Nucleic Acids Research, 2024, **52**, e36
https://doi.org/10.1093/nar/gkae125
Advance access publication date: 26 February 2024
Methods

Symmetry-breaking malachite green as a near-infrared light-activated fluorogenic photosensitizer for RNA proximity labeling

Lan Li[1], Jinghua Han[1,†], Hei-Yong G. Lo[2,†], Winnie Wai Ling Tam[1,3], Han Jia[1], Edmund Chun Ming Tse[1,3,4], J. Matthew Taliaferro[2,*] and Ying Li[1,3,*]

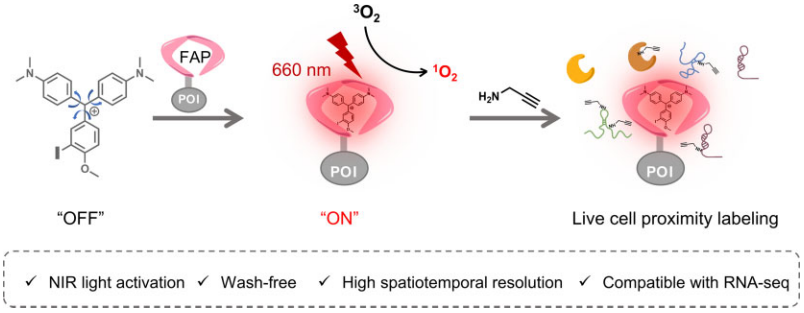
[1]Department of Chemistry, The University of Hong Kong, Hong Kong 999077, China
[2]Department of Biochemistry and Molecular Genetics, RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[3]Laboratory for Synthetic Chemistry and Chemical Biology Limited, New Territories, Hong Kong 999077, China
[4]CAS–HKU Joint Laboratory of Metallomics on Health and Environment, The University of Hong Kong, Hong Kong 999077, China

*To whom correspondence should be addressed. Tel: +852 2219 4824; Email: yingli0e@hku.hk
Correspondence may also be addressed to J. Matthew Taliaferro. Tel: +1 303 724 3274; Email: matthew.taliaferro@cuanschutz.edu
†The second and third authors contributed equally.

Abstract

Cellular RNA is asymmetrically distributed in cells and the regulation of RNA localization is crucial for proper cellular functions. However, limited chemical tools are available to capture dynamic RNA localization in complex biological systems with high spatiotemporal resolution. Here, we developed a new method for RNA proximity labeling activated by near-infrared (NIR) light, which holds the potential for deep penetration. Our method, termed FAP-seq, utilizes a genetically encoded fluorogen activating protein (FAP) that selectively binds to a set of substrates known as malachite green (MG). FAP binding restricts the rotation of MG and rapidly activates its fluorescence in a wash-free manner. By introducing a monoiodo modification to MG, we created a photosensitizer (MG-HI) with the highest singlet oxygen generation ability among various MG derivatives, enabling both protein and RNA proximity labeling in live cells. New insights are provided in the transcriptome analysis with FAP-seq, while a deeper understanding of the symmetry-breaking structural arrangement of FAP–MG-HI was obtained through molecular dynamics simulations. Overall, our wash-free and NIR light-inducible RNA proximity labeling method (FAP-seq) offers a powerful and versatile approach for investigating complex mechanisms underlying RNA-related biological processes.

Graphical abstract



Introduction

Increasing evidence suggests that well-regulated RNA localization in live cells is essential for cell survival (1–3). RNA localization is a complex process that is intimately associated with various aspects of RNA's life cycle, including transcription, processing, translation, degradation and post-transcriptional modifications, many of which rely on the interactions with other biomolecules, especially RNA-binding proteins (4–9). RNA localization plays a crucial role in achieving the biological functions of RNA. For example, architectural noncoding RNAs are involved in chromatin organization for gene expression regulation, while certain messenger RNAs are responsible for local translation in neuronal dendrites (10–13). Dysregulation of RNA localization has been linked to a number of diseases, including but not limited to cancers and neurodegenerative diseases (14–17). Better profiling of RNA localization can lead to comprehensive understanding of the underlying disease mechanisms related to RNA.

Only a limited number of strategies have been developed to study RNA localization on the transcriptomic level via

Received: November 2, 2023. **Revised:** January 17, 2024. **Editorial Decision:** February 6, 2024. **Accepted:** February 8, 2024

proximity labeling methods. These methods can be classified into three categories based on how spatially restricted reactive intermediates are generated: utilization of enzymes (18–20), construction of protein–ligand systems (21–23) and development of small functional molecules (24). Among these, labeling methods relying on singlet oxygen ($^1O_2$) generated via subcellular localized photosensitization offer precise on–off control by light irradiation, leading to RNA and protein labeling with high spatiotemporal resolution (Supplementary Figure S1) (19,21,24–27). The genetically encoded photosensitizer miniSOG has been applied in subcellular transcriptome mapping with blue light irradiation for 15–20 min (19). The synthetic photosensitizer dibromofluorescein (DBF) conjugated to HaloTag ligand (HaloDBF) is able to achieve the transcriptome mapping via green light irradiation within 3–5 min (21,22). However, HaloDBF remains functionally active after cellular uptake, making careful removal of nonspecific HaloDBF crucial to obtain high labeling precision. Nevertheless, neither miniSOG nor HaloDBF can be activated with longer wavelength irradiation, which is essential for deeper penetration in complex biological systems (19,21,28).

A fluorogenic near-infrared (NIR) photosensitizer is the ideal candidate for next-generation proximity labeling methods. The fluorogen activating protein (FAP)–malachite green (MG) system offers a potentially effective approach to construct a wash-free NIR photosensitizer for proximity labeling with high spatiotemporal resolution (29–31). In this system, fluorescence is only activated when FAP restricts the rotation of MG derivatives through binding, and this mechanical force-induced fluorogenic property is tolerant of slight structural modifications, such as heavy-atom substitution for converting a fluorophore into a photosensitizer (31–33). However, other fluorogenic NIR fluorophores, such as silicon rhodamine derivatives, may not exhibit such substrate tolerance, as their fluorogenic property relies on lactone–zwitterion equilibrium controlled by precise molecular electronic distribution (34,35). We recently explored the scope of halogen-substituted MG substrates for the FAP–MG system and discovered that the monobrominated MG exhibits better fluorescence turn-on as well as $^1O_2$ generation than the dibrominated version upon NIR light irradiation (33). We further hypothesize that the monoiodinated MG (MG-HI) can afford the best $^1O_2$ generation yield among other MG derivatives and maintain a decent fluorescence quantum yield, which potentially renders MG-HI an efficient no-wash NIR photosensitizer for RNA proximity labeling, compatible with imaging-assisted studies.

In this study, we reported the development of FAP–MG-HI as a novel fluorogenic NIR photosensitizer for proximity labeling. First, the $^1O_2$ generation capacity of FAP–MG-HI with NIR light irradiation and its RNA proximity labeling ability were thoroughly investigated *in vitro*. Then, cellular imaging studies were carried out to demonstrate the wash-free fluorescence turn-on and proximity labeling ability of MG-HI in FAP-expressed subcellular locations. Protein labeling analysis upon cell fractionation and RNA labeling analysis via quantitative reverse transcription polymerase chain reaction (RT-qPCR) further validated the accuracy of proximity labeling. In addition, *in cellulo* wash-free RNA proximity labeling via NIR light irradiation was validated and discussed at the transcriptomic level, termed as FAP-seq. Lastly, molecular dynamics simulations provided new insight into the structural arrangement of FAP–MG-HI.

Materials and methods

This section describes only the materials and methods relevant to proximity labeling workflow, i.e. NIR light-activated proximity labeling in living cells and subsequent processing for fluorescent confocal imaging, RNA dot blot, RNA library preparation, RNA sequencing data analysis, protein in-gel fluorescence and western blot. A full description of the materials and methods used in this study is provided in Supplementary Data.

Mammalian cell culture

HEK293T cells and the two home-made HEK293T cells stably expressing FAP fusion protein in nucleus (Nucleus-FAP) and endoplasmic reticulum membrane (ER-FAP) were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin, and grown at 37°C and 5% CO_2. Cells were seeded on confocal dishes for live cell imaging and seeded on glass coverslips that were pretreated with poly-D-lysine for the fixed cell imaging process. For Nucleus-FAP and ER-FAP cell line generation, transfection was conducted in HEK293T cells with jetPRIME transfection reagent (Cat. #114-07, PolyPlus transfection) according to the manufacturer's instruction in Opti-MEM medium. The transfection medium was replaced with growth medium 4 h after transfection and the puromycin (1.5–2 μg/ml) selection started 24 h after transfection. After 2 weeks of puromycin selection, limiting dilution was performed on the cells (~0.3 cell/well in 96-well plates) for monoclonal selection. The FAP expression level and localization were verified by the fluorescence images of cells incubating with MG-HI and subcellular organelle trackers.

MG derivative-mediated *in cellulo* labeling

Prior to the labeling process, cells were washed with Hank's balanced salt solution (HBSS) to remove any residual medium. Subsequently, cells were treated with freshly prepared MG probes in a DMEM solution for 20 min. The cells were rinsed once with HBSS and further incubated with propargylamine (PA) in HBSS for 3 min, followed by 660 nm light irradiation. The cells underwent two additional washes with HBSS and were ready for downstream procedures, such as RNA isolation, protein extraction or slide preparation for confocal microscopy imaging.

Imaging of the *in situ* labeling

Following the labeling process, cells adherent to coverslips were fixed with 4% paraformaldehyde (PFA) for 15 min, followed by permeabilization with 0.1% Triton X-100 in Dulbecco's phosphate-buffered saline (DPBS) for another 15 min. Subsequent rinses were performed twice with DPBS for 5 min each on an orbital shaker. Blocking was achieved by incubating cells with a solution of bovine serum albumin (1 mg/ml BSA, 0.45% NaCl and 0.025% NaN_3 in DPBS) for 30 min, followed by two DPBS washes. The coverslips were then incubated with a copper-catalysed azide–alkyne cycloaddition (CuAAC) reaction solution [1 mM $CuSO_4$, 4 mM tris-hydroxypropyltriazolylmethylamine (THPTA) ligand, 10 mM sodium ascorbate and 15 μM Azide-Fluor 545] for 1 h at 37°C in a dark environment. Cells were washed three times with DPBS containing 0.1% Triton X-100 and once more with DPBS alone, each time on an orbital shaker. For nuclear

staining, cells were treated with 5 μg/ml Hoechst stain in DPBS for 3 min, followed by two additional DPBS washes before being mounted for microscopy. Imaging was conducted using fluorescence confocal microscopy with a 63× oil immersion objective on a Leica microscope. The Hoechst stain was visualized using a $\lambda_{ex}/\lambda_{em}$ of 405/410–490 nm, while the Azide-Fluor 545 was imaged with 552/557–735 nm.

