BIOCHEMISTRY

Recording physiological history of cells with chemical labeling

Magnus-Carsten Huppertz[1]†, Jonas Wilhelm[1]†, Vincent Grenier[1], Martin W. Schneider[2], Tjalda Falt[3], Nicola Porzberg[1], David Hausmann[4], Dirk C. Hoffmann[4,5,6], Ling Hai[4,5,7], Miroslaw Tarnawski[8], Gabriela Pino[9], Krasimir Slanchev[2], Ilya Kolb[10], Claudio Acuna[9], Lisa M. Fenk[3], Herwig Baier[2], Julien Hiblot[1]*, Kai Johnsson[1,11]*

Recordings of the physiological history of cells provide insights into biological processes, yet obtaining such recordings is a challenge. To address this, we introduce a method to record transient cellular events for later analysis. We designed proteins that become labeled in the presence of both a specific cellular activity and a fluorescent substrate. The recording period is set by the presence of the substrate, whereas the cellular activity controls the degree of the labeling. The use of distinguishable substrates enabled the recording of successive periods of activity. We recorded protein-protein interactions, G protein–coupled receptor activation, and increases in intracellular calcium. Recordings of elevated calcium levels allowed selections of cells from heterogeneous populations for transcriptomic analysis and tracking of neuronal activities in flies and zebrafish.

The recording of transient cellular events is crucial for the delineation of biological processes. For example, recording changes in intracellular Ca^{2+} concentrations, neurotransmitter concentrations, or ion channel activities enables the tracking of neuronal activity. Ideally, methods to record transient cellular events would access large populations of cells in parallel, across entire tissues, and with high spatiotemporal resolution. Furthermore, transforming transient events into permanent marks for a later analysis would separate the recording from its analysis, which is particularly important for a massive parallel analysis of transient cellular events. However, there is a shortage of methods that fulfill these criteria. At present, most tissue-wide recordings are performed by coupling a given event to the transcription of a reporter gene (*1*–*3*). A widespread example in neuroscience is the use of promotors of immediate early genes (IEGs) to report on neuron activity in vivo by the transcription of a reporter gene and later analysis (*4*, *5*). Using reporter genes that form continuously growing protein filaments as readout also allows continuous recording of physiological activities (*6*, *7*). However, the transcriptional coupling between neuron activity and reporter expression is only indirect (*8*). CRISPR-Cas–based tools can be used to record cellular events on DNA (*9*, *10*), but these approaches are also based on an indirect coupling of signal and readout and have a relatively poor temporal resolution. Optical and optogenetic approaches, which control recording through illumination of live specimens, can directly record activities for in situ and post hoc analyses and offer outstanding spatiotemporal resolution (*11*–*16*). For example, the biosensor CaMPARI undergoes an irreversible color change when illuminated in the presence of high Ca^{2+} levels to record neuronal activity for later analysis and has been successfully used in various model organisms (*11*, *12*). Activity-based labeling of cells with chemical probes could complement the aforementioned approaches because it can directly record the activity of interest, eliminates the need for illumination, and offers a recording window that is determined by addition and washout of an ideally highly permeable and innocuous probe. Yet present approaches for activity-based chemical labeling are not suitable for continuous recordings nor do they allow the recording of multiple successive events (*17*).

We developed split-HaloTag–based protein recorders that are irreversibly labeled only in the presence of both a specific biological activity and an externally applied fluorescent substrate. The sequential use of fluorescent substrates with different colors allows the recording of multiple periods of biological activities for later analysis. The approac… scalable, can be applied to record diffe… biological activities, and is applicable in vivo.

Results

As a starting point for the generation of a recorder of transient cellular events, we used the self-labeling protein HaloTag (*18*). HaloTag can be specifically labeled with fluorescent substrates both in vitro and in vivo using chloroalkane (CA) substrates (*19*–*21*). We reasoned that a recorder could be created by making its self-labeling activity dependent on a transient cellular activity, such that self-labeling only occurred in the presence of the activity and a fluorescent substrate. To design such recorders, we developed a split protein sensor in which functional HaloTag is reconstituted from a folded yet inactive fragment that could be reversibly activated through the binding of a small peptide (Fig. 1A). We circularly permuted HaloTag by connecting the original termini and screened for active variants with new termini close to the substrate binding site (Fig. 1B and fig. S1). We identified circularly permuted HaloTag variants with new C and N termini at positions 141 and 145, respectively, and at positions 154 and 156 that maintained the overall fold and activity of the protein (Fig. 1C and fig. S2). Residues 142 to 155 were truncated to generate cpHaloΔ, which maintains the overall fold of HaloTag but had almost no activity (Fig. 1D and fig. S2). However, activity could be restored through the reversible binding of the decapeptide Hpep1 (residues 145 to 154) (Fig. 1, E and F, and fig. S2). We then tested this split-HaloTag system for the detection of the rapamycin-dependent interaction of FK506-binding protein (FKBP) and the rapamycin-binding domain of mechanistic target of rapamycin (mTOR) (FRB; Fig. 1G) (*22*). When the split-HaloTag fragments were fused to the C termini of FKBP and FRB, split-HaloTag labeling was strictly dependent on the presence of rapamycin (Fig. 1H). When fused to the more distant N termini of FKBP and FRB (47 Å between the N termini versus 14 Å between the C termini), no labeling was detectable (Fig. 1I). We assumed that this was due to the low affinity of Hpep1 for cpHaloΔ [dissociation constant (K_d) = 4.6 mM; fig. S2], which did not allow complementation across larger distances. We developed peptides with higher affinities for cpHaloΔ by setting up a custom computational RosettaScripts protocol. Structures (40,000) with different peptide sequences were generated and scored by Rosetta total score and peptide-binding free energy. Of these, 384 selected peptides were synthesized and assayed for their potency to complement purified cpHaloΔ (table S1). Of those peptides, 80% showed faster labeling of cpHaloΔ than Hpep1 (Fig. 1J). We combined features of highly active peptides, which

[1]Department of Chemical Biology, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany. [2]Department Genes – Circuits – Behavior, Max Planck Institute for Biological Intelligence, Am Klopferspitz 18, 82152 Martinsried, Germany. [3]Active Sensing, Max Planck Institute for Biological Intelligence, Am Klopferspitz 18, 82152 Martinsried, Germany. [4]Clinical Cooperation Unit Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [5]Department of Neurology and Neurooncology Program, National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany. [6]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. [7]Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [8]Protein Expression and Characterization Facility, Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany. [9]Chica and Heinz Schaller Foundation, Institute of Anatomy and Cell Biology, Heidelberg University, Im Neuenheimer Feld 307, 69120 Heidelberg, Germany. [10]GENIE Project Team, Janelia Research Campus, Ashburn, VA 20147, USA. [11]Institute of Chemical Sciences and Engineering (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.
*Corresponding author. Email: johnsson@mr.mpg.de (K.J.); julien.hiblot@mr.mpg.de (J.H.)
†These authors contributed equally to this work.

