Communication

A bio-orthogonal and covalent 5 kDa small protein tag

Ulrich Pabst[1,#], Jana Rossius[2,#], Christina Holmboe Olesen[1,#], Melissa Birol[2,*,$], Johannes Broichhagen[1,*,$]

[1] Leibniz-Forschungsinstitut für Molekulare Pharmakologie, 13125 Berlin, Germany.
[2] Berlin Institute of Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, 10115 Berlin, Germany.

[#] Equal contribution
[$] Co-leads
* Correspondence should be addressed to:
Melissa.birol@mdc-berlin.de and broichhagen@fmp-berlin.de

ABSTRACT

Precise and minimally perturbative protein labelling remains a key challenge for studying biomolecular function in living systems. Here, a minimal way to specifically label proteins based on SNAP-tag and intein-mediated protein splicing reaction is introduced. Termed CLUSTER (for Chemical Label-Unfold-Splice Technology Enables Recombination), this chimeric platform supports efficient labelling across diverse targets in living cells by retaining a fluorescent, 5 kDa sized peptide on a fusion protein of interest after splicing. A bacterial screening workflow was developed to optimize the reaction efficiency and construct design. Quantitative characterization using fluorescence polarization provides mechanistic insight into labelling efficiency and dynamics, while molecular dynamics simulations elucidate its stability, grasping the intricate nature of protein behaviour upon covalent labelling. This bio-orthogonal labelling technology allows for a versatile and minimally invasive approach for protein labelling, providing a powerful tool to probe protein behavior in native cellular systems.

INTRODUCTION

Fluorescence imaging of proteins of interests (POIs) requires specific labeling with chromophores, often achieved via immunohistochemistry or as fusion proteins.[1–3] However, this adds considerable size to the POI, with the disadvantage that it may perturb the proteins native function and behaviour. Antibodies comprise large molecular entities (150 kDa, **Fig. 1A**), substantially exceeding the size of, for instance, their epitopes (HA, 3 kDa), as well as many, transmembrane receptors like GLP1R (55 kDa) and its endogenous ligand GLP1 (4 kDa). Alternatively, genetic fusion to fluorescent proteins (FPs, ~27 kDa) are widely used. However, FPs are inferior in their photophysical properties to organic fluorophores that display augmented brightness and photostability, and which can be specifically and rapidly reacted with self-labeling protein tags (SLPs),[4] like the HaloTag Protein (HTP, 33 kDa) and SNAP-tag (20 kDa). Still comparably large in size, folded tags, however, may disrupt conformations, functions and dynamics of POIs.[5,6] This has been addressed by for instance split enzyme complementation, which has led to improved knock-in strategies[7] and building protein based sensors[8]. Among the pi-clamp system (0.5 kDa) that uses sticky fluorinated substrates[9] that may not be ideal for live cell applications, small covalent tagging systems remain scarce.

RESULTS

Herein, we describe a conceptually new protein-labelling strategy that 1) minimizes tag-size, is 2) controllable in space and time, is 3) bio-orthogonal, and as such 4) applicable in live cells and 5) uses synthetic, bright fluorophores. The approach utilizes the SNAP-tag, which reacts with benzylguanine (BG) substrates, forming a covalent bond on Cys145.[10,11] We inserted the *C*-terminal domain of a *trans*-splicing split intein (gp-41-1^{C})[12] between positions 132/133, just upstream of the reactive cysteine residue (**Fig. 1B**). Upon SNAP-labelling, gp-41-1^{C} will become accessible to its natural interaction partner, gp-41-1^{N}, that resides on the *N*-terminus of SNAP. When the stoichiometric split parts meet, they splice the fluorophore containing peptide (47 amino acids, i.e., 5.1 kDa, excluding cargo) onto the POI.

This engineered system is termed CLUSTER, standing for Chemical Label-Unfold-Splice Technology Enables Recombination. We subcloned the first design (CLUSTER$_{238}$) replacing the SNAP-tag in our reported SNAP-TM-HTP plasmid[13] (TM = transmembrane domain). We first transfected HEK293T cells with the parent SNAP-TM-HTP and observed clean HTP and SNAP labelling using bright TMR-d12-HTL (HTL = HaloTag Ligand) and BG-SiR-d12 (ref[14])

(BG = O^6-benyzlgunaine, i.e. SNAP-tag substrate), respectively (**Fig 1C**). Substitution with CLUSTER$_{238}$–TM–HTP in a similar experiment, yielded detectable dual TMR and SiR signals, albeit at reduced intensity. To assess performance of CLUSTER on intrinsically disordered proteins, we next examined the Tau protein as a benchmark, a microtubule-associated protein implicated in Alzheimer's Disease and tauopathies[15], which shows complex behaviour such as phase separation[16]. We transfected HEK293T cells with Tau-SNAP or Tau-CLUSTER$_{238}$ and labelled both samples successfully with BG-TMR-d12 for subsequent confocal microscopy (**Fig 1D**). While fluorescence signals were recorded for both instances, we harvested the cells and loaded them on SDS-PAGE, for which a spliced product was detectable exclusively for Tau-CLUSTER$_{238}$, with an intein-dead mutant (DM) serving as control (**Fig. 1E**, and see SI). This demonstrates that 1) CLUSTER can be be expressed in mammalian system and 2) being labelled with SNAP-tag substrates, both 3) intra- and extracellularly, and 4) is amenable to intein-driven splicing after it has reacted with a substrate.

