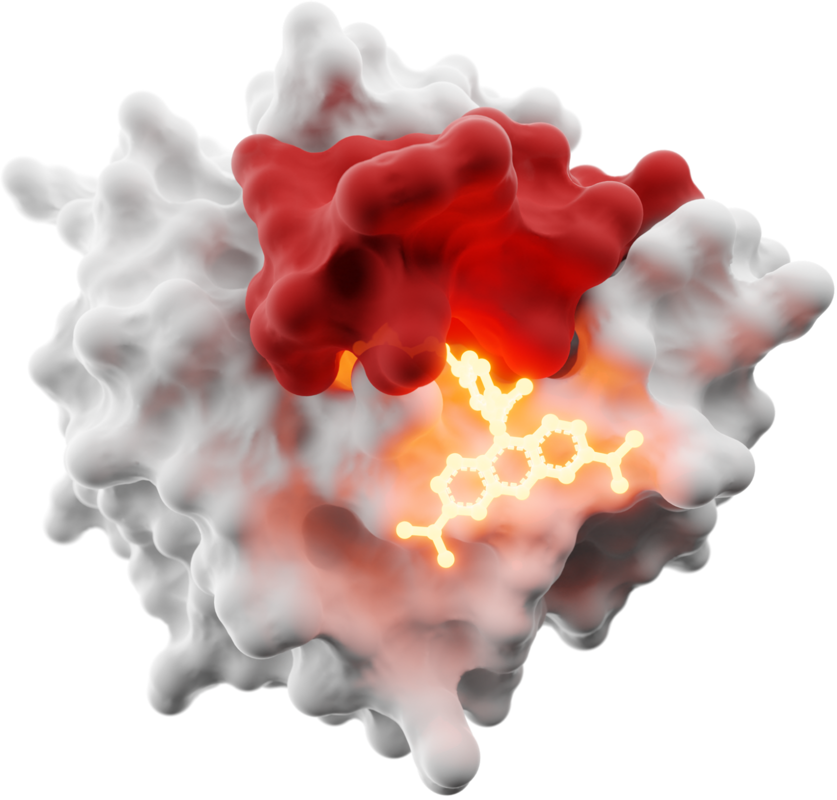
Inaugural Dissertation

Engineering of a Split Self-Labeling Protein for Recording Neuronal Activity and Connectivity

---



Jonas Wilhelm

2023

Inaugural Dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences of the

Ruprecht – Karls – University

Heidelberg

Presented by

M.Sc. Jonas Wilhelm

born in Karlsruhe, Germany

Oral examination: 19.12.2023

Engineering of a Split Self-Labeling Protein for Recording Neuronal Activity and Connectivity

Referees: Prof. Dr. Britta Brügger
Prof. Dr. Kai Johnsson

Abstract

Recording transient cellular events and biomolecular interactions on large scales is pivotal for understanding the mechanisms underlying diverse biological functions. This requires the development of innovative molecular tools that enable the massive parallel analysis of these phenomena in their natural context. To address this challenge, I developed a split version of HaloTag, a self-labeling protein that can be irreversibly labeled with a wide variety of fluorophore ligands. While both split-HaloTag fragments are inactive in isolation, their labeling activity is restored when brought into direct proximity by fusion to interacting proteins.

Initially, I performed an in-depth characterization of HaloTag to gain insights into its biochemical properties, kinetics and substrate preferences. Building upon this knowledge, I engineered a split-HaloTag comprised of a large folded part and a small complementing peptide, and thoroughly characterized this split system to understand its properties and mechanism. By fusing the split-HaloTag parts to calcium-sensing proteins, I created a molecular recorder that becomes irreversibly labeled upon exposure to calcium, a universal second messenger in cellular signaling and indicator of neuronal activity. The recording window is defined by the presence of the fluorescent HaloTag ligand, and thus the successive usage of ligands with different colors enables recording of multiple epochs of calcium activity. The tool has facilitated the recording of brain-wide activity patterns in living flies and zebrafish larvae during visual stimulation in a follow-up study.

I then engineered split-HaloTag to label interactions of synaptic adhesion proteins, in order to develop a highly selective method to stain neuronal connectivity. I used computational protein design methods to fine-tune the intrinsic affinity of the split-HaloTag fragments, and to increase the stability of the large fragment. The refined split-HaloTag successfully labels synaptic adhesion complexes at cell-cell interfaces, laying the foundation for a highly selective approach to large-scale connectivity mapping. Moreover, the usage of HaloTag ligands with different colors should permit the identification of newly formed synapses on a brain-wide scale.

The excellent properties and versatility of the improved split-HaloTag system spurred the development of several other molecular recorders and labeling strategies in our group. Thus, the technology offers a promising pathway for exploring diverse biological activities and functions across various model systems.

Zusammenfassung

Die Aufzeichnung transienter zellulärer Ereignisse und biomolekularer Interaktionen ist von zentraler Bedeutung für das Verständnis von systemischen Prozessen, die den verschiedensten biologischen Funktionen zugrunde liegen. Dies erfordert die Entwicklung innovativer molekularer Werkzeuge, die eine umfangreiche parallele Analyse dieser Phänomene in ihrem natürlichen Kontext erlauben. Daher habe ich eine Split-Version von HaloTag entwickelt, einem selbstmarkierenden Protein, das mit einer Vielzahl von Fluorophorliganden irreversibel markiert werden kann. Während beide Split-HaloTag Fragmente für sich inaktiv sind, wird ihre Markierungsaktivität wiederhergestellt, wenn sie durch Fusion mit interagierenden Proteinen in räumliche Nähe gebracht werden.

Zunächst habe ich eine tiefgehende Analyse von HaloTag durchgeführt, um dessen biochemische Eigenschaften, Reaktionskinetiken und Substratpräferenzen zu charakterisieren. Darauf aufbauend habe ich ein Split-HaloTag System entwickelt, das aus einem großen gefalteten Fragment und einem kleinen komplementären Peptid besteht, und dessen mechanistische Eigenschaften eingehend untersucht. Durch die Fusion der Split-HaloTag Fragmente mit Calcium-sensitiven Proteinen habe ich einen molekularen Rekorder konstruiert, der irreversibel markiert wird, wenn er Calcium ausgesetzt ist, einem universellen sekundären Botenstoff in der zellulären Signalübertragung und Indikator für neuronale Aktivität. Das Aufzeichnungsfenster des Rekorders wird durch das Vorhandensein des fluoreszenten HaloTag-Ligandens definiert, sodass die aufeinanderfolgende Verwendung von Liganden mit unterschiedlichen Farben die Aufzeichnung mehrerer Intervalle mit unterschiedlicher Calciumaktivität erlaubt. Die Verwendung dieses Systems ermöglichte es in einer Folgestudie gehirnweite neuronale Aktivitätsmuster in lebenden Fliegen und Zebrafischlarven während visueller Stimulationen aufzuzeichnen.

Anschließend optimierte ich das Split-HaloTag System für die Markierung interagierender synaptischer Adhäsionsproteine, um eine selektive Methode zur Färbung neuronaler Verbindungen zu entwickeln. Ich verwendete computergestütztes Proteindesign, um die intrinsische Affinität der Split-HaloTag Fragmente anzupassen und um die Stabilität des großen Fragments zu erhöhen. Der verbesserte Split-HaloTag markiert erfolgreich synaptische Adhäsionskomplexe an Zell-Zell Kontakten und legt damit den Grundstein für einen hochselektiven Ansatz zur Kartierung von neuronaler Konnektivität. Darüber hinaus könnte die Verwendung von HaloTag-Liganden mit unterschiedlichen Farben die Identifizierung von neu gebildeten Synapsen im gesamten Gehirn ermöglichen.

Die hervorragenden Eigenschaften sowie die Vielseitigkeit des verbesserten Split-HaloTag Systems haben die Entwicklung mehrerer anderer molekularer Rekorder und Markierungsstrategien in unserer Gruppe unterstützt. Die Technologie bietet somit einen vielversprechenden Weg zur Erforschung diverser biologischer Aktivitäten und Funktionen in verschiedenen Modellsystemen.

Acknowledgements

I would like to express my gratitude to Kai Johnsson, my supervising professor, for the invaluable guidance and support provided throughout my doctoral studies. His dedication and enthusiasm for science have consistently been a source of motivation. I highly appreciate the considerable freedom I was granted, allowing me to explore new methodologies. His trust and openness not only enriched my research but also significantly contributed to my academic and personal development as an independent scientist.

I would like to extend my gratitude to the members of my examination commission, Britta Brügger, Claudio Acuna Goycolea and Jochen Reinstein, for their review and assessment of this thesis and their helpful advice during the annual thesis advisory committee meetings or personal discussions.

Furthermore, I want to thank Jochen Reinstein for generously sharing his equipment and expertise on enzyme kinetics, both of which have been invaluable to my research. I also would like to thank Mirek Tarnawski for solving the crucial crystal structures that formed the foundation for my protein engineering efforts.

While our collaboration is still in its early stages, I would like to acknowledge Peter Scheiffele for his insightful input regarding the application of the synapse detection technology, and I eagerly anticipate where our continued collaboration will lead.

I want to express my gratitude to Julien Hiblot, who has been an exceptional mentor during my doctoral journey. I am grateful for all he has taught me, both inside and outside the lab. Our numerous discussions have been pivotal in shaping my work and refining my scientific perspective. His strong scientific curiosity has always been a motivating force for me.

I would like to thank Magnus Huppertz, with whom I worked in close collaboration on the Caprola project. His dedication and commitment elevated the entire project giving it the potential to be recognized as a landmark study. Also many thanks to Vincent Grenier and Nicola Porzberg for their significant contributions to the project.

I want to express my gratitude to all members of the Chemical Biology Department for creating such a positive working environment and for their valuable input during meetings and informal discussions. I would also like to express a special thanks to Magnus Huppertz, Stefanie Kühn, Yin-Hsi Lin, Lars Hellweg and Konstantin Hinnah for their collaboration, the numerous fruitful discussions we have shared, and for being awesome labmates.

One of the most rewarding aspects of my PhD journey has been the opportunity to supervise and work in collaboration with several students. I want to thank Karoline Schulzen, Jana Tünnermann, Antoni Gralak, Lennart Nickel, Leonhard Kohleick and Rosa Boehler for their significant contributions to my work.

My fascination for chemical biology was sparked early in my studies, and I want to thank Richard Wombacher for being an excellent lecturer and mentor who introduced me to this innovative field.

I want to acknowledge Birgit Koch for her assistance in all cell culture related matters, and for keeping our microscopy and FACS equipment running. Furthermore, I want to thank Andrea Bergner, Bettina Réssy, Carmen Grosskurth, Dominik Schmidt and Jana Kress for providing reagents and technical assistance. Many thanks also to Sebastian Fabritz, Juliana Kling and Tatjana Rudi from the mass spectrometry facility for their support, and to Gwenaëlle Matthies for her assistance in dealing with administrative and organizational matters.

I want to thank Chris Roome and the Biomolecular Mechanisms Department for providing computational resources, as well as for the outstanding assistance and support.

A big thank you to Magnus Huppertz, Julien Hiblot and Sara Becker, who took the time to review this thesis and provided valuable comments and suggestions.

Besides the colleagues and collaborators I worked with, I also would like to thank my friends. Spending time together, whether it was through meetups, vacations, or regular Stammtisch gatherings, provided the necessary balance I needed during my doctoral studies.

Finally, I want to to express my deep gratitude to my parents Andrea Keck-Wilhelm and Ulf Wilhelm. Their constant belief in me, their fostering of my curiosity, and their enduring support have played a pivotal role in shaping me into the researcher and person I am today. I also want to extend my gratitude to my partner Sara Becker. I am truly grateful to have you by my side, and I eagerly look forward to our next chapter together.

Contents

List of Figures

List of Tables

1 | Introduction

1.1 Methods for Recording Neuronal Activity and Connectivity

The brain is the most complex organ of the body. Understanding its structure and how it manages intricate cognitive functions is the primary goal of neuroscience[1]. The anatomy of the brain can be described as a complex communication network, in which neurons represent the nodes and synapses, the functional contact sites between neurons, are the edges. Understanding this intricate network requires structural and functional analysis across different scales, spanning from molecular mechanisms, through cellular processes and local neuronal circuits up to the organ as a whole[1]. This demands innovative methodologies from various disciplines, including genetics and molecular biology, chemistry and pharmacology, physics and optics as well as mathematical modeling and informatics[2]. A selection of influential methods, that facilitate studying the nervous system across different spatial and temporal scales are discussed in this section.

1.1.1 Recording Neuronal Activity in Vivo

1.1.1.1 Real-Time Observation

The first recordings of neuronal activity were made in the nineteenth century using what was later termed electroencephalography (EEG) by directly placing electrodes on the head and spine of animals[3] and later also humans[4]. The method features excellent temporal resolution and is non-invasive, however its spatial resolution is very limited and restricted to larger brain regions at the surface. Nevertheless, the method is still routinely used for diagnostic purposes. The later developed magnetoencephalography (MEG) which detects changes in magnetic field originating from neuronal signal transduction features slightly higher spatial resolution, but is still limited to capturing activity from cortical regions at several millimeter resolution[5].

Other non-invasive methods to trace neuronal activity include functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). These techniques detect the metabolic activity of neurons, such as oxygen consumption or glucose uptake, which increases upon neuronal activity. Their temporal resolution is typically around seconds to minutes. While some fMRI studies are approaching sub-millimeter resolution[6], they do not achieve single-neuron precision and offer only an indirect measure of neuronal activity. Nevertheless, beyond their clinical applications, these methods have revealed the functional connectivity among various brain regions and facilitated the recording of distinct neuronal activity patterns[7,8].

The first technique developed to directly observe action potentials of single neurons was with the patch-clamp method. In this approach a micropipette with an electrode is attached to a neuron which allows the direct measurement of ion currents and potentials across the cell membrane with excellent temporal resolution[9]. This technique has been successfully utilized in living animals, significantly contributing our understanding of neural circuits in various brain regions, such as the sensory cortex[10]. However, despite recent promising advancements in the approach using nanotechnology[11], major limitations include the low number of neurons that can be investigated simultaneously, its invasiveness and its incompatibility with freely moving animals.

Fluorescence microscopy, in combination with probes sensitive to neuronal activity are widely used to monitor in vivo neuronal activity in model organisms. The most direct methods to visualize action potentials by fluorescence microscopy is the use of fluorescent voltage indicators. These can be synthetic, voltage-sensitive dyes (VSDs), which integrate in the cellular membrane and undergo changes in their fluorescent properties in response to variations in membrane potential[12]. However, VSDs suffer from limitations such as their lack of cell-type specificity and a high hydrophobicity, requiring invasive staining procedures[13,14]. To address these challenges, genetically encoded voltage indicators (GEVIs) have been developed[15]. These are usually based on voltage-sensing membrane proteins that undergo conformational changes in response to fluctuations in membrane potential. These proteins can be linked to fluorescent proteins in such a way that the conformational shift affects the fluorescent properties[14]. Alternatively, opsin-based GEVIs report on voltage changes through their associated cofactor, a retinal chromophore, which can be amplified by linked FRET donors[16,17]. While voltage imaging allows to directly observe action potentials, hyperpolarizations and depolarizations, its application remains technically challenging as very high frame rates are needed to capture these fast events (< 5 ms)[13,14]. However, advancements in both GEVI performance and optical setups have recently enabled in vivo voltage imaging of larger neuronal populations[18].

Another possibility to visualize neuronal activity are calcium-sensitive fluorophores. They monitor the action potential-induced influx of calcium ions (Ca^{2+}) into the neuron via voltage gated calcium channels[1,19]. The Ca^{2+} response to neuronal firing is longer than the membrane depolarization, yet in a second timescale[20]. On one hand this restricts the temporal resolution of Ca^{2+}-based activity measurements, but on the other hand it leads to some integration of the signal which facilitates detection via fluorescence microscopy. One type of Ca^{2+}-sentive fluorophores are synthetic organic dyes that change their fluorescent properties when binding to Ca^{2+} ions[21]. These dyes feature excellent photo-physical properties and response kinetics and have been used for functional studies in various model organisms and brain regions[22]. However, they share some of the drawbacks of voltage-sensitive dyes discussed above in regard to loading procedures and specificity of labeling[23]. Several strategies that allow targeted localization of synthetic Ca^{2+} dyes have been described[24–27], for example using self-labeling proteins (which are discussed in section 1.2), but in vivo applicability of these approaches remains to be demonstrated.

As an alternative, genetically encoded calcium indicators (GECIs) have been developed[28] which can be targeted to specific neuron populations or subcellular localizations using genetic methods. They are based on fluorescent proteins which are modulated by Ca^{2+}-sensing domains like calmodulin and have facilitated systematic monitoring of neuronal circuits in different animals and areas of the brain[23,29]. This has not only been enabled by meticulous engineering of the photophysical properties and response kinetics of GECIs[23,30], but also by the development of improved imaging technologies. Especially the development of two-photon laser scanning microscopy[31] has revolutionized in vivo Ca^{2+} imaging[32]. This method uses two near-infrared or infrared photons which cooperatively excite a fluorophore. The long wavelength of the excitation light reduces scattering and absorption in tissues. In combination with the non-linearity of the two-photon excitation process, this enables hundreds of micrometers deep imaging in the brain while resolving small subcellular structures like dendritic spines[32]. The technology has been further refined by usage of adaptive optics and extension of the concept to three-photon microscopy, reaching recording depths of 1.4 mm with excellent resolution[33].

Furthermore, genetically encoded sensors have been developed for real-time imaging of neuronal activation at the synapse level. Localizing fluorescent protein-based pH sensors within the lumen of neurotransmitter vesicles has enabled to study synaptic vesicle release and recycling triggered by neuronal signal transduction[34,35]. Fluorescent neurotransmitter sensors for glutamate[36], acetylcholine[37] and dopamine[38,39] can be used for the live-observation of neurotransmitter release in various organisms. While these reporters provide a more detailed view on neuronal input and output, they are less universal than methods based on voltage or Ca^{2+}.

Despite the considerable advancements in fluorescence microscopy for live observation of neuronal activity in vivo, significant limitations remain. One major challenge is the imaging depth and field of view that can be achieved. Although significant improvements have been made in this regard, the observation of activity across larger brain circuits or entire brains of rodents remains out of reach[40]. Current techniques are relatively invasive and usually require partial removal of the skull and the implantation of cranial windows in mammals. Additionally, issues like phototoxicity, bleaching of fluorophores and thermal injury can become problematic during longer recordings[32]. Furthermore, these methods require the animal to be fixed to the microscope, which significantly restricts which behavioral experiments can be conducted[41].

1.1.1.2 Post-Hoc Readout

To overcome the limitations of real-time neuronal activity monitoring, post-hoc methods have been developed that capture the history of neuronal activity during a specific timeframe. Unlike real-time observations, these methods are not inherently limited in scale or number of neurons that can be investigated in parallel, while still maintaining single-neuron resolution. By permanently tagging active neurons, these methods are compatible with fixation and downstream methods such as clearing, sectioning and large-scale imaging[42–44]. However, post-hoc methods compromise on temporal resolution and provide only limited information on when exactly an activity occurred.

Post-hoc readout of neuronal activity has been achieved by utilizing immediate early genes (IEGs) such as *c-Fos* and *Arc*. Expression levels of these genes are generally very low, but substantially increases upon neural activity[45]. The expression of IEGs can be directly visualized at the RNA level using fluorescence in situ hybridization typically 30 min after the activity, or at the protein level via immunofluorescence roughly 1–2 h post activity[46]. Alternatively, IEG promotors can drive the expression of a reporter gene like lacZ[47] or fluorescent proteins[48] in transgenic mice. This strategy was used to detect neural activity on a brain-wide scale in mice, using automated imaging for the first time[49,50]. The approach has been further refined by constructing genetic circuits driven by IEG promotors, combined with drug-dependent transcription factors such as tTA,[51] or recombinases like CreER[52,53]. This allows for a more precise control over the recording window, started by administration of the drug, and enables stable tagging of active neurons.

Notably, the use of growing protein filaments linked to reporter proteins or epitope tags has recently enabled continuous recording of IEG expression in cultured neurons[44,54]. However, the applicability of this method for recording neuronal activities in vivo remains to be validated. A major drawback of using IEG expression is that they are only indirectly linked to neuronal activity. Their expression is not an unbiased proxy for the number of action potentials but also dependends on distinct activity patterns[55,56]. Moreover, since IEG approaches are based on gene expression, their temporal resolution is limited to the timescale of transcription and translation processes, spanning hours to days[41].

To combat the limitations of IEGs, optogenetic approaches have been developed that use Ca^{2+} as a more direct indicator for neuronal activity to drive reporter gene expression[57,58]. These methods, termed Cal-light and FLARE, as well as their improved versions[59–61], use a transcription factor tethered to the plasma membrane which is released only upon simultaneous exposure to Ca^{2+} and blue light. This is achieved by recruiting or activating a (split-)TEV-protease through the interaction of calmodulin and M13 upon Ca^{2+} binding. Additionally, the recognition sequence of the protease is caged using a blue light-sensitive light-oxygen-voltage (LOV) domain. Upon cleavage of the TEV recognition sequence, the transcription factor is released and induces expression of a reporter gene. These approaches have not only enabled marking neuronal activity in the motor cortex but also facilitated the selective manipulation of active neuron populations by expression of opsins, which allow subsequent light control of neuronal activity.[57,58]. While these approaches do use Ca^{2+} as a trigger, the linkage between Ca^{2+} transients and the output signal is still indirect via gene expression. Moreover, recorded signals are highly dependent on the expression levels of each component[62] making these systems difficult to apply on large scale. Despite the more precise control of the recording window using blue light, these tools still require at least 30 min tagging time in vivo[58,62].

To achieve a direct link between Ca^{2+} signals and fluorescent readout, a Ca^{2+}-modulated photoactivatable ratiometric integrator (CaMPARI) has been designed[63]. This tool is based on a circularly permuted photoconvertible fluorescent protein (mEOS2) fused to the Ca^{2+}-sensitive domains calmodulin and M13. When exposed to Ca^{2+} and UV light at 405 nm, the protein undergoes an irreversible conversion from a green to an orange chromophore. By measuring the ratio of green to orange fluorescence the extend of Ca^{2+} exposure during UV illumination can be delineated. This has been demonstrated by brain-wide recordings of neuronal activity in zebrafish during defined time windows. A refined version (CaMPARI2) which offers improved brightness and contrast[64], has recently been used for cortex-wide recording of neuronal activity in freely-moving mice at cellular resolution[65]. Furthermore, a reversibly switchable version (rsCaMPARI)[66] has been introduced that allows recording of calcium signals under blue light illumination and subsequent erasing of the fluorescence shift using violet light. This enables recording of different activity patterns in response to different stimuli in a single individual[66]. Despite its potential, CaMPARI has notable limitations. It requires illumination of the specimen with short-wavelength light, which has poor tissue penetration, reaching only sub-millimeter depths in a mouse brain[65]. There is also significant background photoconversion, in absence of Ca^{2+}, reducing signal-to-noise ratios[62]. Behavioral experiments present challenges, as many model animals such as zebrafish[67] and mice[68,69] are highly sensitive to UV light, leading to potential disturbances or stress. Furthermore, as it is the case for all optogenetic methods, the approach requires invasive interventions like cranial window surgery or the implantation of fiber optics.

An ideal method for post-hoc readout of neuronal activity would rely directly on the immediate Ca^{2+} or voltage response, be non-invasive and allow continuous, or at least multiple, recordings in a single animal with good temporal resolution. Additionally, the technology should work on a brain-wide scale in different model systems, while maintaining sub-cellular precision. Such method would be transformative for neuroscience, as it would allow to study neural activity in its entirety, unravel neuronal circuits and provide insights into how individual neurons contribute to complex brain functions and behaviors.

1.1.2 High Resolution Mapping of Neuronal Connectivity

1.1.2.1 Microscopy Techniques

To understand the neuronal network in the brain, it is necessary to investigate the wiring of neurons at the resolution of single synapses. However, the small size of synapses (300–500 nm)[70] and the even smaller size of the synaptic cleft (~20 nm)[71], makes localizing them in three-dimensional space challenging. In addition, synapses are highly abundant in many brain areas with up to one billion synapses per cubic millimeter in the cortex[70]. Electron microscopy (EM) has historically been the primary tool for mapping neuronal connectivity due to its ability to reach low nanometer resolution, while classical light microscopy techniques fails to achieve this level of precision. Several advancements in regard to sample preparation, sectioning, data acquisition and analysis have significantly improved the technology. Techniques such as serial block-face scanning EM, which enables three-dimensional reconstruction by sequentially scanning and slicing the sample[72] and focused ion beam scanning EM, which uses an ion beam to mill tissue layers for high-resolution 3D images[73], have facilitated volumetric EM imaging. Similarly, serial-section transmission EM provides detailed 3D images by passing an electron beam through tissue slices[74]. In combination with advanced reconstruction methods, these techniques have spurred large-scale connectomic analyses[75].

Nevertheless, comprehensive mapping of an entire *Drosophila* brain or a cubic millimeter of a rodent's brain still requires the collective efforts of large consortia over several years[76,77]. While EM offers unparalleled resolution, the obtained data suffers from a lack of contrast between different cell types or subcellular structures like synapses, which complicates manual and automatic reconstructions. Consequently, EM also cannot detect the molecular composition of synapses required to reveal the type and strength of synapses[78]. Additionally, there are significant inter-individual connectomic variabilities[79,80], which are challenging to investigate via EM due to the limited throughput. Moreover, the neuronal network in the brain is not static but synaptic connections can changes within hours[81,82]. These dynamics of synaptic plasticity cannot be captured by EM imaging.

Unlike EM, fluorescence microscopy enables detection of signals in multiple channels, which in combination with specific staining techniques allows to distinguish different cell types, sub-cellular structures and molecular compositions. However, the resolution of classical fluorescence microscopy is limited to approximately 250 nm when using high numerical aperture setups with green light, and even lower for longer wavelengths[83]. Hence, fluorescence microscopy cannot effectively resolve small structures like single synapses in dense tissues. Several techniques have been developed that overcome this resolution limit using optical methods, for example by depletion of fluorophores outside the focal volume (stimulated emission depletion, STED)[84], stochastic activation of fluorophores (photo-activated localization microscopy, PALM or stochastic optical reconstruction Microscopy, STORM)[85–87] or via transient binding of fluorophores (points accumulation for imaging in nanoscale topography, PAINT)[88,89]. In particular STED microscopy in combination with labeling of extracellular space, which results in a "shadow image" of all cells, has been shown to facilitate large reconstructions of live or fixed brain tissues at nanometer scales.[90,91].

A complementary approach uses expansion of the specimen[92] to achieve an (apparent) increase in resolution[93]. Therefor, a specimen is embedded in a hydrogel (e.g. a sodium acrylate, acrylamide, copolymer) while proteins are anchored to the matrix via crosslinkers. After homogenization to disrupt complexes, the hydrogel is incubated in water or low salt buffers, which causes isotropic swelling. Expansion microscopy (ExM) is particularly suited for connectomic analysis of larger tissues as it combines the fast scanning speed and throughput of diffraction limited microscopes with the high resolution needed to resolve single synapses. Combining ExM with lattice light-sheet microscopy (ExLLSM) further increases throughput of the method and has enabled mapping of presynaptic sites an the entire *Drosophila* brain, and large volumes in the primary somatosensory cortex of mice[42]. Imaging entire mouse brains is still challenging due to the limited working distances of available objectives. However, as shown in other approaches, sectioning the brain while imaging could circumvent these problems[94]. Furthermore, expansion microscopy can achieve a high level of multiplexability as fluorescent probes can be washed in and out of expanded tissues which enabled co-imaging of several synaptic marker revealing the molecular compositions of synapses[95]. Recently, 45-channel imaging of mouse brain tissue in three staining cycles has been demonstrated using spectral imaging in combination with unmixing algorithms[96].

1.1.2.2 Synapse Labeling Methods

Specific labeling of synapses is needed to enable efficient mapping of their localization in connectomic studies. This can be achieved by labeling of synaptic scaffolding proteins, transporters or receptors. Presynaptic markers include Synapsin-1 & 2, which coat synaptic vesicles[96,97], Bassoon, which regulates the spatial organization of synaptic vesicles[42,98], vGluT1 & 2, the synaptic vesicle glutamate transporters in excitatory synapses[99] and vGAT, the synaptic vesicle GABA transporter in inhibitory synapses[100]. Post-synaptic markers include post-synaptic density protein 95 (PSD-95), which organizes the clustering of synaptic proteins at excitatory synapses[101], Homer1, which binds and cross-links cytoplasmic regions[100,102] and Gephyrin, which is involved in membrane-cytoskeleton interactions at inhibitory synapses[100,103]. These proteins can either be labeled via fluorescent antibody staining or genetically tagged with epitope-tags, fluorescent proteins or self-labeling protein tags[78,90,104] (which are discussed in section 1.2). The reliable identification of synapses usually requires the co-localization of pre- and post-synaptic markers in three-dimensional space.

To unambiguously detect synapses, dual-component systems have been developed that are dependent on pre- and post-synaptic components. In addition to providing more specific signals that are directly located at the synaptic cleft, these methods could reduce the required resolution for connectomic analysis. Synapses can be identified by a single marker instead of the precise detection of adjacent pre- and post-synaptic terminals. The first dual-component synapse marker, GFP reconstitution across synaptic partners (GRASP), uses the reconstitution of split-GFP fragments which are non-fluorescent when isolated, but regain fluorescence after binding to each other. Split-GFP fragments were fused to the synaptic adhesion protein neuroligin-1 expressed in neighboring neurons to identify synaptic contacts[105]. However, the underlying mechanism of the method remains elusive, as neuroligins are localized to post-synaptic terminals and not to presynaptic sites. Furthermore, they do not interact across the synaptic cleft[106]. In another GRASP variant split-GFP parts are fused to truncated CD4 receptors as an inert tether protein, which are also not known to feature synaptic localization or form intercellular complexes. Nevertheless, GRASP has been successfully applied in both *C. elegans* and *Drosophila* in several studies[107–111]. The system has been modified to shift the wavelength of the reporter to enable multi-color experiments[112–114] and the performance of GRASP in *Drosophila* has been further improved by screening for optimal targeting sequences[115]. Another version of the tool (Syb:GRASP) enables labeling of active synapses. By sequestering the pre-synaptic construct to synaptic vesicles via fusion to Synaptobrevin, exposure of split-GFP to the synaptic cleft happens only after neurotransmitter release[112].

For application in mammalian brains the technology has been modified using neurexin and neuroligin proteins[116], which in their native form establish a trans-synaptic adhesion complex that induces synapse formation and maturation[117]. The pre-synaptic mGRASP component is a fusion of split-GFP, a truncated human CD4 receptor and the intracellular domain of rat neurexin-1β, while the post-synaptic counterpart consists in split-GFP fused to a truncated mouse neuroligin-1 that lacks its catalytically inactive esterase domain. The truncation of neuroligin-1, and replacement of the N-terminal neurexin-1β with a CD4 fragment, should render the constructs inert to ensure that they do not interfere with endogenous adhesion and signaling proteins. Although the cytoplasmic sequences of neurexin and neuroligin are not known to drive trafficking to synapses and are dispensable for synaptic localization[118–120], the mGRASP system has been successfully applied in different studies[116,121,122]. An alternative mGRASP approach uses unmodified full-length neurexin and neuroligin and has enabled labeling of synapses in transgenic mice using a Cre-dependent strategy to drive cell-type specific expression of the constructs[123]. Later, an enhanced version of the original mGRASP (eGRASP)[124] has been developed that improves split-GFP complementation by introducing two weakly interacting protein domains to the pre- and post-synaptic components. Moreover, the creation of cyan and yellow variants enabled visualization of two distinct synaptic populations, originating from two different presynaptic neurons[124].

A drawback of the GRASP-based methods is that split-GFP complementation is generally independent of fusion partner interactions that would bring the split fragments into close proximity. Instead, complementation is induced by mere co-localization on proximal membranes[105,106,116,125], which might occur at non-synaptic sites, reducing the specificity of these tools. To address this concern and ensure that split reconstitution happens specifically at genuine synapses the SynView[106] approach was introduced. By inserting the split-GFP fragments into splice loops or inter-domain linkers of neurexin-1β and neuroligin-1, reconstitution became dependent on their trans-synaptic interaction, leading to the first direct demonstration of their association across the synaptic cleft[106]. However, in vivo applicability of the tool still remains to be demonstrated.

Another approach to label synapses, termed ID-PRIME[126], uses a modified lipoic acid ligase fused to neurexin to label a ligase acceptor peptide linked to neuroligin, by ligation of lipoic acid, or an azide-bearing analog of lipoic acid. In a second step, the modification can be revealed either by anti-lipoic acid antibodies or azide-reactive organic dyes. ID-PRIME has been shown to be dependent on neurexin–neuroligin complex formation and does not label when mutations are introduced that interfere with their interaction. However, the method has not yet been demonstrated to label synapses in vivo. Its effectiveness might be limited by the low solubility and high hydrophobicity of the lipoic acid ligase substrate, which might not distribute uniformly in complex tissues.

In a similar technology, named BLINC[126], a biotin ligase and a cognate acceptor peptide are used. When fused to neurexin and neuroligin, the acceptor peptide gets biotinylated at synaptic sites, allowing for subsequent detection by streptavidin-fluorophore conjugates. Despite being described as less robust and giving lower signals than ID-PRIME, a tailored version of the tool (iBLINC)[127] has been successfully applied to detect synapses in *C. elegans*.

Another enzyme-based method uses a split horseradish peroxidase (sHRP)[125] fused to neuroligin and neurexin to label synapses. sHRP is also dependent on protein-protein interactions and the reconstituted protein can be labeled in a single step procedure with a wide variety of commercial tyramide substrates. A conceptual advantage is the signal amplification that is achieved since a single reconstituted sHRP can catalyze the attachment of many fluorophore substrates to nearby proteins. sHRP has enabled labeling of synapses in the visual system of mice, however, it faces some limitations regarding in vivo use. These limitations include that sHRP needs heme as a cofactor, which must be delivered by perfusion, and the background labeling of tissues due to endogenous peroxidases[125].

While split-GFP and split-HRP enable effective labeling of synapses in mouse brains, a potential concern is the irreversible nature of the complementation of these split proteins[125,128,129]. This irreversibility can stabilize the adhesion between pre- and post-synaptic membranes, potentially interfering with synaptic pruning. As a consequence, synapse size and number could be artificially increased, introducing bias into results obtained with these approaches. Another significant limitation of these systems is that their inability to study neuronal network dynamics as they can only provide a static image of connectivity. While in theory, the split-HRP system could label synapses at different time points with distinguishable fluorescent substrates, its dependence on the cofactor heme does not permit such experiments in vivo[125].

An ideal synapse labeling technique would rely on the specificity of trans-synaptic protein-protein interactions but should minimally perturb binding and unbinding of these proteins. This method should be compatible with state of the art imaging technologies such as STED and ExM and have broad applicability in several model organisms. To precisely investigate defined neuronal populations, the approach should be genetically targetable. Furthermore, the ideal approach should allow to obtain data on neuronal network dynamics, for example by indicating the age of synapses through different colors.

1.2 Self-Labeling Proteins

Various methods have been developed that permit specific fluorescent staining of distinct protein species in biological samples and tissues[130]. Although developed already in the mid-twentieth century, immunofluorescence is still a widely applied technology today that uses antibodies conjugated with organic fluorophores to specifically stain their targets. While this method allows for the detection of endogenous proteins without necessitating genetic modifications, it requires that samples are fixed and permeabilized for imaging of intracellular targets. The groundbreaking discovery of the green fluorescent protein (GFP) from *Aequorea victoria*[131,132], introduced entirely new possibilities, including the live observation of labeled proteins in living organisms. This is achieved by genetically fusing a fluorescent protein to a protein of interest. Since the discovery of GFP, numerous fluorescent protein variants with enhanced brightness and stability have been discovered or engineered[133]. Fluorescent proteins with excitation and emission wavelengths spanning the visible and near-infrared spectrum have been developed, greatly increasing versatility, and enabling the simultaneous observation of multiple proteins of interest[133]. Moreover, fluorescent proteins allowed to construct genetically encoded sensors, facilitating real-time monitoring of messenger molecules or metabolites within living cells[134]. However, fluorescent proteins do have certain limitations. Compared to synthetic organic dyes like rhodamines or cyanines, they are usually less bright and have lower photostability. This difference is particularly pronounced in the red and far-red spectrum[135]. Hence, especially in demanding imaging applications with a tight photon budget, such as super-resolution microscopy or deep tissue imaging, synthetic dyes offer advantages over fluorescent proteins[136].

Approaches that combine the genetic targetability and live-cell compatibility of fluorescent proteins with the superior photophysical properties of synthetic dyes have been developed[137]. Self-labeling proteins (SLPs) offer a convenient method to modify fusion proteins in living cells[138,139]. These proteins react specifically with cell-permeable small molecule substrates linked to synthetic fluorophores, forming an irreversible covalent bond in a bio-orthogonal manner. SNAP-tag[138], the first described SLP, is derived from the human O^6-alkylguanine-DNA alkyltransferase (hAGT). This protein is part of the DNA repair machinery and restores alkylated DNA by irreversibly transferring the alkyl moiety to its reactive cysteine residue[140]. SNAP-tag covalently reacts with fluorescent benzylguanine substrates, leading to attachment of the benzyl group bearing the fluorophore, and subsequent release of guanine. In addition, SNAP-tag reacts with chloropyrimidine substrates which are more cell-permeable than benzylguanines[141]. An orthogonal system, CLIP-tag[142], was engineered from SNAP-tag that accepts benzylcytosine substrates, which enabled two-color imaging using SLP technology.

HaloTag[139,143], an alternative SLP, is derived from the *Rhodococcus* dehalogenase DhaA, which catalyses the hydrolization of halogenated alkanes via a covalent intermediate ester, involving a aspartate residue in the active site[144]. The second catalytic step of the reaction mechanism was abolished by introducing point mutations, which lead to trapping of the covalent intermediate. HaloTag can be labeled with synthetic fluorophores linked to a PEG-chloroalkane substrate.

SLPs have been proven to be highly versatile and compatible with a large variety of cell-permeable fluorophore ligands with absorption and emission wavelengths up to the near-infrared spectrum[145–151], and are valuable tools for live-cell super-resolution microscopy[136]. Furthermore, in particular HaloTag technology has been proven to be compatible with in vivo experiments and can be labeled in the brain of living mice after substrate injection[17,147,149,152]. At their core, SLPs serve as a bridge between biology and synthetic chemistry and facilitate the introduction of diverse chemical functionalities into biological systems. In addition to synthetic fluorophores, SLPs have been conjugated to an array of other molecules including affinity handles[153], crosslinkers[154] and drugs[155]. Due to this high flexibility, SLPs have found numerous applications beyond standard fluorescent labeling which include protein isolation[156], molecular interaction assays[157], targeted protein degradation[158], mapping of protein lifetimes[152,159], tethered pharmacology approaches[160] and the design of various semi-synthetic biosensors[17,161–163].

1.3 Protein Complementation Assays

Most proteins within a cell or organism do not function in isolation but interact with a variety of other proteins as part of intricate interaction networks[164]. These protein-protein interactions (PPIs) play a central role in virtually all biological processes including cellular signaling, metabolic pathways, cell adhesion and transport[164,165]. Thus, identification and characterization of PPIs is of large interest. One method to study PPIs in their native cellular environment is via protein fragment complementation assays (PCAs)[166–168]. PCAs are based on reporter proteins, which are split into two fragments, that are inactive individually, but can complement each other when brought into close proximity. Complementation leads to reconstitution of their activity, yielding a detectable signal. To investigate a putative PPI, both fragments of the split reporter protein are genetically fused to the target proteins often denoted as "bait" and "prey". These fusions can either be overexpressed using plasmids or viral vectors, or the split fragments can be fused to their genomic loci via genome editing tools, to study them at near endogenous levels[166,168].

Among the first PCAs developed is the split yeast transcription factor GAL4, which upon reconstitution activates expression of a reporter gene such as the β-galactosidase lacZ[169]. This method, termed yeast two-hybrid, is compatible with high-throughput assays and has aided the discovery of many PPIs. Moreover, the approach facilitated the mapping of PPIs on proteome-wide scales for several organisms including yeast[170], *Drosophila*[171] and humans[172]. However, the yeast two-hybrid system does not allow to study temporal dynamics of protein-protein interactions. To overcome this limitation a PCA based on a split-ubiquitin linked to a reporter protein was developed. The construct is cleaved by ubiquitin-specific proteases upon split ubiquitin assembly, and the cleavage can be detected by immunoblotting[173]. Other PCAs were developed that involve enzymes vital for cell growth or survival under specific conditions, like split murine dihydrofolate reductase (mDHFR)[174] and split yeast N-(5-phosphoribosyl)-anthranilate isomerase (Trp1p)[175]. Especially the split-mDHFR approach has enabled large-scale assays, that enabled probing the complete interactome of *S. cerevisiae*[176].

To facilitate real-time online measurements of PPIs, PCAs featuring a direct optical readout have been developed. These include split-GFP[177] and numerous other split fluorescent proteins[129], which generate a fluorescent signal upon complementation. This signal can be detected in bulk or used for direct visualization of interactions using live cell fluorescence microscopy[168]. While split fluorescent protein-based PCAs are widely used[129], a significant drawback is the irreversible reconstitution of the split-GFP fragments[125,128,129]. Although this can be advantageous for capturing transient and short-lived PPI events due to the integration of the signal, it also interferes with the dynamics of studied protein complexes, introducing unwanted bias by forcing interactions. The same is the case for a split horseradish peroxidase (HRP), which can be used in combination with fluorescent tyramide substrates to detect PPIs. A reversible PCA system that generates a fluorescent readout has been derived from a fluorescence activating protein named FAST. This protein binds to an exogenous fluorogenic ligand, which becomes fluorescent upon binding. Split-FAST[178] fragments have been demonstrated to dissociate when the probed PPI is no longer present. However, the substantial affinity of the split fragments ($K_D \approx 1\,\mu\text{M}$) might still interfere with the behavior of the investigated proteins.

Instead of a fluorescent readout, split luciferase-based PCAs provide a bioluminescent signal upon complementation. These PCAs are particularly well suited for large-scale studies, as bioluminescence features exceptional sensitivity and can be conveniently measured in microtiter plates[168]. However, they are not optimal for live observations using microscopy. Split versions of several luciferases have been engineered[179–181]. Among these, split-NanoLuc[182] stands out, as this system is highly reversible with a K_D of 190 µM while providing robust signals. Additionally, the system features high versatility as different versions are available that enable tuning the K_D down to 700 pM to fit for different applications.

In terms of split protein design, there are two overarching concepts. The first one involves splitting a protein into two relatively large parts, while the second one excises only a small peptide fragment. the former strategy is prone to generate two mostly unfolded fragments, due to the exposure of hydrophobic core residues[183,184]. These fragments can then re-fold once brought into close proximity, which might be the mechanism of many split systems[167,185]. However, this can lead to several drawbacks. Protein refolding can be a relatively slow and largely irreversible process, which hinders observation of dynamic PPIs and can introduce bias by forcing interactions. Furthermore, attaching unfolded fragments to proteins of interest can significantly destabilize the fusion proteins[186,187]. In contrast, the strategy of removing only a small fragment, which might preserve the overall protein fold, can potentially mitigate these issues. Supporting this hypothesis is the observation that many of the reversible split systems, such as split-NanoLuc[182], split-FAST[178] and split-mDHFR[176] adhere to this design strategy.

2 | Objectives, Motivation & Scope

2.1 Objectives and Motivation

The objective of the present thesis was the engineering and application of a protein-fragment complementation assay (PCA) based on the self-labeling protein HaloTag. The assay is based on splitting the HaloTag protein into two fragments, both inactive on their own, which can bind and complement each other in a reversible manner to restore self-labeling activity. This PCA is unique, as it combines reversible complementation with a subsequent in situ permanent modification and a fluorescent readout. By fusing the split-HaloTag fragments to sensing domains, such a system can act as a "molecular recorder" of cellular history (fig. 1). Upon a temporary rise in analyte concentration, e.g. due to a physiological signaling event, split-HaloTag is activated and labeled. This process transforms the transient presence of the analyte into a permanent fluorescent mark. The timeframe of such recording is defined by the presence of a fluorescent HaloTag ligand. The extent of labeling is then read out at a later time point, decoupling recording and measurement. This permits massive parallel analyses of transient cellular events in large cell populations. In addition, the approach is compatible with further downstream analysis, e.g. using (spatial-)omics technologies to correlate signals to specific phenotypes.

I aimed to develop a split-HaloTag-based molecular Ca^{2+} signal recorder, which enables the simultaneous recording of activity of millions of neurons in a freely moving animal. After the experiment, the accumulated signals can be readout with cellular resolution on a brain-wide scale (fig 2). Moreover, utilizing spectrally distinct HaloTag ligands sequentially facilitates recording multiple periods of neurological activities on this scale. While compromising on temporal resolution, such large-scale datasets could complement insights gained from live observation of smaller groups of neurons.

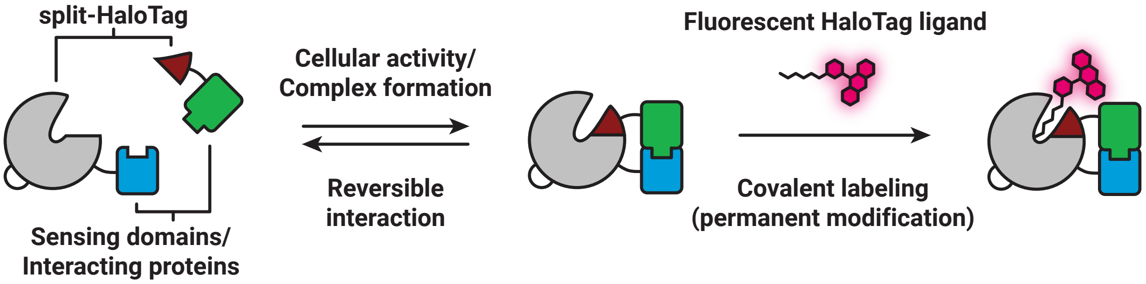


Figure 1: Concept of HaloTag-based protein fragment complementation assay. A split-HaloTag is fused to sensing domains or interacting proteins. During cellular activity (e.g. a Ca^{2+} signaling event or formation of a protein-protein interaction), split-HaloTag fragments complement in a reversible manner. Only then split-HaloTag can be labeled with fluorescent ligands, leading to a permanent trace of the transient cellular activity.

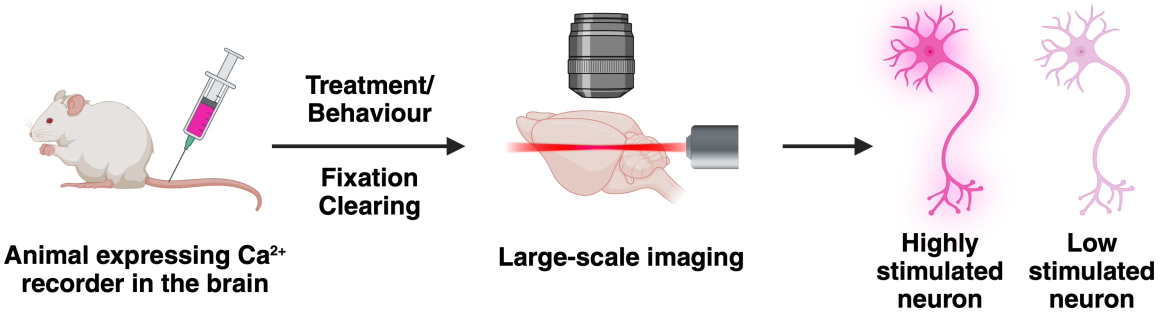


Figure 2: Concept of large-scale recording of neuronal activity in vivo. An animal expressing the split-HaloTag-based Ca^{2+} recorder pan-neuronally is injected with a fluorescent HaloTag ligand to start the recording window. After fixation and clearing, brain tissue is imaged at large scale to obtain activity maps with single neuron resolution.

Moreover, a split-HaloTag-based PCA can be used to record the formation of naturally occurring protein complexes. I aimed to apply this principle to mark protein-protein interactions across synapses, the functional contact sites between neurons. This approach could not only enable exceptionally selective synapse staining but also has the potential to provide unprecedented insights into the history and plasticity of the neuronal network at large scale, by using distinguishable HaloTag ligands administered at different time points.

2.2 Scope of this Thesis

The scope of this thesis encompasses:

- A biochemical characterization of HaloTag, aiming to provide a detailed understanding of its parameters and mechanism as a foundation for subsequent engineering efforts.
- The development and characterization of a split-HaloTag and the engineering of a molecular Ca^{2+} signal recorder based on this system.
- Refinement of the split-HaloTag system in regard to complementation efficiency and stability, and proof-of-principle demonstration of using the improved system to label synaptic protein-protein complexes.

3 Kinetic and Structural Characterization of HaloTag

Given my objective to design split-HaloTag-based molecular recorders, I considered it important to gain a better understanding of the protein's substrate scope by evaluating the rate of HaloTag labeling across various HaloTag ligands. In addition, by consulting structural data, I aimed to delineate a structure-function relationship for HaloTag ligands. While it is known that the chloroalkane-PEG HaloTag ligand core can be modified with a large variety of fluorophores and other functional moieties, little information is available on the effect that these modifications have on labeling efficiency. Hence, this study serves two primary objectives. Firstly, it provides insight into the optimal substrates from a purely biochemical point of view. Secondly, it allows to set an upper limit of labeling speed that can be achieved with current HaloTag technology, which will facilitate benchmarking of the newly developed molecular recorders. In addition, I expected that this work would be an excellent resource for other researchers, who apply HaloTag technology for their studies or develop new HaloTag-based tools, by guiding ligand choice or providing design principles for new HaloTag ligands.

3.1 HaloTag Labeling Kinetics

3.1.1 Rhodamine-derived ligands

Rhodamine derivatives are a common class of HaloTag ligands[189]. I measured HaloTag labeling rates with a selection of rhodamine-derived ligands, including HTL-TMR, HTL-JF$_{549}$, HTL-LIVE580, HTL-CPY and HTL-JF$_{669}$, HTL-TMR-biotin and HTL-Alexa488. Except for HTL-Alexa488 which is negatively charged at physiological pH values, these dyes have a net charge of zero, but are present as zwitterions. The progress of the reaction was monitored by recording the strong increase in fluorescence anisotropy or fluorescence polarization that results from the association of the small fluorophore ligand to the significantly larger HaloTag protein. The rapid labeling kinetics of most substrates, with the exception of the negatively-charged HTL-Alexa488, required the use of a stopped-flow setup to ensures minimal time delay between mixing and measurement as well as fast sampling rates. To analyze the data, I first fit a simple one-step second-order reaction model to the data:

$$P + S \xrightarrow{k_{app}} PS \tag{3.1}$$

where P is the protein, S is the substrate and PS is the protein-substrate conjugate. However, this model did not fit well to the clear biphasic trend (fig. S4) in the data. This suggests that the HaloTag labeling

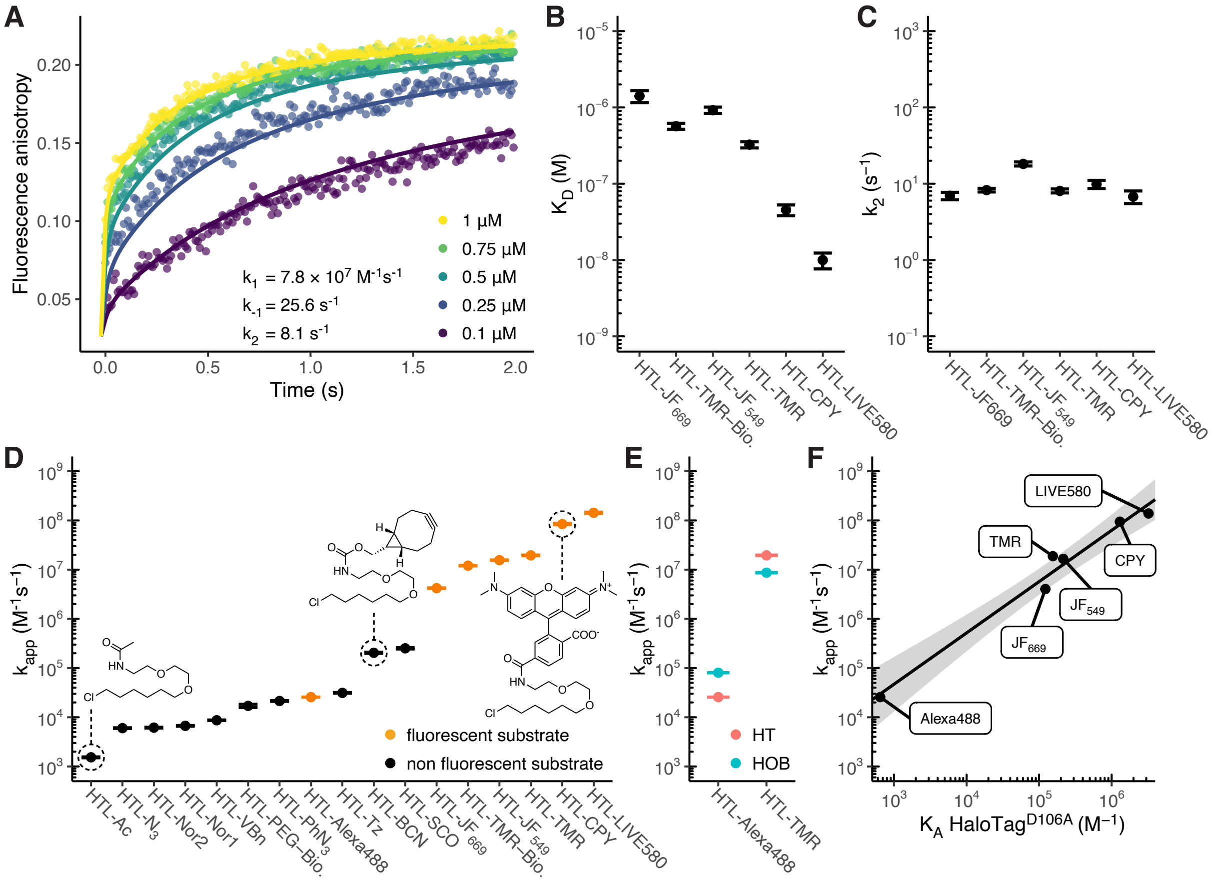


Figure 3: Kinetic characterization of HaloTag. (**A**) Stopped-flow fluorescence anisotropy traces of HaloTag labeling with HTL-TMR in 1:1 stoichiometry at the indicated concentrations fit with a two-step reaction model (equation 3.2). (**B**) Affinities (K_D) of HaloTag for different fluorophore ligands calculated from the kinetic parameters ($K_D = k_{\text{-}1}/k_1$). (Bio.: biotin) (**C**) Labeling rate constants (k_2) for different fluorophore ligands. The minor variations in k_2 show that differences in labeling kinetics are mostly due to differences in K_D. (**D**) Apparent second-order labeling rate constants (k_{app}) of HaloTag labeling with fluorescent and non-fluorescent ligands, spanning over six orders of magnitude. Non-negatively-charged rhodamine derived ligands reach the fastest labeling kinetics, small, non-aromatic ligands have the slowest kinetics. (**E**) Comparison of HaloTag (HT) and HOB labeling kinetics with HTL-TMR and HTL-Alexa488 (Alexa488) reveals faster labeling of HOB with the negatively-charged HTL-Alexa488 but slower reaction with the non-charged HTL-TMR. (**F**) Positive correlation between k_{app} and affinity ($K_A = 1/K_D$, measured using the inactive variant HaloTag$^{\text{D106A}}$) for different fluorophore ligands. A linear model was fit to the log-transformed values ($\log k_{app} = 1.042 \cdot \log K_A + 1.544$, $R^2 = 0.952$). The gray area represents the 95 % confidence bands. Error bars represent 95 % confidence intervals. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

reaction includes at least two relevant steps, an initial binding event and a subsequent covalent reaction:

$$P + S \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} PS^* \xrightarrow{k_2} PS \tag{3.2}$$

where P is the protein, S is the substrate, PS^* is the protein-substrate complex and PS is the protein-substrate conjugate. This model was indeed sufficient to fit the kinetic data effectively and determine the kinetic rate constants k_1, k_{-1} and k_2 independently (fig. 3A, C, S5, table S1). Subsequently, this enabled the calculation of the dissociation constant K_D and the apparent second-order labeling rate constant k_{app}[190,191] (fig. 3B, D):

$$K_D = \frac{k_{-1}}{k_1} \qquad\qquad k_{app} = k_1 \frac{k_2}{k_2 + k_{-1}} \tag{3.3}$$

The K_D value describes the affinity of the HaloTag protein for its substrate, while k_2 defines the maximal rate of protein-substrate conjugate formation at saturating conditions. The tested zwitterionic rhodamine-derived ligands HTL-TMR, HTL-JF_{549}, HTL-LIVE580, HTL-CPY, HTL-JF_{669} and HTL-TMR-biotin featured relatively similar k_2 values around $9.7\,\text{s}^{-1}$ ($4.3\,\text{s}^{-1}$ std. dev., fig. 3C), which could be attributed to the mechanistic similarity of the nucleophilic attack and subsequent covalent binding. These processes happen distant from the fluorophores, at the end of the unvaried chloroalkane-PEG part of the HaloTag ligand. This part, deeply buried in the protein, may adopt a comparable conformation for all substrates, likely explaining the little effects of the substituent on k_2. This suggests that HaloTag ligands, independently of the attached fluorophore, cannot exceed labeling rate constants of ~$10\,\text{s}^{-1}$ resulting in half-labeling times of 0.07 s. Arguably this limit is not relevant for most applications, as incubation times usually exceed the sub-second time scale. Moreover, in many cases, factors such as ligand delivery and permeability across cellular membranes might actually be limiting, hindering the ligand from reaching saturating concentrations at its target site.

In contrast to k_2, K_D values varied significantly depending on the attached fluorophore ranging from ~ 10 nM for HTL-LIVE580 to ~ 1 µM for HLT-JF_{669}, HTL-JF_{549} and HTL-TMR-biotin. This shows that the affinity of the respective fluorophore for the surface of the HaloTag protein plays a crucial role in determining the efficacy of a ligand as a HaloTag substrate. Interestingly, the association rate constants k_1 are relatively similar among different ligands and differences in K_D can be mostly attributed to differences in the dissociation rate constants k_{-1} (fig. S7A,B), indicating that the stability of the initial protein-substrate complex has the largest impact on the overall labeling rate.

Comparing K_D values of HaloTag to average the K_M values of enzymes listed in the BRENDA database[192,193], shows that the affinity of HaloTag for its substrates is on the upper end of the affinity spectrum. This might be due to the fundamental differences in the mechanisms of self-labeling proteins and conventional enzymes. Traditional enzymes must balance substrate affinity and product release, preventing them from indefinitely increasing substrate affinity. Self-labeling proteins, which form a covalent bond to their substrates, avoid the constraints of product inhibition, and can be engineered to reach theoretically unlimited affinity. Besides that, naturally evolved enzymes do not benefit from K_M values far below the physiological concentration of their cognate substrates, which might explain the scarcity of high-affinity enzymes rare.

The k_{app} parameter is the second-order rate constant (or sometimes called covalent efficiency constant) of the labeling reaction in non-saturating conditions with concentrations far below K_D, when no significant accumulation of the intermediate complex PS^* occurs[190,194]. Thus, k_{app} accounts for both, the affinity between the reaction partners and the maximum rate of covalent bond formation, making it my primary metric for comparisons between HaloTag ligands. The k_{app} values for the tested zwitterionic rhodamine-derived substrates were ranging from $4.0 \cdot 10^6\,\text{M}^{-1}\,\text{s}^{-1}$ for HTL-JF_{669} to $1.4 \cdot 10^8\,\text{M}^{-1}\,\text{s}^{-1}$ for HTL-LIVE580. This indicates that with the most performant substrates, HaloTag approaches diffusion-limited labeling rate constants (10^8–$10^{10}\,\text{M}^{-1}\,\text{s}^{-1}$) and can be considered close to a "perfect self-labeling protein", as nearly every encounter with the substrate leads to a labeling event[193,195–197]. Even with the least efficient tested substrate HTL-JF$_{669}$, at a typical ligand concentration of 100 nM for a labeling experiment, this corresponds to half-labeling times of less than 2.5 seconds.

3.1.2 Negatively charged ligands

HTL-Alexa488 features two negatively-charged sulfonic acid groups attached to the xanthene moiety and showed much slower labeling rates with HaloTag than the zwitterionic substrates discussed above. Labeling kinetics of HTL-Alexa488 were measured in a plate reader experiment and did not allow to determine to kinetic parameters k_1, k_{-1} and k_2 of the two-step reaction model (equation 3.2) since no conditions with onset of observable binding site saturation could be reached due to limited solubility and availability of the substrate. As a consequence data was analyzed by fitting directly k_{app} of $2.6 \cdot 10^4$ $\mathrm{M^{-1}\,s^{-1}}$ using the simplified one-step reaction model (equation 3.1, fig. S8, table S2). That means that HaloTag labeling with HTL-Alexa488 is more than tree orders of magnitude slower than with the average zwitterionic fluorophore substrates tested (fig. 3D). Yet, half-labeling time with HTL-Alexa488 at 100 nM concentration would still be below 7 min and on par with the fastest click chemistry reactions[198].

Koßmann et al.[199] have developed a HaloTag variant called halo-based oligonucleotide binder (HOB) which features several mutations close to the ligand binding site that create a more positive electrostatic surface potential which increases labeling rates with negatively-charged oligonucleotide bearing HaloTag ligands. I assessed weather these surface mutations would increase labeling rates with the negatively-charged HTL-Alexa488 as well. A 3.13 ± 0.01-fold increase in k_{app} compared to regular HaloTag was observed (fig. S8, table S2). However, the zwitterionic substrate HTL-TMR had 2.09 ± 0.01-fold slower labeling kinetics with HOB (fig. 3E, S6, table S1). This indicates that the labeling with negatively-charged substrates might be hampered by charge repulsions at the native HaloTag surface, and the introduction of positively charged mutations in HOB mitigates this effect.

3.1.3 HaloTag ligand affinity as a proxy for labeling speed

As discussed above, different rhodamine-derived substrates tend to feature comparable k_2 values for their labeling reaction with HaloTag and mostly vary in affinity of the initial complex formation (K_D). I hypothesized that instead of doing time-consuming and intricate stopped-flow measurement one could estimate the k_{app} of a substrate by assessing its affinity at equilibrium, using the non-reactive HaloTag mutant, HaloTag$^{\mathrm{D106A}}$. The substitution of the nucleophilic aspartate in the active site with a non-reactive alanine residue prevents the covalent attachment of ligands. However, substrates retain their ability to bind reversibly. This allows measuring affinities in a more direct manner through fluorescence polarization titrations using a plate reader (fig. S9, table S4). As hypothesized, the affinities measured with the inactive mutant (K_D^{D106A}) correlate well with affinities calculated from the kinetic data using equation 3.3 (fig. S7C). Consequently, K_A^{D106A} ($= 1/K_D^{D106A}$) correlates with k_{app} (fig. 3F), indicating that measuring affinities to HaloTag$^{\mathrm{D106A}}$ presents a viable strategy to estimate the efficiency of fluorescent HaloTag ligands in a straightforward experiment.

The measurements underlined the preference of HaloTag for zwitterionic rhodamine-derived ligands. However, also HTL-Cy3 and HTL-Cy5, bearing non-sulfonated cyanine dyes, showed decent affinities. Fluorescein, its derivative Oregon green, as well as the rhodol-derived JF$_{503}$ featured lower affinities to

HaloTag^{D106A}. Notably, the discovery of the substantial affinity of HaloTag^{D106A} for rhodamine ligands led to the development of a non-covalent, exchangeable HaloTag system[200]. This system shows reduced photobleaching and improved performance for multi-frame super-resolution microscopy.

3.1.4 Non-fluorescent ligands

After studying rhodamine-derived fluorescent HaloTag ligands, I aimed to determine labeling rates for a diverse set of non-fluorescent ligands. These ligands incorporated various chemical functionalities, such as click chemistry moieties or affinity handles. Due to the non-fluorescent nature of these substrates, direct online monitoring of their reactions, as done with fluorescent substrates, was not possible. Hence, I used an indirect assay in which the non-fluorescent ligands compete against the negatively-charged and slow-reacting HTL-Alexa488 for HaloTag labeling (fig. S10). A simplified competition model allowed to directly determine the apparent second-order labeling rate constant of the non-fluorescent ligand (k^I_{app}):

$$P + S \xrightarrow{k^S_{app}} PS \qquad\qquad P + I \xrightarrow{k^I_{app}} PI \tag{3.4}$$

where P is the protein, S is the fluorescent substrate, I is the non-fluorescent substrate (or “inhibitor”), PS is the protein-fluorescent substrate conjugate and PI is the protein-non-fluorescent substrate conjugate. k_{app} values of non-fluorescent ligands ranged from $10^3\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ to $10^6\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ (fig. 3D, table S3). These values are significantly lower than those observed for zwitterionic rhodamine derivatives, indicating a strong preference of HaloTag for rhodamines. This finding is not unexpected, given that during the development of the HaloTag, ligands such as TMR and fluorescein were predominantly used[139,143], which might have lead to this substrate preference. Among the tested non-fluorescent ligands, large alkynes (cyclooctyne and bicyclononyne) and aromatic structures (tetrazine, phenylazide and vinylbenzene) were favored over alkenes (norbonenes) and small moieties (acetyl and methyl azide).

3.1.5 Improving labeling performance by ligand design

Since HaloTag shows relatively inefficient labeling with click chemistry moieties and affinity handles compared to rhodamine ligands, I speculated that labeling with such groups could be drastically improved by linking them to a rhodamine HaloTag ligand scaffold. Indeed I found that the labeling reaction with a HTL-TMR-biotin ligand was more than 500-fold faster than with the commercially available HTL-PEG-biotin ligand (fig. 3D). I expect that this approach for improving performance of sub-par HaloTag ligands can be applied to other moieties. Interestingly, a more recent study has shown that intracellular labeling of a biotin HaloTag ligand can also be enhanced through the incorporation of a silicon-rhodamine fluorophore in place of the TMR utilized in my experiments[201], validating the approach.