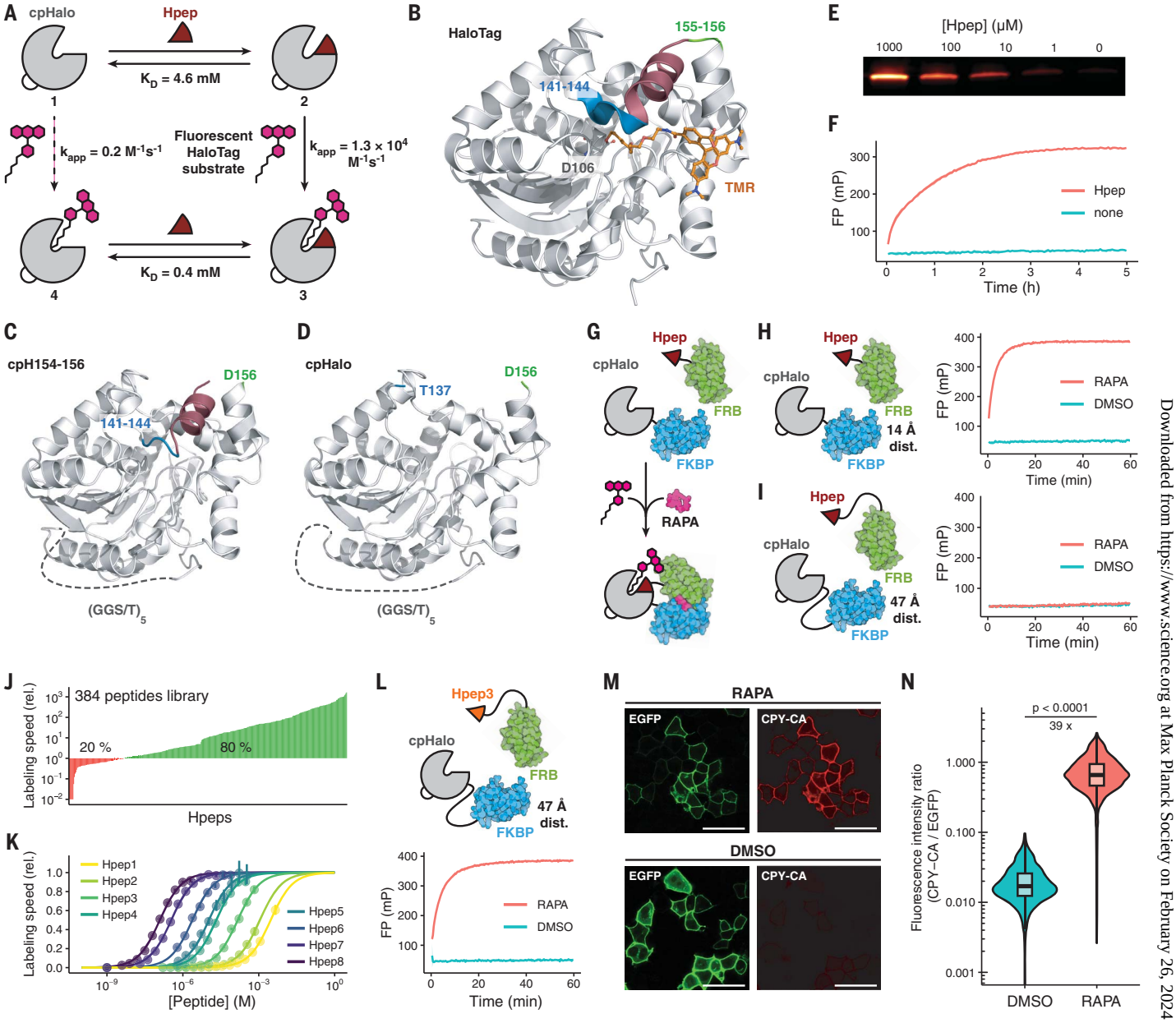

Fig. 1. Split-HaloTag characterization and application as a recorder of protein-protein interactions. (**A**) Scheme of the split-HaloTag system. Binding of Hpep to cpHalo increases labeling speed 26,000-fold. Numbers correspond to reaction with the fluorescent tetramethylrhodamine HaloTag substrate (TMR-CA). k_{app}, apparent second-order rate constant. (**B**) X-ray structure of HaloTag labeled with TMR-CA [Protein Data Bank (PDB) ID 6Y7A]. Positions of the new termini of the circularly permuted HaloTag are highlighted in blue and green. The segment corresponding to Hpep1 is highlighted in red. D, Asp. (**C** and **D**) X-ray structures of cpHaloTag154-156 (cpH154-156; PDB ID 8B6P) and cpHaloΔ (PDB ID 8B6N). Both proteins display the same overall fold as HaloTag. T, Thr. (**E**) Hpep-dependency of cpHaloΔ (1 μM) labeling with TMR-CA (2 μM) analyzed by SDS-PAGE (SDS–polyacrylamide gel electrophoresis). (**F**) Labeling kinetics of split-HaloTag (500 nM, cpHaloΔ) with TMR-CA (50 nM) in presence or absence of Hpep (250 μM) measured by fluorescence polarization (FP). (**G**) Detection of the rapamycin-dependent interaction of FKBP and FRB with split-HaloTag. (**H**) Labeling kinetics of FKBP-(GGS)-cpHaloΔ (250 nM) and FRB-(GGS)-Hpep1 (250 nM) (14-Å distance between FKBP and FRB C termini) with TMR-CA (50 nM) as measured by FP. Split-HaloTag labels itself only in the presence of rapamycin (RAPA; 500 nM). DMSO, dimethyl sulfoxide. (**I**) Labeling kinetics of cpHaloΔ-$(GGS)_9$-FKBP (250 nM) and Hpep1-$(GGS)_3$-FRB (250 nM) (47-Å distance between FKBP and FRB N termini) with TMR-CA (50 nM) in the presence or absence of rapamycin (500 nM) as measured by FP. (**J**) Screening of 384 Rosetta-designed Hpep variants reveals a large number of highly active hits. Labeling rates with TMR-CA (100 nM) in the presence of cpHaloΔ (500 nM) are given relative to the labeling rate of the original Hpep1. (**K**) Titrations of eight selected Hpeps with EC_{50} values ranging from high nanomolar (EC_{50}^{Hpep8}: 124 nM) to low millimolar (EC_{50}^{Hpep1}: 3.0 mM). (**L**) Labeling kinetics of cpHaloΔ-$(GGS)_9$-FKBP (250 nM) and Hpep3-$(GGS)_3$-FRB (250 nM) (47 Å distance between FKBP and FRB N termini) with TMR-CA (50 nM) in the presence or absence of rapamycin (500 nM) as measured by FP. (**M**) Fluorescence micrographs of HeLa cells coexpressing Lyn11-EGFP-cpHaloΔ-$(GGS)_9$-FKBP and Hpep3-$(GGS)_3$-FRB-mScarlet. Labeling with CPY-CA (100 nM, 1 hour) is observed only in the presence of rapamycin (100 nM). Scale bars are 50 μm. (**N**) Flow cytometry analysis of HeLa cells coexpressing Lyn11-EGFP-cpHaloΔ-$(GGS)_9$-FKBP and Hpep3-$(GGS)_3$-FRB-mScarlet incubated with CPY-CA (100 nM, 1 hour) in the presence or absence of rapamycin (100 nM). The presence of rapamycin leads to a 39-fold higher median labeling ratio (CPY-CA/EGFP) ($N > 1900$ cells; $p < 0.0001$; Welch's t test). The center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values.

