How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 21007–21015
International Edition: doi.org/10.1002/anie.202006822
German Edition: doi.org/10.1002/ange.202006822

Protein Modifications **Hot Paper**

In Cellulo Protein Semi-Synthesis from Endogenous and Exogenous Fragments Using the Ultra-Fast Split Gp41-1 Intein

*Maniraj Bhagawati, Simon Hoffmann, Katharina S. Höffgen, Jacob Piehler, Karin B. Busch, and Henning D. Mootz**

Abstract: *Protein semi-synthesis inside live cells from exogenous and endogenous parts offers unique possibilities for studying proteins in their native context. Split-intein-mediated protein trans-splicing is predestined for such endeavors and has seen some successes, but a much larger variety of established split inteins and associated protocols is urgently needed. We characterized the association and splicing parameters of the Gp41-1 split intein, which favorably revealed a nanomolar affinity between the intein fragments combined with the exceptionally fast splicing rate. Following bead-loading of a chemically modified intein fragment precursor into live mammalian cells, we fluorescently labeled target proteins on their N- and C-termini with short peptide tags, thus ensuring minimal perturbation of their structure and function. In combination with a nuclear-entrapment strategy to minimize cytosolic fluorescence background, we applied our technique for super-resolution imaging and single-particle tracking of the outer mitochondrial protein Tom20 in HeLa cells.*

Introduction

Total chemical synthesis and semi-synthesis of proteins by backbone ligation strategies enable unique possibilities to study structure-function relationships and biophysical properties of proteins.[1] In protein semi-synthesis, at least one backbone fragment originates from recombinant production, which can be achieved in different expression hosts and be followed by separate chemical modification. However, in contrast to other modern methods to manipulate native or recombinant proteins on single side chains,[2] it remains a great challenge to perform a protein backbone ligation inside a living cell on an endogenously produced protein with a second polypeptide of exogenous origin, to study proteins in their native context. The most widely used native chemical ligation (NCL)[3] and expressed protein ligation (EPL)[4] reactions cannot be used in the cell in a straight-forward way,[5] because they require high concentrations of the two polypeptides to be ligated and the involved functional groups of a C-terminal thioester and an N-terminal cysteine are not stable in the typical cellular milieu. Alternatively, synthetic or semi-synthetic proteins can be prepared outside the cell and then be delivered to the cellular environment by a variety of techniques.[6] However, this latter approach is not feasible when the protein in question requires endogenous expression, at least in part, for example to facilitate proper processing or correct subcellular localization, or because it is not amenable to ex vivo preparation for reasons pertaining to its size or folding.

To address these challenges suitable reactions to semi-synthesize the protein of interest (POI) inside the cell from endogenously expressed and transduced parts are urgently sought for. Split inteins provide a powerful option, because they can tracelessly ligate two proteins or peptides (exteins) by protein trans-splicing under self-excision with a native bond (Figure S1).[7] They exist as two separate fragments (Int^N and Int^C) that first associate with each other to reconstitute the active intein. Split inteins do not require any strict consensus sequence at the ligation site, except for a single Cys, Ser or Thr residue, and their fragments typically display high-affinity interactions with equilibrium dissociation constants K_d ranging from single digit nanomolar to single digit micromolar concentrations.[8] These advantages contrast to other transpeptidation catalysts that have been repurposed for protein ligation, like sortase[9] and butelase,[10] which recognize a conserved, and thus mostly invariable, ligation site, and exhibit comparably low affinity towards their substrates. However, different split inteins exhibit widely varying and unpredictable properties. Therefore, it is challenging to identify split inteins with the best combination of desired characteristics such as high affinity between the intein fragments, fast splicing kinetics, high flexibility in terms of the accepted extein sequence, high degree of stability and activity at a wide range of physicochemical conditions (temperature, pH, etc.).

[*] Dr. M. Bhagawati, S. Hoffmann, K. S. Höffgen, Prof. H. D. Mootz
Department of Chemistry and Pharmacy, Institute of Biochemistry, University of Münster
Corrensstrasse 36, 48149 Münster (Germany)
E-mail: Henning.Mootz@uni-muenster.de
Prof. J. Piehler
Department of Biology and Center for Cellular Nanoanalytics, University of Osnabrück
Barbarastrasse 11, 49076 Osnabrück (Germany)
Prof. K. B. Busch
Institute of Molecular Cell Biology, University of Münster
Schlossplatz 5, 48149 Münster (Germany)

Supporting information, including Figures S1–S15, details about preparation of the proteins used in this study, protein transduction, splice assays and other methods and Video S1, and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.202006822.

The idea to use split inteins for in cellulo semi-synthesis of proteins is to combine an endogenously expressed part with an exogenously prepared part that is transduced inside the cells (Figure 1; left part). Only a handful of such studies have been reported so far, which utilized either one of the alleles of the naturally split DnaE inteins[11] or an artificially split version of the *Ssp* DnaB intein.[11e] In pioneering work, Giriat and Muir endogenously expressed the *Ssp* DnaE Int^N fragment (123 amino acids (aa)) in fusion to a protein of interest (POI), while the Int^C fragment (36 aa) with a synthetic fluorophore in the $extein^C$ sequence was transduced using a cell-penetrating peptide (CPP). Intracellular splicing furnished labeling of the POI.[11a] The groups of Camarero and Kwon exploited the faster splicing and higher yielding *Npu* allele of the DnaE intein ($t_{1/2} \approx 60$ s compared to 1 to 3 h),[12] and delivered the Int^C precursor by protein transfection.[11b,c] Tampé and co-workers[11e] used an artificially split version of the evolved M86 mutant of the *Ssp* DnaB intein.[13] The short Int^N fragment (11 aa) was synthesized with a fluorophore in the $extein^N$ sequence and was introduced into mammalian cells using cell squeezing, while the Int^C part (143 aa) was endogenously expressed in fusion with the POI. To improve on the moderate affinity between the fragments of this intein (≈ 0.1 μM)[13a] the Int^N fragment was equipped with a tris-NTA chelating unit to enable high-affinity interaction with the His-tagged Int^C.[11e]

While split inteins provide a possible pathway to the in cellulo protein reconstitution from inactive recombinant and synthetic pieces, almost all studies so far have focused on splicing short unrelated tag sequences to intact proteins, typically to attach a synthetic fluorophore. Only shortly before and after we submitted this work for publication the first two examples for truly semi-synthetic proteins obtained in cellulo by protein trans-splicing were reported by the Pless and Muir groups, respectively.[14] Nevertheless, fluorescent tag labeling of cellular proteins is also an important goal of the split intein approach, because such extein tags can be of very small size and variable sequence, consisting of only a few amino acids[15] or even only the fluorophore itself.[16] Such short tags are potentially much less perturbing for protein function than the often larger peptide or even protein tags that have been established for imaging studies of intracellular proteins in cell biology.[17] Short tags are also beneficial in super-resolution microscopy techniques to minimize the increase of the observed sizes of cellular structures.