3.2 HaloTag Structure

3.2.1 Structural analysis of rhodamine-labeled HaloTag and HOB

To better understand HaloTag’s affinity for rhodamine-based CA substrates, HTL-TMR and HTL-CPY labeled HaloTag and HTL-TMR labeled HOB were crystalized and their structures were solved[188] (fig. 4A). All these structures show a consistent α/β-hydrolase fold common to the dehalogenase super-

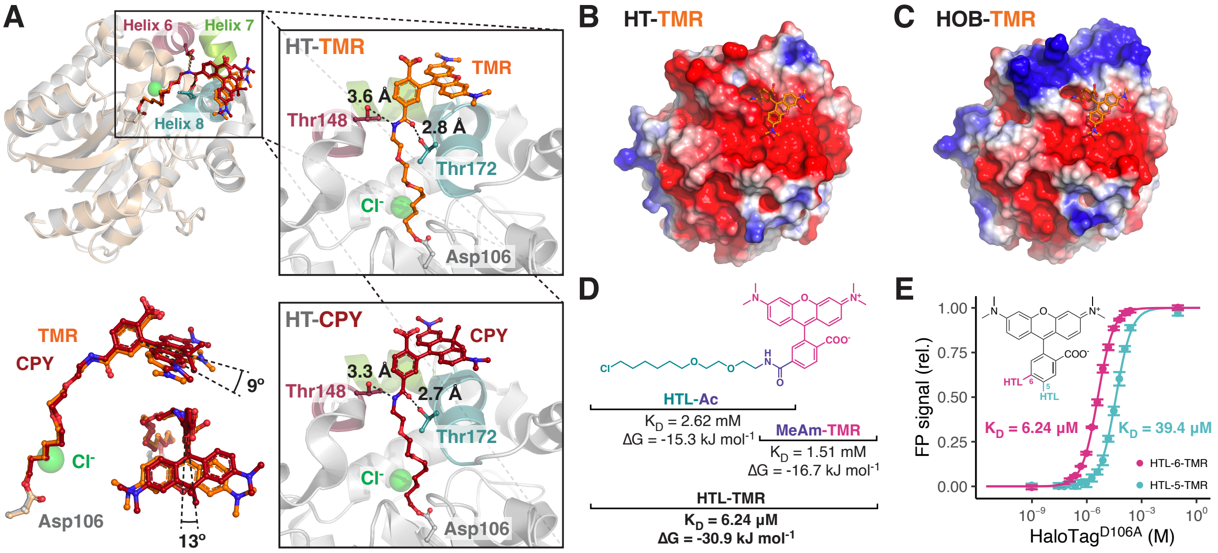


Figure 4: Structure-function analysis of HaloTag--ligand interactions. (**A**) Overlay of HaloTag-TMR (PDB-ID 6Y7A, gray)[188] and HaloTag-CPY (PDB-ID 6Y7B, chain A, gold)[188] structures and close-ups of ligand binding sites, confined by the three α-Helices 6–8. Potential hydrogen bonds between the capping domain and the ligand are annotated with distances. HTL-TMR and HTL-CPY adopt very similar protein bound conformations as depicted in the bottom left. (**B** and **C**) Electrostatic surface potentials of HaloTag-TMR (PDB entry 6Y7A) and HOB-TMR (PDB entry 6ZCC) show a negatively-charged patch above the ligand binding site of HOB. Potentials are colored from −2.0 (red) to 2.0 (blue) kT/e and were calculated using the APBS software[202] with standard parameters. (**D**) Affinities (K_D) and free binding energies (ΔG) for the interaction between the inactive HaloTag$^{\mathrm{D106A}}$ variant and HTL-TMR and its fragments. (**E**) Affinities of HaloTag$^{\mathrm{D106A}}$ for HTL-6-TMR and HTL-5-TMR show a clear preference for ligands linked via the 5-position. Error bars represent 95 % confidence intervals. Reprinted in modified version with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

family. HaloTag features an additional capping domain of six α-helices which partially cover the active site and constitute an entry channel for the HaloTag ligands. After the reaction, the PEG-alkane ligand is enclosed within the protein, the amide linkage to the fluorophore is positioned by two hydrogen bonds to the capping domain and the xanthene component of the dye is positioned on top of α-helix 8 of the capping domain. Thereby, a large interface between the xanthene and the protein surface is formed, potentially stabilized by van der Waals forces. Both HTL-TMR and HTL-CPY have very similar configurations, but the additional methyl groups of CPY, one of which extends the protein-fluorophore interface, could explain why HaloTag features a higher affinity towards HTL-CPY than to HTL-TMR.

HOB has four surface mutations differing from HaloTag, leading to an increased labeling rate with negatively-charged CA substrates. While HaloTag has a negatively-charged surface near the substrate entry, HOB features a positively charged patch opposite to the fluorophore binding site (fig. 4B,C). Even though ligands like HTL-Alexa488 could not directly form charge-charge interactions to this patch if they bind in the same mode as HTL-TMR, a potential electrostatic steering effect[203] could explain the increased labeling rates of HOB with this ligand.

3.2.2 Dissecting HaloTag ligand affinity

To gain a better understanding of how much the chloroalkane-PEG and the rhodamine part of a ligand contribute to HaloTag binding, I dissected the ligand into two fragments: an acetylated chloroalkane-PEG (HTL-Ac) and N-methylamide-fluorophores (MeAm-TMR/CPY). Even though the chloroalkane-PEG is deeply buried within the protein, having a large contact area for van der Waals forces and potential hydrogen bonds via its amide moiety, I observed a moderate affinity for the inactive HaloTag$^{\mathrm{D106A}}$ variant with a K_D of 2.62 mM which is in line with the low k_{app} of HTL-Ac. The affinity for MeAm-TMR and MeAm-CPY were slightly higher with a K_D of 1.51 mM and 0.71 mM, respectively. Hence, binding to HaloTag is not only mediated by the chloroalkane-PEG part of a ligand but can also be substantially driven by the affinity for the rhodamine part.

The free binding energies of the fragments calculated from the K_D values (using the relationship $\Delta G = RT \cdot \ln K_D$) are $-15.3\,\mathrm{kJ\,M^{-1}}$ for HTL-Ac and $-16.7\,\mathrm{kJ\,M^{-1}}$ for MeAm-TMR. Comparison of the sum of these energies ($-32.0\,\mathrm{kJ\,M^{-1}}$) to the free binding energy of the full HTL-TMR ligand of $-30.9\,\mathrm{kJ\,M^{-1}}$ reveals only a very minor synergistic effect[204] (fig. 4D, table S4, S5). A similar observation were made when comparing MeAm-CPY and HTL-Ac ($-18.7\,\mathrm{kJ\,M^{-1}}$ and $-15.3\,\mathrm{kJ\,M^{-1}}$) to HTL-CPY ($-36.4\,\mathrm{kJ\,M^{-1}}$) (fig. S12, table S4, S5).

3.2.3 Effect of linkage position on rhodamine substrates

Commercially available carboxy-rhodamines, which can be readily used to synthesize HaloTag ligands, are often sold as mixtures of 5- and 6-carboxy regioisomers (fig. 4E) as both molecules are formed during the classical synthesis route using substituted phthalic anhydrides[205]. Usually HaloTag ligands bearing rhodamines are linked via the 6-position to the chloroalkane-PEG and ligands linked in this way are used throughout this thesis. However, to test whether the linkage position of rhodamine ligands matters, I measured affinities of HaloTag$^{\mathrm{D106A}}$ for HTL-5-TMR, HTL-5-CPY, MeAm-5-TMR and MeAm-5-CPY. In all cases, ligands linked via the 5-position had lower affinities than their cognate 5-linked ligands. A 6.31- and 22.7-fold decrease was observed for HTL-TMR and HTL-CPY, respectively (fig. 4E, S12, table S4). Even MeAm-TMR and MeAm-CPY showed a 1.28- and 2.81-fold reduced affinity when modified at the 5-position (S12, table S5).

These results highlight the benefits of using 6-linked rhodamine HaloTag ligands. In addition they underline the relevance of the xanthene–α-helix 8 interface for efficient ligand binding, as this interaction could not be formed with a 5-linked rhodamine, which would be tilted 60° away from the protein surface.

3.3 Conclusion and Outlook

This study presents a thorough characterization of HaloTag labeling kinetics with a number of fluorescent and non-fluorescent ligands. Alongside the structural analysis of the rhodamine labeled protein, these results provide insights into the structure-activity relationship for HaloTag and its ligands.

I observed that the labeling reactions of HaloTag with fluorescent ligands can be best described by a two-step process, including a reversible binding phase, which is followed by the irreversible formation of an covalent bond. Using stopped-flow kinetics, I was able to determine the kinetic parameters of this model for a number of rhodamine ligands. Most ligands featured affinities in the sub-micromolar range for the reversible binding and rates of covalent bond formation with sub-second half-lifes. Hence, HaloTag features very rapid overall labeling rates with many rhodamine-derived ligands, reaching full labeling after few seconds at nanomolar ligand concentrations. The high affinity and consequently fast labeling rates can be attributed to the fact that HaloTag does not only bind to the chloroalkane-PEG of the substrates, but also to the xanthene ring of the attached fluorophore. This interaction of HaloTag with the aromatic xanthene could also be a significant factor for its ability to shift the equilibrium of fluorogenic rhodamines from a non-fluorescent spirocyclic form to a zwitterionic and fluorescent one, that contains the xanthene moiety.

In contrast, when labeled with the negatively-charged HTL-Alexa488 or non-rhodamine ligands, HaloTag displayed up to five orders of magnitude slower labeling rates, compared to zwitterionic rhodamine ligands. This finding might be due to the use of rhodamine ligands like TMR during HaloTag's development and engineering. However, labeling rate of a non-fluorescent biotin ligand could be drastically improved by linking the biotin group to HTL-TMR rather than directly to the chloroalkane-PEG. I expect this strategy to be generalizable and might improve the performance of HaloTag for applications beyond fluorescent tagging such as pull-downs[206], crosslinking[154] or tethered pharmacology[207].

In parallel to my study, Stefanie Kühn characterized the alternative self-labeling proteins, SNAP-tag and CLIP-tag[188]. SNAP-tag and CLIP-tag exhibited considerably slower labeling rates with rhodamine ligands compared to HaloTag. Yet, they displayed less variability across ligands with different chemistries, and with some non-fluorescent ligands, SNAP-tag outperformed HaloTag. However, the slower labeling rates with rhodamines and the lower cell-permeability reported for SNAP-tag substrates compared to their HaloTag counterparts[208], highlight the necessity of further engineering of SNAP-tag to get on par with HaloTag for demanding applications.

For my overarching goal of creating molecular recorders for in vivo applications using a split-HaloTag, the fast labeling speed of HaloTag is highly advantageous, as it facilitates efficient labeling even at low ligand concentrations. Given the in vivo context, ligand dose and the duration of its presence are constrained by administration methods and pharmacokinetics, emphasizing the importance of rapid labeling. Furthermore, it is expected that splitting HaloTag into two fragments would lead to a decrease in activity due to more flexibility at the split site, making it even more important to have a robust and highly active protein as a starting point.

A limitation of the presented study is that the characterization of HaloTag labeling was approached exclusively from a biochemical standpoint, ignoring other aspects like membrane permeability and bioavailability of HaloTag ligands. Yet these key parameters need to be considered for experiments in cultured cells and even more so for in vivo applications. However, some rhodamine ligands characterized in my work, such as HTL-JF$_{669}$, which showcased decent labeling rates, are also reported to be highly bioavailable[149], making them promising ligands for in vivo labeling of molecular recorders. Nonetheless, a comprehensive assessment of permeability and bioavailability of HaloTag ligands is lacking and would significantly benefit the scientific community.

Despite the remarkable labeling rates of HaloTag with rhodamine ligands, there is still room for improvement. The fraction of binding events that lead to covalent attachment ($k_2/(k_{-1} \cdot k_2)$) is on average 35 % for the tested ligands. This parameter could be boosted by either decreasing the dissociation rate constant k_{-1} or by increasing the rate constant of covalent bond formation k_2 to get towards a "perfect" self-labeling protein. Recently alternative HaloTag ligands were described that are reported increase labeling efficiencies for non-fluorescent substrates by incorporating a methoxybenzene moiety within the chloroalkane-PEG linker[209]. One could speculate that this structure with increased surface area and rigidity might indeed have a beneficial effect on the dissociation rate constant. On the other hand exchanging the chloride in the HaloTag linker to a better leaving group like bromine or iodine could be a means to increase in the rate of covalent bond formation. However, such substates might be less stable and might participate in unwanted side-reactions. Lastly further protein engineering of HaloTag might also be a useful strategy to perfect the system.

4 | A Molecular Calcium Signal Recorder

To unravel the complexities of the brain and understand the mechanisms enabling sophisticated tasks such as cognition and memory, it is necessary to study the activity of the neuronal network under varied conditions. Tools enabling live observation of neuronal activity, such as voltage or calcium (Ca^{2+}) sensors, have provided invaluable insights by offering excellent temporal and spatial resolution[30,210]. However, a major constraint of these approaches is their limited spatial scale, as only a restricted field of view is accessible for observation at a given moment. Hence, there is a demand for molecular recorders of neuronal activity that capture signals during a defined time window for later readout[44,54,63]. Such tools can overcome limitations of live recordings, providing a view on neuronal activity throughout the brain, including the deeper structures. While compromising on temporal resolution, they are critical for unraveling larger circuits that span across brain regions and complement current live observation tools. Moreover, post-hoc readouts can benefit from fixed-tissue preparations like sectioning, clearing or expansion that further enhance scale and resolution[42–44] and are compatible with further downstream analyses like highly multiplexed imaging[96,211] or spatial transcriptomics[212,213].

The development of a novel molecular Ca^{2+} recorder, constructed from a modified HaloTag protein, presents a promising candidate. A central requirement for a molecular recorder is the ability to establish a permanent mark, a feature inherent to HaloTag. The availability of fluorescent HaloTag ligands across the spectral range, which have been shown to penetrate tissues in vivo and distribute within the brain[17,147,149,152], renders the technology versatile and could enable the recording of successive epochs with different fluorescent colors. In addition rhodamines used in HaloTag ligands are bright, stable and red-shifted organic dyes, optimal for imaging in tissues.

To transform HaloTag into a molecular Ca^{2+} recorder, it was necessary to gain control over the HaloTag labeling reaction. A straightforward strategy to achieve such control is the development of a split-HaloTag system. As the dynamics of such split system have to be compatible with the time scales of neuronal Ca^{2+} waves, excising only a small fragment from HaloTag to prevent a complete unfolding of the protein (and the necessity to re-fold upon complementation), seemed the be the most promising approach.

A molecular Ca^{2+} recorder could then be constructed from the split-HaloTag by fusing it to Ca^{2+}-reactive proteins, such as the calmodulin–M13 pair. These domains, which are used in a plethora of Ca^{2+} sensors, induce dimerization and split complementation upon elevations in Ca^{2+} concentration, rendering the system responsive to Ca^{2+} transients associated with neuronal activity.

4.1 Development and Characterization of a Split-HaloTag

4.1.1 Circular permutation of HaloTag

As attempts to generate a split-HaloTag by truncating short N- or C-terminal fragments were unsuccessful, an alternative strategy with the goal to excise internal fragments was pursued. To achieve this HaloTag was circular permuted by connecting the original termini with a flexible $(GGS/T)_5$ linker and opening new termini within the protein (cpHaloTag). New termini were introduced at surface-exposed loops flanking α-helices adjacent to the fluorophore ligand binding site (fig. 5A). Ten different constructs were generated and screened for their labeling kinetics and thermostability (fig. 5B). Circular permutation leads to a significant decrease in labeling speed (> 19-fold) and melting temperature (> 17 °C). However the best performing variants with new termini at positions 154–156 and 141–145 still featured relatively fast labeling rate constants (> 2.23 $M^{-1}\,s^{-1}$) and good thermostabilities (> 42.8 °C).

I suspected that The N- and C-terminal purification tags of the cpHaloTag proteins could interfere with their function as the tags would be in direct proximity to the ligand binding site. Thus, cpHaloTag 154–156 and 141–145 were expressed from a modified pET-51b(+) vector with only a N-terminal His-tag followed by a TEV cleavage site. This allows to obtain clean, non-tagged protein using the TEV-protease. Stopped-flow labeling kinetics with HTL-TMR were measured and analyzed using the two-step reaction model (equation 3.2, fig. S14). cpHaloTag 154–156 without the purification tags exhibited a k_{app} of 9.22 $M^{-1}\,s^{-1}$, which is nine-fold higher than with the tags and only two-fold slower than the native HaloTag (table S7). In contrast, cpHaloTag 141–145 exhibited comparable labeling rates irrespective of the presence or absence of tags. cpHaloTag 154–156 featured a similar affinity (k_{-1}/k_1) for HTL-TMR as HaloTag. However, the rate of covalent bond formation (k_2) was three-fold lower compared

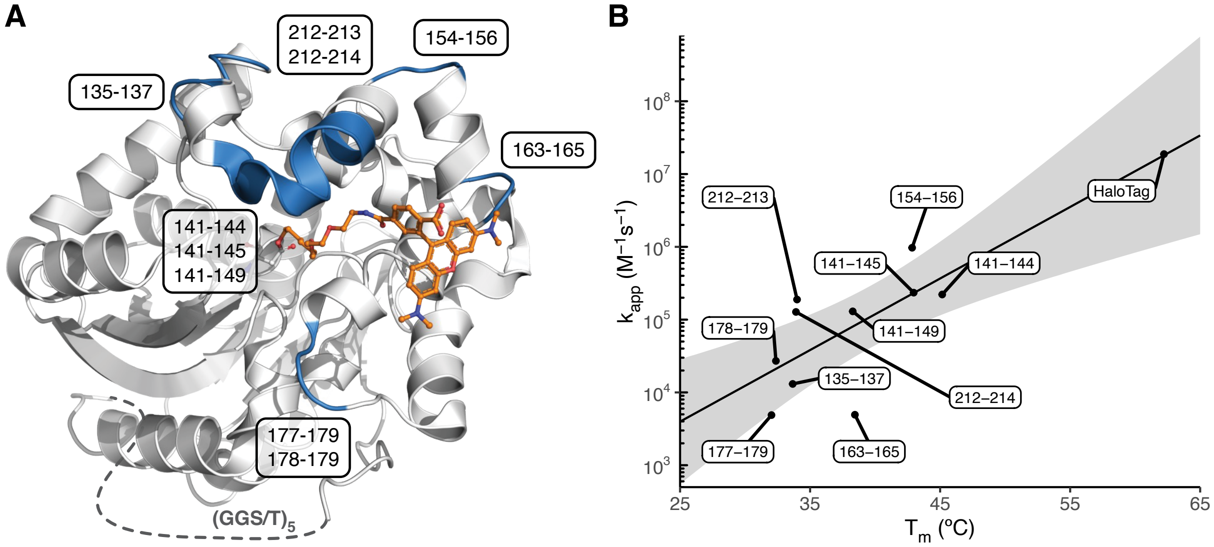


Figure 5: Circular permutation of HaloTag. (**A**) Structure of HTL-TMR labeled HaloTag (PDB-ID 6Y7A)[188]. The termini were connected with a flexible $(GGS/T)_5$ linker and new termini were opened at the annotated sites (blue) to generate circular permuted HaloTags (cpHaloTag). (**B**) Apparent second order rate constants (k_{app}) with HTL-TMR and melting temperatures (T_m) of cpHaloTag variants in comparison to the native HaloTag. A linear model was fit to the data ($\log k_{app} = 0.0983 \cdot T_m + 1.142$, $R^2 = 0.645$). The gray area represents the 95 % confidence bands.

to HaloTag. Furthermore, cpHaloTag 141–145 not only showed a reduced k_2 (seven-fold) but also an decreased affinity (16-fold). These observations show that despite the modifications (opening of new termini) being located at the surface, they influence the covalent bond formation which happens at the core of the protein.

To further investigate the effects of circular permutation, cpHaloTag 154–156 was crystalized and its structure was solved via an in-house collaboration. Except for the new termini, the structure overlays almost perfectly with the HaloTag structure (C_α-RMSD = 0.195 Å). Also the residues of the active site, surrounding the reactive Asp^{106}, are oriented nearly identically. Thus, the differences in the rate of covalent bond formation might be due to subtle dynamic changes within the core of the protein as indicated by the lower melting temperature of the circular permuted proteins. Such changes could affect the active site and might not be represented in the crystal structure.

4.1.2 Split-HaloTag design

Having the two well-performing circular permuted variants (cpHaloTag 154–156 and 141–145) in hands, I proceeded to generate a split-HaloTag based on them. This was accomplished by excising the small helical fragment located between these two circular permutation sites, which participates in forming the ligand entry channel of the protein (fig. 6A). The resulting protein, hereafter called cpHaloΔ, exhibits negligible activity on its own. However, its self-labeling activity can be restored upon the reversible binding to the small excised fragment, hereafter called Hpep. I initially used a 9-mer Hpep (145–153) which I had characterized in my master's thesis. I hypothesized that extending this fragment by incorporating an additional C-terminal residue, Thr^{154}, to obtain a 10-mer Hpep (145–154) could increased its performance. In the HaloTag crystal structure Thr^{154} forms four potential hydrogen bonds to residues outside of the Hpep segment – two via the sidechain hydroxyl group and two via its backbone carbonyl group – which might assist binding and correct orientation of the Hpep in the cpHaloΔ groove (fig. 6A). Indeed, I measured a five-fold increase in cpHaloΔ labeling activity in the presence of the 10-mer Hpep relative to the 9-mer Hpep (fig. 6B). Given this observation, all subsequent characterizations and applications of split-HaloTag were conducted using the 10-mer Hpep. Furthermore, the kinetics of the split-HaloTag labeling reaction could also be monitored online via fluorescence polarization measurements, which confirmed the anticipated Hpep dependency and minimal background activity (fig. 6C).

4.1.3 Structural analysis of cpHaloΔ

cpHaloΔ was crystalized and its structure was solved via an in-house collaboration. The protein retains the overall fold of HaloTag (C_α-RMSD = 0.370 Å), except for the missing Hpep segment. This indicates that despite the removal of the Hpep the protein can still adopt a well-folded structure while exhibiting minimal background activity. This stands in contrast to many other split proteins which are thought to be unfolded when on their own[167,185] and only adopt a distinct fold when brought into proximity of their split partner. A well folded split fragment as cpHaloΔ would be beneficial since it could mitigate destabilization and degradation of fusion proteins[168,186,214]. In addition a pre-folded fragment might accelerate the dynamics of the system as no large-scale folding processes are needed for split complementation.

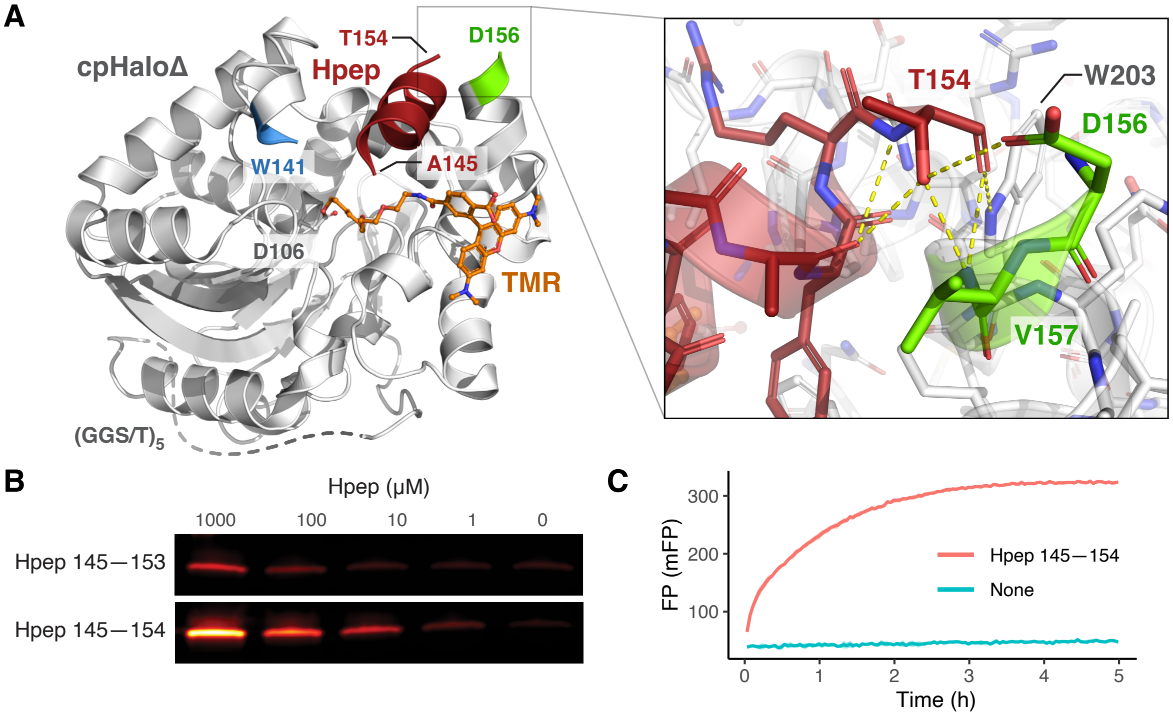


Figure 6: Split-HaloTag design. (**A**) Model of split-HaloTag based on the HTL-TMR labeled HaloTag structure (PDB-ID 6Y7A)[188]. The original termini were connected with a flexible (GGS/T)$_5$ linker and new termini were opened at positions Trp^{141} (blue) and Asp^{156} (green) to generate the inactive protein fragment cpHaloΔ (grey), excising the α-helical Hpep fragment (145–154) located between the new termini (left). Zoom in on Thr^{154}, which forms potential hydrogen bonds to Asp^{156} and Val^{157} via its hydroxyl side chain and to Val^{157} and Trp^{203} via its backbone carbonyl group (right). (**B**) Labeling of 1 µM cpHaloΔ with 2 µM HTL-TMR in the presence or absence of Hpep variants at the specified concentrations for 1 h, analyzed by SDS-PAGE and subsequent in-gel fluorescence scanning. Minimal labeling was detected without Hpeps, while concentration-dependent labeling was observed in the presence of Hpeps. The 10-mer Hpep including Thr^{154} (145–154) showed higher activity compared to the 9-mer Hpep (145–153). (**C**) Labeling kinetics of 100 nM cpHaloΔ with 20 nM HTL-TMR in presence or absence of 100 µM Hpep measured by fluorescence polarization readout (FP).

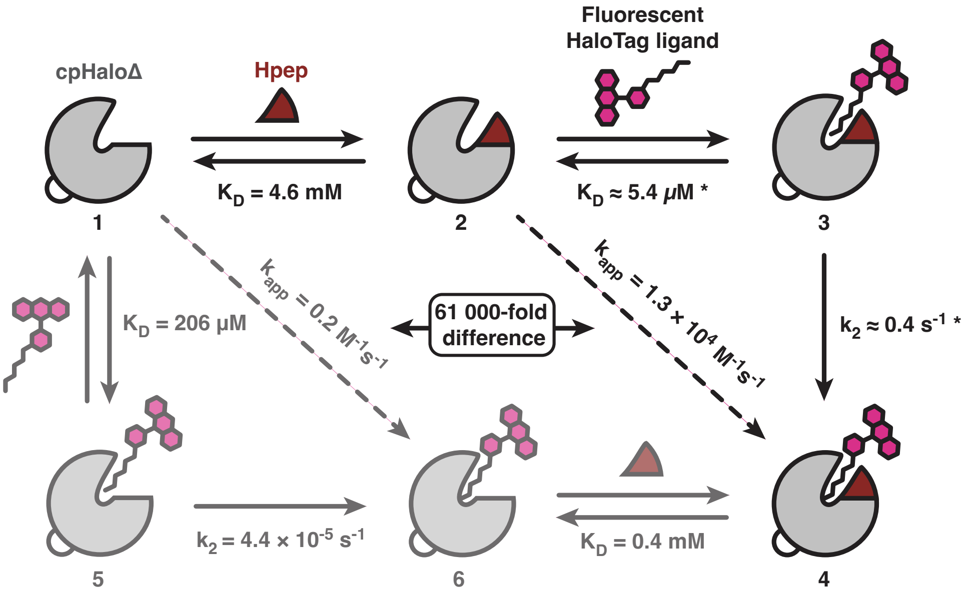
4.1.4 Intrinsic affinity of split-HaloTag

A critical parameter in protein complementation assays is the intrinsic affinity between the split fragments. This becomes particularly important in the context of a split-HaloTag system. Any spontaneous complementation, not induced by fusion proteins bringing the fragments into proximity, results in undesired background labeling. Given the covalent nature of split-HaloTag labeling, this background would accumulate over time, and as a result diminish the sensitivity of the assay. In addition spontaneous complementation could bias fusion proteins as it forces their proximity or interaction.

I measured a K_D value of 4.61 mM for the reversible interaction between cpHaloΔ and Hpep using isothermal titration calorimetry (ITC, fig. S16). This very low affinity would most likely not lead to spontaneous complementation when both parts are co-expressed in the same cell, as intracellular protein concentrations are usually below a low micromolar level[215]. Furthermore, I determined the affinity of Hpep for the HTL-CPY labeled cpHaloΔ. I hypothesized that presence of the ligand might influence Hpep binding, since there is a direct interface between Hpep and the ligand. As anticipated, the presence of the label increased binding affinity, resulting in a K_D value of 0.39 mM. However, this is still a relatively low affinity, which should not significantly bias the interaction of most fusion proteins in a cellular context.

4.1.5 Split-HaloTag labeling kinetics

Another source of background labeling would be residual activity of the cpHaloΔ protein. I measured cpHaloΔ labeling kinetics in absence of Hpep in a fluorescence polarization assay and analyzed the data using a two step reaction model (equation 3.2). This enabled calculation of the affinity of cpHaloΔ (K_D) towards fluorophore ligands, the rate of covalent conjugate formation (k_2) and the apparent second order rate constant (k_{app}) of the complete reaction. cpHaloΔ featured 630-fold higher K_D and a >180 000-fold lower k_2 compared to HaloTag, indicating that circular permutation and deletion of the Hpep segment had drastic impact on ligand binding, but an even more pronounced effect on the reactivity of the active



** approximated values*

Figure 7: Kinetic model of split-HaloTag. (1→2→3→4) The labeling reaction upon association of the Hpep to cpHaloΔ, subsequent binding of the fluorophore ligand and covalent attachment. **(1→5→6)** The background labeling reaction consisting in binding of the fluorophore ligand to cpHaloΔ followed by its covalent attachment. Labeling of complemented cpHaloTag **(2→4)** is 61 000-fold faster than the background labeling **(1→6)**. All values were obtained with the 10-mer Hpep and HTL-TMR as fluorophore ligand. The individual values were determined as follows:
(1→2) Dissociation constant (K_D) was determined by ITC. **(2→4)** Apparent second order rate constant (k_{app}) of complemented split-HaloTag was determined by fluorescence polarization kinetics. **(2→3)** The dissociation constant (K_D) was derived from the labeling reaction of cpHaloTag 141–145. The complemented split-HaloTag is a combination of circular permutations at positions 141–145 and 154–156. Because the circular permutation at positions 154–156 had no impact on K_D, the K_D of cpHaloTag 141–-145 was used as an approximate value. **(3→4)** As described above, the complemented split-HaloTag combines the circular permutations at positions 141-–145 and 154-–156. These two cpHaloTags exhibit a decrease in k_2 by factors of 6.7 and 3.0, respectively. Consequently, one can estimate that the complemented split-HaloTag has a k_2 value that is approximately 20-fold lower than that of the native HaloTag. **(4→6)** Dissociation constant (K_D) was determined by ITC. **(1→5→6)** K_D, k_2 and k_{app} were determined by fluorescence polarization kinetics.

site of the protein. Overall cpHaloΔ features very low residual activity with a k_{app} of 0.5 M^{-1} s^{-1}.

I measured labeling kinetics of complemented split-HaloTag using a fluorescence polarization assay. Direct determination of the labeling rate for the complemented species would require the addition of Hpep at concentrations significantly higher than the dissociation constant (K_D) of 4.6 mM. However, this is not feasible due to the limited solubility of Hpep and the effect high peptide concentrations have on the viscosity of the solution. Thus, I measured kinetics at sub-millimolar Hpep concentrations and fit a model to the data which accounts for both, the activation of cpHaloΔ by Hpep and the subsequent labeling reaction with a HaloTag ligand:

$$cpHalo\Delta + Hpep \xrightleftharpoons[k_{-1}^{Hpep}]{k_1^{Hpep}} P \qquad\qquad P + S \xrightarrow{k_{app}} PS \tag{4.1}$$

where P is the complemented split-HaloTag, S is the HaloTag ligand and PS is the protein-ligand conjugate. By fixing the ratio k_{-1}^{Hpep}/k_1^{Hpep} (which is equivalent to K_D^{Hpep}) to the K_D value (4.6 mM) as measured by ITC, a robust estimate for the apparent second order rate constant (k_{app}) could be obtained. This yielded a value of $1.3 \cdot 10^4$ M^{-1} s^{-1} for the complemented split-HaloTag which is 61 000-fold higher than the labeling rate of cpHaloΔ in absence of Hpep.

The comprehensive biochemical characterization of split-HaloTag enabled the development of a working model for the system, showing the most relevant mechanistic steps depicted in figure 7. This should not only aid the practical application of this tool but also provide a framework for subsequent engineering and optimization of the system.

4.2 Caprola: A Split-HaloTag-Based Molecular Calcium Recorder

4.2.1 Design and preliminary results

A split-HaloTag recorder for Ca^{2+}-dependent protein labeling (Caprola) was constructed by fusing the Hpep to the Ca^{2+}-sensing domain calmodulin (CaM) and cpHaloΔ to the calmodulin-binding peptide M13 using short flexible linkers. Based on this design, labeling of the split-HaloTag should occur exclusively upon the binding of four Ca^{2+} ions to calmodulin. This triggers a conformational change in calmodulin, allowing it to subsequently bind to the M13 peptide[216,217]. Consequently, the split-HaloTag parts are brought into close proximity, allowing for the reconstitution of their self-labeling activity (fig. 8A).

In the initial Caprola constructs, which are described in my master's thesis, I used the nonoptimal 9-mer Hpep. Nonetheless, important engineering steps were performed using these preliminary variants. The lengths of the flexible linkers between Hpep and calmodulin, as well as between M13 and cpHaloΔ were optimized, with short linkers showing the best performance in this context. An intramolecular Caprola was constructed by fusing the two components into a single polypeptide chain (Hpep–CaM–M13–cpHaloΔ, fig. 8A) and the rigidity of the linker between calmodulin and M13 was optimized. Using a helical Pro$_{30}$ linker drastically decreasing background labeling of the intramolecular construct in absence of Ca^{2+}. Additionally, first efforts in tuning the Ca^{2+}-sensitivity of Caprola were made by introducing mutations in the M13 peptide known to . However, due to usage of the slow 9-mer Hpep these designs suffered from low labeling rates in the presence of Ca^{2+} ($k_{app}^{Ca^{2+}}$ = $2.3 \cdot 10^3$ M^{-1} s^{-1}), needing several hours to complete labeling at nanomolar concentrations.

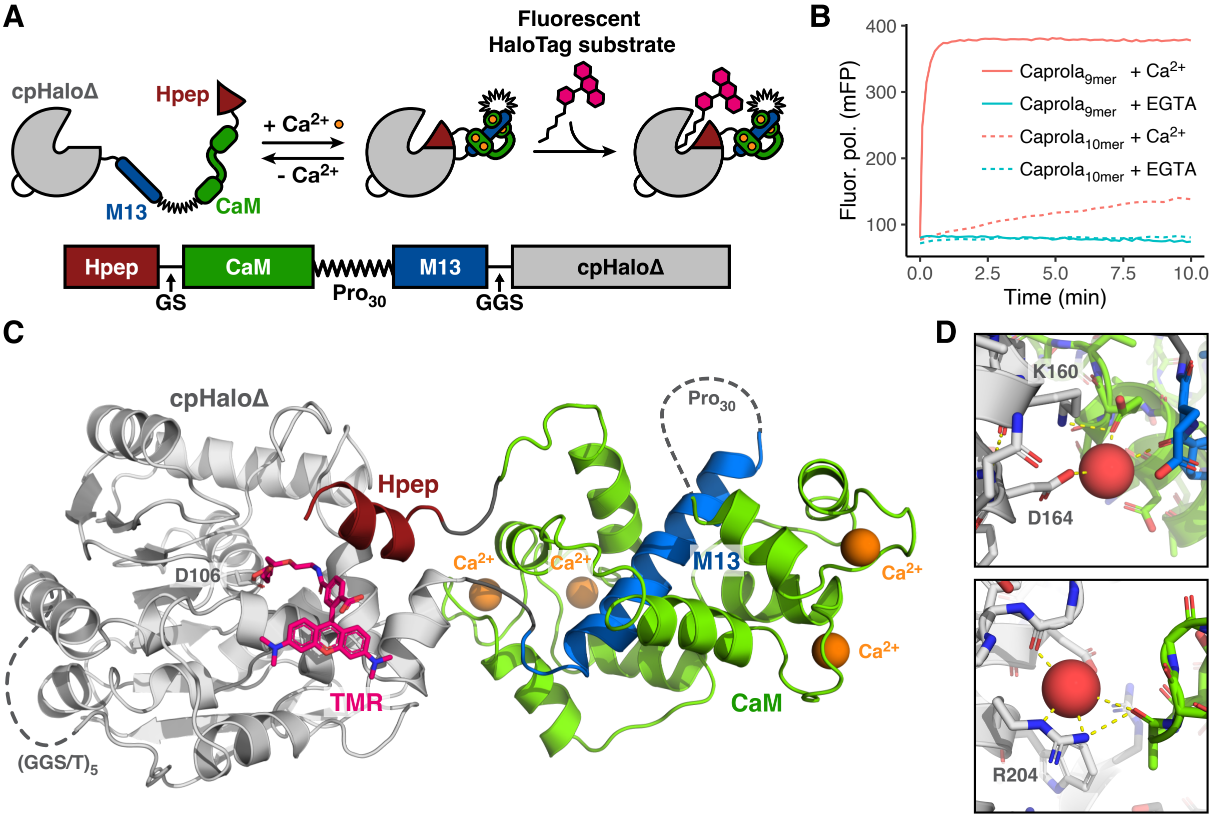


Figure 8: Caprola design and Ca^{2+}-dependent labeling. (**A**) Scheme of Caprola design (top). Calmodulin (CaM) binding of Ca^{2+} leads to increased affinity for M13 which enables splitHaloTag complementation and subsequent labeling. Genetic architecture of Caprola (bottom). (**B**) Ca^{2+}-dependent Caprola (200 nM) labeling kinetics with HTL-TMR (50 nM). The improved Caprola version using the 10-mer Hpep shows 68-fold faster labeling in presence of Ca^{2+} and no background labeling within 10 min. (**C**) Model of the labeled Ca^{2+}-bound Caprola structure. The crystal structure of the Caprola mimic described in section 4.2.3 was modified by removing the direct link between cpHaloΔ and the Hpep. The HTL-TMR label was added by overlay with the HTL-TMR labeled HaloTag structure (PDB-ID 6Y7A)[188]. (**D**) Zoom ins on the interface between cpHaloΔ and CaM. Polar contacts are highlighted in yellow, coordinated water molecules are depicted as red spheres. Residue numbers according to standard HaloTag numbering.

4.2.2 Improving Caprola labeling speed

Upon further characterization of the split-HaloTag system and the discovery of the more effective 10-mer Hpep, I engineered a new generation of Caprola constructs incorporating this improved Hpep. I used the calmodulin domain and M13 peptide from the light-activated Ca^{2+} integrator CaMPARI2[64] since they showed good performance in a related application. This refined variant featured a $k_{app}^{Ca^{2+}}$ of $2.61 \cdot 10^5$ M^{-1} s^{-1} (table S10), representing a 68-fold faster labeling rate compared to the initial constructs previously described (fig. 8B). This results in significantly shorter half-labeling times in the range of minutes instead of hours, which is a crucial step towards application of Caprola for recording neuronal Ca^{2+} waves. Recording times of several hours might not align well with many behavioral experiments and even in highly stimulated neurons elevated Ca^{2+} levels persist only transiently, as neurons use rapid Ca^{2+} buffering and active transport to restore Ca^{2+} concentrations to the baseline level[218,219]. Despite this increase in $k_{app}^{Ca^{2+}}$, the construct maintained minimal activity in absence of Ca^{2+} (k_{app}^{EGTA} = 38 M^{-1} s^{-1}, table S10), achieved by addition of the Ca^{2+}-chelator EGTA (ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N'*,*N'*-tetraacetic acid).

Interestingly, the labeling rate of the improved Caprola is 20-fold faster than the labeling rate determined for the complemented split-HaloTag (fig. 7). This pronounced difference in activity might be caused by reduced flexibility at the fusion sites. The stable CaM–M13 complex potentially anchors the C-terminus of Hpep and the N-terminus of cpHaloΔ in a fixed orientation. Such constraints could alter the inherent dynamics of the protein, favoring conformations that enhance labeling activity. Therefore, the maximal labeling rate of the split-HaloTag might be achieved when the C-terminus of Hpep and the N-terminus of cpHaloΔ are precisely aligned by rigid fusion partners. Based on this reasoning, one might speculate that the maximal labeling rate could approach the labeling rate of cpHalo 141–145 (k_{app} = $3.02 \cdot 10^6$ M^{-1} s^{-1}), where these two sites are directly linked, leading to a rigid configuration.

4.2.3 Caprola structure and stability

To investigate the structure of Caprola in more depth and identify possible targets for further engineering, crystallization trials were conducted. However, no crystals of Caprola protein were obtained, which might be due to the two large unstructured segments: the flexible $(GGS/T)_5$ linker of cpHaloΔ and the Pro_{30} between calmodulin and M13. To omit the Pro_{30} linker crystallization of a split version of Caprola, comprising Hpep–CaM and M13–cpHaloΔ, was attempted. Although the assembled Caprola did not crystalize, crystals of the M13–cpHaloΔ fusion were obtained allowing the structure to be solved via an in-house collaboration (fig. S18).

Interestingly, this structure revealed that the M13 peptide lies within the Hpep binding groove of cpHaloΔ. Specifically, residue Arg^{17} of M13 extends through the ligand entry channel and interacts with the reactive Asp^{106} residue of cpHaloΔ. This interaction presumably hampers binding of HaloTag ligands (fig. S18), potentially explaining the low labeling activity of Caprola in the absence of Ca^{2+}. Additionally, this interaction might increase the stability of cpHaloΔ in Caprola. Indeed M13–cpHaloΔ had a 8.6 °C higher melting temperature than cpHaloΔ alone (T_m = 31.3 °C, table S8). This stabilizing effect appeared to be consistent in the full-length Caprola, which had a melting temperature of 39.9 °C in presence of Ca^{2+} and 42.5 °C in its absence. It seems that within Caprola, binding of M13 to cpHaloΔ in the absence of Ca^{2+} has an even stronger stabilizing effect than binding of Hpep to cpHaloΔ in the presence of Ca^{2+}. This could potentially be due to the ability of M13 to occupy the ligand entry channel mediated by Arg^{17}.

The M13–cpHaloΔ structure leads to some insights on the mechanism of Caprola, but the structure of the full protein remained to be elucidated. Hence, crystallization of Caprola mimic, consisting of HaloTag with an insertion of CaM–M13 at position 155, was attempted. This protein largely resembles the Caprola architecture in the closed state. However, the C-terminal cpHaloΔ fragment is linked directly to the Hpep, where usually the termini of Caprola are located and the construct does not feature the circular permutation linker. In addition the rigid Pro_{30} was exchanged for a short and flexible $(GGS)_2$ linker. Crystals of the Caprola mimic were obtained and its structure was solved via an in-house collaboration, which allowed to create a model of Caprola by removing the direct link between cpHaloΔ and the Hpep (fig. 8C).

The structure showed the expected fold of cpHaloΔ in complex with the Hpep as well as calmodulin bound to four Ca^{2+} ions with its EF-hands wrapped around the M13 peptide. The short linkers between the split-HaloTag fragments and CaM–M13 seem to fit well, with no pronounced excess in length and resulting floppiness observed. In addition the structure revealed an interface between cpHaloΔ and calmodulin potentially mediated by hydrogen bonds or charged interactions involving residues Lys^{160}, Asp^{164} and Arg^{204} of cpHaloΔ (numbers in standard HaloTag numbering, fig 8D). To rule out that this interface is a crystallization artifact, these residues were mutated to either an alanine or to amino acids invert their native charge. All mutants showed impaired labeling of Caprola in presence of Ca^{2+} (fig. S19), corroborating the importance of this interface for the optimal folding of Caprola upon Ca^{2+} binding.

4.2.4 Reversibility of Caprola activation

While labeling of active Caprola is permanent, allowing it to function as a molecular recorder, the Ca^{2+}-induced activation of the protein should be a reversible process. This ensures that exposure to waves of increased Ca^{2+} concentrations occurring prior to the the recording window do not lead to premature labeling. Reversibility of the Caprola activation was tested in an in vitro assay using alternating additions of Ca^{2+} and the Ca^{2+}-chelator EGTA while monitoring the labeling reaction via fluorescence polarization. When Ca^{2+} was added, it triggered the onset of the labeling reaction. In return, addition of EGTA halted the labeling reaction by rapidly sequestering free Ca^{2+} ions from the solution, demonstrating the reversibility of Caprola activation. Starting and stopping the reaction could be successfully executed multiple times (fig. 9A).

It should be noted that when the reaction is stopped by adding EGTA, the fluorescence polarization signal does not remain completely stable but decreases slightly (fig. 9A). This could be due to the dissociation of Hpep from labeled cpHaloΔ after Caprola opens in the absence of Ca^{2+}. This could lead to an increase in flexibility of the covalently bound fluorophore ligand, and a change in hydrodynamic radius of the entire protein due to the change in shape. Moreover, there could be an additional contribution to the fluorescence polarization decrease by non-covalently bound ligands, that dissociate from cpHaloΔ after

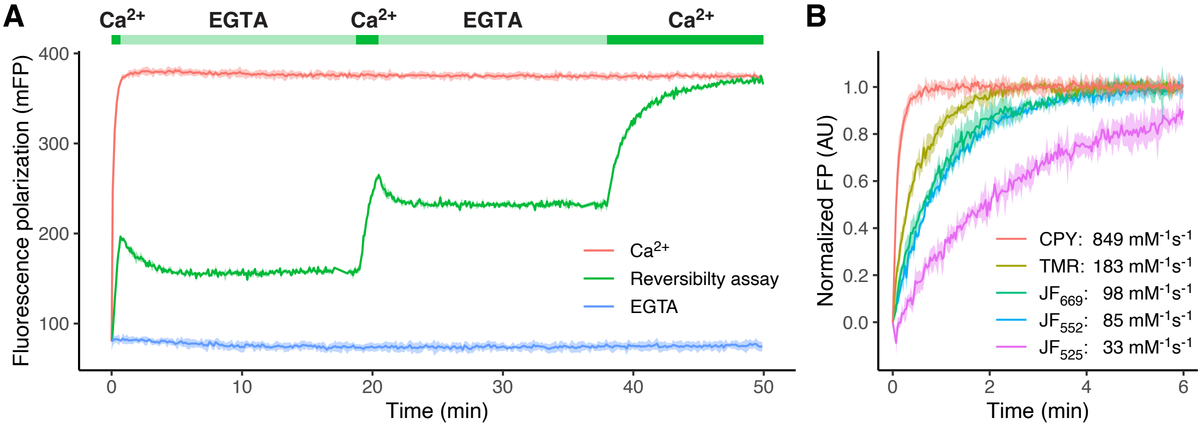


Figure 9: Reversibility of Caprola activation and Caprola ligand dependency. (**A**) Caprola (200 nM) shows fast labeling kinetics with HTL-TMR (50 nM) when saturated with Ca^{2+} (5 mM), low background activity with the Ca^{2+}-chelator EGTA (100 µM) and can be repeatedly activated and deactivated by successive epochs of Ca^{2+} and EGTA exposure. This highlights the reversible nature of Ca^{2+}-induced Caprola activation. (**B**) Labeling kinetics and apparent second-order rate constants of Caprola (200 nM) with various fluorescent ligands (50 nM) saturated with Ca^{2+} (5 mM). Fluorescence polarization values were normalized to the unbound and fully bound state (Normalized FP).

the EGTA induced opening of Caprola.

The number of activation and inactivation cycles that can be demonstrated in this assay is limited by the solubility of Ca^{2+} and EGTA and the fact that for complete removal of free Ca^{2+} a high excess of EGTA is required. It was later shown by Magnus Huppertz that Caprola can be exposed to the natural Ca^{2+} spiking of cultured neurons for several days and still be fully functional once recording is started by HaloTag ligand addition.

4.2.5 pH dependency of Caprola labeling

I investigated the effect of pH variations (pH 6.7–8.2) on the labeling reaction of Caprola at saturating Ca^{2+} concentrations, as previous assays were exclusively conducted at pH 7.3 (fig. S20A). Only minor differences in labeling rates were observed at different pH values, with low pH leading so slightly slower labeling (fig. S20B). No increased background activity in absence of EGTA was detected at any pH tested. Within the physiological intracellular pH of neurons between 7.03 and 7.46[220], Caprola showed robust and consistent labeling kinetics, which should render it applicable in different cellular environments.

4.2.6 Ligand dependency and sequential multi-color labeling of Caprola

The versatility of HaloTag technology and the availability of HaloTag ligands across the visible and near-infrared spectrum could enable the use of Caprola to record Ca^{2+} responses across multiple epochs in a multi-color experiment. By using sequential exposure to spectrally distinguishable probes, multiple activity patterns caused by different conditions or stimuli could be recorded within one specimen.

So far, characterization of Caprola has been conducted using HTL-TMR, due to robust performance of this dye in plate reader assays and availability of the ligand. I tested weather Caprola can be efficiently labeled with a set of four additional, spectrally distinct rhodamine-derived ligands, for which I had preliminary indications that they perform well for in vivo labeling[159] (fig. 9B). HTL-CPY demonstrated five-fold faster Ca^{2+}-dependent labeling compared to HTL-TMR, which is consistent with its five times faster labeling kinetics with HaloTag discussed in chapter 3. HTL-JF_{669} and HTL-JF_{552} also showed fast labeling rates, albeit slower in comparison to HTL-TMR. The labeling reaction with HTL-JF_{525} was approximately six times slower compared to HTL-TMR, rendering this ligand less favorable for applications involving Caprola. None of the tested dyes exhibited increased Caprola labeling in the absence of Ca^{2+} during the assay (fig. S21).

The ability of Caprola to provide recordings of distinct Ca^{2+} exposures over sequential time periods in different colors was demonstrated in an in vitro proof-of-concept experiment (fig. 10). Caprola protein was immobilized via its His_{10}-tag on an IMAC column and consecutively exposed to HTL-JF_{552} and HTL-CPY. Dyes were either applied in either presence or absence of Ca^{2+}. The labeling patterns revealed by SDS-PAGE were coherent with the Ca^{2+} exposure of Caprola, indicating Caprola's ability to record variable Ca^{2+} exposures in different colors across multiple time intervals.

In summary, HTL-CPY, HTL-JF_{669} and HTL-JF_{552} are three spectrally distinguishable ligands that have proven to be compatible with Caprola. These ligands could potentially be utilized to record Ca^{2+} waves in up to three successive epochs of neuronal activity in vivo. The differences in labeling speed between these ligands could possibly be compensated by adjusting the administered concentration. However, in an in vivo setting other parameters such as delivery, distribution, metabolism and excretion

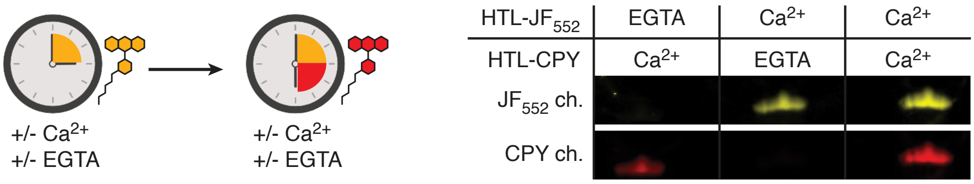


Figure 10: Multi-color in vitro recording of sequential Ca^{2+} events. Demonstration of the capacity of Caprola$_9$ to record Ca^{2+} exposure during successive time intervals in different colors. The protein was immobilized on an IMAC column and exposed to HTL-JF$_{552}$ during a first time window and to HTL-CPY during a second time window. Dye exposure was accompanied by either the presence or absence (EGTA) of Ca^{2+} (scheme, left). After elution, labeled proteins were analyzed by SDS-PAGE followed by fluorescence scanning (right). The observed labeling patterns align with the Ca^{2+} exposure of Caprola.

of each ligand might be more critical than the differences in labeling speed. It is likely that for each model system dye selection and concentrations have to be carefully optimized.

4.2.7 Calcium Sensitivity of Caprola

Using Caprola to record Ca^{2+} responses across different cell types, organelles and species will require the ability to adjust the Ca^{2+} sensitivity of Caprola. It has been shown for other CaM–M13-based sensors that introducing point mutations in the M13 peptide allows for such tuning of the Ca^{2+} response. For example, Moeyaert et al. described modifications that alter the Ca^{2+} affinity of the light-gated Ca^{2+} recorder CaMPARI2 from 145 nM to 2107 nM[64]. By applying these mutations I created 15 Caprola variants (Caprola$_1$–Caprola$_{15}$) and measured their labeling kinetics at defined free Ca^{2+} concentrations, using an EGTA / Ca·EGTA buffer system (fig. 11A, S22A). This data allowed to calculate the half-maximal effective Ca^{2+} concentration EC_{50}^{Ca2+}, at which a variant reaches its half-maximal labeling rate, for all 15 variants (fig. 11B, C, S22B, table S10). EC_{50}^{Ca2+} values for the 15 Caprola variants ranged

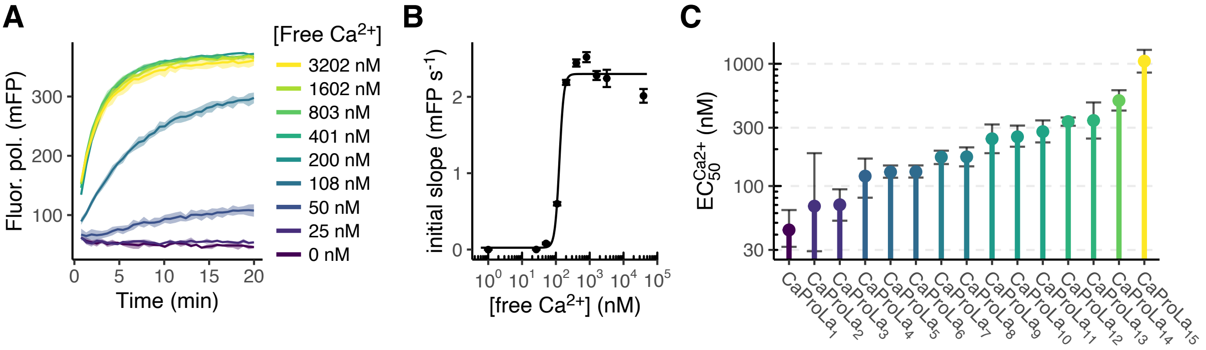


Figure 11: Ca^{2+}-sensitivity of Caprola. (A) Exemplary labeling reactions of Caprola$_9$ (200 nM) with HTL-TMR (50 nM) at different free Ca^{2+} concentrations monitored by fluorescence polarization. Labeling rates are dependent on the free Ca^{2+} concentration. (B) Initial slopes of the Caprola$_9$ labeling reactions from (A) plotted against the free Ca^{2+} concentrations. A four-parameter logistic model was fit to the data to determine the Ca^{2+}-sensitivity (EC_{50}^{Ca2+}). The steep hill slope of the fit model indicates strong cooperativity of the Ca^{2+} mediated Caprola activation. The EC_{50}^{Ca2+} represents the Ca^{2+} concentration at which Caprola$_9$ exhibits half of its maximum labeling speed. (C) EC_{50}^{Ca2+} values of the 15 Caprola variants featuring different M13 mutations as determined in (B). Values range from 44 nM to 1050 nM. Error bars represent 95 % confidence intervals.

from 44 nM to 1050 nM, which should render Caprola applicable in many different cellular contexts and models.

Most neurons have a cytosolic Ca^{2+} concentration of approximately 75 nM at rest which can increase up to the micromolar range upon electrical activity[221–223]. Many of the engineered Caprola variants feature an EC50 between these two extremes and should be useful for the recording of Ca^{2+} transients evoked by different intensities of neuronal activity.

Moreover, the Ca^{2+} response of Caprola exhibited a sharp, switch-like behavior at the EC_{50}^{Ca2+} with an average Hill slope of 3.9. This behavior is typical for CaM–M13-based sensors, due to the cooperative binding of up to four Ca^{2+} ions by calmodulin[224,225]. This effect is advantageous for Caprola for two main reasons: firstly, it ensures minimal activity at the cellular resting Ca^{2+} concentration; and secondly, upon an increase in Ca^{2+} concentration, Caprola quickly reaches high activity and efficiently records the Ca^{2+} transient.

4.2.8 Characterization of final Caprola variants

Labeling rates under saturating Ca^{2+} concentrations (k_{app}^{Ca2+}) and background labeling activity in absence of Ca^{2+} (k_{app}^{EGTA}) were assessed for all 15 final Caprola variants (fig. 12, S23, table S10). All variants featured fast labeling kinetics in presence of Ca^{2+} with an average k_{app}^{Ca2+} value of $1.56 \cdot 10^5$ M^{-1} s^{-1} and minimal background labeling with an average k_{app}^{EGTA} value of 41.1 M^{-1} s^{-1}. The ratio $k_{app}^{Ca2+}/k_{app}^{EGTA}$ denotes the relative increase in Caprola labeling rate upon Ca^{2+} exposure. The 15 final Caprola variants had an average $k_{app}^{Ca2+}/k_{app}^{EGTA}$ ratio of $4.25 \cdot 10^3$, which should facilitate excellent signal-to-noise ratios for in cellulo and in vivo applications of Caprola.

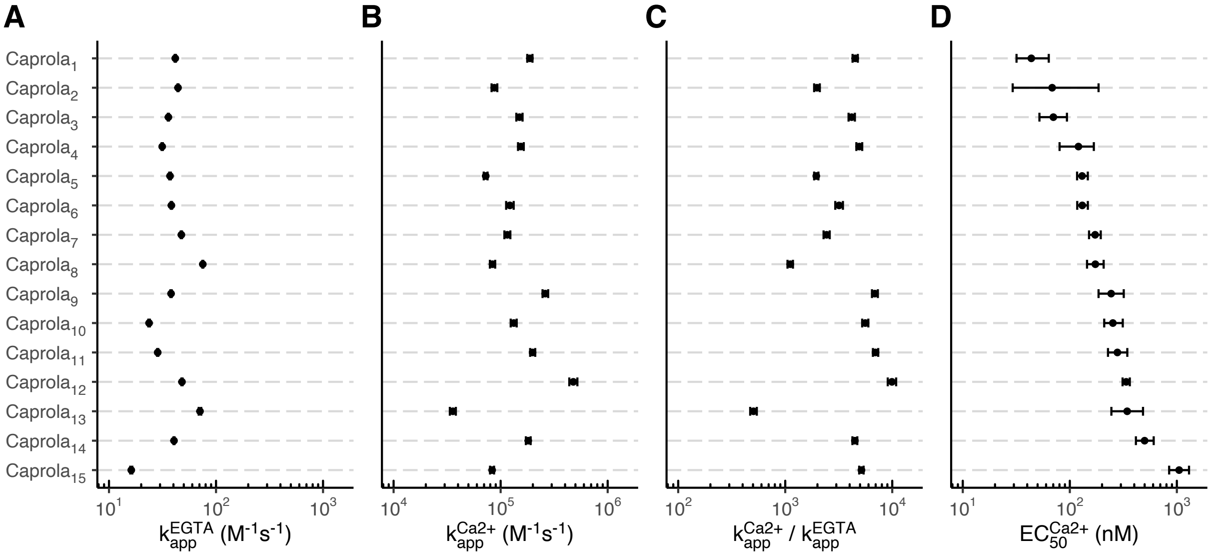


Figure 12: Overview of final Caprola variants. (**A, B**) Apparent second-order rate constants of the labeling reaction with HTL-TMR in presence (k_{app}^{Ca2+}) or absence (k_{app}^{EGTA}) of Ca^{2+}. (**C**) Relative increase in labeling rate of Caprola upon Ca^{2+} exposure ($k_{app}^{Ca2+}/k_{app}^{EGTA}$). (**D**) Half-maximal effective Ca^{2+} concentrations (EC_{50}^{Ca2+}). Error bars represent 95 % confidence intervals.

4.3 Application of Caprola

in, suggesting nonuniform The Caprola project documented in this thesis was an inherently collaborative project. While my role revolved around the creation and optimization of the tool, Magnus Huppertz, a fellow PhD student in the Johnsson group, validated Caprola in cultured cells and subsequently applied it in both, cell culture and in vivo. To provide a perspective on how Caprola enables new insights in neuro- and cancer-biology, this section will briefly describe the experiments conduced by Magnus Huppertz in collaboration with the groups of Herwig Baier, Lisa Fenk (MPI for Biological Intelligence, Munich), Claudio Acuna Goycolea (Heidelberg University) and Frank Winkler (University clinic, Heidelberg & DKFZ, Heidelberg).

4.3.1 Validation in cultured cells

Before using Caprola to record Ca^{2+} spikes in vivo, the tool was validated in cultured cells. To be able to normalize signals from Caprola labeling, the protein was C-terminally fused to EGFP as an expression control. Thapsigargin-induced increases of cytosolic Ca^{2+} in cultured HeLa cells were reliably recorded by Caprola and could be read out via fluorescence microscopy and flow cytometry. These results could be confirmed with a palette of fluorescent HaloTag ligands and different Caprola versions. Caprola was then expressed in cultured primary hippocampal rat neurons, to investigate if it facilitates recording of Ca^{2+} waves induced by external electrical stimulation. Indeed, robust labeling of caprola was observed after as little as 10 action potentials delivered via an electrical field stimulation setup.

4.3.2 Recording neuronal activity in living flies

The capability of Caprola to record neuronal activity in vivo was demonstrated by recording Ca^{2+} transients in the visual system of *Drosophila melanogaster*. Caprola$_5$-EGFP was expressed in T4 and T5 neurons, which are known to have a direction-selective response to local visual motion perceived by the ipsilateral eye[226,227]. Tethered flies were subjected to visual stimulation targeting the left, the right, or both eyes by exposing them to moving sinusoidal gratings for a period of 20 min, in the presence of HTL-CPY (fig. 13A). The HaloTag ligand was administered through a minimally invasive incision made at the posterior region of the head. When both eyes were stimulated, analogous signals were observed in both optic lobes. When the visual stimulation was unilateral, stronger signals were observed on the ipsilateral side (fig. 13B). These experiments show that Caprola is capable of selectively recording neuronal activity in *Drosophila melanogaster* within a concise timeframe of 20 min.

4.3.3 Recording neuronal activity in living zebrafish

Zebrafish (*Danio rerio*) lines were generated which express Caprola$_1$-EGFP pan-neuronally. These larvae (4–6 dpf) could be effectively stained with several HaloTag ligands: HTL-CPY, HTL-JF$_{669}$, HTL-JF$_{552}$ and HTL-JF$_{525}$. Labeling was initially observed in the forebrain at 30 min and later in the hindbrain at approximately 60 min, suggesting nonuniform absorption of the fluorescent ligands. Hence, comparisons of Caprola labeling have to be made within specific brain regions due to the different exposure to the ligands in different areas of the brain.

To record distinct neuronal activity patterns in vivo, zebrafish larvae were embedded in agarose and

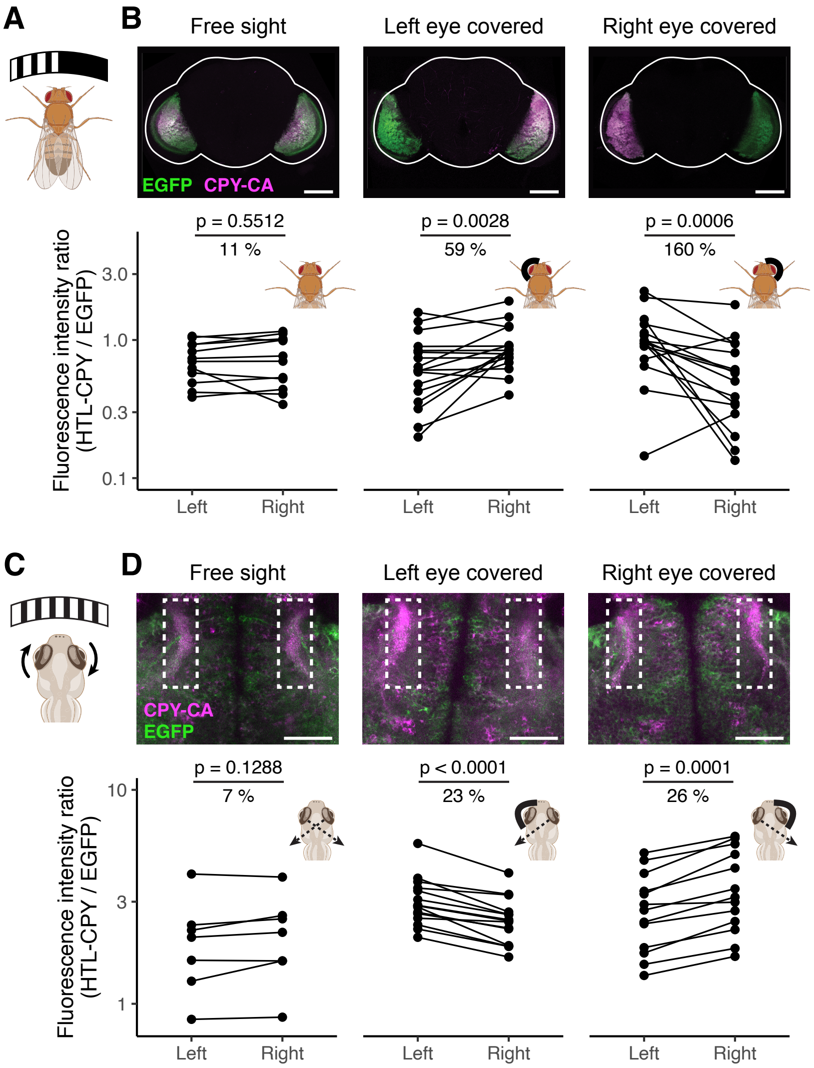


Figure 13: In vivo recording of neuronal activity. (**A**) Cartoon of the experimental set-up for visual stimulation of flies. A curved monitor was positioned in front of adult flies expressing Caprola$_5$-EGFP to selectively stimulate both or only one of the eyes using moving sine wave gratings. (**B**) Asymmetrical visual stimulation of flies recorded with Caprola$_5$-EGFP in presence of HTL-CPY (5 μM, 20 min, N ≥ 12 individuals per condition). (**C**) Cartoon of the experimental set-up 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. (**D**) 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 HTL-CPY (5 μM, 6 h, N ≥ 7 larvae per condition). Data was recorded and analyzed by Magnus-Carsten Huppertz, Martin W. Schneider and Tjalda Falt.

subjected to visual stimulation (fig. 13C), inducing eye movements (optokinetic response)[228]. Post 6 h, analysis of the arborization field 9 (AF9) revealed distinct labeling of neurons. AF9 receives direct input from the retina and exhibits sensitivity to alterations in ambient light, like whole-field motions[229,230]. Obscuring one eye with opaque agarose resulted in decreased labeling in the contralateral AF9, in alignment with the contralateral processing of visual inputs in zebrafish (fig. 13D).