We next created a construct (CLUSTER$_{238}$–HTP) (**Fig. 2A**) as a model system for screening in *E. coli*. BG-SiR-d12 and TMR-d12-HTL were used for labelling SNAP and HTP, respectively, the latter for expression control and normalization before cells were lysed and subjected to SDS-PAGE. We successfully observed tagged proteins under a fluorescent scanner, both spliced (lower band) and non-spliced (upper band) (**Fig. 2B**). Encouraged by this, we performed a glycine linker deletion screen (i.e., G95, G96, G232, G270 and combinations thereof) that reside as linkers for gp-41-1^N (G95/96) and flank the inserted gp-41-1^C (G232/270) to find mutant CLUSTER$_{277}$ working binary, i.e. where SiR-signals are only detectable at the lower molecular weight band (**Fig. 2B**). This indicates that labelling occurs on the full-length protein before splicing. To gain a deeper insight into the mechanism, we performed kinetic measurements by fluorescence polarization on recombinantly expressed SNAP, CLUSTER$_{277}$–HTP and SNAP^N–gp-41-1^C–SNAP^C (i.e. CLUSTER$_{340}$) ± gp41-1^N–HTP and when exposed to BG-TMR-d12. While SNAP showed kinetics towards full labelling within minutes (k_{obs} = 22.2 mHz) (**Fig. 2C**), CLUSTER$_{277}$–HTP gave a rise in polarization (k_{obs} = 0.932 mHz), after which a drop occurred, displaying complex parallel reactions of labelling and splicing (**Fig. 2D**). To gain a better understanding of this complexity, we tested CLUSTER$_{340}$ that showed a steady increase in polarization due to covalent reaction, however, with slower kinetics (k_{obs} = 0.721 mHz), reaching a plateau after approximately 2 hours (**Fig. 2E**). As the signal remained stable, we added gp41-1^N–HTP, which lead to a decrease in polarization indicating 1) successful splicing and, similarly important, 2) dropping polarization values to approximately 55 mP. The drop in polarization can be attributed to the fluorophore's more disordered environment with several more degrees of freedom compared to when bound to folded structures. Importantly, BG-TMR-d12 was unable to react with only the *C*-terminal SNAP^C, as demonstrated by orthogonal control reactions of SNAP^C (synthesized by SPPS). This was achieved with maleimide-linked TMR-d12 (for cysteine reaction) and BG-TMR-d12 (that does not react) and subsequent fluorescence polarization readouts, where the strongly significant observed signal difference to the more reactive maleimide species indicated no significant labelling reactivity (**Fig. S1, Table S1**). Next, we aimed to translate our in vitro observations to live cell imaging, and as such, we fused CLUSTER$_{277}$ to a HaloTag bearing a nuclear localization sequence (NLS) for transfection in HEK293T cells, with HTP–SNAP–NLS (ref[17]) serving as a control (**Fig. 2F**). Indeed, for both cases we observed very good signal overlap from the nuclei, quantified by co-staining (**Fig. S2A**), and SDS-PAGE revealed spliced fragments only in the CLUSTER-expressing cells (**Fig. S2B**). Similar behaviour was observed when we expressed CLUSTER$_{277}$–HTP in the cytosol (**Fig. 2F**).

To further investigate CLUSTER dynamics, as well as to provide evidence for the anticipated mechanism, the *trans*-splicing variant, CLUSTER$_{340}$, was subjected to a series of computational calculations. The sequential reaction of CLUSTER should consist of 1) labelling reaction with the fluorophore, 2) "unfolding" (or, more general, *destabilization*) of the construct, specifically reducing the rigidity of the SNAP domains to enable association of the intein domains, and 3) splicing, which is conducted via the assembled intein fragments. While both the first and last step of this reaction sequence could be directly measured by time-dependent fluorescence polarization as described above, the second step – the intrinsic destabilization of the construct upon covalent reaction with the fluorophore – remains more speculative than evidence-based. Countering this, the post-labelling (and pre-splicing) stability of CLUSTER$_{340}$ was investigated using molecular dynamics (MD) simulations. For this, the protein structure was predicted using Boltz2[18] (root mean squared deviation (RMSD) vs. native SNAP-tag, PDB-6Y8P = 0.54 Å, ref[19]), and the ligand moiety of BG-TMR was covalently docked to the active cysteine residue using GNINA 1.3[20] (**Fig. 3A**). All MD simulations were performed using GROMACS 2025.4 and its included subprograms.[21–31] After relaxation and equilibration, the unlabelled CLUSTER$_{340}$ protein and the labelled CLUSTER$_{340}$:TMR complex were simulated for 1000 ns in isolation, where the trajectories were produced in triplicates. Alignment of the unwrapped trajectories and analysis of the RMSD of all C-α atoms throughout the trajectories reveals a significantly higher overall perturbation introduced upon covalent binding of TMR (**Fig. 3B**). To further elucidate the spatial destabilization of the protein upon binding, the root mean squared flexibility (RMSF) of the same target atoms was calculated across all recorded MD trajectories, and the difference between the unlabelled and labelled construct was compared (**Fig. 3C**). For a more graphic illustration, the calculated average RMSF values were mapped onto the initially predicted protein structure of CLUSTER$_{340}$ (**Fig. 3D**). Apparently, throughout all recorded trajectories, the unlabelled protein consistently exhibits a higher degree of stability and internal retention of the tertiary structure. Comparison of the RMSD traces reveals a stronger deviation from the equilibrated structure for the labelled construct, which could potentially be associated with stronger drive for spatial rearrangement to properly accommodate the covalently attached ligand. Inspection of the RMSF traces also attributes the arrangements primarily to the residues of the SNAP domains and not the circularly inserted *C*-terminal intein, which further demonstrates the geometric perturbation introduced by the ligand. The insignificant difference in the gp41-1^{C} region could be interpreted as an indicator for the retention of intein function and the potential for post-labelling splicing upon association with the respective *N*-terminal fragment. This association is facilitated and enabled by the increased flexibility of the otherwise rigid SNAP domains after labelling, so that premature splicing is potentially disfavored relative to the labelled construct.