RNA isolation

Following the labeling process, cells were immediately lysed for total RNA extraction using RNAiso Plus (Cat. #9109, Takara) according to the manufacturer's protocol for adherent cells. The resulting RNA pellet was resuspended in RNase-free water (NF H_2O) and the RNA sample is suitable for long-term storage at −80°C until further processing. Subsequently, the RNA was treated with Turbo DNase (Cat. #2238, Thermo Fisher) at 37°C for 20 min with agitation at 300 rpm to remove any contaminating DNA. This step was followed by a proteinase K digestion (Cat. #25530049, Invitrogen) at 42°C for 20 min with agitation at 300 rpm, to eliminate protein contaminants. The RNA was purified using a Zymo-Spin column, choosing the small capacity (Cat. #1004) or large capacity (Cat. #1011) option from Zymo Research, in accordance with the manufacturer's guidelines. The purified RNA sample, resuspended in NF H_2O, was aliquoted into 1.5-ml tubes and stored at −80°C for future experiments.

RNA biotinylation through CuAAC

To evaluate the effect of the CuAAC reaction on RNA integrity, RNA samples were treated with different click chemistry solutions. The solution was prepared with either 0 or 0.1 mM $CuSO_4$, 2 mM THPTA ligand, 10 mM sodium ascorbate, and either 2 mM biotin-picolyl-azide or 2 mM biotin-EG3-azide, in 10 mM Tris buffer at pH 7.0. The RNA concentration was set at 0.2 μg/μl, and the reaction was carried out with agitation at 500 rpm and a temperature of 25°C for durations ranging from 10 to 25 min. After optimizing the conditions for RNA biotinylation, the RNA was incubated in the click solution containing 0.1 mM $CuSO_4$, 2 mM THPTA ligand, 10 mM sodium ascorbate and 2 mM biotin-picolyl-azide, in 10 mM Tris buffer at pH 7.0. This was also done at a RNA concentration of 0.2 μg/μl, with agitation at 500 rpm, at 25°C for 10 min. Following the reaction, the RNA was purified using a Zymo-Spin column and eluted with NF H_2O. The RNA concentration was then adjusted to ∼0.5–1 μg/μl. The samples are suitable for long-term storage at −80°C before proceeding to the next step in the experimental workflow.

RNA dot blot analysis

A HybondN+ membrane (Cat. #83-378, GE Healthcare, Genesee Scientific) was first soaked in a solution containing 1.5 M NaCl and 0.15 M sodium citrate for 10 min and then allowed to air dry. To ensure consistency, RNA samples loaded to the membrane were labeled and biotinylated in the same batch. Equal amounts of RNA samples (either 0.5 or 1 μg with similar concentrations) were loaded onto the membrane. After air drying for 5 min, the RNA was UV cross-linked to the membrane using a Stratalinker UV cross-linker set to 2500 μJ/m^2. Subsequently, the membrane was blocked in a blocking buffer [0.12 M NaCl, 0.016 M Na_2HPO_4, 0.008 M NaH_2PO_4 and 0.17 M sodium dodecyl sulfate (SDS)] for 30 min. The membrane was then incubated with a streptavidin–HRP (horseradish peroxidase) conjugate from Pierce (1:10 000 dilution, Cat. #ab7403, Abcam) for 10 min in a fresh blocking buffer. The membrane was washed twice with a diluted form of the blocking buffer (0.1×), and then twice with a washing buffer (10 mM Tris base, 10 mM NaCl, 2 mM $MgCl_2$). For detection, the membrane was treated with Clarity Max Western ECL Substrate (Cat. #1705062, Bio-Rad) and the signal was captured using a ChemiDoc MP imaging system (Bio-Rad). To visualize the loading control, the used membrane was stained with methylene blue [0.04% (w/v) methylene blue in deionized water, 0.3 M sodium acetate] for 30 min, followed by several rinses with deionized water. The relative intensity of the biotin signal was quantified using the volume tool in the Bio-Rad Image Lab Software.

RNA enrichment

Cells expressing Nucleus-FAP and ER-FAP in 10-cm plates were labeled *in situ* with either 1.5 or 0 mM PA, in the presence of 125 nM MG-HI and 1 min of 660 nm light irradiation. To minimize technical variation, RNA samples from both Nucleus-FAP and ER-FAP cells were processed for biotinylation and enrichment simultaneously. RNA was subsequently extracted and treated with DNase and then with proteinase K. Biotinylation was carried out at 25°C with agitation at 500 rpm for 10 min. Dynabeads MyOne Streptavidin C1 beads (Cat. #65001, Invitrogen) (100 μl for 50 μg biotinylated RNA) were prewashed three times with 1 ml W&B buffer (1 M NaCl, 100 mM Tris, pH 7. 5, 10 mM EDTA, 0.2% Tween 20), for 10 min per wash. The beads were then washed twice with 1 ml of Wash A (0.1 M NaOH, 0.05 M NaCl, 0.1% Tween 20) for 2 min each, followed by a single 2-min wash with 1 ml of Wash B (0.1 M NaCl, 0.1% Tween 20). Post-washing, the beads were resuspended in 100 μl of 2× W&B buffer (2 M NaCl, 200 mM Tris, pH 7.5, 200 mM EDTA, 0.4% Tween 20). For binding, the C1 beads in 100 μl of 2× W&B buffer were combined with 100 μl of RNA samples, to which 2 μl of RNaseOut was added. The mixture was incubated at room temperature for 1.5 h with gentle rotation. Samples were placed on a magnetic stand to separate the beads, and the supernatant, designated as flow-through (FT), was transferred to a new tube. The FT was then purified using a Zymo column and set aside for dot blot analysis. The magnetic beads were washed three times with 1× W&B buffer, for 10 min per wash. Beads were transferred to a new tube for the final wash. The enriched RNA was eluted off the beads with 50 μl elution buffer (44 μl formamide, 1 μl of 500 mM EDTA, pH 8.0, 5 μl of 50 mM D-biotin) by heating at 65°C for 5 min, and the supernatant containing the eluted RNA was collected. A second round of elution was performed by heating the beads with another 50 μl of elution buffer at 90°C for 5 min. The combined eluted RNA solution (100 μl total) was then mixed with 50 μl of NF H_2O, 300 μl of RNA binding buffer (Cat. #1013-2, Zymo Research) and 450 μl of 100% ethanol, followed by purification using a Zymo column. The RNA was eluted in two portions with 8 μl of NF H_2O each, yielding a total of 16 μl. From this, 12 μl was set aside for library preparation, while the remaining (∼2.5 μl) was diluted with 7.5 μl of NF H_2O for further dot blot and RT-qPCR. Specifically, 0.5 μl of the diluted RNA solution was utilized in dot blot analysis as a means of quality control for the enrichment process.

RT-qPCR

The complementary DNA (cDNA) synthesis for both 'Enrich' RNA (post-enrichment) and 'Input' RNA (total biotinylated RNA prior to enrichment) was performed using the PrimeScript RT Reagent Kit (Cat. #RR037A, Takara). For the 'Enrich' RNA, 6 μl of the eluted sample was further diluted with 1.5 μl NF H_2O to reach a total volume of 7.5 μl. Similarly, 250 ng of 'Input' RNA was diluted to the same volume with NF H_2O. Each RNA sample was combined with 0.5 μl of oligo(dT) primers and 0.5 μl of random 6-mer primers. These mixtures were then heated in a thermocycler at 65°C for 5 min. Once the thermocycler is cooled to 4°C, 2 μl of 5× PrimeScript Buffer and 0.5 μl of PrimeScript RT Enzyme Mix I were added to each sample on ice. The samples were mixed thoroughly and returned to the thermocycler for cDNA synthesis, following a program of 25°C for 10 min, 42°C for 45 min, 85°C for 5 s and a final hold at 4°C. The resulting cDNA was then diluted with 20 μl of NF H_2O in the preparation for qPCR. For the semiquantitative SYBR Green PCR, 3 μl of the diluted 'Input' or 'Enrich' cDNA was mixed with 1 μl of a forward/reverse primer mixture, 2.25 μl NF H_2O and 6.25 μl of 2× TB Green Ex Taq (Cat. # RR420, Takara). The qPCR reactions were carried out on a Bio-Rad CFX Duet Real-Time PCR System. To calculate the fold of enrichment, the following equation was used: fold enrichment = $2^{\wedge}[(Ct_{Enrich\text{-}PA} - Ct_{Input\text{-}PA}) - (Ct_{Enrich+PA} - Ct_{Input+PA})]$. In this equation, Ct represents the cycle threshold value in qPCR, while +/–PA indicate the presence or absence of PA during the RNA labeling step, respectively. The RT-qPCR data were plotted with four replicates.

FAP-seq library preparation

Positive samples (labeled with 1.5 mM PA in FAP-seq) were used for library preparation due to the inability to enrich the negative controls (samples labeled without PA) to a detectable level, as determined by concentration measurements using the Qubit 4 Fluorometer (Invitrogen). Specifically, 12 μl of 'Enrich' RNA and 30 ng of 'Input' RNA were used in the library preparation. Ribosomal RNA (rRNA) depletion, an essential step to enhance the sequencing of non-rRNA, was conducted using the NEBNext rRNA Depletion Kit v2 (Cat. #E7400, NEB) in conjunction with RNA Sample Purification Beads (Cat. #E7405, NEB), following the manufacturer's guidelines. Subsequently, the RNA samples underwent FAP-seq library preparation employing the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (Cat. #E7760, NEB). This process also involved the use of NEBNext Sample Purification Beads (Cat. #E7767, NEB) for purification steps and NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) (Cat. #E7335, NEB) for indexing. The manufacturer's protocol was closely followed except for a fragmentation time of 9 min, and a library size selection ratio of 1.5 and 3.0, to ensure proper library quality and size range for subsequent high-throughput sequencing.

FAP-seq library data analysis

FAP-seq reads were trimmed of their adapters using cutadapt v4.2.0 (36). Transcript abundances were then calculated using salmon v1.9.0 (37) against transcript sequences from Gencode v28. Transcript enrichments between samples were calculated using DESeq2 (38). For comparisons between the data from Nucleus-FAP and ER-FAP cell lines, DESeq2 was again used with a model that contained an interaction term allowing direct comparison between the enrichments in Nucleus-FAP and ER-FAP cell lines. For the splicing analysis, transcript abundances were calculated using salmon and a fasta file containing spliced and unspliced sequences for all transcripts. For every gene, unspliced/spliced ratios were then calculated by comparing the summed abundances of all unspliced transcript forms to the summed abundance of all spliced transcript forms.

Protein isolation and fractionation

Cells expressing Nucleus-FAP and ER-FAP in 10-cm plates were labeled *in situ* using 250 nM MG-HI and 3 mM PA, with either no light exposure or a 2-min irradiation with 660 nm light. The cells were washed once with cold DPBS and then pipetted off the plates with 2 ml cold DPBS. A 250 μl aliquot of this cell suspension was set aside in a separate tube for total protein extraction, while the remaining suspension was used for protein fractionation. For labeling mediated by HoeDBF, cells were incubated with 5 μM HoeDBF in HBSS for 30 min. After rinsing once with HBBS, the cells were incubated with fresh medium for 20 min twice. Subsequently, the cells were treated with 2 mM PA and exposed to 513 nm light irradiation for 2 min.