Another aspect of intracellular labeling of proteins with fluorophores is the signal-to-background ratio stemming from the excess fluorophore reagent. For some approaches the removal of excess fluorophores can be achieved by washing the cells, however, this limits the range of fluorophores to those that are cell-permeable. Turn-on fluorescence strategies,[18] switchable fluorophores,[19] or the use of fluorophore-quencher pairs[20] can improve the signal-to-background ratio. Also for split intein approaches, the problem of background from the fluorescent signal of transduced and unspliced synthetic intein precursor was addressed by attaching a quencher molecule.[11b,c,e] However, these additional modifications complicate the synthesis and limit suitable split intein precursors to those that can be prepared by total synthesis. Together, the lack of versatile and straightforward split inteins protocols for intracellular protein modification and fluorescent labeling has prohibited the widespread application of in cellulo protein semi-synthesis in cell biological research.

In this work, we biophysically characterized the fragment association and splicing parameters of the natively split Gp41-1 intein, the fastest splicing split intein known to date.[21] We then explored its favorable properties to establish new protocols to utilize it for in cellulo protein semi-synthesis. Finally, we applied our new technologies to label a mitochondrial outer membrane protein on a very short tag with a synthetic fluorophore to investigate it by super-resolution microscopy and single molecule tracking in live cells. Our work provides a significant expansion of the toolbox and shows new applications for proteins that are prepared in this way.

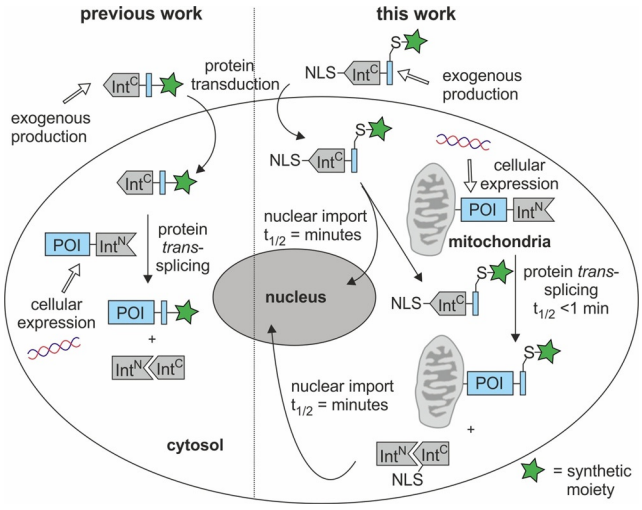

Figure 1. Scheme of split-intein-mediated protein trans-splicing for protein semi-synthesis in live cells. The left part of the cell depicts the general concept for the C-terminal modification of a POI expressed inside mammalian cells with an exogenously produced Int^C protein fragment carrying non-proteinogenic functionalities. The section on the right shows how this general strategy has been expanded in this work with a nuclear entrapment strategy to achieve rapid and nearly background-free fluorescent labeling of the POI. NLS = nuclear localization sequence.

Results and Discussion

Strategic Considerations

We aimed to explore the Gp41-1 intein[21a] for in cellulo protein semi-synthesis. We focused on this naturally split intein because with a half-life time ($t_{1/2}$) of ≈4 s at 45 °C, the highest rate in trans-splicing known to date, and nearly quantitative splicing yields[21b,c] it was supposed to have great potential for the envisaged applications. The rate of protein trans-splicing at 37 °C ($t_{1/2} = \approx 5$ s) is more than 10 times faster than for the *Npu* DnaE intein.[12] However, no biophysical characterization of the split Gp41-1 intein (Int^N: 88 aa, Int^C:

37 aa) has been reported so far and its general suitability for in cellulo protein semi-synthesis is not known.

Interestingly, the catalytically critical +1 residue as the first amino acid of the extein^C is a serine, in contrast to a cysteine in the widely used DnaE inteins.[22] This residue is involved in forming the branched (oxoester) intermediate in the protein splicing pathway[7a,23] and is the only residue strictly required to remain at the ligation site. Together with the absence of any further cysteines in the Gp41-1-Int^C sequence, this intein is therefore suitable for the CysTag approach,[15a,24] which explores a single cysteine residue in an extein part for site-specific bioconjugation and subsequent transfer of the introduced label in the protein trans-splicing reaction.

Biophysical Characterization of the Split Gp41-1 Intein

We first carried out microscale thermophoresis (MST) assays to evaluate the affinity between intein precursor fragments inactivated for splicing by C1A and N125A mutations. We prepared the purified proteins H_6-Smt3-Int^C-(N125A)-CysTag-Trx-H_6 (**1**; Smt3 = SUMO homolog from yeast as expression fusion; CysTag = SSSDVCRS; Trx = thioredoxin) and MBP-Int^N(C1A)-H_6 (**2**; MBP = maltose-binding protein) following recombinant expression in *E. coli* and introduced the required fluorophore by chemoselective bioconjugation of the CysTag sequence of **1** with Alexa647-maleimide to get **1a** (Figure 2A). An equilibrium dissociation constant K_d of 22.4 ± 1.8 nM was determined for the reversible association of the intein fragments (Figure 2B).

We then aimed to deconvolve association, dissociation and splicing kinetics. To this end, we monitored the Gp41-1 intein reaction in real time using surface-sensitive detection by total internal reflection fluorescence spectroscopy coupled with reflectance interferometry (TIRFS/RIf).[25] First, we investigated the complementation of inactivated split intein fragments. For this purpose, we immobilized H_6-Int^N(C1A)-mCherry (**3**) via the His-tag on a tris-NTA functionalized biocompatible PEG-layer on glass surfaces and measured its interaction with SBP-Int^C(N125A)-eGFP (**4**, SBP = streptavidin binding peptide) (Figure 2C). Analyzing the RIf signal, which reflected the binding and dissociation of **4**, we measured association (k_a) and dissociation (k_d) rate constants of $(8.5 \pm 0.8) \times 10^4$ M^{-1}s^{-1} and $(8.2 \pm 0.8) \times 10^{-4}$ s^{-1} respectively, corresponding to a K_d of (9.2 ± 1.3) nM (Figure 2D). This value compares well with the affinity that was obtained from MST (Figure 2A and B). A K_d of 10–20 nM is comparable to the affinity of the split *Ssp* DnaE intein parts,[26] about 10-fold higher than reported for the *Npu* DnaE intein,[27] and 10 to 35-fold lower than reported for the split M86 DnaB intein.[11e,13a] Interestingly, the k_a of the Gp41-1 intein is ≈1,500-fold higher than the k_a of the M86 DnaB intein, but ≈10-fold and ≈300-fold lower than observed for the *Npu* and *Ssp* DnaE inteins, respectively.