To showcase the ability of Caprola to record neuronal activity during two successive epochs of visual stimulation in different colors, initial stimulation was conducted with one eye obscured, followed by a repetition the subsequent day with the alternate eye obscured. Both periods of asymmetric visual stimulation were captured by incubation with spectrally distinguishable fluorescent ligands (HTL-CPY and HTL-JF_{669}). Subsequent analysis showed the expected asymmetric labeling patterns for each fluorophore, underscoring the capability of Caprola to be applicable for sequential multi-color labeling in vivo.

The initiation of Caprola recording is achieved chemically, by addition or injection of a fluorescent ligand. This eliminates the necessity to attach the animal to a device like a microscope, which is required for most optical methods, and hence enables recording of neuronal activity in freely moving and naturally behaving animals. As a proof of concept, neuronal activity was recorded in non-embedded zebrafish larvae during prey hunting, induced by addition of paramecia[231]. It is known that the visual perception of prey activates neurons in the AF7 neuropil[230,232]. Subsequent analysis did indeed reveal increased labeling in this area compared to control fish without exposure to prey. These findings illustrate that Caprola enables the direct recording of neuronal activity in freely moving animals.

4.3.4 Analysis of heterogeneous cell population

Caprola was used to sort heterogenous cell populations according to the intensity of their Ca^{2+} signaling for subsequent analysis of their transcriptional profiles via RNA sequencing. This was performed with a patient-derived glioblastoma cell line, a prevalent and aggressive primary brain tumor in adults. These cells organize in functional networks that communicate via intracellular Ca^{2+} waves[233,234]. The study was motivated by the hypothesis that driving factors of elevated Ca^{2+} transients in these cells might present potential therapeutic targets.

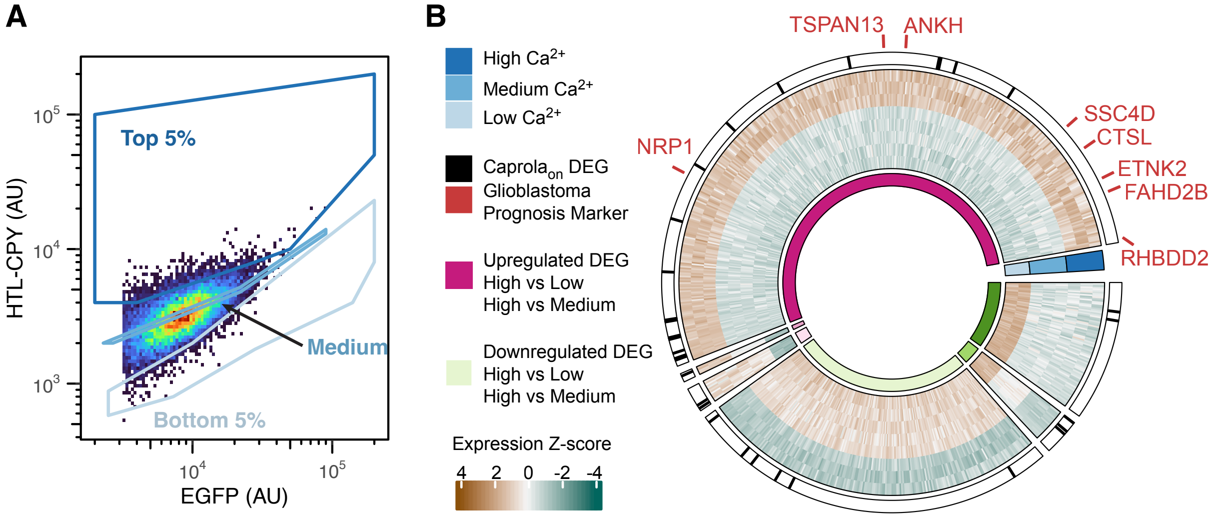


Figure 14: Transcriptomic profiling of glioblastoma cells sorted by Ca^{2+} levels. (A) Gating strategy of CPY-labeled (125 nM, 90 min) glioblastoma cells expressing $Caprola_6$-EGFP for later RNA sequencing analysis. AU: arbitrary unit. (**B**) Transcriptional profiles of the three sorted groups from (A). DEGs identified by RNA sequencing analysis are color-coded according to Z-scores. DEGs were arranged in six groups according to their pairwise comparison pattern (e.g., upregulated high vs. low and high vs 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)x Data was recorded by Magnus-Carsten Huppertz, David Hausmann and Dirk C. Hoffmann and analyzed by Ling Hai.

Glioblastoma cells expressing Caprola$_6$-EGFP were cultured and marked with HTL-CPY depending on their intracellular Ca^{2+} concentrations, before being sorted via fluorescence activated cell sorting (FACS) into three categories: high, medium and low Ca^{2+} (fig. 14A). RNA sequencing and transcriptomic analysis revealed 757 differentially expressed genes (DEGs) between the three pools (fig. 14B). A control experiment was conducted with a constitutively active (non-Ca^{2+}-dependent) Caprola variant, termed Caprola$_{on}$, to rule out non-Ca^{2+} mediated effects some genes might exhibit, e.g. affecting HaloTag ligand permeability. This lead to exclusion of 31 of the 757 originally identified DEGs.

Gene ontology (GO) term analysis revealed that the remaining 726 DEGs were primarily associated with cell division and neurogenesis and eight of the DEGs which were exclusively upregulated in the high Ca^{2+} pool, were known unfavorable prognostic markers. These findings illustrate that heterogeneous cell populations can be categorized into groups with distinct transcriptomes based on their Ca^{2+} levels, demonstrating Caprola's potential in studying medically relevant phenotypes.

4.4 Conclusion and Outlook

A split-HaloTag was developed by circular permutation of the protein and subsequent excision of a small internal α-helix, termed Hpep. The remaining protein, termed cpHaloΔ maintains its three dimensional structure but exhibits negligible labeling activity. However, covalent labeling activity can be restored upon reversible binding to the Hpep. The affinity between cpHaloΔ and the Hpep is low, with a K_D in the millimolar range, which ensures that the two components do not spontaneously complement each other when expressed in the same system. Thorough characterization of the split system revealed that in its unbound form, cpHaloΔ demonstrated a drastic decrease in both, ligand binding affinity and rate of covalent bond formation. Both parameters increase substantially upon the binding of Hpep, leading to a 61 000-fold difference in the overall labeling speed when compared to its absence. While the labeling speed of the complemented protein does not match that of the original HaloTag, half-labeling times are still in the minutes range at nanomolar ligand concentrations.

The split-HaloTag system was then used for the design of a molecular Ca^{2+} recorder named Caprola. This was accomplished by fusing the split-HaloTag fragments to the Ca^{2+}-sensing domains, calmodulin and M13. In the absence of Ca^{2+}, Caprola remains inactive, but it regains its labeling activity upon binding to Ca^{2+}. The Ca^{2+}-induced activation is reversible, and features a sharp threshold due to the cooperative binding of four Ca^{2+} ions to calmodulin. Remarkably, CaProLa's labeling rates surpass those of the unmodified split-HaloTag which might be due to the decreased flexibility at the split junctions after formation of the CaM–M13 complex. A panel of 15 Caprola variants was engineered which feature different Ca^{2+} sensitivities, spanning from the low nanomolar to the micromolar range. This should render Caprola applicable in many different cell types and environments with different Ca^{2+} dynamics.

In a subsequent project led by Magnus Huppertz, Caprola was used to record distinct patterns of neuronal activity in the visual systems of flies and zebrafish after specific visual stimulations. The use of spectrally distinguishable fluorescent HaloTag ligands enabled to record different neuronal activity patterns in a singular organism, in different colors. Importantly, recordings were also feasible in freely moving animals, highlighting the non-intrusive nature of the tool. Moreover, the applicability of Caprola was not restricted to recording neuronal activity. The tool was employed to sort in vitro network-forming and heterogeneous glioblastoma cells based on their Ca^{2+} levels. This enabled subsequent RNA sequencing analysis of the subpopulations and lead to the discovery of 726 differentially expressed genes among the groups. Correlating gene expression profiles to Ca^{2+} phenotypes, could potentially lead to a better understanding of the molecular mechanisms of the tumor network which could assist for the development of new therapeutic strategies.

I expect Caprola to be a useful tool for generating large-scale activity profiles with single cell or neuron resolution as it uniquely combines several beneficial traits. It offers direct recording of Ca^{2+} signals, bypassing the need for indirect coupling to downstream gene expression. It is also compatible with freely moving animals, due to minimal invasiveness. Additionally, its capacity to record multiple activity patterns within one individual under different conditions could help to overcome challenges and better understand inter-individual differences[235]. Future applications of Caprola could also include recording Ca^{2+} transients with subcellular precision by targeting the protein to specific organelles or structures. For instance, targeting Caprola to pre- or post-synaptic sites could enable mapping of neuronal inputs or outputs and contribute to our understanding of the mechanisms of neuronal circuits.

While split-HaloTag and Caprola has showcased significant potential, some aspects of the tool could benefit from further improvements. A potential issue is the stability of the cpHaloΔ protein. When incorporated in Caprola, its stability was satisfactory, however isolated the protein displayed melting temperatures below the physiological temperature of 37 °C, which might complicate generalization of the molecular recorder approach. Moreover, the low affinity between the split components might restrict their utilization to complexes like CaM–M13 that feature a small distance between their termini. Hence, enhancing robustness, speed and tunability in terms of intrinsic affinity could be important steps to improve the versatility of the tool. Efforts to address these challenges are discussed in the following chapter of this thesis.

In vivo labeling of Caprola still presents limitations. For instance, in zebrafish experiments, incubation periods of several hours were required to observe clear activity patterns. In addition, labeling was not uniform throughout the brain, making comparisons between certain brain regions impossible. A comprehensive study of HaloTag ligand distribution in different model organisms could help to identify better suited ligands. Such analysis and might enable finding overarching principles influencing bioavailability, which could inform the design of tailored ligands. A complementary approach would be to interfere with the active export of rhodamine dyes across the blood-brain barrier mediated by multidrug resistance proteins like P-glycoprotein[236]. Indeed, a recent report of a zebrafish knockout model of *abcb4* (a P-glycoprotein homolog), showed elevated accumulation of rhodamine 123 in the brain[237]. However, when inhibiting multidrug resistance proteins either pharmacologically or via knock-out/down, potential secondary effects should be carefully considered and monitored.

Nevertheless, good signals were observed in zebrafish brains after shorter incubation times below one hour when using control constructs with constitutively active cpHaloTag. This suggests that the tissue penetration of the ligands is not the only limiting factor, but also Caprola labeling rates at the achieved ligand concentrations could be improved to enable recordings in shorter timeframes. This could go hand in hand with the improvements of the split-HaloTag system proposed above. Another relevant factor beyond overall labeling speed of Caprola is the speed at which Ca^{2+} induced CaM–M13 binding and subsequent spilt-HaloTag complementation occur. As Ca^{2+} spikes caused by neuronal signal transduction are often in the millisecond time scale[218], rapid activation kinetics of Caprola are a prerequisite for efficient recording of these events. The rate of Ca^{2+} induced activation should be characterized and possibly optimized if they are a limiting factor, e.g. by incorporating mutations from fast CaM–M13-based sensors[30].

Furthermore it would be attractive to use Caprola, or an improved version of it, in other in vivo models such as mice or rats which exhibit more complex behaviors and offer a broader range of relevant disease models. A suitable proof-of-concept experiment could be carried out in the primary somatosensory cortex. In this brain region, there are spatially organized neuronal populations, termed "barrels", responsive to specific individual whiskers[238]. By removing one or multiple whiskers, certain populations no longer receive input and, consequently, their activity significantly decreases. This paradigm offers a straightforward way to showcase the functionality of Caprola, particularly its ability to record successive periods of neuronal activity, in rodents.

Beyond these initial experiments, the scalability and other advantages of Caprola cold facilitate capturing brain-wide activity data from freely behaving rodents, potentially revealing novel neural pathways and disease-related mechanisms. It will be intriguing to see how the capacity of Caprola to record Ca^{2+} activity can be integrated with different systems neuroscience approaches to deepen our understanding of the brain.

5 Engineering of a Reporter for Neuronal Connectivity and Plasticity

Understanding neuronal circuits and how they enable the brain to process complex information requires – among other aspects – two key facets: high-resolution structural mapping of the architecture of the neuronal network down to the synapse level, and functional recordings of the network's activity under varying conditions. The preceding chapter presented a molecular tool developed to facilitate brain-wide functional recordings while this chapter shifts focus to a novel chemo-genetic approach to study neuronal connectivity and plasticity.

Synapses are fundamental to neuronal communication, acting as junctions where neurons transfer signals. However, detecting these microscopic structures in tissues poses significant challenges due to their small size (300–500 nm, with ~20 nm distance at the synaptic cleft)[70,71] and high density, with up to one billion synapses per cubic millimeter in the cortex[70]. Historically, mapping synapses in brain tissues has been achieved with serial section electron microscopy (SSEM), which offers unparalleled spatial resolution[72,77,239–241]. Yet, application of SSEM to reconstruct neuronal circuits is limited by a lack of contrast distinguishing different cell types and structures like synapses. This necessitates elaborate data analysis involving years of work of large consortia[77], which constrains the scale and throughput that can be achieved with this method, especially in larger specimens[242].

Fluorescence microscopy has emerged as a complementary technology, offering the capacity to target specific cell types and subcellular structures which can then be imaged in spectrally distinct channels. The introduction of advanced fluorescence imaging technologies like expansion and lattice light-sheet microscopy (ExLLSM)[42] and comprehensive analysis of tissues across scales (CATS)[91] has spurred advancements in this field, overcoming prior limitations by enhancing resolution and scale. However, accurately localizing the synapses connecting two neurons remains challenging, typically requiring the detection of the co-localization of pre- and post-synaptic markers.

I propose a novel chemo-genetic approach for labeling of synapses using split-HaloTag technology. By fusing split-HaloTag fragments (cpHaloΔ and Hpep) to proteins that interact across the synaptic cleft (e.g. NRX and NLG)[243,244], split complementation and the subsequent labeling with fluorescent ligands would be strictly confined to synaptic sites (fig. 15A). This method for <u>syn</u>apse specific <u>pro</u>tein <u>la</u>beling (SynProLa) could offer multiple advantages including genetic targetability and strong, red-shifted signals from synthetic organic dyes, which are compatible with demanding super resolution imaging technologies. Furthermore, the approach should feature high specificity, due to the dependency on trans-synaptic protein-protein interactions instead of mere co-localization, and introduce minimal bias due to the low affinity and reversibility of split-HaloTag.

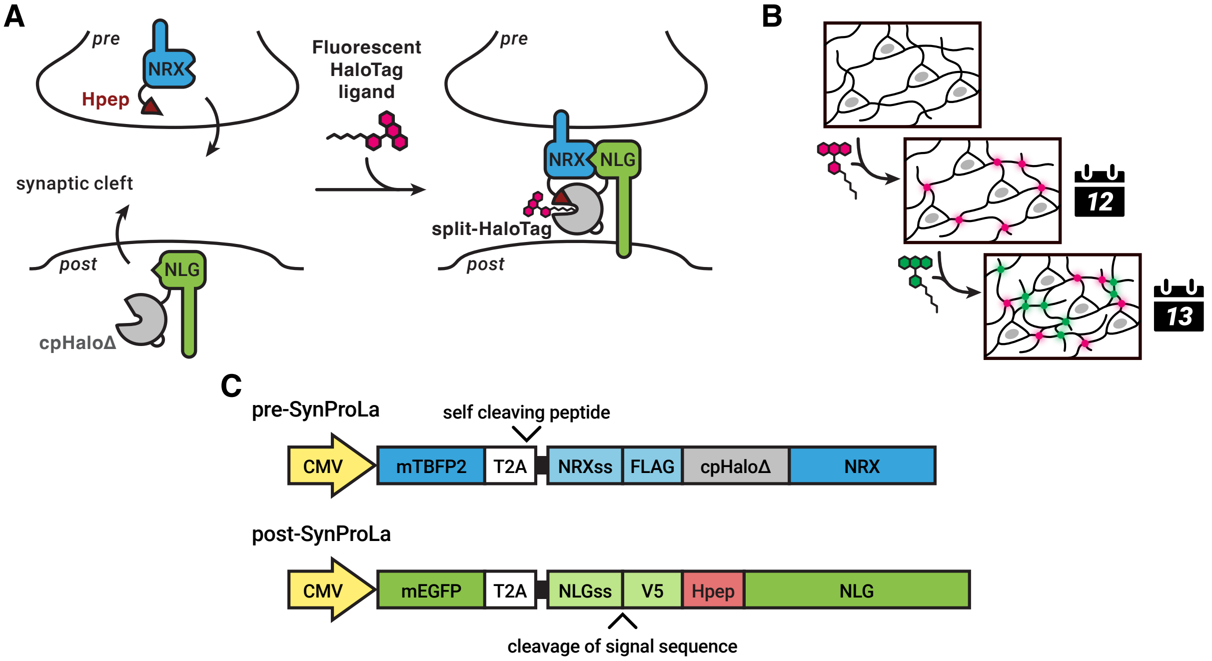


Figure 15: SynProLa concept and design. (**A**) Concept of split-HaloTag-based synapse specific protein labeling (SynProLa). Split-HaloTag fragments cpHaloΔ and Hpep are fused to neurexin-3-β (NRX) and neuroligin-1 (NLG) proteins, which form a trans-synaptic complex. (**B**) Concept of multi-color SynProLa labeling to reveal synaptic plasticity an synaptogenesis. (**C**) Genetic architecture of pre- and post-SynProLa constructs.

Importantly, the use of spectrally distinct fluorescent HaloTag ligands for synapse labeling in a sequential procedure could enable the identification of newly formed synapses or synapses that increase in size (fig. 15B). This would enable mapping of synaptogenesis and synaptic plasticity on a tissue or organ wide scale, which is not possible with currently available methods. Availability of such data might lead to new discoveries in neuroscience and a better understanding of previously unclear mechanisms and processes.

5.1 First Generation SynProLa Constructs

5.1.1 SynProLa design and initial trials

For the creation of split-HaloTag-based synapse specific protein labeling (SynProLa), I decided to utilize the neurexin–neuroligin (NRX–NLG) complex, a central trans-synaptic adhesion complex[245]. This decision was motivated by the high-affinity interactions of neurexins and neuroligins, the availability of structural data[246–248] and the proven success of other split protein approaches using this complex[125,249]. Both, NRX and NLG present their C-termini to the inside of the neuron and their N-termini to the synaptic cleft. Consequently, the split-HaloTag fragments cpHaloΔ and Hpep were inserted after the N-terminal signal peptides of human neurexin-3-β (isoform 2b) and murine neuroligin-1 (isoform 1) to create pre- and post-SynProLa constructs (fig. 15C). Different constructs were generated coupling either Hpep or cpHaloΔ to NRX or NLG using flexible $(GSG)_{2\text{–}4}$ linkers of varying length.

SynProLa constructs were evaluated utilizing a cellular co-culture assay with HEK-293 Flp-In™

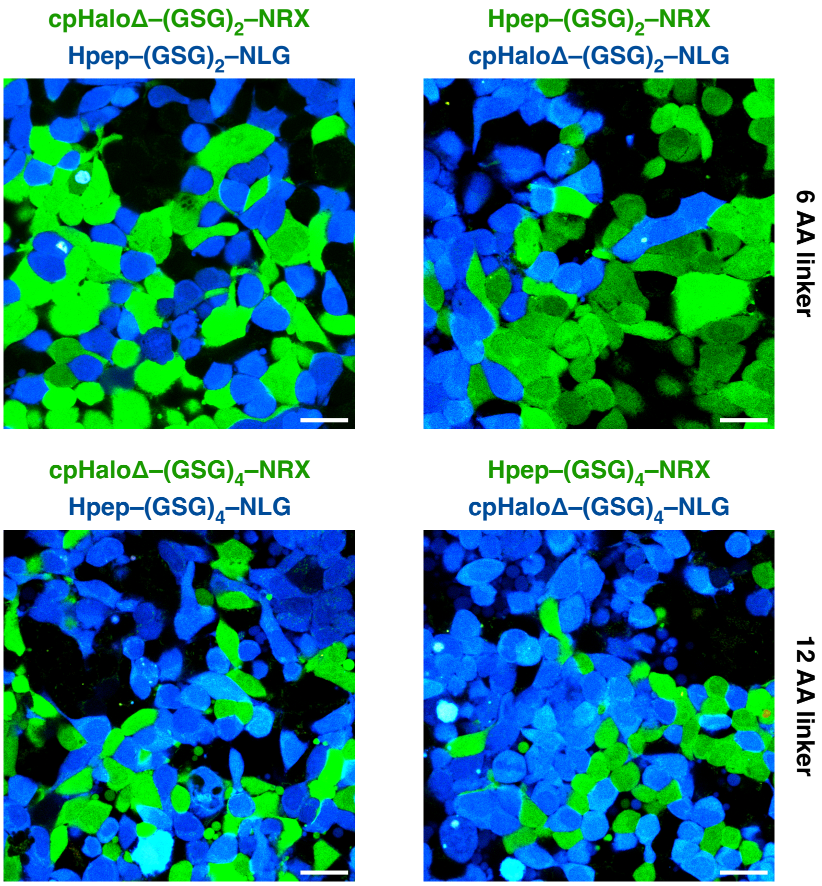


Figure 16: HEK-293 co-culture assay with first generation SynProLa constructs. Confocal microscopy images of co-cultured HEK-293 Flp-In™ T-REx™ cells stably expressing first generation SynProLa constructs with different orientations and linkers. Cells were incubated with HTL-CPY (100 nM, 1 h) prior to PFA fixation and imaging. Pre- and post-SynProLa expressing cells can be clearly distinguished by the co-expressed fluorescent proteins (mTagBFP2: blue, EGFP: green). However, no labeling of cell-cell contact sites was observed. AA: amino acid, scale bars: 25 µm.

T-REx™ cells. It has been shown that NRX–NLG complexes are formed at the cell-cell interface of co-cultured HEK-293 cells, expressing either NRX or NLG[106]. To facilitate the identification of NRX or NLG expressing cells, fluorescent proteins (mTagBFP2[250] and EGFP[251], respectively) followed by three copies of the self cleaving peptide T2A[252] were incorporated into the constructs described above. In addition epitope tags (FLAG-tag and V5-tag) were inserted before the split-HaloTag parts to enable detection via immunofluorescence. Constructs were expressed separately under CMV promotor in HEK-293 cells either by transient transfection or via genomic integration using the Flp-In™ system to generate stable cell lines. Pre- and post-SynProLa expressing cells were then co-cultured at high confluency to promote cell-cell contacts between them before being incubated with fluorescent HaloTag ligand and imaged via confocal fluorescence microscopy.

None of the combinations of the designed constructs enabled labeling of cell-cell contact sites between pre- and post-SynProLa cells (fig. 16). To confirm that the constructs are expressed in HEK-293 cells localized to the cellular membrane, immunofluorescence staining of the fused epitope tags was performed (fig. S24). While some protein was detected inside of the cells, most was found at at the plasma membrane, confirming full length expression and correct trafficking of the co-expression constructs. To

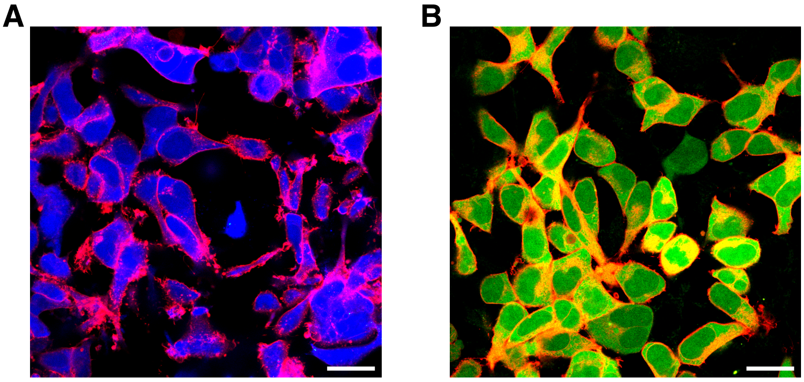


Figure 17: SynProLa cpHaloTag control constructs. Confocal microscopy images of HEK-293 cells expressing SynProLa positive control constructs, featuring the constitutively active cpHaloTag 141–145 instead of cpHaloΔ (**A**: mTagBFP2–F2A–NRXss–FLAG-tag–cpHaloTag141-145–hNRX-3-β, **B**: EGFP–F2A–NLGss–V5-tag–cpHaloTag141-145–mNLG-1). Cells were labeled with HTL-CPY (100 nM) for 1 h and PFA fixed prior to imaging. Plasma membrane localization and functionality of HaloTag labeling on NRX/NLG were confirmed. Scale bars: 25 µm.

investigate whether NRX/NLG fusions can be labeled on the cell surface with HaloTag technology, I generated positive control constructs for pre- and post-SynProLa that contain the permanently active cpHaloTag 141–145 instead of cpHaloΔ. When expressed in HEK-293 cells, the proteins can be labeled with HTL-CPY (fig. 17) and show predominantly the expected localization at the plasma membrane.

5.1.2 Structural considerations

The findings described above suggest that the inability of SynProLa to label cell-cell contact sites is most likely due to inefficient split-HaloTag complementation on NRX–NLG complexes. A critical factor for split protein complementation is how close the split fragments are to each other. Determining the precise distance between the N-termini in the NRX–NLG complex is challenging since neurexin-3-β possess a putatively disordered N-terminal 47 amino acid domain, which is not resolved in available crystal structures or AlphaFold models[253,254]. Furthermore, there is no available crystal structure of neurexin-3-β in complex with neuroligin-1. Nonetheless, the structure of neurexin-1-β in complex with neuroligin-1 (PDB-ID 3BIW)[246] can serve as a suitable reference, as it aligns very closely with the structure of free neurexin-3-β (0.495 Å RMSD, PDB-ID 3MW4)[255].

The distance between the first resolved residue of NRX and the N-terminus of NLG in the complex is approximately 56 Å (fig. 18A). Notably, this distance is substantially larger than the ~ 15 Å distance between the split-HaloTag fusion sites in Caprola. The effective concentration (C_{eff}) that is reached at a certain distance and linker length can be estimated using worm-like chain (WLC) models[256,257]. For Caprola, with a distance of ~ 15 Å and 5 linker residues the predicted C_{eff} is 24 mM. In contrast, for SynProLa with a distance of 56 Å between fusion sites and a combined length of 24 linker residues and 47 disordered residues, the predicted C_{eff} is just 0.23 mM (fig. 18B). These calculations are based on an idealized model and do not account for several factors, including the fact that the linkers in both SynProLa and Caprola are distributed across two polypeptides. Nonetheless, these estimates provide an indication why SynProLa constructs failed to be labeled. Considering that the split-HaloTag fragments

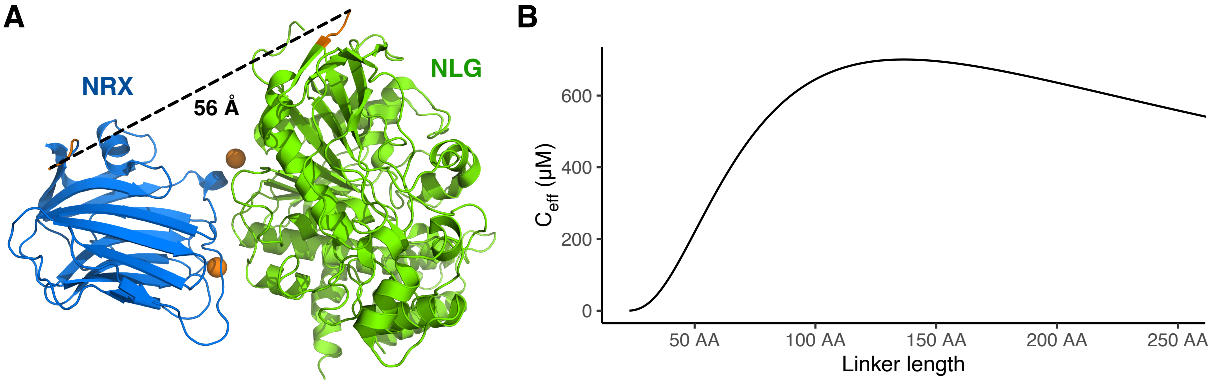


Figure 18: NRX–NLG complex structure and estimated effective concentrations. (**A**) Crystal structure of a homolog NRX–NLG complex (PDB-ID 3BIW)[246]. The distance between the extracellular N-termini (orange) is annotated. (**B**) Estimated effective concentration (C_{eff}) at 56 Å distance against flexible linker length calculated using a worm-like chain (WLC) model.[257].

have a relatively low affinity to each other with a K_D value of 4.6 mM, they might be predominantly complemented in Caprola but mostly unbound in the SynProLa complex.

5.1.3 FKBP/FRB model system

To further characterize the split-HaloTag system and corroborate the hypothesis that complementation is inefficient across larger distances, I used a model system comprising the two protein domains FKBP and FRB (fig. 19A). These two proteins interact only upon addition of the small molecule dimerizer rapamycin and can be expressed and purified from *E. coli*. When in complex, the distance between their C-termini is rather short at 15 Å, the distance between the N-termini is at 49 Å and the distances between N- and C-termini are 33 Å and 39 Å (fig. 19B). Hence the system provides an ideal model to study how distance and linker length influence split-HaloTag activity.

N- and C-terminal fusions of cpHaloΔ and Hpep to either FKBP or FRB using flexible $(GGS)_{1\text{-}3}$ linkers were generated and labeling rates in presence of the dimerizer rapamycin were measured for all possible combinations of these constructs (fig. 19C, S25, S26). Fastest labeling rates were observed when the split-HaloTag fragments were fused to the C-termini of FKBP and FRB wich are close in space at 15 Å distance. The reactions reached rate constants (k_{app}) of up to $1.0 \cdot 10^4\,M^{-1}\,s^{-1}$ which are close to the maximum labeling rate of the split-HaloTag system of $1.3 \cdot 10^4\,M^{-1}\,s^{-1}$. These findings align with the predicted C_{eff} values for this distance, ranging from 30 mM to 34 mM depending on the linker length. At this concentration it is anticipated that split-HaloTag is mostly in the complemented state. The labeling rates appeared to be mostly uninfluenced by the linker length, consistent with the little effect of linker length in the WLC model for the given distance.

For combinations where one split component is attached to the N-terminus of FKBP or FRB and the other to the C-terminus (33–39 Å distance), labeling rates were approximately one to two orders of magnitude slower and largely dependent on linker length. The predicted C_{eff} values with two $(GGS)_3$ at these distances are between 0.14 mM and 1.1 mM. This should cause most of the split-HaloTag to be in a non-complemented state, reducing the observed labeling rate. When both split parts are fused to the relatively distant N-termini of FKBP and FRB (49 Å distance), minimal labeling was observed even with the longest linkers. The predicted C_{eff} with two $(GGS)_3$ at this distance is 1.0 µM, indicating minimal

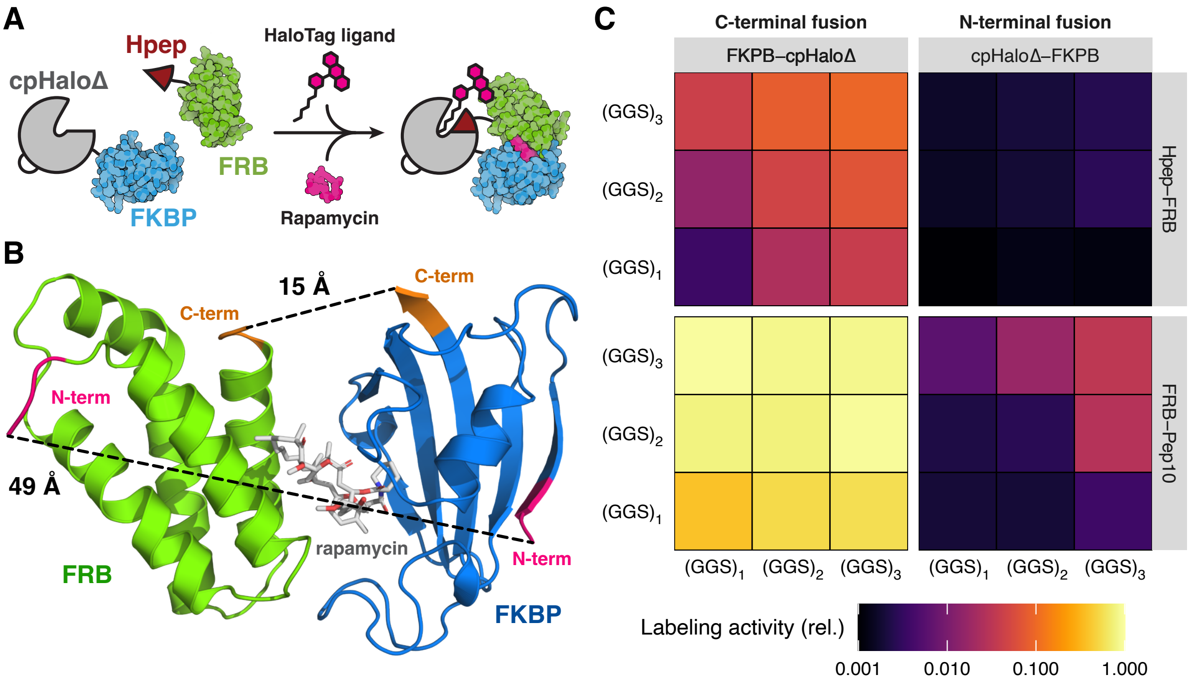


Figure 19: FKBP/FRB split-HaloTag model system. (**A**) Scheme of the FKBP/FRB split-HaloTag model system. (**B**) Structure of the FKBP–FRB–rapamycin complex (PDB-ID 1FAP)[258]. Distances between N-termini (orange) and C-termini (magenta) are indicated. (**C**) Relative labeling activities (initial slopes) of combinations of cpHaloΔ–FKBP and Hpep–FRB fusions in presence of rapamycin with HTL-TMR. Split-HaloTag fragments (cpHaloΔ and Hpep) are linked to either N-termini or C-termini of FKBP and FRB via flexible linkers with varying length $(GGS)_{1\text{-}3}$. When both split fragments are linked to the close C-termini of the complex activity is highest. Minimal activity is observed when split-HaloTag is linked to the distant N-termini of the complex. Especially at longer fusion site distance, longer linkers lead to higher activity. Performance of Hpep–FKBP and cpHaloΔ–FRB fusions is shown in figure S25.

split-HaloTag complementation. No labeling activity in absence of rapamycin was detected in any of the tested combinations (fig. S26).

These results underscore the need to improve the split-HaloTag system for SynProLa, where split-HaloTag needs to complement across a 56 Å distance. Specifically, I aimed to achieve a mid-micromolar K_D for the interaction between the split fragments. This should lead to efficient complementation at effective concentrations in the high micromolar range, while still keeping the introduced bias to the system minimal.

5.2 Improving Split-HaloTag Performance and Versatility

5.2.1 Computational design of higher affinity Hpep variants

A computational approach was pursued to increase the affinity between Hpep and cpHaloΔ using the Rosetta modeling and design software suite[259]. I decided to tackle this task by re-designing the Hpep while keeping the current cpHaloΔ fixed. This decision was informed by two factors. Firstly, I anticipated that redesigning a peptide should be a simpler task than redesigning a protein, due to the smaller number of residues and the decreased probability to encounter unpredictable issues like interfering with the folding dynamics. Secondly, designed peptides would be easier and faster to screen in a biochemical assay, since custom peptide libraries are commercially available for a fraction of the price of a custom protein library.

A structural model of HTL-TMR labeled split-HaloTag, as shown in figure 6A, was generated from the HTL-TMR labeled HaloTag structure (PDB-ID 6Y7A)[188] by deleting residues 140–142 and 153, which leads to excision of the Hpep (residues 143–152). The cpHaloΔ structure was extended by two glycine spacers at both termini. This was done to increase the distance of the charged N- and C-termini to the Hpep (which would also not be present in fusion constructs) in order to prevent bias for complementary charged residues during Hpep design. It was anticipated that the presence of the covalently attached HTL-TMR ligand is vital for the design process, as the Hpep participates in the formation of the ligand entry channel of split-HaloTag. Without the ligand, the software might design Hpep variants that have good complementarity to cpHaloΔ, but also obstruct the ligand binding site and thereby deplete labeling activity. The HTL-TMR was parameterized for Rosetta and the non-canonical attachment to Asp[106] was realized with a constraints file, fixing distance, angles and dihedral of the bond. The structure was then minimized using the Rosetta relax application and the lowest energy structure was used for further steps.

New Hpep variants were designed using a RosettaScripts[260] protocol. Briefly, the conformations of the glycine spacers that were added to cpHaloΔ were randomized, and docking of the Hpep was performed using a soft scoring function with decreased van der Waals radii. Hpep residues were then redesigned as the structure was minimized using the FastRelax module. In addition to newly designed 10-mer Hpep variants, I designed variants with either N-terminal or C-terminal extensions of four residues. These added residues could potentially form an additional turn in the α-helical Hpep, expanding the interface to cpHaloΔ. This approach might be an additional mean to increase binding affinity. A total of 40 000 structures with different Hpep sequences were generated. Structures were scored by Rosetta total score and peptide binding free energy ($\Delta\Delta G$), and the top 20 % of structures were further analyzed.

5.2.2 Sequence analysis of designed Hpep variants

Sequence analysis of the designed 10-mer Hpep variants revealed a relatively low sequence recovery (fig. 20). However, the two phenylalanine residues at positions 5 and 8, which are oriented towards the protein core, are highly conserved. Notably, phenylalanine at position 5 plays a pivotal role in forming the ligand entry channel. Another conserved residue is arginine at position 9, which establishes a complex hydrogen bond network with the loop-helix motif spanning residues 206–214 (standard HaloTag numbering). There is an overall trend towards hydrophobic residues, which might allow to form more van der Waals interactions with the rather hydrophobic Hpep binding groove of cpHaloΔ. However, residue 3 which is oriented away from the protein, predominantly features charged and hydrophilic residues. At residue 10, despite the low recovery of the original threonine, amino acids that function as hydrogen bond donors or acceptors are enriched which could also form hydrogen bonds with the N-terminal residues of cpHaloΔ.

As intended, N-terminal extension of the Hpep by four residues lead to the formation of an additional helix turn. This results in a larger interface with cpHaloΔ, fully enclosing the ligand entry channel which is partially exposed with the original 10-mer Hpep (fig. 20). The sequence logo of the N-terminal extensions reveals that residue 1, which points towards the hydrophobic core of the protein and aids in forming the ligand entry channel, predominantly features hydrophobic and aromatic residues. In contrast, residues 2–4, which are oriented towards the protein surface or away from it, mostly feature charged and

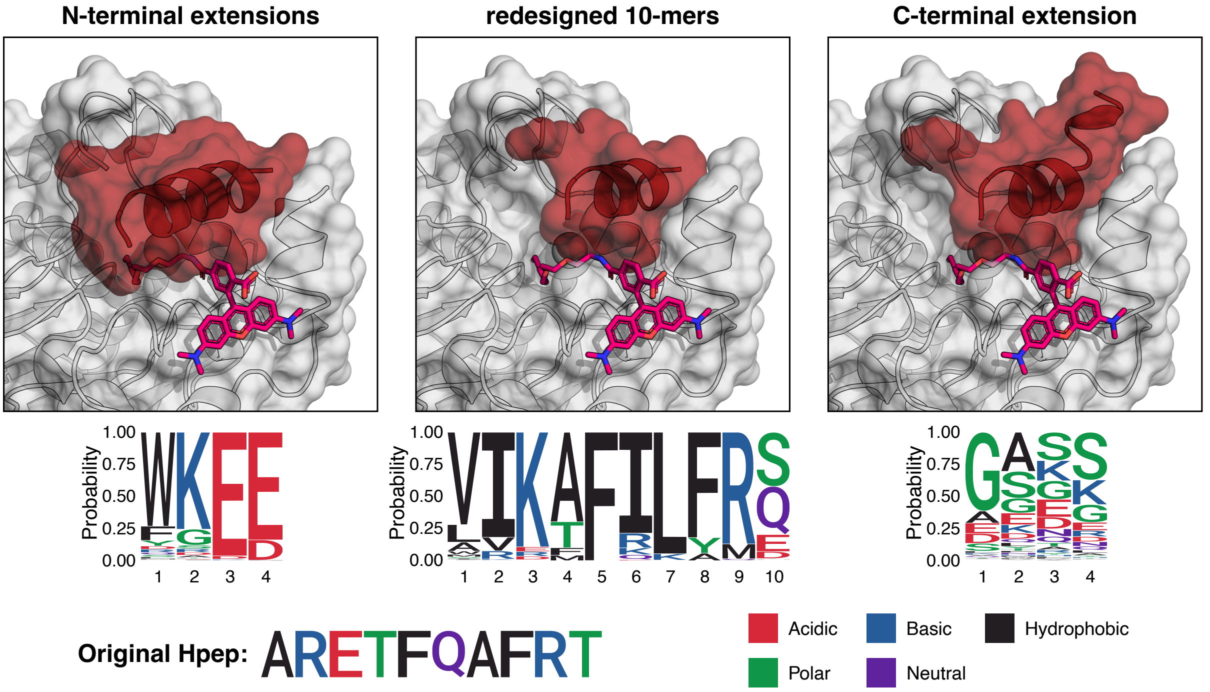


Figure 20: Structure and sequence analysis of designed Hpep variants. Exemplary structures (top) of designed Hpep variants (red) in complex with HTL-TMR (pink) bound cpHaloΔ (grey) and sequence logos (bottom) of the top 20 % of designed sequences for each category. The N-terminal extensions add another turn to the helical peptide, covering the ligand entry channel. C-terminal extensions protrude from the protein.

hydrophilic residues.

On the other hand, the C-terminal Hpep extensions seem less promising. They do not lead to an extended helix, due to steric clashes with the N-terminus of cpHaloΔ. They do not form additional interactions with cpHaloΔ but protrude from the protein in most designs (fig. 20). The sequence logo of the C-terminal extensions shows mainly small and polar residues.

A library of 384 Hpep variants was compiled by combining the best scoring 10-mers (155 variants), N-terminally (160 variants) or C-terminally (9 variants) extended Hpep sequences, combinations of the best 10-mers and N-terminal extensions (25 variants) and peptides with single point mutations of the most prevalent residues in high-scoring designs (32 variants). In addition controls were added including the original 10-mer Hpep (ARETFQAFRT), the 9-mer Hpep (ARETFQAFR) which features substantially decreased activity (fig. 6) and a peptide unable to activate cpHaloΔ (ARETAQAFRT)[261].

5.2.3 Evaluation of the Hpep library and further Hpep optimization

The Hpep library was obtained as synthetic peptides in crude quality and was screened for its efficiency to activate labeling of purified cpHaloΔ protein. 80 % of peptides showed higher activity than the native Hpep with some reaching more than a thousand fold higher labeling rates (fig. 21, table S11). Among both, the redesigned 10-mer Hpep variants and the N-terminally extended native Hpep variants, there were numerous variants with high activities. The highest average performance was seen for Hpep variants combining redesigned sequences and N-terminal extension. However, peptides featuring redesigned sequences and N-terminal extensions had the highest average activity, suggesting that effects of both design

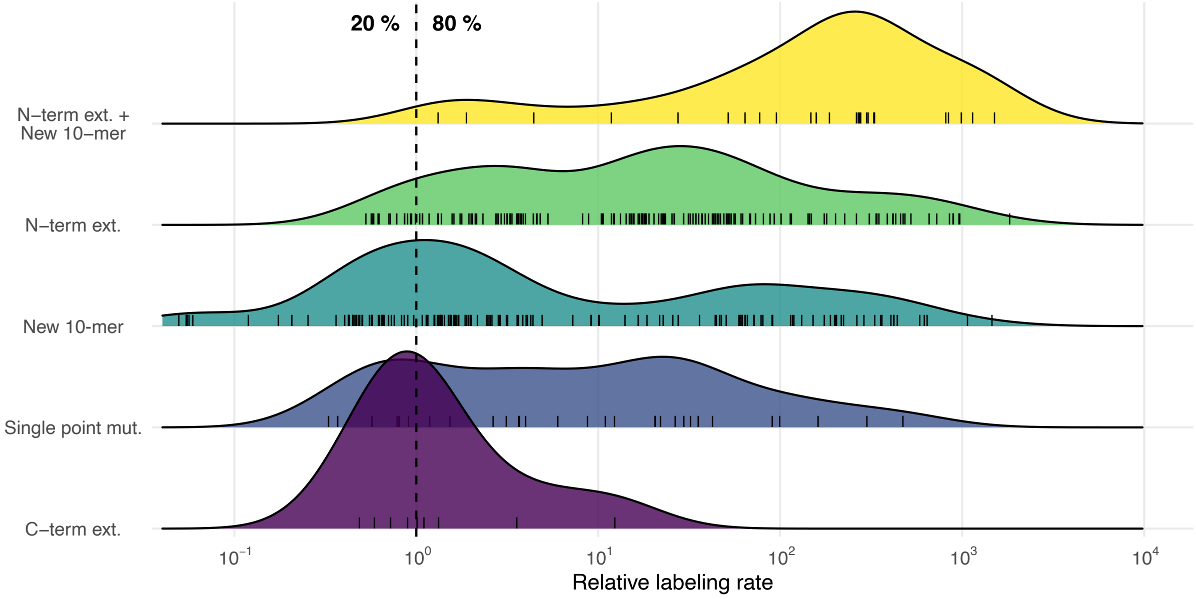


Figure 21: Hpep library screen. Density plots of cpHaloΔ labeling rates in presence of the individual Hpep variants (each small line representing one peptide, areas normalized for each category) relative to the labeling rate with the original Hpep (dashed line). Peptides with no detectable labeling were placed at 0.05 to fit on the logarithmic scale. 80 % of Hpep variants had a higher labeling rate than the original Hpep, with some leading to three orders of magnitude faster reactions. Best performance was observed for peptides combining new 10-mer Hpep sequences with N-terminal extensions.

strategies can be additive. Notably, also a number of highly active single point mutants were identified in the screen. These results are particularly useful as they enable rational design by combining successful point mutations. As expected, C-terminal extension of the Hpep did not lead to significant increases in labeling speed and this strategy was not further pursued.

While the results of the Hpep screen were encouraging, they did not provide a clear understanding of the underlying mechanisms responsible for the observed increase in labeling activity. The observed effects could be the results of a combination of higher affinity and faster labeling speed after complementation. In addition, secondary factors including peptide purity, synthesis yield and solubility, may also have significantly influenced the screening results. Hence, a number of promising candidates were purified from the library or re-synthesized to test them under more controlled conditions and estimate their affinity for cpHaloΔ. In addition, new peptides were developed, incorporating features (i.e. point mutations and extensions) of highly active Hpep variants. In a parallel endeavor by Konstantin Hinnah, new N-terminal extensions were identified through phage display, with the goal of identifying Hpep variants with increased activity. These extensions were subsequently also integrated into highly active 10-mer Hpep variants.

26 different Hpep variants were characterized using HTL-TMR to determine their half-maximal effective concentration (EC_{50}), which can be used as a proxy for the affinity (K_D) for cpHaloΔ, and their maximal apparent second-order labeling rate constants (k_{app}^{max}) for the labeling reaction of cpHaloΔ at saturating Hpep concentrations (fig. S27, S28, table S12). EC_{50} values ranged from 40.3 nM to 2.98 mM and were well distributed throughout this range. Hpep variants were named Hpep1–Hpep27 on basis of their EC_{50} values, with Hpep1 being the native peptide. The highest affinity peptides (Hpep19–

Hpep27), were ones that combine four beneficial point mutations (A1W, T4M, Q6R and A7L) and extension identified via the computational design or the phage display selections.

For most Hpep variants, k_{app}^{max} values were around the same value as for the native Hpep around $10^4\,M^{-1}\,s^{-1}$ or slightly higher with some exceptions. Hpep13 (WRETFQLFRT) had a lower k_{app}^{max} of $8.5 \cdot 10^2\,M^{-1}\,s^{-1}$, while Hpep5 (SKRDAREMFQAFRT) featured a substantially faster labeling rate at $7.5 \cdot 10^5\,M^{-1}\,s^{-1}$, which is 58-fold faster than the native Hpep1. This renders Hpep5 an especially interesting candidate for SynProLa and possibly other potential split-HaloTag applications. Besides its fast kinetics, Hpep5 has an EC_{50} value of 149 µM which is considerably lower compared to the native Hpep, and could be in a good range for complementation across larger distances using extended flexible linkers like in SynProLa (predicted $C_{eff} \approx 0.23$ mM).

The availability of Hpep variants across such a wide span of affinities renders split-HaloTag a very versatile protein complementation system. By choosing different Hpep variants, the sensitivity, introduced bias (by enforcing interactions) and extent of spontaneous complementation can be well balanced to fit the requirements of a certain application.

5.2.4 FKBP/FRB model system with improved Hpep

The FKBP/FRB model system described above was used to test whether Hpep variants with increased affinity could facilitate split-HaloTag labeling over larger distances. Constructs were used where split-HaloTag fragments are fused to the distant N-termini of FKBP and FRB. With the native Hpep1, these constructs showed virtually no labeling activity. However, upon incorporating Hpep5, efficient rapamycin-dependent labeling was observed within 20 min (fig. 22). With a reaction rate constant of $2.7 \cdot 10^4\,M^{-1}\,s^{-1}$ the reaction was slower than the maximum labeling rate measured for Hpep5, indicating only partial split-HaloTag complementation. Nevertheless, this labeling speed is comparable to the maximal labeling rates of other higher-affinity Hpep variants, even though the latter might have a greater propensity for rapamycin-independent complementation and labeling.

Following these results, I tested the performance of the improved split-HaloTag system in a more relevant context, inside living cells. I created constructs for co-expression of membrane localized EGFP-cpHaloΔ-FKBP and cytosolic Hpep1/5-FRB-mScarlett fusion proteins. After incubation with HTL-CPY, cells with the Hpep5 construct showed clear rapamycin-dependent labeling in microscopy (fig. 23,

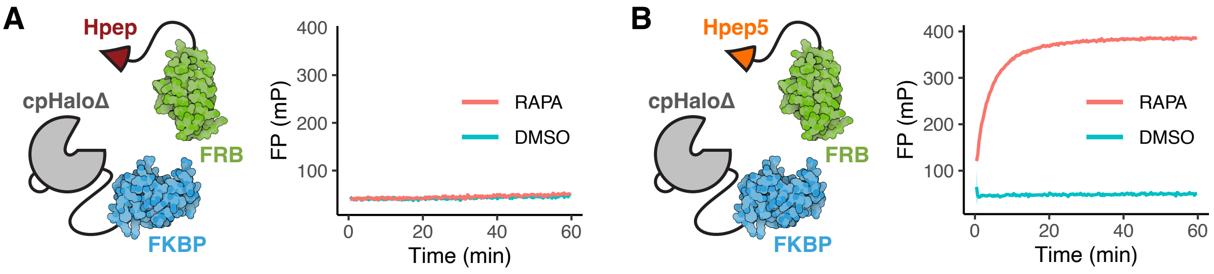


Figure 22: FKBP/FRB model system with improved Hpep variants in vitro. Scheme and labeling kinetics of cpHaloΔ-FKBP (250 nM) and Hpep1-FRB (250 nM, **A**) or Hpep5-FRB (250 nM, **B**) with HTL-TMR (50 nM) monitored by fluorescence polarization. Due to the large distance between the N-termini of FKBP and FRB, only the improved Hpep5 enables rapamycin-dependent split-HaloTag labeling.

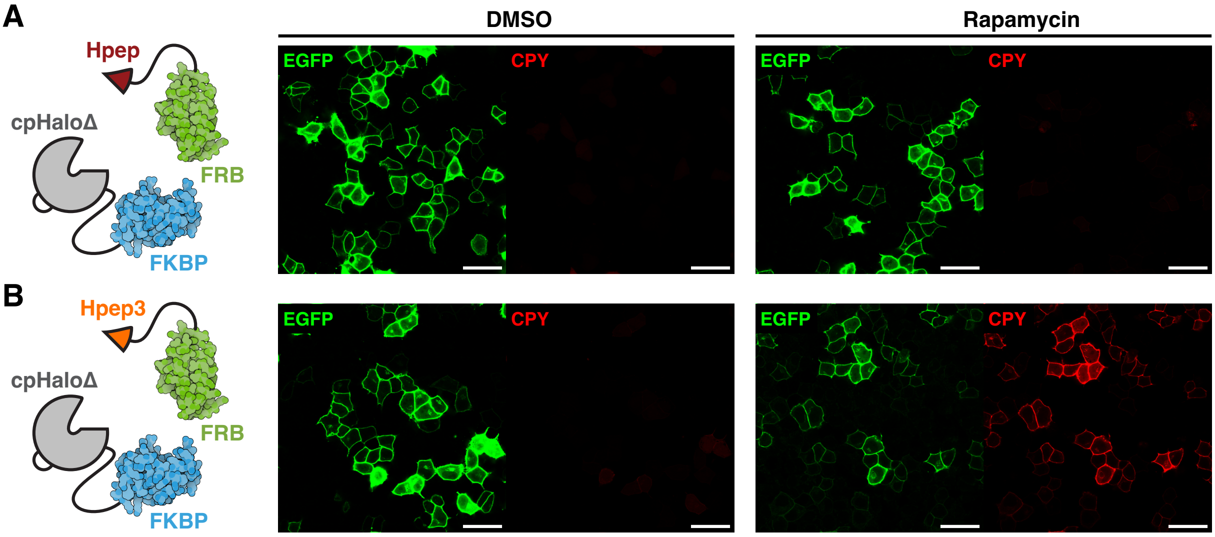


Figure 23: FKBP/FRB model system with improved Hpep variants in HeLa cells. Confocal microscopy images of HeLa cells co-expressing Lyn11-EGFP-cpHaloΔ-FKBP and Hpep1-FRB-mScarlet (**A**) or Hpep5-FRB-mScarlet (**B**). Labeling of EGFP (green) positive cells with HTL-CPY (red, 100 nM, 1 h) is observed only in the presence of rapamycin (100 nM). Scale bars: 50 µm.

S30) and flow cytometry (fig. S31) experiments, while Hpep1 cells showed only marginal signals. Control constructs where the split-HaloTag was fused to the much closer N-termini of FKBP and FRB showed rapamycin-dependent labeling for both Hpep variants, albeit signal strength with Hpep5 was higher.

These experiments validate that the designed Hpep variants increase the versatility of the split-HaloTag approach, by making it applicable to more complex protein-protein interactions with non-optimal location of termini in a cellular context. Consequently, the improved Hpep variants might also enable split-HaloTag labeling at synaptic NRX–NLG complexes in SynProLa constructs.

5.2.5 Computational design of a cpHaloΔ with improved stability

Another key factor for the performance of split-HaloTag may be the stability of the cpHaloΔ protein. As described in section 4.2.3, the naive cpHaloΔ protein has a melting temperature of 31.3 °C. Hence, at the physiological temperature of 37 °C the majority of the protein might exist in an unfolded and inactive state. While fusion with other proteins can substantially enhance its stability, as observed for M13-cpHaloΔ in Caprola, it is uncertain if NRX- and NLG-fusions would cause similar stabilizing effects. Therefore, it would be desirable to increase cpHaloΔ thermostability above 37 °C.

Two strategies for improving cpHaloΔ stability were developed: identifying stabilizing point mutations within the protein, and redesigning the flexible $(GGS/T)_5$ circular permutation linker (CP-linker) that connects the C- and N-terminal domains of the protein. I hypothesized that a well-folded and less flexible linker could promote a stronger association between the N- and C-terminal fragments that constitute cpHaloΔ. This could be achieved by forming a shared interface between the linker and both fragments. Moreover, a more rigid linker could maintain the optimal distance between the original termini.

The computational and practical work for this project was conducted by Lennart Nickel for his master's thesis, which I supervised. Hence, these efforts are only briefly outlined in this thesis. Candidates

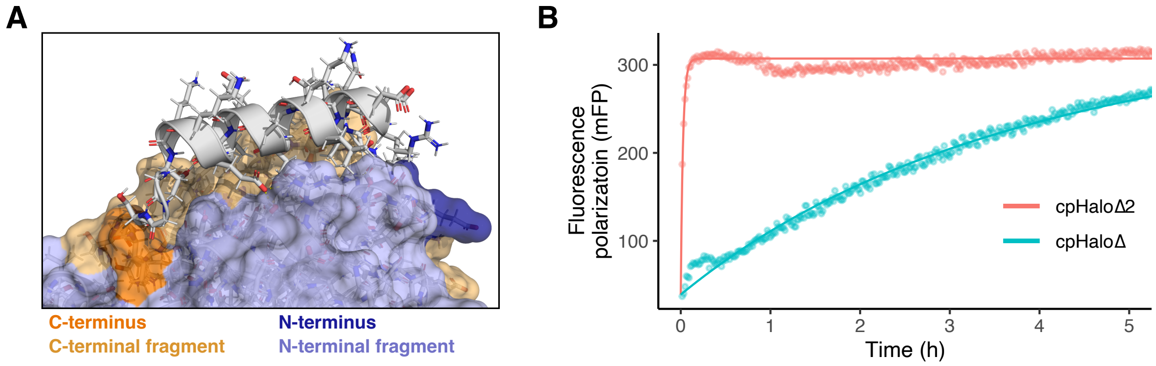


Figure 24: cpHaloΔ2 engineering and characterization. (**A**) Structure of cpHaloΔ with Rosetta designed α-helical CP-linker (white) connecting the original N- (red) and C-termini (orange) of HaloTag. The helical linker forms hydrophobic interactions and hydrogen bonds to both fragments (light-pink and light-orange) of cpHaloΔ. (**B**) Comparative labeling kinetics of cpHaloΔ and the engineered cpHaloΔ2 (100 nM) with HTL-TMR (20 nM) in presence of 2.6 µM Hpep5 followed by fluorescence polarization. Data was recorded by Lennart Nickel.

for stabilizing point mutations were identified using a computational pipeline based on atomistic Rosetta modeling and phylogenetic sequence information[262]. A position-specific substitution matrix (PSSM) for the protein is calculated based on a multiple sequence alignment with homologs. Structure-based computational scanning is then performed for all mutations with a positive PSSM score to identify beneficial mutations using Rosetta software. The HTL-TMR labeled HaloTag structure (PDB-ID: 6Y7A)[188] was used as an input. From the results, 12 mutations were selected and experimentally tested as a single point mutation in cpHaloΔ. Four mutations (E20S, N119H, V184E, V197K, standard HaloTag numbering) that had positive impact on both, melting temperature and labeling rate in presence of Hpep5 were combined. The resulting protein, featured a melting temperature of 43.4 °C and a 34-fold faster labeling rate in presence of 12.5 µM Hpep5 at 37 °C.

New, well-folded α-helical CP-linkers were designed using the Rosetta Remodel[263] application. 23 linkers with different topologies were tested experimentally and 22 of these had a positive impact on melting temperature and labeling rate in presence of Hpep5. The best performing linker, which had a 5 °C increase in thermostability compared to the original linker, was incorporated into the quadruple cpHaloΔ mutant described above (fig. 24A). The resulting protein, herein called cpHaloΔ2, featured a melting temperature of 45.4 °C and a 96-fold faster labeling rate in presence of 12.5 µM Hpep5 at 37 °C (fig. 24B).

The EC_{50} and k_{app}^{max} values for Hpep5, Hpep7 and Hpep10 with cpHaloΔ2 were determined (fig. S33, S34, table S13). Notably, the EC_{50} values were, on average, 8-fold lower while k_{app}^{max} values were 3-fold higher compared to those of the original cpHaloΔ. However, these differences might be even more pronounced in a cellular environment, where the proteins are continuously at 37 °C in contrast to the in vitro conditions where proteins are stored at low temperatures and only warmed up during the assay. Preliminary data indicates that a pre-incubation at 37 °C for 20 min, significantly reduces split-HaloTag labeling activity with the original cpHaloΔ. In the absence of Hpep, the observed labeling rate constant for cpHaloΔ2 remains low at 5.5 M^{-1} s^{-1} (fig. S35).

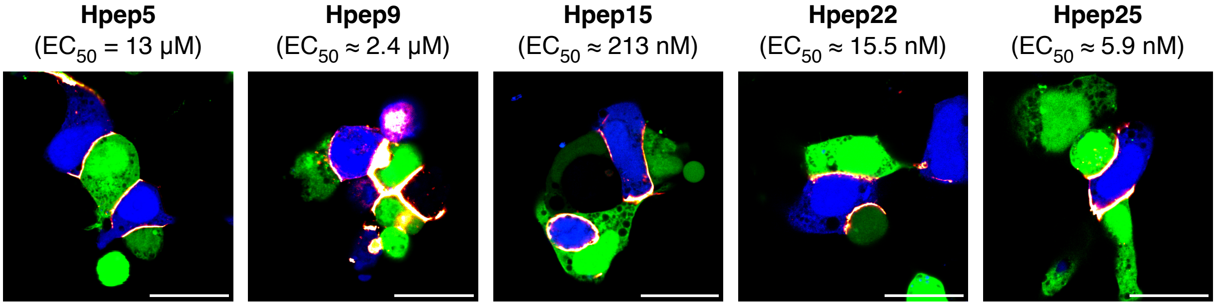


Figure 25: HEK-293 co-culture assay with second generation SynP roLa constructs. Confocal microscopy images of co-cultured HEK-293 Flp-In™ T-REx™ cells expressing second generation SynProLa constructs mTagBFP2–F2A–cpHaloΔ2–(GSG)$_4$–NRX and EGFP–F2A–Hpep–(GSG)$_4$–NLG with different Hpep variants. EC_{50} values for Hpep variants are provided. Except for Hpep5, the EC_{50} values were not directly characterized for cpHaloΔ2. Instead, they were estimated based on the values for the original cpHaloΔ, factoring in the average 8-fold increase in EC_{50} when transitioning from cpHaloΔ to cpHaloΔ2. Cells were stained with HTL-CPY (100 nM) for 1 h prior to PFA fixation and imaging. Pre- and post-SynProLa expressing cells can be clearly distinguished by the co-expressed fluorescent proteins (mTagBFP2: blue, EGFP: green). Contact sites between pre- and post-SynProLa cells are selectively stained with HTL-CPY (red-hot lookup table). Scale bars: 25 µm.

5.3 Second Generation SynProLa Constructs

The SynProLa constructs previously described (section 5.1.1) were modified by substituting cpHaloΔ with the more thermostable cpHaloΔ2 and the original Hpep1 was replaced with either Hpep5, Hpep9, Hpep15, Hpep22 or Hpep25, exhibiting estimated EC_{50} values ranging from 5.9 nM to 13 µM. Not every combination of split-HaloTag NRX/NLG fusions was generated, but Hpep was always fused to NLG, while cpHaloΔ2 was fused to NRX. This was done to limit the genetic construct size. Given that the NLG gene is more than twice the size of NRX, fusing it to cpHaloΔ2, rather than the shorter Hpep, could complicate future delivery strategies such as AAV-vectors, due to size constraints. Moreover, (GSG)$_4$ linkers were used consistently between split-HaloTag and NRX/NLG instead of shorter linkers which were tried in first generation SynProLa constructs, due to the expected positive impact on the effective concentration.

When expressed in separate pools of co-cultured HEK-293 cells, SynProLa constructs lead to a clear labeling of intercellular NRX–NLG complexes after 1 h incubation with HTL-CPY. No differences between Hpep variants with different EC_{50} values was observed. These results highlight the importance of the development of an improved split-HaloTag as these experiments were unsuccessful with the previous version. However, while these findings are promising, additional controls, further optimization (i.e. optimal choice of Hpep) and final characterizations are necessary to establish a robust and final version of SynProLa.

5.4 Conclusion and Outlook

I aimed to develop a synapse specific protein labeling approach (SynProLa) utilizing the split-HaloTag system. To achieve this, I fused the split-HaloTag components Hpep and cpHaloΔ with NRX and NLG, which are known to form a trans-synaptic complex. This should restrict split-HaloTag complementation and the subsequent labeling with fluorescent ligands to synaptic sites.

However, in initial trials using a HEK-293-based assay, none of the constructs yielded successful labeling of intercellular NRX–NLG complexes. Using a protein-protein interaction model system, I could show that the problem is likely the large distance between the attachment sites of the split-HaloTag fragments on the NRX–NLG complex and the resulting low effective concentrations of the fragments. Since the intrinsic affinity of the split-HaloTag fragments is in the millimolar range, complementation cannot be achieved under these conditions. To address this issue, I used a computational strategy, utilizing the Rosetta software, to design new Hpep variants with enhanced binding affinities. This lead to the identification of 26 Hpep variants with affinities spanning from low nanomolar to millimolar EC_{50} values.

Another known limitation of the original split-HaloTag system was the relatively low thermal stability of cpHaloΔ. Using a phylogeny- and Rosetta-based computational approach, facilitated the identification of point mutations and the design of a new circular permutation linker that resulted in a significant increase in melting temperature of cpHaloΔ, from 31 °C to 45 °C. Further characterization of the improved split-HaloTag system showed that beyond the tunable affinity and enhanced stability, there was also a notable acceleration in the labeling speed of the complemented protein with certain Hpep variants.

The improved split-HaloTag system was then fused to NRX and NLG which enabled successful staining of intercellular NRX–NLG complexes at the interface of HEK-293 cells. While these findings are encouraging, their are still preliminary. Additional controls, further optimization (e.g. finding the optimal choice of Hpep) and final characterizations are necessary to establish a robust and final version of SynProLa.

As the next step, testing lower affinity Hpep variants would be important, to keep the bias caused by split-HaloTag fusion on the NRX–NLG interaction to a minimum, while still achieving robust labeling. In addition, a negative control for the co-culture assay should be established, either by truncating NRX or NLG or by employing a mock receptor like CD4. This would allow to ensure that the contact sites between cells are not labeled upon mere co-localization, and demonstrate that SynProLa is only labeled upon NLG–NRX complex formation. Furthermore, labeling could be observed in real-time to get insights into the kinetics of SynProLa labeling which could reveal differences between Hpep variants and might help find an optimal balance between affinity and efficient labeling. To quantify the effect that split-HaloTag has on the NRX–NLG interaction, soluble domains of both proteins with and without split-HaloTag fusion could be purified, which would enable measuring precise affinities.