Total protein (labeled 'Total') was obtained by lysing cells in 300 μl of 1% SDS–HEPES buffer (150 mM NaCl, 50 mM HEPES, 2 mM $MgCl_2$, 10% glycerol, pH 7.4, 1% SDS) followed by sonication for 1 min and 20 s (with cycles of 5 s on and 5 s off at 20% amplitude). The reagents used for fractionation were pre-chilled on ice, and the fractionation procedures were carried out on ice or in a centrifuge set to 4°C. The Mem-PER Plus Membrane Protein Extraction Kit (Cat. # 89842, Thermo) was employed for protein fractionation, following the manufacturer's instructions. For protein fractionation using the HLB buffer, the steps were executed in accordance with a previously reported protocol, with minor modifications (39). Briefly, cells were centrifuged at 300 × *g* for 5 min, and the supernatant DPBS was discarded. Subsequently, 0.4 ml of HLB buffer [10 mM Tris, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 10% glycerol (v/v)] with 0.05% Tween 20 was added to resuspend the pellet. The cells were gently agitated to ensure a uniform suspension and incubated on ice for 10 min with shaking. The mixture was centrifuged at 1000 × *g* for 4 min. The supernatant containing the cytosolic proteins (labeled 'Cytosol') was collected. The remaining pellet was then resuspended in 0.4 ml of HLB buffer supplemented with 0.05% NP-40, pipetted until homogeneous and incubated on ice with shaking for 10 min. After a subsequent centrifugation at 1000 × *g* for 4 min, the supernatant (labeled 'Wash') was collected. Finally, the remaining pellet was resuspended in 0.4 ml of 1% SDS–HEPES buffer and sonicated for 1 min and 20 s (with cycles of 5 s on and 5 s off at 20% amplitude) to solubilize the chromatin-bound proteins (labeled 'Nucleus').

In-gel fluorescence

Protein fractions at the same concentration were incubated in a click chemistry solution (0.33 mM $CuSO_4$, 0.66 mM THPTA ligand, 2.5 mM sodium ascorbate and 20 μM Azide-Fluor 545, achieving a protein concentration of ∼0.6 μg/μl). This mixture was agitated at 500 rpm and 37°C for 1 h in dark. After the click chemistry reaction, the protein fractions were mixed with loading dye, denatured at 95°C for 5 min

and then rapidly cooled on ice. An equal volume (~10 μl) from each protein fraction was loaded onto 12% SDS–PAGE (polyacrylamide gel electrophoresis) gels. For total protein analysis, one piece of gel was run with a protein ladder. For analysis of proteins after fractionation, three pieces of gels were run concurrently, one for in-gel fluorescence and Coomassie blue staining without a protein ladder, and the other two for western blot analysis of subcellular localization markers, including a protein ladder. The protein gels were run in an SDS-containing buffer and the run was halted when the front-loading dye exited the gels. The gel designated for in-gel fluorescence was washed with deionized water and visualized using the Alexa 546 channel on the imaging system. Subsequently, the gel was incubated with PageBlue protein staining solution (Cat. #24620, Thermo) and imaged to capture the Coomassie staining.

Western blot analysis

The proteins resolved in the gel were transferred to a nitrocellulose membrane (Cat. #1620112, Bio-Rad) for western blot analysis using a Trans-Blot Turbo Transfer System (Bio-Rad). The membrane was blocked in the blocking buffer [0.5% (w/v) non-fat milk, 0.1% Tween 20 in PBS (PBST)] for 0.5 h and then washed twice with PBST. The membrane was then incubated with a primary antibody (1:1000 dilution, rabbit mAb, Cell Signaling Technology) in the blocking buffer overnight at 4°C. After washing twice with PBST, the membrane was incubated with an HRP-conjugated secondary antibody (1:10 000 dilution, anti-rabbit IgG) in PBST for 1 h. After three additional washes with PBST, the membrane was developed using Clarity Max Western ECL Substrate (Cat. #1705062, Bio-Rad) and imaged on the ChemiDoc MP imaging system (Bio-Rad).

Results

Characterization of FAP–MG-HI *in vitro*

The photophysical properties of the newly synthesized MG-HI were characterized and compared to other reported MG derivatives (Figure 1A) (33). As shown in Figure 1B, the free MG-HI was in a fluorescent off state because of the fast rotation upon excitation. Upon binding to FAP, MG-HI exhibited an absorption redshift and intense fluorescence turn-on due to encapsulation-induced rotation restriction (32,33). The fluorescence activation of MG-HI occurred within 1 min after being added to a FAP solution and remained stable for 3 h (Figure 1C). In contrast, ~70% of the free MG-HI decayed during the 3-h incubation in DPBS buffer, possibly due to the reaction between the cationic central carbon ion in MG-HI and the OH^- in aqueous solution (Supplementary Figure S2A) (33,40). After 24-h incubation in DPBS buffer, the decayed MG-HI showed weak fluorescence in a FAP solution, and the fluorescence was restored by adding extra fresh MG-HI, suggesting weak interaction between decayed MG-HI and FAP (Supplementary Figure S2B). The binding studies based on fluorescence turn-on demonstrated a tight 1:1 binding between MG-HI and FAP, with a dissociation constant (K_D) of 61.7 nM, similar to the other halogen-modified MG derivatives (Supplementary Figure S2C–E and Supplementary Table S1). These combined results indicate that the strong binding between MG-HI and FAP can sufficiently protect MG-HI from degradation, making FAP–MG-HI a promising candidate for live cell applications. The viscosity response of MG-HI and its interaction with BSA were also evaluated (Supplementary Figure S3). The weak fluorescence emission of MG-HI in BSA-containing solutions and viscous glycerol/water mixtures suggests the low possibility of nonspecific labeling induced by local subcellular environments under 660 nm NIR light exposure. Similar to the other monohalogen-substituted MG-HBr, the symmetry-breaking MG-HI also features a blueshifted absorption wavelength as well as a fluorescence quantum yield ($\Phi_F = 10.3\%$) higher than those of the dihalogen-substituted MG derivatives, including dichloro-substituted ($\Phi_F = 7.3\%$), dibromo-substituted ($\Phi_F = 7.6\%$) and diiodo-substituted ($\Phi_F = 8.5\%$) MG derivatives (Figure 1D and E, and Supplementary Table S1).

We further assessed the complexes' RNA proximity labeling capacity and mechanism *in vitro*. The $^1O_2$ generation capability was measured and compared via the anthracene-9,10-dipropionic acid disodium salt (ADPA) photobleaching assay (41). Even with a 675-nm light source whose wavelength coverage fits the absorbance of dihalogen-substituted MG derivatives better, FAP–MG-HI exhibited the highest ADPA bleaching rate (Figure 2A). FAP–MG-HI demonstrated high fluorescence quantum yield as well as $^1O_2$ generation yield, confirming our original hypothesis that a symmetry-breaking heavy-atom modification on MG leads to both higher $^1O_2$ generation and fluorescence emission. The ADPA bleaching studies conducted with MG derivatives in acetonitrile and DPBS buffer showed that the free MG derivatives had negligible $^1O_2$ generation ability (Supplementary Figure S4). Dot blot analysis further verified the *in vitro* labeling efficiency at the RNA level. Total RNA extracted from HEK293T cells was tagged with PA in the presence of FAP–MG complexes upon 675 nm light irradiation. The PA tagging efficiency was assessed by biotinylation through CuAAC. Consistent with the ADPA bleaching assay, FAP–MG-HI demonstrated the highest biotinylation level (Figure 2B). The RNA dot blot analysis suggested that FAP–MG-HI afforded a much more significant RNA labeling than the other five complexes when RNA labeling was carried out with a light source centered at 660 nm, which is close to the maximum absorbance of FAP–MG-HI (Supplementary Figure S5). This 660 nm NIR light irradiation also exhibits significantly deeper tissue penetration than blue and green ones (Supplementary Figure S6). The result further underscores the high efficiency of the FAP–MG-HI system over blue and green light-triggered proximity labeling tools within complex biological systems.

The photosensitizing pathway for proximity labeling was examined by conducting and comparing $^1O_2$-mediated ADPA bleaching assay and radical-mediated DHR123 fluorescence turn-on assay (42–44). Compared to free MG-HI, FAP–MG-HI exhibited a much stronger effect on ADPA bleaching (Figure 2C). According to the bleaching assay, FAP alone and MG-HI alone afforded no observable $^1O_2$ generation, displaying a similar trend as dimethyl sulfoxide control. As for radical generation evaluated with DHR123, free MG-HI exhibited a linear correlation of DHR123 fluorescence increase over the course of 8 min light exposure (660 nm), while FAP–MG-HI induced a fluorescence signal increase within the first 2 min with similar efficiency as free MG-HI and plateaued with longer light exposure time (Figure 2D). The comparative study between ADPA and DHR123 suggests that FAP–MG-HI is a type II photosensitizer that predominantly generates $^1O_2$

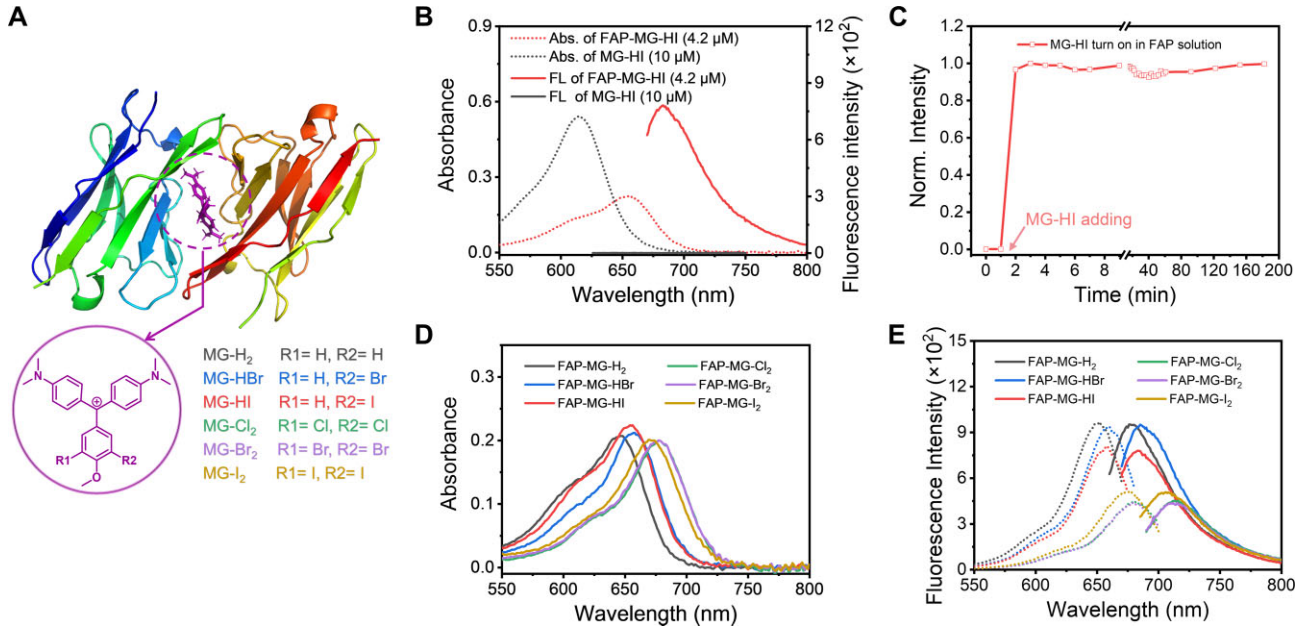


Figure 1. Optical properties of FAP–MG. (**A**) Global structure of the FAP–MG complexes and the chemical structure of MG-H_2, MG-HBr, MG-HI, MG-Cl_2, MG-Br_2 and MG-I_2. (**B**) UV–Vis absorption and fluorescence spectra of 10 μM MG-HI and 4.2 μM FAP–MG-HI in DPBS buffer. (**C**) The fluorescence turn-on kinetics of 4.2 μM MG-HI in 5 μM FAP solution. Comparison of the (**D**) UV–Vis absorption and (**E**) fluorescence spectra of 4.2 μM FAP–MG complexes in DPBS buffer.