To investigate the combined association and splicing kinetics we used the splice-active fragments of the Gp41-1 intein. We first immobilized H_6-Int^N-mCherry (**5**) and injected SBP-Int^C-eGFP (**6**) onto the surface (Figure 3A).

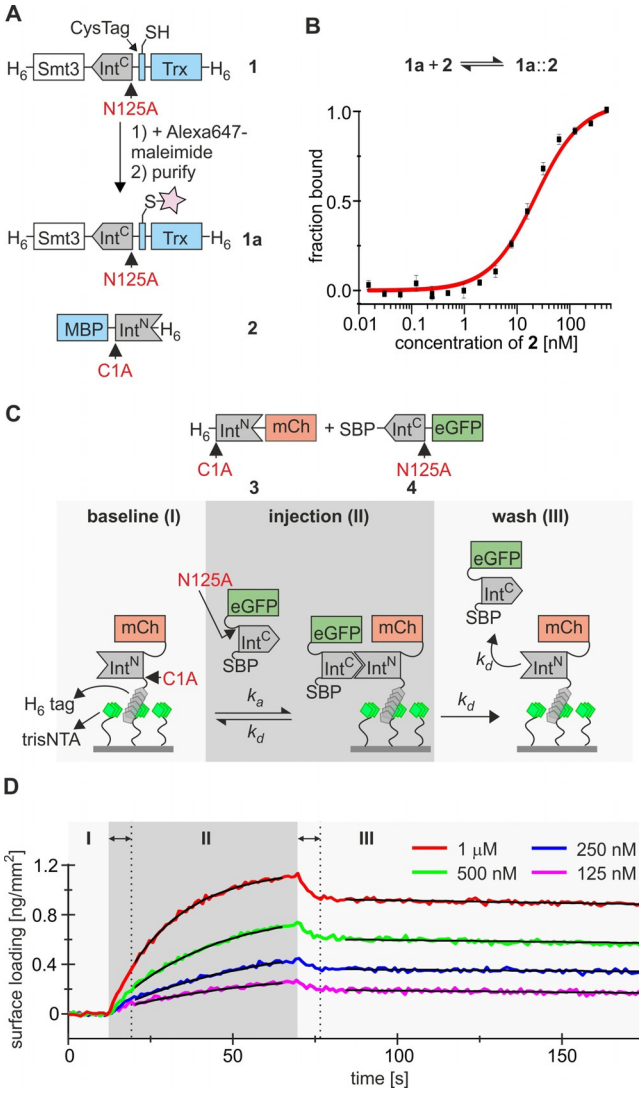

Figure 2. Kinetic and thermodynamic analysis of the interaction between Int^N and Int^C. A) Proteins used for the MST assays. B) Affinity curve for the interaction between Int^N and Int^C measured via MST. 0.25 nM of **1a** and varying concentrations of **2** were used. The data points represent averages and standard deviations obtained from five experiments. C) Schematic representation of the TIRFS/RIf interaction assay. D) RIf curves showing the mass signal for the binding and dissociation of **4** from surface immobilized **3**. The black curves show fits of the data. Initial time periods of the injection and wash phases were not used for fitting (indicated with double-headed arrows).

The RIf curves now showed distinctly different characteristics with the formation of a plateau and subsequent decrease in signal during the injection phase (Figure 3B). This observation is consistent with two events in the injection phase, namely association of **6** with surface-immobilized **5**, which leads to an increase in signal intensity, and splicing of the C-extein eGFP to the surface-immobilized His-tag with a concomitant loss of Int^N-mCherry, which leads to a decrease in signal (Figure 3A). In order to separate these two events, we analyzed the mCherry-derived fluorescence signal, which, following an initial lag period in the injection phase, displayed a significant decrease that continued during the wash phase of

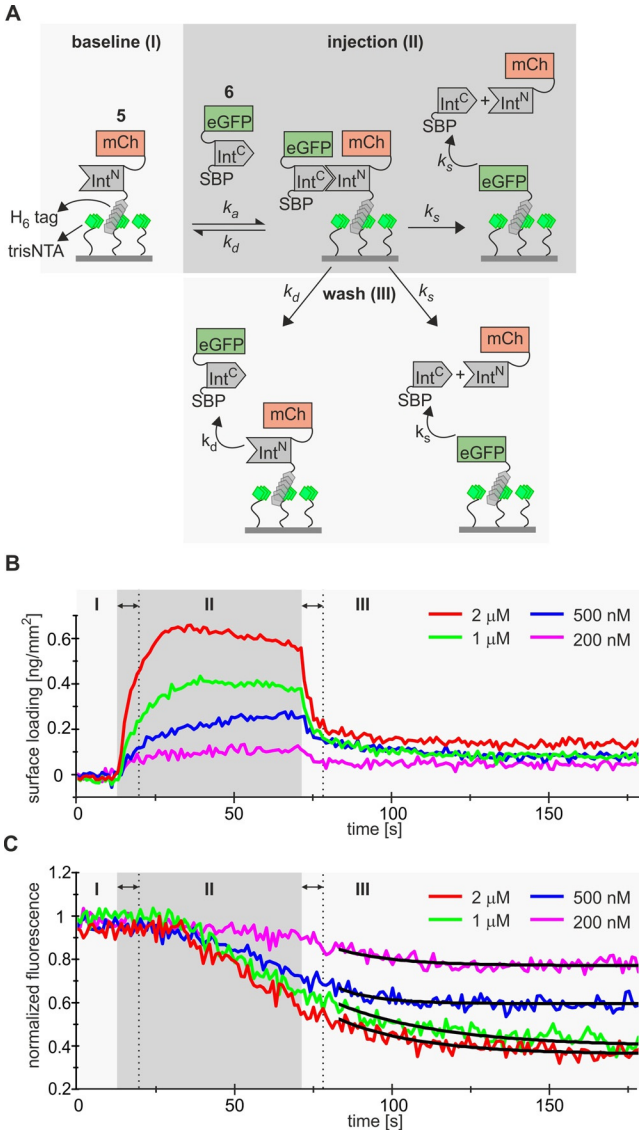

Figure 3. Analysis of the splicing kinetics of the Gp41-1 intein. A) Schematic representation of the TIRFS/RIf splicing assay. B) RIf curves showing the mass signal, which is a convolution between interaction and splicing. C) Normalized mCherry fluorescence intensity curves, which selectively reflect the splicing reaction. The black curves show fits of the data in the wash phase (III).