Subsequently, SynProLa constructs should be tested in cultured primary hippocampal neurons to assess their ability to label NRX–NLG complexes at actual synapses rather than at HEK-293 cell-cell interfaces. In this context it would also be possible to test the capacity of SynProLa to reveal changes in the connectivity of a neuronal network by labeling with two spectrally distinguishable fluorophore ligands at different time points. During the first weeks of culture, neurons consistently develop and establish synapses[264]. Synapses that are formed between the applications of the first and second fluorophore ligand should be exclusively labeled with the second ligand. It may also be possible to detect changes in strength of synapses which has been correlated to synaptic size[265] by evaluating the labeling ratios of the sequentially used ligands.

If SynProLa shows satisfactory results in cultured neurons, in vivo labeling of synapses could be attempted using a mouse model. In collaboration with the group of Peter Scheiffele, we intend to investigate the synapses of thalamic neurons that project to inhibitory interneurons in layer 4 of the cerebral cortex.

The thalamus receives sensory input and motor signals, subsequently relaying this information to specific regions within the cerebral cortex, which are responsible for certain senses[266]. Expression of SynProLa constructs could be confined to these cells by defined injection sites and usage of specific promotors. Transduced cells could be identified by co-expression of fluorescent proteins. After administration of a fluorescent HaloTag ligand and tissue fixation distinct labeling of synapses between the marked cells should be observed. Correct localization of SynProLa signal could be confirmed by immunofluorescence.

As a next step the ability of SynProLa to label newly formed synapses or detect differences in synapse size by using spectrally distinguishable substrates at different time points could be investigated. This could be performed in the same model system, as layer 4 neurons exhibit pronounced synaptic plasticity in response to the intensity of sensory input from the thalamus[267]. Variations in input could either be achieved pharmacologically using a DREADD (designer receptors activated only by designer drugs) system co-transduced into the thalamus, or by sensory modulation via deprivation or enrichment of the environment. If these experiments are successful, they could pave the way for comprehensive studies that explore neuronal connectivity and synaptic plasticity between different neuron types and brain region at large scale, using state of the art imaging technologies (fig. 26). Such studies could contribute to our understanding of how the neuronal network develops, functions and adapts to environmental stimuli, learning processes and potential pathological conditions.

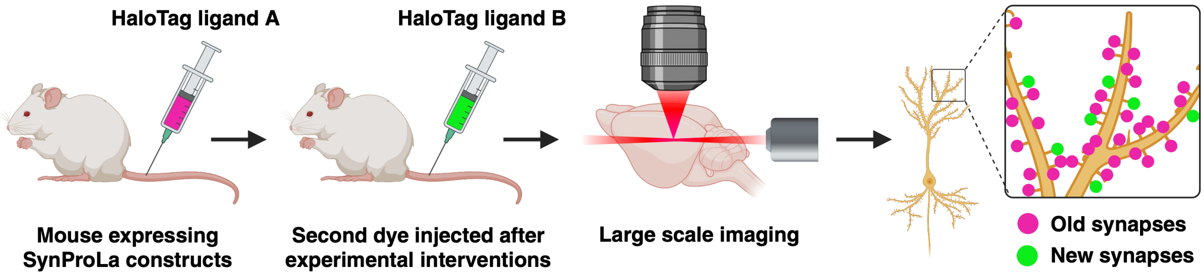


Figure 26: Outline of envisioned multi-color SynProLa experiments to reveal synaptogenesis throughout the brain. Experimental interventions could include pharmacological treatments, sensory modulation or behavioral training.

A valid concern when using SynProLa in the proposed form are the potential effects that overexpression of NRX and NLG might have on the neuronal network. Especially NLG overexpression can enhance the recruitment of NRX to newly formed terminals, thereby promoting synapse formation, which might heavily bias SynProLa experiments. To circumvent these problems, the short Hpep could be fused to endogenous NLG by generating transgenic mice. To gain precise control over transgene expression, the approach could be combined with a Cre recombinase dependent strategy to restrict SynProLa expression to specific neuronal subtypes using the large number of available Cre driver lines. Minimal perturbation of synapse morphology could be confirmed by EM or correlative light and electron microscopy (CLEM).

Apart from further development of SynProLa and its potential applications, the described improvement of the split-HaloTag system could be a solid foundation for the development of countless other new labeling strategies or molecular recorders. Indeed, several split-HaloTag-based tools are currently under development in our laboratory that benefit greatly from the described improvements, further highlighting the versatility of the system. However, a limitation of these tools is that they cannot easily be combined as they react with the same ligands. Although it would be very interesting, for example, to use Caprola and SynProLa in parallel to show correlations between neuronal activity and synaptic plasticity. To some extend this could be overcome by using distinct localizations, e.g. sequestering Caprola into the nucleus of neurons to not overlap with SynProLa signals. Nevertheless, this only partially solves the problem and in the long run it would be of great interest to develop orthogonal split self-labeling proteins that allow unrestricted combination. This could possibly be achieved by developing an orthogonal (split-)HaloTag system, or by improving and splitting the SNAP-tag or CLIP-tag.

In conclusion and despite pending validation in neurons, the SynProLa system represents a promising avenue for extending the toolbox for mapping neuronal connectivity and could provide unprecedented, large-scale data on synaptic plasticity throughout the brain. Such knowledge could greatly advance our understanding of the foundational mechanisms governing processes like cognitive flexibility, learning and memory formation under physiological or pathological conditions.

6 | Materials & Methods

6.1 Buffers, Media and Chemicals

Chemicals and reagents were purchased from the following commercial providers: Merck KGaA, Honeywell International Inc., Thermo Fisher Scientific Inc., Carl Roth GmbH & Co. KG, VWR International, Acros Organics B.V.B.A., TCI Co. Ltd.. Custom oligonucleotides were purchased from Merck KGaA or Eurofins Scientific SE. Custom peptides were either synthesized as described below (section 6.5.2) or purchased from GenScript Biotech Corp.. HaloTag substrates were either synthesized in-house according to published procedures[145–148,188,268–271] by Bettina Réssy, Dominik Schmidt, Nicole Mertes, Lin Xue or Julie Karpenko, purchased from Promega Corp. or kind gifts from Dr. L. D. Lavis (HHMI, Janelia research campus). Chemical structures of HaloTag substrates are depicted in figures S1 & S2. The term "water" (H_2O) stands exclusively for ultrapure water form a Milli-Q® Reference A+ (Merck KGaA) system. The composition of buffers and media used in this thesis are listed in table 1.

Table 1: Buffers and media

Name	Composition
IMAC lysis buffer	50 mM KH_2PO_4, 150 mM NaCl, 5 mM imidazole, 1 mM PMSF, 0.25 mg/mL lyzozyme, pH 8.0
IMAC wash buffer	50 mM KH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 7.5
IMAC elution buffer	50 mM KH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 7.5
Strep-tag wash buffer	100 mM TRIS · HCl, 150 mM NaCl, 1 mM EDTA, pH 8
Strep-tag elution buffer	100 mM TRIS · HCl, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin, pH 8
TEVp buffer	25 mM TRIS · HCl, 200 mM NaCl, 1 mM DTT, 2.5 % glycerol, pH 8
Activity buffer	50 mM HEPES, 50 mM NaCl, pH 7.3
FP buffer	50 mM HEPES, 50 mM NaCl, 0.5 g/L BSA, pH 7.3
PBS	10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4
PBS-T	10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, 0.01 % Triton™ X-100, pH 7.4
Zero free Ca^{2+} buffer	30 mM MOPS, 100 mM KCl, 10 mM EGTA, pH 7.2

39 µM free Ca^{2+} buffer	30 mM MOPS, 100 mM KCl, 10 mM Ca·EGTA, pH 7.2
5 × Isothermal Gibson assembly buffer	25 % PEG-8000, 500 mM TRIS·HCl, 50 mM MgCl$_2$, 50 mM DTT, 1 mM of each dNTP, 5 mM NAD, pH 7.5
LB medium	10 g/L tryptone, 5 g/L yeast extract, 171.1 mM NaCl
General growth medium	DMEM, 4.5 g/L glucose, 110 mg/L pyruvate, 1 × GlutaMAX™, 10 % FBS

6.2 Bacterial and Mammalian Cells

Bacterial and mammalian cells used in this thesis are listed in table 2.

Table 2: Cell lines used in this thesis

Name	Type	Provider
E. coli E. cloni 10G	bacterial	Lucigen Corp.
E. coli NEB 5-alpha	bacterial	New England Biolabs Inc.
E. coli BL21(DE3)pLysS	bacterial	Merck KGaA
HEK-293 Flp-In™ T-REx™	mammalian	Thermo Fisher Scientific Inc.
HeLa	mammalian	DSMZ (German Collection of Microorganisms and Cell Cultures GmbH)

6.3 Equipment and Devices

Equipment and devices used in this thesis are listed in table 3.

Table 3: Equipment and devices

Device	Name/description	Company
Plate reader	Tecan Spark® 20M	Tecan Group Ltd.
FPLC	ÄKTA™ Pure	Cytiva Europe GmbH
Flow cytometer	BD LSRFortessa™ X-20	BD Biosciences Corp.
Stopped flow	BioLogic SFM-400	BioLogic Science Instruments Ltd.
ITC	MicroCal PEAQ-ITC	Malvern Panalytical GmbH.
Confocal fluorescence microscope	Stellaris 5 (white light laser, 405 nm diode, hybrid photodetectors)	Leica Microsystems GmbH
nanoDSF	Prometheus NT 48	NanoTemper Technologies GmbH
96/384-channel pipettor	VIAFLO 96 | 384	Integra Biosciences GmbH.

Microvolume spectrophotometer	NanoDrop™ 2000c	Thermo Fisher Scientific Inc.
Analytical HPLC-MS	Shimadzu Nexera X2 , LCMS-2020	Shimadzu Corp.
Preparative HPLC-MS	Shimadzu Prominence, LCMS-2020	Shimadzu Corp.
Peptide synthesizer	Liberty Blue™	CEM Corp.
Microfluidizer	LM10 Microfluidizer®	Microfluidics Int. Corp.
Sonicator	Sonoplus HD 2070	Bandelin electronic GmbH & Co. KG
Mass spectrometer	Bruker maXis II ETD	Bruker Corp.
Ultrapure water system	Milli-Q® Reference A+	Merck KGaA
Fluorescence laser scanner	Amersham™ Typhoon™ 5	Cytiva Europe GmbH
Gel imager	ChemiDoc™ MP & ChemiDoc™ EZ	Bio-Rad Laboratories Inc.

6.4 Software and Data Analysis

Construct and primer design for molecular cloning was performed with Geneious (Dotmatics Ltd.). Protein structures were analyzed and visualized with PyMOL[272] (Schrödinger LLC.). Images were processed and analyzed with FIJI[273]. Fitting of ordinary differential equation systems to kinetic data was performed with DynaFit 4.08[274] (BioKin Ltd.). ITC data was analyzed using the MicroCal PEAQ-ITC analysis software (Malvern Panalytical GmbH, version 1.30). Flow cytometry data was processed with FlowJo™ 10.8.1 (BD Biosciences Corp.). Macromolecular modeling and protein design was performed with the Rosetta software suite[259] (version 3.13). Chemical structures were drawn with ChemDraw 20.1 (PerkinElmer Informatics Inc.). Schemes and figures were created with Adobe Illustrator CC 2023 (Adobe Inc.) or BioRender.

If not stated otherwise, data was analyzed and represented using R[275], with the following packages: ggplot2[276] dplyr[277] tidyr[278] readr[279] purrr[280] furrr[281] tibble[282] stringr[283] forcats[284] tsoutliers[285] minpack.lm[286] Bio3D[287] Biostrings[288] viridis[289] cowplot[290] ggseqlogo[291].

6.5 Chemical Procedures

6.5.1 General information

Preparative HPLC-MS was performed with a Shimadzu prominence preparative HPLC system coupled to a LCMS-2020 mass spectrometer. A Shimazu shim-pack GIS column (5 µm, C18, 50 × 250 mm) was used with a flow rate of 50 mL min^{-1}. Analytical HPLC-MS was performed on a Shimadzu Nexera X2 analytical HPLC system coupled to a LCMS-2020 mass spectrometer. A Supleco® Titan™ column (Merck KGaA, 5 µm, C18, 50 × 1.9 mm) was used with a flow rate of 1 mL min^{-1}. High resolution mass spectrometry (HRMS) measurements were performed by the MS core facility of the Max Planck

Institute for Medical Research Heidelberg on a Bruker maXis II ETD connected to a Shimadzu Nexera X2 HPLC.

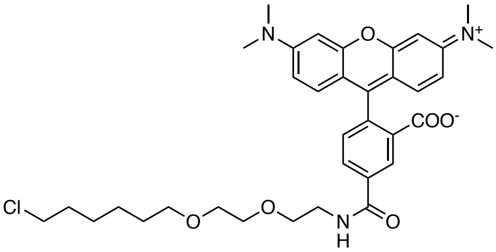
6.5.2 Solid phase peptide synthesis

Solid phase peptide synthesis was conducted using standard Fmoc chemistry on an automated microwave peptide synthesizer (Liberty Blue, CEM Corp.). N,N'-diisopropylcarbodiimide (DIC) was used as coupling reagent, ethyl cyano(hydroxyimino)acetate (oxyma) as activating base, 20 % piperidine in dimethylformamid (DMF) as deprotecting reagent and DMF as a solvent. Preloaded Wang resin (0.25 mmol loading) was allowed to swell in DMF for 5 min, followed by cycles of deprotection, washing and coupling.

1. Deprotection: 10 mL 20 % piperidine, 100 s, 90 °C;
2. Wash: 4 × 10 mL DMF
3. Coupling: 2.5 mL 1 M DIC, 1.25 mL 0.5 M oxyma and 6.25 mL 0.2 M protected amino acid, 265 s, 90 °C

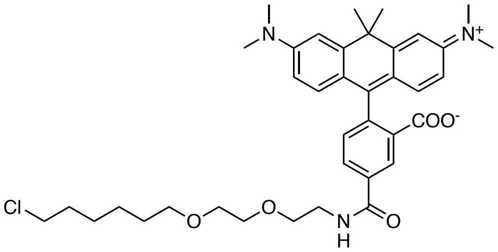
For arginine the coupling step was repeated twice. After a final deprotection and washing step, resins were washed with dichloromethane (DCM) and dried under reduced pressure. Peptides were deprotected and cleaved from the resin with 5 mL of 95 % TFA, 2.5 % H_2O and 2.5 % TIS (RT, 4 h). The volume of cleavage mixture was reduced under a stream of air to 1 mL and peptides were precipitated by pouring the mixture into 50 mL -20 °C cold diethyl ether. The precipitated peptides were separated from the supernatant by centrifugation (15 min, 4000 RCF, -10 °C) and dried under reduced pressure. Crude products were dissolved in 10 % MeCN, 0.1 % formic acid in H_2O and purified via preparative HPLC (0 % to 60 % MeCN in H_2O, 0.1 % formic acid, 60 min). After lyophilization, purity above 95 % and correct mass were confirmed via analytical HPLC-MS and HRMS.

6.5.3 HTL-5-TMR



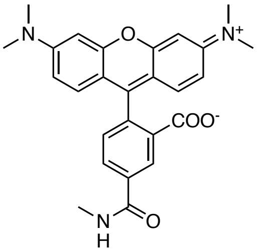
The synthesis of HTL-5-TMR was performed as described by Wilhelm & Kühn et al.[188]. To a solution of TMR-5-COOH (2.5 mg, 5.81 µmol, 1.0 eq.) in dry DMSO (500 µL), BOP (0.5 M in DMSO, 16.4 µL, 8.21 µmol, 1.5 eq.) was added and the reaction was shaken at RT at 500 rpm for 5 min. DIPEA (3.84 µL, 23.2 µmol, 4.0 eq.) and HTL−NH_2 (1 M in DMSO, 8.71 µL, 8.71 µmol, 1.5 eq.) were added and the reaction was shaken at RT at 500 rpm for 4 h. The crude product was acidified with acetic acid and purified over preparative HPLC (10 % to 90 % MeCN in H_2O, 0.1 % formic acid, 50 min) to give 1.2 mg (1.89 µmol) of the desired product in 33 % yield after lyophilization.
HRMS (ESI): expected $C_{36}H_{44}N_2O_6Cl^+$ $[M+H]^+$: 635.2887; found: 635.2882.

6.5.4 HTL-5-CPY



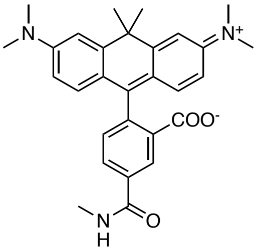
The synthesis of HTL-5-CPY was performed as described by Wilhelm & Kühn et al.[188]. To a solution of CPY-5-COOH (2.5 mg, 5.81 µmol, 1.0 eq.) in dry DMSO (1 mL), BOP (0.5 M in DMSO, 16.4 µL, 8.21 µmol, 1.5 eq.) was added and the reaction was shaken at RT at 500 rpm for 5 min. DIPEA (3.84 µL, 23.2 µmol, 4.0 eq.) and HTL−NH_2 (1 M in DMSO, 8.71 µL, 8.71 µmol, 1.5 eq.) were added and the reaction was shaken at RT at 500 rpm for 4 h. The crude product was acidified with acetic acid and purified over preparative HPLC (10 % to 90 % MeCN in H_2O, 0.1 % formic acid, 50 min) to give 0.38 mg (0.57 µmol) of the desired product in 10 % yield after lyophilization.
HRMS (ESI): expected $C_{38}H_{49}N_3O_5Cl^+$ $[M+H]^+$: 662.3360; found: 662.3349.

6.5.5 MeAm-5-TMR



The synthesis of MeAm-5-TMR was performed as described by Wilhelm & Kühn et al.[188]. To a solution of TMR-5-COOH (2.5 mg, 5.81 µmol, 1.0 eq.) in dry DMSO (500 µL), BOP (2.59 mg, 8.71 µmol, 1.5 eq.) was added and the reaction was shaken at RT at 500 rpm for 5 min. DIPEA (3.84 µL, 23.2 µmol, 4.0 eq.) and methylamine (2 M in THF, 4.36 µL, 8.71 µmol, 1.5 eq.) were added and the reaction was shaken at RT at 500 rpm for 4 h. The crude product was acidified with acetic acid and purified over preparative HPLC (0 % to 90 % MeCN in H_2O, 0.1 % formic acid, 50 min) to give 0.97 mg (2.19 µmol) of the desired product in 38 % yield after lyophilization.
HRMS (ESI): expected $C_{26}H_{26}N_3O_4^+$ $[M+H]^+$: 444.1923; found: 444.1914.

6.5.6 MeAm-5-CPY



The synthesis of MeAm-5-CPY was performed as described by Wilhelm & Kühn et al.[188]. To a solution of CPY-5-COOH (2.5 mg, 5.48 µmol, 1.0 eq.) in dry DMSO (1 mL), BOP (0.5 M in DMSO, 17.4 µL,

8.71 µmol, 1.5 eq.) was added and the reaction was shaken at RT at 500 rpm for 5 min. DIPEA (3.62 µL, 21.9 µmol, 4.0 eq.) and methylamine (2 M in THF, 4.11 µL, 8.21 µmol, 1.5 eq.) were added and the reaction was shaken at RT at 500 rpm for 4 h. The crude product was acidified with acetic acid and purified over preparative HPLC (0 % to 90 % MeCN in H_2O, 0.1 % formic acid, 50 min) to give 0.77 mg (1.64 µmol) of the desired product in 30 % yield after lyophilization.
HRMS (ESI): expected $C_{29}H_{32}N_3O_3^+$ $[M+H]^+$: 470.2443; found: 470.2437.

6.6 Biochemistry and Molecular Biology

6.6.1 General information

DNA and protein concentrations were determined with a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc.) by measuring absorption at 260 nm or 280 nm, respectively. DNA solutions were stored at -20 °C and proteins were either stored at -80 °C after flash freezing in liquid nitrogen or mixed 1:1 with 90 % (w/v) glycerol in activity buffer (table 1) and stored at -20 °C. Protein mass spectrometry (intact mass) was performed by the MS core facility of the Max Planck Institute for Medical Research Heidelberg on a Bruker maXis II ETD connected to a Shimadzu Nexera X2 HPLC.

All plate reader experiments were performed using a Spark20M microplate reader (Tecan Group Ltd.). Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates (Corning Inc.) or black, low-binding, flat bottom, 96-well plates (PerkinElmer Inc.). Filters used for plate reader experiments are summarized in table 4. Fluorescence polarization was calculated according to equation 6.1. The device-specific correction factor (G-factor)[292] was calibrated for each fluorophore and measurement.

$$FP = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} + GI_{\perp}} \tag{6.1}$$

where:

$I_{\parallel}$ = fluorescence intensity parallel to the excitation light polarization
$I_{\perp}$ = fluorescence intensity perpendicular to the excitation light polarization
G = G-factor (correction factor)

Table 4: Filters settings for plate reader experiments

Fluorophore	**Excitation filter (BW[a])**	**Emission filter (BW[a])**
Alexa488, fluorescein, Oregon green, JF_{503}, 500R	485 nm (20 nm)	535 nm (25 nm)
TMR, JF_{549}, JF_{525}, TMR-az-F_2, TMR-CN, TMR-SCH_3, TMR-SNH_2, MaP555, 510R, 515R, 580CP, Cy3	535 nm (25 nm)	595 nm (35 nm)
CPY, SiR, LIVE580, JF_{608}, JF_{646}, JF_{669}, Cy5	620 nm (20 nm)	680 nm (30 nm)

[a] BW: bandwidth

6.6.2 Molecular cloning

Molecular cloning was performed as described by Wilhelm & Kühn et al.[188] by Gibson assembly[293] or with the NEB Q5 site-directed mutagenesis kit (New England Biolabs Inc.) according to the manufacturer's protocol. For Gibson assemblies, DNA fragments were amplified by polymerase chain reactions (PCR) using either the KOD Hot Start Master Mix (Merck KGaA) with 5 % DMSO added and a reduced primer concentration of 75 nM or the Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs Inc.) with a reduced primer concentration of 250 nM. Agarose gel electrophoresis was performed to evaluate success of the PCR and correct length of amplified fragments. Template DNA was digested with FastDigest DpnI (Thermo Fisher Scientific Inc.) in FastDigest Buffer for 15-60 min at 37 °C followed by heat inactivation at 80 °C for 5 min. PCR amplified fragments were purified with the MinElute PCR Purification Kit (Qiagen N.V.) according to the manufacturer's protocol. In case major side products were observed, fragments with correct length were purified by agarose gel extraction using the QIAquick Gel Extraction Kit (Qiagen N.V.) according to the manufacturer's protocol. Assembly reactions were prepared in a total volume of 20 µL with 0.133 U T5 exonuclease, 0.275 U Phusion® High-Fidelity DNA Polymerase, 2.50 U *Taq* ligase (all from New England Biolabs Inc.) and 1-100 ng of each DNA fragment in 1 × Isothermal Gibson assembly buffer (table 1) and incubated at 50 °C for 15-60 min.

Plasmids generated by Gibson assembly were either directly used to transform *E. coli* bacteria (E. cloni 10G, Lucigen Corp.) or precipitated (2 mL isopropanol, -20 °C, 20 000 RCF) and washed (2 mL 70 % ethanol, -20 °C, 20 000 RCF) beforehand. Plasmids were purified from *E. coli* cultures using the QIAprep Spin Miniprep Kit (Qiagen N.V.) according to the manufacturer's protocol. Coding sequences of cloned plasmids were verified by Sanger sequencing (Eurofins Scientific SE or Microsynth Seqlab GmbH). Genes for bacterial protein expression and purification were cloned into the pET-51b(+) vector (Novagen®, Merck KGaA) with a N-terminal Strep-tag and a C-terminal His-tag or a modified version of the pET-51b(+) vector with an N-terminal His-tag followed by a TEVp cleavage site and no C-terminal tags. Genes for expression in mammalian cells or for generation of mammalian cell lines were cloned into the pCDNA™ 5/FRT vector (Thermo Fisher Scientific Inc.). DNA templates for PCR were obtained from in-house collaborators, ordered at Addgene or custom synthesized at TWIST bioscience.

6.6.3 Protein expression and purification

Expression and purification of recombinant proteins was performed as described by Wilhelm & Kühn et al.[188]. *E. coli* BL21(DE3) (Merck KGaA) were transformed with the respective expression plasmids and grown over night on agar plates with selection antibiotics. Single colonies were picked to inoculate 10 mL of lysogenic broth (LB) medium and pre-cultures were grown over night at 37 °C at 220 rpm. Main-cultures (100–1000 mL LB medium) were inoculated with the pre-cultures (1:100 dilution) and grown at 37 °C at 220 rpm until they reached an optical density at 600 nm of 0.8. Transgene expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and cells were grown at 16 °C overnight or at 30 °C for 4 h in the presence of 1 mM $MgCl_2$. Cells were harvested by centrifugation (10 min, 4000 RCF, 4 °C), resuspended in IMAC lysis buffer (table 1) and lysed by sonication or with a microfluidizer (10 000 psi) on ice. Lysates were cleared by centrifugation (10 min, 75 000 RCF, 4 °C) and proteins were purified via immobilized metal affinity chromatography (IMAC) using a HisTrap FF crude column (Cytiva Europe GmbH) on an ÄktaPure FPLC system (Cytiva Europe GmbH). Strep-tagged proteins were further purified on a StrepTrap HP column (Cytiva Europe GmbH) on the ÄktaPure FPLC system. Buffer was exchanged using a HiPrep 26/10 desalting column (Cytiva Europe GmbH) to activity buffer (table 1).

The purification tag of N-terminally His$_{10}$-TEV tagged HaloTag, cpHaloTag and cpHaloΔ variants, was removed by cleavage with tobacco etch virus protease (TEVp) at 30 °C overnight in TEVp buffer (table 1). Cleaved proteins were purified with a HisTrap FF crude column (Cytiva Europe GmbH) on the ÄktaPure FPLC system (Cytiva Europe GmbH) by collecting the flow-through. TEVp cut proteins were further purified by size exclusion chromatography (HiLoad 26/600 Superdex 75, Cytiva Europe GmbH) in activity buffer (table 1).

All proteins were concentrated using Amicon Ultra-15 centrifugal filter devices (Merck KGaA) with a molecular weight cut-off (MWCO) smaller than the size of the protein to a final concentration of 50-1000 μM. Correct size and purity of proteins were confirmed by SDS-PAGE and mass spectrometry analysis. Protein sequences are listed in the appendix (page 145).

6.6.4 Quantification of peptide solutions

Concentrations of peptides in aqueous solutions or buffers were determined using the Pierce™ Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. The kit contains an amine-reactive fluorescent detection reagent which specifically labels the N-terminus of peptides, enabling quantification by measuring fluorescence intensities (ex: 390 nm, em: 475 nm) in a plate reader against a standard with known peptide concentration. A quadratic function (equation 6.2) was fit to the standard curve and used to calculate peptide concentrations of samples.

$$F([peptide]) = a(b + [peptide])^2 + c \tag{6.2}$$

where:

$F([peptide])$ = fluorescence intensity
$[peptide]$ = peptide concentration

6.6.5 Isothermal titration calorimetry

Proteins used for ITC were dialyzed overnight at 4 °C using Slide-A-Lyzer™ dialysis cassettes (10K MWCO, Thermo Fisher Scientific Inc.) against 400 mL lightly agitated activity buffer (table 1). Potential protein aggregates were removed by centrifugation (21 000 RCF, 10 min, 4 °C) and the protein concentration was adjusted to the target concentration by addition of dialysis buffer. Lyophilized peptides were dissolved in the dialysis buffer by incubation at 50 °C for 5 min with regular vortexing. The peptide was quantified using the Pierce Quantitative fluorometric Peptide Assay (Thermo Fisher Scientific Inc.) as described above (section 6.6.4). The protein was loaded into the cell of the MicroCal PEAQ-ITC (Malvern Panalytical GmbH) instrument, the syringe was loaded with peptide solution. Measurements consisted of 12 3 μL injections. In cases where no plateau could be achieved, a second series of 12 injections was recorded after re-loading the syringe with peptide solution and data from both series were concatenated (ConCat32 software, Malvern Panalytical GmbH). To control for the heat of dilution of the peptide, measurements were performed with dialysis buffer in the cell and peptide solution in the syringe.

Data was analyzed with the MicroCal PEAQ-ITC analysis software (version 1.30) using point-to-point subtraction of the control to fit K_D, ΔH and ΔH offset. Final K_D values were determined by averaging K_D values from three independent measurements and uncertainties were estimated by calculating standard deviations.

To determine the affinity of HTL-TMR labeled cpHaloΔ (cpHaloΔ-TMR) for peptides, cpHaloΔ protein (5 µM) was labeled overnight at 4 °C in presence of 1 mM Hpep1 and 10 mM HTL-TMR substrate in activity buffer (table 1). The labeled protein was concentrated with an Amicon Ultra-15 centrifugal filter device (Merck KGaA) and purified by size exclusion chromatography (HiLoad 26/600 Superdex 75, Cytiva) using activity buffer (table 1). The successful removal of Hpep1 and completion of cpHaloΔ labeling was confirmed by mass spectrometry analysis. ITC measurements and analysis were performed as described above.

6.6.6 Thermostability of proteins

Thermostabilities of proteins were measured by Miroslaw Tarnawski from the Protein core facility of the Max Planck Institute for Medical Research Heidelberg. Melting temperatures were measured in in technical duplicates at 20 µM protein in activity buffer (table 1) on a Prometheus NT 48 nanoscale differential scanning fluorimeter (NanoTemper Technologies GmbH) over a temperature range from 20 °C to 95 °C with a heating rate of 1 °C min^{-1} by monitoring changes in the ratio of the fluorescence intensities at 350 nm and 330 nm. Melting temperatures were determined as the inflection point (maximum of the first derivative) of the fluorescence ratio.

6.6.7 Affinities of HaloTag$^{\text{D106A}}$ and HOB$^{\text{D106A}}$ for fluorophore ligands

Experiments were performed as described by Wilhelm & Kühn et al.[188]. The binding affinities of HaloTag$^{\text{D106A}}$ or HOB$^{\text{D106A}}$ for fluorophore ligands were determined by fluorescence polarization measurements using a plate reader. Fluorophore ligands (10 nM) were titrated against different protein concentrations (0–250 µM) in FP buffer (table 1). Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final volume of 20 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. FP values were averaged and a single-site binding model (equation 6.3) was fit to the data to estimate K_D values for each fluorophore ligand. The FP value of each dye fully reacted with native HaloTag or HOB was used to improve fitting of the upper plateau of the curves by adding an extra data point at a protein concentration of 0.1 M.

$$FP = FP_{free} + \frac{FP_{bound} - FP_{free}}{1 + \dfrac{K_D}{[\text{protein}]}} \tag{6.3}$$

where:

FP = fluorescence polarization
FP_{bound} = maximal fluorescence polarization of fully bound fluorophore (upper plateau)
FP_{free} = fluorescence polarization of the free fluorophore (lower plateau)
K_D = dissociation constant
[protein] = protein concentration

6.6.8 Affinities of HaloTag for methyl-amide fluorophores

Experiments were performed as described by Wilhelm & Kühn et al.[188]. Binding affinities of HaloTag for methyl-amide fluorophores were determined by fluorescence polarization as described above (section 6.6.7) for fluorescent HaloTag ligands with the following changes: Methyl-amide fluorophores were used at a final concentration of 50 nM, and measurements were performed at room temperature. The FP values of HTL derivatives of the respective dyes fully bound to HaloTag protein were used to improve fitting of the upper plateau of the curves by adding an extra data point at a protein concentration of 1 M.

6.6.9 Affinity of HaloTag$^{\mathrm{D106A}}$ for HTL-Ac

Experiments were performed as described by Wilhelm & Kühn et al.[188]. The binding affinity of HaloTag$^{\mathrm{D106A}}$ for HTL-Ac was determined by a fluorescence polarization competition assay against HTL-TMR. 5 µM HaloTag$^{\mathrm{D106A}}$ and 50 nM HTL-TMR were titrated against HTL-Ac concentrations ranging from 80 µM to 10 mM in FP buffer. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final volume of 20 µL using a plate reader. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. FP values were averaged and a four parameter logistic model (equation 6.4) was fit to the data to estimate the I_{50} value. The lower plateau was fixed to the measured FP value of the free dye to improve the fit. The dissociation constant (K_D) of HTL-Ac was calculated as described by Rossi and Taylor[294].

$$FP = FP_{max} + \frac{FP_{free} - FP_{max}}{1 + \left(\frac{I_{50}}{\text{[ligand]}}\right)^{HillCoef}} \tag{6.4}$$

where:

FP = fluorescence polarization
FP_{max} = maximal fluorescence polarization in absence of competing ligand (upper plateau)
FP_{free} = fluorescence polarization of the free fluorophore (lower plateau)
I_{50} = concentration of competing ligand causing 50 % displacement
[ligand] = ligand concentration

6.6.10 Stopped-flow labeling kinetics of HaloTag variants

Experiments were performed as described by Wilhelm & Kühn et al.[188]. Stopped-flow labeling kinetics were measured on a BioLogic SFM-400 (BioLogic Science Instruments Ltd.). Protein and fluorophore substrates in activity buffer (table 1) were mixed in defined ratios and the reaction course was followed by recording fluorescence anisotropy over time at 37 °C. Monochromator excitations wavelengths and emission filters used for different fluorophore substrates are listed in table 5. Baseline anisotropy of the free fluorophore substrates was measured. The dead time of the instrument was determined according to the manufacturer's protocol (BioLogic Technical Note 53) by recording the fluorescence decay during the pseudo-first-order reaction of N-acetyl-L-tryptophanamide with a large excess of N-bromosuccinimide and fitting a first-order rate equation (equation 6.5) to the data.

$$F(t) = Plateau + (Yzero - Plateau) * e^{-k(t-delay)} \tag{6.5}$$

where:

$F(t)$ = fluorescence intensity at time t
$Plateau$ = fluorescence intensity at $\lim_{t\to\infty}$
$Yzero$ = fluorescence intensity at $t = 0$
k = first-order rate constant
$delay$ = dead time of the instrument

Pre-trigger time points were removed, and time values were adjusted according to the dead time of the instrument. Values from replicates were averaged and a two-step reaction model (equation 6.6), described by differential equations 6.7–6.10, was fit to the data globally using DynaFit[274].

$$P + S \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} PS^* \xrightarrow{k_2} PS \tag{6.6}$$

where:

P = protein
S = substrate
PS^* = protein-substrate complex
PS = protein-substrate conjugate

The anisotropy of the free dye was set as fixed offset and it was assumed that the protein substrate complex and the reacted product are contributing equally to the anisotropy signal. Hence, the response for both species was set equal in DynaFit and fit together with the kinetic constants. Standard deviations and confidence intervals of the fit parameters were estimated with the Monte Carlo method[295] ($N = 1000$, 5 % worst fits discarded). Data points and predictions based on the fit models were plotted using R. Fluorescence intensity changes upon protein binding were verified to be minimal (<12 %) and hence not noticeably biasing the fluorescence anisotropy readout.

$$\frac{\mathrm{d[P]}}{\mathrm{d}t} = -k_1[\mathrm{P}][\mathrm{S}] + k_{-1}[\mathrm{PS}^*] \tag{6.7}$$

$$\frac{\mathrm{d[S]}}{\mathrm{d}t} = -k_1[\mathrm{P}][\mathrm{S}] + k_{-1}[\mathrm{PS}^*] \tag{6.8}$$

$$\frac{\mathrm{d[PS^*]}}{\mathrm{d}t} = k_1[\mathrm{P}] - k_{-1}[\mathrm{PS}^*] - k_2[\mathrm{PS}^*] \tag{6.9}$$

$$\frac{\mathrm{d[PS]}}{\mathrm{d}t} = k_2[\mathrm{PS}]^* \tag{6.10}$$

The dissociation constant K_D and the apparent first-order reaction rate constant k_{app} were calculated using equations 6.11 and 6.12.

$$K_D = \frac{k_{-1}}{k_1} \tag{6.11}$$

$$k_{app} = k_1 \frac{k_2}{k_2 + k_{-1}} \tag{6.12}$$

Table 5: Excitation and emission wavelengths for stopped-flow measurements

Fluorophore	Excitation wavelength	Emission filter
TMR, JF_{549}	555 nm	570 nm longpass
CPY	610 nm	630 nm longpass
LIVE580	603 nm	630 nm longpass
JF_{669}	669 nm	690 nm longpass

6.6.11 Labeling kinetics of HaloTag variants measured via plate reader

Experiments were performed as described by Wilhelm & Kühn et al.[188]. Labeling kinetics of HaloTag and HOB with HTL-Alexa488 were measured in FP buffer (table 1) by recording fluorescence polarization over time in a plate reader. The final concentration of the fluorophore substrate was kept constant at 50 nM, protein concentrations were varied from 200 nM–256 µM. Labeling reactions were started by addition of the fluorophore substrate using either a multichannel pipet or the injector module of the plate reader. Assays were performed in black, low-binding, flat bottom, 96-well plates with a final reaction volume of 200 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. The FP of the free HTL-Alexa488 was measured to obtain a baseline.

A simplified one-step reaction model (equation 6.13), described by differential equations 6.14–6.16, was fit to the data globally using DynaFit[274].

$$P + S \xrightarrow{k_{app}} PS \tag{6.13}$$

where:

P = protein
S = substrate
PS = protein-substrate conjugate

The dead time of the setup and the baseline FP value of the fluorophore were set as fixed parameters. Standard deviations and confidence intervals of the fit parameters were estimated with the Monte Carlo method[295] ($N = 1000$, 5 % worst fits discarded). Data points and predictions based on the fit models were plotted using R. Fluorescence intensity changes upon protein binding were verified to be minimal (<12 %) and hence not noticeably biasing the FP readout.

$$\frac{\mathrm{d[P]}}{\mathrm{d}t} = -k_{app}[\mathrm{P}][\mathrm{S}] \tag{6.14}$$

$$\frac{\mathrm{d[S]}}{\mathrm{d}t} = -k_{app}[\mathrm{P}][\mathrm{S}] \tag{6.15}$$

$$\frac{\mathrm{d[PS]}}{\mathrm{d}t} = k_{app}[\mathrm{P}][\mathrm{S}] \tag{6.16}$$

6.6.12 HaloTag competitive labeling kinetics

Experiments were performed as described by Wilhelm & Kühn et al.[188]. Labeling kinetics of non-fluorescent HaloTag ligands were determined indirectly by measuring competition kinetics against HTL-Alexa488 in FP buffer (table 1). Competitive kinetics were measured by recording fluorescence polarization over time in a plate reader. The final concentrations of HTL-Alexa488 (50 nM) and HaloTag protein (200 nM) remained constant with varying concentrations of non-fluorescent substrates (0–1 µM). Assays were performed in black, low-binding, flat bottom, 96-well plate plates with a final reaction volume of 200 µL. Labeling reactions were started by addition of HaloTag protein to wells containing HTL-Alexa488 and non-fluorescent substrates using an electronic 96-channel pipettor. Measurements were conducted in triplicated at 37 °C, filter settings are listed in table 4. The FP value of free HTL-Alexa488 was measured to obtain a baseline.

A simplified competition model (equations 6.17 and 6.18), described by differential equations 6.19–6.23, was fit to the data globally using DynaFit[274].

$$P + S \xrightarrow{k^S_{app}} PS \tag{6.17}$$

$$P + I \xrightarrow{k^I_{app}} PI \tag{6.18}$$

where:

P = protein
S = fluorescent substrate
I = non-fluorescent substrate (inhibitor)
PS = protein-fluorescent substrate conjugate
PI = protein-non-fluorescent substrate conjugate

The dead time of the setup and the baseline FP value of the fluorophore were set as fixed parameters. Standard deviations and confidence intervals of the fit parameters were estimated with the Monte Carlo method[295] ($N = 1000$, 5 % worst fits discarded). Data points and predictions based on the fit models were plotted using R.

$$\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} = -k^S_{app}[\mathrm{P}][\mathrm{S}] - k^I_{app}[\mathrm{P}][\mathrm{I}] \tag{6.19}$$

$$\frac{\mathrm{d}[\mathrm{S}]}{\mathrm{d}t} = -k^S_{app}[\mathrm{P}][\mathrm{S}] \tag{6.20}$$

$$\frac{\mathrm{d}[\mathrm{I}]}{\mathrm{d}t} = -k^I_{app}[\mathrm{P}][\mathrm{I}] \tag{6.21}$$

$$\frac{\mathrm{d}[\mathrm{PS}]}{\mathrm{d}t} = k^S_{app}[\mathrm{P}][\mathrm{S}] \tag{6.22}$$

$$\frac{\mathrm{d}[\mathrm{PI}]}{\mathrm{d}t} = k^I_{app}[\mathrm{P}][\mathrm{I}] \tag{6.23}$$

6.6.13 Background labeling kinetics of cpHaloΔ

Labeling kinetics of cpHaloΔ with HTL-TMR in absence of the Hpep were measured in activity buffer (table 1) by recording fluorescence polarization over time in a plate reader. Final concentration of cpHaloΔ was varied from 2 to 250 µM while HTL-TMR substrate concentration was kept constant at 50 nM. Labeling reactions were started by addition of the fluorophore substrate. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final reaction volume of 40 µL. Measurements were conducted in triplicates at 37 °C, using a humidity cassette to limit evaporation over extended periods of time. Filter settings are listed in table 4. The fluorescence polarization of free HTL-TMR was measured to obtain a baseline.

Data was analyzed as described above for stopped flow kinetic data. However, during model fitting cpHaloΔ concentrations (except for the 250 µM starting solution) were allowed to vary up to 30 % to account for inaccuracies in the dilution series and instead of fitting k_1 and k_{-1} individually only the ratio (i.e., $K_D = k_{-1}/k_1$) was fit. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$, 5 % worst fits discarded). Data points and predictions based on the fit models were plotted using R.

6.6.14 Split-HaloTag labeling by SDS-PAGE

1 µM cpHaloΔ was incubated with an excess of 2 µM HTL-TMR at Hpep concentrations ranging from 0 µM to 1000 µM in activity buffer (table 1) for 1 h at 37 °C. Reactions were stopped by adding 4x Laemmli buffer (Bio-Rad Laboratories, Inc.) and heat denaturated at 95 °C for 10 min. 11.25 pmol of the samples were loaded on polyacrylamide gels (Mini-PROTEAN TGX stain-free 4-20 %, Bio-Rad Laboratories, Inc.) and run for 20 min at 300 V. Gels were imaged on a Typhoon fluorescence laser-scanner (532 nm excitation laser, 570 nm emission filter, 20 nm bandwidth).

6.6.15 Labeling kinetics of complemented split-HaloTag

Labeling kinetics of complemented split-HaloTag were measured in activity buffer (table 1) by recording fluorescence polarization over time in a plate reader. Final concentrations were 500 nM cpHaloΔ protein, 50 nM HTL-TMR substrate and 0 µM to 250 µM Hpep1. Labeling reactions were started by addition of the fluorophore substrate. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final reaction volume of 40 µL. Measurements were conducted in triplicates at 37 °C, using a humidity cassette to limit evaporation over extended periods of time. Filter settings are listed in table 4. The FP of the free HTL-TMR was measured to obtain a baseline.

Data from replicates was averaged and baseline fluorescence polarization was determined. A simplified model describing the reaction mechanism (equations 6.24 and 6.25) was fit to the data using DynaFit[274] to obtain estimates for the apparent second-order rate constant k_{app}. As the K_D of the cpHaloΔ–Hpep1 interaction was known from ITC experiments, the ratio of k_{-1}/k_1 was fixed to 4.64 mM. During model fitting, Hpep concentrations (except for the 250 µM starting solution) were allowed to vary up to 30 % to account for inaccuracies in the dilution series. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$, 5 % worst fits discarded). Data points and predictions based on the fit models were plotted using R.

$$cpHalo\Delta + Hpep \underset{k_{-1}^{Hpep}}{\overset{k_1^{Hpep}}{\longleftrightarrow}} P \tag{6.24}$$

$$P + S \xrightarrow{k_{app}} PS \tag{6.25}$$

where:

P = complemented split-HaloTag

S = fluorescent substrate

PS = complemented split-HaloTag-substrate conjugate

6.6.16 Substrate dependency of split-HaloTag labeling

Labeling kinetics of complemented split-HaloTag were measured in activity buffer (table 1) by recording fluorescence polarization over time in a plate reader. Final concentrations were 500 nM cpHaloΔ protein, 50 nM HaloTag substrate (HTL-TMR, HTL-CPY, HTL-JF$_{669}$, HTL-JF$_{552}$ or HTL-JF$_{525}$) and 1 mM Hpep5. Labeling reactions were started by addition of HaloTag substrate. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final reaction volume of 40 µL. Measurements were conducted in triplicates at 37 °C, using a humidity cassette to limit evaporation over extended periods of time. Filter settings are listed in table 4. The FP of the free substrates was measured to obtain a baseline. A second-order reaction rate equation (equation 6.26) was fit to the data to obtain estimates for the apparent second-order rate constant k_{app}. FP_{free} was fixed to the FP of the free dye in buffer. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

$$FP(t) = FP_{bound} + \frac{FP_{free} - FP_{bound}}{[\mathrm{A}]_0} \cdot \frac{[\mathrm{A}]_0([\mathrm{A}]_0 - [\mathrm{B}]_0)e^{([\mathrm{A}]_0-[\mathrm{B}]_0)k_{app}\cdot t}}{[\mathrm{A}]_0 e^{([\mathrm{A}]_0-[\mathrm{B}]_0)k_{app}\cdot t} - [\mathrm{B}]_0} \tag{6.26}$$

where:

t = time

$FP(t)$ = FP at time t

FP_{free} = FP of the free dye

FP_{bound} = FP of the bound dye

$[\mathrm{A}]_0$ = dye concentration at $t = 0$

$[\mathrm{B}]_0$ = protein concentration at $t = 0$

k_{app} = apparent second-order rate constant

6.6.17 Labeling kinetics of circular permuted HaloTag variants

Labeling kinetics of circular permuted HaloTag variants were measured in FP buffer (table 1) by recording fluorescence polarization over time in a plate reader. Final concentrations were 50 nM protein and 10 nM HTL-TMR substrate. Labeling reactions were started by injection of 100 µL HTL-TMR substrate using the injector module of the plate reader. Assays were performed in black, low-binding, flat bottom, 96-well plates with a final reaction volume of 200 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. The FP of the free HTL-TMR was measured to obtain a baseline. A second-order reaction rate equation (equation 6.26) was fit to the data to obtain estimates for the apparent second-order rate constant k_{app}. FP_{free} was fixed to the FP of the free dye in buffer. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

6.6.18 Caprola labeling kinetics

Labeling kinetics of Caprola variants were measured by recording fluorescence polarization over time in a plate reader. Labeling reactions were performed in FP buffer (table 1) supplemented with 100 µM EGTA and with or without 5 mM $CaCl_2$. Final concentrations of 200 nM Caprola protein and 50 nM HaloTag substrate (HTL-TMR, HTL-CPY, HTL-JF_{669}, HTL-JF_{552} or HTL-JF_{525}) were used. Caprola protein and substrates were prepared in 100 µL and the labeling reactions were started by addition of 100 µL 10 mM $CaCl_2$ in FP buffer (table 1) (5 mM final concentration). Control experiments were conducted where 100 µL of buffer without $CaCl_2$ was added. Assays were performed in black, low-binding, flat bottom, 96-well plates with a final reaction volume of 200 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. The FP of the free fluorophore substrates was measured to obtain a baseline. A second-order reaction rate equation (equation 6.26) was fit to the data to obtain estimates for the apparent second-order rate constant k_{app}. FP_{free} was fixed to the FP of the free dye in buffer. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

To measure the reaction speed of the background labeling reaction in absence of $CaCl_2$ which is orders of magnitudes slower, fluorescence polarization assays were conducted as described above for the control experiments with HTL-TMR. However, a humidity cassette and a much lower sampling rate (1/300s) were used to limit evaporation and enable measurements over 24 h. Data analysis was performed as described above, but since FP did not plateau after 24 h FP_{bound} was fixed to 371 mFP which is the average plateau reached after full labeling of Caprola in presence of Ca^{2+}.

6.6.19 Caprola reversibility assay

Caprola labeling kinetics with HTL-TMR were measured as described above (section 6.6.18) in a plate reader. To start or pause the reaction, 15 µL of $CaCl_2$ or EGTA in FP buffer (table 1) were injected to reach the following final concentrations:

- 150 µM $CaCl_2$
- 300 µM EGTA (when FP reached 185 mFP)
- 50 µM $CaCl_2$
- 2250 µM EGTA (when FP reached 260 mFP)
- 3375 µM $CaCl_2$

6.6.20 pH dependence of Caprola labeling

The pH dependence of Caprola labeling was determined by performing labeling kinetics as described above (section 6.6.18) using FP buffer (table 1) with adjusted pH ranging from 6.8 to 8.

6.6.21 Ca^{2+}-dependence of Caprola labeling

The half-maximal effective Ca^{2+} concentration (EC_{50}) for Caprola activation was determine by measuring labeling kinetics at different, buffered free Ca^{2+} concentrations. Kinetics were recorded by measuring fluorescence polarization over time in a plate reader. To obtain buffers with precise and buffered free Ca^{2+} concentrations from 0 µM to 39 µM a calcium calibration kit (Thermo Fisher Scientific Inc.) was used (100 mM KCl, 30 mM MOPS, pH 7.2 and K_2EGTA / CaEGTA in different ratios). Labeling reactions were started by mixing equal volumes of Caprola protein (200 nM final) and HTL-TMR (50 nM final) substrate in the respective Ca^{2+} kit buffers. Assays were performed in black, low-binding, flat bottom, 96-well plates with a final reaction volume of 200 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. The FP of the free HTL-TMR was measured to obtain a baseline.

Data from replicates were averaged and a second-order reaction rate equation (equation 6.26) was fit to the data in conditions where at least 200 mFP were reached to obtain estimates for the apparent second-order rate constant k_{app}. The initial slope at time 0 ($s_{t=0}$) was calculated using the derivative of equation 6.26 at $t = 0$ (equation 6.27).

$$s_{t=0} = \frac{\mathrm{d}FP}{\mathrm{d}t}(t=0) = [\mathrm{B}]_0 \cdot k_{app}(FP_{free} - FP_{bound}) \tag{6.27}$$

where:

t = time

$s_{t=0}$ = initial slope at $t = 0$

FP_{free} = FP of the free dye

FP_{bound} = FP of the bound dye

$[\mathrm{B}]_0$ = protein concentration at $t = 0$

k_{app} = apparent second-order rate constant

In conditions where 200 mFP were not reached, due to the reaction being too slow, a linear model (equation 6.28) was fit to the data to determine the initial slope $s_{t=0}$.

$$FP(t) = FP_{free} + t \cdot s_{t=0} \tag{6.28}$$

where:

t = time

$s_{t=0}$ = initial slope at $t = 0$

FP_{free} = FP of the free dye

Initial slopes were plotted against the free Ca^{2+} concentration and a sigmoidal dose response model (equation 6.29) was fit to this data to estimate EC_{50} values. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

$$s_{t=0}([\mathrm{Ca}^{2+}_{\mathrm{free}}]) = \frac{s_{max}}{1 + 10^{(log_{10}(EC_{50} - log10([\mathrm{Ca}^{2+}_{\mathrm{free}}])) * HillCoef}} \tag{6.29}$$

where:

$s_{t=0}$ = initial slope at $t = 0$

s_{max} = maximal initial slope at $t = 0$

$[\mathrm{Ca}^{2+}_{\mathrm{free}}]$ = free Ca^{2+} concentration

EC_{50} = half-maximal effective concentration

$HillCoef$ = Hill coefficient

6.6.22 Multi-color sequential Caprola labeling

His_{10}-tagged Caprola9 protein (5 nmol) in activity buffer (table 1) was immobilized on a 5 mL HisTrap FF crude column (Cytiva) on an ÄktaPure FPLC system (Cytiva). The column was washed with 5 column volumes of EGTA buffer (50 mM HEPES, 50 mM NaCl, 100 µM EGTA, pH 7.3). Then two rounds of substrate incubation – each with 2 column volumes of HTL-JF_{552} or HTL-CPY (1 nmol) in either EGTA or Ca^{2+} buffer (50 mM HEPES, 50 mM NaCl, 5 mM $CaCl_2$, pH 7.3) – were applied with a constant flow of 5 mL min^{-1}. The column was washed with 5 column volumes of EGTA buffer in between substrate incubation rounds. Labeled Caprola protein was eluted with 3 column volumes of IMAC elution buffer (table 1) and concentrated using Amicon Ultra-15 centrifugal filter devices (Merck). Labeling was revealed by SDS-PAGE followed by in-gel fluorescence scanning as described above for split-HaloTag labeling (section 6.6.14). JF_{552}: 532 nm excitation laser, 570 nm emission filter, 20 nm bandwidth; CPY: 635 nm excitation laser, 670 nm emission filter, 30 nm bandwidth.

6.6.23 Labeling kinetics of split-HaloTag–FKBP/FRB fusions

Labeling kinetics of split-HaloTag–FKBP/FRB fusions were measured in FP buffer (table 1) by recording fluorescence polarization over time in a plate reader. Final concentrations were 250 nM of each protein, 50 nM HaloTag substrate (HTL-TMR, HTL-CPY, HTL-JF_{669}, HTL-JF_{552} or HTL-JF_{525}) and

500 nM rapamycin. Labeling reactions were started by addition of rapamycin. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final reaction volume of 40 µL. Measurements were conducted in triplicates at 37 °C, filter settings are listed in table 4. The FP of the free fluorophores was measured to obtain a baseline.

A second-order reaction rate equation (equation 6.26) was fit to the data in conditions where at least 200 mFP were reached to obtain estimates for the apparent second-order rate constant k_{app}. The initial slope at time 0 ($s_{t=0}$) was calculated using the derivative of equation 6.26 at $t = 0$ (equation 6.27). In conditions where 200 mFP were not reached, due to the reaction being too slow, a linear model (equation 6.28) was fit to the data to determine the initial slope $s_{t=0}$. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

6.6.24 Hpep library screening

The Hpep library consisting of 384 synthetic peptides was screened for the ability to activate the labeling reaction of cpHaloΔ. Peptides were obtained in crude quality in 1–4 mg scale. To dissolve the peptides, 0.5 mL MQ water were added to each tube and tubes were shaken for 1 h at 1800 rpm at 45 °C. Tubes were centrifuged for 10 min, 4000 RCF, at room temperature to sediment any undissolved peptide. 20 µL of the supernatants were transferred into a black, non-binding, flat bottom, low volume, 384-well plate. 10 µL cpHaloΔ protein (0.5 µM final) in activity buffer (table 1) was added and the labeling reaction was started by addition of 10 µL of HTL-TMR substrate (0.1 µM final) in activity buffer (table 1). Labeling kinetics were measured by recording fluorescence polarization over time in a plate reader at 37 °C. A humidity cassette was used to limit evaporation over extended periods of time.

A second-order reaction rate equation (equation 6.26) was fit to the data in conditions where at least 200 mFP were reached to obtain estimates for the apparent second-order rate constant k_{app}. The initial slope at time 0 ($s_{t=0}$) was calculated using the derivative of equation 6.26 at $t = 0$ (equation 6.27). In conditions where 200 mFP were not reached, due to the reaction being too slow, a linear model (equation 6.28) was fit to the data to determine the initial slope $s_{t=0}$. Initial slopes were normalized to the parental Hpep1 to obtain relative labeling speeds. Negative slopes were arbitrarily set to $5 \cdot 10^{-2}$ for plotting on a logarithmic scale.

6.6.25 Half-maximal effective concentration (EC_{50}) of Hpep variants

To determine the half-maximal effective concentration (EC_{50}) of purified Hpep variants for split-HaloTag activation, labeling kinetics of cpHaloΔ were measured in FP buffer (table 1) as a function of peptide concentration by recording fluorescence polarization over time in a plate reader. For peptides with high EC_{50} (> 1 µM) final concentrations of 500 nM cpHaloΔ protein and 100 nM HTL-TMR substrate were used, for peptides with low EC_{50} (< 1 µM) final concentrations of 20 nM cpHaloΔ protein and 4 nM HTL-TMR substrate were used. The Hpep was titrated in concentrations from 0 mM to 5 mM, depending on the EC_{50} of the respective peptide. Labeling reactions were started by addition of HTL-TMR substrate. Assays were performed in black, non-binding, flat bottom, low volume, 384-well plates with a final reaction volume of 40 µL. Measurements were conducted in triplicates at 37 °C, using a humidity cassette to limit evaporation over extended periods of time. Filter settings are listed in table 4. The FP of the free HTL-TMR was measured to obtain a baseline.

A second-order reaction rate equation (equation 6.26) was fit to the data in conditions where at least 200 mFP were reached to obtain estimates for the apparent second-order rate constant k_{app}. In conditions where 200 mFP were not reached, due to the reaction being too slow, a linear model (equation 6.28) was fit to the data to determine the initial slope $s_{t=0}$. Apparent second-order rate constants were back calculated from initial slopes using the derivative of equation 6.26 at $t = 0$ (equation 6.27). FP_{bound} was set to the median plateau observed at higher Hpep concentrations.

Apparent second-order rate constants k_{app} were plotted against the Hpep concentration and a sigmoidal dose response model (equation 6.30) was fit to this data to estimate EC_{50} values and maximal apparent second-order rate constants (k_{app}^{max}) at saturating Hpep concentrations. Uncertainties and confidence intervals of the fit parameters were estimated using the Monte Carlo method ($N = 1000$).

$$s_{t=0}([Hpep]) = \frac{s_{max}}{1 + 10^{(log_{10}(EC_{50}) - log_{10}([Hpep]))}} \tag{6.30}$$

where:

$s_{t=0}$ = initial slope at $t = 0$
s_{max} = maximal initial slope at $t = 0$
$[Hpep]$ = Hpep concentration
EC_{50} = half-maximal effective concentration

6.7 Computational Methods

6.7.1 Design of C-terminal Hpep extensions

The crystal structure of HaloTag labeled with HTL-TMR (PDB-ID: 6Y7A)[188] was downloaded from the PDB. Water, glycerol, chloride ion and covalently bound HTL-TMR (“OEH”) residues were deleted and the file was modified by changing the “ASP A 106” residue to “ASX A 106”. The oxygen connecting the aspartate with the covalently bound HaloTag ligand (“OD2”) was removed. Residue numbering was adjusted to start at 1 and residues 140, 141, 142 and 153 which are not in split-HaloTag were removed. The numbering was adjusted to resemble the circular permuted cpHaloΔ with Hpep1 (ARETFQAFRT) at the end of the sequence (C-terminal fragment → N-terminal fragment → Hpep1). A Rosetta remodel blueprint file was generated with the `getBluePrintFromCoords.pl` script.

```
~/Rosetta/main/tools/remodel/getBluePrintFromCoords.pl -pdbfile split_halo.pdb
    -chain A > resfile
```

The first lines of the file were modified to extend the protein by two glycine residues in order to limit potential bias by the charged terminus of cpHaloΔ in direct proximity to the Hpep1 extension.

```
1 x L PIKAA G
1 x L PIKAA G
1 D L PIKAA D
2 V H PIKAA V
3 G H PIKAA G
4 R H PIKAA R
5 K H PIKAA K
```

```
6 L .
7 I .
[...]
```

The last lines of the file were modified to extend Hpep1 by four residues.

```
[...]
286 A .
287 F .
288 R H PIKAA R
289 T L PIKAA T
0   X L ALLAAxc
0   X L ALLAAxc
0   X L ALLAAxc
0   X L ALLAAxc
```

RosettaRemodel[263] was run with the following command and flags.

```
~/Rosetta/main/source/bin/remodel.mpi.linuxgccrelease @flags
```

```
-nstruct 10000
-run:chain A
-use_clusters false
-hb_srbb 1.0
-find_neighbors
-packing:linmem_ig 10
-use_input_sc
-no_optH false
-flip_HNQ
-ex1
-ex2
-packing:extrachi_cutoff 3
-packing:ex1aro
-packing:ex2aro
-renumber_pdb true
-in:file:fullatom
-in:file:s in/split_halo.pdb
-remodel:blueprint in/resfile
--out:path:all out
```

6.7.2 Design of new 10-mer and N-terminally extended 14mer Hpep variants

6.7.2.1 Preparation of the split-HaloTag structure

The crystal structure of HaloTag labeled with HTL-TMR (PDB-ID: 6Y7A)[188] was downloaded from the PDB. Water, glycerol, chloride ion and covalently bound HTL-TMR ("OEH") residues were deleted and the file was modified by changing the "ASP A 106" residue to "ASX A 106". The oxygen connecting the aspartate with the covalently bound HaloTag ligand ("OD2") was removed. Residue numbering was adjusted to start at 1 and residues 140, 141, 142 and 153 which are not in split-HaloTag were removed. The Hpep1 10-mer (ARETFQAFRT) was moved to the start of the sequence and the `getBluePrintFromCoords.pl` script was used to generate a Rosetta remodel blueprint file.

```
~/Rosetta/main/tools/remodel/getBluePrintFromCoords.pl -pdbfile split_halo.pdb
    -chain A > resfile
```

The first lines of the file were modified to extend Hpep1 by four glycine residues.

```
1 x H PIKAA G
1 x H PIKAA G
1 x H PIKAA G
1 x H PIKAA G
1 A H PIKAA A
2 R .
3 E .
[...]
```

RosettaRemodel[263] was run with the following command and flags.

```
~/Rosetta/main/source/bin/remodel.mpi.linuxgccrelease @flags
```

```
-nstruct 1
-run:chain A
-use_clusters false
-hb_srbb 1.0
-find_neighbors
-packing:linmem_ig 10
-use_input_sc
-no_optH false
-renumber_pdb true
-auto_setup_metals true
-in:file:fullatom
-in:file:s in/split_halo.pdb
-remodel:blueprint in/resfile
--out:path:all out
```

The output structure was used for the subsequent steps.

In order to limit potential bias by the charged termini of cpHaloΔ in direct proximity to Hpep1, glycine-glycine spacers were attached to the termini. For N-terminal spacers residue numbering was adjusted to resemble the circular permuted cpHaloΔ with Hpep1 (ARETFQAFRT) at the end of the sequence (C-terminal fragment → N-terminal fragment → Hpep1) and a Rosetta remodel blueprint file was generated with the `getBluePrintFromCoords.pl` script. The first lines of the file were modified to extend the protein by two glycine residues.

```
1 x L PIKAA G
1 x L PIKAA G
1 D L PIKAA D
2 V H PIKAA V
3 G H PIKAA G
4 R H PIKAA R
5 K H PIKAA K
6 L .
7 I .
[...]
```

RosettaRemodel[263] was run with the following command and flags.

```
~/Rosetta/main/source/bin/remodel.mpi.linuxgccrelease @flags
```

```
-nstruct 100
-run:chain A
-use_clusters false
-hb_srbb 1.0
-find_neighbors
-packing:linmem_ig 10
-use_input_sc
-no_optH false
-flip_HNQ
-ex1
-ex2
-packing:extrachi_cutoff 3
-packing:ex1aro
-packing:ex2aro
-renumber_pdb true
-in:file:fullatom
-in:file:s in/split_halo.pdb
-remodel:blueprint in/resfile
--out:path:all out
```

The best scoring structure was used for the subsequent steps.

For C-terminal spacers residue numbering was adjusted to place Hpep1 at the start of the sequence and the `getBluePrintFromCoords.pl` script was used to generate a Rosetta remodel blueprint file. The last lines of the file were modified to extend the protein by two glycine residues.

```
[...]
293 D .
294 E .
295 W L PIKAA W
0   X L PIKAA G
0   X L PIKAA G
```

RosettaRemodel[263] was run with the same flags and commands as for the N-terminal spacers. The best scoring structure was modified by moving Hpep1 to chain B and renumbering all residues.

6.7.2.2 Preparation of the covalent ligand and relaxing the split-HaloTag-TMR structure

In order to parameterize the covalently bound HTL-TMR ligand for usage in Rosetta, the ligand including the ester oxygen was extracted from the HTL-TMR labeled HaloTag structure (PDB-ID: 6Y7A)[188]. Hydrogen atoms were added in PyMOL and the hydrogen on the former ester oxygen was removed. The molecule was saved as a pdb file and converted to a mol file with Avogadro. The mol file was edited manually, changing the bond types to “4” (aromatic) for all aromatic bonds in the molecule. The ligand was parameterized using the `molfile_to_params.py` script.

```
~/Rosetta/main/source/scripts/python/public/molfile_to_params.py -n TMR -p HTL-TMR
    04_TMR.mol
```

The resulting params file was modified in the following manner:

- adding a “CONNECT O1” line between the “BOND_TYPE” and “CHI” lines
- removing the “CHARGE O1 FORMAL -1” line
- adding last “ICOOR_INTERNAL” line with the internal coordinates of the gamma carbon atom of ASP106 seen from O1, C21, C20 of the ligand to build the connection between ligand and protein

The content of “HTL-TMR_0001.pdb” was copied into the previously prepared split-HaloTag pdb file (as residue number 284, chain A). The chloride ion atom from the 6Y7A structure was added (as residue 285, chain A). The “ASX.params” file from the Rosetta database was copied and modified by removing the virtual atom “V1” (including “BOND” and “CHI” record). (`Rosetta/main/database/chemical/residue_type_sets/fa_standard/residue_types/sidechain_conjugation/ASX.params`). A constraints file was set up to fix the bond geometry between the ASX and the covalently bound HTL-TMR residue.

```
AtomPair  CG 104 O1 294                HARMONIC 1.265      0.001
Angle     CB 104 CG 104 O1 294         HARMONIC 2.05034299 0.034906585
Angle     CG 104 O1 294 C21 294        HARMONIC 2.0356822  0.034906585
Dihedral  CA 104 CB 104 CG 104 O1 294  HARMONIC 1.32936493 0.034906585
```

The structure was relaxed with the following command and flags. The best scoring structure was used for the subsequent steps.

```
~/Rosetta/main/source/bin/relax.mpi.linuxgccrelease @flags
```

```
-nstruct 10000
-cst_fa_file in/chemical_bond.cst
-constraints:cst_fa_weight 1
-relax:ramp_constraints false
-relax:constrain_relax_to_start_coords
-ex1
-ex2
-use_input_sc
-flip_HNQ
-packing:extrachi_cutoff 16
-packing:ex1aro
-packing:ex2aro
-no_optH false
-renumber_pdb true
-auto_setup_metals true
-in:file:extra_res_path in/params
-in:file:fullatom
-in:file:s in/gg_split_halo_gg.pdb
--out:path:all out
```

6.7.2.3 Design of new 10-mer Hpep variants

To design new 10-mer Hpep variants the relaxed split-HaloTag-TMR structure from the previous steps was used. The four glycine residues extending Hpep1 were removed and a RosettaScripts[260] protocol was set up with the following steps.