DISCUSSION

In summary, this engineered CLUSTER cassette adds a small, fluorescent label at virtually any position (i.e. *N*/*C*-terminally or in between domains) with its strictly stoichiometric reaction. Moreover, the exogenous fluorophore application allows for exceptional temporal control. After splicing, the size (~5.1 kDa + cargo) of CLUSTER is substantially smaller than the HaloTag Protein (33 kDa), SNAP-tag (20 kDa), BromoCatch (12.8 kDa)[32] and FPs (27 kDa).[33] Still bigger than unnatural amino acids,[34] CLUSTER however overcomes other challenges in terms of costs, toxicity, and putative off-target incorporation.[35] In contrast, other methods (e.g. Lap-tag/LplA^{W37V} ref.[36] or TC-tag/ReAsH[37]) offer restricted cargos, or do not covalently react with a dye (e.g. FAST[38], miniVIPER[39]). While the initial construct made use of a *pseudo-cis* splicing

mechanism, leveraging the intramolecular association of originally *trans*-splicing intein domains, a conceptually different approach was demonstrated in CLUSTER$_{340}$. Being originally planned as a diagnostic tool to study the feasibility and discern the contributing reaction rates to the apparent kinetics, CLUSTER$_{340}$ enables true *trans*-splicing of the labelled *C*-terminal SNAP domain onto a POI. By effectively replacing the genetically fused *N*-terminal intein domain with the labelled residue, this strategy mitigates structural perturbations of the POI arising from additional protein bulk. Together, we envision CLUSTER$_{340}$ as a potent labelling agent for biorthogonal and minimally perturbative labelling of accessible protein targets, e.g., cell surface receptors like GLP1R. CLUSTER$_{277}$ demonstrated to function in challenging biological environments like the cytoplasm or inside the mammalian nucleus being fused either *N*- or *C*-terminally. Generally, we envision the utility of CLUSTER to extend to challenging targets and localized spaces, enabling direct observations of small and disordered POIs like tau, synuclein, and GLP-1. The to date limitations remain 1) its comparably slow reaction kinetics (k_{SNAP} ~ 10^4 M^{-1} s^{-1}; k_{HTP} ~ 10^7 M^{-1} s^{-1}) and the 2) not full splicing reaction in mammalian systems. Similar to SNAP- and HaloTags,[40–43] further protein engineering is needed to address these drawbacks, however, our system demonstrates a plug-and-play approach for proteins by merging a self-labelling tag with intein splicing. We anticipate CLUSTER to further demonstrate superior performance in currently impractical experimental setups, and as such, we envision wide adoption of this method across scales in biomedical and life science applications. In addition, we further anticipate that the reduced steric footprint of CLUSTER, combined with the use of bright synthetic fluorophores, will enable advanced fluorescence-based imaging applications. In particular, these properties may facilitate quantitative single-molecule approaches such as fluorescence correlation spectroscopy (FCS) and fluorescence lifetime imaging microscopy (FLIM), where probe size, concentrations and photophysical stability are critical determinants of performance.[44] Moreover, the small tag size and compatibility with high-performance dyes are expected to support super-resolution microscopy, potentially improving localization precision and reducing linkage error compared to conventional protein-based labels.[45]

SUMMARY

In summary, we have designed and cloned a protein chimera consisting of a SNAP-tag in which an *N*-intein was added to its *N*-terminus, while its parent *C*-intein was circularly inserted upstream of the reactive cysteine residue for covalent labelling with BG substrates. While the first system, CLUSTER$_{238}$, was labelled in mammalian cells *N*-terminally on a localized transmembrane protein and C-terminally on a disordered protein (Tau), we optimized the linkers in an *E. coli* screen to yield a variant (CLUSTER$_{277}$) that only showed signals stemming from SNAP-tag substrates in its spliced form. We lastly show the mechanism of reaction by means of fluorescence polarization measurements, including a split version (CLUSTER$_{340}$), for which we lastly performed MD simulations that indicate destabilization after covalent reaction. We anticipate updated CLUSTERs in the future for applications in which small tags are in high demand.

REFERENCES

(1) Xue, L.; Karpenko, I. A.; Hiblot, J.; Johnsson, K. Imaging and Manipulating Proteins in Live Cells through Covalent Labeling. *Nat. Chem. Biol.* **2015**, *11* (12), 917–923. https://doi.org/10.1038/nchembio.1959.