the experiment (Figure 3C). The Gp41-1 intein is believed to follow a "capture and collapse" mechanism, whereby the individual intein fragments contain extensive disordered regions that first bind and then undergo folding into splice-active conformations.[27b,28] In keeping with this hypothesis, the initial lag period in the injection phase before decrease of the mCherry signal is observed would reflect the time required for these folding and conformational changes to occur. To quantitatively determine splicing kinetics with this method, we analyzed the decrease in the fluorescence signal in the wash period, which selectively reflects the loss of mCherry from the surface due to splicing. This analysis afforded a splicing rate constant of $0.05 \pm 0.004\ s^{-1}$, which perfectly matches the value previously obtained for this intein at 25 °C[21b] and further corroborates the ultra-fast splicing behavior.

Together, the high k_{on} and the high affinity of the Gp41-1 intein in combination with its ultra-fast and highly efficient protein trans-splicing activity fully supports its potential for the intended in cellulo semi-synthesis approaches.

Chemical Tag Modification in Vitro and in Cellulo Using the Split Gp41-1 Intein

We next investigated the utility of the Gp41-1 intein for site-specific chemical labeling of target proteins. We fused the Int^C fragment with a short extein tag containing a unique cysteine (CysTag) to be labeled via classical thiol bioconjugation. Purified SBP-Smt3-Int^C-CysTag (**7**) was conjugated with fluorescein-maleimide to give **7a** (Figure S2A). A splice assay with MBP-Int^N-H_6 showed that **7a** could be completely consumed within one minute and resulted in the transfer of the fluorescently labeled tag to MBP (Figure S2B). Similarly, we modified an anti-EGFR nanobody (V_HH) with fluorescein and showed its specific binding in a cellular assay (Figure S2C–E).

We then explored the Gp41-1 intein for labeling proteins inside live mammalian cells by in cellulo protein semi-synthesis. Intracellular splicing of two co-expressed proteins containing the Int^N and Int^C fragments has been reported before[29] and is a standard application of split inteins that underlines their selective recognition in the cellular milieu. To recapitulate such an entirely endogenous process, we transfected HeLa cells with two plasmids encoding Tom20-Int^N-mCherry-NLS (**8**) and Int^C-eGFP (Figure S3; NLS = nuclear localization sequence). Tom20 is a transmembrane protein that localizes in the outer membrane of mitochondria with the fused Int^N part facing the cytoplasm. We observed protein trans-splicing, by which eGFP became attached to the mitochondria, while Int^N-mCherry-NLS got released from Tom20 and changed its localization to the nucleus, as expected (Figure S3).

We then exogenously prepared the Int^C precursor by expression in *E. coli* and subsequent chemical modification. Transduction into mammalian cells was explored using bead-loading, which relies on a mechanical procedure by which small and transient holes are introduced into the plasma membrane by sheering forces exerted with small glass beads.[30] We first bead-loaded untransfected HeLa cells with the fluorescein-modified construct **7a**. We observed that the procedure led to the detachment of around 50% of the cells from the culture plate. Despite this, around 30% of the remaining cells were transduced with the protein (Figure S4). Furthermore, observation of the cells via bright field and fluorescence microscopy for durations of up to 5 h after bead loading showed that the majority of the cells did not demonstrate any difference in appearance when compared to non-bead loaded cells, thus indicating that bead-loading did not lead to significant loss of cell viability. We next transduced **7a** into cells transfected with a plasmid encoding for **8**. Different outcomes in single cells were observed by fluorescence microscopy depending on whether or not a cell was