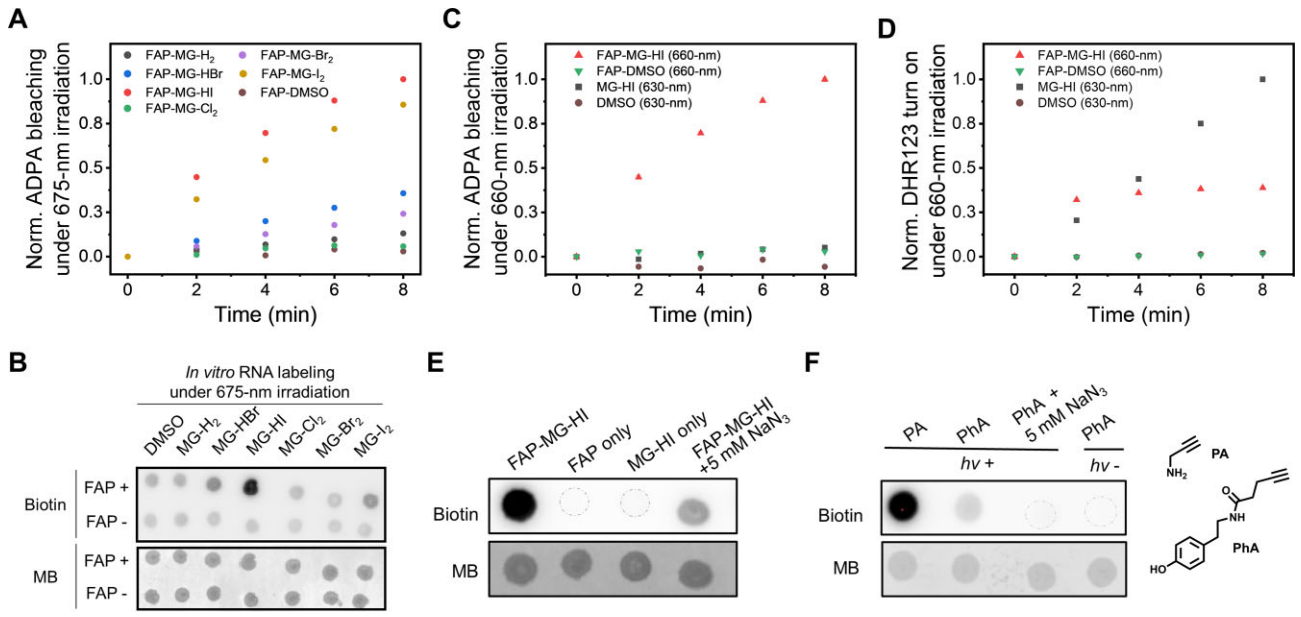


Figure 2. Photooxidation properties of FAP–MG. (**A**) ADPA bleaching assay of 5 μM FAP–MG complexes under 675 nm irradiation. The ADPA bleaching percentage was normalized to the result for FAP–MG-HI at 8 min. (**B**) RNA dot blot of *in vitro* labeled RNA with 5 μM FAP–MG complexes under 675 nm irradiation. (**C**) ADPA bleaching assay of 5 μM FAP–MG-HI and MG-HI. The ADPA bleaching percentage was normalized to the result for FAP–MG-HI at 8 min. (**D**) DHR123 turn-on assay of 5 μM FAP–MG-HI and MG-HI. The DHR123 turn-on percentage was normalized to the result for MG-HI at 8 min. (**E**) RNA dot blot of *in vitro* labeled RNA with 1 μM FAP–MG-HI, FAP, MG-HI or FAP–MG-HI containing 5 mM NaN_3 under 660 nm irradiation. (**F**) RNA dot blot of *in vitro* labeled RNA with PA, PhA, PhA containing 5 mM NaN_3 or PhA without light irradiation. The ADPA fluorescence blenching and DHR123 fluorescence turn-on were monitored at $\lambda_{ex}/\lambda_{em}$ = 380/420 nm and $\lambda_{ex}/\lambda_{em}$ = 480/520 nm, respectively.

instead of radicals. Parallel to the DHR123 fluorescence turn-on in the first 2 min, the fluorescence of MG-HI initially exhibited a slight decrease and then plateaued after 2 min of light exposure (Supplementary Figure S7). The quasi-linear ADPA bleaching over 8 min, coupled with the quickly plateaued fluorescence bleaching of MG-HI, suggests that FAP–MG-HI remains relatively photostable throughout the light exposure period.

To fully investigate the primary mechanism for RNA proximity labeling, RNA dot blot analysis was conducted. A strong RNA biotinylation signal was detected when FAP–MG-HI was employed, while neither FAP nor MG-HI produced any

detectable signals, suggesting that the trace amount of radicals generated by free MG-HI has little effect on the RNA labeling with PA. The $^1O_2$ quencher NaN_3 (45) significantly inhibited RNA biotinylation, further confirming that the proximity labeling was mediated by $^1O_2$ produced during NIR light irradiation (Figure 2E). Phenol derivatives have been widely used in APEX-mediated labeling by forming reactive phenol radicals (18,46). A genetically encoded photosensitizer has achieved RNA proximity labeling with phenol biotin (20,47). To evaluate the feasibility of phenol radicals on RNA proximity labeling mediated by FAP–MG-HI, dot blot analysis was conducted with phenol alkyne (PhA) (46) in comparison to PA. As shown in Figure 2F, PhA-tagged RNA exhibited a much lower signal than that of PA, and the weak labeling signal of PhA was further quenched by adding NaN_3. The observation is consistent with the literature report that phenol radical can be generated by electron transfer from $^1O_2$ to phenol with relatively lower kinetics than direct electron transfer via type I photosensitizing (48). These findings provide strong evidence that FAP–MG-HI mediates RNA proximity labeling through the type II photosensitizing pathway, instead of type I (42).

Evaluation of FAP–MG-HI *in cellulo*

After verifying that FAP–MG-HI is capable of inducing RNA proximity labeling by generating $^1O_2$, we performed *in cellulo* evaluations in two cell lines stably expressing Nucleus-FAP and ER-FAP as proof-of-concept studies (30). The fast cellular uptake and wash-free fluorescence turn-on of MG-HBr and MG-HI were validated by live cell imaging (Figure 3A), together with MG-H_2, which indicated that the halogen atom modifications on MG had no observable impact on uptake kinetics (Supplementary Figure S8). Confocal imaging studies confirmed the excellent co-localization between FAP and MG-HI fluorescence signals in live cells without washing steps, as well as the successful subcellular localization of FAPs to designated locations by validating with known subcellular location markers (Figure 3B and C, and Supplementary Figure S9). In addition, the fluorescence signals of MG-HI in live cells remained observable after the labeling process with 1 min light exposure (Supplementary Figure S10). This stability may enable multistep labeling and fluorescence imaging-assisted analysis when necessary.

The proximity labeling was validated by *in situ* fluorescence imaging. After MG-HI uptake, cells were incubated with PA for 3 min, and then exposed to 660 nm light irradiation (1 min) for proximity labeling. The *in situ* labeling was visualized by clicking with Azide-Fluor 545, and the signal presented by Azide-Fluor 545 overlapped well with mCer3, the fluorescent protein fused to FAP, indicating high spatial resolution of the labeling method (Figure 3D). Furthermore, no labeling signal was detected in the absence of light irradiation, PA or MG-HI probe (Supplementary Figure S11). The proximity labeling ability of MG-HBr and MG-HI was also compared *in cellulo*. Similar to the *in vitro* RNA experiment, MG-HI demonstrated higher proximity labeling efficiency than MG-HBr (Figure 3E). This diminished signal correlates with MG-HBr's lower efficiency in generating singlet oxygen, resulting in a lower reaction yield between biomolecules and PA and a substantially weaker Azide-Fluor 545 fluorescence intensity.

Next, we carefully evaluate the subcellular specificity of FAP–MG-HI-mediated proximity labeling in the two cell lines at the protein level. In-gel fluorescence of proteins isolated from the two cell lines after *in situ* labeling demonstrated sharp signal differences (23,46). Comparing the protein labeling pattern between Nucleus-FAP and ER-FAP, the histone proteins (marked by ***) were clearly more significantly tagged in the Nucleus-FAP cell line (Figure 4A). In each positive protein lane, a significant band was observed slightly above the 50 kDa marker. The sizes of the bands (marked by * and **) match with the exogenous proteins Nuclear-FAP (57.0 kDa) and ER-FAP (57.3 kDa), respectively. Their strong labeling signals fit the proximity labeling mechanism, as these two proteins were closest to the activated MG-HI photosensitizer. Besides, the band in ER-FAP featured higher labeling signals compared to that in Nucleus-FAP, probably due to the higher expression level of ER-FAP protein. To further support our protein assignment hypothesis, a chemical photosensitizer HoeDBF (24) that targets double-stranded DNA in the chromatin region of the nucleus was used as a reference. The clear and strong protein band (marked by *) in the MG-HI-mediated labeling was not observed in the HoeDBF-mediated labeling in the Nucleus-FAP cell line because of the longer distance between FAP and HoeDBF compared to that between FAP and MG-HI, which strongly supports our hypothesis (Supplementary Figure S12).