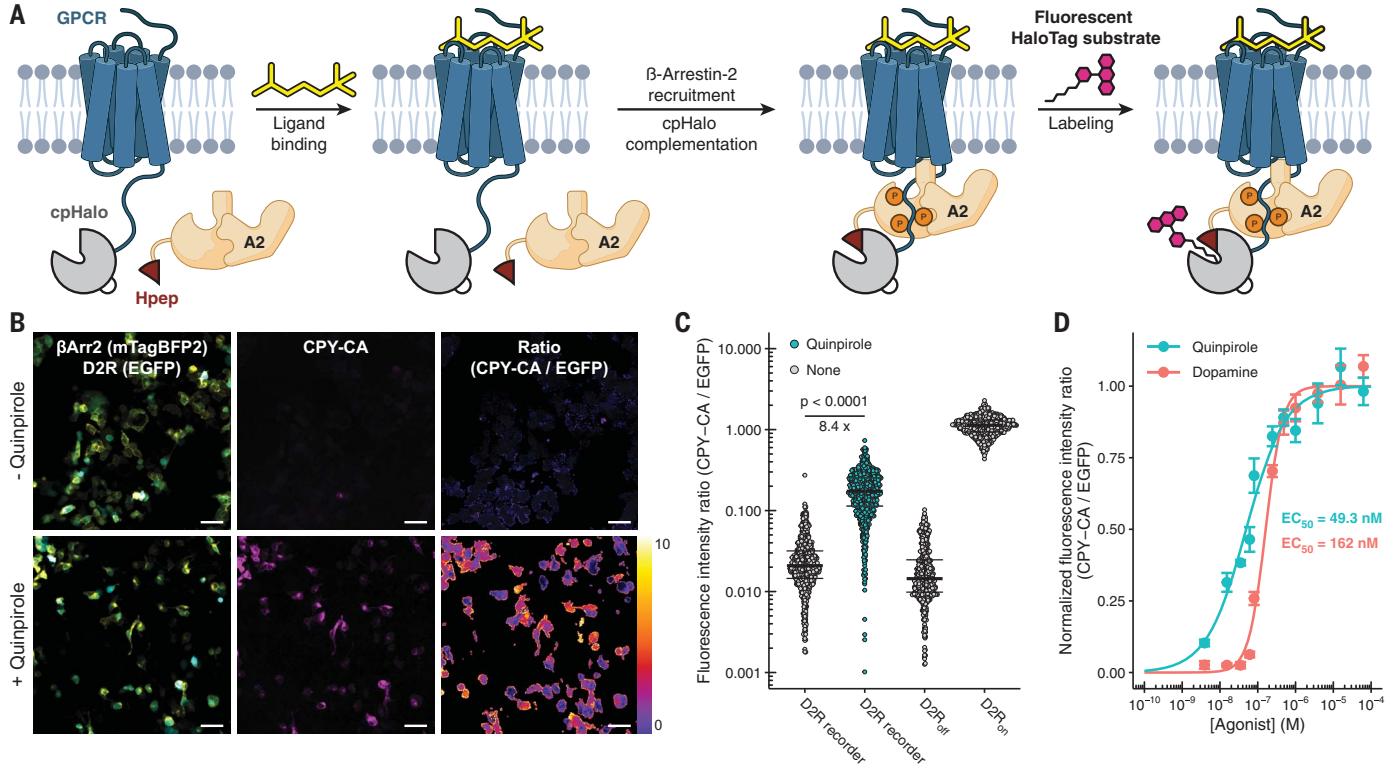

Fig. 2. Recording of GPCR signaling using split-HaloTag. (**A**) Scheme of split-HaloTag–based GPCR recorder. Fusion of Hpep to βArr2 and cpHaloΔ to a GPCR leads to labeling of split-HaloTag after ligand binding and downstream βArr2 recruitment. (**B**) Fluorescence micrographs of human embryonic kidney (HEK) 293 cells coexpressing Hpep1-βArr2-T2A-mTagBFP2 and D2R-cpHaloΔ-GGS-EGFP treated with or without quinpirole (100 μM, 1 hour) in the presence of CPY-CA (200 nM). The color scale bar represents the ratio of fluorescence intensities. Scale bars are 50 μm. (**C**) Quantification of D2R-cpHaloΔ-EGFP labeling ($N > 500$ cells; $p < 0.0001$; Welch's t test). Error bars indicate median and 25 and 75% quantiles. $D2R_{off}$, D2R lacking cpHaloΔ; $D2R_{on}$, D2R fused to cpHaloTag154-156. (**D**) Labeling of HEK293 cells coexpressing D2R-cpHaloΔ and Hpep-βArr2 in the presence of varying concentrations of dopamine or quinpirole. Data are shown as mean ± SEM. A sigmoidal model was applied to determine EC_{50} values.

resulted in an additional eight peptides. Most of these eight peptides showed labeling even at low μM concentrations, conditions at which Hpep1 showed no detectable labeling, and featured median effective concentration (EC_{50}) values ranging from nano- to millimolar (Fig. 1K, fig. S3, and table S2). We fused Hpep3 (EC_{50} = 149 μM) and cpHaloΔ to the distant N termini of FKBP and FRB and observed efficient labeling in the presence of rapamycin, whereas no labeling was detectable in its absence (Fig. 1L). To evaluate the performance of split-HaloTag in live cells, we expressed membrane-localized cpHaloΔ-FKBP and cytosolic Hpep1-FRB or Hpep3-FRB in HeLa cells. Membrane-specific labeling of cpHaloΔ-FKBP with a fluorescent carbopyronine HaloTag substrate (CPY-CA) was only observed with Hpep3-FRB and was strictly dependent on the presence of rapamycin. Labeled cells could be identified by either fluorescence microscopy or flow cytometry, mirroring the results obtained in vitro (Fig. 1, M and N, and figs. S4 and S5). The access to different Hpeps enables the selection of peptides with optimal affinities for each split-HaloTag recorder.

We extended the use of split-HaloTag to receptors of physiologically relevant small molecules (Fig. 2A). G protein–coupled receptors (GPCRs) are activated by various physiologically important ligands such as neurotransmitters and hormones. Recording the activation of GPCRs thus offers the opportunity to record signals transmitted by their cognate ligands. Activation of GPCRs results in downstream binding of arrestins to the cytosolic side of the GPCR, and the interaction between the two proteins can be used to track GPCR activation. This strategy has been used in screening assays (*23*), as well as in light-triggered and chemically triggered recorders (*3*, *24*). As a model system, we fused cpHaloΔ to the C-terminal tail of the human dopamine receptor D2 (D2R) and different Hpeps to the N terminus of β-Arrestin 2 (βArr2). We fused D2R to monomeric enhanced green fluorescent protein (EGFP) to normalize for differences in expression levels. In cells coexpressing D2R and βArr2 split-HaloTag constructs, application of the D2R agonist quinpirole in the presence of CPY-CA led to an eightfold higher fluorescence intensity than in the absence of quinpirole (Fig. 2, B and C). The response was dose-dependent for quinpirole and the native ligand, dopamine (Fig. 2D). In these experiments, Hpep1 yielded the best results because higher-affinity peptides led to greater agonist-independent labeling (fig. S6). We assessed the generality of this method by fusing cpHaloΔ to the C termini of monomeric class A GPCRs muscarinic acetylcholine receptor m3 (mAChR) and β2 adrenergic receptor (β2AR), as well as the heteromeric class C GPCRs γ-aminobutyric acid type B receptor subunit 1 ($GABA_{B1}R$) and metabotropic glutamate receptor 2 (mGluR2). We observed responses to agonists ranging from 1.5- to 18-fold labeling over control, with the exception of baclofen-stimulated $GABA_{B1}R$ (fig. S7). The split-HaloTag–based recorders of GPCR activation thus enable the labeling of cells, depending on received input from physiologically relevant small molecules such as neurotransmitters, for later analysis.