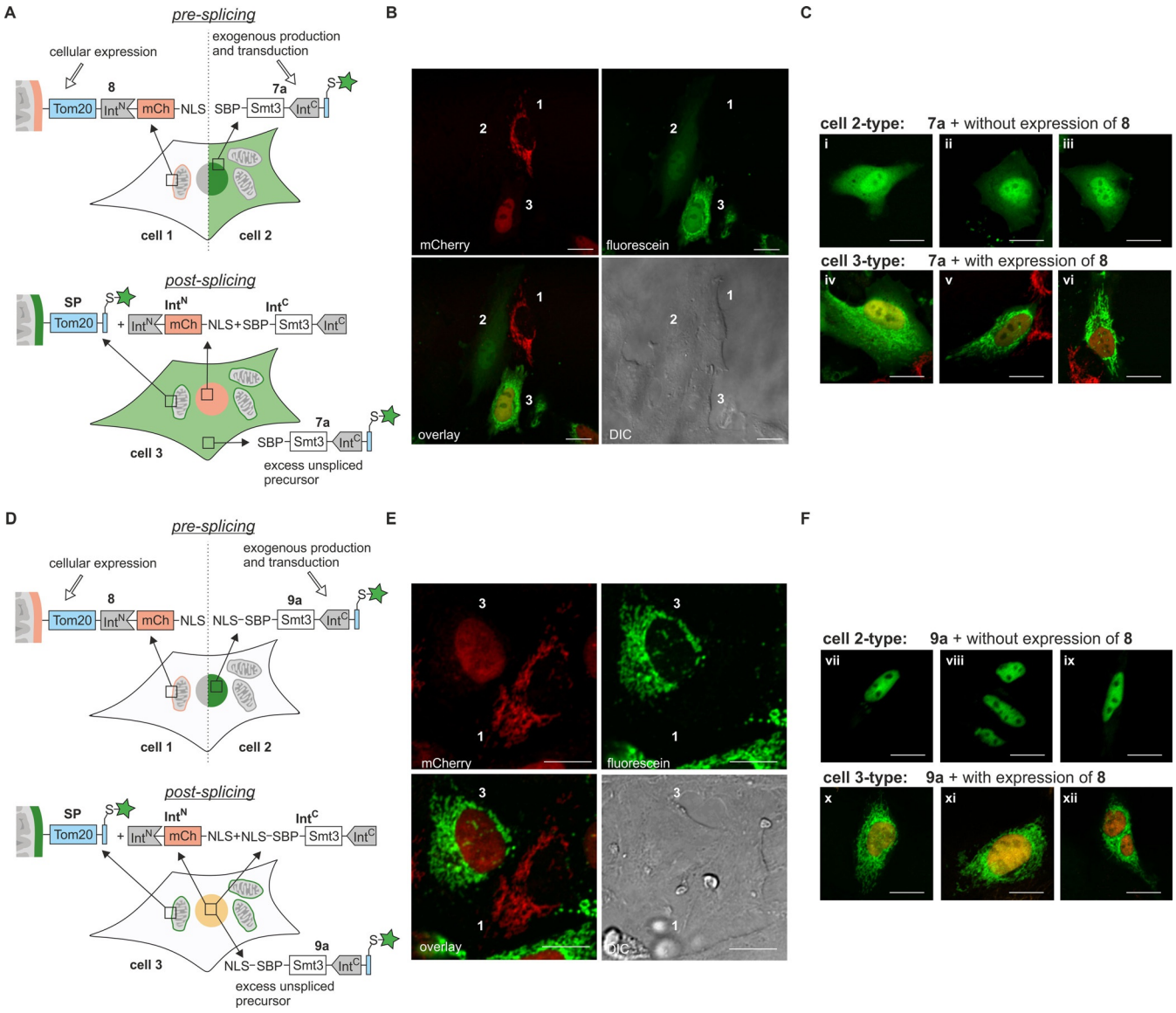

Figure 4. C-terminal labeling of Tom20 with fluorescein in HeLa cells using the split Gp41-1 intein. A) Schematic representation of the protocol for labeling Tom20 using construct **7a**. B,C) Representative microscopy images according to Scheme shown in (A) representing different cell types: Cell 1 expresses **8** but is not transduced with **7a**; cell 2 does not express **8** but is transduced with **7a**; and cell 3 expresses **8** and is transduced with **7a**. Panels i–iii and iv–vi show more examples as discussed in the main text. D) Schematic representation of the protocol for labeling Tom20 using construct **9a**. Construct **9a** is like construct **7a** but carries an NLS to direct the precursor **9a** and its remaining intein fragment after splicing, NLS-SBP-Int^C, into the nucleus. E,F) Representative microscopy images according to scheme shown in (D) representing different cell types: Cell 1 expresses **8** but is not transduced with **9a**; cell 2 does not express **8** but is transduced with **9a**; and cell 3 expresses **8** and is transduced with **9a**. Panels vii–ix and x–xii show more examples as discussed in the main text.

transduced with **7a** and/or expressing **8**, i.e., contained just one or both splice partners (Figure 4A, cells 1–3), and depending on the relative intracellular amounts of the proteins. We could observe significant localization of the fluorescein signal on the mitochondria and a concomitant restriction of the mCherry signal to the nucleus via fluorescence microscopy (Figure 4B, cell 3, Figure S5), clearly supporting the notion that protein trans-splicing had occurred. Significantly, in cells that were fixed only 30 min after transduction of **7a** the complete lack of any mCherry signal on the mitochondria could be observed (Figure 4B, cell 3, Figure S5), suggesting the complete consumption of **8** in the splicing reaction is possible when the Int^C precursor **7a** is present at equimolar or excess concentrations. Similar observations were made in more than 95% of the cells (>50 cells) that expressed **8** and were also transduced with **7a**, thus confirming that close to complete intracellular splicing efficiency could be achieved under these conditions (see Figure S6 for more examples). Residual fluorescein signal could be detected in the cytosol and nucleus of cell 3 in Figure 4B, indicating the stoichiometric excess of **7a** over **8**. Cells expressing **8**, but that had not been transduced with **7a** (Figure 4B, cell 1), showed a markedly different outcome, as expected, with the mCherry signal still being localized

exclusively on the mitochondria. Finally, in cells that did not express **8**, but were transduced with **7a**, the fluorescein signal of **7a** could be observed in the cytosol and nucleus without any mitochondrial accumulation, as expected (Figure 4B, cell 2). Further representative confocal microscopy images of such cells are shown in Figure 4C (i–iii), where a similar cytosolic and nuclear accumulation of **7a** could be observed (see Figure S7A–F for cross-sectional quantification). Importantly, the background cytosolic fluorescence derived from unspliced **7a** in cells expressing **8** (as seen in cell 3 in Figure 4B), was strongly dependent on the relative ratio of the two proteins as observed in Figure 4C (iv–vi), with significant cytosolic fluorescence at higher abundance of **7a** (Figure 4C (iv), Figure S7G&H), decreased fluorescence as the relative abundance of **7a** decreases (Figure 4C (v), Figure S7I&J), and finally with negligible fluorescence when the amount of **7a** was almost equal to or lesser than the amount of **8** (Figure 4C (vi), Figure S7K&L). We also confirmed splicing using SDS-PAGE analysis with either in-gel fluorescence or immune detection (Figure S8).

We further demonstrated the application of this approach for proteins localized to other organelles besides the mitochondria, namely the endoplasmic reticulum (Figure S9). We next established a similar strategy for cellular labeling of proteins of interest on their N terminus by switching the endogenous and exogenous origins of the Int^N and Int^C precursors (Figure S10 and S11). The essential catalytic cysteine at the N-terminal splice junction of the Gp41-1 Int^N fragment[31] prohibited the use of the same simple CysTag strategy. We therefore resorted to a chemical conjugation for fluorophore attachment based on copper catalyzed alkyne-azide cycloaddition[32] of a short ClickTag[15b] sequence on the Int^N precursor containing the unnatural amino acid p-azidophenylalanine (Figure S10 and S11).[33]

A Compartmental Entrapment Strategy to Reduce Background from Unspliced Intein Precursor