- setting up constraints and custom fold tree
- perturbing the N- and C-terminal glycine-glycine spacers
- minimizing the spacers with FastRelax
- docking the Hpep to cpHaloΔ
- designing all Hpep residues using conservative design based on the BLOSUM80 matrix with FastDesign using a custom relax script

The XML Rosetta script can be found in the appendix (page 154). The same params file and bond constraints file as described above for relaxing the structure were used.

Fold tree:

```
FOLD_TREE  EDGE 246 1 -1  EDGE 246 283 -1 EDGE 246 284 1  EDGE 246 285 2  EDGE 246 291 3  EDGE 291 286 -1  EDGE 291 295 -1
```

Custom relax script:

```
repeat %%nrepeats%%
coord_cst_weight 1.0
scale:fa_rep 0.8
repack
scale:fa_rep 0.805
min 0.01
coord_cst_weight 0.5
scale:fa_rep 0.847
repack
scale:fa_rep 0.853
min 0.01
coord_cst_weight 0.0
scale:fa_rep 0.908
repack
scale:fa_rep 0.917
min 0.01
coord_cst_weight 0.0
scale:fa_rep 1
repack
min 0.00001
accept_to_best
endrepeat
```

Resfile:

```
NATAA
START

286 B ALLAAxc
287 B ALLAAxc
288 B ALLAAxc
289 B ALLAAxc
290 B ALLAAxc
291 B ALLAAxc
292 B ALLAAxc
```

```
293 B ALLAAxc
294 B ALLAAxc
295 B ALLAAxc
```

Command and flags:

```
~/Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease @flags
```

```
-nstruct 10000
-parser:protocol in/design_Hpep.xml
-in:file:extra_res_path in/par
-in:file:fullatom
-in:file:s in/gg_split_halo_gg_relaxed.pdb
-out:path:all out
```

6.7.2.4 Design of N-terminally extended 14mer Hpep variants

To design new N-terminally extended 14mer Hpep variants the relaxed split-HaloTag-TMR structure from the previous steps was used. Three different RosettaScripts[260] protocol were set up with the following overall structure.

- setting up constraints and custom fold tree
- perturbing the N- and C-terminal glycine-glycine spacers
- minimizing the spacers with FastRelax
- docking the Hpep to cpHaloΔ
- designing all Hpep residues with FastDesign using the custom relax script described above

In the first run the script was set up as described above, in the second run the FastDesign step was conducted with more repeats and in a third run the backbone of cpHaloΔ was completely fixed. The XML Rosetta scripts can be found in the appendix (page 154). The same params file and bond constraints file as described above for relaxing the structure were used.
Fold tree:

```
FOLD_TREE  EDGE 246 1 -1  EDGE 246 283 -1 EDGE 246 284 1  EDGE 246 285 2  EDGE 246
    293 3  EDGE 293 286 -1  EDGE 293 299 -1
```

Resfile:

```
NATAA
START

286 B ALLAAxc
287 B ALLAAxc
288 B ALLAAxc
289 B ALLAAxc
290 B ALLAAxc
291 B ALLAAxc
292 B ALLAAxc
293 B ALLAAxc
```

```
294 B ALLAAxc
295 B ALLAAxc
296 B ALLAAxc
297 B ALLAAxc
298 B ALLAAxc
299 B ALLAAxc
```

Command and flags:

```
~/Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease @flags
```

```
-nstruct 10000
-parser:protocol in/design_Hpep.xml
-in:file:extra_res_path in/params
-in:file:fullatom
-in:file:s in/gg_split_halo_gg_relaxed.pdb
-out:path:all out
```

6.7.3 Analysis of computational Hpep design and Hpep library compilation

For each design of 10-mer and N-terminally extended 14mer Hpep variants generated by Rosetta, sequences were extracted and a final score was calculated for each design by adding total score and ΔΔG values output by Rosetta. Hpep variants were ranked by this final score and the top 20 % designs of each Rosetta run described above (sections 6.7.1 and 6.7.2) were analyzed for the most beneficial Hpep mutations relative to the parental sequence. All occurring mutations were identified and a mutation score was calculated by summing up the final scores of all designs the mutation appeared in subtracted by the median final score off all designs of this run. The top 20 % designs with the N- or C-terminally extended 14mer Hpep variants were analyzed for the most beneficial four amino acid extensions similarly as described above for mutations. Extensions were ranked by the sum of their final scores subtracted by the median final total score of all designs of this run.

A list of Hpep candidates was compiled by pooling the top ranked sequences from the different Rosetta runs, the top scoring mutations as single point mutants of the parental Hpep and parental Hpeps extended N- or C-terminally with top scoring extensions. This resulted in 1039 sequences in total out of which 402 were unique. Control sequences were added to the list: the parental peptide (ARETFQAFRT), a shortened parental peptide (ARETFQAFR) known to have strongly decreased activity and negative control peptide (ARETAQAFRT, double F→A mutation) known to have no activity. All sequences were analyzed with the ThermoFisher Scientific "peptide synthesis and proteotypic peptide analyzing tool". Peptides were ranked by hydrophobicity and the 384 least hydrophobic peptides were chosen for the final Hpep library. The Hpep library was ordered as synthetic peptides in 1–4 mg scale in crude quality (GenScript).

6.7.4 Calculation of estimated effective concentrations

Effective concentrations (C_{eff}) were estimated using a worm-like chain (WLC) model via the C$_{\mathrm{eff}}$-Calculator application[257]. A persistence length (L_p) of 3 Å and a amino acid size (b) of 3.8 Å was used to resemble the behavior flexible linkers.

6.8 Cell Biology

6.8.1 General information

Mammalian cells were cultured in high glucose (4.5 g L^{-1}) DMEM with GlutaMax™ (Gibco™, Thermo Fisher Scientific Inc.) supplemented with 110 mg L^{-1} sodium pyruvate, 10 % FBS and phenol red at 37 °C (herein termed "medium") in a humidified incubator with 5 % CO_2 atmosphere. Cells were passaged every 2-4 days before reaching 100 % confluency. Therefor, cells were washed with PBS and incubated with 1X TrypLE™ Express Enzyme (Gibco™, Thermo Fisher Scientific Inc.) for 5 min at 37 °C. After dissociating the cells by pipetting up and down, they were seeded in medium to the desired density. Cultures were regularly tested for mycoplasma contamination by PCR. Cells used for imaging experiments were maintained in medium without phenol red. Cell lines were stored in FBS containing 10 % DMSO at -80 °C. Mammalian cell lines used in this thesis are listed in table 2.

6.8.2 Generation of stable cell lines

HEK-293 cell lines constitutively expressing transgenes were prepared using the Invitrogen™ Flp-In™ system (Thermo Fisher Scientific Inc.). HEK-293 Flp-In™ were cultured to 80 % confluency in 8-well dishes and transfected with a 1:10 mixture of donor pCDNA™ 5/FRT plasmid (encoding the transgene and the Flp recombination target, FRT) to pOG44 plasmid (encoding the Flp recombinase) using lipofectamine™ 3000 (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. After 8–16 h medium was exchanged and cells were maintained in medium containing 100 µg mL^{-1} Hygromycin B (Gibco™, Thermo Fisher Scientific Inc.) for 3 days to select cells with successful transgene integration. Cells were then recovered in medium without Hygromycin B for 2 days. In order to obtain clonal lines, cells were seeded into 96-well plates at a densities of 2–4 cells per well and grown for 5 days until colonies were well observable. Wells with single colonies expressing the transgene were identified by fluorescence microscopy and clonal cell lines were cultured as described above.

6.8.3 Chemical fixation of cells for fluorescence microscopy

Cells were washed with PBS and then incubated in PBS containing 4 % paraformaldehyde for 15 min at room temperature. Cells were again washed with PBS and kept at 4 °C until imaging.

6.8.4 Immunofluorescence staining and confocal imaging

For immunofluorescence, cells were seeded in 96-well glass bottom imaging plates (Eppendorf) and grown for at least 2 days. Staining of cells was performed at room temperature in the following steps:

- cells were chemically fixed as described above (section 6.8.3)
- washed once with PBS
- incubated in PBS with 0.1 % Triton™ X-100 for 10 min
- washed with PBS-T (table 1)
- blocked in 3 % BSA in PBS-T for 1 h
- incubated with primary antibody in 3 % BSA in PBS-T for 2 h

- washed three times with PBS-T
- incubated with secondary antibody in 3 % BSA in PBS-T for 1 h
- washed twice with PBS-T
- washed twice with PBS

Antibodies used for immunofluorescence are listed in table 6. Confocal microscopy was performed on a Stellaris 5 inverted microscope (Leica) equipped with a white line laser, a 405 nm diode laser and hybrid photodetectors at 37 °C. A 63x/1.20 water immersion objective was used at to image a single plane at 512x512 pixel (92x92 μm) resolution (2 x zoom, 400 Hz scan speed, 112 μm pinhole, 8-fold line average, mTagBFP2: 405 nm excitation 430–478 nm emission, EGFP: 488 nm excitation 500–550 nm emission, Alexa568: 580 nm excitation 600–740 nm emission). Due to the high variability of expression levels, brightness and contrast of images were adjusted using FIJI[273].

Table 6: Antibodies used in this thesis

Name	Source	Clonality	Conc.[a]	Provider
Anti-V5-tag	mouse	monoclonal	2 μg mL^{-1}	Thermo Fisher Sci. Inc.
Anti-FLAG-tag	rabbit	polyclonal	2 μg mL^{-1}	Merck KGaA
Anti-Rabbit IgG-Alexa568	goat	polyclonal	1 μg mL^{-1}	Thermo Fisher Sci. Inc.
Anti-Mouse IgG-Alexa568	donkey	polyclonal	1 μg mL^{-1}	Thermo Fisher Sci. Inc.

[a] Conc: working concentration

6.8.5 Split-HaloTag based labeling of FKPB-FRB interactions in HeLa cells

Constructs for co-expression of Split-HaloTag-FKBP/FRB were transiently transfected into HeLa cells using lipofectamine™ 3000 (Thermo Fisher Scientific Inc.) according to the manufacturer's protocol. After 8 h incubation, the medium was replaced and cells were maintained for additional 24 h. Cells were labeled with 100 nM HTL-CPY in presence or absence of 100 nM rapamycin in medium for 1 h. Cells were washed 3 times with fresh medium containing 1 μM HaloTag protein to scavenge the HTL-CPY substrate before analysis via flow cytometry or confocal fluorescence microscopy.

For flow cytometry cells were detached using 1X TrypLE™ Express Enzyme (Gibco™, Thermo Fisher Scientific Inc.) for 5 min at 37 °C. Detached cells were suspended in PBS containing 2 % FBS and analyzed with a BD Fortessa X-20 flow cytometer (mEGFP: 488 nm excitation, 530 nm emission filter with 30 nm bandwidth, CPY: 648 nm excitation, 660 nm emission filter with 20 nm bandwidth). Live, single and mEGFP positive cells were gated (see fig. S3 for gating strategy) and fluorescence intensity ratios (HTL-CPY/mEGFP) were calculated.

For confocal microscopy cells were chemically fixed in 96-well glass bottom imaging plates (Eppendorf) as described above (section 6.8.3). Confocal microscopy was performed on a Stellaris 5 inverted microscope (Leica) equipped with a white line laser, a 405 nm diode laser and hybrid photodetectors at 37 °C. A 40x/1.10 water immersion objective was used at to image a single plane at 1024x1024 pixel (194x194 μm) resolution (1.5 x zoom, 600 Hz scan speed, 68.8 μm pinhole, 32-fold line average, mEGFP: 488 nm excitation 494–556 nm emission, CPY: 610 nm excitation 618–750 nm emission). Images were used without further processing.

6.8.6 SynProLa co-culture assay

HEK-293 cells co-expressing fluorescent markers (mTagBFP2 or mEGFP) and SynProLa constructs (split-HaloTag fusions to NRX or NLG) were obtained either by generating stable cell lines as described above (section 6.8.2) or by transient transfection using lipofectamine™ 3000 (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. 96-well glass bottom imaging plates (Eppendorf) were coated with bovine fibronectin (25 µg mL^{-1} in PBS, 2 h, 37 °C) and washed once with PBS. For co-culture, NRX and NLG expressing cells were detached using 1X TrypLE™ Express Enzyme (Gibco™, Thermo Fisher Scientific Inc.), mixed in 1:1 stoichiometry and seeded in the fibronectin coated plates. Cells were then grown for 2–4 days to reach almost 100 % confluency and form cell-cell contacts. Complemented split-HaloTag at intercellular NRX-NLG complexes was labeled by incubation with 100 nM HTL-CPY in growth medium for 1 h. Cells were washed twice with fresh medium containing 1 µM HaloTag protein to scavenge the HTL-CPY substrate before being chemically fixed as described above (section 6.8.3). Confocal microscopy was performed on a Stellaris 5 inverted microscope (Leica) equipped with a white line laser, a 405 nm diode laser and hybrid photodetectors at 37 °C. A 40x/1.10 water immersion objective was used at to image a single plane at 1024x1024 pixel (145x145 µm) resolution (2 x zoom, 200 Hz scan speed, 77.2 µm pinhole, 4-fold line average, mTagBFP2: 405 nm excitation 430–494 nm emission, EGFP: 488 nm excitation 494–590 nm emission, CPY: 612 nm excitation 617–730 nm emission). To limit crosstalk between the mTagBFP2 and EGFP channels, channels were acquired sequentially (mTagBFP2 and EGFP plus CPY). Due to the high variability of expression levels, brightness and contrast of images were adjusted using FIJI[273].

Bibliography

[1] Kandel, E. R., Schwartz, J. H., Jessell, T. M. *et al.* (eds.) *Principles of Neural Science* (McGraw-Hill, New York, 2013), 5th ed edn.

[2] Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* **20**, 353–364 (2017).

[3] Coenen, A., Fine, E. & Zayachkivska, O. Adolf Beck: A Forgotten Pioneer in Electroencephalography. *Journal of the History of the Neurosciences* **23**, 276–286 (2014).

[4] Berger, H. Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten* **87**, 527–570 (1929).

[5] Hämäläinen, M., Hari, R., Ilmoniemi, R. J. *et al.* Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* **65**, 413–497 (1993).

[6] Sharoh, D., Van Mourik, T., Bains, L. J. *et al.* Laminar specific fMRI reveals directed interactions in distributed networks during language processing. *Proceedings of the National Academy of Sciences* **116**, 21185–21190 (2019).

[7] Glover, G. H. Overview of Functional Magnetic Resonance Imaging. *Neurosurgery Clinics of North America* **22**, 133–139 (2011).

[8] Lameka, K., Farwell, M. D. & Ichise, M. Positron Emission Tomography. In *Handbook of Clinical Neurology*, vol. 135, 209–227 (Elsevier, 2016).

[9] Hamill, O. P., Marty, A., Neher, E. *et al.* Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv - European Journal of Physiology* **391**, 85–100 (1981).

[10] Tao, C., Zhang, G., Xiong, Y. & Zhou, Y. Functional dissection of synaptic circuits: In vivo patch-clamp recording in neuroscience. *Frontiers in Neural Circuits* **9** (2015).

[11] Gao, J., Liao, C., Liu, S. *et al.* Nanotechnology: New opportunities for the development of patch-clamps. *Journal of Nanobiotechnology* **19**, 97 (2021).

[12] Chemla, S. & Chavane, F. Voltage-sensitive dye imaging: Technique review and models. *Journal of Physiology-Paris* **104**, 40–50 (2010).

[13] Looger, L. L. & Griesbeck, O. Genetically encoded neural activity indicators. *Current Opinion in Neurobiology* **22**, 18–23 (2012).

[14] Knöpfel, T. & Song, C. Optical voltage imaging in neurons: Moving from technology development to practical tool. *Nature Reviews Neuroscience* **20**, 719–727 (2019).

[15] Dimitrov, D., He, Y., Mutoh, H. *et al.* Engineering and Characterization of an Enhanced Fluorescent Protein Voltage Sensor. *PLoS ONE* **2**, 1–5 (2007).

[16] Gong, Y., Huang, C., Li, J. Z. *et al.* High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor. *Science* **350**, 1361–1366 (2015).

[17] Abdelfattah, A. S., Kawashima, T., Singh, A. *et al.* Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science* **365**, 699–704 (2019).

[18] Platisa, J., Ye, X., Ahrens, A. M. *et al.* High-speed low-light in vivo two-photon voltage imaging of large neuronal populations. *Nature Methods* **20**, 1095–1103 (2023).

[19] Jaffe, D. B., Johnston, D., Lasser-Ross, N. *et al.* The spread of Na+ spikes determines the pattern of dendritic Ca2+ entry into hippocampal neurons. *Nature* **357**, 244–246 (1992).

[20] Smetters, D., Majewska, A. & Yuste, R. Detecting Action Potentials in Neuronal Populations with Calcium Imaging. *Methods* **18**, 215–221 (1999).

[21] Tsien, R. Y. New calcium indicators and buffers with high selectivity against magnesium and protons: Design, synthesis, and properties of prototype structures. *Biochemistry* **19**, 2396–2404 (1980).

[22] Russell, J. T. Imaging calcium signals *in vivo* : A powerful tool in physiology and pharmacology. *British Journal of Pharmacology* **163**, 1605–1625 (2011).

[23] Inoue, M. Genetically encoded calcium indicators to probe complex brain circuit dynamics in vivo. *Neuroscience Research* **169**, 2–8 (2021).

[24] Tour, O., Adams, S. R., Kerr, R. A. *et al.* Calcium Green FlAsH as a genetically targeted small-molecule calcium indicator. *Nature Chemical Biology* **3**, 423–431 (2007).

[25] Kamiya, M. & Johnsson, K. Localizable and Highly Sensitive Calcium Indicator Based on a BODIPY Fluorophore. *Analytical Chemistry* **82**, 6472–6479 (2010).

[26] Best, M., Porth, I., Hauke, S. *et al.* Protein-specific localization of a rhodamine-based calcium-sensor in living cells. *Organic & Biomolecular Chemistry* **14**, 5606–5611 (2016).

[27] Mertes, N., Busch, M., Huppertz, M.-C. *et al.* Fluorescent and Bioluminescent Calcium Indicators with Tuneable Colors and Affinities. *Journal of the American Chemical Society* **144**, 6928–6935 (2022).

[28] Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nature Biotechnology* **19**, 137–141 (2001).

[29] Yang, W. & Yuste, R. In vivo imaging of neural activity. *Nature Methods* **14**, 349–359 (2017).

[30] Zhang, Y., Rózsa, M., Liang, Y. *et al.* Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature* **615**, 884–891 (2023).

[31] Denk, W., Strickler, J. H. & Webb, W. W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science* **248**, 73–76 (1990).

[32] Grienberger, C., Giovannucci, A., Zeiger, W. & Portera-Cailliau, C. Two-photon calcium imaging of neuronal activity. *Nature Reviews Methods Primers* **2**, 67 (2022).

[33] Streich, L., Boffi, J. C., Wang, L. *et al.* High-resolution structural and functional deep brain imaging using adaptive optics three-photon microscopy. *Nature Methods* **18**, 1253–1258 (2021).

[34] Li, Z., Burrone, J., Tyler, W. J. *et al.* Synaptic vesicle recycling studied in transgenic mice expressing synaptopHluorin. *Proceedings of the National Academy of Sciences* **102**, 6131–6136 (2005).

[35] Li, Y. & Tsien, R. W. pHTomato, a red, genetically encoded indicator that enables multiplex interrogation of synaptic activity. *Nature Neuroscience* **15**, 1047–1053 (2012).

[36] Marvin, J. S., Borghuis, B. G., Tian, L. *et al.* An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nature Methods* **10**, 162–170 (2013).

[37] Jing, M., Zhang, P., Wang, G. *et al.* A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. *Nature Biotechnology* **36**, 726–737 (2018).

[38] Patriarchi, T., Cho, J. R., Merten, K. *et al.* Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* **360**, eaat4422 (2018).

[39] Sun, F., Zeng, J., Jing, M. *et al.* A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **174**, 481–496.e19 (2018).

[40] Ji, N., Freeman, J. & Smith, S. L. Technologies for imaging neural activity in large volumes. *Nature Neuroscience* **19**, 1154–1164 (2016).

[41] Franceschini, A., Costantini, I., Pavone, F. S. & Silvestri, L. Dissecting Neuronal Activation on a Brain-Wide Scale With Immediate Early Genes. *Frontiers in Neuroscience* **14**, 569517 (2020).

[42] Gao, R., Asano, S. M., Upadhyayula, S. *et al.* Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* **363**, eaau8302 (2019).

[43] Cai, R., Pan, C., Ghasemigharagoz, A. *et al.* Panoptic imaging of transparent mice reveals whole-body neuronal projections and skull–meninges connections. *Nature Neuroscience* **22**, 317–327 (2019).

[44] Linghu, C., An, B., Shpokayte, M. *et al.* Recording of cellular physiological histories along optically readable self-assembling protein chains. *Nature Biotechnology* **41**, 640–651 (2023).

[45] Sheng, M. & Greenberg, M. E. The regulation and function of c-fos and other immediate early genes in the nervous system. *Neuron* **4**, 477–485 (1990).

[46] Guzowski, J. F., Timlin, J. A., Roysam, B. *et al.* Mapping behaviorally relevant neural circuits with immediate-early gene expression. *Current Opinion in Neurobiology* **15**, 599–606 (2005).

[47] Smeyne, R. J., Schilling, K., Robertson, L. *et al.* Fos-IacZ transgenic mice: Mapping sites of gene induction in the central nervous system. *Neuron* **8**, 13–23 (1992).

[48] Barth, A. L., Gerkin, R. C. & Dean, K. L. Alteration of Neuronal Firing Properties after *In Vivo* Experience in a FosGFP Transgenic Mouse. *The Journal of Neuroscience* **24**, 6466–6475 (2004).

[49] Vousden, D. A., Epp, J., Okuno, H. *et al.* Whole-brain mapping of behaviourally induced neural activation in mice. *Brain Structure and Function* **220**, 2043–2057 (2015).

[50] Kim, Y., Venkataraju, K. U., Pradhan, K. *et al.* Mapping Social Behavior-Induced Brain Activation at Cellular Resolution in the Mouse. *Cell Reports* **10**, 292–305 (2015).

[51] Reijmers, L. G., Perkins, B. L., Matsuo, N. & Mayford, M. Localization of a Stable Neural Correlate of Associative Memory. *Science* **317**, 1230–1233 (2007).

[52] Guenthner, C. J., Miyamichi, K., Yang, H. H. *et al.* Permanent Genetic Access to Transiently Active Neurons via TRAP: Targeted Recombination in Active Populations. *Neuron* **78**, 773–784 (2013).

[53] DeNardo, L. A., Liu, C. D., Allen, W. E. *et al.* Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nature Neuroscience* **22**, 460–469 (2019).

[54] Lin, D., Li, X., Moult, E. *et al.* Time-tagged ticker tapes for intracellular recordings. *Nature Biotechnology* **41**, 631–639 (2023).

[55] Sheng, H. Z., Fields, R. D. & Nelson, P. G. Specific regulation of immediate early genes by patterned neuronal activity. *Journal of Neuroscience Research* **35**, 459–467 (1993).

[56] Fields, R. D., Eshete, F., Stevens, B. & Itoh, K. Action Potential-Dependent Regulation of Gene Expression: Temporal Specificity in Ca $^{2+}$, cAMP-Responsive Element Binding Proteins, and Mitogen-Activated Protein Kinase Signaling. *The Journal of Neuroscience* **17**, 7252–7266 (1997).

[57] Lee, D., Hyun, J. H., Jung, K. *et al.* A calcium- and light-gated switch to induce gene expression in activated neurons. *Nature Biotechnology* **35**, 858–863 (2017).

[58] Wang, W., Wildes, C. P., Pattarabanjird, T. *et al.* A light- and calcium-gated transcription factor for imaging and manipulating activated neurons. *Nature Biotechnology* **35**, 864–871 (2017).

[59] Kim, C. K., Sanchez, M. I., Hoerbelt, P. *et al.* A Molecular Calcium Integrator Reveals a Striatal Cell Type Driving Aversion. *Cell* **183**, 2003–2019.e16 (2020).

[60] Sanchez, M. I., Nguyen, Q.-A., Wang, W. *et al.* Transcriptional readout of neuronal activity via an engineered Ca $^{2+}$-activated protease. *Proceedings of the National Academy of Sciences* **117**, 33186–33196 (2020).

[61] Hyun, J. H., Nagahama, K., Namkung, H. *et al.* Tagging active neurons by soma-targeted Cal-Light. *Nature Communications* **13**, 7692 (2022).

[62] Wang, W., Kim, C. K. & Ting, A. Y. Molecular tools for imaging and recording neuronal activity. *Nature Chemical Biology* **15**, 101–110 (2019).

[63] Fosque, B. F., Sun, Y., Dana, H. *et al.* Labeling of active neural circuits in vivo with designed calcium integrators. *Science* **347**, 755–760 (2015).

[64] Moeyaert, B., Holt, G., Madangopal, R. *et al.* Improved methods for marking active neuron populations. *Nature Communications* **9**, 4440 (2018).

[65] Das, A., Holden, S., Borovicka, J. *et al.* Large-scale recording of neuronal activity in freely-moving mice at cellular resolution. *Nature Communications* **14**, 6399 (2023).

[66] Sha, F., Abdelfattah, A. S., Patel, R. & Schreiter, E. R. Erasable labeling of neuronal activity using a reversible calcium marker. *eLife* **9**, e57249 (2020).

[67] Nava, S. S., An, S. & Hamil, T. Visual detection of UV cues by adult zebrafish (Danio rerio). *Journal of Vision* **11**, 2–2 (2011).

[68] Jacobs, G. H., Neitz, J. & Deegan, J. F. Retinal receptors in rodents maximally sensitive to ultraviolet light. *Nature* **353**, 655–656 (1991).

[69] Lei, B. & Yao, G. Spectral attenuation of the mouse, rat, pig and human lenses from wavelengths 360nm to 1020nm. *Experimental Eye Research* **83**, 610–614 (2006).

[70] Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1998).

[71] Peters, A., Palay, S. L., deF Webster, H. & Palay, S. L. *The Fine Structure of the Nervous System: Neurons and Their Supporting Cells* (Oxford University Press, New York Oxford, 1991), third edition edn.

[72] Denk, W. & Horstmann, H. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biology* **2**, e329 (2004).

[73] Heymann, J. A., Hayles, M., Gestmann, I. *et al.* Site-specific 3D imaging of cells and tissues with a dual beam microscope. *Journal of Structural Biology* **155**, 63–73 (2006).

[74] Bock, D. D., Lee, W.-C. A., Kerlin, A. M. *et al.* Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177–182 (2011).

[75] Kubota, Y., Sohn, J. & Kawaguchi, Y. Large Volume Electron Microscopy and Neural Microcircuit Analysis. *Frontiers in Neural Circuits* **12**, 98 (2018).

[76] Yin, W., Brittain, D., Borseth, J. *et al.* A petascale automated imaging pipeline for mapping neuronal circuits with high-throughput transmission electron microscopy. *Nature Communications* **11**, 4949 (2020).

[77] Dorkenwald, S., Matsliah, A., Sterling, A. R. *et al.* Neuronal wiring diagram of an adult brain. *bioRxiv* (2023).

[78] Sneve, M. A. & Piatkevich, K. D. Towards a Comprehensive Optical Connectome at Single Synapse Resolution via Expansion Microscopy. *Frontiers in Synaptic Neuroscience* **13**, 754814 (2022).

[79] Bergmann, E., Gofman, X., Kavushansky, A. & Kahn, I. Individual variability in functional connectivity architecture of the mouse brain. *Communications Biology* **3**, 738 (2020).

[80] Witvliet, D., Mulcahy, B., Mitchell, J. K. *et al.* Connectomes across development reveal principles of brain maturation. *Nature* **596**, 257–261 (2021).

[81] Xu, T., Yu, X., Perlik, A. J. *et al.* Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462**, 915–919 (2009).

[82] Berry, K. P. & Nedivi, E. Spine Dynamics: Are They All the Same? *Neuron* **96**, 43–55 (2017).

[83] Abbe, E. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie* **9**, 413–468 (1873).

[84] Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Optics Letters* **19**, 780 (1994).

[85] Betzig, E., Patterson, G. H., Sougrat, R. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642–1645 (2006).

[86] Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophysical Journal* **91**, 4258–4272 (2006).

[87] Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–796 (2006).

[88] Sharonov, A. & Hochstrasser, R. M. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proceedings of the National Academy of Sciences* **103**, 18911–18916 (2006).

[89] Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T. *et al.* Super-resolution microscopy with DNA-PAINT. *Nature Protocols* **12**, 1198–1228 (2017).

[90] Velicky, P., Miguel, E., Michalska, J. M. *et al.* Dense 4D nanoscale reconstruction of living brain tissue. *Nature Methods* (2023).

[91] Michalska, J. M., Lyudchik, J., Velicky, P. *et al.* Imaging brain tissue architecture across millimeter to nanometer scales. *Nature Biotechnology* (2023).

[92] Hama, H., Kurokawa, H., Kawano, H. *et al.* Scale: A chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. *Nature Neuroscience* **14**, 1481–1488 (2011).

[93] Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).

[94] Winnubst, J., Bas, E., Ferreira, T. A. *et al.* Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain. *Cell* **179**, 268–281.e13 (2019).

[95] Shen, F. Y., Harrington, M. M., Walker, L. A. *et al.* Light microscopy based approach for mapping connectivity with molecular specificity. *Nature Communications* **11**, 4632 (2020).

[96] Seo, J., Sim, Y., Kim, J. *et al.* PICASSO allows ultra-multiplexed fluorescence imaging of spatially overlapping proteins without reference spectra measurements. *Nature Communications* **13**, 2475 (2022).

[97] Thiel, G. Synapsin I, Synapsin II, and Synaptophysin: Marker Proteins of Synaptic Vesicles. *Brain Pathology* **3**, 87–95 (1993).

[98] Maas, C., Torres, V. I., Altrock, W. D. *et al.* Formation of Golgi-Derived Active Zone Precursor Vesicles. *The Journal of Neuroscience* **32**, 11095–11108 (2012).

[99] Ku, T., Swaney, J., Park, J.-Y. *et al.* Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nature Biotechnology* **34**, 973–981 (2016).

[100] Chang, J.-B., Chen, F., Yoon, Y.-G. *et al.* Iterative expansion microscopy. *Nature Methods* **14**, 593–599 (2017).

[101] El-Husseini, A. E.-D., Schnell, E., Chetkovich, D. M. *et al.* PSD-95 Involvement in Maturation of Excitatory Synapses. *Science* **290**, 1364–1368 (2000).

[102] Luo, P., Li, X., Fei, Z. & Poon, W. Scaffold protein Homer 1: Implications for neurological diseases. *Neurochemistry International* **61**, 731–738 (2012).

[103] Dejanovic, B., Semtner, M., Ebert, S. *et al.* Palmitoylation of Gephyrin Controls Receptor Clustering and Plasticity of GABAergic Synapses. *PLoS Biology* **12**, e1001908 (2014).

[104] Choquet, D., Sainlos, M. & Sibarita, J.-B. Advanced imaging and labelling methods to decipher brain cell organization and function. *Nature Reviews Neuroscience* **22**, 237–255 (2021).

[105] Feinberg, E. H., VanHoven, M. K., Bendesky, A. *et al.* GFP Reconstitution Across Synaptic Partners (GRASP) Defines Cell Contacts and Synapses in Living Nervous Systems. *Neuron* **57**, 353–363 (2008).

[106] Tsetsenis, T., Boucard, A. A., Arac, D. *et al.* Direct Visualization of Trans-Synaptic Neurexin-Neuroligin Interactions during Synapse Formation. *Journal of Neuroscience* **34**, 15083–15096 (2014).

[107] Gordon, M. D. & Scott, K. Motor Control in a Drosophila Taste Circuit. *Neuron* **61**, 373–384 (2009).

[108] Gong, Z., Liu, J., Guo, C. *et al.* Two Pairs of Neurons in the Central Brain Control *Drosophila* Innate Light Preference. *Science* **330**, 499–502 (2010).

[109] Yuan, Q., Xiang, Y., Yan, Z. *et al.* Light-Induced Structural and Functional Plasticity in *Drosophila* Larval Visual System. *Science* **333**, 1458–1462 (2011).

[110] Kim, W. J., Jan, L. Y. & Jan, Y. N. A PDF/NPF Neuropeptide Signaling Circuitry of Male Drosophila melanogaster Controls Rival-Induced Prolonged Mating. *Neuron* **80**, 1190–1205 (2013).

[111] Shao, Z., Watanabe, S., Christensen, R. *et al.* Synapse Location during Growth Depends on Glia Location. *Cell* **154**, 337–350 (2013).

[112] Macpherson, L. J., Zaharieva, E. E., Kearney, P. J. *et al.* Dynamic labelling of neural connections in multiple colours by trans-synaptic fluorescence complementation. *Nature Communications* **6**, 10024 (2015).

[113] Li, Y., Guo, A. & Li, H. CRASP: CFP reconstitution across synaptic partners. *Biochemical and Biophysical Research Communications* **469**, 352–356 (2016).

[114] Feng, S., Varshney, A., Coto Villa, D. *et al.* Bright split red fluorescent proteins for the visualization of endogenous proteins and synapses. *Communications Biology* **2**, 344 (2019).

[115] Shearin, H. K., Quinn, C. D., Mackin, R. D. *et al.* T-GRASP, a targeted GRASP for assessing neuronal connectivity. *Journal of Neuroscience Methods* **306**, 94–102 (2018).

[116] Kim, J., Zhao, T., Petralia, R. S. *et al.* mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nature Methods* **9**, 96–102 (2012).

[117] Cao, X. & Tabuchi, K. Functions of synapse adhesion molecules neurexin/neuroligins and neurodevelopmental disorders. *Neuroscience Research* **116**, 3–9 (2017).

[118] Dresbach, T., Neeb, A., Meyer, G. *et al.* Synaptic targeting of neuroligin is independent of neurexin and SAP90/PSD95 binding. *Molecular and Cellular Neuroscience* **27**, 227–235 (2004).

[119] Ko, J., Zhang, C., Arac, D. *et al.* Neuroligin-1 performs neurexin-dependent and neurexin-independent functions in synapse validation. *The EMBO Journal* **28**, 3244–3255 (2009).

[120] Gokce, O. & Südhof, T. C. Membrane-Tethered Monomeric Neurexin LNS-Domain Triggers Synapse Formation. *The Journal of Neuroscience* **33**, 14617–14628 (2013).

[121] Druckmann, S., Feng, L., Lee, B. *et al.* Structured Synaptic Connectivity between Hippocampal Regions. *Neuron* **81**, 629–640 (2014).

[122] Song, J. H., Lucaci, D., Calangiu, I. *et al.* Combining mGRASP and Optogenetics Enables High-Resolution Functional Mapping of Descending Cortical Projections. *Cell Reports* **24**, 1071–1080 (2018).

[123] Yamagata, M. & Sanes, J. R. Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Frontiers in Molecular Neuroscience* **5** (2012).

[124] Choi, J.-H., Sim, S.-E., Kim, J.-i. *et al.* Interregional synaptic maps among engram cells underlie memory formation. *Science* **360**, 430–435 (2018).

[125] Martell, J. D., Yamagata, M., Deerinck, T. J. *et al.* A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapses. *Nature Biotechnology* **34**, 774–780 (2016).

[126] Liu, D. S., Loh, K. H., Lam, S. S. *et al.* Imaging Trans-Cellular Neurexin-Neuroligin Interactions by Enzymatic Probe Ligation. *PLoS ONE* **8**, e52823 (2013).

[127] Desbois, M., Cook, S. J., Emmons, S. W. & Bülow, H. E. Directional *Trans* -Synaptic Labeling of Specific Neuronal Connections in Live Animals. *Genetics* **200**, 697–705 (2015).

[128] Shekhawat, S. S. & Ghosh, I. Split-protein systems: Beyond binary protein–protein interactions. *Current Opinion in Chemical Biology* **15**, 789–797 (2011).

[129] Romei, M. G. & Boxer, S. G. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annual Review of Biophysics* **48**, 19–44 (2019).

[130] Coons, A. H., Creech, H. J., Jones, R. N. & Berliner, E. The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody. *The Journal of Immunology* **45**, 159–170 (1942).

[131] Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *Journal of Cellular and Comparative Physiology* **59**, 223–239 (1962).

[132] Chalfie, M., Tu, Y., Euskirchen, G. *et al.* Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263**, 802–805 (1994).

[133] Rodriguez, E. A., Campbell, R. E., Lin, J. Y. *et al.* The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends in Biochemical Sciences* **42**, 111–129 (2017).

[134] Greenwald, E. C., Mehta, S. & Zhang, J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chemical Reviews* **118**, 11707–11794 (2018).

[135] Grimm, J. B. & Lavis, L. D. Caveat fluorophore: An insiders' guide to small-molecule fluorescent labels. *Nature Methods* **19**, 149–158 (2022).

[136] Wang, L., Frei, M. S., Salim, A. & Johnsson, K. Small-Molecule Fluorescent Probes for Live-Cell Super-Resolution Microscopy. *Journal of the American Chemical Society* **141**, 2770–2781 (2019).

[137] Hinner, M. J. & Johnsson, K. How to obtain labeled proteins and what to do with them. *Current Opinion in Biotechnology* **21**, 766–776 (2010).

[138] Keppler, A., Gendreizig, S., Gronemeyer, T. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology* **21**, 86–89 (2003).

[139] Los, G. V., Encell, L. P., McDougall, M. G. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology* **3**, 373–382 (2008).

[140] Pegg, A. E., Dolan, M. E. & Moschel, R. C. Structure, Function, and Inhibition of O6-Alkylguanine-DNA Alkyltransferase. In *Progress in Nucleic Acid Research and Molecular Biology*, vol. 51, 167–223 (Elsevier, 1995).

[141] Correa, I., Baker, B., Zhang, A. *et al.* Substrates for Improved Live-Cell Fluorescence Labeling of SNAP-tag. *Current Pharmaceutical Design* **19**, 5414–5420 (2013).

[142] Gautier, A., Juillerat, A., Heinis, C. *et al.* An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry & Biology* **15**, 128–136 (2008).

[143] Ohana, R. F., Encell, L. P., Zhao, K. *et al.* HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expression and Purification* **68**, 110–120 (2009).

[144] Damborský, J. & Koča, J. Analysis of the reaction mechanism and substrate specificity of haloalkane dehalogenases by sequential and structural comparisons. *Protein Engineering, Design and Selection* **12**, 989–998 (1999).

[145] Lukinavičius, G., Umezawa, K., Olivier, N. *et al.* A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nature Chemistry* **5**, 132–139 (2013).

[146] Butkevich, A. N., Mitronova, G. Y., Sidenstein, S. C. *et al.* Fluorescent Rhodamines and Fluorogenic Carbopyronines for Super-Resolution STED Microscopy in Living Cells. *Angewandte Chemie International Edition* **55**, 3290–3294 (2016).

[147] Grimm, J. B., Muthusamy, A. K., Liang, Y. *et al.* A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nature Methods* **14**, 987–994 (2017).

[148] Wang, L., Tran, M., D'Este, E. *et al.* A general strategy to develop cell permeable and fluorogenic probes for multicolour nanoscopy. *Nature Chemistry* **12**, 165–172 (2020).

[149] Grimm, J. B., Tkachuk, A. N., Xie, L. *et al.* A general method to optimize and functionalize red-shifted rhodamine dyes. *Nature Methods* **17**, 815–821 (2020).

[150] Lardon, N., Wang, L., Tschanz, A. *et al.* Systematic Tuning of Rhodamine Spirocyclization for Super-resolution Microscopy. *Journal of the American Chemical Society* **143**, 14592–14600 (2021).

[151] Grimm, J. B., Tkachuk, A. N., Patel, R. *et al.* Optimized Red-Absorbing Dyes for Imaging and Sensing. *Journal of the American Chemical Society* (2023).

[152] Bulovaite, E., Qiu, Z., Kratschke, M. *et al.* A brain atlas of synapse protein lifetime across the mouse lifespan. *Neuron* S0896627322008145 (2022).

[153] So, M.-k., Yao, H. & Rao, J. HaloTag protein-mediated specific labeling of living cells with quantum dots. *Biochemical and Biophysical Research Communications* **374**, 419–423 (2008).

[154] Ollech, D., Pflästerer, T., Shellard, A. *et al.* An optochemical tool for light-induced dissociation of adherens junctions to control mechanical coupling between cells. *Nature Communications* **11**, 472 (2020).

[155] Broichhagen, J., Damijonaitis, A., Levitz, J. *et al.* Orthogonal Optical Control of a G Protein-Coupled Receptor with a SNAP-Tethered Photochromic Ligand. *ACS Central Science* **1**, 383–393 (2015).

[156] England, C. G., Luo, H. & Cai, W. HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjugate Chemistry* **26**, 975–986 (2015).

[157] Haruki, H., Gonzalez, M. R. & Johnsson, K. Exploiting Ligand-Protein Conjugates to Monitor Ligand-Receptor Interactions. *PLoS ONE* **7**, e37598 (2012).

[158] Buckley, D. L., Raina, K., Darricarrere, N. *et al.* HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chemical Biology* **10**, 1831–1837 (2015).

[159] Mohar, B., Grimm, J. B., Patel, R. *et al.* Brain-wide measurement of protein turnover with high spatial and temporal resolution. *bioRxiv* (2022).

[160] Leippe, P., Koehler Leman, J. & Trauner, D. Specificity and Speed: Tethered Photopharmacology. *Biochemistry* **56**, 5214–5220 (2017).

[161] Sallin, O., Reymond, L., Gondrand, C. *et al.* Semisynthetic biosensors for mapping cellular concentrations of nicotinamide adenine dinucleotides. *eLife* **7**, e32638 (2018).

[162] Yu, Q., Xue, L., Hiblot, J. *et al.* Semisynthetic sensor proteins enable metabolic assays at the point of care. *Science* **361**, 1122–1126 (2018).

[163] Hellweg, L., Edenhofer, A., Barck, L. *et al.* A general method for the development of multicolor biosensors with large dynamic ranges. *Nature Chemical Biology* **19**, 1147–1157 (2023).

[164] Nooren, I. M. Diversity of protein-protein interactions. *The EMBO Journal* **22**, 3486–3492 (2003).

[165] Keskin, O., Tuncbag, N. & Gursoy, A. Predicting Protein–Protein Interactions from the Molecular to the Proteome Level. *Chemical Reviews* **116**, 4884–4909 (2016).

[166] Michnick, S. W., Ear, P. H., Manderson, E. N. *et al.* Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nature Reviews Drug Discovery* **6**, 569–582 (2007).

[167] Morell, M., Ventura, S. & Avilés, F. X. Protein complementation assays: Approaches for the in vivo analysis of protein interactions. *FEBS Letters* **583**, 1684–1691 (2009).

[168] Blaszczak, E., Lazarewicz, N., Sudevan, A. *et al.* Protein-fragment complementation assays for large-scale analysis of protein–protein interactions. *Biochemical Society Transactions* **49**, 1337–1348 (2021).

[169] Fields, S. & Song, O.-k. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).

[170] Uetz, P., Giot, L., Cagney, G. *et al.* A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627 (2000).

[171] Formstecher, E., Aresta, S., Collura, V. *et al.* Protein interaction mapping: A *Drosophila* case study. *Genome Research* **15**, 376–384 (2005).

[172] Luck, K., Kim, D.-K., Lambourne, L. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).

[173] Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proceedings of the National Academy of Sciences* **91**, 10340–10344 (1994).

[174] Pelletier, J. N., Campbell-Valois, F.-X. & Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proceedings of the National Academy of Sciences* **95**, 12141–12146 (1998).

[175] Tafelmeyer, P., Johnsson, N. & Johnsson, K. Transforming a (β/α)8-Barrel Enzyme into a Split-Protein Sensor through Directed Evolution. *Chemistry & Biology* **11**, 681–689 (2004).

[176] Tarassov, K., Messier, V., Landry, C. R. *et al.* An in Vivo Map of the Yeast Protein Interactome. *Science* **320**, 1465–1470 (2008).

[177] Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *Journal of the American Chemical Society* **122**, 5658–5659 (2000).

[178] Tebo, A. G. & Gautier, A. A split fluorescent reporter with rapid and reversible complementation. *Nature Communications* **10**, 2822 (2019).

[179] Paulmurugan, R. & Gambhir, S. S. Monitoring Protein-Protein Interactions Using Split Synthetic Renilla Luciferase Protein-Fragment-Assisted Complementation. *Analytical Chemistry* **75**, 1584–1589 (2003).

[180] Luker, K. E., Smith, M. C. P., Luker, G. D. *et al.* Kinetics of regulated protein–protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proceedings of the National Academy of Sciences* **101**, 12288–12293 (2004).

[181] Remy, I. & Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nature Methods* **3**, 977–979 (2006).

[182] Dixon, A. S., Schwinn, M. K., Hall, M. P. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chemical Biology* **11**, 400–408 (2016).

[183] Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[184] Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Current Opinion in Structural Biology* **14**, 70–75 (2004).

[185] Kerppola, T. K. Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementation. *Chemical Society Reviews* **38**, 2876 (2009).

[186] Ohmuro-Matsuyama, Y., Chung, C.-I. & Ueda, H. Demonstration of protein-fragment complementation assay using purified firefly luciferase fragments. *BMC Biotechnology* **13**, 31 (2013).

[187] Blaszczak, E., Prigent, C. & Rabut, G. Bimolecular Fluorescence Complementation to Assay the Interactions of Ubiquitylation Enzymes in Living Yeast Cells. In Matthiesen, R. (ed.) *Proteostasis*, vol. 1449, 223–241 (Springer New York, New York, NY, 2016).

[188] Wilhelm, J., Kühn, S., Tarnawski, M. *et al.* Kinetic and Structural Characterization of the Self-Labeling Protein Tags HaloTag7, SNAP-tag, and CLIP-tag. *Biochemistry* **60**, 2560–2575 (2021).

[189] Hoelzel, C. A. & Zhang, X. Visualizing and Manipulating Biological Processes by Using HaloTag and SNAP-Tag Technologies. *ChemBioChem* **21**, 1935–1946 (2020).

[190] Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics* (Elsevier Science, 2014).

[191] Malcolm, A. D. B. & Radda, G. K. The Reaction of Glutamate Dehydrogenase with 4-Iodoacetamido Salicylic Acid. *European Journal of Biochemistry* **15**, 555–561 (1970).

[192] Chang, A., Jeske, L., Ulbrich, S. *et al.* BRENDA, the ELIXIR core data resource in 2021: New developments and updates. *Nucleic Acids Research* **49**, D498–D508 (2021).

[193] Bar-Even, A., Noor, E., Savir, Y. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).

[194] McWhirter, C. Kinetic mechanisms of covalent inhibition. In *Annual Reports in Medicinal Chemistry*, vol. 56, 1–31 (Elsevier, 2021).

[195] Kuo-chen, C. & Shou-ping, J. Studies on the rate of diffusion-controlled reactions of enzymes. Spatial factor and force field factor. *Scientia Sinica* **27**, 664–680 (1974).

[196] Schurr, J. M. & Schmitz, K. S. Orientation constraints and rotational diffusion in bimolecular solution kinetics. A simplification. *The Journal of Physical Chemistry* **80**, 1934–1936 (1976).

[197] Samson, R. & Deutch, J. M. Diffusion-controlled reaction rate to a buried active site. *The Journal of Chemical Physics* **68**, 285–290 (1978).

[198] Lang, K. & Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chemical Biology* **9**, 16–20 (2014).

[199] Koßmann, K. J., Ziegler, C., Angelin, A. *et al.* A Rationally Designed Connector for Assembly of Protein-Functionalized DNA Nanostructures. *ChemBioChem* **17**, 1102–1106 (2016).

[200] Kompa, J., Bruins, J., Glogger, M. *et al.* Exchangeable HaloTag Ligands for Super-Resolution Fluorescence Microscopy. *Journal of the American Chemical Society* **145**, 3075–3083 (2023).

[201] Kumar, P., Vevea, J. D., Chapman, E. R. & Lavis, L. D. Multifunctional fluorophores for live-cell imaging and affinity capture of proteins. *bioRxiv* (2022).

[202] Jurrus, E., Engel, D., Star, K. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27**, 112–128 (2018).

[203] Wade, R. C., Gabdoulline, R. R., Lüdemann, S. K. & Lounnas, V. Electrostatic steering and ionic tethering in enzyme–ligand binding: Insights from simulations. *Proceedings of the National Academy of Sciences* **95**, 5942–5949 (1998).

[204] Jencks, W. P. On the attribution and additivity of binding energies. *Proceedings of the National Academy of Sciences* **78**, 4046–4050 (1981).

[205] Grimm, J. B. & Lavis, L. D. Synthesis of Rhodamines from Fluoresceins Using Pd-Catalyzed C–N Cross-Coupling. *Organic Letters* **13**, 6354–6357 (2011).

[206] Daniels, D. L., Méndez, J., Benink, H. *et al.* Discovering Protein Interactions and Characterizing Protein Function Using HaloTag Technology. *Journal of Visualized Experiments* 51553 (2014).

[207] Shields, B. C., Kahuno, E., Kim, C. *et al.* Deconstructing behavioral neuropharmacology with cellular specificity. *Science* **356**, eaaj2161 (2017).

[208] Erdmann, R. S., Baguley, S. W., Richens, J. H. *et al.* Labeling Strategies Matter for Super-Resolution Microscopy: A Comparison between HaloTags and SNAP-tags. *Cell Chemical Biology* **26**, 584–592.e6 (2019).

[209] Shields, B. C., Yan, H., Lim, S. S. *et al.* Thousandfold Cell-Specific Pharmacology of Neurotransmission. *bioRxiv* (2022).

[210] Kannan, M., Vasan, G., Haziza, S. *et al.* Dual-polarity voltage imaging of the concurrent dynamics of multiple neuron types. *Science* **378**, eabm8797 (2022).

[211] Murray, E., Cho, J. H., Goodwin, D. *et al.* Simple, Scalable Proteomic Imaging for High-Dimensional Profiling of Intact Systems. *Cell* **163**, 1500–1514 (2015).

[212] Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nature Methods* **19**, 534–546 (2022).

[213] Kishi, J. Y., Liu, N., West, E. R. *et al.* Light-Seq: Light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nature Methods* **19**, 1393–1402 (2022).

[214] Matthiesen, R. (ed.) *Proteostasis: Methods and Protocols*, vol. 1449 of *Methods in Molecular Biology* (Springer New York, New York, NY, 2016).

[215] Beck, M., Schmidt, A., Malmstroem, J. *et al.* The quantitative proteome of a human cell line. *Molecular Systems Biology* **7**, 549 (2011).

[216] Babu, Y. S., Sack, J. S., Greenhough, T. J. *et al.* Three-dimensional structure of calmodulin. *Nature* **315**, 37–40 (1985).

[217] Ikura, M., Clore, G. M., Gronenborn, A. M. *et al.* Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638 (1992).

[218] Ross, W. N. Understanding calcium waves and sparks in central neurons. *Nature Reviews Neuroscience* **13**, 157–168 (2012).

[219] Huang, L., Ledochowitsch, P., Knoblich, U. *et al.* Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife* **10**, e51675 (2021).

[220] Ruffin, V. A., Salameh, A. I., Boron, W. F. & Parker, M. D. Intracellular pH regulation by acid-base transporters in mammalian neurons. *Frontiers in Physiology* **5** (2014).

[221] Shkryl, V. M. The spatio-temporal properties of calcium transients in hippocampal pyramidal neurons in vitro. *Frontiers in Cellular Neuroscience* **16**, 1054950 (2022).

[222] Larkum, M. E., Watanabe, S., Nakamura, T. *et al.* Synaptically Activated Ca $^{2+}$ Waves in Layer 2/3 and Layer 5 Rat Neocortical Pyramidal Neurons. *The Journal of Physiology* **549**, 471–488 (2003).

[223] Miller, L. D., Petrozzino, J. J., Golarai, G. & Connor, J. A. Ca2+ release from intracellular stores induced by afferent stimulation of CA3 pyramidal neurons in hippocampal slices. *Journal of Neurophysiology* **76**, 554–562 (1996).

[224] Valeyev, N. V., Bates, D. G., Heslop-Harrison, P. *et al.* Elucidating the mechanisms of cooperative calcium-calmodulin interactions: A structural systems biology approach. *BMC Systems Biology* **2**, 48 (2008).

[225] Pepke, S., Kinzer-Ursem, T., Mihalas, S. & Kennedy, M. B. A Dynamic Model of Interactions of Ca2+, Calmodulin, and Catalytic Subunits of Ca2+/Calmodulin-Dependent Protein Kinase II. *PLoS Computational Biology* **6**, e1000675 (2010).

[226] Maisak, M. S., Haag, J., Ammer, G. *et al.* A directional tuning map of Drosophila elementary motion detectors. *Nature* **500**, 212–216 (2013).

[227] Fischbach, K.-F. & Dittrich, A. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. *Cell and Tissue Research* **258** (1989).

[228] Wu, Y., Dal Maschio, M., Kubo, F. & Baier, H. An Optical Illusion Pinpoints an Essential Circuit Node for Global Motion Processing. *Neuron* **108**, 722–734.e5 (2020).

[229] Robles, E., Laurell, E. & Baier, H. The Retinal Projectome Reveals Brain-Area-Specific Visual Representations Generated by Ganglion Cell Diversity. *Current Biology* **24**, 2085–2096 (2014).

[230] Semmelhack, J. L., Donovan, J. C., Thiele, T. R. *et al.* A dedicated visual pathway for prey detection in larval zebrafish. *eLife* **3**, e04878 (2014).

[231] Mearns, D. S., Donovan, J. C., Fernandes, A. M. *et al.* Deconstructing Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop. *Current Biology* **30**, 54–69.e9 (2020).

[232] Antinucci, P., Folgueira, M. & Bianco, I. H. Pretectal neurons control hunting behaviour. *eLife* **8**, e48114 (2019).

[233] Osswald, M., Jung, E., Sahm, F. *et al.* Brain tumour cells interconnect to a functional and resistant network. *Nature* **528**, 93–98 (2015).

[234] Hausmann, D., Hoffmann, D. C., Venkataramani, V. *et al.* Autonomous rhythmic activity in glioma networks drives brain tumour growth. *Nature* **613**, 179–186 (2023).

[235] Pittaras, E., Hamelin, H. & Granon, S. Inter-Individual Differences in Cognitive Tasks: Focusing on the Shaping of Decision-Making Strategies. *Frontiers in Behavioral Neuroscience* **16**, 818746 (2022).

[236] Tsuji, A. & Tamai, I. Blood-brain barrier function of P-glycoprotein. *Advanced Drug Delivery Reviews* **25**, 287–298 (1997).

[237] Park, J., Kim, H., Alabdalla, L. *et al.* Generation and characterization of a zebrafish knockout model of abcb4, a homolog of the human multidrug efflux transporter P-glycoprotein. *Human Genomics* **17**, 84 (2023).

[238] Petersen, C. C. The Functional Organization of the Barrel Cortex. *Neuron* **56**, 339–355 (2007).

[239] White, J., Southgate, E., Thomsom, J. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **314**, 1–340 (1986).

[240] Helmstaedter, M., Briggman, K. L., Turaga, S. C. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).

[241] Takemura, S.-y., Bharioke, A., Lu, Z. *et al.* A visual motion detection circuit suggested by Drosophila connectomics. *Nature* **500**, 175–181 (2013).

[242] Rah, J.-C. & Choi, J. H. Finding Needles in a Haystack with Light: Resolving the Microcircuitry of the Brain with Fluorescence Microscopy. *Molecules and Cells* **45**, 84–92 (2022).

[243] Ichtchenko, K., Hata, Y., Nguyen, T. *et al.* Neuroligin 1: A splice site-specific ligand for β-neurexins. *Cell* **81**, 435–443 (1995).

[244] Scheiffele, P., Fan, J., Choih, J. *et al.* Neuroligin Expressed in Nonneuronal Cells Triggers Presynaptic Development in Contacting Axons. *Cell* **101**, 657–669 (2000).

[245] Missler, M., Sudhof, T. C. & Biederer, T. Synaptic Cell Adhesion. *Cold Spring Harbor Perspectives in Biology* **4**, a005694–a005694 (2012).

[246] Araç, D., Boucard, A. A., Özkan, E. *et al.* Structures of Neuroligin-1 and the Neuroligin-1/Neurexin-1β Complex Reveal Specific Protein-Protein and Protein-Ca2+ Interactions. *Neuron* **56**, 992–1003 (2007).

[247] Fabrichny, I. P., Leone, P., Sulzenbacher, G. *et al.* Structural Analysis of the Synaptic Protein Neuroligin and Its β-Neurexin Complex: Determinants for Folding and Cell Adhesion. *Neuron* **56**, 979–991 (2007).

[248] Chen, X., Liu, H., Shim, A. H. R. *et al.* Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. *Nature Structural & Molecular Biology* **15**, 50–56 (2008).

[249] Lee, H., Oh, W. C., Seong, J. & Kim, J. Advanced Fluorescence Protein-Based Synapse-Detectors. *Frontiers in Synaptic Neuroscience* **8** (2016).

[250] Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An Enhanced Monomeric Blue Fluorescent Protein with the High Chemical Stability of the Chromophore. *PLoS ONE* **6**, e28674 (2011).

[251] Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).

[252] Ryan, M. D., King, A. M. Q. & Thomas, G. P. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology* **72**, 2727–2732 (1991).

[253] Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

[254] Varadi, M., Anyango, S., Deshpande, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).

[255] Koehnke, J., Katsamba, P. S., Ahlsen, G. *et al.* Splice Form Dependence of β-Neurexin/Neuroligin Binding Interactions. *Neuron* **67**, 61–74 (2010).

[256] Zhou, H.-X. The Affinity-Enhancing Roles of Flexible Linkers in Two-Domain DNA-Binding Proteins. *Biochemistry* **40**, 15069–15073 (2001).

[257] Kjaergaard, M., Glavina, J. & Chemes, L. B. Predicting the effect of disordered linkers on effective concentrations and avidity with the "C calculator" app. In *Methods in Enzymology*, vol. 647, 145–171 (Elsevier, 2021).

[258] Choi, J., Chen, J., Schreiber, S. L. & Clardy, J. Structure of the FKBP12-Rapamycin Complex Interacting with Binding Domain of Human FRAP. *Science* **273**, 239–242 (1996).

[259] Leaver-Fay, A., Tyka, M., Lewis, S. M. *et al.* Rosetta3: 3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*, vol. 487, 545–574 (Elsevier, 2011).

[260] Fleishman, S. J., Leaver-Fay, A., Corn, J. E. *et al.* RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS ONE* **6**, e20161 (2011).

[261] Hiblot, J., Huppertz, M.-C., Johnsson, K. & Wilhelm, J. Circularly Permutated Haloalkane Transferase Fusion Molecules (Patent US20220275350A1).

[262] Goldenzweig, A., Goldsmith, M., Hill, S. E. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **63**, 337–346 (2016).

[263] Huang, P.-S., Ban, Y.-E. A., Richter, F. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **6**, e24109 (2011).

[264] Basarsky, T., Parpura, V. & Haydon, P. Hippocampal synaptogenesis in cell culture: Developmental time course of synapse formation, calcium influx, and synaptic protein distribution. *The Journal of Neuroscience* **14**, 6402–6411 (1994).

[265] Holler, S., Köstinger, G., Martin, K. A. C. *et al.* Structure and function of a neocortical synapse. *Nature* **591**, 111–116 (2021).

[266] Schröder, H., Moser, N. & Huggenberger, S. *Neuroanatomy of the Mouse: An Introduction* (Springer International Publishing, Cham, 2020).

[267] Miquelajauregui, A., Kribakaran, S., Mostany, R. *et al.* Layer 4 Pyramidal Neurons Exhibit Robust Dendritic Spine Plasticity *In Vivo* after Input Deprivation. *The Journal of Neuroscience* **35**, 7287–7294 (2015).

[268] Hiblot, J., Yu, Q., Sabbadini, M. D. *et al.* Luciferases with Tunable Emission Wavelengths. *Angewandte Chemie International Edition* **56**, 14556–14560 (2017).

[269] Mudd, G., Pi, I. P., Fethers, N. *et al.* A general synthetic route to isomerically pure functionalized rhodamine dyes. *Methods and Applications in Fluorescence* **3**, 045002 (2015).

[270] Grimm, J. B., English, B. P., Chen, J. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods* **12**, 244–250 (2015).

[271] Grimm, J. B., Brown, T. A., Tkachuk, A. N. & Lavis, L. D. General Synthetic Method for Si-Fluoresceins and Si-Rhodamines. *ACS Central Science* **3**, 975–985 (2017).

[272] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8 (2015).

[273] Schindelin, J., Arganda-Carreras, I., Frise, E. *et al.* Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).

[274] Kuzmič, P. Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. *Analytical Biochemistry* **237**, 260–273 (1996).

[275] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2021).

[276] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

[277] Wickham, H., François, R., Henry, L. *et al.* *dplyr: A Grammar of Data Manipulation* (2023). R package version 1.1.2.

[278] Wickham, H., Vaughan, D. & Girlich, M. *tidyr: Tidy Messy Data* (2023). R package version 1.3.0.

[279] Wickham, H., Hester, J. & Bryan, J. *readr: Read Rectangular Text Data* (2023). R package version 2.1.4.

[280] Wickham, H. & Henry, L. *purrr: Functional Programming Tools* (2023). R package version 1.0.1.

[281] Vaughan, D. & Dancho, M. *furrr: Apply Mapping Functions in Parallel using Futures* (2022). R package version 0.3.1.

[282] Müller, K. & Wickham, H. *tibble: Simple Data Frames* (2023). R package version 3.2.1.

[283] Wickham, H. *stringr: Simple, Consistent Wrappers for Common String Operations* (2022). R package version 1.5.0.

[284] Wickham, H. *forcats: Tools for Working with Categorical Variables (Factors)* (2023). R package version 1.0.0.

[285] de Lacalle, J. L. *tsoutliers: Detection of Outliers in Time Series* (2019). R package version 0.6-8.

[286] Elzhov, T. V., Mullen, K. M., Spiess, A.-N. & Bolker, B. *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm* (2023). R package version 1.2-3.

[287] B.J., G., A.P.C., R., K.M., E. *et al.* Bio3d: An r package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).

[288] Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. *Biostrings: Efficient manipulation of biological strings* (2021). R package version 2.62.0.

[289] Garnier, Simon, Ross *et al.* *viridis(Lite) - Colorblind-Friendly Color Maps for R* (2023). Viridis package version 0.6.3.

[290] Wilke, C. O. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'* (2020). R package version 1.1.1.

[291] Wagih, O. *ggseqlogo: A 'ggplot2' Extension for Drawing Publication-Ready Sequence Logos* (2017). R package version 0.1.

[292] Chen, R. F. Corrections in the Measurement of Fluorescence Polarization. *Analytical Letters* **4**, 459–469 (1971).

[293] Gibson, D. G., Young, L., Chuang, R.-Y. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).

[294] Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols* **6**, 365–387 (2011).

[295] Straume, M. & Johnson, M. L. Monte Carlo Method for determining complete confidence probability distributions of estimated model parameters. In *Methods in Enzymology*, vol. 210, 117–129 (Elsevier, 1992).

List of Publications

Publications in Peer-Reviewed Journals

Baalmann, M., Ziegler, M. J., Werther, P., Wilhelm, J. & Wombacher, R. “**Enzymatic and Site-Specific Ligation of Minimal-Size Tetrazines and Triazines to Proteins for Bioconjugation and Live-Cell Imaging.**” *Bioconjugate Chem.* **30**, 1405–1414 (2019).

Baalmann, M., Neises, L., Bitsch, S., Schneider, H., Deweid, L., Werther, P., Ilkenhans, N., Wolfring, M., Ziegler, M. J., Wilhelm, J., Kolmar, H. & Wombacher, R. “**A Bioorthogonal Click Chemistry Toolbox for Targeted Synthesis of Branched and Well-Defined Protein-Protein Conjugates.**” *Angew. Chem. Int. Ed.* **59**, 12885–12893 (2020).

Wilhelm, J.*, Kühn, S.*, Tarnawski, M., Gotthard, G., Tünnermann, J., Tänzer, T., Karpenko, J., Mertes, N., Xue, L., Uhrig, U., Reinstein, J., Hiblot, J. & Johnsson, K. “**Kinetic and Structural Characterization of the Self-Labeling Protein Tags HaloTag7, SNAP-tag, and CLIP-tag.**” *Biochemistry* **60**, 2560–2575 (2021).

Mertes, N., Busch, M., Huppertz, M.-C., Hacker, C. N., Wilhelm, J., Gürth, C.-M., Kühn, S., Hiblot, J., Koch, B. & Johnsson, K. “**Fluorescent and Bioluminescent Calcium Indicators with Tuneable Colors and Affinities.**” *J. Am. Chem. Soc.* **144**, 6928–6935 (2022).

Kompa, J., Bruins, J., Glogger, M., Wilhelm, J., Frei, M. S., Tarnawski, M., D’Este, E., Heilemann, M., Hiblot, J. & Johnsson, K. “**Exchangeable HaloTag Ligands for Super-Resolution Fluorescence Microscopy.**” *J. Am. Chem. Soc.* **145**, 3075—083 (2023).

Huppertz, M.-C.*, Wilhelm, J.*, Grenier, V., Schneider, M. W., Falt, T., Porzberg, N., Hausmann, D., Hoffmann, D. C., Hai, L., Tarnawski, M., Pino, G., Slanchev, K., Kolb, I., Acuna, C., Fenk, L. M., Baier, H., Hiblot, J. & Johnsson, K. “**Recording Physiological History of Cells with Chemical Labeling.**” *In revision at Science.*

* Authors contributed equally.