(2) Sahl, S. J.; Hell, S. W.; Jakobs, S. Fluorescence Nanoscopy in Cell Biology. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (11), 685–701. https://doi.org/10.1038/nrm.2017.71.

(3) Porzberg, N.; Gries, K.; Johnsson, K. Exploiting Covalent Chemical Labeling with Self-Labeling Proteins. *Annu Rev Biochem* **2025**, *94* (1), 29–58. https://doi.org/10.1146/annurev-biochem-030222-121016.

(4) 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* **2021**, *60* (33), 2560–2575. https://doi.org/10.1021/acs.biochem.1c00258.

(5) Margolin, W. The Price of Tags in Protein Localization Studies. *J Bacteriol* **2012**, *194* (23), 6369–6371. https://doi.org/10.1128/JB.01640-12.

(6) Leonetti, M. D.; Sekine, S.; Kamiyama, D.; Weissman, J. S.; Huang, B. A Scalable Strategy for High-Throughput GFP Tagging of Endogenous Human Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (25). https://doi.org/10.1073/pnas.1606731113.

(7) Makhija, S.; Brown, D.; Rudlaff, R. M.; Doh, J. K.; Bourke, S.; Wang, Y.; Zhou, S.; Cheloor-Kovilakam, R.; Huang, B. Versatile Labeling and Detection of Endogenous Proteins Using Tag-Assisted Split Enzyme Complementation. *ACS Chem. Biol.* **2021**, *16* (4), 671–681. https://doi.org/10.1021/acschembio.0c00925.

(8) 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. *Science* **2024**, *383* (6685), 890–897. https://doi.org/10.1126/science.adg0812.

(9) Zhang, Y.; Park, K.-Y.; Suazo, K. F.; Distefano, M. D. Recent Progress in Enzymatic Protein Labelling Techniques and Their Applications. *Chem. Soc. Rev.* **2018**, *47* (24), 9106–9136. https://doi.org/10.1039/C8CS00537K.

(10) Mollwitz, B.; Brunk, E.; Schmitt, S.; Pojer, F.; Bannwarth, M.; Schiltz, M.; Rothlisberger, U.; Johnsson, K. Directed Evolution of the Suicide Protein O^6 -Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry* **2012**, *51* (5), 986–994. https://doi.org/10.1021/bi2016537.

(11) Mollwitz, B.; Brunk, E.; Schmitt, S.; Pojer, F.; Bannwarth, M.; Schiltz, M.; Rothlisberger, U.; Johnsson, K. Directed Evolution of the Suicide Protein O(6)-Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry* **2012**, *51* (5), 986–994. https://doi.org/10.1021/bi2016537.

(12) Carvajal-Vallejos, P.; Pallissé, R.; Mootz, H. D.; Schmidt, S. R. Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources. *Journal of Biological Chemistry* **2012**, *287* (34), 28686–28696. https://doi.org/10.1074/jbc.M112.372680.

(13) Birke, R.; Ast, J.; Roosen, D. A.; Lee, J.; Roßmann, K.; Huhn, C.; Mathes, B.; Lisurek, M.; Bushiri, D.; Sun, H.; Jones, B.; Lehmann, M.; Levitz, J.; Haucke, V.; Hodson, D. J.; Broichhagen, J. Sulfonated Red and Far-Red Rhodamines to Visualize SNAP- and Halo-Tagged Cell Surface Proteins. *Org. Biomol. Chem.* **2022**, *20* (30), 5967–5980. https://doi.org/10.1039/D1OB02216D.

(14) Roßmann, K.; Akkaya, K. C.; Poc, P.; Charbonnier, C.; Eichhorst, J.; Gonschior, H.; Valavalkar, A.; Wendler, N.; Cordes, T.; Dietzek-Ivanšić, B.; Jones, B.; Lehmann, M.; Broichhagen, J. *N* -Methyl Deuterated Rhodamines for Protein Labelling in Sensitive Fluorescence Microscopy. *Chem. Sci.* **2022**, *13* (29), 8605–8617. https://doi.org/10.1039/D1SC06466E.

(15) Avila, J.; Lucas, J. J.; Pérez, M.; Hernández, F. Role of Tau Protein in Both Physiological and Pathological Conditions. *Physiological Reviews* **2004**, *84* (2), 361–384. https://doi.org/10.1152/physrev.00024.2003.

(16) Wegmann, S.; Eftekharzadeh, B.; Tepper, K.; Zoltowska, K. M.; Bennett, R. E.; Dujardin, S.; Laskowski, P. R.; MacKenzie, D.; Kamath, T.; Commins, C.; Vanderburg, C.; Roe, A. D.; Fan, Z.; Molliex, A. M.; Hernandez-Vega, A.; Muller, D.; Hyman, A. A.; Mandelkow, E.; Taylor, J. P.; Hyman, B. T. Tau Protein Liquid–Liquid Phase Separation Can Initiate Tau Aggregation. *The EMBO Journal* **2018**, *37* (7), e98049. https://doi.org/10.15252/embj.201798049.

(17) Wang, L.; Tran, M.; D'Este, E.; Roberti, J.; Koch, B.; Xue, L.; Johnsson, K. A General Strategy to Develop Cell Permeable and Fluorogenic Probes for Multicolour Nanoscopy. *Nat. Chem.* **2019**, 1–8. https://doi.org/10.1038/s41557-019-0371-1.