We then designed a split-HaloTag recorder for Ca^{2+}-dependent protein labeling (Caprola) by connecting cpHaloΔ and Hpep1 through calmodulin and the peptide M13. Calmodulin binds to M13 in a Ca^{2+}-dependent manner,

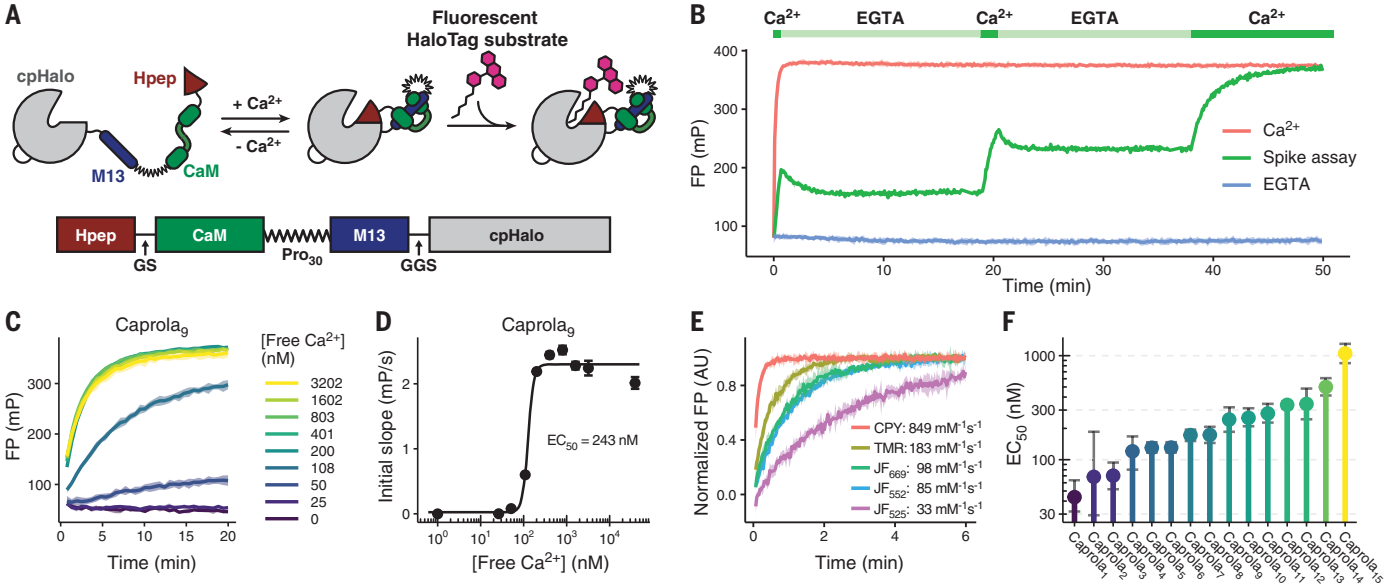

Fig. 3. Design and characterization of the Ca^{2+} recorder Caprola. (**A**) Scheme of Caprola design. GS, Gly-Ser; Pro_{30}, 30 Pro; GGS, Gly-Gly-Ser. (**B**) In vitro labeling kinetics of $Caprola_9$ measured by FP. $Caprola_9$ (200 nM) shows fast labeling kinetics with TMR-CA (50 nM) in the presence of Ca^{2+} (5 mM) and low background activity in the presence of Ca^{2+}-chelator EGTA (100 μM) and can be repeatedly activated and deactivated by successive additions of Ca^{2+} and EGTA, respectively. (**C**) Labeling kinetics of $Caprola_9$ (200 nM) with TMR-CA (50 nM) at different free-Ca^{2+} concentrations measured by FP. (**D**) Ca^{2+} sensitivity of $Caprola_9$ determined by plotting free-Ca^{2+} concentrations against the initial slopes of the labeling reactions from (C). The EC_{50} represents the calcium concentration at which $Caprola_9$ shows half of its maximum labeling speed. (**E**) Labeling kinetics and apparent second-order rate constants of $Caprola_9$ (200 nM) with different fluorescent substrates (50 nM) in the presence of Ca^{2+} (5 mM). FP values were normalized to their unbound and fully bound values (normalized FP). AU, arbitrary units. (**F**) Ca^{2+} sensitivity (EC_{50} values) of the 15 Caprola variants as determined in (D). Values range from 44 to 1050 nM. Error bars represent 95% confidence intervals.

and we used binding of calmodulin to M13 to enable complementation of cpHaloΔ by Hpep1 (Fig. 3A). We chose the low-affinity peptide Hpep1 to minimize Ca^{2+}-independent labeling. Caprola showed negligible labeling in the absence of Ca^{2+} but more than 10^3-fold faster labeling at saturating concentrations of Ca^{2+} [second-order rate constant in the presence of Ca^{2+} (k_{Ca}) = 2.61×10^5 $M^{-1}s^{-1}$; fig. S8 and table S3]. Activation of Caprola by Ca^{2+} was reversible: Addition of the Ca^{2+} chelator EGTA stopped labeling, but it resumed after further addition of Ca^{2+} (Fig. 3B). This reversibility ensures that periods of increased Ca^{2+} before the addition of the fluorescent substrate do not result in labeling. The cooperative binding of four Ca^{2+} ions to calmodulin results in a sharp threshold of Ca^{2+} concentration below which negligible labeling of Caprola was observed (Fig. 3, C and D). As observed for other split-HaloTag constructs (fig. S9), Caprola was efficiently labeled with spectrally distinguishable fluorescent substrates (Fig. 3E), which enabled sequential recording of different periods of Ca^{2+} exposure (fig. S10). Different applications will require Caprola variants with different Ca^{2+} sensitivities. By introducing M13 point mutations that control Ca^{2+} affinity, we created the Caprola variants $Caprola_1$ to $Caprola_{15}$ with EC_{50} values ranging from 44 nM to 1.1 μM (Fig. 3F and fig. S11). All Caprola variants maintained high labeling rates upon Ca^{2+} binding (table S3).

We used Caprola to record changes in cytosolic Ca^{2+} in cultured mammalian cells. We fused Caprola to EGFP to normalize for differences in expression levels. We further constructed constitutively active and inactive versions of Caprola ($Caprola_{on}$ and $Caprola_{off}$). HeLa cells expressing $Caprola_5$-EGFP were incubated with CPY-CA in the presence or absence of thapsigargin, an inhibitor of the sarcoplasmic and endoplasmic reticulum Ca^{2+} adenosine triphosphatase (ATPase) that increases cytosolic Ca^{2+}. Incubation with thapsigargin led to strong labeling, whereas little labeling was observed in its absence (Fig. 4A). Flow cytometry revealed a 35-fold change in normalized fluorescence intensity between treated and untreated cells (Fig. 4B), which approached the difference observed between $Caprola_{on}$ and $Caprola_{off}$. Similar behavior was observed for most other Caprola variants (figs. S12 and S13). The fluorescent marks resisted chemical fixation (Fig. 4A and fig. S14), and labeling times as short as 5 min were sufficient to detect a difference between treated and untreated cells (fig. S15). Moreover, a variety of fluorescent substrates allowed thapsigargin-dependent fluorescent marking of cells (Fig. 4C). To demonstrate the marking of cells over multiple defined recording periods, we first incubated cells expressing $Caprola_5$-EGFP with the fluorescent Janelia Fluor 552 HaloTag substrate [JF_{552}-CA (*25*)] and adenosine triphosphate (ATP) (moderate Ca^{2+} increase), second with CPY-CA in the absence of stimulus, and third with the fluorescent HaloTag substrate JF_{669}-CA (*26*) and thapsigargin (strong Ca^{2+} increase) (Fig. 4D). Control experiments showed that after the first recording period, sufficient sensor protein was still available for subsequent recordings (fig. S16). Imaging after the three recordings revealed the expected pattern of moderate, weak, and strong labeling with JF_{552}, CPY, and JF_{669}, respectively. The differences in labeling efficiencies that were observed for the different periods were independent of which fluorescent substrate was used for which period (fig. S16).