In addition, cell fractionations were conducted to further support the spatially restricted labeling of proteins (39,49). Total cellular protein (marked as 'Total' fraction) was isolated with cell lysis buffer, while cells were fractionated into 'Cytosol', 'Membrane' and 'Nuclear' fractions with a Mem-PER Plus Membrane Protein Extraction Kit (Thermo). The protein samples from each fraction were conjugated with Azide-Fluor 545 via CuAAC. Protein in-gel fluorescence from the ER-FAP cell line clearly indicated that proteins in the 'Membrane' fraction were substantially labeled, exhibiting a much more intensive fluorescence signal compared to the other fractions (Figure 4B and Supplementary Figure S13). The concentrated signal in the 'Membrane' fraction aligns well with the fact that ER is a membrane-rich organelle and supports the high spatial resolution of FAP–MG-HI-mediated protein proximity labeling. We also evaluated the cell fractionation in the Nuclear-FAP cell line, and no such concentrated signal was observed in the 'Membrane' fraction (Supplementary Figures S14 and S15)

We further conducted RNA dot blot analysis to confirm the labeling at the RNA level in live cells with 1 min of 660 nm light exposure. The concentration titration for MG-HI and MG-HBr was performed from 125 nM to 2 mM with 20 min incubation time before labeling. As shown in Figure 4C and the quantified plot in Supplementary Figure S16, no significant increase in the biotinylation signal was observed with increasing concentration of MG probes. This signal saturation suggests that 125 nM MG probes are sufficient to occupy the accessible FAP binding sites in both cell lines due to their rapid cellular uptake and efficient FAP binding, which may benefit the application in biological systems with low expression of FAP fusion proteins. Consistent with aforementioned *in vitro* data, MG-HI afforded higher RNA labeling signals than MG-HBr. Combining with the cytotoxicity results obtained via MTT assay (Supplementary Figure S17) (43), 125 nM MG-HI was selected for downstream analysis. The RNA biotinylation via CuAAC reaction and *in cellulo* labeling process were further optimized in terms of PA concentration and light irradiation time (Supplementary Figures S18–S22). The optimized condition was applied in the stable cell lines to extensively

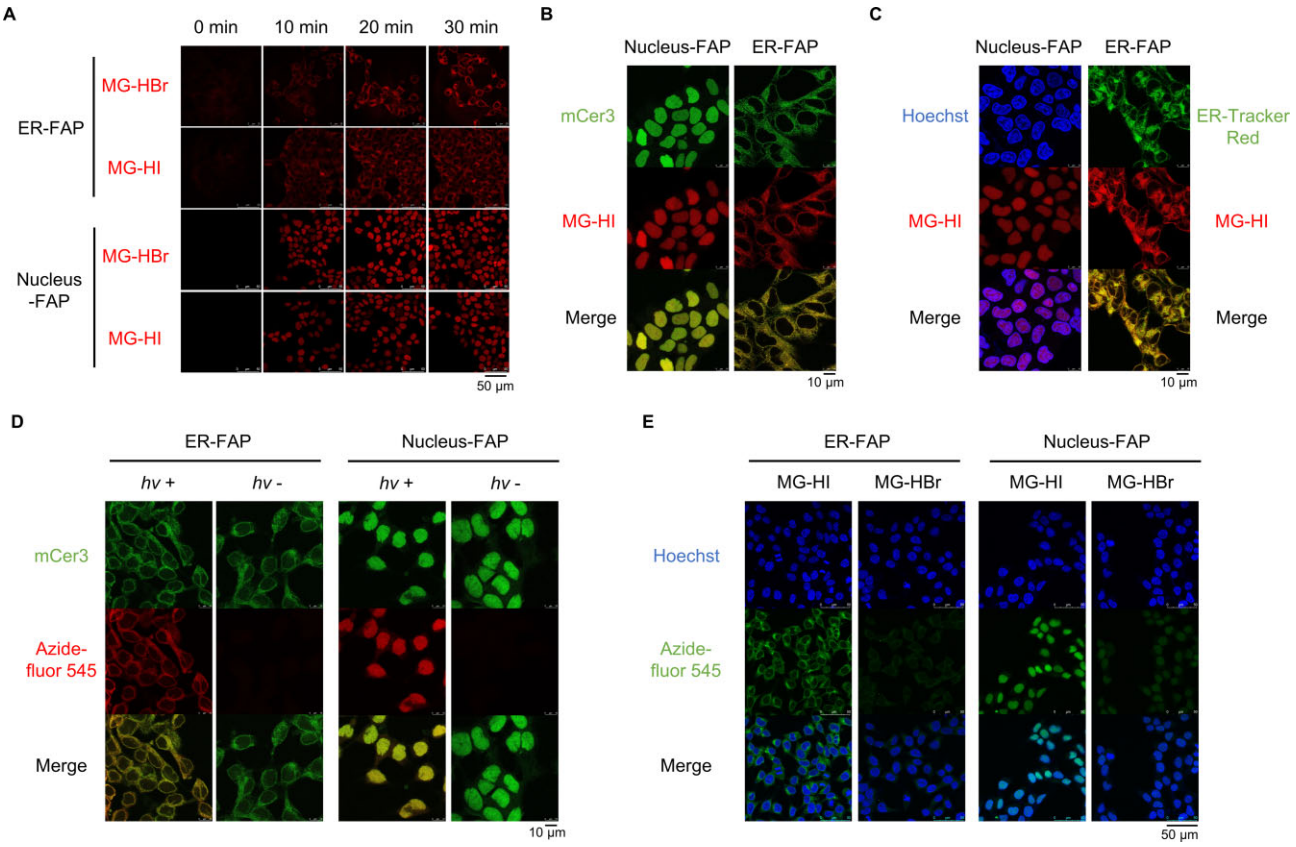


Figure 3. Fluorescence confocal imaging results of FAP–MG-HI. (**A**) Fluorescence turn-on kinetics of 500 nM MG-HBr and MG-HI in live cells expressing ER-FAP and Nucleus-FAP; images were captured before (0 min) and after adding MG probes with a confocal microscope ($\lambda_{ex}/\lambda_{em}$ = 638/650–721 nm). (**B**) The co-localization study on FAP (signal presented by its fusion fluorescence protein mCer3) with MG-HI in live Nucleus-FAP and ER-FAP cell lines (mCer3 channel: $\lambda_{ex}/\lambda_{em}$ = 405/450–550 nm; MG-HI channel: $\lambda_{ex}/\lambda_{em}$ = 638/650–721 nm). (**C**) The co-localization study on MG-HI with nucleus tracker Hoechst in live Nucleus-FAP cell line and ER tracker ER-Tracker Red in ER-FAP cell line (MG-HI channel: $\lambda_{ex}/\lambda_{em}$ = 638/650–721 nm; Hoechst channel: $\lambda_{ex}/\lambda_{em}$ = 405/410–490 nm; ER-Tracker Red channel: $\lambda_{ex}/\lambda_{em}$ = 552/570–631 nm). (**D**) Visualization of *in situ* labeling of live cells with 125 nM MG-HI in live cells with (*hν*+) or without (*hν*−) 660 nm light irradiation (1 min). Azide-Fluor 545 was clicked to alkyne handles of tagged PA after fixing cells with 4% PFA post-labeling (mCer3 channel: $\lambda_{ex}/\lambda_{em}$ = 405/450–550 nm; Azide-Fluro 545 channel: $\lambda_{ex}/\lambda_{em}$ = 552/557–735 nm). (**E**) Comparison of *in situ* labeling ability of 500 nM MG-HI and MG-HBr in live cells (Hoechst channel: $\lambda_{ex}/\lambda_{em}$ = 405/410–490 nm; Azide-Fluro 545 channel: $\lambda_{ex}/\lambda_{em}$ = 552/557–735 nm).

evaluate the negative controls, including eliminating light irradiation, PA and MG-HI one at a time. No observable difference was captured among the three negative conditions, confirming the low background of the light-triggered RNA proximity labeling mediated by FAP–MG-HI (Figure 4D).

We next validated the precision of RNA proximity labeling with RT-qPCR studies. Proximity labeling with and without PA was utilized and further analyzed after RNA being pulled down (Supplementary Figure S23). As shown in Supplementary Figure S24, nucleus RNAs (RN7SK, XIST and MALAT1) and ER-associated RNAs (HSPA5, HSP90B1 and TMX1) were preferentially enriched in the Nucleus-FAP and ER-FAP cell lines, respectively, while cytosolic and mitochondria RNAs (GAPDH, ACTB and MT-CO2) were barely enriched in either cell line. Plotting the relative enrichment fold using RNAs enriched from either Nucleus-FAP or ER-FAP cell line as a reference enabled easy identification of RNAs predominantly enriched in the other cell line (Figure 4E). Specifically, RNAs (GAPDH, ACTB and MT-CO2) that do not localize in these two locations displayed comparable relative fold enrichment scores by using either as a reference. However, RNA enriched in one of these two locations exhibited distinct scores when the ratios of fold enrichment scores were taken.

Transcriptome-wide analysis of FAP–MG-HI-mediated labeling

After confirming the RNA enrichment procedures with RT-qPCR studies, we further conducted FAP-seq to evaluate FAP–MG-HI-mediated RNA proximity labeling at the transcriptomic level. RNA localization was initially quantified by comparing pulldown enrichments in the Nucleus-FAP and ER-FAP cell lines (discussion in Supplementary Figure S25). To distinguish between nucleus- and ER-localized RNAs, we used the relative enrichment of an RNA in the Nucleus-FAP cell line compared to its enrichment in the ER-FAP cell line as a metric, which has been adopted in the RT-qPCR data analysis. As shown in Figure 5A, genes for known nuclear RNAs (RN7SK, SCARNA10, XIST and MALAT1) were more highly enriched with FAP-seq in the Nucleus-FAP cell line. Conversely, genes for known ER-associated RNAs (RN7SL, HSPA5, HSP90B1 and TMX1) were more highly enriched with FAP-seq in the ER-FAP cell line. Other control RNAs, which were expected to be neither particularly nucleus- nor ER-enriched (GAPDH, ACTB and MT-CO2), showed approximately equal enrichments in both cell lines. To quantify subcellular RNA labeling more generally, we calculated relative enrichment fold in Nucleus-FAP to ER-FAP cell lines (Nucleus-FAP/ER-FAP) for all detected long noncoding RNAs (lncRNAs), small

