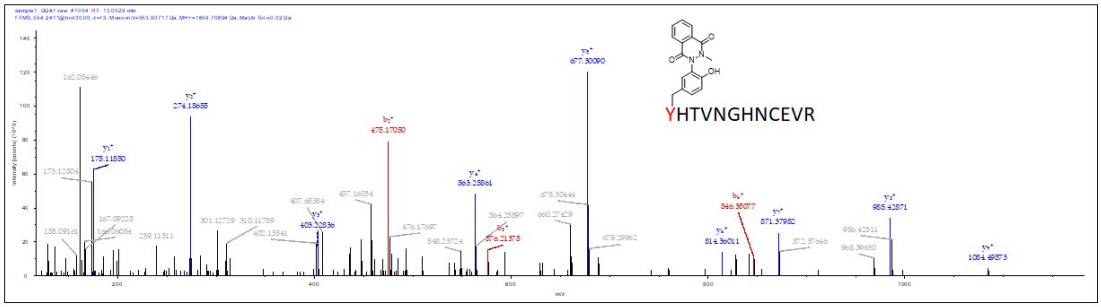


Figure S7. nanoLC-MS/MS analysis of trypsin-digested peptide fragments of labeled UP1. (a) Sequence of UP1 detected by nanoLC-MS/MS. (b) Structure of labeling reagent, N-Me-Lumi. (c) MS/MS fragment assignment of labeled peptide in Y167.

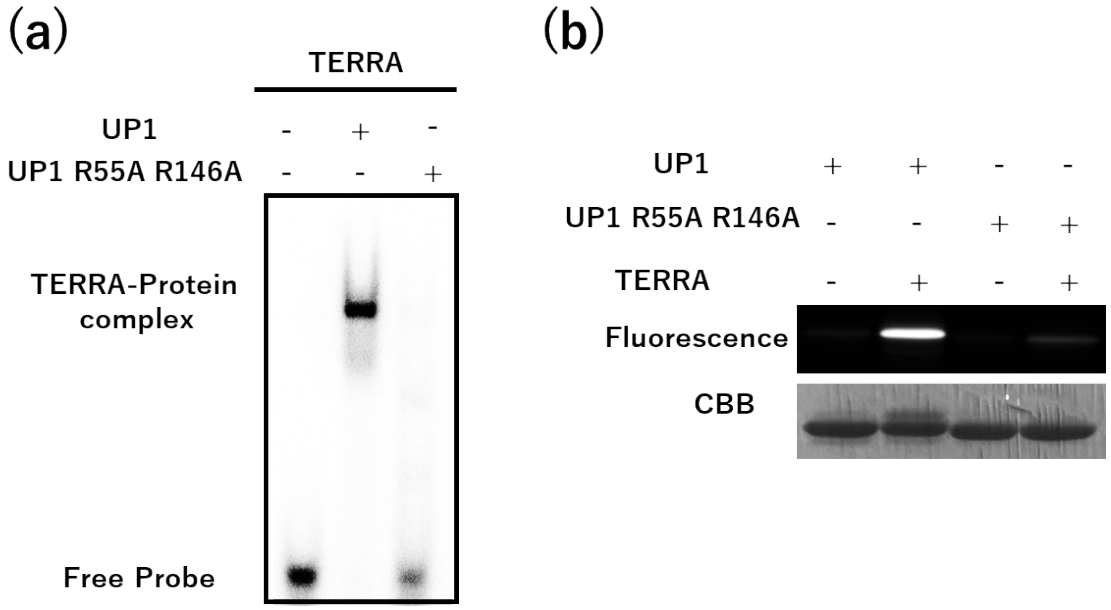


Figure S8. Effect of G-quadruplex binding activities for modification of UP1. (a) EMSA was performed using UP1 or UP1 R55A R146A (arginine 55 and 146 substituted by each alanine) with ^{32}P-labeled TERRA. (b) 3 μM UP1 or UP1 R55A R146A were treated with **1**, hemin, H_2O_2, and 1.0 μM TERRA in 50 mM Tris-HCl pH 7.5 and 100 mM KCl. After modification, each protein was visualized with DBCO-Cy3.