(18) Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V. R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; Kwabi-Addo, D.; Beaini, D.; Jaakkola, T.; Barzilay, R. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRxiv June 18, 2025, p 2025.06.14.659707. https://doi.org/10.1101/2025.06.14.659707.

(19) 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* **2021**, *60* (33), 2560–2575. https://doi.org/10.1021/acs.biochem.1c00258.

(20) McNutt, A. T.; Li, Y.; Meli, R.; Aggarwal, R.; Koes, D. R. GNINA 1.3: The next Increment in Molecular Docking with Deep Learning. *J Cheminform* **2025**, *17* (1), 28. https://doi.org/10.1186/s13321-025-00973-x.

(21) Berendsen, H. J. C.; Van Der Spoel, D.; Van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(22) Abraham, M.; Alekseenko, A.; Andrews, B.; Basov, V.; Bauer, P.; Bird, H.; Briand, E.; Brown, A.; Doijade, M.; Fiorin, G.; Fleischmann, S.; Gorelov, S.; Gouaillardet, G.; Gray, A.; Irrgang, M. E.; Jalalypour, F.; Johansson, P.; Kutzner, C.; Łazarski, G.; Lemkul, J. A.; Lundborg, M.; Merz, P.; Miletić, V.; Morozov, D.; Müllender, L.; Nabet, J.; Páll, S.; Pasquadibisceglie, A.; Pellegrino, M.; Piasentin, N.; Rapetti, D.; Sadiq, M. U.; Santuz, H.; Schulz, R.; Shirts, M.; Shugaeva, T.; Shvetsov, A.; Turner, P.; Villa, A.; Wingbermühle, S.; Hess, B.; Lindahl, E. GROMACS 2025.4 Source Code, 2025. https://doi.org/10.5281/ZENODO.17671793.

(23) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J Comput Chem* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.

(24) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(25) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H.

(26) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J Comput Chem* **1992**, *13* (8), 952–962. https://doi.org/10.1002/jcc.540130805.

(27) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*; Markidis, S., Laure, E., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2015; Vol. 8759, pp 3–27. https://doi.org/10.1007/978-3-319-15976-8_1.

(28) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. https://doi.org/10.1093/bioinformatics/btt055.

(29) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447. https://doi.org/10.1021/ct700301q.

(30) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(31) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(32) Rodriguez-Rios, M.; Craigon, C.; Nakasone, M. A.; Bond, A. G.; Dorward, M.; Edmonds, A. K.; Norley, M. C.; Arnold, R. E.; Wood, P. M.; Reynolds, S. J.; Cresser-Brown, J. O.; Marsh, G. P.; Maple, H. J.; Ciulli, A. BromoCatch: A Self-Labelling Tag Platform for Protein Analysis and Live Cell Imaging. Biochemistry April 8, 2025. https://doi.org/10.1101/2025.04.07.647551.

(33) Erdmann, R. S.; Baguley, S. W.; Richens, J. H.; Wissner, R. F.; Xi, Z.; Allgeyer, E. S.; Zhong, S.; Thompson, A. D.; Lowe, N.; Butler, R.; Bewersdorf, J.; Rothman, J. E.; St Johnston, D.; Schepartz, A.; Toomre, D. Labeling Strategies Matter for Super-Resolution Microscopy: A Comparison between HaloTags and SNAP-Tags. *Cell Chemical Biology* **2019**, *26* (4), 584-592.e6. https://doi.org/10.1016/j.chembiol.2019.01.003.

(34) Elia, N. Using Unnatural Amino Acids to Selectively Label Proteins for Cellular Imaging: A Cell Biologist Viewpoint. *The FEBS Journal* **2021**, *288* (4), 1107–1117. https://doi.org/10.1111/febs.15477.

(35) Zhou, W.; Wesalo, J. S.; Liu, J.; Deiters, A. Genetic Code Expansion in Mammalian Cells: A Plasmid System Comparison. *Bioorganic & Medicinal Chemistry* **2020**, *28* (24), 115772. https://doi.org/10.1016/j.bmc.2020.115772.

(36) Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernández-Suárez, M.; Puthenveetil, S.; Ting, A. Y. A Fluorophore Ligase for Site-Specific Protein Labeling inside Living Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (24), 10914–10919. https://doi.org/10.1073/pnas.0914067107.

(37) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124* (21), 6063–6076. https://doi.org/10.1021/ja017687n.

(38) Gautier, A. Fluorescence-Activating and Absorption-Shifting Tags for Advanced Imaging and Biosensing. *Acc. Chem. Res.* **2022**, *55* (21), 3125–3135. https://doi.org/10.1021/acs.accounts.2c00098.

(39) Doh, J. K.; Tobin, S. J.; Beatty, K. E. MiniVIPER Is a Peptide Tag for Imaging and Translocating Proteins in Cells. *Biochemistry* **2020**, *59* (33), 3051–3059. https://doi.org/10.1021/acs.biochem.0c00526.

(40) Juillerat, A.; Gronemeyer, T.; Keppler, A.; Gendreizig, S.; Pick, H.; Vogel, H.; Johnsson, K. Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules in Vivo. *Chem Biol* **2003**, *10* (4), 313–317. https://doi.org/10.1016/s1074-5521(03)00068-1.