We then aimed to improve the signal background stemming from the excess or unconsumed fraction of the labeled and transduced intein fragment precursor. This can be problematic even with the very efficiently splicing Gp41-1 intein, because it is difficult to control the amount of the exogenously transduced protein partner and because this protein cannot be washed out of the cells. In order to address this challenge, we conceived a novel nuclear entrapment strategy by recombinantly fusing a nuclear localization sequence (NLS) to the exogenously prepared Int^C-fusion construct (Figure 1, right half). The idea was to transfer the transduced Int^C precursor into the nucleus by using the nuclear import machinery and thereby eliminate or reduce the background signal of the attached fluorophore in the cytosol. Only rapid splicing with the Tom20-Int^N partner on the surface of mitochondria, that could compete with the kinetics of transport into the nucleus, would result in delivering the fluorophore to the target protein (Figure 1, right half). Thus, we added an NLS sequence to the Int^C precursor to give NLS-SBP-Smt3-Int^C-CysTag (**9**) and labeled the CysTag with fluorescein-maleimide to get **9a** (Figure 4D). Again, three possible outcomes with either one of the two or both splice partners present in a single cell were possible (Figure 4D, cells 1–3). Upon transduction of **9a** into HeLa cells not expressing **8** we could observe fluorescence signal exclusively in the nucleus, as intended (cell 2-type in Figure 4D&F). A comparison of cells in panels vii–ix of Figure 4F with those in panels i–iii of Figure 4C shows the efficient entrapment of the Int^C precursor in the nucleus (see Figure S12A–F for quantification). When **9a** was transduced into cells expressing Tom20-Int^N-mCherry-NLS (**8**), we observed the expected localization of fluorescein signal on the mitochondria and mCherry fluorescence in the nucleus (cell 3-type in Figure 4D and Figure 4F panels x–xii), which confirmed the occurrence of splicing. Importantly, the mitochondria-to-cytosol contrast was strongly improved compared to our earlier results using **7a**, which lacked the NLS (compare panels x–xii in Figure 4F with panels iv–vi in Figure 4C, as well as the respective cross-sectional quantification in Figure S12 and S7), indicating the successful establishment of the nuclear entrapment strategy for excess **9a**. Unreacted **9a** could also be detected in the nucleus of cells when it was present in a stoichiometric excess over **8** (Figure 4F x–xii, Figure S12G–L). Another important observation was the fact that we could only detect negligible amounts of residual mCherry signal on the mitochondria (Figure 4E, cell 3, Figure S13), indicating virtually quantitative splicing and thereby also confirming that the nuclear import driven removal of **9a** from the cytosol did not negatively impact the extent of splicing. We could observe similarly high intracellular splicing yields in more than 95% of the cells (>50 cells analyzed) that expressed **8** and also were transduced with **9a** (see Figure S14 for more examples). These findings highlight the advantages of the extremely fast kinetics of the association and splicing of the split Gp41-1 intein, which ensured that splicing occurred before the removal and nuclear entrapment of the precursor by cellular importins.

Application of in Cellulo Protein Semi-Synthesis with the Gp41-1 Intein for Single-Molecule Fluorescence Microscopy

We next applied our nuclear entrapment-based labeling strategy with minimal fluorescence background for two different single-molecule fluorescence microscopy methods, namely super-resolution microscopy using dSTORM[34] and single-particle tracking (SPT),[35] to investigate localization and diffusion of the integral mitochondrial outer membrane protein Tom20. For both these applications, we first prepared NLS-SBP-Smt3-Int^C-CysTag(Alexa647) (**9b**) by reacting the CysTag on **9** with Alexa647-maleimide, and transduced it into HeLa cells expressing **8** to label Tom20 with Alexa647. Alexa647 is a photostable fluorophore that has been extensively used for dSTORM due to its superior photoswitching properties.[36] We acquired epifluorescence and dSTORM images of a mitochondrial network with the expected increase in resolution that is obtained via dSTORM (Figure 5A). The localization of Tom20 on the mitochondrial membrane is

evident by the accumulation of signal on the periphery of the mitochondrion. The intensity profiles (Figure 5B) over the entire cross section denoted by the white boxes shown in Figure 5A allowed us to measure a mitochondrial width of ≈325 nm, which compares very well with previous super-resolution and electron microscopy analyses.[37] Using SPT, we could observe diffusion of single AF647-labeled Tom20 molecules (Video S1) on the mitochondrial membrane and reconstruct the tracks to get the trajectory map shown in Figure 5C. Quantitative analysis of the diffusion dynamics yielded a distribution of jump distances (Figure S15) that could be fitted by with a two states model,[38] resulting in a fast diffusion fraction ($D = 0.49\ \mu m^2 s^{-1}$, 81%) and a slow diffusing fraction ($D = 0.02\ \mu m^2 s^{-1}$, 19%). These diffusion constants closely match previously reported values.[39]

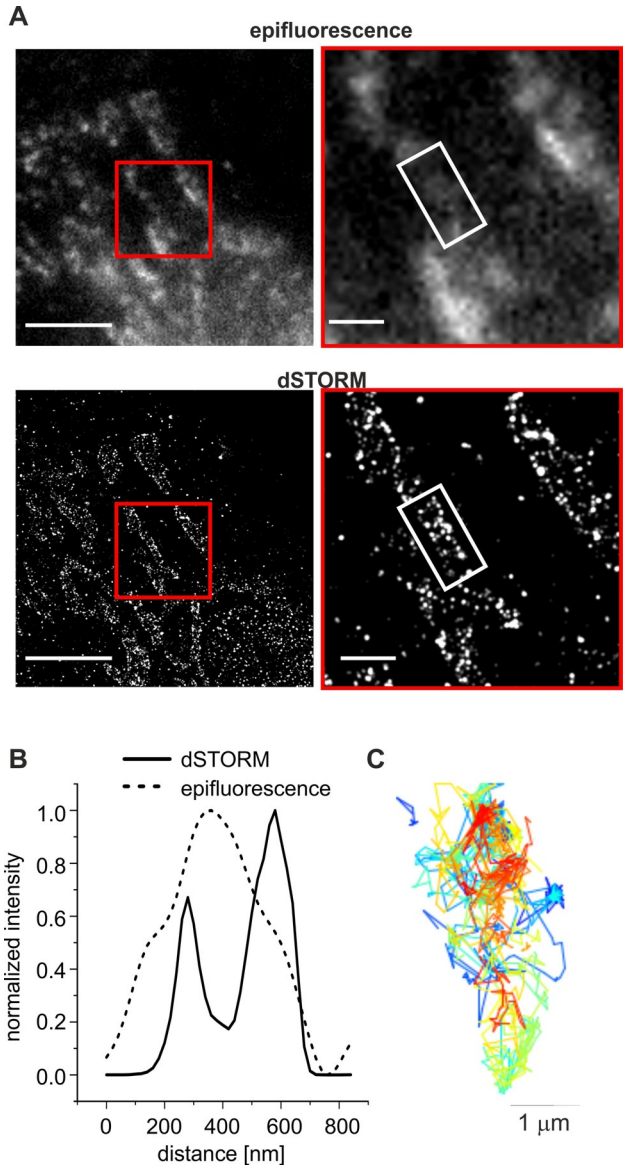

Figure 5. Single-molecule fluorescence microscopy of Tom20 labeled using the Gp41-1 intein. A) Diffraction-limited epifluorescence (top) and super-resolved dSTORM (bottom) images of Alexa647-labeled Tom20. The panels on the right show zooms of the regions marked by red boxes in the left panels. Scale bars represent 5 µm for the images on the left, and 1 µm for the zoomed images on the right. B) Normalized intensity profiles over the white boxes marked in the panels on the right in A. C) Tracks obtained via SPT of Alexa647-labeled Tom20 in a live HeLa cell. 5000 frames were collected at a rate of 32 ms/frame for this analysis.