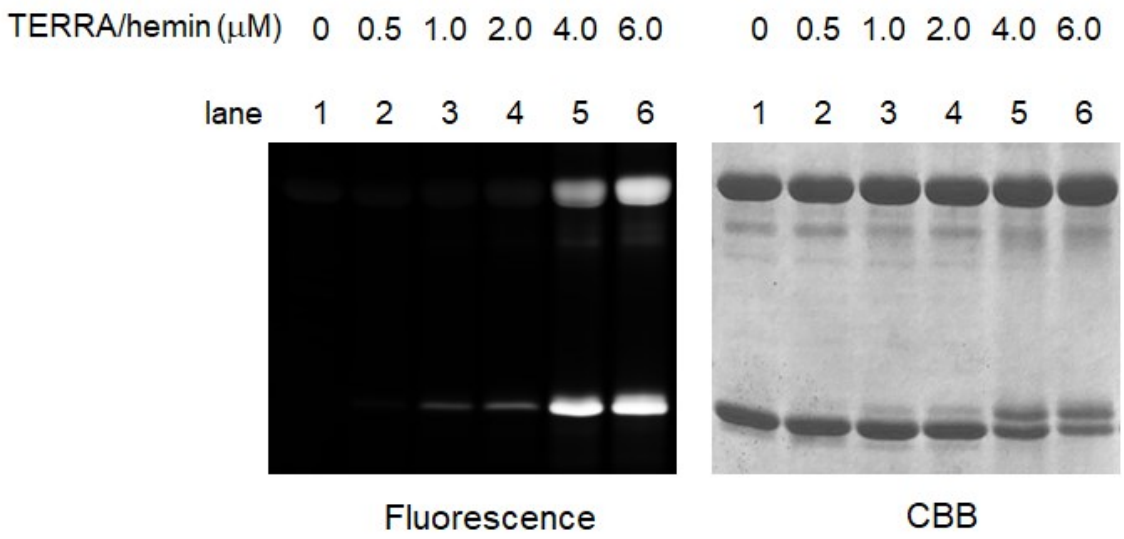


Figure S9. Modification of UP1 with 1 in the presence of BSA. UP1 (3.0 μM) and BSA (3.0 μM) were treated with various amounts of TERRA-hemin complex with 200 μM H_2O_2 in 50 mM Tris-HCl pH 7.5 containing 100 mM KCl. After modification, each protein was visualized with DBCO-Cy3. The TERRA-hemin concentrations were 0 μM (lane 1), 0.5 μM (lane 2), 1.0 μM (lane 3), 2.0 μM (lane 4), 4.0 μM (lane 5), and 6.0 μM (lane 6).