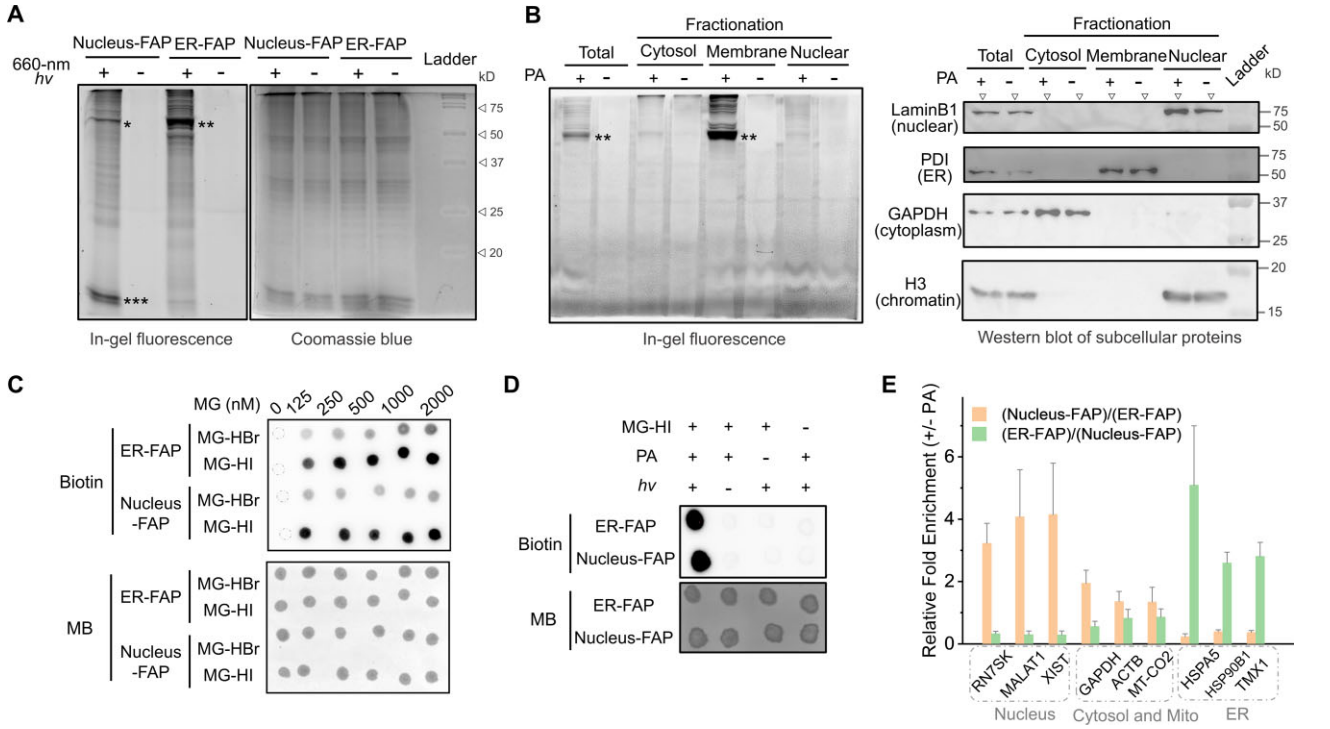


Figure 4. Protein and RNA proximity labeling results. (**A**) In-gel fluorescence and the Coomassie blue staining of proteins to evaluate the live cell protein labeling. Azide-Fluor 545 was clicked to alkyne handles. (**B**) In-gel fluorescence analysis of *in cellulo* protein labeling in the ER-FAP cell line by conjugation with Azide-Fluor 545 via CuAAC and the western blot of subcellular protein markers to validate the fractionation. Protein fractions were separated with Thermo Scientific Mem-PER Plus Membrane Protein Extraction Kit. (**C**) RNA dot blot analysis to evaluate the dosage dependence of live cell RNA labeling efficiency. Azide-biotin was clicked to alkyne handles after RNA extraction. (**D**) Dot blot analysis for *in cellulo* RNA labeling with MG-HI. (**E**) Relative enrichment of RNA in Nucleus-FAP and ER-FAP. '(Nucleus-FAP)/(ER-FAP)' is the enrichment fold of RNA in Nucleus-FAP divided by the enrichment fold of RNA in ER-FAP in each trial and then averaged. '(ER-FAP)/(Nucleus-FAP)' is based on the opposite division.

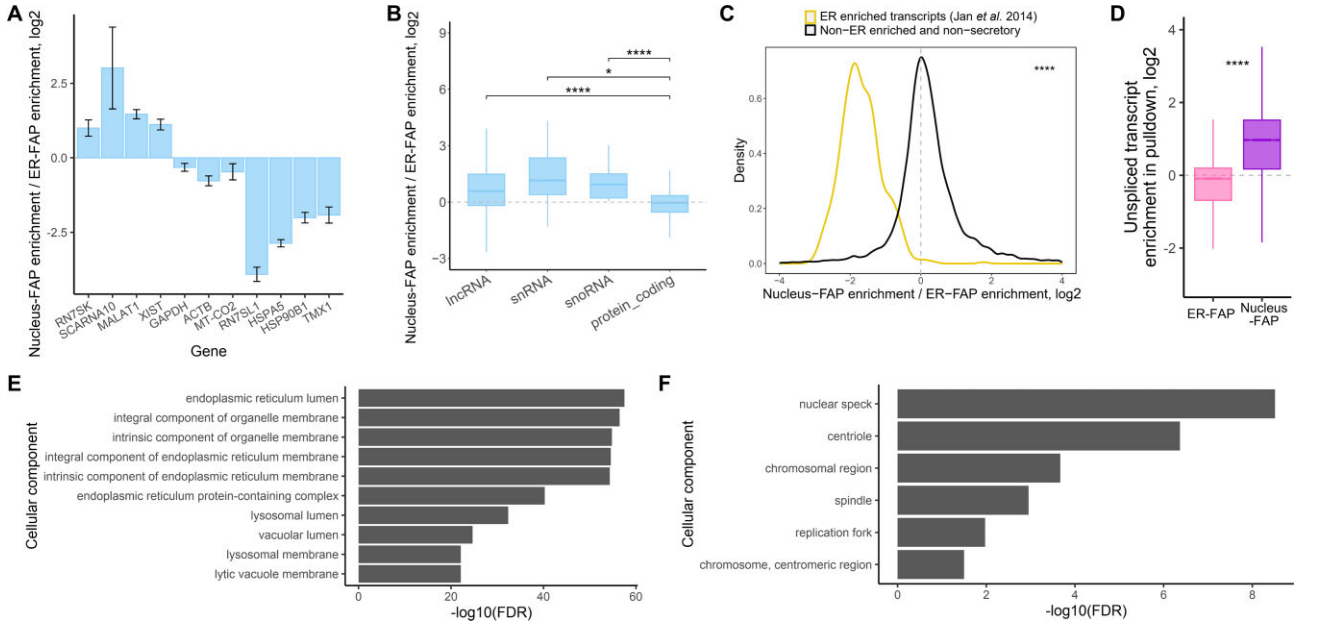


Figure 5. Transcriptome analysis with FAP–MG-HI. (**A**) FAP-seq enrichments for selected RNA species known to be localized to the nucleus or ER. FAP-seq enrichments for (**B**) transcripts of different biotypes (lncRNA, snRNA, snoRNA and protein-coding RNA) and (**C**) ER enriched transcripts, transcripts excluding the ER enriched and secretory transcripts. (**D**) FAP-seq enrichments for unspliced, intron-containing transcripts. (**E**) FAP-seq enriched GO terms derived from RNAs identified as localized to the ER in the ER-FAP cell line. (**F**) FAP-seq enriched GO terms derived from RNAs identified as localized to the nucleus in the Nucleus-FAP cell line. All significance tests were performed using a Wilcoxon rank-sum test: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$.

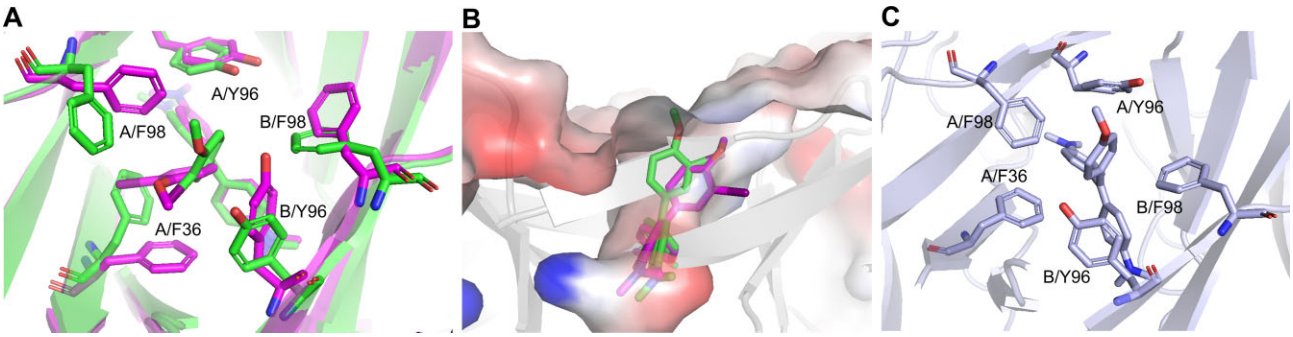


Figure 6. The results of molecular dynamics simulations. (**A**) Representative snapshots of MG-HI bound by FAP in Cfm **1** (green) and Cfm **2** (pink). (**B**) The overall binding position of MG-HI in FAP binding pocket, Cfm **1** (green) and Cfm **2** (pink). (**C**) Representative snapshots of MG-H_2 bound by FAP.

nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and protein-coding RNAs (Figure 5B). Compared to protein-coding RNAs, lncRNAs, snRNAs and snoRNAs were more highly enriched with FAP-seq in the Nucleus-FAP cell line, in line with their known nuclear localization (50,51). Furthermore, we observed a more pronounced enrichment of transcripts previously identified by ER ribosome profiling in the ER-FAP cell line compared to the Nucleus-FAP cell line (Figure 5C) (52). Conversely, we did not observe a similar preferential enrichment for transcripts that excluded the ER-enriched and secretory ones, as predicted by Phobius (53), SignalP 5.0 (54) and TMHMM 2.0 (55,56). In addition to gene types, we further compared the splicing events between the enrichments of the two cell lines. Consistent with the previous report (57), a higher abundance of unspliced transcripts was enriched with FAP-seq in the Nucleus-FAP cell line compared to the ER-FAP cell line (Figure 5D). We then performed gene ontology (GO) analysis of RNAs enriched in the two cell lines. RNAs enriched with FAP-seq in the ER-FAP cell line were associated with ER-related GO terms like ER lumen and integral component of organelle membrane (58,59), while those enriched with FAP-seq in the Nucleus-FAP cell line were associated with nucleus-related GO terms like nuclear speck (60) and chromosomal region (61) (Figure 5E and F). Overall, these results provide a high level of confidence in the ability of FAP-seq to label subcellular populations of RNA.

Molecular dynamics simulations of the FAP–MG-HI system

Lastly, we employed molecular dynamics simulations to further enhance our understanding of the structural arrangement of FAP–MG-HI. Throughout 200 ns of simulations, both the protein and ligand structures were well retained (Supplementary Table S2 and Supplementary Figure S26A–C). The distance between MG-HI and nearby aromatic residues of FAP was monitored during the simulations (Supplementary Figure S26D). Clustering analysis based on the coordinates of MG-HI revealed the presence of two distinct binding modes. In conformation **1** (Cfm **1**, shown in green), MG-HI was ∼5 Å away from F36 on chain A, F98 on chain A and Y96 on chain B. Conversely, in Cfm **2**, MG-HI was pulled closer to F36 on chain A (Supplementary Figure S26D, pink shade). Representative structures from each cluster are depicted in Figure 6A. In Cfm **1** (green), the substituted benzene ring weakly interacted with F98 on chain A and Y96 on chain B, while partially engaging with F36 on chain A through an edge-to-face (T-shaped) π-interaction (62). In Cfm **2** (pink), the substituted benzene ring was more effectively sandwiched between F36 and F98 on chain A, and it also exhibited partial interaction with Y96 on chain B via an edge-to-face π-interaction (63). As the ligand approached the deeply situated F36, it became immersed in a hydrophobic pocket (Figure 6B). Moreover, the MMGBSA (64) binding energies of Cfm **2** (−55.82 kcal/mol) were lower than those of Cfm **1** (−53.03 kcal/mol). However, the complex adopted Cfm **2** for only 25 ns before reverting to Cfm **1** during the 200 ns simulations. Notably, this unique conformational switch was not observed with MG-H_2 (Figure 6C), as it adopts a symmetric distribution within FAP without any π/π or π/σ stacking interactions (32). Consequently, we speculate that MG-HI, being asymmetric with a large monoiodo substitution, can accommodate a previously unexplored binding mode with FAP, thereby facilitating the activation of its photosensitizing capabilities.