Caprola recorded transient increases in cytosolic Ca^{2+} in cultured primary rat hippocampal neurons. We first verified that Caprola expression did not alter neuronal physiology (fig. S17). Stimulating neurons expressing $Caprola_4$-EGFP with glutamate and glycine resulted in a significant increase in the fluorescence intensity ratios relative to basally active neurons, whereas silencing synaptic transmission between neurons with antagonists of the *N*-methyl-D-aspartate (NMDA) and the α-amino-3-hydroxy-5-methyl-isoxazolepropionic acid (AMPA) receptors resulted in decreased labeling relative to that observed in basally active neurons (Fig. 4, E and

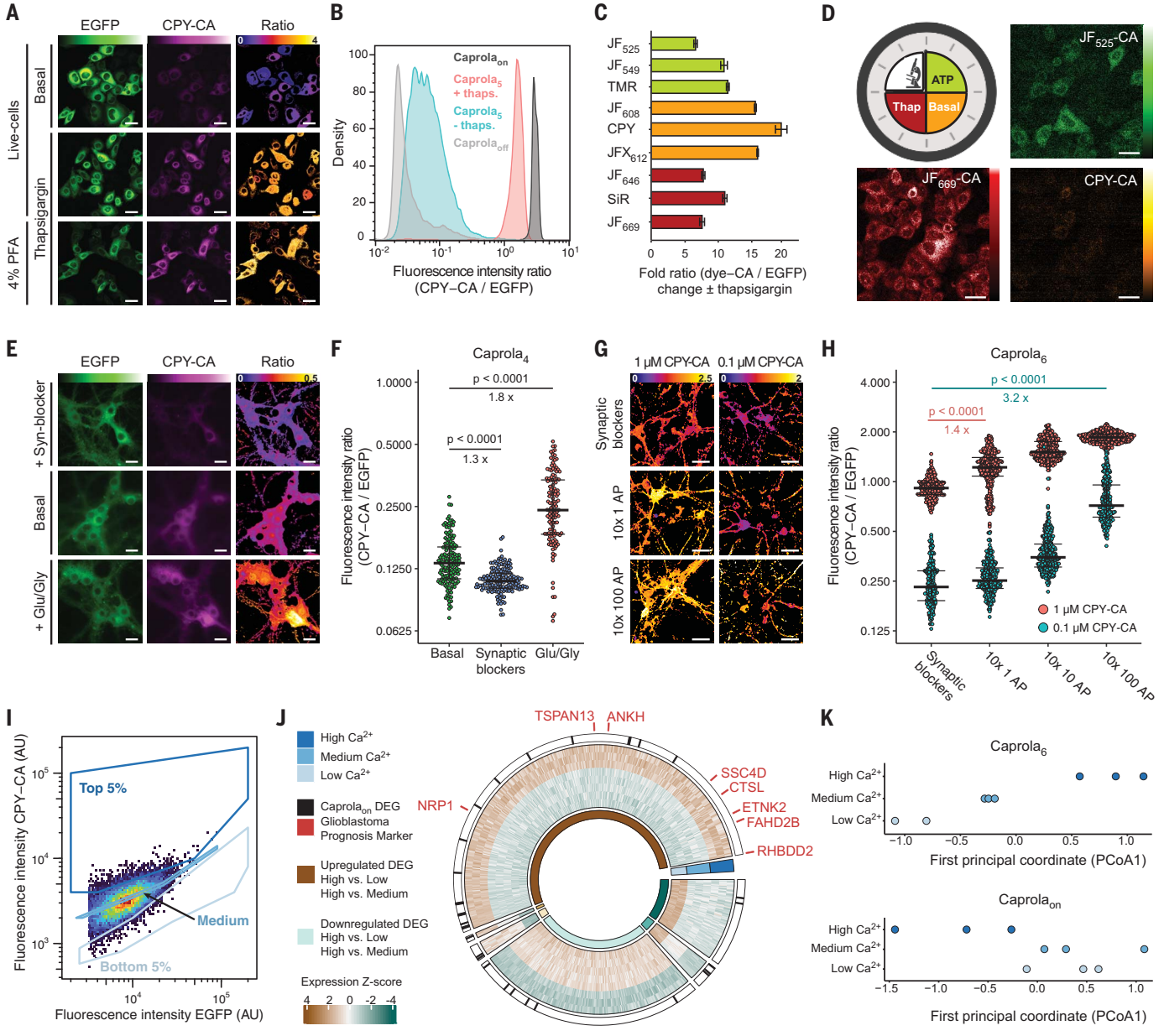

Fig. 4. Recording calcium signaling using Caprola. (**A**) Fluorescence micrographs of HeLa cells expressing Caprola$_5$-EGFP and incubated with CPY-CA (100 nM) in the presence or absence of thapsigargin (100 nM, 1 hour). Live and chemically fixed cells are shown. Ratio indicates CPY-CA/EGFP. (**B**) Flow cytometry analysis of HeLa cells expressing Caprola$_5$-EGFP, Caprola$_{on}$-EGFP, or Caprola$_{off}$-EGFP incubated with CPY-CA (100 nM, 1 hour) in the presence or absence of thapsigargin (100 nM) (N > 10,000 cells). (**C**) Flow cytometry analysis of HeLa cells expressing Caprola$_5$-EGFP incubated with different fluorescent substrates (100 nM, 1 hour) in the presence or absence of thapsigargin (100 nM) (N > 10,000 cells). Error bars represent SEM. (**D**) Recording of three successive periods of Ca^{2+}-activity in HeLa cells expressing Caprola$_5$-EGFP. First period: ATP (100 µM) and JF$_{525}$-CA (300 nM, 1 hour); second period: no treatment and CPY-CA (25 nM, 1 hour); third period: thapsigargin (Thap; 100 nM) and JF$_{669}$-CA (300 nM, 1 hour). Between incubations, cells were allowed to recover for 2 hours. Fluorescence micrographs were acquired after the three recording periods. (**E**) Fluorescence micrographs of primary rat hippocampal neurons expressing Caprola$_4$-EGFP labeled with CPY-CA (250 nM, 30 min) in the presence of synaptic blockers APV/NBQX (25 µM/10 µM) or glutamate/glycine (Glu/Gly; 10 µM/2.5 µM) or without treatment. Ratio indicates CPY-CA/EGFP. (**F**) Quantification of Caprola$_4$ labeling from experiments described in (E) (N > 100 cells). The center line indicates the median, and error bars indicate 25 and 75% quantiles. (**G**) Ratiometric fluorescence micrographs (CPY-CA/EGFP) of primary rat hippocampal neurons expressing Caprola$_6$-EGFP incubated with CPY-CA (1 or 0.1 µM, 30-min preincubation, 1-hour stimulation) upon defined electrical field stimulation. Action potentials (APs) were delivered in 10 short pulses (80 Hz) over the course of 1 hour interspaced with resting periods (approximately 7 min). (**H**) Quantification of Caprola$_6$ labeling from experiments described in (G) (N > 100 cells). The center line indicates the median, and error bars indicate 25 and 75% quantiles. (**I**) Gating strategy of CPY-labeled (125 nM, 90 min) glioblastoma cells expressing Caprola$_6$-EGFP for later RNA-seq analysis. (**J**) Transcriptional profiles of the three sorted groups from (I). DEGs identified by RNA-seq analysis are color coded according to Z-scores. DEGs were arranged in six groups according to their pairwise comparison pattern (e.g., up-regulated high versus low and high versus medium; inner circle). Glioblastoma prognosis markers are indicated in red, and DEGs identified in both Caprola$_6$ and Caprola$_{on}$ experiments are highlighted as black lines (outer circle). (**K**) Dot plots of the first principal coordinate analysis from Caprola$_6$-EGFP– and Caprola$_{on}$-EGFP–expressing glioblastoma cells. Statistical significances in panels (F) and (H) were calculated with Welch's t test, and p values are given for each comparison. In (A), (D), (E), and (G), color scale bars represent either fluorescence intensities or ratios, and scale bars are 20 µm.

F, and fig. S18). In pharmacologically silenced neurons, Caprola was able to record the increase in intracellular calcium elicited by as few as 10 electrically evoked action potentials (Fig. 4G and fig. S17). The concentration of the fluorescent substrate could be adjusted to control the sensitivity of the recording: Higher concentrations of fluorescent substrates recorded weaker stimuli (Fig. 4H and fig. S19). Pharmacological or electrical stimulations as short as 5 min were sufficient to record differences between treated and untreated neurons (fig. S20). Furthermore, the signal in labeled, cultured neurons remained detectable for at least 3 days after recording (fig. S21).

