Characterization of a novel membrane protein insertase in the mitochondrial outer membrane

Thesis by

Taylor Anthony Stevens

In Partial Fulfillment of the Requirements for

the degree of

Doctorate of Philisophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2024

(Defended November 8th, 2023)

Taylor Anthony Stevens
ORCID: 0000-0002-6232-5316

ACKNOWLEDGEMENTS

I want to acknowledge all of my friends and extended family for their unwavering support throughout my life. My parents Chris and Nancy have always been there when I needed them and without them I would not be where I am today. Dad, you've been an inspiring role model who has taught me how to work hard to achieve my dreams, how to problem solve, but also how to find joy in life. Mom, you recognized my potential for science and math at a young age, you pushed me to study hard in school, and also tried to find additional challenges for me to work on when the default coursework wasn't challenging for me. There is no way I would have made it from community college to a Ph.D. at Caltech of all places without either of your support and encouragement.

I owe so much of my present success to my wife Veronica. You've been by my side the entire time I've been a Ph.D. student, although the first four years it was mostly in spirit as you finished your Ph.D. on the other side of the country. Your advice guided me in the right direction during one of the most challenging times in my Ph.D. when I decided to change advisors, and the impact of this decision ended up being positive to a degree I could not have imagined. You listen to me ramble on about all of my nutty ideas, and you let me complain during the inevitable periods when things aren't working in lab. When I am on the verge of overworking myself (or past it) you give me a reason to take a break and are the main reason why I have been able to enjoy my life so much in the past several years.

My brothers Carter, Cameron, and Zach have been a great source of entertainment, fashion advice, and company. One of the highlights of my yearly visits home to Tennessee is visiting with you and seeing how much you are all growing, and I really hope we can all visit more frequently after I finish my Ph.D. In the first 4 years while I lived in student housing, I was lucky to have Sho Harvey, Rory Williams, Tom O'Connell, and Tom Roeschinger as roommates, especially the two Toms who were my sole source of human interaction for multiple months during the 2020 covid lockdown. I also thank Claudie Jette, Lucas Schaus, Sharan Prakash, Alex Hooten, Valentino Constantinou, Emma Brand, Jonathon Haymer, all

members of the BMB bombers softball team, as well as Mauro Santos (Maestre Falcão) and all fellow Caltech capoeira practitioners for their friendship over the past several years.

My advisor Rebecca Voorhees has been a great source of knowledge, encouragement, and inspiration. I could not ask for a better Ph.D. mentor, and I consider myself lucky for just having the opportunity to work under you. I have had more fun working in your lab than anywhere else I have been. It is simply a joy to come in every day, to discuss science and plan experiments, to interpret new data, and get excited about future plans. Alison Inglis and Alina Guna have been great teammates on the quest to understand mitochondrial membrane protein insertion, and I am especially thankful to Alison for teaching me so much when I first joined the lab. I am also thankful to Tino Pleiner who has been a pleasure to work with while developing a nanobody protein purification toolkit. It has been inspiring to witness your trajectory. Erini Galatis has been a wonderful and reliable technician to work with. Charlene DeKalb and Robert Oania have both played a crucial role in my Ph.D. by keeping the lab in order as excellent lab managers. Giovanni Pinton-Tomaleri has been a great source of positivity and helped convince me to join the lab. Katie Page, Masami Hazu, Vy Nguyen, Maxine Wang, Natalie Chen, Sophia Wei, Gina Mowla and Kelly Shaffer have all been wonderful colleagues and have all helped me in one way or another throughout my time in lab.

I would additionally like to thank Evelyn Li, Andy Delaitsch, Przemek Dutka, and Cornelius Gati for scientific discussion and data processing advice. Chris Bley and Ferdinand Huber taught me useful skills in the first half of my Ph.D. Bil Clemons has been an excellent leader of the BMB program. Doug Rees, David Chan, and Pamela Bjorkman have provided much appreciated guidance as members of my thesis committee. I would like to thank Jonathan Weissman and his lab members including Gayathri Muthukumar, Joseph Replogle, and Theodore Esantsi for the fruitful collaborations resulting in multiple publications. From my time at the University of Tennessee at Knoxville, I am thankful to professors Elias Fernandez and Tongye Shen for guidance on independent research projects, and also to graduate students Vandna Gahlot and Quentin Johnson for training. I am thankful to Martin St. Maurice at Marquette University to being a mentor during my summer internship there.

Lastly, I want to thank all professors and staff at Nashville State Community College and Tennessee State University for helping me build the foundation for everything I have done since starting my college education.