Submitted Patents

Hiblot, J., Wilhelm, J., Huppertz, M.-C. & Johnsson, K. “**Circularly Permutated Haloalkane Transferase Fusion Molecules.**” *WO2020212537A1* (2020)

Wilhelm, J., Lin, Y.-H., Hinnah, K., Nickel, L., Hiblot, J. & Johnsson, K. “**Improved Split HaloTags.**” *EP23158085.3* (2023)

Hinnah, K., Wilhelm, J. & Johnsson, K. “**Spatial Proximity Assay.**” *pending* (2023)

Appendix

Supplementary Figures

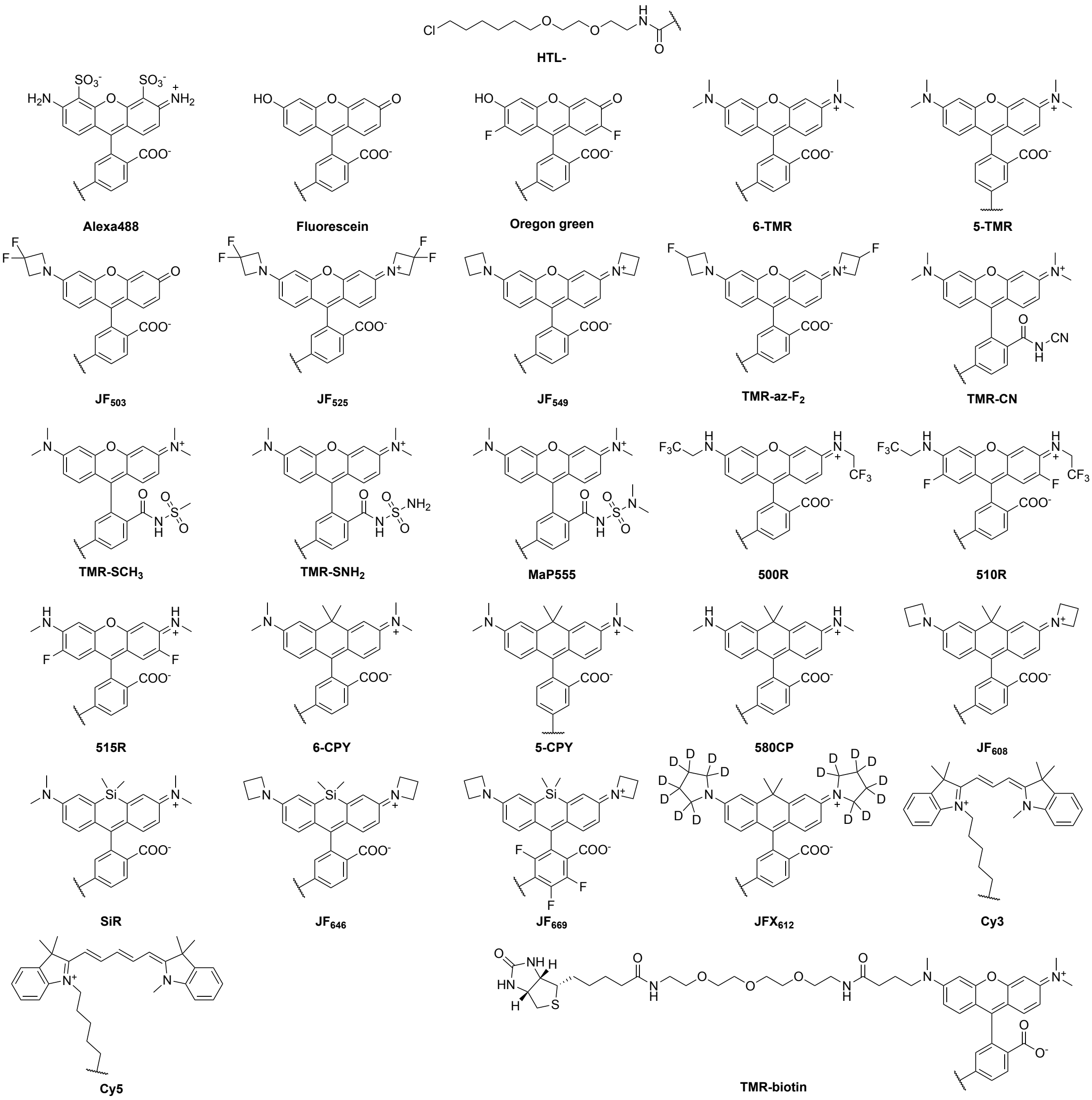


Figure S1: Chemical structures of fluorescent HaloTag ligands used in this thesis.

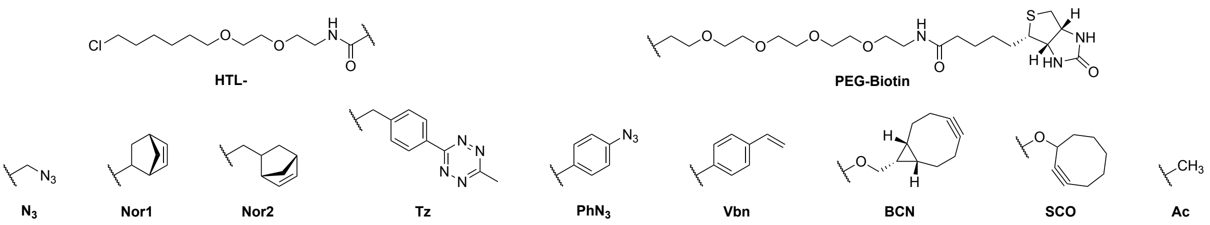


Figure S2: Chemical structures of non-fluorescent HaloTag ligands used in this thesis.

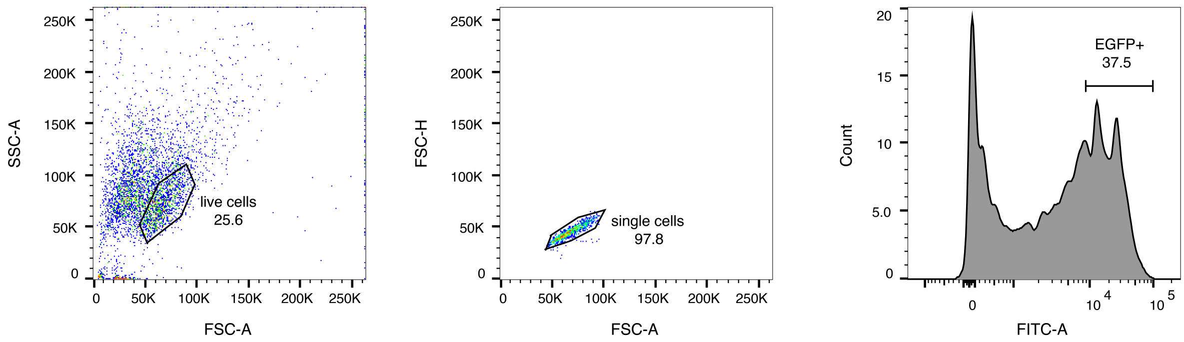


Figure S3: Flow cytometry gating strategy. Events were gated hierarchically for live cells (FSC-A vs. SSC-A), single cells (FSC-A vs. FSC-H) and EGFP positive cells (FITC-A).

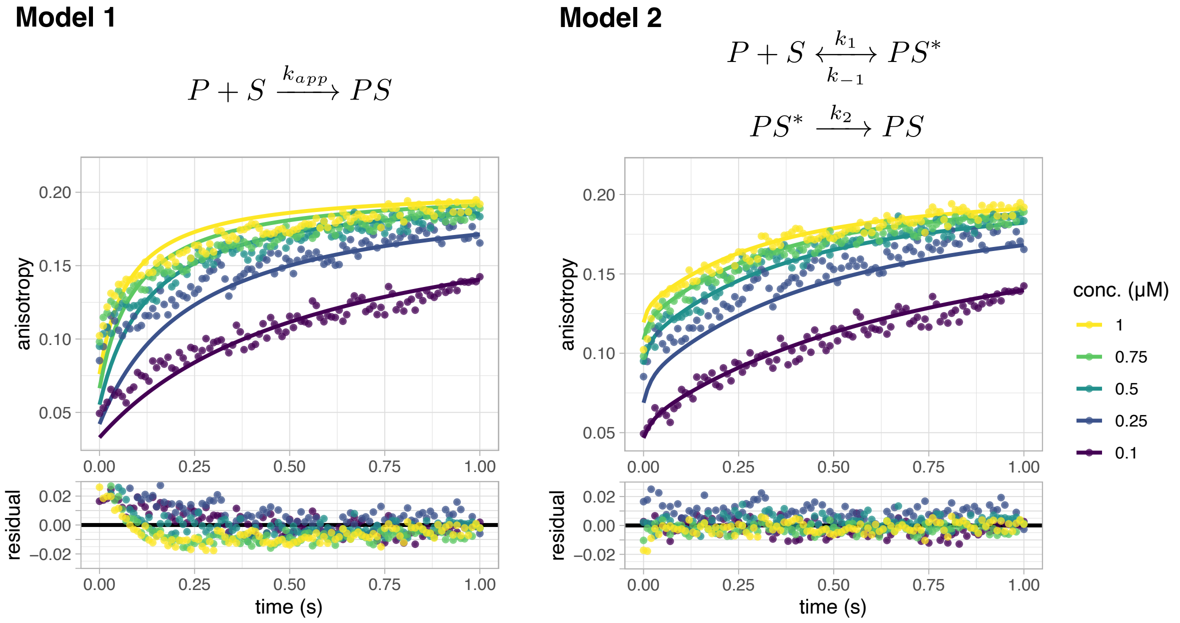


Figure S4: Comparison of reaction models for HaloTag labeling kinetics. Stopped-flow fluorescence anisotropy traces of HaloTag labeling with HTL-TMR and predictions based on the one-step kinetic model 1 or the two-step kinetic model 2. Residuals of the respective fits are depicted in the bottom panels. HaloTag protein and HTL-TMR substrate were mixed in 1:1 stoichiometry at different concentrations (conc.) in a stopped-flow device and fluorescence anisotropy values were recorded. While model 2 fits well to the data, model 1 does not adequately represent the biphasic trend of the traces. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

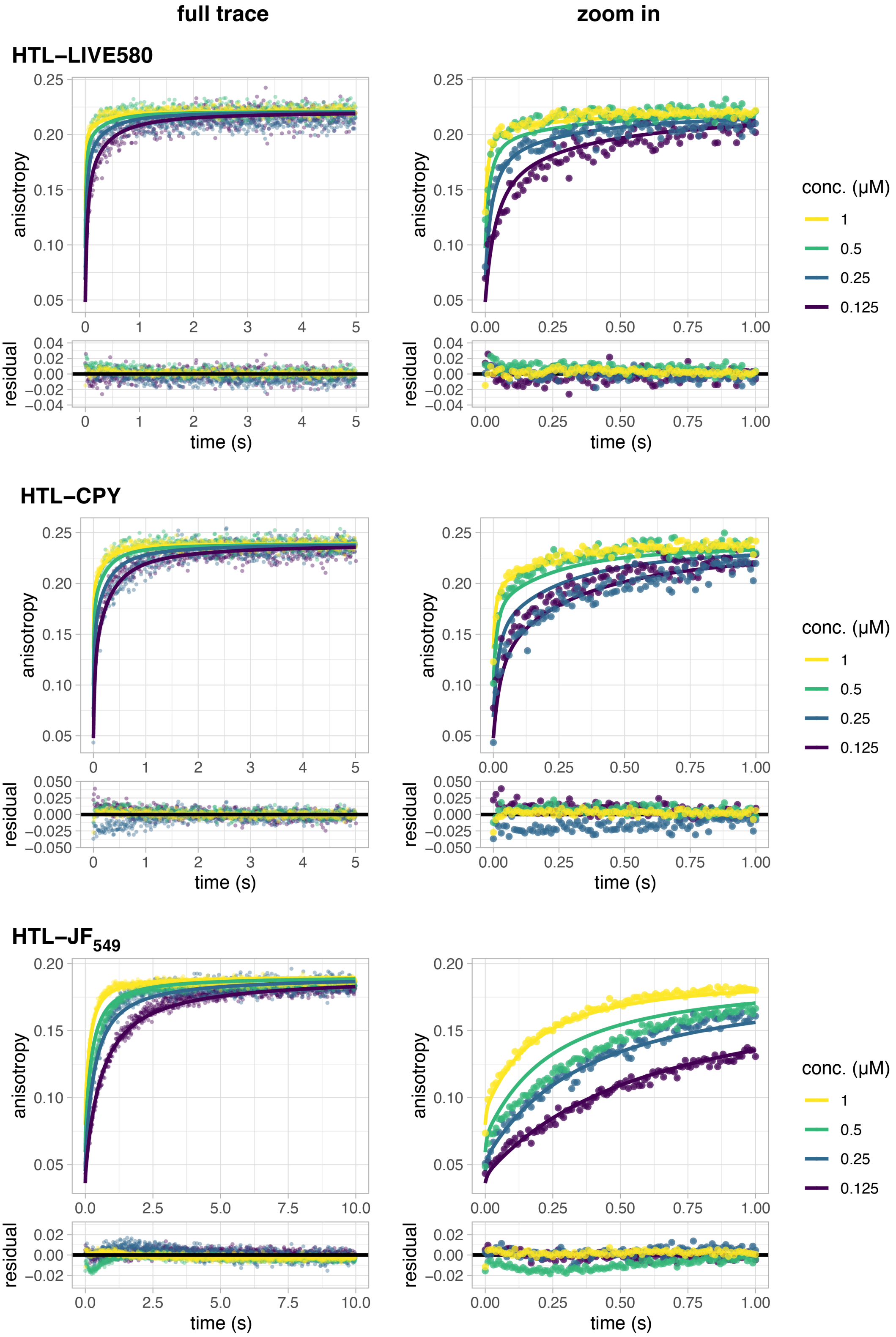


Figure S5: Figure continued on the next page...

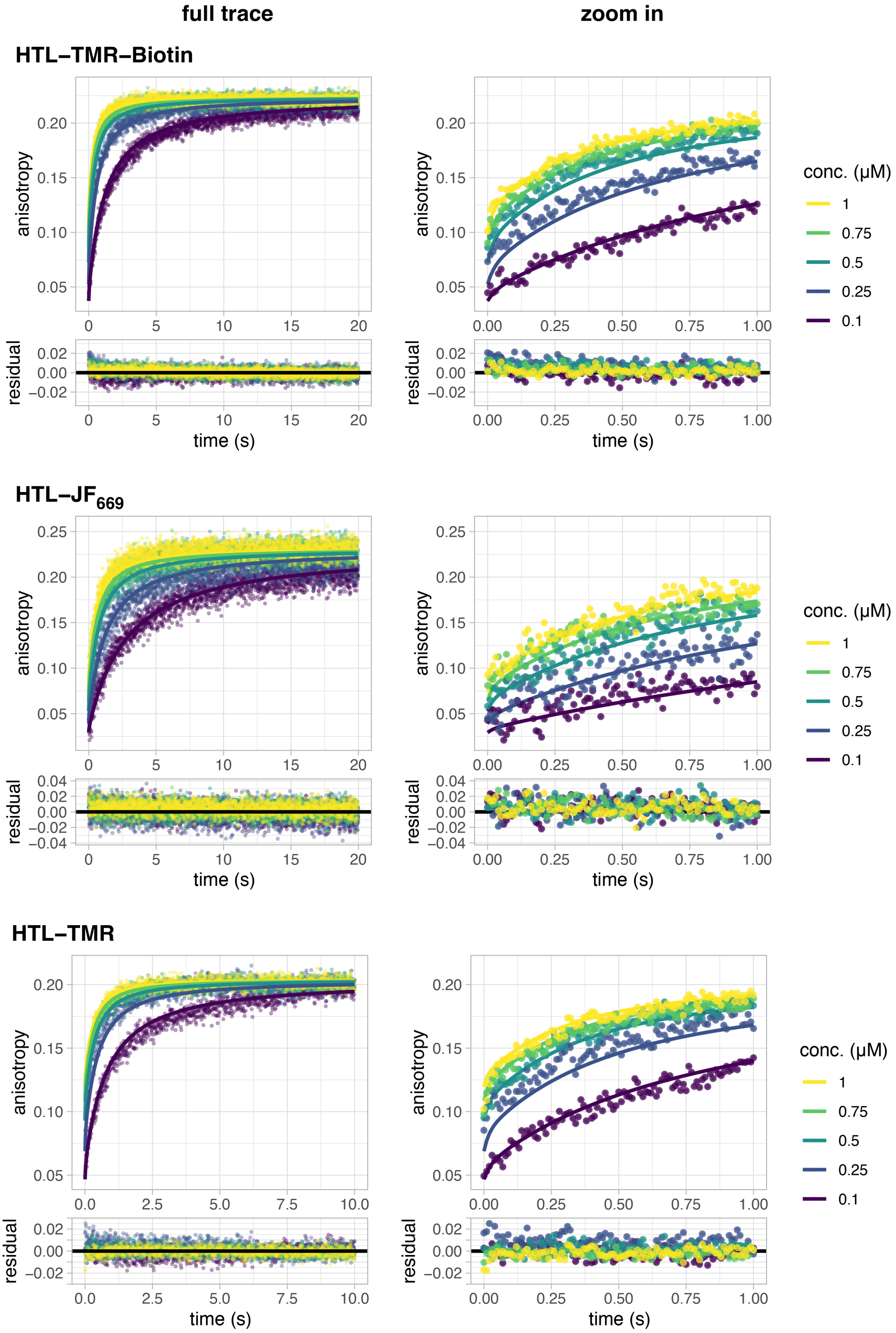


Figure S5: HaloTag labeling kinetics with fluorescent ligands. Stopped-flow anisotropy traces of HaloTag labeling with various zwitterionic rhodamine-derived ligands in 1:1 stoichiometry at the indicated concentrations. A two-step reaction model was fit to the data to determine the rate constants k_1, k_{-1} and k_2. Residuals of the respective fit are depicted in the bottom panels. Full traces are shown on the left, zoom ins to the first second on the right. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

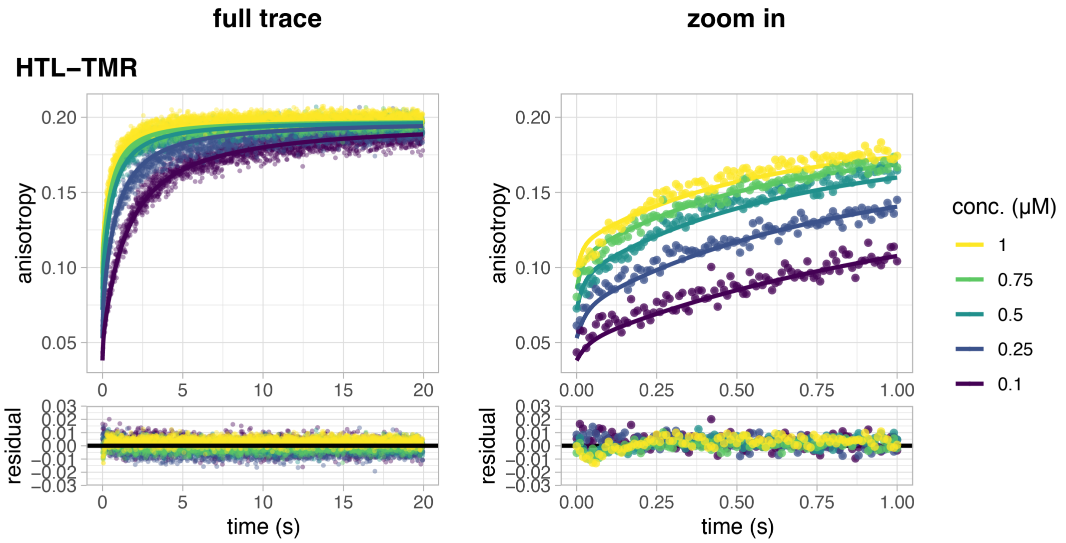


Figure S6: HOB labeling kinetics with HTL-TMR. Stopped-flow anisotropy traces of HOB labeling with HTL-TMR in 1:1 stoichiometry at the indicated concentrations. A two-step reaction model was fit to the data to determine the rate constants k_1, k_{-1} and k_2. Residuals of the respective fit are depicted in the bottom panels. Full traces are shown on the left, zoom ins on the first second on the right. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

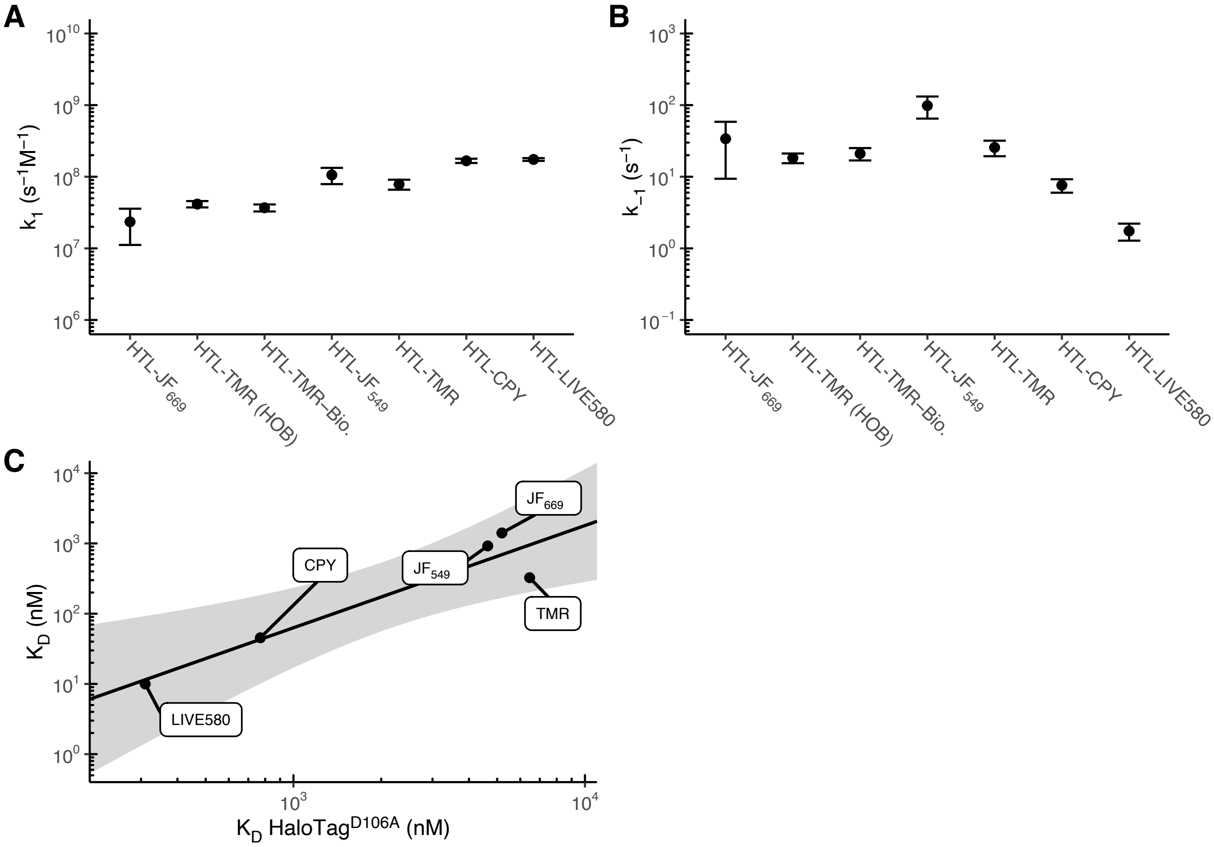


Figure S7: HaloTag and HOB association and dissociation rate constants with fluorophore ligands. Association (k_1, **A**) and dissociation (k_{-1}, **B**) rate constants of HaloTag or HOB (indicated in parentheses) labeling determined by fitting a two-step reaction model (equation 3.2) to stopped-flow anisotropy traces. (**C**) Dissociation constants calculated from the kinetic parameters ($K_D = k_{-1}/k_1$) correlate well with dissociation constants determined using the inactive variant HaloTag$^{\text{D106A}}$. A linear model was fit to the log-transformed values ($\log y = 1.455 \cdot \log x - 2.567$, $R^2 = 0.856$). The gray area represents the 95 % confidence bands. Error bars represent 95 % confidence intervals. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

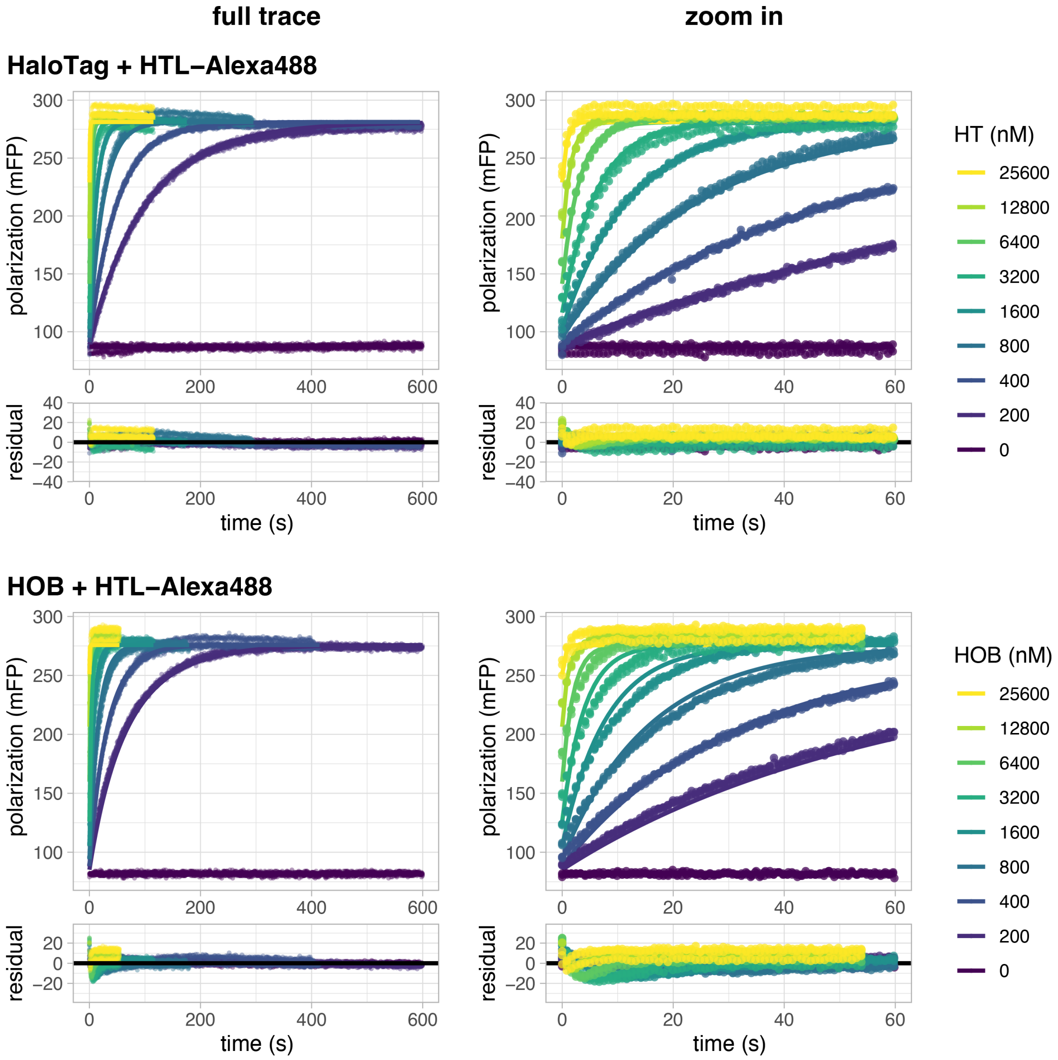


Figure S8: HaloTag and HOB labeling kinetics with HTL-Alexa488. Fluorescence polarization traces of HaloTag (HT) and HOB labeling with HTL-Alexa488 (50 nM). A one-step second-order reaction model was fit to the data to determine k_{app} values. Residuals of the respective fit are depicted in the bottom panels. Full traces are shown on the left, zoom ins on the first second on the right. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

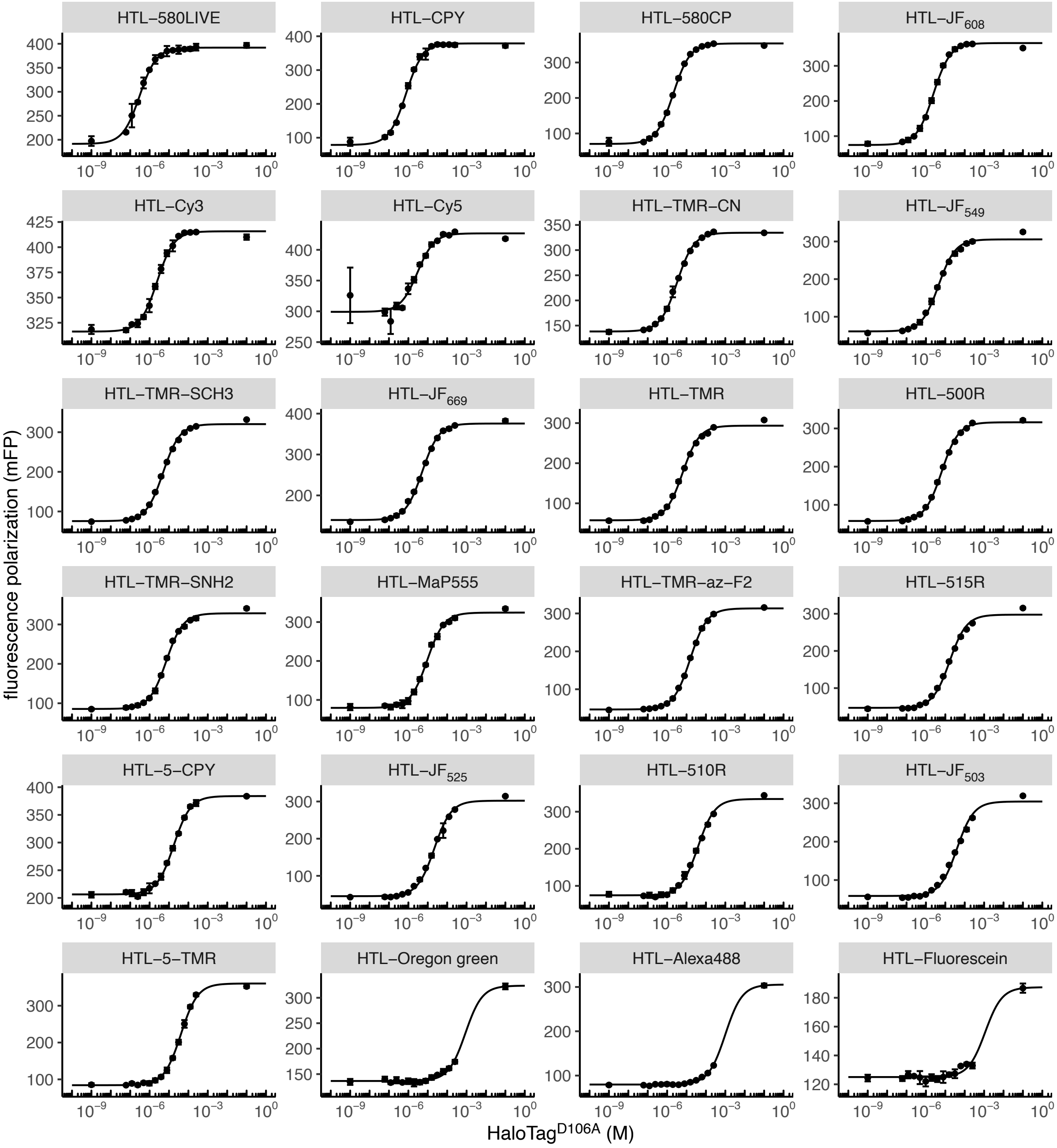


Figure S9: Affinities of fluorophore ligands to HaloTag^{D106A}. Fluorescence polarization titrations of various fluorescent HaloTag ligands (10 nM) with the inactive HaloTag^{D106A} variant. A single-site biding model was fit to the data to determine K_D values. The fluorescence polarization of each dye fully bound to active HaloTag was added at 100 mM to improve fitting of the upper plateau. Error bars represent 95 % confidence intervals. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

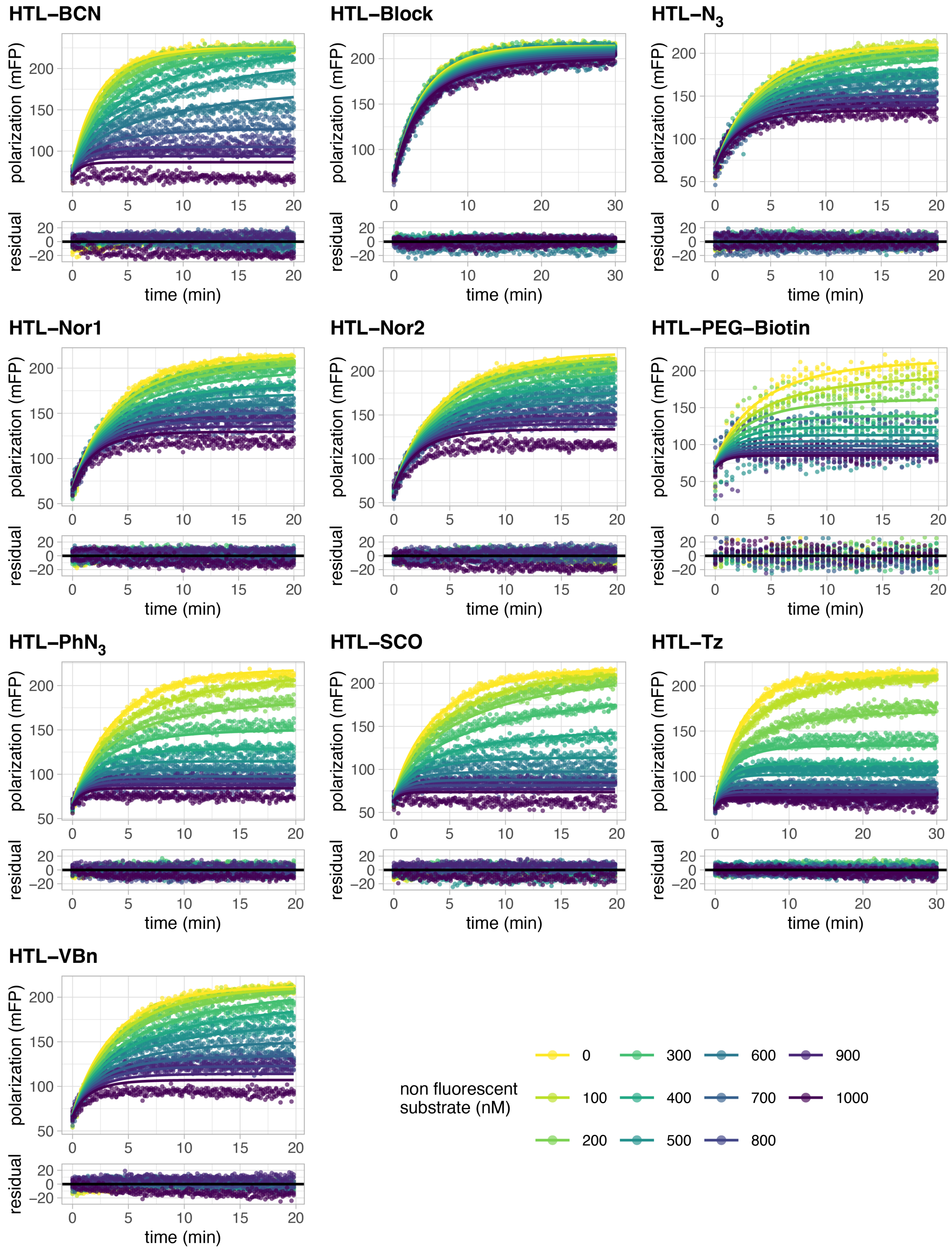


Figure S10: HaloTag labeling kinetics with non-fluorescent ligands via competition assay. Fluorescence polarization traces of HaloTag (200 nM) labeling with various non-fluorescent ligands in competition with HTL-Alexa488 (50 nM). A competition model was fit to the data to determine the apparent second-order rate constants of the non-fluorescent ligands. Residuals of the respective fit are depicted in the bottom panels. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

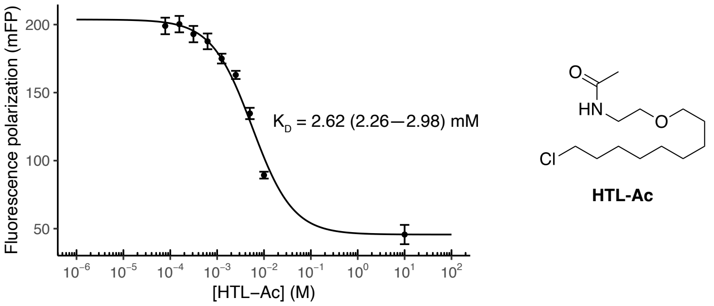


Figure S11: Affinity of HaloTag^{D106A} for HTL-Ac. Fluorescence polarization competition assay of HTL-Ac against HTL-TMR (50 nM) with the inactive HaloTag^{D016A} variant (5 µM). A four parameter logistic model was fit to the data and the dissociation constant (K_D) was calculated as described by Rossi and Taylor[294] Error bars represent 95 % confidence intervals. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

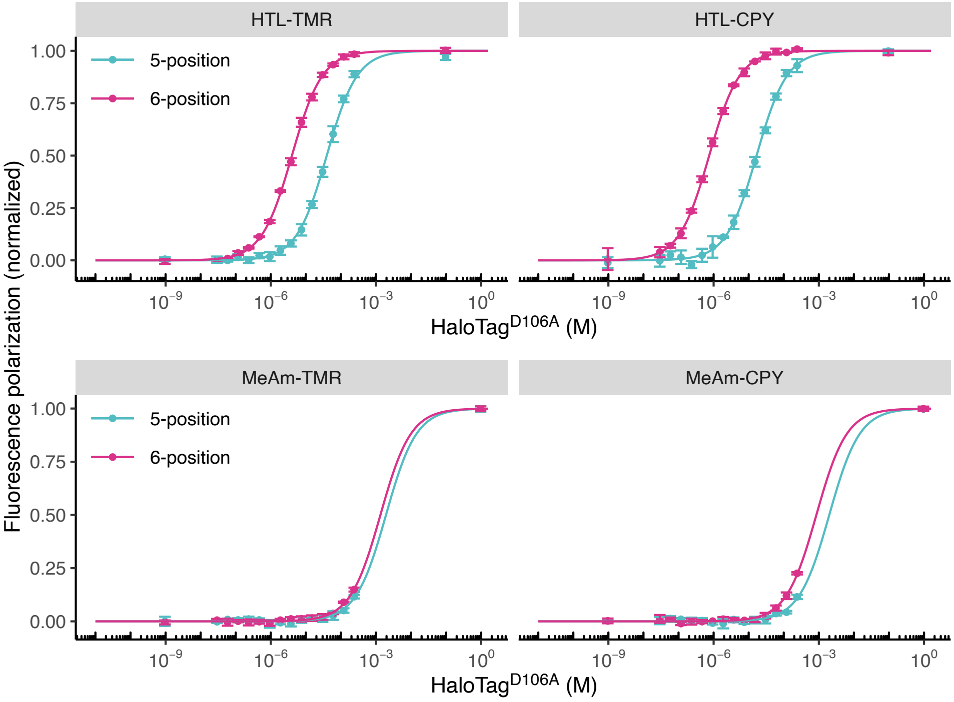


Figure S12: Affinity of HaloTag^{D106A} for 5- or 6-position modified rhodamines. Fluorescence polarization titrations of HaloTag ligands and meAm-rhodamines with HaloTag. A single-site binding model was fit to the data to determine K_D values. The fluorescence polarization of each dye fully bound to active HaloTag was added at 0.1 M (for HaloTag ligands) or 1 M (for MeAm-rhodamines) to improve fitting of the upper plateau. Error bars represent 95 % confidence intervals. Reprinted with permission from Wilhelm & Kühn et al. Biochemistry 2021, 60, 2560–2575[188]. Copyright 2021 American Chemical Society.

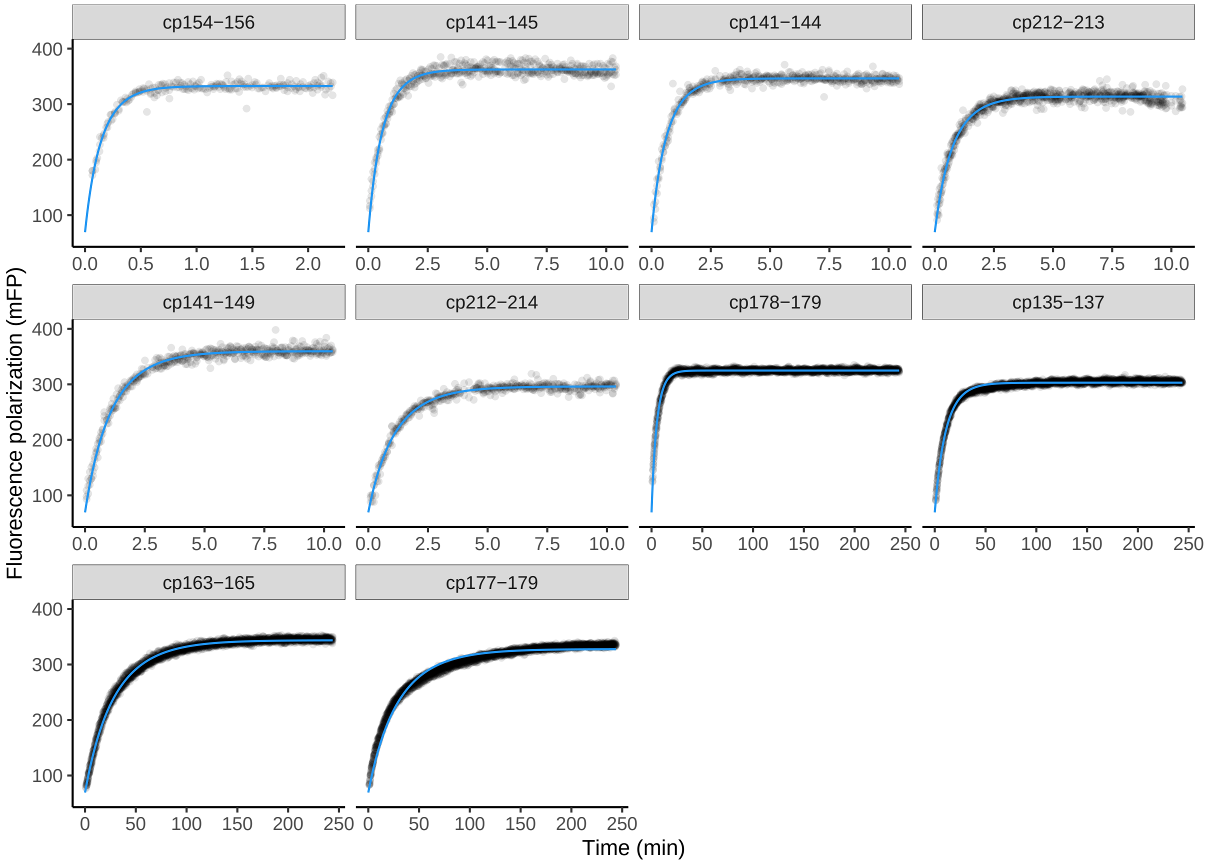


Figure S13: Labeling kinetics of circular permuted HaloTag variants. Fluorescence polarization traces following the labeling reaction of circular permuted HaloTag variants (50 nM) with HTL-TMR (10 nM). A second-order reaction model was fit to the data to determine k_{app} values.

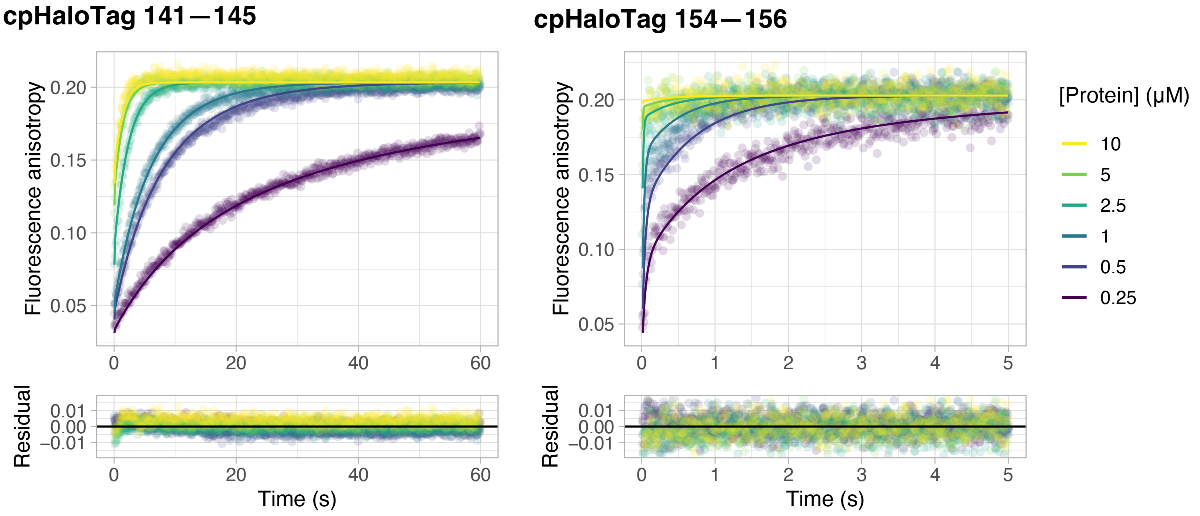


Figure S14: Stopped-flow kinetics of cpHaloTag 141–145 and 154–156. Anisotropy traces of cpHaloTag labeling with HTL-TMR (250 nM). A two-step reaction model was fit to the data to determine the rate constants k_1, k_{-1} and k_2. Residuals of the respective fits are depicted in the bottom panels.

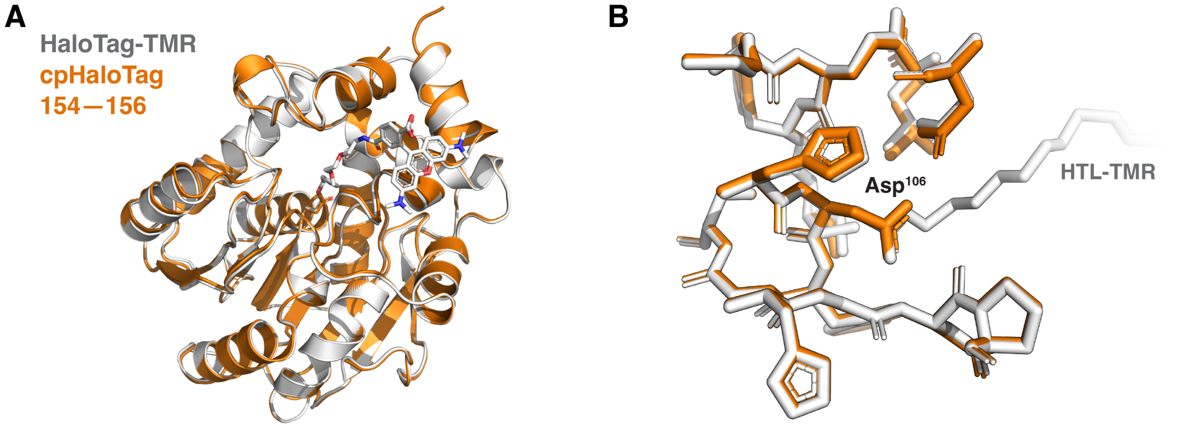


Figure S15: Structure of cpHaloTag 154–156 compared to native HaloTag. (**A**) Overlay of the cpHaloTag 154–156 (orange) and native HaloTag (grey, PDB-ID 6Y7A)[188] structures. Except for the new termini opened in cpHaloTag 154–156, the structures are nearly identical (C_α-RMSD = 0.195 Å). The flexible $(GGS/T)_5$ linker connecting the original termini in cpHaloTag 154–156 is not resolved in the structure. (**B**) Overlay of the core residues in proximity to the active Asp^{106}. The structures are nearly identical (C_α-RMSD = 0.111 Å).

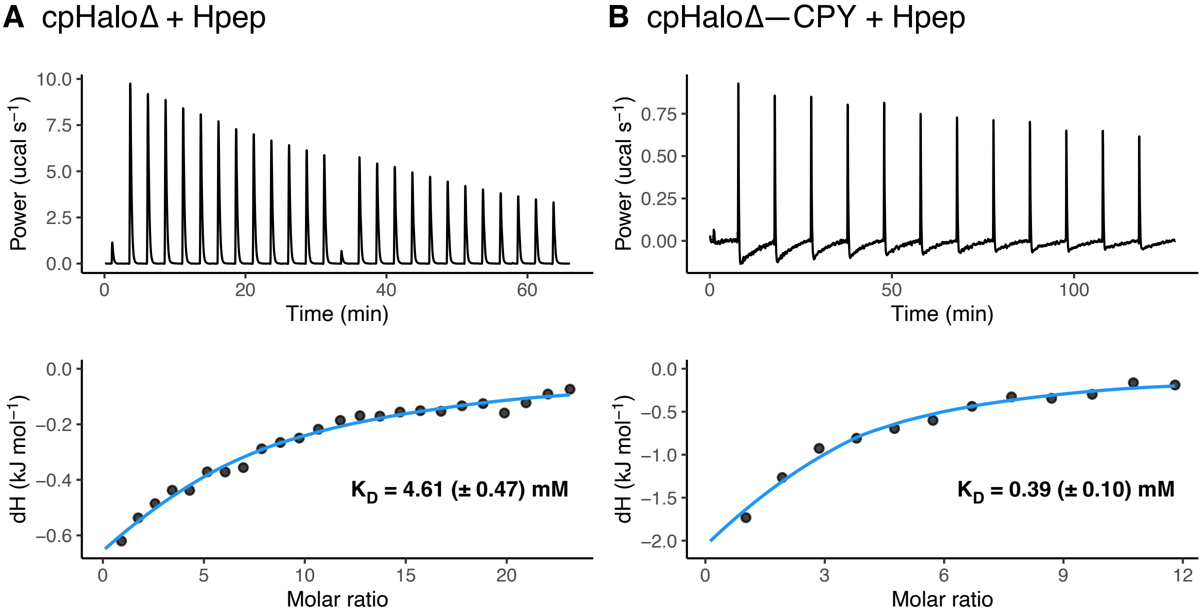


Figure S16: Intrinsic affinity of split-HaloTag fragments. Affinity of Hpep for cpHaloΔ (**A**) or HTL-CPY labeled cpHaloΔ (**B**) measured by ITC. K_D values (± std. dev.) are annotated. Labeling with HTL-CPY increases Hpep affinity 12-fold.

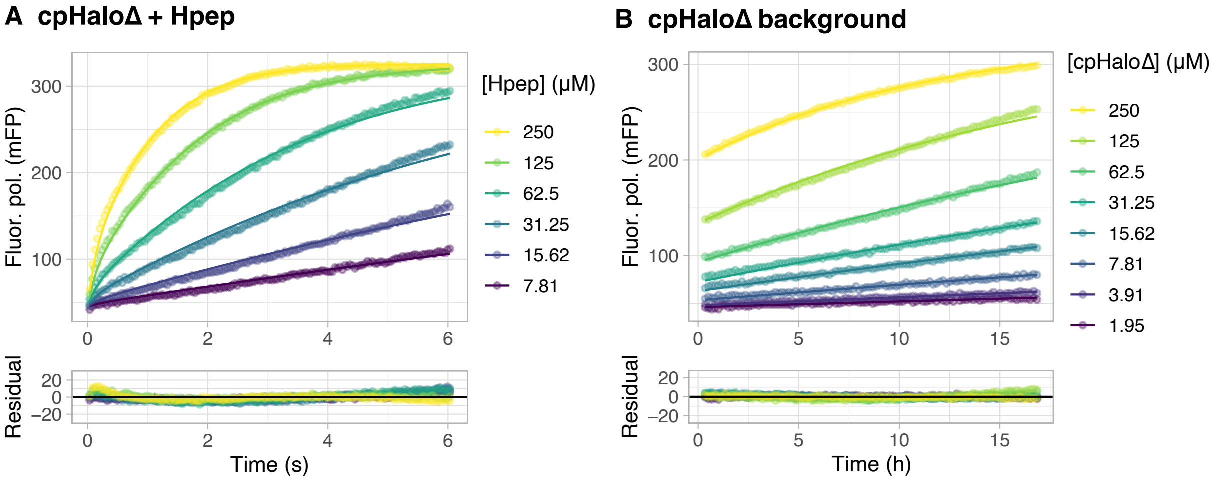


Figure S17: Labeling kinetics of split-HaloTag. (**A**) cpHaloΔ (500 µM) labeling kinetics with HTL-TMR (50 nM) in the presence of Hpep at specified concentrations. A kinetic model that includes cpHaloΔ activation by Hpep binding and subsequent reaction (equations 6.24 and 6.25) with HTL-TMR was fit to the data to determine the apparent second-order rate constant of complemented split-HaloTag labeling. $k_{app} = 1.33$ [1.31–1.35] $\cdot 10^4$ M^{-1} s^{-1}. (**B**) cpHaloΔ background labeling kinetics with HTL-TMR (50 nM) in the absence of Hpep at specified protein concentrations. A two-step reaction model (equation 3.2) was fit to the data to determine K_D, k_2 and k_{app} of the reaction. $K_D = 206$ [201–211] µM, $k_2 = 4.36$ [4.27–4.45] $\cdot 10^{-5}$ s^{-1}, $k_{app} = 0.212$ [0.204–0.219] M^{-1} s^{-1}.

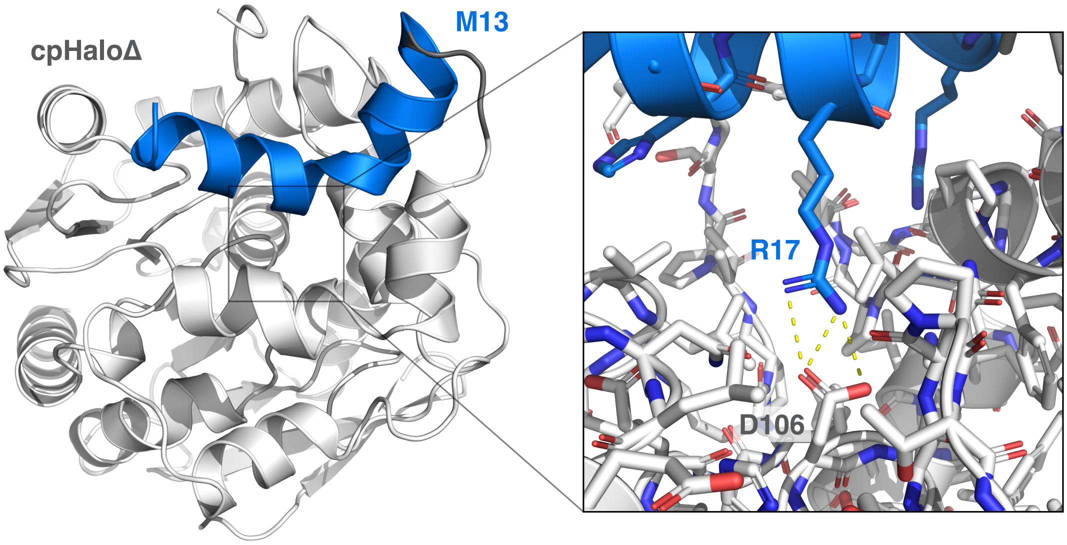


Figure S18: M13–cpHaloΔ crystal structure. The Hpep binding groove of cpHaloΔ is occupied by the M13 peptide (left). The zoom in (right) shows how Arg^{17} of M13 extends through the ligand entry channel and interacts with the reactive Asp^{106} residue, blocking the entry of HaloTag ligands.

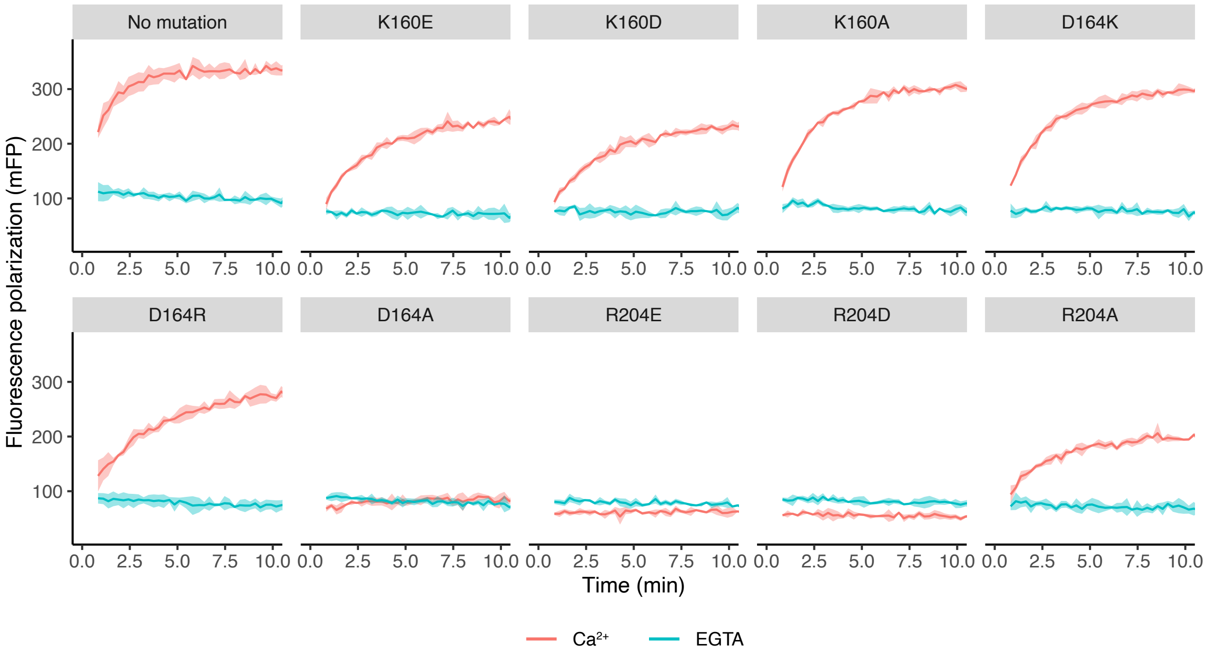


Figure S19: Caprola cpHaloΔ–CaM interface mutants. Labeling kinetics of Caprola variants (200 nM) with different cpHaloΔ–CaM interface mutations with HTL-TMR (50 nM) in presence of absence (EGTA) of Ca^{2+}. All interface mutations lead to a decreased labeling rate in presence of Ca^{2+}, suggesting that they play a role for proper alignment of the domains in the Ca^{2+} bound state. Residue numbers according to standard HaloTag numbering.

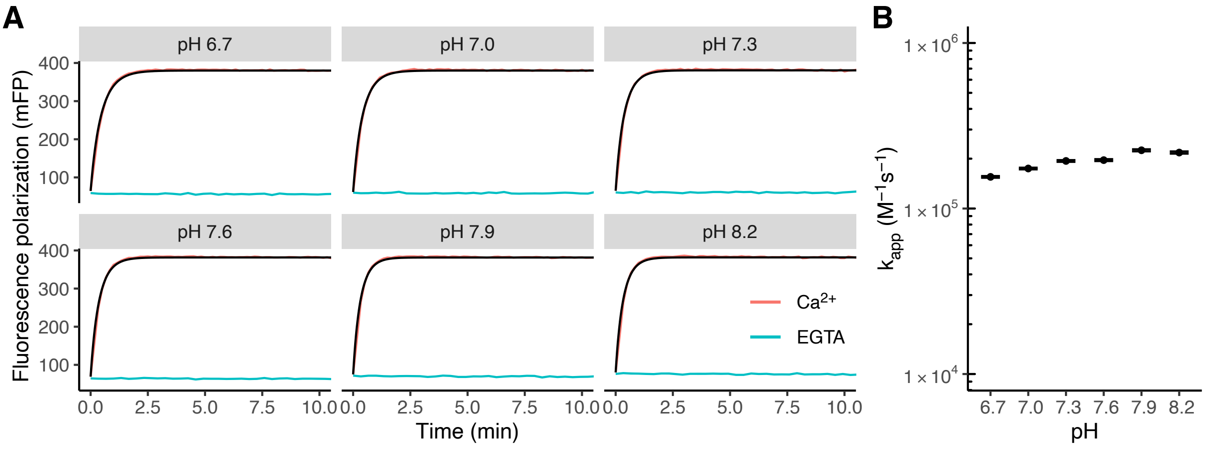


Figure S20: pH dependency of Caprola labeling. (**A**) Fluorescence polarization traces of Caprola (200 nM) labeling with HTL-TMR (50 nM) in presence of Ca^{2+} (5 mM) or EGTA (100 μM) at different pH values. A second-order reaction model was fit to the Ca^{2+} kinetics to determine the apparent second-order rate constant (k_app). (**B**) The comparison of the apparent second-order rate constants (k_app) shows that Caprola is mostly insensitive to pH changes within the physiological range. Error bars represent 95 % confidence intervals.

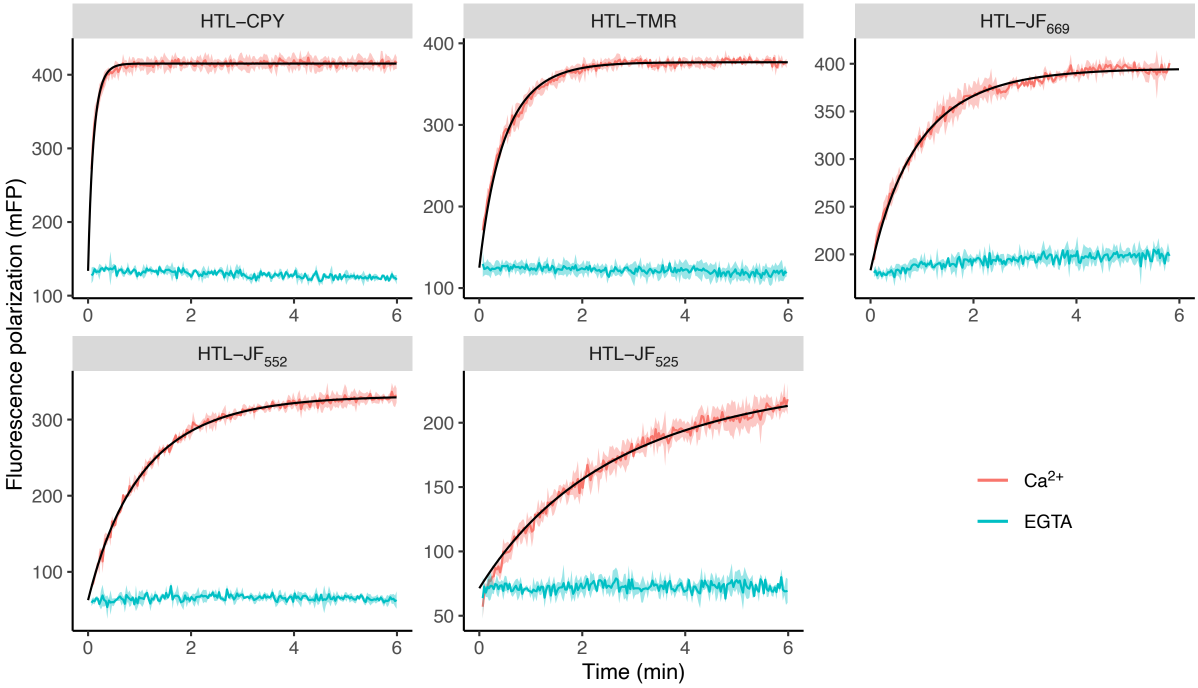


Figure S21: Caprola performance with different fluorescent ligands. Caprola (200 nM) labeling kinetics with HTL-CPY, HTL-TMR, HTL-JF$_{669}$, HTL-JF$_{552}$ and HTL-JF$_{525}$ (50 nM each) in presence of 5 mM Ca^{2+} or 100 µM EGTA. A second-order reaction model was fit to Ca^{2+} kinetics to determine the apparent second-order rate constant (k_{app}).

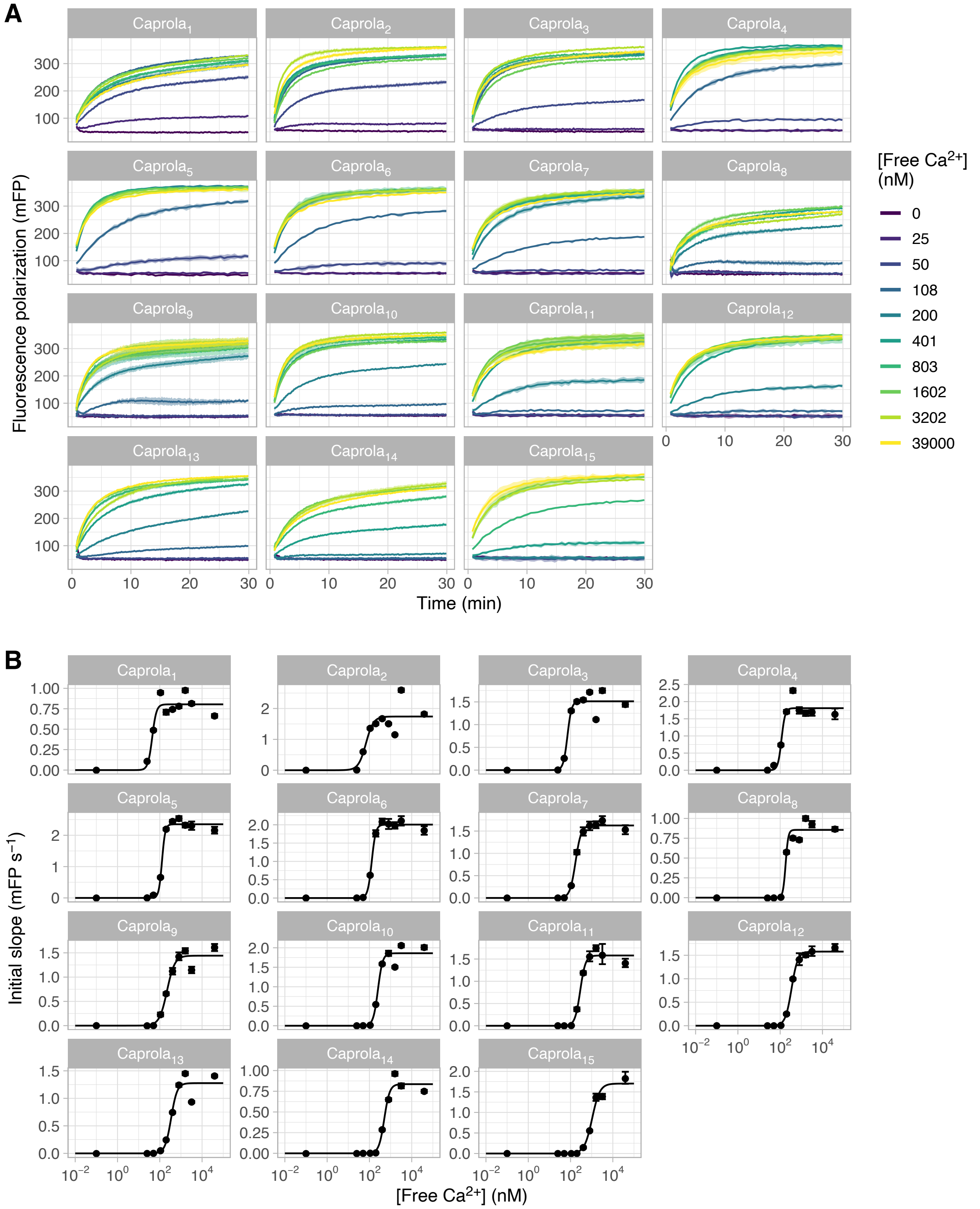


Figure S22: Calcium sensitivity of Caprola$_1$–Caprola$_{15}$ (**A**) Labeling reactions of Caprola$_1$–Caprola$_{15}$ (200 nM) with HTL-TMR (50 nM) at different free Ca^{2+} concentrations monitored by fluorescence polarization. (**B**) Initial slopes of the labeling reactions from (A) plotted against the free Ca^{2+} concentrations. A four-parameter logistic model was fit to the data to determine the Ca^{2+}-sensitivity (EC_{50}^{Ca2+}). The steep hill slope of the fit model indicates strong cooperativity of the Ca^{2+} mediated Caprola activation. The EC_{50}^{Ca2+} represents the calcium concentration at which a variant exhibits half of its maximum labeling speed. Error bars represent 95 % confidence intervals.