We used Caprola to sort cells according to their relative Ca^{2+} concentrations during defined time periods for later analysis. Specifically, Caprola was applied to decipher the molecular heterogeneity of glioblastoma, the most common and aggressive primary brain tumor in adults (*27*, *28*). Glioblastoma cells form multicellular, heterogeneous networks with Ca^{2+} transients that drive tumor progression in vivo (*29*–*33*). Patient-derived glioblastoma cell lines also form cellular networks with Ca^{2+} transients in vitro (*34*). The factors in glioblastoma cells that drive Ca^{2+} transients in such networks remain to be identified and might represent therapeutic targets. We therefore cultured glioblastoma cells under conditions known to result in network formation. We used Caprola to label cells according to their intracellular Ca^{2+} concentrations and later sorted them by flow cytometry for downstream analysis by RNA sequencing (RNA-seq). Glioblastoma cells that stably express $Caprola_6$-EGFP were labeled with CPY-CA and sorted into three groups based on their normalized labeling intensity (high, medium, and low; Fig. 4I and fig. S22). These three groups were subjected to bulk RNA-seq. We identified 757 differentially expressed genes (DEGs) between all groups in pairwise group analysis (Fig. 4J). To rule out that heterogeneities in the permeability of CPY-CA contributed to the observed differences in $Caprola_6$ labeling, we used glioblastoma cells expressing $Caprola_{on}$-EGFP and sorted the cells into three groups as described above (figs. S22 and S23). Principal coordinate analysis of the transcriptomic profiles clearly separated three groups in $Caprola_6$, but not in $Caprola_{on}$, experiments (Fig. 4K). Only 31 of the 757 $Caprola_6$-identified DEGs were also differentially regulated in the positive control experiment and hence were excluded from further analysis. Gene ontology (GO) term analysis revealed that the remaining 726 DEGs were associated with cell division and neurogenesis, whereas the 31 DEGs also identified through $Caprola_{on}$ were mostly associated with transporters (fig. S24). Among the 726 $Caprola_6$-specific DEGs, eight DEGs are unfavorable prognostic genes of glioblastoma (*35*), all of which were exclusively up-regulated in the high Ca^{2+} group (Fig. 4J and fig. S23). These findings demonstrate that cells that form a heterogeneous network can be individually labeled according to Ca^{2+} levels within the intact network and later be sorted based on the label into subpopulations with different transcriptomes.

Recording transient physiological parameters for later analysis is particularly valuable for in vivo studies. Therefore, we used Caprola to record cytosolic calcium transients in two popular model organisms in neuroscience: adult flies and zebrafish larvae.

We used Caprola to record cytosolic calcium activity in the visual system of *Drosophila melanogaster*. Specifically, we generated lines that express $Caprola_6$-EGFP in T4 and T5 neurons, which respond in a direction-selective manner to local visual motion (*36*). We stimulated the left, the right, or both eyes of tethered flies by exposing them to moving sinusoidal gratings for 20 min in the presence of CPY-CA (Fig. 5A), which was applied through an incision at the back of the head. T4 and T5 cells predominantly receive visual input from the ipsilateral eye (*37*). Despite interindividual variabilities, we found similar signals in both optic lobes when stimulating both eyes (Fig. 5B and fig. S25) and stronger signals on the ipsilateral side with unilateral visual stimulation (Fig. 5B and fig. S25). Caprola can thus record differences in neuronal activity in flies on timescales as short as 20 min.

We generated zebrafish lines that pan-neuronally expressed either $Caprola_1$-EGFP, $Caprola_6$-EGFP, $Caprola_{on}$-EGFP, or $Caprola_{off}$-EGFP. Zebrafish larvae expressing active Caprola variants could be stained by simply incubating the fish with CPY-CA, JF_{669}-CA, JF_{552}-CA, or JF_{525}-CA, whereas zebrafish larvae expressing $Caprola_{off}$-EGFP showed little background staining (fig. S26). Staining of the nervous system of zebrafish larvae expressing active Caprola variants was first observed in the forebrain after 30 min and in the hindbrain after 60 min, reflecting the uptake of the fluorescent substrate (fig. S27). The kinetics of staining observed in these experiments determine the minimal temporal resolution of the approach. Because labeling efficiency of Caprola generally depends on the concentration of both Ca^{2+} and fluorescent substrate, comparisons should only be made within the same brain regions. To use Caprola for recordings of neuronal activity in vivo, we mounted zebrafish larvae in agarose in the presence of CPY-CA and exposed them to sideways-drifting black-and-white gratings (Fig. 5C). This visual stimulation evokes horizontal eye movements, the so-called optokinetic responses (*38*). After 6 hours, we analyzed the arborization field 9 (AF9) ventral to the tectum, which receives direct input from the retina and is sensitive to changes in ambient luminance, including whole-field motion (*39*, *40*). The visual stimulation resulted in labeling of neurons in AF9 (Fig. 5D). Occluding the left eye of the larvae with opaque agarose resulted in a modest, yet statistically significant, decrease in labeling in the AF9 of the right hemisphere, and vice versa, consistent with contralateral processing of visual inputs (Fig. 5E and figs. S28 and S29). To record two periods of visual stimulation in the same larvae, we first visually stimulated them with one eye occluded and repeated the stimulation the next day with the other eye occluded. These two different periods of asymmetric visual stimulation were recorded by incubation with distinguishable fluorescent substrates. Later analysis revealed the expected asymmetric labeling pattern for both fluorescent substrates, highlighting that Caprola can record different periods of neuronal activity in vivo (Fig. 5F and fig. S30). We recorded neuronal activity in freely swimming zebrafish larvae that were hunting prey (*41*). The AF7 neuropil is known to be activated by the sight of prey objects (*40*, *42*). In a later analysis, we indeed observed increased fluorescence labeling in this region relative to that in control fish that were not exposed to prey (Fig. 5, G and H, and fig. S31). These experiments demonstrate that Caprola allows direct recording of neuronal activity with cellular resolution in immobilized and freely moving animals for later analysis.

Discussion

We developed a split-HaloTag system comprising a folded domain and a palette of small peptides in which the fluorescent labeling of the folded domain is dependent on its binding to the small peptide. Key characteristics of split-HaloTag are the reversibility of the interaction between the folded domain and the small peptide as well as the availability of small peptides with affinities ranging from nM to mM. This enables the rational design of recorders for various transient cellular activities, such as protein-protein interactions, activation of GPCRs, and elevations in intracellular Ca^{2+} levels, where the presence of the cellular activity and a fluorescent substrate leads to irreversible fluorescent labeling of the recorder.

Our recorders allow the recording of transient activities for later analysis. The recording period is determined by the addition and wash-out of the fluorescent substrate; activities occurring before addition or after wash-out do not result in the labeling of the recorder. The temporal resolution of the recording is determined by the permeability of the fluorescent substrates in cell culture experiments and by their pharmacokinetics in animal experiments. The availability of fluorescent substrates of different colors enables sequential recording of different periods of activity in the same sample or organism. Because the fluorescent