Together, these experiments demonstrate the suitability of our approach for labeling intracellular proteins for single molecule fluorescence based techniques. The low background achieved using the nuclear entrapment strategy, in addition to being crucial for minimizing fluorescence background levels, also eliminates the need for washing steps to remove excess fluorophore, thus simplifying the experimental process. Moreover, currently existing methods for intracellular protein labeling that are used for SPT are limited to use with membrane permeable fluorophores that can be efficiently washed out post-labeling. This prohibits the use of a wide variety of non-membrane permeable photostable dyes (e.g. Alexa647). Thus, our split-intein-mediated technique can mitigate this limitation by extending the dye palette that is available for live cell applications.

Conclusion

In cellulo protein semi-synthesis holds great potential to study proteins and their biological pathways in ways that are out of reach with other technologies. To expand the very limited toolbox to assemble protein backbones from endogenously and exogenously provided parts in live cells we established the currently fastest splicing split intein for this purpose. We carried out detailed biophysical characterization of the Gp41-1 intein fragment complementation and trans-splicing kinetics that unraveled favorably high affinity and rapid association kinetics of the two parts. Following protein delivery of one exogenously prepared intein fragment precursor inside mammalian cells using a bead-loading protocol, we ligated proteins located in or facing the cytoplasm with chemically modified tags at their N or C termini. Using a nuclear entrapment strategy, we were able to achieve almost background free labeling of Tom20 on mitochondria, which allowed us to acquire super-resolution images of this protein, as well as to perform single particle tracking on it. For applications in which fluorescent labeling of nuclear proteins is desired, or where the translocation of the precursor into the nucleus could affect the biological process under study, we envision an extension of our entrapment strategy that aims to direct the excess intein fragment precursor to another localization in the cell, for example to the inner leaflet of the plasma membrane or the mitochondria, by exchanging the NLS sequence with a lipidation or a mitochondrial anchor sequence, respectively. An exchange with a degron sequence tag would trigger proteasomal removal of the intein precursor that did not splice in due course and appears promising for endeavors to install native sequence segments of a protein, including post-translational modifications. Notably, the orthogonality of the Gp41-1 intein to other split inteins,[21b] like the split DnaE inteins that was previously explored for in cellulo semi-synthesis, could be exploited for multiplexed intracellular manipulation of proteins in the future. In

conclusion, the developments described here should greatly expand the range of applications of split-intein based protein modification for cell biological applications.

Acknowledgements

We gratefully acknowledge financial support by the DFG (SPP1623, MO1073/5-2 to H.D.M. and Cluster-of-excellence EXC1003 'Cells in Motion', FF project 2017-07 to K.B.B. and H.D.M.) and the Alexander von Humboldt Foundation (postdoctoral fellowship to M.B.). We thank Timo Dellmann for help with SPT data acquisition and the iBiOs Facility (PI 405/14-1) for help with dSTORM data acquisition and analysis (Dr. Rainer Kurre), and Christian P. Richter for developing softwares for analysis of SPT data. Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: dSTORM · protein splicing · protein transduction · single-molecule studies · split intein

[1] a) T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249 – 289; b) S. B. H. Kent, *Protein Sci.* **2019**, *28*, 313 – 328.

[2] a) E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, *48*, 6974 – 6998; *Angew. Chem.* **2009**, *121*, 7108 – 7133; b) D. D. Young, P. G. Schultz, *ACS Chem. Biol.* **2018**, *13*, 854 – 870.

[3] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779.

[4] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705 – 6710.

[5] S. Chattopadhaya, R. Srinivasan, D. S. Yeo, G. Y. Chen, S. Q. Yao, *Bioorg. Med. Chem.* **2009**, *17*, 981 – 989.

[6] V. J. Bruce, B. R. McNaughton, *Cell Chem. Biol.* **2017**, *24*, 924 – 934.

[7] a) C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem. Int. Ed.* **2000**, *39*, 450 – 466; *Angew. Chem.* **2000**, *112*, 458 – 476; b) O. Novikova, N. Topilina, M. Belfort, *J. Biol. Chem.* **2014**, *289*, 14490 – 14497.

[8] a) H. D. Mootz, *ChemBioChem* **2009**, *10*, 2579 – 2589; b) N. H. Shah, T. W. Muir, *Chem. Sci.* **2014**, *5*, 446 – 461.

[9] a) M. Fottner, A. D. Brunner, V. Bittl, D. Horn-Ghetko, A. Jussupow, V. R. I. Kaila, A. Bremm, K. Lang, *Nat. Chem. Biol.* **2019**, *15*, 276 – 284; b) J. E. Glasgow, M. L. Salit, J. R. Cochran, *J. Am. Chem. Soc.* **2016**, *138*, 7496 – 7499.

[10] X. Bi, J. Yin, G. K. T. Nguyen, C. Rao, N. B. A. Halim, X. Hemu, J. P. Tam, C. F. Liu, *Angew. Chem. Int. Ed.* **2017**, *56*, 7822 – 7825; *Angew. Chem.* **2017**, *129*, 7930 – 7933.

[11] a) I. Giriat, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 7180 – 7181; b) R. Borra, D. Dong, A. Y. Elnagar, G. A. Woldemariam, J. A. Camarero, *J. Am. Chem. Soc.* **2012**, *134*, 6344 – 6353; c) D. Jung, K. Sato, K. Min, A. Shigenaga, J. Jung, A. Otaka, Y. Kwon, *Chem. Commun.* **2015**, *51*, 9670 – 9673; d) Y. David, M. Vila-Perello, S. Verma, T. W. Muir, *Nat. Chem.* **2015**, *7*, 394 – 402; e) M. Braner, A. Kollmannsperger, R. Wieneke, R. Tampe, *Chem. Sci.* **2016**, *7*, 2646 – 2652.