(41) Gronemeyer, T.; Chidley, C.; Juillerat, A.; Heinis, C.; Johnsson, K. Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Applications in Protein Labeling. *Protein Eng Des Sel* **2006**, *19* (7), 309–316. https://doi.org/10.1093/protein/gzl014.

(42) Frei, M. S.; Tarnawski, M.; Roberti, M. J.; Koch, B.; Hiblot, J.; Johnsson, K. Engineered HaloTag Variants for Fluorescence Lifetime Multiplexing. *Nat Methods* **2022**, *19* (1), 65–70. https://doi.org/10.1038/s41592-021-01341-x.

(43) Kühn, S.; Nasufovic, V.; Wilhelm, J.; Kompa, J.; De Lange, E. M. F.; Lin, Y.-H.; Egoldt, C.; Fischer, J.; Lennoi, A.; Tarnawski, M.; Reinstein, J.; Vlijm, R.; Hiblot, J.; Johnsson, K. SNAP-Tag2 for Faster and Brighter Protein Labeling. *Nat Chem Biol* **2025**, *21* (11), 1754–1761. https://doi.org/10.1038/s41589-025-01942-z.

(44) Birol, M.; Wojcik, S. P.; Miranker, A. D.; Rhoades, E. Identification of N-Linked Glycans as Specific Mediators of Neuronal Uptake of Acetylated α-Synuclein. *PLoS Biol* **2019**, *17* (6), e3000318. https://doi.org/10.1371/journal.pbio.3000318.

(45) Früh, S. M.; Matti, U.; Spycher, P. R.; Rubini, M.; Lickert, S.; Schlichthaerle, T.; Jungmann, R.; Vogel, V.; Ries, J.; Schoen, I. Site-Specifically-Labeled Antibodies for Super-Resolution Microscopy Reveal *In Situ* Linkage Errors. *ACS Nano* **2021**, *15* (7), 12161–12170. https://doi.org/10.1021/acsnano.1c03677.