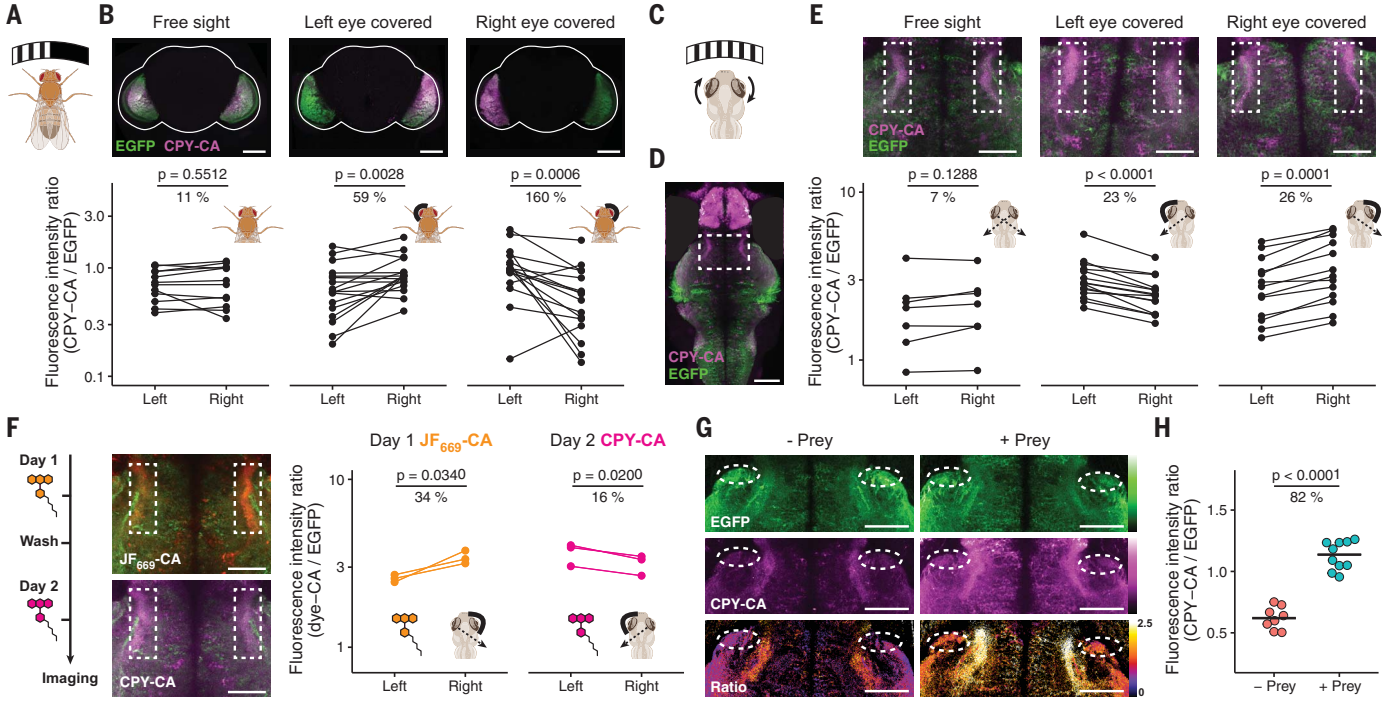

Fig. 5. Recording of neuronal calcium signaling in vivo. (**A**) Cartoon of the experimental setup for visual stimulation of flies. A curved monitor was positioned in front of adult flies expressing Caprola$_5$-EGFP to stimulate both or only one of the eyes using moving sine wave gratings. (**B**) Asymmetrical visual stimulation of flies recorded with Caprola$_5$-EGFP in presence of CPY-CA (5 μM, 20 min). Representative fluorescence micrographs for each condition and pair plots of mean ratios (CPY-CA/EGFP) of left and right lobes are shown ($N \geq 12$ individuals per condition). The shape of the whole brain is outlined in white. (**C**) Cartoon of the experimental setup for visual stimulation of zebrafish larvae. Black-and-white gratings, moving in both horizontal directions in front of the agarose-embedded larvae, evoke optokinetic eye movements (indicated by curved arrows). (**D**) Fluorescence micrograph of a zebrafish larva brain expressing Caprola$_1$-EGFP labeled with CPY-CA (5 μM, 6 hours) during visual stimulation (maximum intensity projection). The white box indicates the region of interest containing the AF9. Fluorescent signal from the zebrafish's eyes was removed. (**E**) Asymmetrical visual stimulation of zebrafish larvae recorded with Caprola$_1$-EGFP. Individual eyes of larvae were occluded or left unobstructed during visual stimulation in presence of CPY-CA (5 μM, 6 hours). Representative fluorescence micrographs of AF9 (white boxes) are shown. Pair plots of mean ratios (CPY-CA/EGFP) from AF9 regions are shown for each condition ($N \geq 7$ larvae per condition). (**F**) Sequential recordings of asymmetrical visual stimulation from AF9 on two consecutive days using Caprola$_1$-EGFP with JF$_{669}$-CA (5 μM, 6 hours) and CPY-CA (5 μM, 6 hours), respectively. Fluorescence micrographs of AF9 (white boxes) were acquired after the two recording periods. Pair plots of mean ratios (JF$_{669}$-CA/EGFP and CPY-CA/EGFP) from AF9 are shown for both fluorescent substrates ($N = 3$ larvae per condition). (**G**) Freely swimming zebrafish larvae expressing Caprola$_1$-EGFP labeled with CPY-CA (5 μM, 2 hours) in the presence or absence of prey (paramecia). Fluorescence micrographs of AF7 (white ovals) are shown. Ratio indicates CPY-CA/EGFP. Color scale bars represent either fluorescence intensities or ratios. (**H**) Dot plots of mean ratios from AF7 are shown ($N \geq 4$ larvae per condition). Statistical significances in (B), (E), (F), and (H) were calculated with one-tailed Welch's *t* test (or two-tailed for symmetric stimulation), and *p* values are given for each comparison. Pair-plots percentages indicate the mean ratio increase from the lower to the higher condition. Scale bars are 100 μm [(B) and (D)] and 50 μm [(E) to (G)].

signal remains detectable over days and is resistant to fixation strategies, the approach is scalable and allows the simultaneous analysis of large populations of cells or organisms. The persistence of the fluorescent signal also enables the sorting of cells according to the labeling of the recorder and subsequent transcriptomic analysis. Because our recorders are directly activated by the activity of interest and can be targeted to defined cellular locations, they offer the opportunity to record cellular activities with spatial resolution.

All of these features are exemplified in our experiments with Caprola, which we have used to record neuronal activity in zebrafish larvae and adult flies as well as for the sorting and transcriptomic analysis of heterogeneous cell populations according to their Ca^{2+} levels. Caprola does not offer the same spatiotemporal resolution as light-dependent recorders such as CaMPARI because it requires addition and wash -out of the chemical probe, but it enables multiple distinguishable recordings of Ca^{2+} transients in the same animal over periods of minutes and hours. Caprola should thus become a complementary tool to link behavior to the activity of neuronal subpopulations and, more generally, to investigate physiological processes that involve Ca^{2+} signaling. Considering the prevalence of Ca^{2+} signaling and the successful labeling of HaloTag in many model systems (*16*, *19*), Caprola may have applications in various areas of biology.

Furthermore, bearing in mind the design principle of our three recorders, that is, linking a cellular activity to an increase in proximity of the two components of split-HaloTag, it should be possible to design recorders for additional cellular activities by borrowing from the design of other existing protein-based fluorescent biosensors that use similar principles. The split-HaloTag system thus creates the opportunity to generate various recorders that may help to link the physiological history of cells with biological phenotypes.

REFERENCES AND NOTES

1. C. J. Guenthner, K. Miyamichi, H. H. Yang, H. C. Heller, L. Luo, *Neuron* **78**, 773–784 (2013).
2. X. J. Gao *et al.*, *Nat. Neurosci.* **18**, 917–925 (2015).
3. H. K. Inagaki *et al.*, *Cell* **148**, 583–595 (2012).
4. M. Sheng, M. E. Greenberg, *Neuron* **4**, 477–485 (1990).
5. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).
6. D. Lin *et al.*, *Nat. Biotechnol.* **41**, 631–639 (2023).