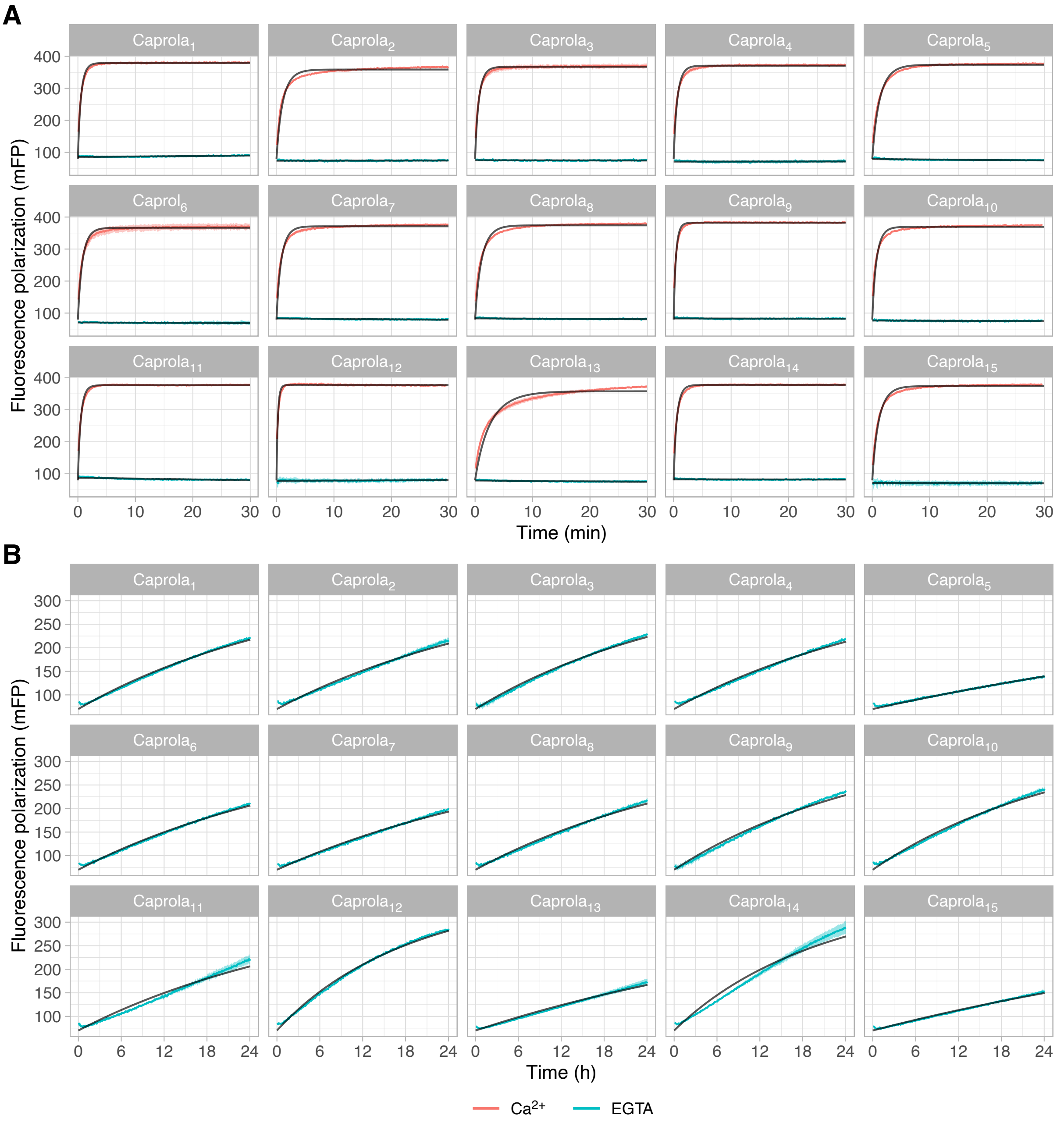


Figure S23: Labeling kinetics of Caprola$_1$–Caprola$_{15}$ (**A**) Labeling reactions of Caprola$_1$–Caprola$_{15}$ (200 nM) with HTL-TMR (50 nM) in presence or absence (EGTA) of Ca^{2+} (5 mM) monitored by fluorescence polarization. A second-order reaction model was fit to the Ca^{2+} kinetics to determine the apparent second order rate constant (k_{app}^{Ca2+}). (**B**) Rate constants in the absence of Ca^{2+} were determined in a long term 24 h fluorescence polarization assay. A second-order reaction model with fixed plateau was fit to the kinetics to determine the apparent second order rate constant (k_{app}^{EGTA}).

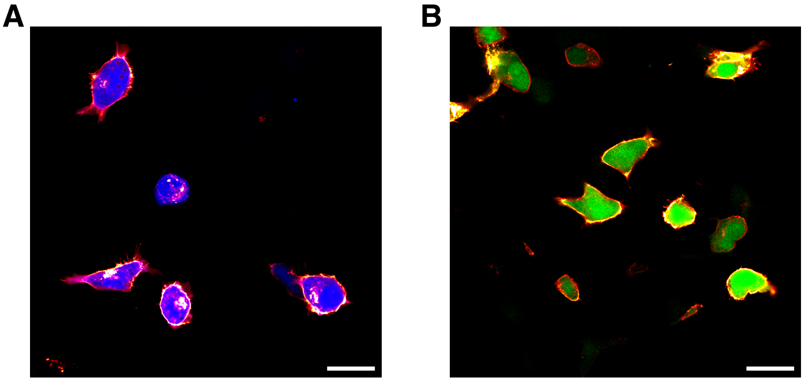


Figure S24: Immunofluorescence of HEK-293 cells expressing first generation pre- and post-SynProLa constructs. Exemplary confocal microscopy images of PFA fixed HEK-293 cells expressing first generation pre-SynProLa (**A**: mTagBFP2–F2A–NRXss–FLAG-tag–cpHaloΔ–hNRX-3-β) and post-SynProLa (**B**: EGFP–F2A–NLGss–V5-tag–Hpep–mNLG-1) constructs. Alexa568 stained FLAG-tag and V5-tag (red-hot lookup table) show the expected localization of both SynProLa constructs at the plasma membrane. The co-expressed mTagBFP (blue) and EGFP (green) reporter proteins are localized in the cytosol. Scale bars: 25 µm.

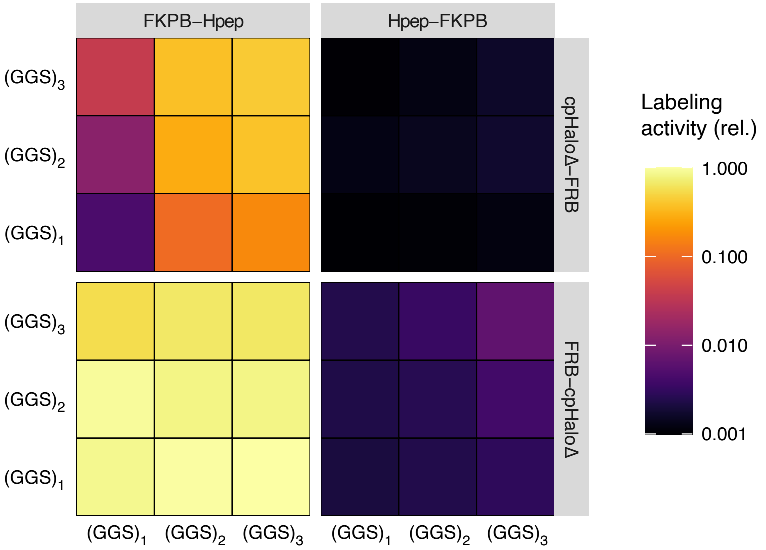


Figure S25: FKBP/FRB split-HaloTag model system (continued from fig. 19). Relative labeling activities of combinations of Hpep–FKBP and cpHaloΔ–FRB fusions in presence of rapamycin with HTL-TMR. split-HaloTag fragments (cpHaloΔ and Hpep) are linked to either N-termini or C-termini of FKBP and FRB via flexible linkers with varying length $(GGS)_{1\text{-}3}$. When both split fragments are linked to the close C-termini of the complex activity is highest. Minimal activity is observed when split-HaloTag is linked to the distant N-termini of the complex. Especially at longer fusion site distance, longer linkers lead to higher activity.

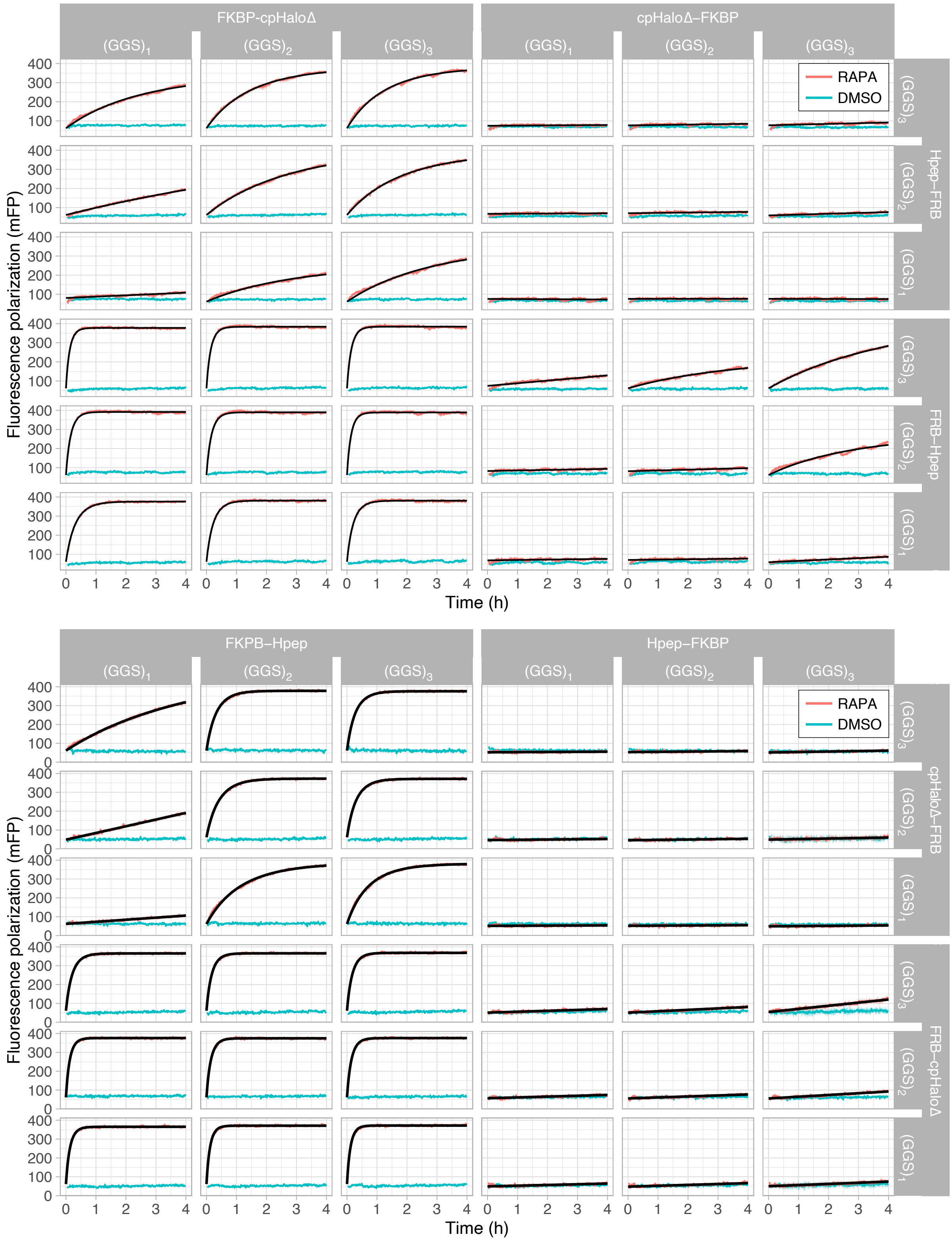


Figure S26: FKBP/FRB split-HaloTag model system kinetics. Labeling reactions of combinations of split-HaloTag FKBP/FRB fusions (250 nM each) with HTL-TMR (50 nM) in presence or absence of the dimerizer rapamycin (RAPA, 500 nM) monitored by fluorescence polarization. A second order order reaction model or a linear model (for traces that did not reach a plateau) was fit to the rapamycin kinetics.

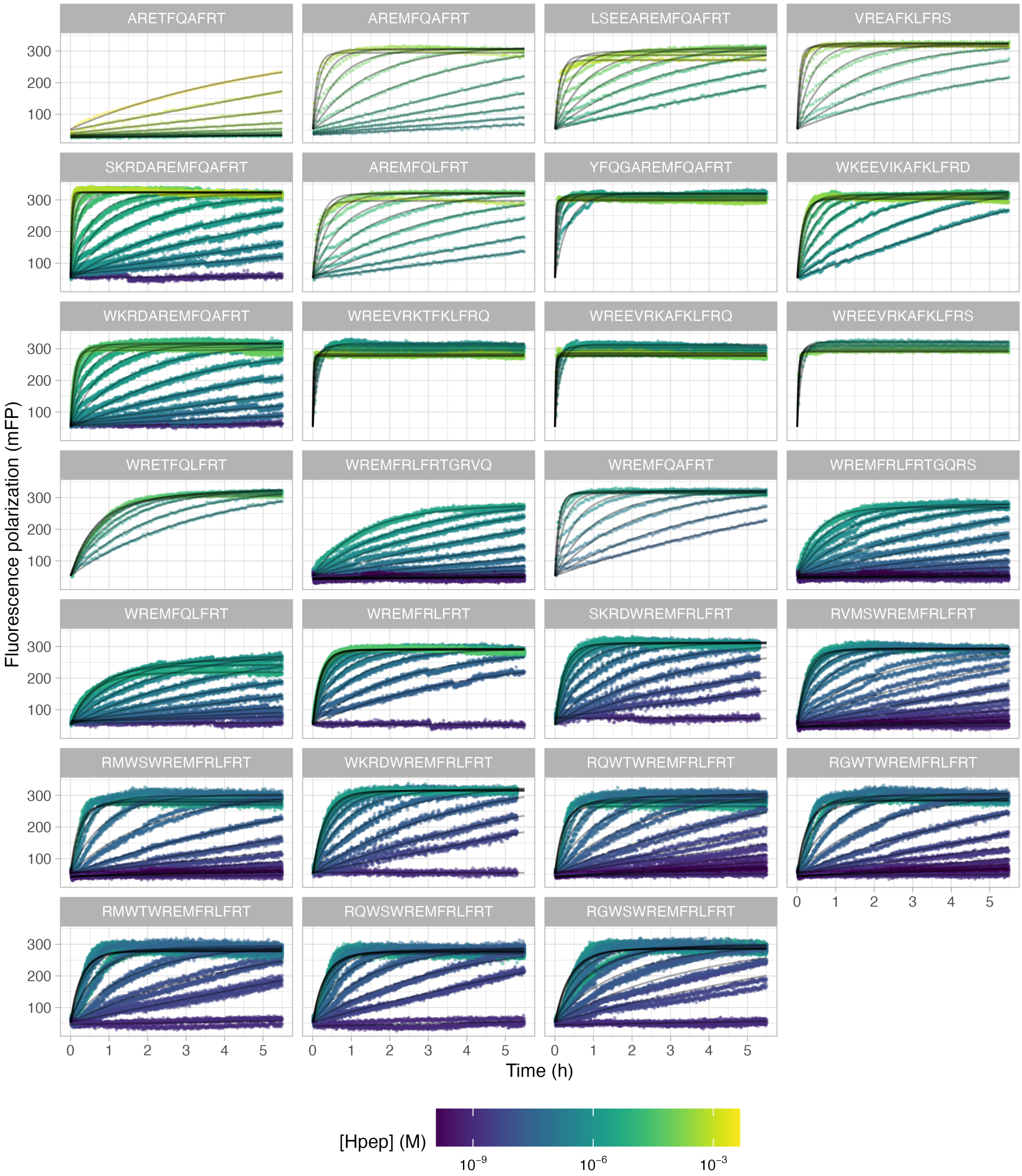


Figure S27: Split-HaloTag labeling kinetics with different Hpep variants. Labeling reactions of split-HaloTag with HTL-TMR at different concentrations of Hpep variants monitored by fluorescence polarization. A second-order reaction model (or a linear model for reactions not reaching a plateau) was fit to the data to determine apparent reaction rate constants.

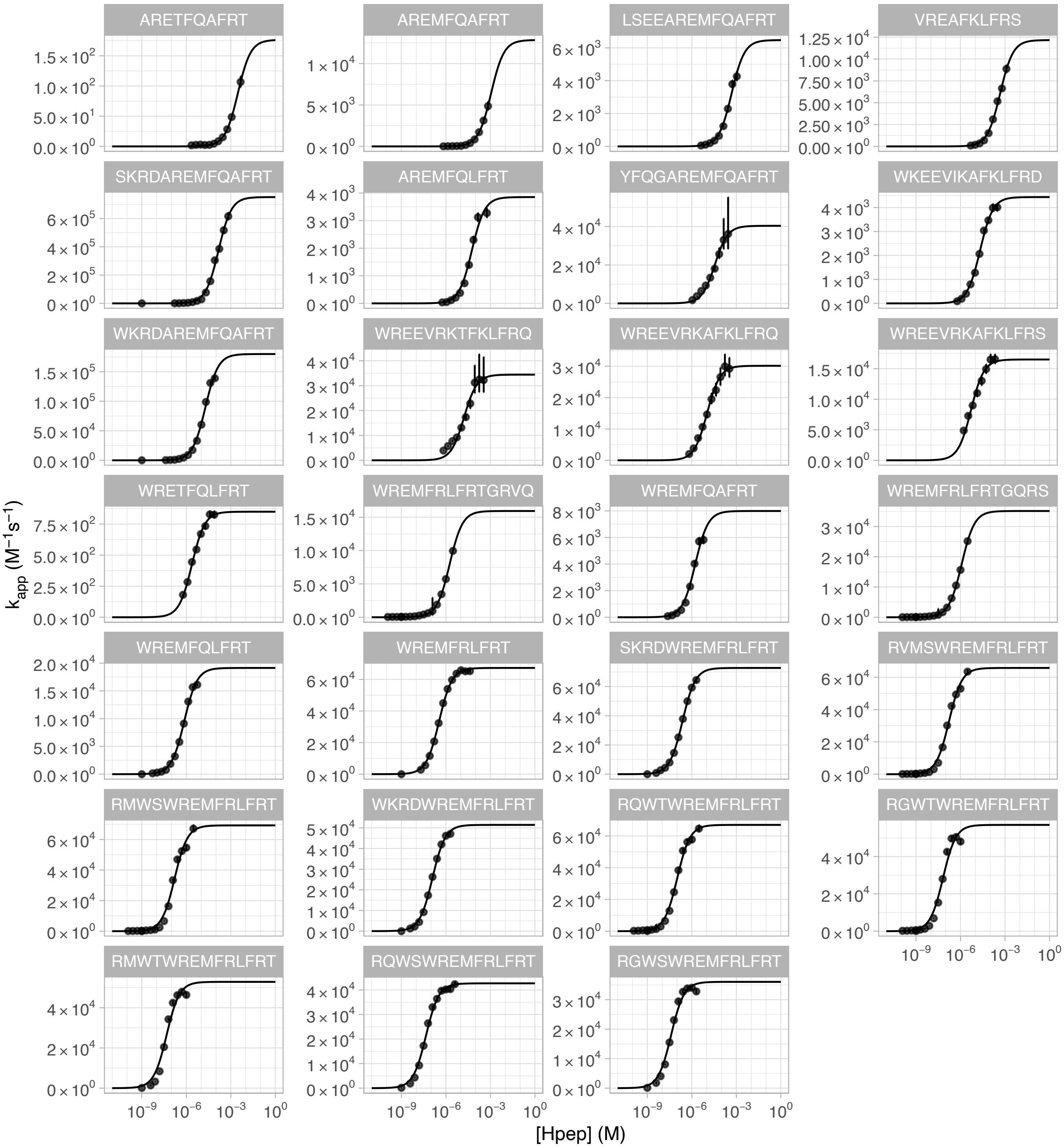


Figure S28: Analysis of split-HaloTag labeling kinetics with different Hpep variants. Apparent second-order reaction rate constants from split-HaloTag labeling kinetics shown in figure S27 plotted against Hpep concentration. A sigmoidal model was fit to the data to estimate half maximal effective concentrations (EC_{50}) and maximal apparent labeling rate constants (k_{app}^{max}). Error bars represent 95 % confidence intervals. Fit values are listed in table S12.

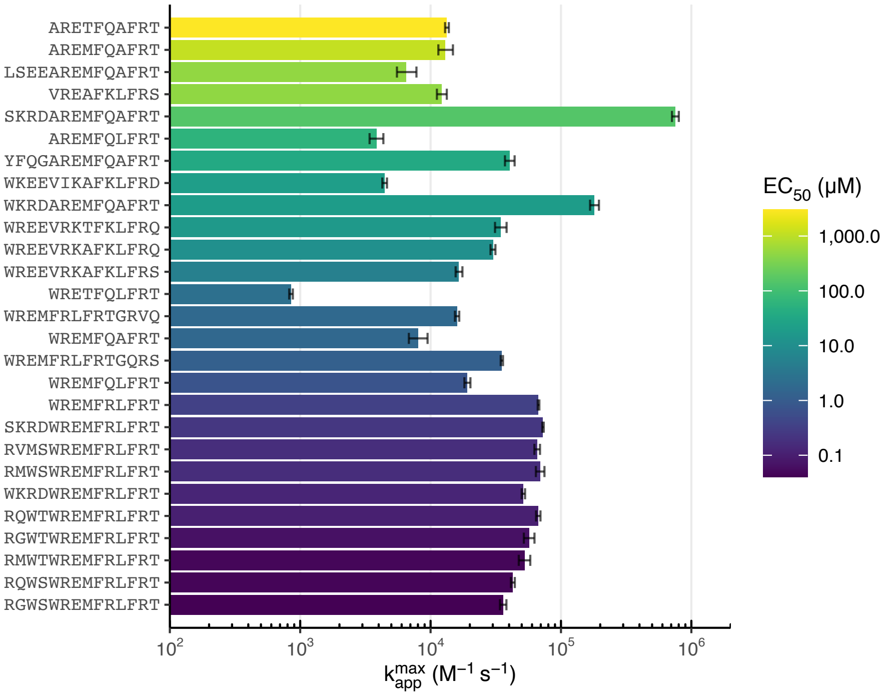


Figure S29: Maximal second-order labeling rate constants of Hpep variants. Values are listed in table S12. Error bars represent 95 % confidence intervals.

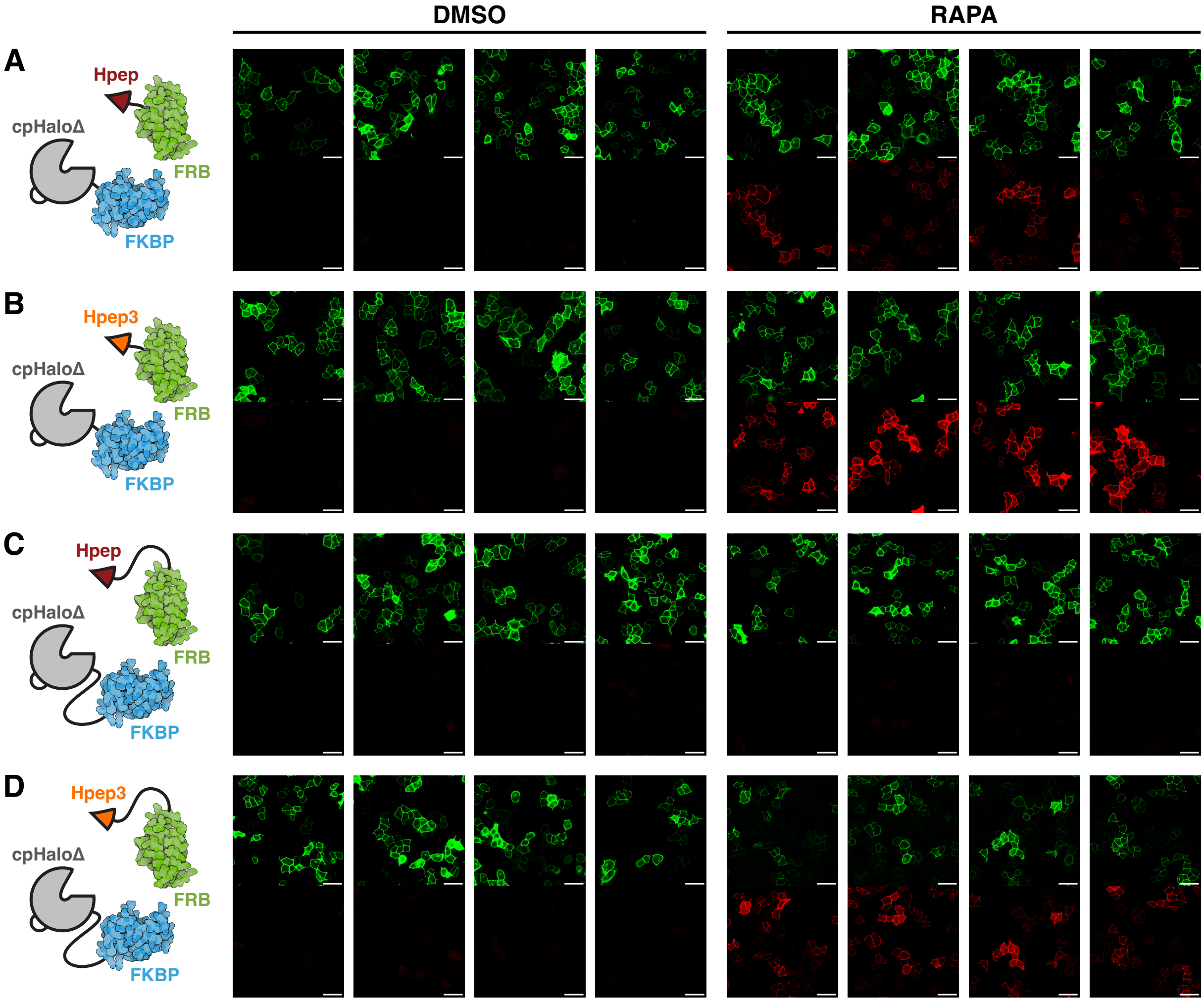


Figure S30: FKBP/FRB model system with improved Hpep variants in HeLa cells analyzed by confocal microscopy. Confocal microscopy images of HeLa cells co-expressing different split-HaloTag FKPB and FRB fusion proteins with different Hpep variants and orientations: (**A**) Lyn11-EGFP-FKBP-cpHaloΔ & FRB-Hpep1-mScarlet, (**B**) Lyn11-EGFP-FKBP-cpHaloΔ & FRB-Hpep5-mScarlet, (**C**) Lyn11-EGFP-cpHaloΔ-FKBP & Hpep1-FRB-mScarlet or (**D**) Lyn11-EGFP-cpHaloΔ-FKBP & Hpep5-FRB-mScarlet. Cells were labeled in presence (RAPA) or absence (DMSO) of 100 nM rapamycin for 1 h. Scale bars: 50 µm.

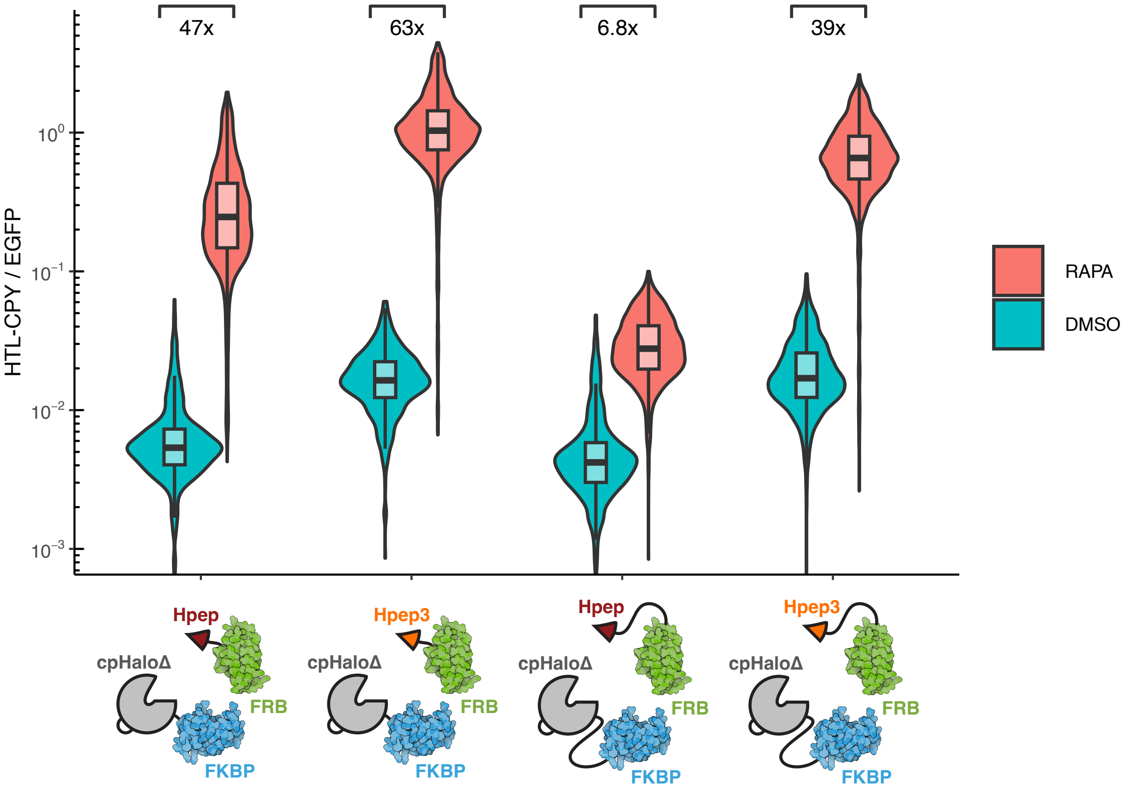


Figure S31: FKBP/FRB model system with improved Hpep variants in HeLa cells analyzed by flow cytometry. Flow cytometry analysis of HeLa cells co-expressing different split-HaloTag FKPB and FRB fusion proteins with different Hpep variants and orientations: Lyn11-EGFP-FKBP-cpHaloΔ & FRB-Hpep1-mScarlet, Lyn11-EGFP-FKBP-cpHaloΔ & FRB-Hpep5-mScarlet, Lyn11-EGFP-cpHaloΔ-FKBP & Hpep1-FRB-mScarlet or Lyn11-EGFP-cpHaloΔ-FKBP & Hpep5-FRB-mScarlet. Cells were labeled in presence (RAPA) or absence (DMSO) of 100 nM rapamycin for 1 h.

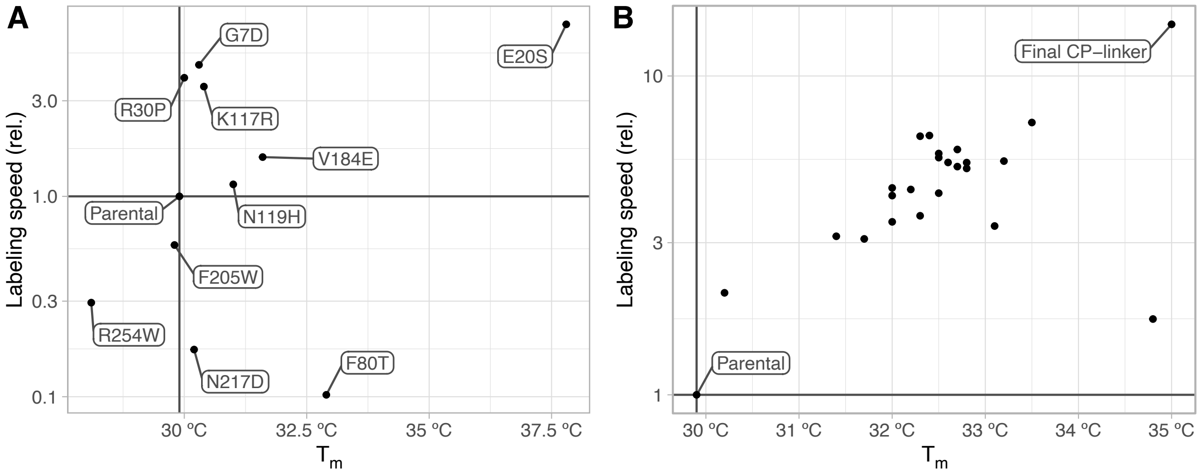


Figure S32: cpHaloΔ point mutation and CP-linker screening. Screening results from PROSS[262] predicted point mutations and Rosetta Remodel[263] redesigned CP-linkers. Data was recorded by Lennart Nickel.

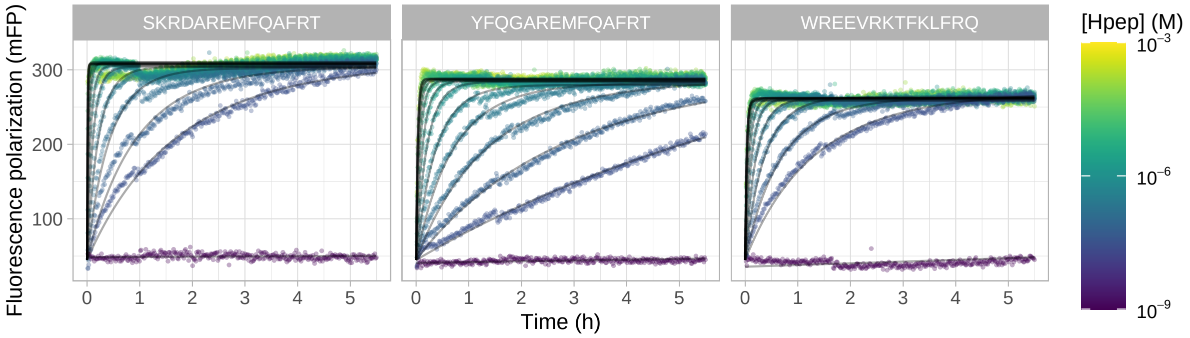


Figure S33: cpHaloΔ2 labeling kinetics with different Hpep variants. Labeling reactions of cpHaloΔ2 with HTL-TMR at different concentrations of Hpep variants monitored by fluorescence polarization. A second-order reaction model (or a linear model for reactions not reaching a plateau) was fit to the data to determine reaction rate constants. Data was recorded by Lennart Nickel.

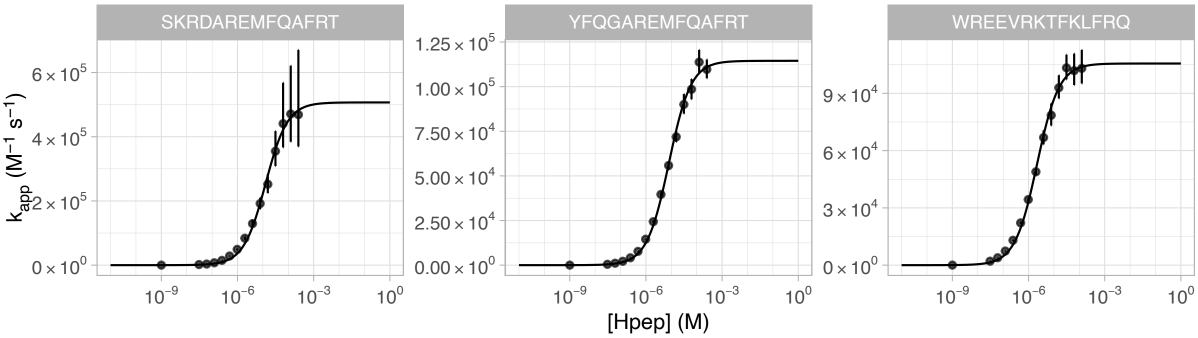


Figure S34: Analysis of cpHaloΔ2 labeling kinetics with different Hpep variants. Apparent second-order reaction rate constants from split-HaloTag labeling kinetics shown in figure S27 plotted against Hpep concentration. A sigmoidal model was fit to the data to estimate half maximal effective concentrations (EC_{50}) and maximal apparent labeling rate constants (k_{app}^{max}). Error bars represent 95 % confidence intervals. Fit values are listed in table S13. Data was recorded by Lennart Nickel.

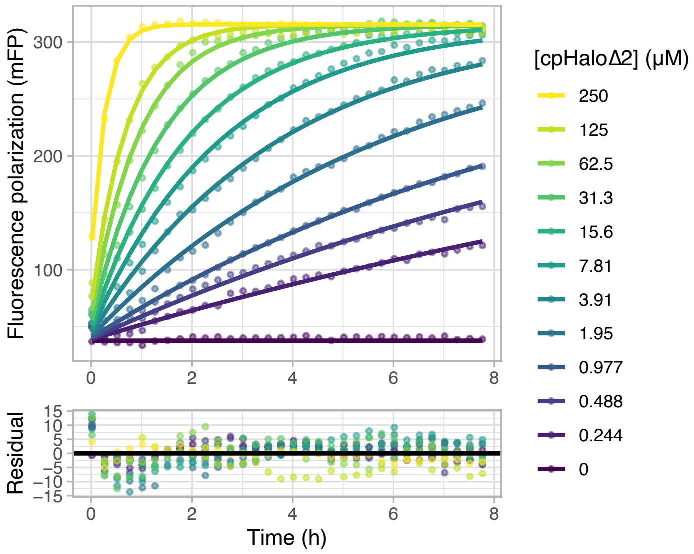


Figure S35: Background labeling kinetics of cpHaloΔ2 in absence of Hpep. cpHaloΔ2 background labeling kinetics with HTL-TMR (50 nM) in the absence of Hpep at specified protein concentrations. A two-step reaction model (equation 3.2) was fit to the data to determine k_{app} of the reaction. $k_{app} =$ 5.49 [5.04–5.94] M^{-1} s^{-1}. Data was recorded by Lennart Nickel.

Supplementary Tables

Table S1: Kinetic constants (k_1, k_{-1}, k_2) of HaloTag and HOB (indicated in parentheses) labeling fit with the two-step reaction model. For experimental data and fits see figures S5 and S6.

Ligand	k_1 **[95 % CI[a]] (M^{-1} s^{-1})**	k_{-1} **[95 % CI[a]] (s^{-1})**	k_2 **[95 % CI[a]] (s^{-1})**	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**
HTL-LIVE580	1.74 [1.67–1.82] $\cdot 10^8$	1.75 [1.28–2.21]	6.77 [5.50–8.04]	1.39 [1.34–1.43] $\cdot 10^8$
HTL-CPY	1.67 [1.56–1.78] $\cdot 10^8$	7.60 [5.98–9.22]	9.86 [8.66–11.1]	9.44 [9.15–9.73] $\cdot 10^7$
HTL-TMR	7.84 [6.59–9.09] $\cdot 10^7$	25.6 [19.3–31.8]	8.06 [7.58–8.54]	1.88 [1.86–1.90] $\cdot 10^7$
HTL-TMR (HOB)	4.15 [3.73–4.57] $\cdot 10^7$	18.3 [15.5–21.1]	5.05 [4.83–5.26]	8.99 [8.92–9.06] $\cdot 10^6$
HTL-TMR-Biotin	3.69 [3.28–4.10] $\cdot 10^7$	21.0 [16.9–25.1]	8.24 [7.79–8.70]	1.04 [1.03–1.05] $\cdot 10^7$
HTL-JF_{549}	1.06 [0.79–1.33] $\cdot 10^8$	98.3 [64.8–131]	18.2 [17.1–19.3]	1.66 [1.64–1.67] $\cdot 10^7$
HTL-JF_{669}	2.35 [1.12–3.58] $\cdot 10^7$	33.9 [9.42–58.5]	6.94 [6.17–7.71]	4.03 [3.99–4.07] $\cdot 10^6$

[a]CI: confidence interval

Table S2: Apparent second-order rate constants of HaloTag and HOB labeling with HTL-Alexa488 fit with a one-step second-order reaction model. For experimental data and fits see figure S8.

Ligand	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**
HaloTag	2.57 [2.56–2.57] $\cdot 10^4$
HOB	8.04 [8.01–8.07] $\cdot 10^4$

[a]CI: confidence interval

Table S3: Apparent second-order rate constants of HaloTag labeling with non-fluorescent ligands determined via a kinetic competition assay against HTL-Alexa488. For experimental data and fits see figure S10.

Ligand	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**
HTL-SCO	2.52 [2.44–2.60] $\cdot 10^5$
HTL-BCN	2.04 [1.98–2.10] $\cdot 10^5$
HTL-VBn	4.27 [4.09–4.45] $\cdot 10^4$
HTL-Tz	3.13 [3.08–3.17] $\cdot 10^4$
HTL-PhN_3	2.14 [2.11–2.18] $\cdot 10^4$
HTL-PEG-Biotin	1.70 [1.57–1.83] $\cdot 10^4$
HTL-Nor1	6.68 [6.58–6.78] $\cdot 10^3$
HTL-Nor2	6.15 [6.04–6.26] $\cdot 10^3$
HTL-N_3	6.00 [5.91–6.10] $\cdot 10^3$
HTL-Block	1.53 [1.50–1.56] $\cdot 10^3$

[a]CI: confidence interval

Table S4: Affinities of fluorophore ligands to HaloTag^{D106A}. For titration data and fits see figure S9.

Ligand	K_D **[95 % CI[a]] (μM)**
HTL-580LIVE	0.31 [0.27–0.35]
HTL-CPY	0.74 [0.65–0.85]
HTL-580CP	2.08 [1.94–2.23]
HTL-JF$_{608}$	2.43 [2.16–2.73]
HTL-Cy3	2.44 [2.08–2.87]
HTL-Cy5	2.92 [1.70–5.08]
HTL-TMR-CN	3.29 [2.98–3.62]
HTL-JF$_{549}$	4.48 [3.61–5.55]
HTL-TMR-SCH3	4.94 [4.36–5.61]
HTL-JF$_{669}$	5.16 [4.59–5.79]
HTL-TMR	6.24 [5.25–7.41]
HTL-500R	6.51 [5.92–7.16]
HTL-TMR-SNH2	7.23 [6.20–8.43]
HTL-MaP555	9.36 [8.06–10.9]
HTL-TMR-az-F2	15.5 [14.8–16.3]
HTL-515R	16.7 [13.5–20.8]
HTL-5-CPY	17.6 [15.9–19.6]
HTL-JF$_{525}$	21.8 [17.9–26.5]
HTL-510R	38.4 [32.1–45.9]
HTL-JF$_{503}$	39.7 [30.1–52.0]
HTL-5-TMR	40.9 [36.9–45.4]
HTL-Oregon green	866 [697–1109]
HTL-Alexa488	992 [894–1109]
HTL-Fluorescein	1135 [736–2098]

[a] CI: confidence interval

Table S5: Affinities of methyl-amide (MeAm) fluorophores to HaloTag. For titration data and fits see figure S12.

Ligand	K_D **[95 % CI[a]] (μM)**
MeAm-6-CPY	$8.52\ [7.84–9.30] \cdot 10^2$
MeAm-6-TMR	$1.34\ [1.23–1.47] \cdot 10^3$
MeAm-5-CPY	$1.99\ [1.69–2.40] \cdot 10^3$
MeAm-5-TMR	$1.93\ [1.66–2.29] \cdot 10^3$

[a] CI: confidence interval

Table S6: Apparent second-order rate constants of cpHaloTag screening with HTL-TMR and melting temperatures in comparison to native HaloTag. For experimental data and fits see figure S13.

Protein	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**	T_m **(°C)**
HaloTag	1.88 [1.86–1.90] · 10^7	62.2
cpHaloTag 154–156	9.72 [9.33–10.1] · 10^5	42.8
cpHaloTag 141–145	2.35 [2.27–2.43] · 10^5	43.0
cpHaloTag 141–144	2.21 [2.15–2.28] · 10^5	45.2
cpHaloTag 212–213	1.89 [1.84–1.94] · 10^5	34.0
cpHaloTag 141–149	1.30 [1.27–1.33] · 10^5	38.3
cpHaloTag 212–214	1.28 [1.24–1.31] · 10^5	33.9
cpHaloTag 178–179	2.71 [2.68–2.73] · 10^4	32.4
cpHaloTag 135–137	1.31 [1.30–1.32] · 10^4	33.7
cpHaloTag 163–165	4.94 [4.90–4.97] · 10^3	38.4
cpHaloTag 177–179	4.88 [4.82–4.95] · 10^3	32.0

[a] CI: confidence interval

Table S7: Kinetic constants (k_1, k_{-1}, k_2) for the labeling reaction of cpHaloTag variants with HTL-TMR in comparison to native HaloTag. Data was analyzed with a two-step reaction model. For experimental data and fits see figure S14.

Ligand	k_1 **[95 % CI[a]] (M^{-1} s^{-1})**	k_{-1} **[95 % CI[a]] (s^{-1})**	k_2 **[95 % CI[a]] (s^{-1})**	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**
HaloTag	7.84 [6.59–9.09] · 10^7	25.6 [19.3–31.8]	8.06 [7.58–8.54]	1.88 [1.86–1.90] · 10^7
cpHaloTag 154–156	3.43 [3.11–3.74] · 10^7	7.40 [6.13–8.67]	2.72 [2.47–2.98]	9.22 [8.89–9.55] · 10^6
cpHaloTag 141–145	3.02 [2.36–3.67] · 10^6	16.3 [12.2–20.3]	1.20 [1.14–1.26]	2.08 [2.06–2.09] · 10^5

[a] CI: confidence interval

Table S8: Melting temperatures of HaloTag and split-HaloTag variants.

Protein	T_m **[± std. dev.[a]]**
HT7	62.0 [± 0.06] °C
cpHaloTag 141–145	44.7 [± 0.04] °C
cpHaloTag 154–156	47.3 [± 0.01] °C
cpHaloΔ	31.3 [± 0.14] °C
Caprola9 + Ca^{2+}	39.9 [± 0.06] °C
Caprola9 + EGTA	42.5 [± 0.02] °C
M13–cpHaloΔ	39.9 [± 0.02] °C

[a] std. dev.: standard deviation

Table S9: Apparent second-order rate constants of Caprola9 labeling with HTL-TMR at different pH values. For experimental data and fits see figure S20.

Protein	k_{app} **[95 % CI[a]] (M^{-1} s^{-1})**
pH 6.7	1.55 [1.54–1.57] · 10^5
pH 7.0	1.75 [1.73–1.77] · 10^5
pH 7.3	1.94 [1.91–1.96] · 10^5
pH 7.6	1.96 [1.94–1.99] · 10^5
pH 7.9	2.25 [2.22–2.28] · 10^5
pH 8.2	2.18 [2.15–2.21] · 10^5

[a] CI: confidence interval

Table S10: Biochemical characterization of Caprola$_1$–Caprola$_{15}$. For experimental data and fits see figures S21 and S22

Variant	k_{Ca2+} [95 % CI[a]] (M^{-1} s^{-1})	k_{EGTA} [95 % CI[a]] (M^{-1} s^{-1})	k_{Ca2+}/k_{EGTA} [95 % CI[a]]	EC_{50} [95 % CI[a]] (nM)	$Hill$[b] [95 % CI[a]]
Caprola$_1$	1.88 [1.77–1.98] · 10^5	41.6 [41.1–42.2]	4.51 [4.26–4.77] · 10^3	43.7 [31.8–63.6]	4.09 [1.52–8.49]
Caprola$_2$	8.74 [8.23–9.31] · 10^4	44.2 [43.7–44.6]	1.98 [1.86–2.10] · 10^3	68.4 [29.3–185]	2.82 [0.46–8.29]
Caprola$_3$	1.50 [1.40–1.60] · 10^5	35.9 [35.5–36.3]	4.17 [3.91–4.47] · 10^3	70.2 [52.0–94.0]	4.62 [1.95–9.98]
Caprola$_4$	1.55 [1.45–1.64] · 10^5	31.4 [31.0–31.8]	4.92 [4.62–5.24] · 10^3	120 [80.2–167]	4.61 [1.87–10.2]
Caprola$_5$	7.26 [6.95–7.55] · 10^4	37.2 [36.6–37.8]	1.95 [1.86–2.04] · 10^3	130 [116–147]	5.92 [3.94–9.73]
Caprola$_6$	1.22 [1.13–1.32] · 10^5	38.3 [37.8–38.8]	3.19 [2.93–3.46] · 10^3	130 [117–147]	4.79 [3.51–6.92]
Caprola$_7$	1.16 [1.08–1.23] · 10^5	47.5 [46.8–48.1]	2.44 [2.28–2.60] · 10^3	172 [151–194]	3.31 [2.58–4.25]
Caprola$_8$	8.39 [7.94–8.89] · 10^4	75.3 [74.6–76.0]	1.11 [1.05–1.18] · 10^3	173 [144–207]	5.65 [2.75–10.4]
Caprola$_9$	2.61 [2.46–2.78] · 10^5	38.0 [37.4–38.5]	6.87 [6.46–7.33] · 10^3	243 [185–319]	2.64 [1.39–5.12]
Caprola$_{10}$	1.33 [1.24–1.41] · 10^5	23.7 [23.3–24.1]	5.60 [5.23–5.97] · 10^3	252 [209–312]	4.30 [2.23–8.32]
Caprola$_{11}$	1.99 [1.89–2.10] · 10^5	28.5 [28.2–28.9]	6.97 [6.60–7.39] · 10^3	278 [227–344]	3.62 [2.40–5.23]
Caprola$_{12}$	4.77 [4.38–5.21] · 10^5	48.1 [47.5–48.7]	9.93 [9.08–10.9] · 10^3	336 [311–363]	2.98 [2.44–3.63]
Caprola$_{13}$	3.58 [3.35–3.81] · 10^4	70.8 [69.0–72.7]	5.05 [4.71–5.43] · 10^2	343 [244–482]	3.43 [1.39–7.48]
Caprola$_{14}$	1.81 [1.72–1.91] · 10^5	40.5 [40.1–40.9]	4.47 [4.24–4.71] · 10^3	499 [414–607]	3.69 [2.06–6.73]
Caprola$_{15}$	8.33 [7.97–8.68] · 10^4	16.2 [15.8–16.5]	5.15 [4.91–5.40] · 10^3	1050 [847–1299]	2.51 [1.69–3.85]

[a] CI: confidence interval, [b] Hill: Hill coefficient

Table S11: Screening results of the 384 Hpep library. cpHaloΔ Labeling activities in presence of Hpep variants are given relative to the native peptide (ARETFQAFRT). The table is continued on the next pages.

Sequence	Rel. act.[a]	Sequence	Rel. act.[a]	Sequence	Rel. act.[a]
AIRTFILFRS	< 0.01	VIETFIQFRS	0.63	EKEEARETFQAFRT	1.2
AIRAFIKFRS	< 0.01	SREAFEAYRS	0.63	VIKAFIKFRE	1.2
WIETFKLFRS	< 0.01	QKEEARETFQAFRT	0.64	LAEEARETFQAFRT	1.23
SIRAFILFRS	< 0.01	WKEEARETFQAFRT	0.65	VVRTFKLFRS	1.24
ARKTFILFRS	< 0.01	AVKTFILFRD	0.65	FWEEARETFQAFRT	1.27
AIRTFILFRQ	< 0.01	TRDTFKAFQS	0.65	AQETFKAFRS	1.28
ARKAFILFRS	0.06	ARETFQAFRTAKEK	0.65	ARDTFKAFRS	1.31
ARKTFILFRQ	0.12	FEEEARETFQAFRT	0.66	SKEDARETFQAFRT	1.31
VRKTFKLFRS	0.18	AIRAFQLYRS	0.66	IREAFEAYRS	1.31
IRDTFQAFQS	0.21	ARETFEAFRS	0.67	IREAFEAFRS	1.32
ARDTFKAFQS	0.28	YAEEARETFQAFRT	0.72	SSEDARETFQAFRT	1.33
IRETFEAYQS	0.34	SAEEARETFQAFRT	0.72	ARETFQAFRTAHQK	1.34
ARETFQAFQT	0.36	TRDAFRAYQS	0.73	VIKAFQLFRS	1.36
ARETFQAFMT	0.39	SRDAFKAFQS	0.74	YKEDLIDTFRKFRQ	1.36
GRDTFQAFQS	0.39	AQETFQAFRT	0.76	VRDTFILFRS	1.38
IRETFEAFQS	0.39	SRDAFQAYQS	0.79	ARDTFILFRE	1.42
SRDAFQAFQS	0.42	ARETFQAFRTTSDS	0.79	VVETFKLFRS	1.45
TRDTFKAFKS	0.43	WFEEARETFQAFRT	0.80	AIETFKLFRE	1.50
AIRFFILARD	0.44	ARETAQAFRT	0.80	SRDTFKAFRS	1.51
AIRAFILFRE	0.44	FAEEARETFQAFRT	0.83	VRTAFKLFRQ	1.51
SREAFEAFQS	0.44	ARETFEAFRT	0.84	WGKEARETFQAFRT	1.51
GRDAFKAFQS	0.45	AVETFQAFRT	0.86	VIRAFQLFRQ	1.54
IRDTMKAFQN	0.45	TGEEARETFQAFRT	0.86	LIEAFRQFRS	1.56
ARETFQAFRTPPEG	0.48	ARETFQAFRE	0.88	KDEEARETFQAFRT	1.60
ARETFKAFQS	0.48	AVKTFILFRE	0.90	AREAFQAFRT	1.61
TREAFEAFQS	0.50	ARETFQAFRTAKDS	0.91	AIKTFILFRE	1.64
IRDTFQAFRS	0.50	QEEEARETFQAFRT	0.93	YKEELIKTFRLFMQ	1.7
VVKAFIKFRD	0.50	ARETFQAFRTGAES	0.94	IRETFEAFRS	1.71
ARETFQAART	0.51	YSEDARETFQAFRT	0.94	VRETFKLFRS	1.71
AIRTFILFRE	0.52	TRETFKAFRD	0.96	NRDTFQAFRS	1.76
YSEIARETFQAFRT	0.52	SRKAFELFRQ	0.96	WAKEARETFQAFRT	1.77
IRDTFRAFQS	0.53	LIEAFIQFRS	0.97	VREFFKLARD	1.81
SRDAFQAFQT	0.53	ARETFQAFRT	1.00	LIETFKLFRT	1.83
AIRAFILFRD	0.54	YGRDARETFQAFRT	1.02	SRDAFQAYRS	1.86
VRETFEAFRS	0.54	AIRAFQLFRQ	1.03	VIRTFQLFRQ	1.88
TGEIARETFQAFRT	0.55	GRDTFKAFQS	1.03	SRDTFQAFQS	1.89
TRDTFRAFQS	0.55	MRDEARETFQAFRT	1.04	STEIARETFQAFRT	1.92
IRDTFRAFKS	0.56	QKEDARETFQAFRT	1.05	MGEEARETFQAFRT	1.95
VRKTFILFRD	0.58	ARKTFQAFRT	1.13	VRKTFKLFRQ	1.97
AIETFQAFRT	0.58	LIKTFILFRD	1.15	VRDTFQAFRS	1.98
ARETFQAFR	0.58	ARETFQAFRTGASS	1.17	WTKEARETFQAFRT	2.01
LIETFIQFRS	0.60	KKEDARETFQAFRT	1.18	WAEEARETFQAFRT	2.02
FGEDARETFQAFRT	0.60	GRDTFQAFRS	1.18	AIETFKLFRQ	2.14

[a] Rel. act.: Relative labeling activity

Table S11: continued

Sequence	Rel. act.[a]	Sequence	Rel. act.[a]	Sequence	Rel. act.[a]
KSEEARETFQAFRT	2.15	GRDTFKAFRS	4.22	LIETFKLFRE	18.12
QKDEARETFQAFRT	2.26	LIETFKKFRD	4.25	FREEARETFQAFRT	18.20
AIEAFKLFRQ	2.27	YKEEARETFQAFRT	4.45	ARETFRAFRT	18.37
YTEEARETFQAFRT	2.30	MGEDARETFQAFRT	4.65	DAEDARETFQAFRT	18.45
SRDTFKAFQS	2.35	VRDTFQAFRT	4.82	WLEEARETFQAFRT	18.52
VRKTFELFRS	2.46	FSDEARETFQAFRT	4.89	FGEEARETFQAFRT	18.53
MRETFQAFRT	2.46	FKEIARETFQAFRT	4.90	WDEIARETFQAFRT	18.78
AVEMFKLFRD	2.51	YGEIARETFQAFRT	4.90	FTEEARETFQAFRT	19.13
WWEEARETFQAFRT	2.70	WREIARETFQAFRT	4.92	LRETFQAFRT	20.58
SRDTFKAFQT	2.73	YKEDARETFQAFRT	4.96	WTEEARETFQAFRT	20.71
YNEEARETFQAFRT	2.74	ARDTFQAFRT	6.39	STEEARETFQAFRT	21.15
KEEEARETFQAFRT	2.80	AIEAFKLFRE	7.82	LIEAFKLFRD	21.29
TRETFEAFQS	2.82	TRETFQAFRT	8.35	EKEDARETFQAFRT	21.39
AIKAFQLFRE	2.86	YQEEARETFQAFRT	8.63	ARETFKAFRS	21.42
AREAFEAFRS	2.87	SRETFEAFRS	9.16	FAEDARETFQAFRT	22.18
SKEEARETFQAFRT	3.04	WTDEARETFQAFRT	9.44	ARETFQAYRT	22.28
WQEEARETFQAFRT	3.07	ARETFKAFRD	9.45	WNEEARETFQAFRT	23.71
TRETFEAFRS	3.11	WADEARETFQAFRT	9.69	QKDVARETFQAFRT	23.94
NGEDARETFQAFRT	3.11	GRETFQAFRT	10.24	FSEEARETFQAFRT	24.49
FKDDARETFQAFRT	3.20	LIETFIQYRQ	10.35	WSEDARETFQAFRT	24.96
WSETARETFQAFRT	3.20	VIETFKLFRD	10.55	LIDTFRKFRS	25.18
SSEEARETFQAFRT	3.23	FKEEARETFQAFRT	11.30	WDEEARETFQAFRT	25.36
LIETFKLFRQ	3.26	DGDDARETFQAFRT	11.46	WKDEVIDAFRKFRE	25.88
ARETFQAFRD	3.40	DKEEARETFQAFRT	11.55	VRETFQAFRT	26.65
WRDEARETFQAFRT	3.47	WEDEARETFQAFRT	11.57	WKEIARETFQAFRT	27.26
WSDEARETFQAFRT	3.51	ARETFQAFRTALKD	11.63	AREFFQAFRT	27.40
WGQEARETFQAFRT	3.52	AKETFQAFRT	12.29	YWEEARETFQAFRT	28.73
WKDEARETFQAFRT	3.57	DGEEARETFQAFRT	12.44	IRDTFQAYRS	29.49
ARETFIAFRT	3.57	WFDEARETFQAFRT	12.73	YGEDARETFQAFRT	29.80
WREEARETFQAFRT	3.65	WKEEVVKAFIKFRD	12.73	KKEEARETFQAFRT	30.01
ARETFKAFRT	3.69	TREAFEAFRS	12.80	WGEDARETFQAFRT	31.15
ARETFQAFRTIFEH	3.77	FQEEARETFQAFRT	13.37	ARRTFQAFRT	33.23
WSEMARETFQAFRT	3.77	QAEEARETFQAFRT	13.55	WGEEARETFQAFRT	33.52
ARDTFRAFRD	3.79	TGEDARETFQAFRT	14.29	WGEIARETFQAFRT	34.30
LIETFRQFRE	3.87	QGEEARETFQAFRT	14.38	LKDIARETFQAFRT	36.26
DKEIARETFQAFRT	3.90	FGDEARETFQAFRT	15.25	FKEDARETFQAFRT	36.39
LIKTFQLFRD	3.95	PKEDARETFQAFRT	15.57	VIRMFQLFRE	36.43
ARETFQAFRQ	3.97	ESEDARETFQAFRT	15.75	KDETARETFQAFRT	37.37
AQETFKAFRD	4.01	YGEEARETFQAFRT	15.99	ARETFQLFRT	37.74
SRETFILFRQ	4.03	DGEIARETFQAFRT	16.29	YKEIARETFQAFRT	38.27
YEEEARETFQAFRT	4.13	VIKAFQLFRE	17.13	RKEEARETFQAFRT	39.66
YAEIARETFQAFRT	4.15	YREEARETFQAFRT	17.82	MREDARETFQAFRT	40.68
WKEEVIKAFILYRE	4.19	WKDDARETFQAFRT	18.10	SRETFQAFRT	41.71

[a] Rel. act.: Relative labeling activity

Table S11: continued

Sequence	Rel. act.[a]	Sequence	Rel. act.[a]	Sequence	Rel. act.[a]
TRDTFQAFRT	43.34	WYEEARETFQAFRT	88.83	DGRDARETFQAFRT	308.00
SRDEARETFQAFRT	43.98	WKEEVIKAFILYRS	89.82	WKKEARETFQAFRT	309.63
WEEEARETFQAFRT	44.20	QKKEARETFQAFRT	91.08	SRDTFQAFRS	309.89
AIRTFQLFRQ	44.92	IRETFQAFRT	92.72	WKEEVIKAFKLFRS	322.27
SKEIARETFQAFRT	47.40	ARETFQKFRT	94.09	YKEDLIKTFILFRQ	348.07
FSEDARETFQAFRT	47.58	GRETFQAFRS	96.35	LGEEARETFQAFRT	350.13
MSEEARETFQAFRT	47.86	WGRDARETFQAFRT	96.68	YKDDLIKTFILFRQ	360.10
WSEEARETFQAFRT	47.92	HKEEARETFQAFRT	97.93	WRKEARETFQAFRT	361.66
YGEELIKMFIKFRD	48.04	LIETFRLFRQ	106.52	ARDMFIKFRD	367.76
LGEDARETFQAFRT	48.10	LAEDARETFQAFRT	118.39	FKQDARETFQAFRT	376.88
LKEDARETFQAFRT	48.49	VIKTFQLFRS	119.52	WDREARETFQAFRT	380.13
TREAFKAFRA	48.88	WKEDARETFQAFRT	119.99	AIKAFQLFRS	381.00
SGEDARETFQAFRT	50.85	SREAFEAFRS	124.17	IRDAFQAFRS	394.10
WAQEARETFQAFRT	51.44	VREAFKLFRQ	128.90	AQEAFKAFRS	405.63
ARDTFQAFRS	52.15	WKEEVIKAFIKFRE	134.28	WKETARETFQAFRT	409.06
LKEEARETFQAFRT	52.52	WKELARETFQAFRT	135.29	LIKTFQLFRQ	419.30
WDEDARETFQAFRT	53.04	FGKEARETFQAFRT	138.74	WEETARETFQAFRT	441.21
FIETFKLFRD	53.08	ARETFQAFRS	144.80	LIETFKLFRS	444.81
DKEDARETFQAFRT	53.45	VVKTFQLFRS	156.13	WRETARETFQAFRT	453.59
AVETFKLFRQ	54.13	MKEDARETFQAFRT	163.83	LRKTFKLFRS	476.55
WHEEARETFQAFRT	54.83	LGDDARETFQAFRT	165.98	QERDARETFQAFRT	515.29
MKEEARETFQAFRT	57.76	WKEDVIDAFRKFRD	172.46	AREMFQAFRT	519.10
ARDTFRAFRS	58.16	LVETFKLFRS	176.98	DGEDARETFQAFRT	536.24
WDRDARETFQAFRT	58.70	WKEEVIKAFRLYRE	177.93	ARLFFQLFRT	554.77
AQETFKAYRS	60.80	WEKEARETFQAFRT	184.16	DKETARETFQAFRT	567.46
QREEARETFQAFRT	61.04	TRDTFQAFQS	192.59	AREAFKAFRS	592.67
WKEEVIKAFILYRQ	61.60	SRDAFKAYRS	194.89	VREAFKLFRS	599.08
SRDAFQAFRS	64.06	VRTAFKLFRS	197.26	WKEKARETFQAFRT	697.62
VIKTFQLFRQ	66.44	RDEDARETFQAFRT	203.92	RKKEARETFQAFRT	711.89
SRDTFQAFRT	66.56	YSEEARETFQAFRT	208.67	SKRDARETFQAFRT	808.46
TRETFKAFRS	67.28	LIKTFQLYRS	217.70	WREEVRKTFKLFRS	828.94
HKDEARETFQAFRT	67.42	SRDAFKAFRS	222.76	WDKEARETFQAFRT	887.88
SRDTFRAFRS	68.65	ARKAFELFRS	223.61	WREEVRKAFKLFRS	930.10
ARETFQAYRS	68.70	LIKTFQLFRS	243.50	WSKEARETFQAFRT	943.09
AIKTFQLFRS	71.38	WKEEVIKAFRLFRE	253.66	WKREARETFQAFRT	962.45
WGDEARETFQAFRT	72.85	WDEEVKDAFKKFRD	259.31	LSEEARETFQAFRT	976.73
YKDDARETFQAFRT	78.52	WREEVIKAFKLFRQ	265.38	WREEVRKTFKLFRQ	989.97
TRDTFQAFRS	78.67	SKKEARETFQAFRT	268.50	AIRMFQLFRE	1038.03
WKEEVIKAFRLYRD	79.07	LIKAFQLFRS	276.56	WREEVRKAFKLFRQ	1189.69
FGREARETFQAFRT	79.77	WKEEVIKAFKLFRD	285.96	WIETFKLYRE	1333.05
VIKAFQLFRQ	85.12	WKEEVIKAFKLFRQ	293.44	WREEVRKAFKLFRD	1525.72
VIETFKLFRS	88.76	WRETFQAFRT	294.43	WKRDARETFQAFRT	1641.32

[a] Rel. act.: Relative labeling activity

Table S12: Biochemical characterization of Hpep variants. Half maximal effective concentrations (EC_{50}) and maximal apparent split-HaloTag labeling rate constants (k_{app}^{max}) are listed. For experimental data and fits see figures S27 and S28.