[12] J. Zettler, V. Schütz, H. D. Mootz, *FEBS Lett.* **2009**, *583*, 909 – 914.

[13] a) J. H. Appleby-Tagoe, I. V. Thiel, Y. Wang, Y. Wang, H. D. Mootz, X. Q. Liu, *J. Biol. Chem.* **2011**, *286*, 34440 – 34447; b) J. C. J. Matern, K. Friedel, J. Binschik, K. S. Becher, Z. Yilmaz, H. D. Mootz, *J. Am. Chem. Soc.* **2018**, *140*, 11267 – 11275.

[14] a) K. K. Khoo, I. Galleano, F. Gasparri, R. Wieneke, H. Harms, M. H. Poulsen, H. C. Chua, M. Wulf, R. Tampe, S. A. Pless, *Nat. Commun.* **2020**, *11*, 2284; b) A. J. Burton, M. Haugbro, E. Parisi, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 12041 – 12049.

[15] a) T. Kurpiers, H. D. Mootz, *Angew. Chem. Int. Ed.* **2007**, *46*, 5234 – 5237; *Angew. Chem.* **2007**, *119*, 5327 – 5330; b) V. Schütz, H. D. Mootz, *Angew. Chem. Int. Ed.* **2014**, *53*, 4113 – 4117; *Angew. Chem.* **2014**, *126*, 4197 – 4201.

[16] A. Wasmuth, C. Ludwig, H. D. Mootz, *Bioorg. Med. Chem.* **2013**, *21*, 3495 – 3503.

[17] J. Lotze, U. Reinhardt, O. Seitz, A. G. Beck-Sickinger, *Mol. Biosyst.* **2016**, *12*, 1731 – 1745.

[18] S. Hauke, A. von Appen, T. Quidwai, J. Ries, R. Wombacher, *Chem. Sci.* **2017**, *8*, 559 – 566.

[19] H. Li, J. C. Vaughan, *Chem. Rev.* **2018**, *118*, 9412 – 9454.

[20] X. Sun, A. Zhang, B. Baker, L. Sun, A. Howard, J. Buswell, D. Maurel, A. Masharina, K. Johnsson, C. J. Noren, M. Q. Xu, I. R. Correa, Jr., *ChemBioChem* **2011**, *12*, 2217 – 2226.

[21] a) B. Dassa, N. London, B. L. Stoddard, O. Schueler-Furman, S. Pietrokovski, *Nucleic Acids Res.* **2009**, *37*, 2560 – 2573; b) P. Carvajal-Vallejos, R. Pallisse, H. D. Mootz, S. R. Schmidt, *J. Biol. Chem.* **2012**, *287*, 28686 – 28696; c) J. K. Böcker, W. Dörner, H. D. Mootz, *Chem. Commun.* **2019**, *55*, 1287 – 1290.

[22] H. Iwai, S. Zuger, J. Jin, P. H. Tam, *FEBS Lett.* **2006**, *580*, 1853 – 1858.

[23] G. Volkmann, H. D. Mootz, *Cell. Mol. Life Sci.* **2013**, *70*, 1185 – 1206.

[24] a) M. Bhagawati, T. M. E. Terhorst, F. Fusser, S. Hoffmann, T. Pasch, S. Pietrokovski, H. D. Mootz, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22164 – 22172; b) T. Kurpiers, H. D. Mootz, *ChemBioChem* **2008**, *9*, 2317 – 2325; c) A. L. Bachmann, H. D. Mootz, *J. Pept. Sci.* **2017**, *23*, 624 – 630.

[25] M. Gavutis, S. Lata, P. Lamken, P. Muller, J. Piehler, *Biophys. J.* **2005**, *88*, 4289 – 4302.

[26] J. Shi, T. W. Muir, *J. Am. Chem. Soc.* **2005**, *127*, 6198 – 6206.

[27] a) N. H. Shah, M. Vila-Perello, T. W. Muir, *Angew. Chem. Int. Ed.* **2011**, *50*, 6511 – 6515; *Angew. Chem.* **2011**, *123*, 6641 – 6645; b) N. H. Shah, E. Eryilmaz, D. Cowburn, T. W. Muir, *J. Am. Chem. Soc.* **2013**, *135*, 18673 – 18681.

[28] a) J. A. Gramespacher, A. J. Stevens, D. P. Nguyen, J. W. Chin, T. W. Muir, *J. Am. Chem. Soc.* **2017**, *139*, 8074 – 8077; b) H. Michael Beyer, K. Malgorzata Mikula, M. Li, A. Wlodawer, H. Iwai, *FEBS J.* **2020**, *287*, 1886 – 1898.

[29] H. Wang, J. Liu, K. P. Yuet, A. J. Hill, P. W. Sternberg, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3900 – 3905.

[30] a) P. L. McNeil, E. Warder, *J. Cell Sci.* **1987**, *88 (Pt 5)*, 669 – 678; b) D. Humphrey, Z. Rajfur, B. Imperiali, G. Marriott, P. Roy, K. Jacobson, *CSH Protoc.* **2007**, *2007*, pdb prot4658.

[31] A. L. Bachmann, H. D. Mootz, *J. Biol. Chem.* **2015**, *290*, 28792 – 28804.

[32] a) V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596 – 2599; *Angew. Chem.* **2002**, *114*, 2708 – 2711; b) C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057 – 3064.

[33] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026 – 9027.

[34] S. J. Sahl, S. W. Hell, S. Jakobs, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 685 – 701.

[35] F. R. Beinlich, C. Drees, J. Piehler, K. B. Busch, *ACS Chem. Biol.* **2015**, *10*, 1970 – 1976.

[36] G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, X. Zhuang, *Nat. Methods* **2011**, *8*, 1027 – 1036.

[37] A. Dlasková, H. Engstová, T. Špaček, A. Kahancová, V. Pavluch, K. Smolková, J. Špačková, M. Bartoš, L. P. Hlavatá, P. Ježek, *Biochim. Biophys. Acta Bioenerg.* **2018**, *1859*, 829 – 844.
[38] A. S. Hansen, M. Woringer, J. B. Grimm, L. D. Lavis, R. Tjian, X. Darzacq, *eLife* **2018**, *7*, e331250.
[39] T. Appelhans, C. P. Richter, V. Wilkens, S. T. Hess, J. Piehler, K. B. Busch, *Nano Lett.* **2012**, *12*, 610 – 616.

Manuscript received: May 12, 2020
Revised manuscript received: July 15, 2020
Accepted manuscript online: August 10, 2020
Version of record online: September 11, 2020