7. C. Linghu *et al*., *Nat. Biotechnol.* **41**, 640–651 (2023).
8. R. Y. Tsien, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12456–12461 (2013).
9. F. Schmidt *et al*., *Science* **376**, eabm6038 (2022).
10. F. Farzadfard *et al*., *Mol. Cell* **75**, 769–780.e4 (2019).
11. B. F. Fosque *et al*., *Science* **347**, 755–760 (2015).
12. A. Das *et al*., *Nat. Commun.* **14**, 6399 (2023).
13. D. Lee, J. H. Hyun, K. Jung, P. Hannan, H. B. Kwon, *Nat. Biotechnol.* **35**, 858–863 (2017).
14. W. Wang *et al*., *Nat. Biotechnol.* **35**, 864–871 (2017).
15. T. W. Chen *et al*., *Nature* **499**, 295–300 (2013).
16. A. S. Abdelfattah *et al*., *Science* **365**, 699–704 (2019).
17. J. Ohata *et al*., *ACS Cent. Sci.* **6**, 32–40 (2020).
18. G. V. Los *et al*., *ACS Chem. Biol.* **3**, 373–382 (2008).
19. E. Bulovaite *et al*., *Neuron* **110**, 4057–4073.e8 (2022).
20. J. B. Grimm *et al*., *Nat. Methods* **14**, 987–994 (2017).
21. J. Wilhelm *et al*., *Biochemistry* **60**, 2560–2575 (2021).
22. L. A. Banaszynski, C. W. Liu, T. J. Wandless, *J. Am. Chem. Soc.* **127**, 4715–4721 (2005).
23. W. K. Kroeze *et al*., *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).
24. D. Lee *et al*., *Nat. Methods* **14**, 495–503 (2017).
25. Q. Zheng *et al*., *ACS Cent. Sci.* **5**, 1602–1613 (2019).
26. J. B. Grimm *et al*., *Nat. Methods* **17**, 815–821 (2020).
27. M. Weller *et al*., *Nat. Rev. Dis. Primers* **1**, 15017 (2015).
28. K. D. Miller *et al*., *CA Cancer J. Clin.* **71**, 381–406 (2021).
29. M. Osswald *et al*., *Nature* **528**, 93–98 (2015).
30. V. Venkataramani *et al*., *Nature* **573**, 532–538 (2019).
31. H. S. Venkatesh *et al*., *Nature* **573**, 539–545 (2019).
32. P. G. Gritsenko *et al*., *Nat. Cell Biol.* **22**, 97–107 (2020).
33. F. Winkler, W. Wick, *Science* **359**, 1100–1101 (2018).
34. J. Lee *et al*., *Cancer Cell* **9**, 391–403 (2006).
35. J. N. Weinstein *et al*., *Nat. Genet.* **45**, 1113–1120 (2013).
36. M. S. Maisak *et al*., *Nature* **500**, 212–216 (2013).
37. K. F. Fischbach, A. P. M. Dittrich, *Cell Tissue Res.* **258**, 441–475 (1989).
38. Y. Wu, M. Dal Maschio, F. Kubo, H. Baier, *Neuron* **108**, 722–734.e5 (2020).
39. E. Robles, E. Laurell, H. Baier, *Curr. Biol.* **24**, 2085–2096 (2014).
40. J. L. Semmelhack *et al*., *eLife* **3**, e04878 (2014).
41. D. S. Mearns, J. C. Donovan, A. M. Fernandes, J. L. Semmelhack, H. Baier, *Curr. Biol.* **30**, 54–69.e9 (2020).
42. P. Antinucci, M. Folgueira, I. H. Bianco, *eLife* **8**, e48114 (2019).
43. J. Wilhelm, johnsson-lab/splitHaloTag_Hpep_design. Zenodo (2024); https://doi.org/10.5281/zenodo.8113621.

ACKNOWLEDGMENTS

We thank I. Schlichting for x-ray data collection. Diffraction data were collected at the Swiss Light Source, beamline X10SA, of the Paul Scherrer Institute, Villigen, Switzerland. We thank J. Reinstein for help with stopped-flow experiments. We thank A. Andres-Pons (European Molecular Biology Laboratory, Heidelberg) for providing the HeLa Kyoto Flp-In cell line, the Flow Cytometry Core Facility of the German Cancer Research Center (DKFZ) (Heidelberg) for support, the Genomics and Proteomics Core Facility of the DKFZ for sequencing, the Omics IT and Data Management Core Facility of the DKFZ for processing RNA-seq data and data storage, and the Optical Microscopy facility of the Max Planck Institute for Medical Research for support with microscopy experiments. We thank L. D. Lavis for providing Janelia Fluor Dyes; A. Bergner, B. Koch, A. Herold, B. Réssy, D. Schmidt, E. D'Este, J. Hubrich, and T. Schilling for reagents or materials; B. Solak and A. Friedrich for assistance with fly experiments; K. Finger-Baier and J. Huppert for assistance with animal protocols; S. Prech for building the monitor mount; and F. Broch, J. Tünnermann, M. Ziegler, and I. Arnold-Ammer for technical assistance. M.C.H., J.W., and N.P. are members of the Heidelberg Biosciences International Graduate School (HBIGS). **Funding:** This work was funded by the Max Planck Society (M.-C.H., J.W., V.G., M.W.S., N.P., M.T., K.S., T.F., L.M.F., H.B., J.H., K.J.); Ecole Polytechnique Federale de Lausanne (EPFL) (K.J.); Swiss National Science Foundation grant 31003A_166316 (K.J.); Deutsche Forschungsgemeinschaft (DFG) [SFB TRR 186 (K.J.), SPP 1926 (H.B.), SFB 1398 (D.C.H., D.H., L.H.), and SFB1158 (C.A.)]; Boehringer Ingelheim Fonds (M.C.H., M.W.S.); Max Planck School Matter to Life (J.W.); a Humboldt Research Fellowship (V.G.); European Union's Horizon 2020 grant 955623 (N.P.); the Chica and Heinz Schaller Stiftung (C.A.); the Fritz Thyssen Stiftung (C.A.); and a NARSAD 2019 Young Investigator Award (C.A.). **Author contributions:** K.J. conceived the project. J.H. and K.J. planned and supervised the work. J.W. and J.H. performed the in vitro experiments. J.W. performed computational modeling and data analysis. M.-C.H., V.G., and N.P. performed the mammalian cell experiments. M.-C.H., D.C.H., and D.H. performed the glioblastoma experiments. L.H. conducted RNA-seq data analysis. M.-C.H. and I.K. performed the primary neuron–related experiments. G.P. and C.A. performed the electrophysiology measurements. T.F., M.-C.H., and L.M.F. designed and performed the *Drosophila* experiments. M.-C.H., M.W.S., K.S., and H.B. designed and performed the zebrafish experiments. M.T. solved the crystal structures. J.W. and J.H. analyzed the crystal structures. M.-C.H., J.W., V.G., J.H., and K.J. wrote the manuscript with input from all authors. **Competing interests:** M.-C.H., J.W., J.H., and K.J. are inventors of patents entitled "Circular permuted haloalkane transferase fusion molecules" and "Improved Split-HaloTag components" filed by the Max Planck Society. **Data and materials availability:** All data are available in the paper or the supplementary materials. The x-ray crystal structures of cpHaloTag154-156 and cpHaloTagΔ have been deposited to the Protein Data Bank (PDB) under accession codes 8B6P and 8B6N, respectively. Plasmids of interest from the study have been deposited at Addgene; accession codes are provided in the supplementary materials. Scripts and input files for the computational design are available at https://github.com/johnsson-lab/splitHaloTag_Hpep_design and at Zenodo (*43*). RNA-seq data are available at the European Genome-Phenome Archive (https://ega-archive.org/) database under accession number EGAS50000000056. Sources of the HaloTag substrates are listed in the supplementary materials. Reagents and materials are available from the corresponding authors upon request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg0812
Materials and Methods
Supplementary Text
Figs. S1 to S32
Tables S1 to S7
References (*44*–*81*)
Data S1

Submitted 1 December 2022; accepted 22 January 2024
10.1126/science.adg0812