Discussion

In this study, we have comprehensively demonstrated the high efficiency of FAP–MG-HI across a spectrum of scenarios, ranging from *in vitro* to *in cellulo* conditions, and from experimental setups to computational simulations. Inspired by our previous spectroscopic findings and theoretical calculations (33), we designed MG-HI and subjected it to extensive comparative analyses with MG-HBr. These analyses included assessments of $^1O_2$ generation via ADPA assays, *in vitro* RNA dot blot assays, fluorescence confocal imaging, *in cellulo* RNA dot blot analysis and cytotoxicity evaluations using MTT assays. MG-HI was chosen for downstream evaluation due to its higher $^1O_2$ yield. Both western blot and RT-qPCR analyses validated the high spatial resolution of FAP–MG-HI-mediated proximity labeling. We further performed FAP-seq to demonstrate the compatibility of FAP–MG-HI with subcellular transcriptomic profiling. The high $^1O_2$ generation efficiency of FAP–MG-HI is crucial for achieving effective RNA proximity labeling with FAP-seq in a narrow time frame, which we have successfully demonstrated with a labeling duration as short as 1 min. Additionally, we conducted molecular dynamics simulations to shine light on the potential distinct structural arrangement of FAP–MG-HI, revealing previously underexplored possibilities of introducing symmetry-breaking MG derivatives. This symmetry-breaking concept forges new avenues for developing specialized FAP–MG systems with unconventional functionalities that could uncover new insights in biological systems.

Our study has significantly enriched the utility of FAP technology by incorporating proximity labeling capabilities. The previously established FAP–MG platform focused primarily on fluorescence-based protein tracking (30,65,66) and reactive oxygen species-mediated protein inactivation (31,67,68). These two functions require different MG derivatives and cannot be achieved simultaneously. Here, we introduced an innovative dimension to FAP–MG technology: biomolecular profiling via proximity labeling. We established a new method, FAP-seq, which utilizes our newly developed symmetry-breaking FAP–MG-HI to achieve RNA proximity labeling predominantly through the $^1O_2$ pathway upon NIR light irradiation. In this aspect, MG-HI has demonstrated superior labeling efficiency over all other MG derivatives tested. Moreover, FAP–MG-HI preserves a robust fluorescence intensity, which enables a 'one-stone-three-birds' strategy that concurrently allows for protein tracking, targeted protein ablation and biomolecular profiling of cellular microenvironments. Using FAP-seq, we successfully profiled nuclear and ER transcriptomes, showcasing the potential of this methodology.

FAP-seq stands out as a versatile tool applicable in various biological systems, offering unique advantages over other RNA proximity labeling methods. By tailoring FAP expression to specific subcellular locations, we can correspondingly achieve RNA proximity labeling across various cellular compartments. Our research presents a compelling proof of concept using cell lines stably expressing ER-FAP and Nucleus-FAP. A distinctive feature of the FAP-seq is its capacity to generate $^1O_2$ upon NIR light irradiation, which is detectable only when MG-HI is bound to FAP due to restricted rotation. This characteristic precludes the necessity for pre-labeling washes (21,24) or post-labeling quenching steps (18), thus providing FAP-seq with high spatiotemporal resolution. Compared to APEX-seq, FAP-seq eliminates the usage of high concentrations of hydrogen peroxide and subsequent quenching steps in live cells (18). Moreover, FAP-seq surpasses other light-triggered proximity labeling tools by providing a unique fluorogenic turn-on capability, a feature not commonly found in conventional synthetic photosensitizers (21,22,28,69). The NIR light activation of FAP-seq distinguishes it from flavin-dependent genetically encoded photosensitizers (19,20), enabling deeper tissue penetration (28).

In addition to transcriptomic profiling, we envision widespread applications of FAP–MG-HI-mediated proximity labeling in complex biological systems. The NIR light control of FAP–MG-HI allows for orthogonal activation in conjunction with other proximity labeling techniques, such as APEX (18) and miniSOG (19). This combinatorial approach provides exciting opportunities for integrated applications in studying RNA localization (70). Furthermore, the reactivity of $^1O_2$ with guanosine residues in RNA (19,21) and electron-rich amino acids in proteins (25–27) enables simultaneous mapping of both RNA and protein distributions via FAP–MG-HI, enriching our understanding of RNA and protein interplay within specific cellular domains. This dual reactivity also sets the stage for innovative cross-linking chemistries, potentially revolutionizing the exploration of RNA–protein interactions and expanding the utility of FAP–MG-HI in cellular dynamics research. With attributes such as fast turn-on response, NIR light activation, short irradiation time required, and compatibility with both protein and RNA proximity labeling, FAP–MG-HI holds great potential for diverse applications in complex biological systems (71,72).

Data availability

All data other than Illumina-type sequencing data are incorporated into the article and its online supplementary material. The sequencing data information is listed at https://www.ncbi.nlm.nih.gov/sra/PRJNA1033228.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

RNA sequencing data analysis was conducted on HKU High Performance Computing facilities. The authors appreciate Prof. Dr Kou Okuro (Department of Chemistry, The University of Hong Kong) for sharing with us the plate reader.

Funding

Research Grants Council, University Grants Committee [207301212 and 207301251 to E.C.M.T., 27302220 and 17318422]; National Natural Science Foundation of China [22107088]; Laboratory for Synthetic Chemistry and Chemical Biology under the Health@InnoHK Program launched by the Innovation and Technology Commission—Hong Kong; National Institutes of Health [R35GM13385]. Funding for open access charge: Research Grants Council, University Grants Committee [17318422].

Conflict of interest statement

None declared.

References

1. Buxbaum,A.R., Haimovich,G. and Singer,R.H. (2015) In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.*, **16**, 95–109.
2. Chin,A. and Lécuyer,E. (2017) RNA localization: making its way to the center stage. *Biochim. Biophys. Acta Gen. Subj.*, **1861**, 2956–2970.
3. Carlevaro-Fita,J. and Johnson,R. (2019) Global positioning system: understanding long noncoding RNAs through subcellular localization. *Mol. Cell*, **73**, 869–883.
4. Hachet,O. and Ephrussi,A. (2004) Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature*, **428**, 959–963.
5. Wutz,A., Rasmussen,T.P. and Jaenisch,R. (2002) Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.*, **30**, 167–174.
6. Madugalle,S.U., Meyer,K., Wang,D.O. and Bredy,T.W. (2020) RNA N^6-methyladenosine and the regulation of RNA localization and function in the brain. *Trends Neurosci.*, **43**, 1011–1023.
7. Wang,E.T., Cody,N.A., Jog,S., Biancolella,M., Wang,T.T., Treacy,D.J., Luo,S., Schroth,G.P., Housman,D.E. and Reddy,S. (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell*, **150**, 710–724.
8. Katahira,J. (2015) Nuclear export of messenger RNA. *Genes*, **6**, 163–184.
9. Lipshitz,H.D. and Smibert,C.A. (2000) Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genet. Dev.*, **10**, 476–488.
10. Statello,L., Guo,C.-J., Chen,L.-L. and Huarte,M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, **22**, 96–118.