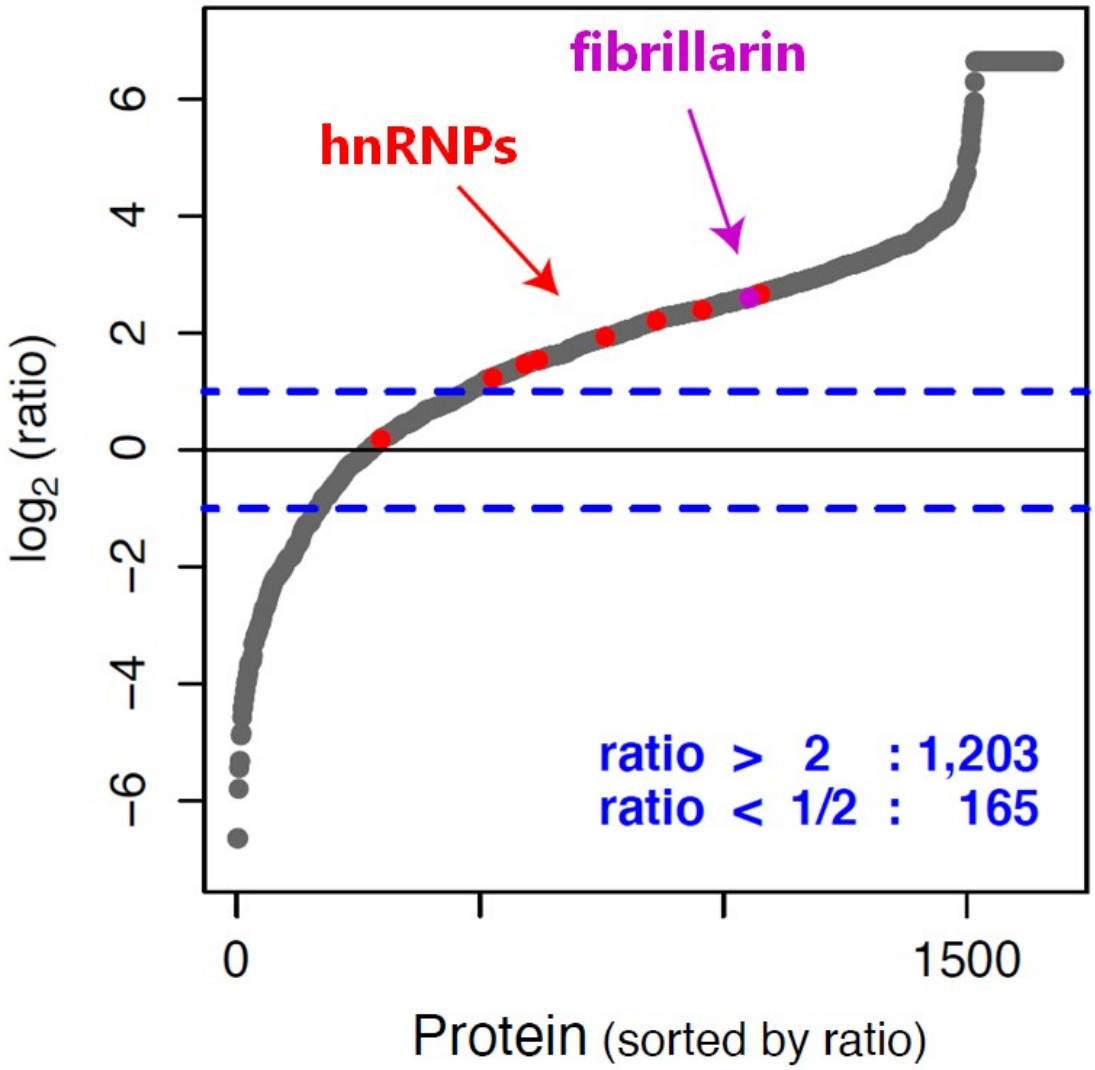


Figure S10. Modification of TERRA-binding proteins with **1** in HeLa cell lysate. Plot of abundance ratio values of detected proteins in the presence/absence of TERRA (100 μM). A control sample was prepared in the presence of **1** and in the absence of TERRA.