FIGURES

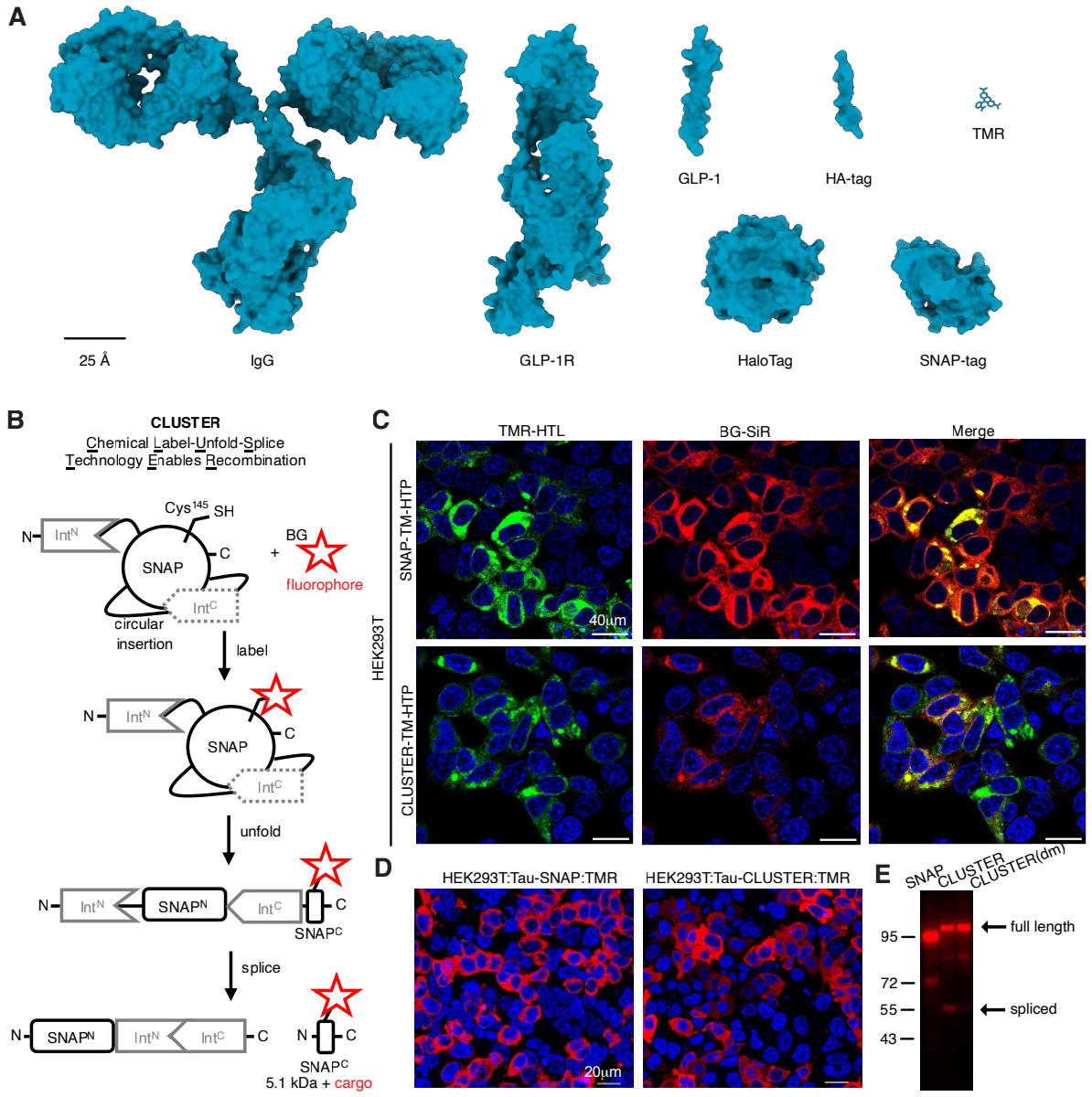

Figure 1. Protein sizes, CLUSTER design and expression. A) Size comparison of various relevant molecular structures, including immunoglobulin, GLP-1 and GLP-1R, various tags (HaloTag, SNAP-tag, HA-tag), and tetramethylrhodamine (TMR). **B)** SNAP-tag is endowed *N*-terminally with gp41-1^{N} and inserted with gp41-1^{C} between position 132-133, upstream of the reactive residue Cys145. Upon SNAP-tag reaction with its BG substrate, the construct becomes less stable and the inteins assemble for splicing, transferring the labelled *C*-terminal SNAP residue onto a *N*-terminal protein of interest. **C)** Live HEK293T cells transiently transfected with SNAP–TM–HTP or CLUSTER–TM–HTP, stained with TMR-d12-HTL and BG-SiR-d12. **D)** Live HEK293T cells transiently transfected with Tau–SNAP or Tau–CLUSTER, stained with TMR-d12-HTL and BG-SiR-d12. **E)** SDS-PAGE of protein extract from (D), including a dead mutant (dm), showing only spliced and fluorescently labelled product for Tau-CLUSTER.