11. Thakur,J. and Henikoff,S. (2020) Architectural RNA in chromatin organization. *Biochem. Soc. Trans.*, **48**, 1967–1978.
12. Martin,K.C. and Ephrussi,A. (2009) mRNA localization: gene expression in the spatial dimension. *Cell*, **136**, 719–730.
13. Holt,C.E., Martin,K.C. and Schuman,E.M. (2019) Local translation in neurons: visualization and function. *Nat. Struct. Mol. Biol.*, **26**, 557–566.
14. Winkler,L. and Dimitrova,N. (2022) A mechanistic view of long noncoding RNAs in cancer. *Wiley Interdiscip. Rev. RNA*, **13**, e1699.
15. Wang,X., He,Y., Mackowiak,B. and Gao,B. (2021) MicroRNAs as regulators, biomarkers and therapeutic targets in liver diseases. *Gut*, **70**, 784–795.
16. Liu,E.Y., Cali,C.P. and Lee,E.B. (2017) RNA metabolism in neurodegenerative disease. *Dis. Models Mech.*, **10**, 509–518.
17. Li,J., Sun,D., Pu,W., Wang,J. and Peng,Y. (2020) Circular RNAs in cancer: biogenesis, function, and clinical significance. *Trends Cancer*, **6**, 319–336.
18. Fazal,F.M., Han,S., Parker,K.R., Kaewsapsak,P., Xu,J., Boettiger,A.N., Chang,H.Y. and Ting,A.Y. (2019) Atlas of subcellular RNA localization revealed by APEX-seq. *Cell*, **178**, 473–490.
19. Wang,P., Tang,W., Li,Z., Zou,Z., Zhou,Y., Li,R., Xiong,T., Wang,J. and Zou,P. (2019) Mapping spatial transcriptome with light-activated proximity-dependent RNA labeling. *Nat. Chem. Biol.*, **15**, 1110–1119.
20. Hananya,N., Ye,X., Koren,S. and Muir,T.W. (2023) A genetically encoded photoproximity labeling approach for mapping protein territories. *Proc. Natl Acad. Sci. U.S.A.*, **120**, e2219339120.
21. Li,Y., Aggarwal,M.B., Nguyen,K., Ke,K. and Spitale,R.C. (2017) Assaying RNA localization *in situ* with spatially restricted nucleobase oxidation. *ACS Chem. Biol.*, **12**, 2709–2714.
22. Engel,K.L., Lo,H.-Y.G., Goering,R., Li,Y., Spitale,R.C. and Taliaferro,J.M. (2022) Analysis of subcellular transcriptomes by RNA proximity labeling with Halo-seq. *Nucleic Acids Res.*, **50**, e24.
23. Li,Y., Aggarwal,M.B., Ke,K., Nguyen,K. and Spitale,R.C. (2018) Improved analysis of RNA localization by spatially restricted oxidation of RNA–protein complexes. *Biochemistry*, **57**, 1577–1581.
24. Li,L., Liang,J., Luo,H., Tam,K.M., Edmund,C. and Li,Y. (2019) A new chemical approach for proximity labelling of chromatin-associated RNAs and proteins with visible light irradiation. *Chem. Commun.*, **55**, 12340–12343.
25. Zheng,F., Yu,C., Zhou,X. and Zou,P. (2023) Genetically encoded photocatalytic protein labeling enables spatially-resolved profiling of intracellular proteome. *Nat. Commun.*, **14**, 2978.
26. Zhai,Y., Huang,X., Zhang,K., Huang,Y., Jiang,Y., Cui,J., Zhang,Z., Chiu,C.K., Zhong,W. and Li,G. (2022) Spatiotemporal-resolved protein networks profiling with photoactivation dependent proximity labeling. *Nat. Commun.*, **13**, 4906.
27. Nakane,K., Sato,S., Niwa,T., Tsushima,M., Tomoshige,S., Taguchi,H., Ishikawa,M. and Nakamura,H. (2021) Proximity histidine labeling by umpolung strategy using singlet oxygen. *J. Am. Chem. Soc.*, **143**, 7726–7731.
28. Buksh,B.F., Knutson,S.D., Oakley,J.V., Bissonnette,N.B., Oblinsky,D.G., Schwoerer,M.P., Seath,C.P., Geri,J.B., Rodriguez-Rivera,F.P., Parker,D.L., *et al.* (2022) μMap-Red: proximity labeling by red light photocatalysis. *J. Am. Chem. Soc.*, **144**, 6154–6162.
29. Szent-Gyorgyi,C., Schmidt,B.F., Creeger,Y., Fisher,G.W., Zakel,K.L., Adler,S., Fitzpatrick,J.A., Woolford,C.A., Yan,Q. and Vasilev,K.V. (2008) Fluorogen-activating single-chain antibodies for imaging cell surface proteins. *Nat. Biotechnol.*, **26**, 235–240.
30. Telmer,C.A., Verma,R., Teng,H., Andreko,S., Law,L. and Bruchez,M.P. (2015) Rapid, specific, no-wash, far-red fluorogen activation in subcellular compartments by targeted fluorogen activating proteins. *ACS Chem. Biol.*, **10**, 1239–1246.
31. He,J., Wang,Y., Missinato,M.A., Onuoha,E., Perkins,L.A., Watkins,S.C., St Croix,C.M., Tsang,M. and Bruchez,M.P. (2016) A genetically targetable near-infrared photosensitizer. *Nat. Methods*, **13**, 263–268.
32. Szent-Gyorgyi,C., Stanfield,R.L., Andreko,S., Dempsey,A., Ahmed,M., Capek,S., Waggoner,A.S., Wilson,I.A. and Bruchez,M.P. (2013) Malachite green mediates homodimerization of antibody VL domains to form a fluorescent ternary complex with singular symmetric interfaces. *J. Mol. Biol.*, **425**, 4595–4613.
33. Deng,Z., Li,L., Jia,H., Li,N.F., He,J., Li,M.D., Phillips,D.L. and Li,Y. (2023) Insights into the photodynamics of fluorescence emission and singlet oxygen generation of fluorogen activating protein–malachite green systems. *Chem. Eur. J.*, **29**, e202203684.
34. Wang,L., Tran,M., D'Este,E., Roberti,J., Koch,B., Xue,L. and Johnsson,K. (2020) A general strategy to develop cell permeable and fluorogenic probes for multicolour nanoscopy. *Nat. Chem.*, **12**, 165–172.
35. Zheng,Q., Ayala,A.X., Chung,I., Weigel,A.V., Ranjan,A., Falco,N., Grimm,J.B., Tkachuk,A.N., Wu,C. and Lippincott-Schwartz,J. (2019) Rational design of fluorogenic and spontaneously blinking labels for super-resolution imaging. *ACS Cent. Sci.*, **5**, 1602–1613.
36. Magoč,T. and Salzberg,S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.
37. Stancil,I.T., Michalski,J.E., Davis-Hall,D., Chu,H.W., Park,J.-A., Magin,C.M., Yang,I.V., Smith,B.J., Dobrinskikh,E. and Schwartz,D.A. (2021) Pulmonary fibrosis distal airway epithelia are dynamically and structurally dysfunctional. *Nat. Commun.*, **12**, 4566.
38. Ross-Innes,C.S., Stark,R., Teschendorff,A.E., Holmes,K.A., Ali,H.R., Dunning,M.J., Brown,G.D., Gojis,O., Ellis,I.O. and Green,A.R. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389–393.
39. Gagnon,K.T., Li,L., Janowski,B.A. and Corey,D.R. (2014) Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.*, **9**, 2045–2060.
40. Yong,L., Zhanqi,G., Yuefei,J., Xiaobin,H., Cheng,S., Shaogui,Y., Lianhong,W., Qingeng,W. and Die,F. (2015) Photodegradation of malachite green under simulated and natural irradiation: kinetics, products, and pathways. *J. Hazard. Mater.*, **285**, 127–136.
41. Gandin,E., Lion,Y. and Van de Vorst,A. (1983) Quantum yield of singlet oxygen production by xanthene derivatives. *Photochem. Photobiol.*, **37**, 271–278.
42. Baptista,M.S., Cadet,J., Greer,A. and Thomas,A.H. (2021) Photosensitization reactions of biomolecules: definition, targets and mechanisms. *Photochem. Photobiol.*, **97**, 1456–1483.
43. Li,M., Xia,J., Tian,R., Wang,J., Fan,J., Du,J., Long,S., Song,X., Foley,J.W. and Peng,X. (2018) Near-infrared light-initiated molecular superoxide radical generator: rejuvenating photodynamic therapy against hypoxic tumors. *J. Am. Chem. Soc.*, **140**, 14851–14859.
44. Dichmann,L., Bregnhøj,M., Liu,H., Westberg,M., Poulsen,T.B., Etzerodt,M. and Ogilby,P.R. (2021) Photophysics of a protein-bound derivative of malachite green that sensitizes the production of singlet oxygen. *Photochem. Photobiol. Sci.*, **20**, 435–449.
45. Ehrenberg,B., Anderson,J.L. and Foote,C.S. (1998) Kinetics and yield of singlet oxygen photosensitized by hypericin in organic and biological media. *Photochem. Photobiol.*, **68**, 135–140.
46. Li,Y., Tian,C., Liu,K., Zhou,Y., Yang,J. and Zou,P. (2020) A clickable APEX probe for proximity-dependent proteomic profiling in yeast. *Cell Chem. Biol.*, **27**, 858–865.
47. Westberg,M., Holmegaard,L., Pimenta,F.M., Etzerodt,M. and Ogilby,P.R. (2015) Rational design of an efficient, genetically encodable, protein-encased singlet oxygen photosensitizer. *J. Am. Chem. Soc.*, **137**, 1632–1642.
48. Hope,T., Reyes-Robles,T., Ryu,K.A., Removski,N., Maisonneuve,J., Oslund,R., Fadeyi,O. and Frenette,M. (2022) Mechanistic evidence

for a radical–radical recombination pathway of flavin-based photocatalytic tyrosine labeling. ChemRxiv doi: http://doi.org/10.26434/chemrxiv-2022-w8xz2, 05 August 2022, preprint: not peer reviewed.
49. Qoronfleh,M.W., Benton,B., Ignacio,R. and Kaboord,B. (2003) Selective enrichment of membrane proteins by partition phase separation for proteomic studies. *J. Biomed. Biotechnol.*, **2003**, 249–255.
50. Kopp,F. and Mendell,J.T. (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell*, **172**, 393–407.
51. Deogharia,M. and Gurha,P. (2022) The "guiding" principles of noncoding RNA function. *Wiley Interdiscip. Rev. RNA*, **13**, e1704.
52. Jan,C.H., Williams,C.C. and Weissman,J.S. (2014) Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, **346**, 1257521.
53. Käll,L., Krogh,A. and Sonnhammer,E.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.*, **338**, 1027–1036.
54. Almagro Armenteros,J.J., Tsirigos,K.D., Sønderby,C.K., Petersen,T.N., Winther,O., Brunak,S., von Heijne,G. and Nielsen,H. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.*, **37**, 420–423.
55. Sonnhammer,E.L., Von Heijne,G. and Krogh,A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: *Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology*. Montreal, Quebec,Canada, pp. 175–182.
56. Krogh,A., Larsson,B., Von Heijne,G. and Sonnhammer,E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
57. Maquat,L.E. and Carmichael,G.G. (2001) Quality control of mRNA function. *Cell*, **104**, 173–176.
58. Pyhtila,B., Zheng,T., Lager,P.J., Keene,J.D., Reedy,M.C. and Nicchitta,C.V. (2008) Signal sequence- and translation-independent mRNA localization to the endoplasmic reticulum. *RNA*, **14**, 445–453.
59. Cui,X.A. and Palazzo,A.F. (2014) Localization of mRNAs to the endoplasmic reticulum. *Wiley Interdiscip. Rev. RNA*, **5**, 481–492.
60. Bardou,F., Ariel,F., Simpson,C.G., Romero-Barrios,N., Laporte,P., Balzergue,S., Brown,J.W. and Crespi,M. (2014) Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Dev. Cell*, **30**, 166–176.
61. Xiao,R., Chen,J.-Y., Liang,Z., Luo,D., Chen,G., Lu,Z.J., Chen,Y., Zhou,B., Li,H. and Du,X. (2019) Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell*, **178**, 107–121.
62. Wilson,K.A., Kellie,J.L. and Wetmore,S.D. (2014) DNA–protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Res.*, **42**, 6726–6741.
63. Pisani,P., Caporuscio,F., Carlino,L. and Rastelli,G. (2016) Molecular dynamics simulations and classical multidimensional scaling unveil new metastable states in the conformational landscape of CDK2. *PLoS One*, **11**, e0154066.
64. Chen,F., Liu,H., Sun,H., Pan,P., Li,Y., Li,D. and Hou,T. (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein–protein binding free energies and re-rank binding poses generated by protein–protein docking. *Phys. Chem. Chem. Phys.*, **18**, 22129–22139.
65. Bulgari,D., Deitcher,D.L., Schmidt,B.F., Carpenter,M.A., Szent-Gyorgyi,C., Bruchez,M.P. and Levitan,E.S. (2019) Activity-evoked and spontaneous opening of synaptic fusion pores. *Proc. Natl Acad. Sci. U.S.A.*, **116**, 17039–17044.
66. Carpenter,M.A., Wang,Y., Telmer,C.A., Schmidt,B.F., Yang,Z.P. and Bruchez,M.P. (2020) Protein proximity observed using fluorogen activating protein and dye activated by proximal anchoring (FAP–DAPA) system. *ACS Chem. Biol.*, **15**, 2433–2443.
67. Fouquerel,E., Barnes,R.P., Uttam,S., Watkins,S.C., Bruchez,M.P. and Opresko,P.L. (2019) Targeted and persistent 8-oxoguanine base damage at telomeres promotes telomere loss and crisis. *Mol. Cell*, **75**, 117–130.
68. Barnes,R.P., de Rosa,M., Thosar,S.A., Detwiler,A.C., Roginskaya,V., Van Houten,B., Bruchez,M.P., Stewart-Ornstein,J. and Opresko,P.L. (2022) Telomeric 8-oxo-guanine drives rapid premature senescence in the absence of telomere shortening. *Nat. Struct. Mol. Biol.*, **29**, 639–652.
69. Geri,J.B., Oakley,J.V., Reyes-Robles,T., Wang,T., McCarver,S.J., White,C.H., Rodriguez-Rivera,F.P., Parker,D.L., Hett,E.C., Fadeyi,O.O., *et al.* (2020) Microenvironment mapping via Dexter energy transfer on immune cells. *Science*, **367**, 1091–1097.
70. Liang,J., Han,J., Gao,X., Jia,H., Li,R., Tse,E.C.M. and Li,Y. (2024) Clickable APEX2 probes for enhanced RNA proximity labeling in live cells. *Anal. Chem.*, **96**, 685–693.
71. Dutertre,M. and Vagner,S. (2017) DNA-damage response RNA-binding proteins (DDRBPs): perspectives from a new class of proteins and their RNA targets. *J. Mol. Biol.*, **429**, 3139–3145.
72. Anderson,P. and Kedersha,N. (2008) Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.*, **33**, 141–150.

Received: November 2, 2023. **Revised:** January 17, 2024. **Editorial Decision:** February 6, 2024. **Accepted:** February 8, 2024