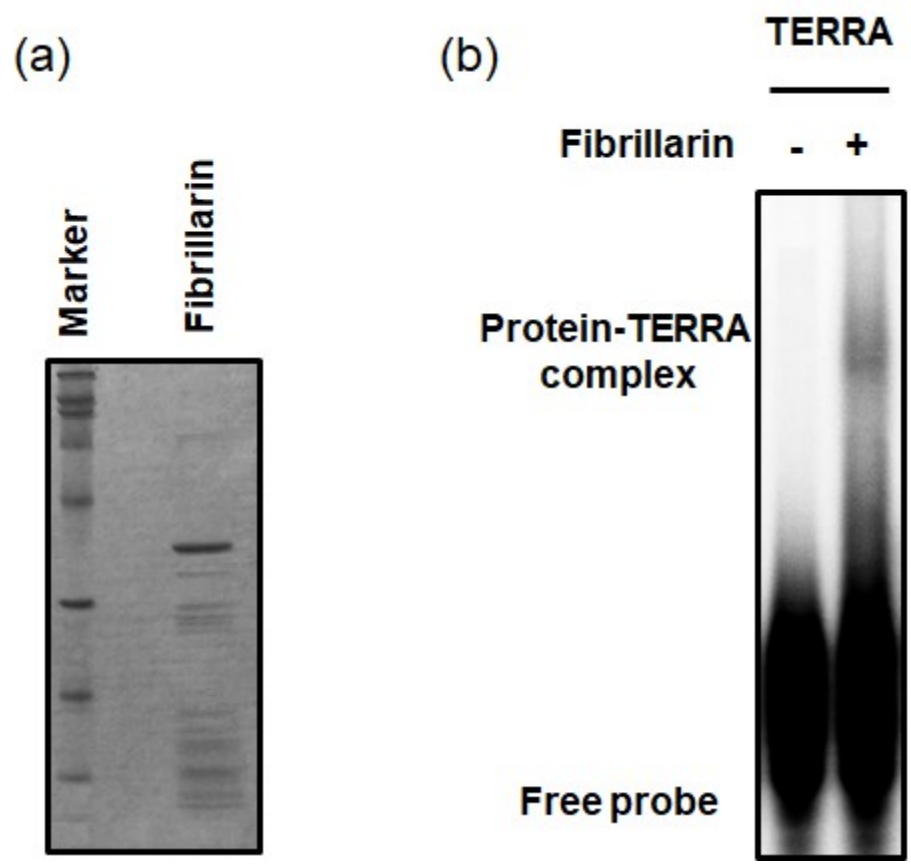


Figure S11. TERRA binding of fibrillarin. (A) 14% SDS-PAGE, Lane 1, molecular weight makers; Lane 2, fibrillarin (33.8 kDa). Molecular weight markers (kDa): pig myosin (200), Escherichia coli b-galactosidase (116), rabbit muscle phosphorylase B (97.2), bovine serum albumin (66.4), hen egg white ovalbumin (44.3), bovin carbonic anhydrase (29), soybean trypsin inhibitor (20.1), hen egg white lysozyme (14.3). (B) An EMSA was performed with ^{32}P-labeled TERRA and fibrillarin. The RNA-protein complexes ware resolved by 6% polyacrylamide gel electrophoresis and visualized by autoradiography.

2. Experimental section

BSA labeling using G4-hemin.

DNAs were customized by Integrated DNA Technologies. DNA solution (100 μM) in a buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) was added hemin (100 μM) and incubated at room temperature for 5 min. The resulting G4-hemin complexes were used without purification. BSA (final 10 μM), **1** (final 1 mM) were added hemin or G4-hemin (final 1.0 or 10 μM) and H_2O_2 (final 1 mM). The solution was incubated at room temperature for 60 min. To avoid the thiol-yne reaction, iodoacetamide (final 10 mM) was treated at room temperature for 15 min in dark. DBCO-Cy3 (from 10 mM in DMF, final concentration 100 μM) was treated and shaken at 37 °C for 1 h. The resulting solutions were analyzed by SDS-PAGE.