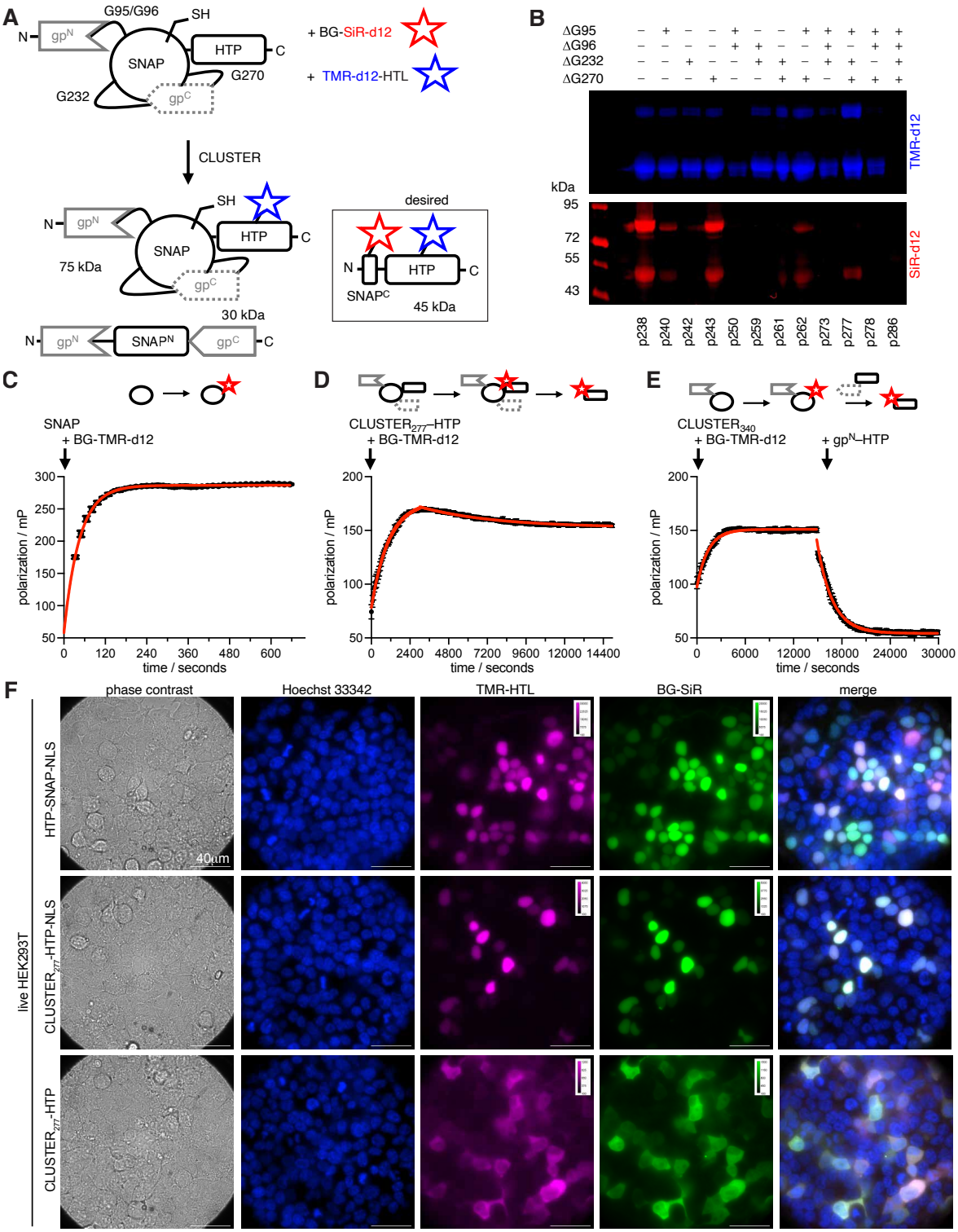

Figure 2. A) CLUSTER is equipped with a *C*-terminal HaloTag for orthogonal control, as well as glycine linkers at domain intersections (G95, G96, G232, G270) for subsequent mutational screening. Upon labelling with respective fluorophores, the construct is spliced, releasing the dual-stained SNAP^{C}-HTP fragment. **B)** SDS-PAGE gel under a fluorescence imager after CLUSTER-HTP-expression and labelling in *E. coli.* CLUSTER$_{277}$ shows clean labelling of only spliced product. **C-E)** Fluorescence polarization assay to mechanistically determine protein labelling and dynamics. **F)** Live-cell confocal microscopy of HEK293T cells transiently transfected with HTP–SNAP–NLS, CLUSTER$_{277}$–HTP–NLS, and CLUSTER$_{277}$–HTP. Staining was performed overnight using Hoechst 33342, BG-SiR-d12, and TMR-d12-HTL.

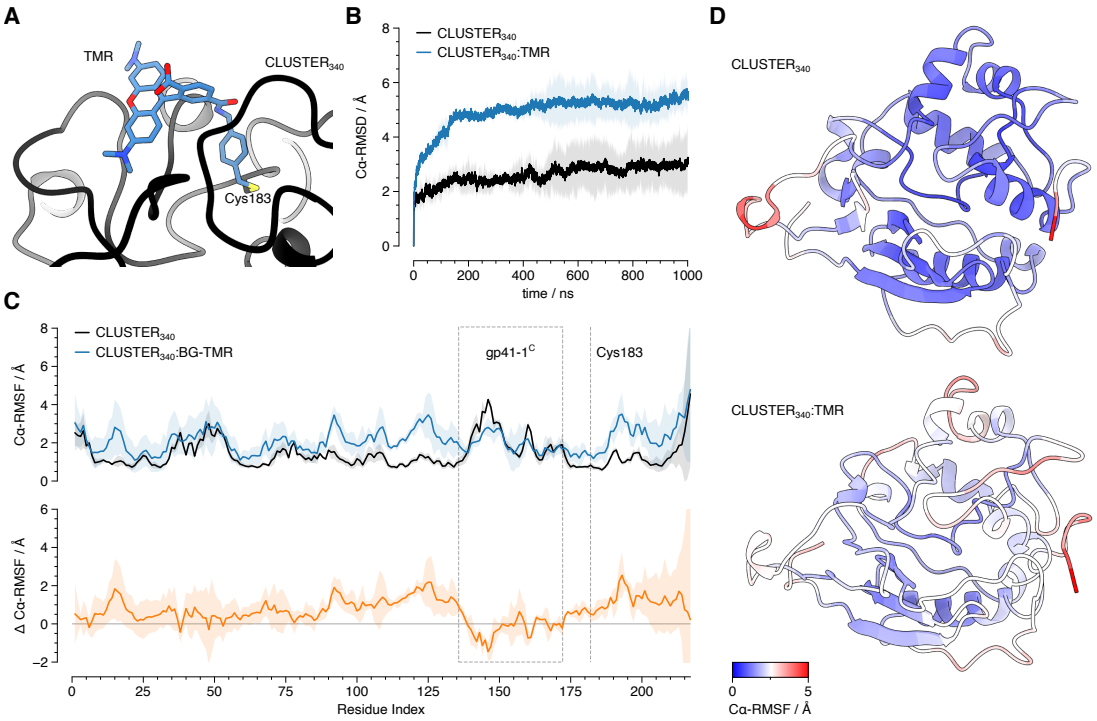

Figure 3. A) BG-TMR was covalently docked to the computationally predicted structure of $CLUSTER_{340}$ for subsequent molecular dynamics simulation. **B)** C-α RMSD traces over molecular dynamics simulation trajectories of $CLUSTER_{340}$ and $CLUSTER_{340}$:TMR (opaque: average of triplicate runs, light: absolute spread of measured values across all trajectories). **C)** Top: RMSF of $CLUSTER_{340}$ and $CLUSTER_{340}$:TMR across triplicate molecular dynamics simulation trajectories (opaque: average of triplicate runs, light: absolute spread of measured values across all trajectories). Bottom: Difference of trajectory RMSF between $CLUSTER_{340}$ and $CLUSTER_{340}$:TMR. **D)** Predicted structure of $CLUSTER_{340}$ with color-mapped average RMSF values from **C**.

MATERIALS and METHODS

can be found in the SUPPORTING INFORMATION

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon Europe Framework Program (deuterON, grant agreement no. 101042046 to JB). We thank Ramona Birke for technical assistance. Funded by Boehringer Ingelheim Foundation (Exploration Grant to JB). We thank Ines Kretzschmar (Leibniz-FMP) for support on SPPS.

COMPETING INTERESTS

JB receives licensing revenue from Celtarys Research for provision of chemical probes. MDC and FVB have filed a patent application for the CLUSTER technology. The remaining authors declare no competing interests.