Hpep	Sequence	EC_{50} [95 % CI[a]]	k_{app}^{max} [95 % CI[a]] (M^{-1} s^{-1})
Hpep01	ARETFQAFRT	2979 [2605–3442] µM	1.33 [1.29–1.38] · 10^4
Hpep02	AREMFQAFRT	1116 [933–1356] µM	1.29 [1.14–1.47] · 10^4
Hpep03	LSEEAREMFQAFRT	483 [349–717] µM	6.54 [5.57–7.86] · 10^3
Hpep04	VREAFKLFRS	473 [387–584] µM	1.22 [1.11–1.34] · 10^4
Hpep05	SKRDAREMFQAFRT	149 [125–174] µM	7.53 [7.08–7.99] · 10^5
Hpep06	AREMFQLFRT	56.6 [39.1–79.6] µM	3.86 [3.40–4.34] · 10^3
Hpep07	YFQGAREMFQAFRT	35.6 [27.7–46.0] µM	4.05 [3.72–4.40] · 10^4
Hpep08	WKEEVIKAFKLFRD	21.0 [18.3–23.9] µM	4.44 [4.26–4.61] · 10^3
Hpep09	WKRDAREMFQAFRT	19.3 [15.7–24.2] µM	1.80 [1.66–1.96] · 10^5
Hpep10	WREEVRKTFKLFRQ	16.6 [11.1–23.6] µM	3.45 [3.12–3.81] · 10^4
Hpep11	WREEVRKAFKLFRQ	10.7 [8.9–13.0] µM	3.02 [2.89–3.15] · 10^4
Hpep12	WREEVRKAFKLFRS	5.12 [3.86–6.67] µM	1.65 [1.55–1.76] · 10^4
Hpep13	WRETFQLFRT	2.39 [2.10–2.75] µM	8.50 [8.24–8.78] · 10^2
Hpep14	WREMFRLFRTGRVQ	1.80 [1.68–1.94] µM	1.59 [1.54–1.65] · 10^4
Hpep15	WREMFQAFRT	1.70 [1.12–2.57] µM	8.05 [6.80–9.55] · 10^3
Hpep16	WREMFRLFRTGQRS	1.20 [1.15–1.25] µM	3.51 [3.45–3.59] · 10^4
Hpep17	WREMFQLFRT	730 [624–865] nM	1.92 [1.82–2.02] · 10^4
Hpep18	WREMFRLFRT	350 [326–376] nM	6.73 [6.63–6.82] · 10^4
Hpep19	SKRDWREMFRLFRT	235 [225–245] nM	7.27 [7.17–7.37] · 10^4
Hpep20	RVMSWREMFRLFRT	169 [141–199] nM	6.58 [6.26–6.90] · 10^4
Hpep21	RMWSWREMFRLFRT	163 [126–208] nM	6.93 [6.44–7.42] · 10^4
Hpep22	WKRDWREMFRLFRT	124 [111–138] nM	5.14 [5.00–5.30] · 10^4
Hpep23	RQWTWREMFRLFRT	104 [90–121] nM	6.70 [6.44–6.95] · 10^4
Hpep24	RGWTWREMFRLFRT	65.9 [47.0–91.2] nM	5.70 [5.20–6.28] · 10^4
Hpep25	RMWTWREMFRLFRT	46.8 [30.3–69.9] nM	5.29 [4.76–5.87] · 10^4
Hpep26	RQWSWREMFRLFRT	45.2 [38.6–52.8] nM	4.27 [4.13–4.41] · 10^4
Hpep27	RGWSWREMFRLFRT	40.3 [31.5–50.7] nM	3.60 [3.42–3.81] · 10^4

[a] CI: confidence interval

Table S13: Biochemical characterization of selected Hpep variants with cpHaloΔ2. Half maximal effective concentrations (EC_{50}) and maximal apparent split-HaloTag labeling rate constants (k_{app}^{max}) are listed. For experimental data and fits see figures S33 and S34.

Hpep	Sequence	EC_{50} [95 % CI[a]]	k_{app}^{max} [95 % CI[a]] (M^{-1} s^{-1})
Hpep05	SKRDAREMFQAFRT	12.7 [10.9–14.8] µM	5.07 [4.88–5.28] · 10^5
Hpep07	YFQGAREMFQAFRT	8.05 [7.14–9.12] µM	1.15 [1.11–1.18] · 10^5
Hpep10	WREEVRKTFKLFRQ	2.16 [1.93–2.41] µM	1.06 [1.03–1.08] · 10^5

[a] CI: confidence interval

Protein Sequences

HaloTag variants and split-HaloTag

Color code:

His-tag TEVp cleavage site Strep-tag enterokinase cleavage site linker

HaloTag7

```
  1 MHHHHHHHHH HENLYFQGIG TGFPFDPHYV EVLGERMHYV DVGPRDGTPV LFLHGNPTSS YVWRNIIPHV
 71 APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL GLEEVVLVIH DWGSALGFHW AKRNPERVKG
141 IAFMEFIRPI PTWDEWPEFA RETFQAFRTT DVGRKLIIDQ NVFIEGTLPM GVVRPLTEVE MDHYREPFLN
211 PVDREPLWRF PNELPIAGEP ANIVALVEEY MDWLHQSPVP KLLFWGTPGV LIPPAEAARL AKSLPNCKAV
281 DIGPGLNLLQ EDNPDLIGSE IARWLSTLEI
```

HOB

```
  1 MHHHHHHHHH HENLYFQGIG TGFPFDPHYV EVLGERMHYV DVGPRDGTPV LFLHGNPTSS YVWRNIIPHV
 71 APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL GLEEVVLVIH DWGSALGFHW AKRNPERVKG
141 IAFMEFIRPI PTWDEWPKFA RKTFQAFRTK KVGRKLIIDQ NVFIEGTLPM GVVRPLTEVE MDHYREPFLN
211 PVDREPLWRF PNELPIAGEP ANIVALVEEY MDWLHQSPVP KLLFWGTPGV LIPPAEAARL AKSLPNCKAV
281 DIGPGLNLLQ EDNPDLIGSE IARWLSTLEI SG
```

cpHaloTag 135-137

```
  1 MASWSHPQFE KGADDDDKVP HTWDEWPEFA RETFQAFRTT DVGRKLIIDQ NVFIEGTLPM GVVRPLTEVE
 71 MDHYREPFLN PVDREPLWRF PNELPIAGEP ANIVALVEEY MDWLHQSPVP KLLFWGTPGV LIPPAEAARL
141 AKSLPNCKAV DIGPGLNLLQ EDNPDLIGSE IARWLSTLEI GGTGGSGGTG GSGGSIGTGF PFDPHYVEVL
211 GERMHYVDVG PRDGTPVLFL HGNPTSSYVW RNIIPHVAPT HRCIAPDLIG MGKSDKPDLG YFFDDHVRFM
281 DAFIEALGLE EVVLVIHDWG SALGFHWAKR NPERVKGIAF MEFIRPIAPG FSSISAHHHH HHHHHH
```

cpHaloTag 141-144

```
  1 MASWSHPQFE KGADDDDKVP HFARETFQAF RTTDVGRKLI IDQNVFIEGT LPMGVVRPLT EVEMDHYREP
 71 FLNPVDREPL WRFPNELPIA GEPANIVALV EEYMDWLHQS PVPKLLFWGT PGVLIPPAEA ARLAKSLPNC
141 KAVDIGPGLN LLQEDNPDLI GSEIARWLST LEIGGTGGSG GTGGSGGSIG TGFPFDPHYV EVLGERMHYV
211 DVGPRDGTPV LFLHGNPTSS YVWRNIIPHV APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL
281 GLEEVVLVIH DWGSALGFHW AKRNPERVKG IAFMEFIRPI PTWDEWAPGF SSISAHHHHH HHHHH
```

cpHaloTag 141-145

```
  1 MASWSHPQFE KGADDDDKVP HARETFQAFR TTDVGRKLII DQNVFIEGTL PMGVVRPLTE VEMDHYREPF
 71 LNPVDREPLW RFPNELPIAG EPANIVALVE EYMDWLHQSP VPKLLFWGTP GVLIPPAEAA RLAKSLPNCK
141 AVDIGPGLNL LQEDNPDLIG SEIARWLSTL EIGGTGGSGG TGGSGGSIGT GFPFDPHYVE VLGERMHYVD
211 VGPRDGTPVL FLHGNPTSSY VWRNIIPHVA PTHRCIAPDL IGMGKSDKPD LGYFFDDHVR FMDAFIEALG
281 LEEVVLVIHD WGSALGFHWA KRNPERVKGI AFMEFIRPIP TWDEWAPGFS SISAHHHHHH HHHH
```

cpHaloTag 141-149

```
  1 MASWSHPQFE KGADDDDKVP HFQAFRTTDV GRKLIIDQNV FIEGTLPMGV VRPLTEVEMD HYREPFLNPV
 71 DREPLWRFPN ELPIAGEPAN IVALVEEYMD WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI
141 GPGLNLLQED NPDLIGSEIA RWLSTLEIGG TGGSGGTGGS GGSIGTGFPF DPHYVEVLGE RMHYVDVGPR
211 DGTPVLFLHG NPTSSYVWRN IIPHVAPTHR CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV
281 VLVIHDWGSA LGFHWAKRNP ERVKGIAFME FIRPIPTWDE WAPGFSSISA HHHHHHHHHH
```

cpHaloTag 154-156

```
  1 MASWSHPQFE KGADDDDKVP HDVGRKLIID QNVFIEGTLP MGVVRPLTEV EMDHYREPFL NPVDREPLWR
 71 FPNELPIAGE PANIVALVEE YMDWLHQSPV PKLLFWGTPG VLIPPAEAAR LAKSLPNCKA VDIGPGLNLL
141 QEDNPDLIGS EIARWLSTLE IGGTGGSGGT GGSGGSIGTG FPFDPHYVEV LGERMHYVDV GPRDGTPVLF
211 LHGNPTSSYV WRNIIPHVAP THRCIAPDLI GMGKSDKPDL GYFFDDHVRF MDAFIEALGL EEVVLVIHDW
281 GSALGFHWAK RNPERVKGIA FMEFIRPIPT WDEWPEFARE TFQAFRTAPG FSSISAHHHH HHHHHH
```

cpHaloTag 163-165

```
  1 MASWSHPQFE KGADDDDKVP HQNVFIEGTL PMGVVRPLTE VEMDHYREPF LNPVDREPLW RFPNELPIAG
 71 EPANIVALVE EYMDWLHQSP VPKLLFWGTP GVLIPPAEAA RLAKSLPNCK AVDIGPGLNL LQEDNPDLIG
141 SEIARWLSTL EIGGTGGSGG TGGSGGSIGT GFPFDPHYVE VLGERMHYVD VGPRDGTPVL FLHGNPTSSY
211 VWRNIIPHVA PTHRCIAPDL IGMGKSDKPD LGYFFDDHVR FMDAFIEALG LEEVVLVIHD WGSALGFHWA
281 KRNPERVKGI AFMEFIRPIP TWDEWPEFAR ETFQAFRTTD VGRKLIIAPG FSSISAHHHH HHHHHH
```

cpHaloTag 177-179

```
  1 MASWSHPQFE KGADDDDKVP HRPLTEVEMD HYREPFLNPV DREPLWRFPN ELPIAGEPAN IVALVEEYMD
 71 WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI GPGLNLLQED NPDLIGSEIA RWLSTLEIGG
141 TGGSGGTGGS GGSIGTGFPF DPHYVEVLGE RMHYVDVGPR DGTPVLFLHG NPTSSYVWRN IIPHVAPTHR
211 CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV VLVIHDWGSA LGFHWAKRNP ERVKGIAFME
281 FIRPIPTWDE WPEFARETFQ AFRTTDVGRK LIIDQNVFIE GTLPMGVAPG FSSISAHHHH HHHHHH
```

cpHaloTag 178-179

```
  1 MASWSHPQFE KGADDDDKVP HRPLTEVEMD HYREPFLNPV DREPLWRFPN ELPIAGEPAN IVALVEEYMD
 71 WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI GPGLNLLQED NPDLIGSEIA RWLSTLEIGG
141 TGGSGGTGGS GGSIGTGFPF DPHYVEVLGE RMHYVDVGPR DGTPVLFLHG NPTSSYVWRN IIPHVAPTHR
211 CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV VLVIHDWGSA LGFHWAKRNP ERVKGIAFME
281 FIRPIPTWDE WPEFARETFQ AFRTTDVGRK LIIDQNVFIE GTLPMGVVAP GFSSISAHHH HHHHHHH
```

cpHaloTag 212-213

```
  1 MASWSHPQFE KGADDDDKVP HGEPANIVAL VEEYMDWLHQ SPVPKLLFWG TPGVLIPPAE AARLAKSLPN
 71 CKAVDIGPGL NLLQEDNPDL IGSEIARWLS TLEIGGTGGS GGTGGSGGSI GTGFPFDPHY VEVLGERMHY
141 VDVGPRDGTP VLFLHGNPTS SYVWRNIIPH VAPTHRCIAP DLIGMGKSDK PDLGYFFDDH VRFMDAFIEA
211 LGLEEVVLVI HDWGSALGFH WAKRNPERVK GIAFMEFIRP IPTWDEWPEF ARETFQAFRT TDVGRKLIID
281 QNVFIEGTLP MGVVRPLTEV EMDHYREPFL NPVDREPLWR FPNELPIAAP GFSSISAHHH HHHHHHH
```

cpHaloTag 212-214

```
  1 MASWSHPQFE KGADDDDKVP HEPANIVALV EEYMDWLHQS PVPKLLFWGT PGVLIPPAEA ARLAKSLPNC
 71 KAVDIGPGLN LLQEDNPDLI GSEIARWLST LEIGGTGGSG GTGGSGGSIG TGFPFDPHYV EVLGERMHYV
141 DVGPRDGTPV LFLHGNPTSS YVWRNIIPHV APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL
211 GLEEVVLVIH DWGSALGFHW AKRNPERVKG IAFMEFIRPI PTWDEWPEFA RETFQAFRTT DVGRKLIIDQ
281 NVFIEGTLPM GVVRPLTEVE MDHYREPFLN PVDREPLWRF PNELPIAAPG FSSISAHHHH HHHHHH
```

His-TEV-cpHaloTag 154-156

```
  1 MHHHHHHHHH HENLYFQGGG DVGRKLIIDQ NVFIEGTLPM GVVRPLTEVE MDHYREPFLN PVDREPLWRF
 71 PNELPIAGEP ANIVALVEEY MDWLHQSPVP KLLFWGTPGV LIPPAEAARL AKSLPNCKAV DIGPGLNLLQ
141 EDNPDLIGSE IARWLSTLEI GGTGGSGGTG GSGGSIGTGF PFDPHYVEVL GERMHYVDVG PRDGTPVLFL
211 HGNPTSSYVW RNIIPHVAPT HRCIAPDLIG MGKSDKPDLG YFFDDHVRFM DAFIEALGLE EVVLVIHDWG
281 SALGFHWAKR NPERVKGIAF MEFIRPIPTW DEWPEFARET FQAFRT
```

His-TEV-cpHaloTag 141-145

```
  1 MHHHHHHHHH HENLYFQGGG ARETFQAFRT TDVGRKLIID QNVFIEGTLP MGVVRPLTEV EMDHYREPFL
 71 NPVDREPLWR FPNELPIAGE PANIVALVEE YMDWLHQSPV PKLLFWGTPG VLIPPAEAAR LAKSLPNCKA
141 VDIGPGLNLL QEDNPDLIGS EIARWLSTLE IGGTGGSGGT GGSGGSIGTG FPFDPHYVEV LGERMHYVDV
211 GPRDGTPVLF LHGNPTSSYV WRNIIPHVAP THRCIAPDLI GMGKSDKPDL GYFFDDHVRF MDAFIEALGL
281 EEVVLVIHDW GSALGFHWAK RNPERVKGIA FMEFIRPIPT WDEW
```

cpHaloΔ

```
  1 MHHHHHHHHH HENLYFQGDV GRKLIIDQNV FIEGTLPMGV VRPLTEVEMD HYREPFLNPV DREPLWRFPN
 71 ELPIAGEPAN IVALVEEYMD WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI GPGLNLLQED
141 NPDLIGSEIA RWLSTLEIGG TGGSGGTGGS GGSIGTGFPF DPHYVEVLGE RMHYVDVGPR DGTPVLFLHG
211 NPTSSYVWRN IIPHVAPTHR CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV VLVIHDWGSA
281 LGFHWAKRNP ERVKGIAFME FIRPI
```

cpHaloΔ2

```
  1 MHHHHHHHHH HENLYFQGDV GRKLIIDQNV FIEGTLPMGV VRPLTEEEMD HYREPFLNPK DREPLWRFPN
 71 ELPIAGEPAN IVALVEEYMD WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI GPGLNLLQED
141 NPDLIGSEIA RWLSTLEIKS KYDRDQILKI IAELEKKTGG SIGTGFPFDP HYVEVLGSRM HYVDVGPRDG
211 TPVLFLHGNP TSSYVWRNII PHVAPTHRCI APDLIGMGKS DKPDLGYFFD DHVRFMDAFI EALGLEEVVL
281 VIHDWGSALG FHWAKRHPER VKGIAFMEFI RPIPTWDEW
```

Hpep variants

Variant	Sequence
Hpep01	`ARETFQAFRT`
Hpep02	`AREMFQAFRT`
Hpep03	`LSEEAREMFQAFRT`
Hpep04	`VREAFKLFRS`
Hpep05	`SKRDAREMFQAFRT`
Hpep06	`AREMFQLFRT`
Hpep07	`YFQGAREMFQAFRT`
Hpep08	`WKEEVIKAFKLFRD`
Hpep09	`WKRDAREMFQAFRT`
Hpep10	`WREEVRKTFKLFRQ`
Hpep11	`WREEVRKAFKLFRQ`
Hpep12	`WREEVRKAFKLFRS`
Hpep13	`WRETFQLFRT`
Hpep14	`WREMFRLFRTGRVQ`
Hpep15	`WREMFQAFRT`
Hpep16	`WREMFRLFRTGQRS`
Hpep17	`WREMFQLFRT`
Hpep18	`WREMFRLFRT`
Hpep19	`SKRDWREMFRLFRT`
Hpep20	`RVMSWREMFRLFRT`
Hpep21	`RMWSWREMFRLFRT`
Hpep22	`WKRDWREMFRLFRT`
Hpep23	`RQWTWREMFRLFRT`
Hpep24	`RGWTWREMFRLFRT`
Hpep25	`RMWTWREMFRLFRT`
Hpep26	`RQWSWREMFRLFRT`
Hpep27	`RGWSWREMFRLFRT`

Split-HaloTag FKBP/FRB fusions

Color code:

His-tag Strep-tag enterokinase cleavage site linker cpHaloΔ Hpep FKBP FRB

cpHaloΔ-(GGS/T)$_3$-FKBP

```
  1 MASWSHPQFE KGADDDDKVP HGGDVGRKLI IDQNVFIEGT LPMGVVRPLT EVEMDHYREP FLNPVDREPL
 71 WRFPNELPIA GEPANIVALV EEYMDWLHQS PVPKLLFWGT PGVLIPPAEA ARLAKSLPNC KAVDIGPGLN
141 LLQEDNPDLI GSEIARWLST LEIGGTGGSG GTGGSGGSIG TGFPFDPHYV EVLGERMHYV DVGPRDGTPV
211 LFLHGNPTSS YVWRNIIPHV APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL GLEEVVLVIH
281 DWGSALGFHW AKRNPERVKG IAFMEFIRPI PTWDEWGGSG GTGGSMGVQV ETISPGDGRT FPKRGQTCVV
351 HYTGMLEDGK KFDSSRDRNK PFKFMLGKQE VIRGWEEGVA QMSVGQRAKL TISPDYAYGA TGHPGIIPPH
421 ATLVFDVELL KLEAPGFSSI SAHHHHHHHH HH
```

FKBP-(GGS/T)$_3$-cpHaloΔ

```
  1 MASWSHPQFE KGADDDDKVP HMGVQVETIS PGDGRTFPKR GQTCVVHYTG MLEDGKKFDS SRDRNKPFKF
 71 MLGKQEVIRG WEEGVAQMSV GQRAKLTISP DYAYGATGHP GIIPPHATLV FDVELLKLEG GSGGTGGSDV
141 GRKLIIDQNV FIEGTLPMGV VRPLTEVEMD HYREPFLNPV DREPLWRFPN ELPIAGEPAN IVALVEEYMD
211 WLHQSPVPKL LFWGTPGVLI PPAEAARLAK SLPNCKAVDI GPGLNLLQED NPDLIGSEIA RWLSTLEIGG
281 TGGSGGTGGS GGSIGTGFPF DPHYVEVLGE RMHYVDVGPR DGTPVLFLHG NPTSSYVWRN IIPHVAPTHR
351 CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV VLVIHDWGSA LGFHWAKRNP ERVKGIAFME
421 FIRPIPTWDE WGGAPGFSSI SAHHHHHHHH HH
```

cpHaloΔ-(GGS/T)$_3$-FRB

```
  1 MASWSHPQFE KGADDDDKVP HGGDVGRKLI IDQNVFIEGT LPMGVVRPLT EVEMDHYREP FLNPVDREPL
 71 WRFPNELPIA GEPANIVALV EEYMDWLHQS PVPKLLFWGT PGVLIPPAEA ARLAKSLPNC KAVDIGPGLN
141 LLQEDNPDLI GSEIARWLST LEIGGTGGSG GTGGSGGSIG TGFPFDPHYV EVLGERMHYV DVGPRDGTPV
211 LFLHGNPTSS YVWRNIIPHV APTHRCIAPD LIGMGKSDKP DLGYFFDDHV RFMDAFIEAL GLEEVVLVIH
281 DWGSALGFHW AKRNPERVKG IAFMEFIRPI PTWDEWGGSG GTGGSAILWH EMWHEGLEEA SRLYFGERNV
351 KGMFEVLEPL HAMMERGPQT LKETSFNQAY GRDLMEAQEW CRKYMKSGNV KDLLQAWDLY YHVFRRISKA
421 PGFSSISAHH HHHHHHHH
```

FRB-(GGS/T)$_3$-cpHaloΔ

```
  1 MASWSHPQFE KGADDDDKVP HAILWHEMWH EGLEEASRLY FGERNVKGMF EVLEPLHAMM ERGPQTLKET
 71 SFNQAYGRDL MEAQEWCRKY MKSGNVKDLL QAWDLYYHVF RRISKGGSGG TGGSDVGRKL IIDQNVFIEG
141 TLPMGVVRPL TEVEMDHYRE PFLNPVDREP LWRFPNELPI AGEPANIVAL VEEYMDWLHQ SPVPKLLFWG
211 TPGVLIPPAE AARLAKSLPN CKAVDIGPGL NLLQEDNPDL IGSEIARWLS TLEIGGTGGS GGTGGSGGSI
281 GTGFPFDPHY VEVLGERMHY VDVGPRDGTP VLFLHGNPTS SYVWRNIIPH VAPTHRCIAP DLIGMGKSDK
351 PDLGYFFDDH VRFMDAFIEA LGLEEVVLVI HDWGSALGFH WAKRNPERVK GIAFMEFIRP IPTWDEWGGA
421 PGFSSISAHH HHHHHHHH
```

Hpep1-(GGS/T)$_3$-FKBP

```
  1 MASWSHPQFE KGADDDDKVP HGGARETFQA FRTGGSGGTG GSMGVQVETI SPGDGRTFPK RGQTCVVHYT
 71 GMLEDGKKFD SSRDRNKPFK FMLGKQEVIR GWEEGVAQMS VGQRAKLTIS PDYAYGATGH PGIIPPHATL
141 VFDVELLKLE APGFSSISAH HHHHHHHHH
```

FKBP-(GGS/T)$_3$-Hpep1

```
  1 MASWSHPQFE KGADDDDKVP HMGVQVETIS PGDGRTFPKR GQTCVVHYTG MLEDGKKFDS SRDRNKPFKF
 71 MLGKQEVIRG WEEGVAQMSV GQRAKLTISP DYAYGATGHP GIIPPHATLV FDVELLKLEG GSGGTGGSAR
141 ETFQAFRTGG APGFSSISAH HHHHHHHHH
```

Hpep1-(GGS/T)$_3$-FRB

```
  1 MASWSHPQFE KGADDDDKVP HGGARETFQA FRTGGSGGTG GSAILWHEMW HEGLEEASRL YFGERNVKGM
 71 FEVLEPLHAM MERGPQTLKE TSFNQAYGRD LMEAQEWCRK YMKSGNVKDL LQAWDLYYHV FRRISKAPGF
141 SSISAHHHHH HHHHH
```

FRB-(GGS/T)$_3$-Hpep1

```
  1 MASWSHPQFE KGADDDDKVP HAILWHEMWH EGLEEASRLY FGERNVKGMF EVLEPLHAMM ERGPQTLKET
 71 SFNQAYGRDL MEAQEWCRKY MKSGNVKDLL QAWDLYYHVF RRISKGGSGG TGGSARETFQ AFRTGGAPGF
141 SSISAHHHHH HHHHH
```

Split-HaloTag FKBP/FRB co-expression constructs

Color code:

Lyn11 EGFP mScarlet P2A cpHaloΔ Hpep FKBP FRB

Lyn11-EGFP-FKBP-(GGS)-cpHaloΔ-P2A-FRB-(GGS)-Hpep1-mScarlet

```
  1 MGCIKSKGKD SAGADSAGSA GMVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT YGKLTLKFIC
 71 TTGKLPVPWP TLVTTLTYGV QCFSRYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT RAEVKFEGDT
141 LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKIR HNIEDGSVQL ADHYQQNTPI
211 GDGPVLLPDN HYLSTQSALS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK GSGMGVQVET ISPGDGRTFP
281 KRGQTCVVHY TGMLEDGKKF DSSRDRNKPF KFMLGKQEVI RGWEEGVAQM SVGQRAKLTI SPDYAYGATG
351 HPGIIPPHAT LVFDVELLKL EGGSDVGRKL IIDQNVFIEG TLPMGVVRPL TEVEMDHYRE PFLNPVDREP
421 LWRFPNELPI AGEPANIVAL VEEYMDWLHQ SPVPKLLFWG TPGVLIPPAE AARLAKSLPN CKAVDIGPGL
491 NLLQEDNPDL IGSEIARWLS TLEIGGTGGS GGTGGSGGSI GTGFPFDPHY VEVLGERMHY VDVGPRDGTP
561 VLFLHGNPTS SYVWRNIIPH VAPTHRCIAP DLIGMGKSDK PDLGYFFDDH VRFMDAFIEA LGLEEVVLVI
631 HDWGSALGFH WAKRNPERVK GIAFMEFIRP IPTWDEWGSG ATNFSLLKQA GDVEENPGPG GSAILWAILW
701 HEMWHEGLEE ASRLYFGERN VKGMFEVLEP LHAMMERGPQ TLKETSFNQA YGRDLMEAQE WCRKYMKSGN
771 VKDLLQAWDL YYHVFRRISK GGSARETFQA FRTGSGGSGV SKGEAVIKEF MRFKVHMEGS MNGHEFEIEG
841 EGEGRPYEGT QTAKLKVTKG GPLPFSWDIL SPQFMYGSRA FTKHPADIPD YYKQSFPEGF KWERVMNFED
911 GGAVTVTQDT SLEDGTLIYK VKLRGTNFPP DGPVMQKKTM GWEASTERLY PEDGVLKGDI KMALRLKDGG
981 RYLADFKTTY KAKKPVQMPG AYNVDRKLDI TSHNEDYTVV EQYERSEGRH STG
```

Lyn11-EGFP-FKBP-(GGS)-cpHaloΔ-P2A-FRB-(GGS)-Hpep5-mScarlet

```
  1 HHHHHHHHHH MGCIKSKGKD SAGADSAGSA GMVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT
 71 YGKLTLKFIC TTGKLPVPWP TLVTTLTYGV QCFSRYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT
141 RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKIR HNIEDGSVQL
211 ADHYQQNTPI GDGPVLLPDN HYLSTQSALS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK GSGMGVQVET
281 ISPGDGRTFP KRGQTCVVHY TGMLEDGKKF DSSRDRNKPF KFMLGKQEVI RGWEEGVAQM SVGQRAKLTI
351 SPDYAYGATG HPGIIPPHAT LVFDVELLKL EGGSDVGRKL IIDQNVFIEG TLPMGVVRPL TEVEMDHYRE
421 PFLNPVDREP LWRFPNELPI AGEPANIVAL VEEYMDWLHQ SPVPKLLFWG TPGVLIPPAE AARLAKSLPN
491 CKAVDIGPGL NLLQEDNPDL IGSEIARWLS TLEIGGTGGS GGTGGSGGSI GTGFPFDPHY VEVLGERMHY
561 VDVGPRDGTP VLFLHGNPTS SYVWRNIIPH VAPTHRCIAP DLIGMGKSDK PDLGYFFDDH VRFMDAFIEA
631 LGLEEVVLVI HDWGSALGFH WAKRNPERVK GIAFMEFIRP IPTWDEWGSG ATNFSLLKQA GDVEENPGPG
701 GSAILWHEMW HEGLEEASRL YFGERNVKGM FEVLEPLHAM MERGPQTLKE TSFNQAYGRD LMEAQEWCRK
771 YMKSGNVKDL LQAWDLYYHV FRRISKGGSS KRDAREMFQA FRTGSGGSGV SKGEAVIKEF MRFKVHMEGS
841 MNGHEFEIEG EGEGRPYEGT QTAKLKVTKG GPLPFSWDIL SPQFMYGSRA FTKHPADIPD YYKQSFPEGF
```

```
 911 KWERVMNFED GGAVTVTQDT SLEDGTLIYK VKLRGTNFPP DGPVMQKKTM GWEASTERLY PEDGVLKGDI
 981 KMALRLKDGG RYLADFKTTY KAKKPVQMPG AYNVDRKLDI TSHNEDYTVV EQYERSEGRH STG
```

Lyn11-EGFP-cpHaloΔ-(GGS)9-FKBP-P2A-Hpep1-(GGS)3-FRB-mScarlet

```
   1 HHHHHHHHHH MGCIKSKGKD SAGADSAGSA GMVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT
  71 YGKLTLKFIC TTGKLPVPWP TLVTTLTYGV QCFSRYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT
 141 RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKIR HNIEDGSVQL
 211 ADHYQQNTPI GDGPVLLPDN HYLSTQSALS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK GSGGSGDVGR
 281 KLIIDQNVFI EGTLPMGVVR PLTEVEMDHY REPFLNPVDR EPLWRFPNEL PIAGEPANIV ALVEEYMDWL
 351 HQSPVPKLLF WGTPGVLIPP AEAARLAKSL PNCKAVDIGP GLNLLQEDNP DLIGSEIARW LSTLEIGGTG
 421 GSGGTGGSGG SIGTGFPFDP HYVEVLGERM HYVDVGPRDG TPVLFLHGNP TSSYVWRNII PHVAPTHRCI
 491 APDLIGMGKS DKPDLGYFFD DHVRFMDAFI EALGLEEVVL VIHDWGSALG FHWAKRNPER VKGIAFMEFI
 561 RPIPTWDEWG SGGTGGSGGS GGTGGSGGSG GTGGSGMGVQ VETISPGDGR TFPKRGQTCV VHYTGMLEDG
 631 KKFDSSRDRN KPFKFMLGKQ EVIRGWEEGV AQMSVGQRAK LTISPDYAYG ATGHPGIIPP HATLVFDVEL
 701 LKLEGSGATN FSLLKQAGDV EENPGPGGSA RETFQAFRTG GSGGTGGSAI LWHEMWHEGL EEASRLYFGE
 771 RNVKGMFEVL EPLHAMMERG PQTLKETSFN QAYGRDLMEA QEWCRKYMKS GNVKDLLQAW DLYYHVFRRI
 841 SKGSGVSKGE AVIKEFMRFK VHMEGSMNGH EFEIEGEGEG RPYEGTQTAK LKVTKGGPLP FSWDILSPQF
 911 MYGSRAFTKH PADIPDYYKQ SFPEGFKWER VMNFEDGGAV TVTQDTSLED GTLIYKVKLR GTNFPPDGPV
 981 MQKKTMGWEA STERLYPEDG VLKGDIKMAL RLKDGGRYLA DFKTTYKAKK PVQMPGAYNV DRKLDITSHN
1051 EDYTVVEQYE RSEGRHSTG
```

Lyn11-EGFP-cpHaloΔ-(GGS)9-FKBP-P2A-Hpep5-(GGS)3-FRB-mScarlet

```
   1 HHHHHHHHHH MGCIKSKGKD SAGADSAGSA GMVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT
  71 YGKLTLKFIC TTGKLPVPWP TLVTTLTYGV QCFSRYPDHM KQHDFFKSAM PEGYVQERTI FFKDDGNYKT
 141 RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKIR HNIEDGSVQL
 211 ADHYQQNTPI GDGPVLLPDN HYLSTQSALS KDPNEKRDHM VLLEFVTAAG ITLGMDELYK GSGGSGDVGR
 281 KLIIDQNVFI EGTLPMGVVR PLTEVEMDHY REPFLNPVDR EPLWRFPNEL PIAGEPANIV ALVEEYMDWL
 351 HQSPVPKLLF WGTPGVLIPP AEAARLAKSL PNCKAVDIGP GLNLLQEDNP DLIGSEIARW LSTLEIGGTG
 421 GSGGTGGSGG SIGTGFPFDP HYVEVLGERM HYVDVGPRDG TPVLFLHGNP TSSYVWRNII PHVAPTHRCI
 491 APDLIGMGKS DKPDLGYFFD DHVRFMDAFI EALGLEEVVL VIHDWGSALG FHWAKRNPER VKGIAFMEFI
 561 RPIPTWDEWG SGGTGGSGGS GGTGGSGGSG GTGGSGMGVQ VETISPGDGR TFPKRGQTCV VHYTGMLEDG
 631 KKFDSSRDRN KPFKFMLGKQ EVIRGWEEGV AQMSVGQRAK LTISPDYAYG ATGHPGIIPP HATLVFDVEL
 701 LKLEGSGATN FSLLKQAGDV EENPGPGGSS KRDAREMFQA FRTGGSGGTG GSAILWHEMW HEGLEEASRL
 771 YFGERNVKGM FEVLEPLHAM MERGPQTLKE TSFNQAYGRD LMEAQEWCRK YMKSGNVKDL LQAWDLYYHV
 841 FRRISKGSGV SKGEAVIKEF MRFKVHMEGS MNGHEFEIEG EGEGRPYEGT QTAKLKVTKG GPLPFSWDIL
 911 SPQFMYGSRA FTKHPADIPD YYKQSFPEGF KWERVMNFED GGAVTVTQDT SLEDGTLIYK VKLRGTNFPP
 981 DGPVMQKKTM GWEASTERLY PEDGVLKGDI KMALRLKDGG RYLADFKTTY KAKKPVQMPG AYNVDRKLDI
1051 TSHNEDYTVV EQYERSEGRH STG
```

Caprola

Color code:

His-tag Strep-tag enterokinase cleavage site linker cpHaloΔ M13 CaM

Caprola$_1$

```
  1 MASWSHPQFE KGADDDDKVP HGGSARETFQ AFRTGSDQLT EEQIAEFKEA FSLFDKDGDG TITTKELGTV
 71 MRSLGQNPTE AELQDMINEV DADGDGTIDF PEFLTMMARK MKDTDSEEEI REAFRVFDKD GNGYISAAEL
141 RHVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVVMM TAKEFPPPPP PPPPPPPPPP PPPPPPPPPP
211 PPPPPGGSRV DSSRRKMNKT GHALRAIGRL SSLEGGSDVG RKLIIDQNVF IEGTLPMGVV RPLTEVEMDH
281 YREPFLNPVD REPLWRFPNE LPIAGEPANI VALVEEYMDW LHQSPVPKLL FWGTPGVLIP PAEAARLAKS
351 LPNCKAVDIG PGLNLLQEDN PDLIGSEIAR WLSTLEIGGT GGSGGTGGSG GSIGTGFPFD PHYVEVLGER
421 MHYVDVGPRD GTPVLFLHGN PTSSYVWRNI IPHVAPTHRC IAPDLIGMGK SDKPDLGYFF DDHVRFMDAF
491 IEALGLEEVV LVIHDWGSAL GFHWAKRNPE RVKGIAFMEF IRPIPTWDEW APGFSSISAH HHHHHHHHH
```

M13 sequences

Caprola version	M13 sequence
Caprola$_1$	RVDSSRRKWNKTGHALRAIGRLSSLE
Caprola$_2$	RVDSSRRKWNKTGHAVRAIGRLSSLE
Caprola$_3$	RVDSSRRKFNKAGHALRAIGRLSSLE
Caprola$_4$	RVDSSRRKFNKTGHAVRAIGRLSSLE
Caprola$_5$	RVDSSRRKFNKTAHALRAIGRLSSLE
Caprola$_6$	RVDSSRRKFNKTGHALRAIGRLSSLE
Caprola$_7$	RVDSSRRKLNKTGHALRAIGRLSSLE
Caprola$_8$	RVDSSRRKWNKTGHATRAIGRLSSLE
Caprola$_9$	RVDSSRRKMNKTGHALRAIGRLSSLE
Caprola$_{10}$	RVDSSRRKVNKTGHALRAIGRLSSLE
Caprola$_{11}$	RVDSSRRKWNKTDHALRAIGRLSSLE
Caprola$_{12}$	RVDSSRRKYNKTGHALRAIGRLSSLE
Caprola$_{13}$	RVDSSRRKFNKTGKALRAIGRLSSLE
Caprola$_{14}$	RVDSSRRKFNKTGHATRAIGRLSSLE
Caprola$_{15}$	RVDSSRRKFNKDGHALRAIGRLSSLE

Split-HaloTag NRX/NLG fusions

Color code:

EGFP mTagBFP2 F2A$_3$ NLG NRX Hpep cpHaloΔ V5-tag FLAG-tag

pre-SynProLa: mTagBFP2-F2A$_3$-NRXss-FLAG-cpHaloΔ2-NRX

```
  1 MVSKGEELIK ENMHMKLYME GTVDNHHFKC TSEGEGKPYE GTQTMRIKVV EGGPLPFAFD ILATSFLYGS
```

```
  71 KTFINHTQGI PDFFKQSFPE GFTWERVTTY EDGGVLTATQ DTSLQDGCLI YNVKIRGVNF TSNGPVMQKK
 141 TLGWEAFTET LYPADGGLEG RNDMALKLVG GSHLIANAKT TYRSKKPAKN LKMPGVYYVD YRLERIKEAN
 211 NETYVEQHEV AVARYCDLPS KLGHKLNSST NFDLLKLAGD VESNPGPNFD LLKLAGDVES NPGPNFDLLK
 281 LAGDVESNPG PMHLRIHARR SPPRRPAWTL GIWFLFWGCI VSSVWSGSGD YKDDDDKGSG DVGRKLIIDQ
 351 NVFIEGTLPM GVVRPLTEEE MDHYREPFLN PKDREPLWRF PNELPIAGEP ANIVALVEEY MDWLHQSPVP
 421 KLLFWGTPGV LIPPAEAARL AKSLPNCKAV DIGPGLNLLQ EDNPDLIGSE IARWLSTLEI KSKYDRDQIL
 491 KIIAELEKKT GGSIGTGFPF DPHYVEVLGS RMHYVDVGPR DGTPVLFLHG NPTSSYVWRN IIPHVAPTHR
 561 CIAPDLIGMG KSDKPDLGYF FDDHVRFMDA FIEALGLEEV VLVIHDWGSA LGFHWAKRHP ERVKGIAFME
 631 FIRPIPTWDE WGSGGSGGSG GSGSSNVASS SSTSSSPGSH SQHEHHFHGS KHHSVPISIY RSPVSLRGGH
 701 AGATYIFGKS GGLILYTWPA NDRPSTRSDR LAVGFSTTVK DGILVRIDSA PGLGDFLQLH IEQGKIGVVF
 771 NIGTVDISIK EERTPVNDGK YHVVRFTRNG GNATLQVDNW PVNEHYPTGR QLTIFNTQAQ IAIGGKDKGR
 841 LFQGQLSGLY YDGLKVLNMA AENNPNIKIN GSVRLVGEVP SILGTTQTTS MPPEMSTTVM ETTTTMATTT
 911 TRKNRSTASI QPTSDDLVSS AECSSDDEDF VECEPSTGRS ANPTEPGIRR VPGASEVIRE SSSTTGMVVG
 981 IVAAAALCIL ILLYAMYKYR NRDEGSYQVD ETRNYISNSA QSNGTLMKEK QQSSKSGHKK QKNKDREYYV
1051 **
```

post-SynProLa: EGFP-F2A$_3$-NLGss-V5-Hpep5-NLG

```
   1 MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT TGKLPVPWPT LVTTLTYGVQ
  71 CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH
 141 KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSALSK
 211 DPNEKRDHMV LLEFVTAAGI TLGMDELYKS STNFDLLKLA GDVESNPGPN FDLLKLAGDV ESNPGPNFDL
 281 LKLAGDVESN PGPMALPRCM WPNYVWRAMM ACVVHRGSGA PLTLCLLGCL LQTFHVLSQK GSGGKPIPNP
 351 LLGLDSTGSG SKRDAREMFQ AFRTGSGGSG GSGGSGQLLD DVDPLVTTNF GKIRGIKKEL NNEILGPVIQ
 421 FLGVPYAAPP TGEHRFQPPE PPSPWSDIRN ATQFAPVCPQ NIIDGRLPEV MLPVWFTNNL DVVSSYVQDQ
 491 SEDCLYLNIY VPTEDGPLTK KHTDDLGDND GAEDEDIRDS GGPKPVMVYI HGGSYMEGTG NLYDGSVLAS
 561 YGNVIVITVN YRLGVLGFLS TGDQAAKGNY GLLDLIQALR WTSENIGFFG GDPLRITVFG SGAGGSCVNL
 631 LTLSHYSEGN RWSNSTKGLF QRAIAQSGTA LSSWAVSFQP AKYARILATK VGCNVSDTVE LVECLQKKPY
 701 KELVDQDVQP ARYHIAFGPV IDGDVIPDDP QILMEQGEFL NYDIMLGVNQ GEGLKFVENI VDSDDGVSAS
 771 DFDFAVSNFV DNLYGYPEGK DVLRETIKFM YTDWADRHNP ETRRKTLLAL FTDHQWVAPA VATADLHSNF
 841 GSPTYFYAFY HHCQTDQVPA WADAAHGDEV PYVLGIPMIG PTELFPCNFS KNDVMLSAVV MTYWTNFAKT
 911 GDPNQPVPQD TKFIHTKPNR FEEVAWTRYS QKDQLYLHIG LKPRVKEHYR ANKVNLWLEL VPHLHNLNDI
 981 SQYTSTTTKV PSTDITLRPT RKNSTPVTSA FPTAKQDDPK QQPSPFSVDQ RDYSTELSVT IAVGASLLFL
1051 NILAFAALYY KKDKRRHDVH RRCSPQRTTT NDLTHAPEEE IMSLQMKHTD LDHECESIHP HEVVLRTACP
1121 PDYTLAMRRS PDDIPLMTPN TITMIPNTIP GIQPLHTFNT FTGGQNNTLP HPHPHPHSHS TTRV
```

XML Rosetta Scripts

Conservative design of new 10mer Hpep variants

```
<ROSETTASCRIPTS>
<SCOREFXNS>

    <ScoreFunction name="sfxn_stand" weights="ref2015">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

    <ScoreFunction name="sfxn_soft" weights="ref2015_soft">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

</SCOREFXNS>

<RESIDUE_SELECTORS>

    <Index name="Nterm" resnums="1-3"/>
    <Index name="Cterm" resnums="281-283"/>
    <And name="termini" selectors="Nterm,Cterm"/>
    <Not name="not_termini" selector="termini"/>
    <Index name="Hpep" resnums="286-295"/>

</RESIDUE_SELECTORS>

<TASKOPERATIONS>

    <ReadResfile name="resfile" filename="in/resfile"/>

    <ExtraRotamersGeneric name="extra_rot"
    ex1="true" ex2="true" ex3="false" ex4="false"
    ex1aro="true" ex2aro="true" ex1aro_exposed="false" ex2aro_exposed="false"
    extrachi_cutoff="0"/>

    <OperateOnResidueSubset name="to_Nterm" selector="Nterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_Cterm" selector="Cterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_not_termini" selector="not_termini">
        <PreventRepackingRLT/>
    </OperateOnResidueSubset>
```

```

    <ConservativeDesignOperation name="conservative_design" residue_selector="Hpep"
        data_source="blosum80" include_native_aa="true"/>

</TASKOPERATIONS>

<FILTERS>

    <Ddg name="ddG" scorefxn="sfxn_stand" threshold="-15" repeats="3" jump="3"
        repack_bound="false"/>

    <Sasa name="sasa" threshold="800" jump="3"/>

    <Rmsd name="rmsd" confidence="0"/>

    <CompoundStatement name="ddg_sasa">
        <AND filter_name="ddG"/>
        <AND filter_name="sasa"/>
    </CompoundStatement>

    <ScoreType name="tot_score"
        threshold="-250" scorefxn="sfxn_stand" confidence="1" />

    <ScoreType name="tot_score_soft"
        threshold="-250" scorefxn="sfxn_soft" confidence="1" />

</FILTERS>

<MOVERS>

    <AddConstraints name="constraint" >
        <FileConstraintGenerator name="gen_my_csts"
    filename="in/chemical_bond.cst" />
    </AddConstraints>

    <DeclareBond name="ligand_bond" res1="246" res2="284" atom1="CG" atom2="O1"/>

    <AtomTree name="set_foldtree" fold_tree_file="in/foldtree" />

    <SetTorsion name="perturb_Nterm">
        <Torsion residue="1" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="2" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="3" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <SetTorsion name="perturb_Cterm">
        <Torsion residue="281" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="282" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="283" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

```

```
    <FastRelax name="relax_term" scorefxn="sfxn_stand"
        disable_design="true"
        task_operations="to_Nterm,to_Cterm,to_not_termini,extra_rot"
        repeats="5"
        cst_file="in/chemical_bond.cst"
        ramp_down_constraints="false" bondangle="false" bondlength="false">
        <MoveMap name="mov1" bb="false" chi="false" jump="false" >
            <Span begin="1" end="3" chi="true" bb="true" bondangle="false"
                bondlength="false" />
            <Span begin="281" end="283" chi="true" bb="true" bondangle="false"
                bondlength="false" />
        </MoveMap>
    </FastRelax>

    <Docking name="docking" score_high="sfxn_soft" fullatom="1" local_refine="1"
        jumps="1" optimize_fold_tree="true" conserve_foldtree="true"/>

    <FastDesign name="design" scorefxn="sfxn_stand"
        disable_design="false"
        task_operations="extra_rot,resfile,conservative_design"
        repeats="3" relaxscript="in/InterfaceDesign2019_mod.txt"
        cst_file="in/chemical_bond.cst" cartesian="false"
        ramp_down_constraints="false" bondangle="false"
        bondlength="false"
        min_type="lbfgs_armijo_nonmonotone">
        <MoveMap name="mov2" bb="true" chi="true" jump="false" >
            <Jump number="3" setting="true" />
            <Span begin="1" end="295"
                chi="true" bb="true" bondangle="false" bondlength="false" />
        </MoveMap>
    </FastDesign>

</MOVERS>

<PROTOCOLS>

    <Add mover="constraint"/>
    <Add mover="ligand_bond"/>
    <Add mover="set_foldtree"/>
    <Add mover="perturb_Nterm"/>
    <Add mover="perturb_Cterm"/>
    <Add mover="relax_term"/>
    <Add mover="docking"/>
    <Add filter="tot_score_soft"/>
    <Add mover="design"/>
    <Add filter="ddg_sasa"/>
    <Add filter="ddG"/>
    <Add filter="sasa"/>
    <Add filter="rmsd"/>
    <Add filter="tot_score"/>

```

```
</PROTOCOLS>
</ROSETTASCRIPTS>
```

Design of 14mer Hpep variants

```
<ROSETTASCRIPTS>

<SCOREFXNS>

    <ScoreFunction name="sfxn_stand" weights="ref2015">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

    <ScoreFunction name="sfxn_soft" weights="ref2015_soft">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

</SCOREFXNS>

<RESIDUE_SELECTORS>

    <Index name="Nterm" resnums="1-3"/>
    <Index name="Cterm" resnums="281-283"/>
    <And name="termini" selectors="Nterm,Cterm"/>
    <Not name="not_termini" selector="termini"/>

</RESIDUE_SELECTORS>

<TASKOPERATIONS>

    <ReadResfile name="res" filename="in/resfile"/>

    <ExtraRotamersGeneric name="extra_rot"
    ex1="true" ex2="true" ex3="false" ex4="false"
    ex1aro="true" ex2aro="true" ex1aro_exposed="false" ex2aro_exposed="false"
    extrachi_cutoff="0"/>

    <OperateOnResidueSubset name="to_Nterm" selector="Nterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_Cterm" selector="Cterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_not_termini" selector="not_termini">
        <PreventRepackingRLT/>
```

```
    </OperateOnResidueSubset>

</TASKOPERATIONS>

<FILTERS>

    <Ddg name="ddG" scorefxn="sfxn_stand" threshold="-15" repeats="3" jump="3"
        repack_bound="false"/>

    <Sasa name="sasa" threshold="800" jump="3"/>

    <Rmsd name="rmsd" confidence="0"/>

    <CompoundStatement name="ddg_sasa">
        <AND filter_name="ddG"/>
        <AND filter_name="sasa"/>
    </CompoundStatement>

    <ScoreType name="tot_score"
        threshold="-250" scorefxn="sfxn_stand" confidence="1" />

    <ScoreType name="tot_score_soft"
        threshold="-250" scorefxn="sfxn_soft" confidence="1" />

</FILTERS>

<MOVERS>

    <AddConstraints name="constraint" >
        <FileConstraintGenerator name="gen_my_csts"
    filename="in/chemical_bond.cst" />
    </AddConstraints>

    <DeclareBond name="ligand_bond" res1="246" res2="284" atom1="CG" atom2="O1"/>

    <AtomTree name="set_foldtree" fold_tree_file="in/foldtree" />

    <SetTorsion name="perturb_Nterm">
        <Torsion residue="1" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="2" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="3" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <SetTorsion name="perturb_Cterm">
        <Torsion residue="281" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="282" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="283" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <FastRelax name="relax_term" scorefxn="sfxn_stand"
        disable_design="true"
```

```
        task_operations="to_Nterm,to_Cterm,to_not_termini,extra_rot"
        repeats="5"
        cst_file="in/chemical_bond.cst"
        ramp_down_constraints="false" bondangle="false" bondlength="false">
        <MoveMap name="mov1" bb="false" chi="false" jump="false" >
            <Span begin="1" end="3" chi="true" bb="true" bondangle="false"
                bondlength="false" />
            <Span begin="281" end="283" chi="true" bb="true" bondangle="false"
                bondlength="false" />
        </MoveMap>
    </FastRelax>

    <Docking name="docking" score_high="sfxn_soft" fullatom="1" local_refine="1"
        jumps="1" optimize_fold_tree="true" conserve_foldtree="true"/>

    <FastDesign name="design" scorefxn="sfxn_stand"
        disable_design="false"
        task_operations="extra_rot,res"
        repeats="1" relaxscript="in/InterfaceDesign2019_mod.txt"
        cst_file="in/chemical_bond.cst" cartesian="false"
        ramp_down_constraints="false" bondangle="false"
        bondlength="false"
        min_type="lbfgs_armijo_nonmonotone">
        <MoveMap name="mov2" bb="true" chi="true" jump="false" >
            <Jump number="3" setting="true" />
            <Span begin="1" end="299"
                chi="true" bb="true" bondangle="false" bondlength="false" />
        </MoveMap>
    </FastDesign>

</MOVERS>

<PROTOCOLS>

    <Add mover="constraint"/>
    <Add mover="ligand_bond"/>
    <Add mover="set_foldtree"/>
    <Add mover="perturb_Nterm"/>
    <Add mover="perturb_Cterm"/>
    <Add mover="relax_term"/>
    <Add mover="docking"/>
    <Add filter="tot_score_soft"/>
    <Add mover="design"/>
    <Add filter="ddg_sasa"/>
    <Add filter="ddG"/>
    <Add filter="sasa"/>
    <Add filter="rmsd"/>
    <Add filter="tot_score"/>
</PROTOCOLS>

</ROSETTASCRIPTS>
```

Design of 14mer Hpep variants with more FastDesign repetitions

```
<ROSETTASCRIPTS>

<SCOREFXNS>

    <ScoreFunction name="sfxn_stand" weights="ref2015">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

    <ScoreFunction name="sfxn_soft" weights="ref2015_soft">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

</SCOREFXNS>

<RESIDUE_SELECTORS>

    <Index name="Nterm" resnums="1-3"/>
    <Index name="Cterm" resnums="281-283"/>
    <And name="termini" selectors="Nterm,Cterm"/>
    <Not name="not_termini" selector="termini"/>

</RESIDUE_SELECTORS>

<TASKOPERATIONS>

    <ReadResfile name="res" filename="in/resfile"/>

    <ExtraRotamersGeneric name="extra_rot"
    ex1="true" ex2="true" ex3="false" ex4="false"
    ex1aro="true" ex2aro="true" ex1aro_exposed="false" ex2aro_exposed="false"
    extrachi_cutoff="0"/>

    <OperateOnResidueSubset name="to_Nterm" selector="Nterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_Cterm" selector="Cterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_not_termini" selector="not_termini">
        <PreventRepackingRLT/>
    </OperateOnResidueSubset>

</TASKOPERATIONS>

```

```xml
<FILTERS>

    <Ddg name="ddG" scorefxn="sfxn_stand" threshold="-15" repeats="3" jump="3"
        repack_bound="false"/>

    <Sasa name="sasa" threshold="800" jump="3"/>

    <Rmsd name="rmsd" confidence="0"/>

    <CompoundStatement name="ddg_sasa">
        <AND filter_name="ddG"/>
        <AND filter_name="sasa"/>
    </CompoundStatement>

    <ScoreType name="tot_score"
        threshold="-250" scorefxn="sfxn_stand" confidence="1" />

    <ScoreType name="tot_score_soft"
        threshold="-250" scorefxn="sfxn_soft" confidence="1" />

</FILTERS>

<MOVERS>

    <AddConstraints name="constraint" >
        <FileConstraintGenerator name="gen_my_csts" filename="in/chemical_bond.cst" />
    </AddConstraints>

    <DeclareBond name="ligand_bond" res1="246" res2="284" atom1="CG" atom2="O1"/>

    <AtomTree name="set_foldtree" fold_tree_file="in/foldtree" />

    <SetTorsion name="perturb_Nterm">
        <Torsion residue="1" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="2" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="3" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <SetTorsion name="perturb_Cterm">
        <Torsion residue="281" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="282" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="283" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <FastRelax name="relax_term" scorefxn="sfxn_stand"
        disable_design="true"
        task_operations="to_Nterm,to_Cterm,to_not_termini,extra_rot"
        repeats="5"
        cst_file="in/chemical_bond.cst"
        ramp_down_constraints="false" bondangle="false" bondlength="false">
```

```
        <MoveMap name="mov1" bb="false" chi="false" jump="false" >
            <Span begin="1" end="3" chi="true" bb="true" bondangle="false"
                bondlength="false" />
            <Span begin="281" end="283" chi="true" bb="true" bondangle="false"
                bondlength="false" />
        </MoveMap>
    </FastRelax>

    <Docking name="docking" score_high="sfxn_soft" fullatom="1" local_refine="1"
        jumps="1" optimize_fold_tree="true" conserve_foldtree="true"/>

    <FastDesign name="design" scorefxn="sfxn_stand"
        disable_design="false"
        task_operations="extra_rot,res"
        repeats="3" relaxscript="in/InterfaceDesign2019_mod.txt"
        cst_file="in/chemical_bond.cst" cartesian="false"
        ramp_down_constraints="false" bondangle="false"
        bondlength="false"
        min_type="lbfgs_armijo_nonmonotone">
        <MoveMap name="mov2" bb="true" chi="true" jump="false" >
            <Jump number="3" setting="true" />
            <Span begin="1" end="299"
                chi="true" bb="true" bondangle="false" bondlength="false" />
        </MoveMap>
    </FastDesign>

</MOVERS>

<PROTOCOLS>

    <Add mover="constraint"/>
    <Add mover="ligand_bond"/>
    <Add mover="set_foldtree"/>
    <Add mover="perturb_Nterm"/>
    <Add mover="perturb_Cterm"/>
    <Add mover="relax_term"/>
    <Add mover="docking"/>
    <Add filter="tot_score_soft"/>
    <Add mover="design"/>
    <Add filter="ddg_sasa"/>
    <Add filter="ddG"/>
    <Add filter="sasa"/>
    <Add filter="rmsd"/>
    <Add filter="tot_score"/>
</PROTOCOLS>

</ROSETTASCRIPTS>
```

Design of 14mer Hpep variants with fixed cpHaloΔ backbone

```
<ROSETTASCRIPTS>
```

```xml

<SCOREFXNS>

    <ScoreFunction name="sfxn_stand" weights="ref2015">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

    <ScoreFunction name="sfxn_soft" weights="ref2015_soft">
        <Reweight scoretype="atom_pair_constraint" weight="10"/>
        <Reweight scoretype="angle_constraint" weight="10"/>
        <Reweight scoretype="dihedral_constraint" weight="10"/>
    </ScoreFunction>

</SCOREFXNS>

<RESIDUE_SELECTORS>

    <Index name="Nterm" resnums="1-3"/>
    <Index name="Cterm" resnums="281-283"/>
    <And name="termini" selectors="Nterm,Cterm"/>
    <Not name="not_termini" selector="termini"/>

</RESIDUE_SELECTORS>

<TASKOPERATIONS>

    <ReadResfile name="res" filename="in/resfile"/>

    <ExtraRotamersGeneric name="extra_rot"
    ex1="true" ex2="true" ex3="false" ex4="false"
    ex1aro="true" ex2aro="true" ex1aro_exposed="false" ex2aro_exposed="false"
    extrachi_cutoff="0"/>

    <OperateOnResidueSubset name="to_Nterm" selector="Nterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_Cterm" selector="Cterm">
        <RestrictToRepackingRLT/>
    </OperateOnResidueSubset>

    <OperateOnResidueSubset name="to_not_termini" selector="not_termini">
        <PreventRepackingRLT/>
    </OperateOnResidueSubset>

</TASKOPERATIONS>

<FILTERS>

```

```
    <Ddg name="ddG" scorefxn="sfxn_stand" threshold="-15" repeats="3" jump="3"
        repack_bound="false"/>

    <Sasa name="sasa" threshold="800" jump="3"/>

    <Rmsd name="rmsd" confidence="0"/>

    <CompoundStatement name="ddg_sasa">
        <AND filter_name="ddG"/>
        <AND filter_name="sasa"/>
    </CompoundStatement>

    <ScoreType name="tot_score"
        threshold="-250" scorefxn="sfxn_stand" confidence="1" />

    <ScoreType name="tot_score_soft"
        threshold="-250" scorefxn="sfxn_soft" confidence="1" />

</FILTERS>

<MOVERS>

    <AddConstraints name="constraint" >
        <FileConstraintGenerator name="gen_my_csts"
    filename="in/chemical_bond.cst" />
    </AddConstraints>

    <DeclareBond name="ligand_bond" res1="246" res2="284" atom1="CG" atom2="O1"/>

    <AtomTree name="set_foldtree" fold_tree_file="in/foldtree" />

    <SetTorsion name="perturb_Nterm">
        <Torsion residue="1" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="2" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="3" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <SetTorsion name="perturb_Cterm">
        <Torsion residue="281" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="282" torsion_name="rama" angle="rama_biased"/>
        <Torsion residue="283" torsion_name="rama" angle="rama_biased"/>
    </SetTorsion>

    <FastRelax name="relax_term" scorefxn="sfxn_stand"
        disable_design="true"
        task_operations="to_Nterm,to_Cterm,to_not_termini,extra_rot"
        repeats="5"
        cst_file="in/chemical_bond.cst"
        ramp_down_constraints="false" bondangle="false" bondlength="false">
        <MoveMap name="mov1" bb="false" chi="false" jump="false" >
            <Span begin="1" end="3" chi="true" bb="true" bondangle="false"
```

```
                bondlength="false" />
            <Span begin="281" end="283" chi="true" bb="true" bondangle="false"
                bondlength="false" />
        </MoveMap>
    </FastRelax>

    <Docking name="docking" score_high="sfxn_soft" fullatom="1" local_refine="1"
        jumps="1" optimize_fold_tree="true" conserve_foldtree="true"/>

    <FastDesign name="design" scorefxn="sfxn_stand"
        disable_design="false"
        task_operations="extra_rot,res"
        repeats="3" relaxscript="in/InterfaceDesign2019_chill.txt"
        cst_file="in/chemical_bond.cst" cartesian="false"
        ramp_down_constraints="false" bondangle="false"
        bondlength="false"
        min_type="lbfgs_armijo_nonmonotone">
        <MoveMap name="mov2" bb="true" chi="true" jump="false" >
            <Jump number="3" setting="true" />
            <Span begin="1"   end="285" chi="true" bb="false" bondangle="false"
                bondlength="false" />
            <Span begin="286" end="299" chi="true" bb="true"  bondangle="false"
                bondlength="false" />
        </MoveMap>
    </FastDesign>

</MOVERS>

<PROTOCOLS>

    <Add mover="constraint"/>
    <Add mover="ligand_bond"/>
    <Add mover="set_foldtree"/>
    <Add mover="perturb_Nterm"/>
    <Add mover="perturb_Cterm"/>
    <Add mover="relax_term"/>
    <Add mover="docking"/>
    <Add filter="tot_score_soft"/>
    <Add mover="design"/>
    <Add filter="ddg_sasa"/>
    <Add filter="ddG"/>
    <Add filter="sasa"/>
    <Add filter="rmsd"/>
    <Add filter="tot_score"/>
</PROTOCOLS>

</ROSETTASCRIPTS>
```

List of Abbreviations

AF	arborization field
BCN	bicyclononyne
BLOSUM	blocks substitution matrix
BOP	(benzotriazol-1-yloxy)-tris-(dimethylamino)-phosphonium-hexafluorophosphat
BSA	bovine serum albumin
BW	bandwidth
CaM	calmodulin
CATS	comprehensive analysis of tissues across scales
CD	cluster of differentiation
CLEM	correlative light and electron microscopy
CMV	cytomegalovirus
CP-linker	circular permutation linker
cpHaloTag	circular permuted HaloTag
CPY	carbopyronine
DCM	dichloromethane
DEG	differentially expressed gene
DHFR	dihydrofolate reductase
DIC	*N*,*N*'-diisopropylcarbodiimide
DMEM	Dulbecco's Modified Eagle Medium
DMF	dimethylformamid
DNA	deoxyribonucleic acid
dNTP	deoxynucleosidtriphosphate
dpf	days post fertilization
DREADD	designer receptors activated only by designer drugs
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EEG	electroencephalography
EGFP	enhanced green fluorescent protein
EGTA	ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*',*N*'-tetraacetic acid

EM	electron microscopy
ESI	electron spray ionisation
ExLLSM	expansion lattice light-sheet microscopy
ExM	expansion microscopy
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FKBP	FK506-binding protein
Fmoc	fluorenylmethoxycarbonyl
fMRI	functional magnetic resonance imaging
FP	fluorescence polarization
FPLC	fast protein liquid chromatography
FRB	FKBP--rapamycin binding domain
FRET	fluorescence resonance energy transfer
FRT	Flp recombination target
GABA	γ-aminobutyric acid
GECI	genetically encoded calcium indicator
GEVI	genetically encoded voltage indicator
GFP	green fluorescent protein
GO	gene ontology
GRASP	GFP reconstitution across synaptic partners
hAGT	human O^6-alkylguanine-DNA alkyl transferase
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOB	HaloTag-based oligonucleotide binder
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
HRP	horseradish peroxidase
HT	HaloTag
HTL	HaloTag ligand
IEG	immediate early gene
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
JF	Janelia Fluor
k_{app}	apparent second-order rate constant
LB	lysogenic broth

MeCN	acetonitrile
MEG	magnetoencephalography
MOPS	3-(*N*-morpholino)propanesulfonic acid
MS	mass spectrometry
NAD	nicotinamide adenine dinucleotide
nanoDSF	nano differential scanning fluorimetry
NCS	newborn calf serum
NLG	neuroligin
NLGss	neuroligin signal sequence
Nor	norbornene
NRX	neurexin
NRXss	neurexin signal sequence
PAGE	polyacrylamide gel electrophoresis
PAINT	points accumulation for imaging in nanoscale topography
PALM	photo-activated localization microscopy
PBS	phosphate buffered saline
PCA	protein-fragment complementation assay
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PET	positron emission tomography
PMSF	phenylmethylsulfonyl fluoride
PPI	protein–protein interaction
PROSS	protein repair one-stop shop
PSSM	position-specific substitution matrix
RAPA	rapamycin
RCF	relative centrifugal force
RNA	ribonucleic acid
RT	room temperature
SCO	cyclooctyne
SDS	sodium dodecyl sulfate
SLP	self-labeling protein
SSEM	serial section electron microscopy
STED	stimulated emission depletion
STORM	stochastic optical reconstruction microscopy
TEVp	tobacco etch virus protease

TIS	triisopropyl silane
TMR	tetramethylrhodamine
TRIS	tris(hydroxymethyl)aminomethane
Trp1p	N-(5-phosphoribosyl)-anthranilate isomerase
Tz	tetrazine
$\mathbf{T_m}$	melting temperature
U	enzyme unit
UV	ultraviolet
Vbn	vinylbenzene
VSD	voltage-sensitive dyes
WLC	worm-like chain