Mass spectrometry of DNA after labeling reaction

The mixture of EAD DNA (1 μM), BSA (10 μM) and hemin (1 μM) in a buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) was treated with H_2O_2 (final 0 or 1 mM) at room temperature for 1 h. After the reaction, the solutions were desalted using C18 pipette tips (Agilent) (equilibration and wash: 0.1 M triethylammonium acetate pH 7.0, elution: 50% CH_3CN/water). Each sample was mixed with 1 μL of saturated 3-hydroxypicolinic acid (3-HPA) solution in 50% CH_3CN/0.1% TFA aq. on MALDI-TOF plate. The peak of DNA was detected by MALDI-TOF analysis (Bruker, UltrafleXtreme)

Plasmid constructs

The hnRNP A1 plasmid was used as a template for polymerase chain reaction (PCR). The hnRNP A1 (1-319) and UP1 (1-194) cDNAs were cloned into the pGEX6P-1 vector (GE Healthcare, USA) between the BamHI and XhoI sites using the following stets of primers to express an *N*-terminal glutathione S-transferase (GST) fusion protein, respectively: for hnRNP A1 forward 1 d(CGG GAT CCA TGT CTA AGT CAG AGT CTC CTA AAG AG) and reverse 1 d(CGC TCG AGT TAA AAT CTT CTG CCA CTG CC); for UP1, forward 1 and reverse 2 d(CGC TCG AGT TAA CCT CTT TGG CTG GA). UP1 R55A R146A and UP1 Y167F were obtained by replacing Arg55, 146 to Ala or Tyr167 to Phe in UP1 using a KOD-Plus- Mutagenesis Kit (Toyobo) with the UP1 in the pGEX6P-1 vector used as the template and the following primers, respectively: for UP1 R55A R146A, R55A R146A forward 1 d(CTC CGC GGG CTT TGG GTT TGT CAC A), R55A R146A reverse 1 d(CGC TTG GTG TTT GGA TCT CTC ATT ACC A), R55A R146A forward 2 d(GAA AGC GGG CTT TGC CTT TGT AAC C) and R55A R146A reverse 2 d(TTG CCA CTG CCT CGG TCA GTC ATG A); for UP1 Y167F, Y167F forward 1 d(TCT GAA TGA CAA TCT TAT CCA CGG AGT CAT GGT CG) and Y167F reverse 1 d(AAT TCC ATA CTG TGA ATG GCC ACA ACT GTG AAG TT). The fibrillarin plasmid was used as a template for PCR. The fibrillarin cDNAs were cloned into the

pGEX6P-1 vector between the BamHI and XhoI sites using the following stets of primers to express an *N*-terminal GST fusion protein: for fibrillarin forward d(CGG GAT CCA TGA AGC CAG GAT TCA GTC CCC) and reverse d(CGC TCG AGT CAG TTC TTC ACC TTG GGG GG). All constructs were verified by automated DNA sequencing. All DNA oligomers were obtained from Operon Biotechnologies.

Expression and purification of glutathione S-transferase fusion proteins

All recombinant proteins fused to glutathione S-transferase (GST) at the *N*-terminus and overexpressed in *Escherichia coli* as described previously.[1] The *E. coli* strain BL21 (DE3) pLysS-component cells were transformed with the vectors, and transformants were grown at 37 °C in Luria Bertani medium containing ampicillin (0.1 mg/mL). Protein expression was induced at $A_{600} = 0.6$ with 0.1 mM isopropyl β-D-1-thiogalactopyranoside. The cells were then grown for an additional 16 h at 25 °C and harvested by centrifugation (6400 *g* for 15 min). The *E. coli* pellets were resuspended in W buffer (100 mM Tris-HCl [pH 7.5], 150 mM KCl, 1 mM ethylenediaminetetraacetic acid [EDTA], 1 mM dithiothreitol [DTT], 0.1 mM phenylmethanesulfonyl fluoride and 12.5 μg/mL RNase G.S. [Nippon Gene]) (hnRNP A1 and UP1) or buffer B (100 mM Tris-HCl [pH 7.5], 150 mM KCl, 1 mM EDTA acid, 1 mM DTT and 1 v/v% Triton X-100) with 0.1 mM phenylmethanesulfonyl fluoride and 12.5 μg/mL RNase G.S. (fibrillarin). The supernatants containing the expressed proteins were lysed by sonication (model UR-20P, Tomy Seiko) and centrifuged at 16200 *g* for 10 min at 4 °C. The supernatant and glutathione agarose (Sigma-Aldrich) were incubated with gentle mixing for 1 h at 4°C; resin was washed with buffer B and KCl buffer (50 mM Tris-HCl pH 7.5 and 100 mM KCl) at 4 °C. GST-tags were cleaved using buffer containing 8 units/mL (hnRNP A1 and UP1) or 32 units/mL (fibrillarin) PreScission protease (GE healthcare) on a resin for 16 h (hnRNP A1 and UP1) or 4 h (fibrillarin) at 4 °C, and the protein was eluted by KCl buffer. The protein concentrations were determined using a BCA Protein Assay Kit (Thermo Scientific).

Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were performed as described previously.[1] 2 nM ^{32}P-labeled telomeric repeat-containing RNA (TERRA) were annealed by incubation at 95 °C, followed by cooling to 4 °C at a rate of 2 °C /min in 50 mM Tris-HCl buffer (pH 7.5) containing 100 mM KCl. Binding reactions were performed in a final volume of 20 μL using 20 fmol of the ^{32}P-labeled TERRA in KCl buffer with or without 1 μM purified UP1 or 13.5 μM purified fibrillarin. Samples were incubated for 30 min at 4 °C, and then loaded onto a 6% polyacrylamide (acrylamide/bisacrylamide, 19:1) nondenaturing gel. The gel and the electrophoresis buffer contained 0.5×TBE (45 mM Tris base, 45 mM boric acid, and 0.5 mM EDTA). Electrophoresis was performed at 10 V/cm for 100 min at 25 °C. The gel was exposed in a phosphorimager cassette and imaged (Personal Molecular Imager FX, Bio-Rad Laboratories).

Protein modification for SDS-PAGE

TERRA was annealed by incubation at 95 °C, followed by cooling to 4 °C at a rate of 2 °C /min in 50 mM Tris-HCl buffer (pH 7.5) containing 100 mM KCl. Modification reaction was performed in a final volume 30 μL in the KCl buffer. Hemin (from 100 μM stock solution in *N,N*-Dimethylformamide (DMF); final concentration, 1 μM) and TERRA (final concentration 0.2, 0.6 or 1.0 μM) were incubated for 30 min at 4 °C, then purified proteins (final concentration 3 μM), H_2O_2 (final concentration 200 μM) and **1** (from 100 mM stock solution in dimethyl sulfoxide (DMSO); final concentration, 1 mM) were added to the reaction mixture and incubated for 1 h at 4 °C. For Cys alkylation, iodoacetamide (final concentration 10 mM) was added to the reaction mixture, and incubated for 15 min at room temperature in dark place. To visualize the modifications, DBCO-Cy3 (Sigma-Aldrich, from 10 mM stock solution in DMF; final concentration, 100 μM) was added to the reaction mixture and incubated for 1 h at 37 °C. After reactions, the modified proteins were separated by Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) with 14% gels. Fluorescence image was obtained with a Molecular Imager PharosFX Plus (Bio-Rad Laboratories). The gel was visualized with Coomassie brilliant blue (CBB) stain.

UP1 modification for identification of modification site

TERRA was annealed by incubation at 95 °C, followed by cooling to 4 °C at a rate of 2 °C /min in 50 mM Tris-HCl buffer (pH 7.5) containing 100 mM KCl. Modification reaction was performed in a final volume 30 μL in the KCl buffer. Hemin (from 100 μM stock solution in DMF; final concentration, 1 μM) and TERRA (final concentration 1 μM) were incubated for 30 min at 4 °C, then purified UP1 (final concentration 3 μM), H_2O_2 (final concentration 200 μM) and N-Me-Lumi (from 100 mM stock solution in DMSO; final concentration, 1 mM) were added to the reaction mixture and incubated for 1 h at 4 °C. After reactions, the modified proteins were separated by SDS-PAGE with 14% gels. Separated bands by SDS-PAGE were cut out, then cut excised bands (ca. 1 mm pieces). Gel pieces were transferred into tubes, then water 1 mL was added to the tubes. The tubes were incubated at 37 °C for 10 min, then the solution was removed (repeat 3 times). A solution of 50% CH_3CN in 100 mM NH_4HCO_3 aq. was added to the tubes for gel de-staining. The tubes were incubated at 37 °C for 10 min, then removed the solution in the tubes. CH_3CN was added to the tubes for dehydration. The tubes were incubated at 37 °C for 10 min, then the solution was removed. Dithiothreitol (100 mM) in 100 mM NH_4HCO_3 aq. was added to the tubes for Cys reduction. The tubes were incubated at 37 °C for 30 min, then the solution was removed. Iodoacetamide (250 mM) in 100 mM NH_4HCO_3 aq. was added to the tubes for Cys alkylation. The tubes were incubated at room temperature for 30 min in the dark, then the solution was removed. The gels were washed with 100 mM NH_4HCO_3 aq. and 50%

CH_3CN in 100 mM NH_4HCO_3 aq. CH_3CN was added to the tubes for dehydration. The tubes were incubated at 37 °C for 10 min, then the solution was removed. A trypsin solution was added. The tubes were incubated at 37 °C overnight. Obtained solution was quenched by TFA aq. (final conc. 0.5%). The sample was desalted using C18 pipette tips (Nikkyo Technos Co., Ltd.). After desalting, the solvent was removed with a centrifugal evaporator. The resulting peptide was dissolved in 5% acetonitrile solution in 0.5 %TFA aq. (15 μL).

NanoLC-MS/MS for identification of UP1 modification site

NanoLC-MS/MS analysis was performed by LC-nano-ESI-MS composed of a quadrupole-orbitrap hybrid mass spectrometer (Q-Exactive; Thermo Fisher Scientific) equipped with a nanospray ion source and a nano HPLC system (Easy-nLC 1000; Thermo Fisher Scientific). The trap column used for the nano HPLC was a 2 cm × 75 μm capillary column packed with 3 μm C18-silica particles (Thermo Fisher Scientific) and the separation column was a 12.5 cm × 75 μm capillary column packed with 3 μm C18-silica particles (Nikkyo Technos Co., Ltd., Japan). The flow rate of the nano HPLC was 300 nL/min. The separation was conducted using a 10−40% linear acetonitrile gradient at 30 min in the presence of 0.1% formic acid. The nanoLC-MS/MS data were acquired in data-dependent acquisition mode controlled by Xcalibur 4.0 (Thermo Fisher Scientific). The settings of data-dependent acquisition were as follows: the resolution was 70,000 for a full MS scan and 17,500 for MS2 scan; the AGC target was 3.0E6 for a full MS scan and 5.0E5 for MS2 scan; the maximum IT was 60 ms for both a full MS scan and MS2 scan; the scan range was 310−1,500 m/z for a full MS scan and 200−2,000 m/z for MS2 scan; and the top 10 signals were selected for MS2 scan per one full MS scan. The MS/MS spectra were searched against the UP1 amino acid sequence (shown in Figure S5 (a)) using the Sequest algorithm within the Proteome Discoverer 2.1 (Thermo Fisher Scientific). To identify the modification site in UP1 protein, +174.0429 Da ($C_9H_6N_2O_2$) was set as a dynamic modification for Tyr, Trp, and Cys residues.

Measurements of UV Absorption Spectra

TERRA (final concentration 1μM), which was folded by heating, and hemin (from 100 μM stock solution in DMF; final concentration, 0.2, 0.4, 0.6, 0.8 and 1.0 μM) were incubated for 30 min at 4 °C, then purified proteins (final concentration 3 μM) were added to the reaction mixture and incubated for 1 h at 4 °C. UV Absorption spectra were recorded on a model V-630 UV-VIS spectrophotometer (Jasco) using a 1 cm path length cell at room temperature.

Protein modification of cell lysate

HeLa S3 cells were maintained in Roswell Park Memorial Institute medium supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. For assays, cells were cultured in 6-well plates. The cells were suspended in K-Ca buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 2.5 mM $CaCl_2$). The supernatants were lysed by sonication (model UR-20P, Tomy Seiko) and centrifuged at 16200×g for 10 min at 4 °C. The lysate concentrations were determined using a BCA Protein Assay Kit (Thermo Scientific). Modification reaction was performed in a final volume 400 μL in the K-Ca buffer. Hemin (from 10 mM stock solution in DMF; final concentration, 100 μM) and TERRA (final concentration 10 μM) were incubated for 30 min at 4 °C, then cell lysate (final concentration 2.0 mg/mL), H_2O_2 (final concentration 1 mM) and **1** (from 100 mM stock solution in DMSO; final concentration, 1 mM) were added to the reaction mixture with 0.2 U/mL SUPERase In RNase Inhibitor (Thermo Scientific) and incubated for 1 h at 4 °C. For Cys alkylation, iodoacetamide (final concentration 10 mM) was added to the reaction mixture and incubated for 90 min at room temperature in dark place. To collect the modified proteins, DBCO-Biotin ((DBCO-PEG4-biotin conjugate, Sigma-Aldrich) from 10 mM stock solution in DMF; final concentration, 100 mM) was added to the reaction mixture and incubated for 1 h at 37 °C.

NanoLC-MS/MS sample preparation

The labeled lysate (100 μL) was used for next step. Excess DBCO reagent was removed using a Sephadex G-25 gel (GE Healthcare) filtration column (2000×g, 4 min). The solution was added Dynabeads™MyOne™Streptavidin C1 (Invitrogen) (1.0 mg) and shaken at room temperature for 60 min. The beads were washed with 1 M urea in PBS (200 μL) five times and 100 mM Tris buffer pH 8.0 twice. Then, the beads were treated with 0.5 μg Trypsin Gold (Promega) in 100 mM Tris buffer pH 8.0 (200 μL), and incubated at 37 °C overnight. Obtained solution was quenched by TFA aq. (final conc. 0.5%). The sample was desalted using C18 pipette tips (Nikkyo Technos Co., Ltd.). After desalting, the solvent was removed with a centrifugal evaporator. The resulting peptide was dissolved in 5% acetonitrile solution in 0.1 %TFA aq. (15 μL).

NanoLC-MS/MS for identification of labeled protein in HeLa cell lysate

NanoLC-MS/MS analysis was performed by Q-Exactive (Thermo Fisher Scientific) and Easy-nLC 1000 (Thermo Fisher Scientific). The settings for the measurement was almost the same as the condition to identify the UP1 modification site; only the gradient time was changed into 70 min because the samples were assumed to be more complex (i.e. contain many kinds of proteins). Each sample was measured three times. The MS/MS spectra search and the calculation of the relative abundance was conducted by Proteome Discoverer 2.4 (Thermo Fisher Scientific). The sequence library for the peptide search was obtained from UniProt database (Proteome ID: UP000005640). The relative abundance in each protein was calculated by MS1 quantification

(LFQ) method using “Pairwise Ratio Based” mode in Proteome Discoverer 2.4. To ensure the validity of the results, the proteins whose detected peptide numbers (Number of Peptides) were only 1 were omitted. The information about Nucleotide- or RNA-binding properties for the proteins whose increase rates were more than the third quartile score (178 proteins) were manually curated with UniProt database.

3. References

1. R. Yagi, T. Miyazaki, T. Oyoshi, G-quadruplex binding ability of TLS/FUS depends on the β-spiral structure of the RGG domain. *Nucleic Acids Res.* 2018, **46**, 5894-5901.