ABSTRACT

Mitochondria are eukaryotic organelles derived from the endosymbiosis of an ancient bacteria. As a result of their endosymbiotic origin, the mitochondrial proteome is composed of a mixture of ancient bacterial derived genes and others which are unique to eukaryotes. This dual bacterial/eukaryotic protein origin results in a complicated landscape for biogenesis of mitochondrial proteins. This is particularly true for mitochondrial membrane proteins, since mitochondria have both an inner and outer membrane each with unique protein composition. Proteins localized to the outer mitochondrial membrane (OMM) are of particular interest due to their connection to many important physiological pathways in humans. OMM proteins are known to be inserted into the lipid bilayer by the MIM complex in yeast and by ATOM36 in trypanosomes, however it is not known how they are inserted in human cells. In my Ph.D. thesis, I describe the development of improved biochemical tools for protein purification and characterization, and then use them as part of an effort to characterize MTCH2, which we identify as the human gene responsible for OMM protein insertion. After identifying MTCH2 in a genome-wide screen, we use a variety of cell biology and biochemical experiments to show that MTCH2 is both necessary and sufficient for OMM protein insertion. We further show that endogenous OMM proteins are affected by MTCH2 depletion, and that apoptosis, a pathway relying on OMM proteins, is sensitive to MTCH2 modulation. Additional work in my thesis demonstrates that MTCH2 is a deeply conserved gene across metazoans, that other OMM insertases likely evolved independently in separate multi-cellular eukaryotic lineages, and that MTCH2 is regulated by nucleotide binding. Lastly I determine the structure of a homologue of the neuronal signaling protein Ca^{2+}/Calmodulin-dependent protein kinase II found in the thermophilic worm *A. pompejana* and propose the use of proteins from this organism as tools to study their human counterparts in experimental methods that require enhanced protein stability, such as structure determination.

PUBLISHED CONTENT AND CONTRIBUTIONS

Alina Guna*, Taylor A. Stevens*, Alison J. Inglis*, Joseph M. Replogle, Theodore K. Esantsi, Gayathri Muthukumar, Kelly C.L. Shaffer, Maxine L. Wang, Angela N. Pogson, Jeff J. Jones, Brett Lomenick, Tsui-Fen Chou, Jonathan S. Weissman, and Rebecca M. Voorhees. (2022). "MTCH2 is a mitochondrial outer membrane protein insertase". In: *Science* 6617, pp. 317-233. doi: 10.1126/science.add1856

* denotes equal contribution

T.A.S participated in the design, analysis, and interpretation of experiments as well as manuscript preparation. Specific experimental contributions include: mitochondria isolation and proteoliposome reconstitution for *in vitro* assays, crosslinking mass-spectrometry analysis, and mitochondrial proteomic analysis.

Taylor Anthony Stevens, Giovani Pinton Tomaleri, Masami Hazu, Sophia Wei, Vy N. Nguyen, Charlene DeKalb, Rebecca M. Voorhees, and Tino Pleiner. (2023). A nanobody-based strategy for rapid and scalable purification of human protein complexes. In press: *Nature Protocols*.

T.A.S participated in the design, analysis, and interpretation of experiments as well as manuscript preparation. Specific experimental contributions include: plasmid design, cloning, and validation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

Chapter 1

INTRODUCTION

Membrane protein biogenesis requires many dedicated factors

One of the core principals to understanding life at a molecular level is the central dogma, which states that our cells store genetic information in the form of DNA, which is transcribed into messenger RNA (mRNA) before being translated by the ribosome into proteins, which are ultimately responsible for carrying out the gene's function. However, the vast majority of proteins initially emerge from the ribosome as unstructured polypeptides that are incapable of carrying out their ultimate function. In order to carry out their functions, most proteins need to fold into specific 3-dimensional conformations which yield functional surfaces such as protein-protein interaction interfaces or enzyme active sites. Protein folding is driven by a variety of physical forces, most notably the need to shield hydrophobic amino acids from the aqueous environment of the cytoplasm where proteins are translated. Additionally, many proteins function as obligate members of multi-protein complexes, and cannot be functional until they are assembled with interaction partners in the correct stoichiometry. Lastly, many proteins function in different cellular compartments such as the nucleus, the endoplasmic reticulum (ER), the mitochondria, or in cellular membranes, and they need to be transported to these locations in order to carry out their functions. This collective process of correct protein folding, assembly, and localization is referred to as protein biogenesis. The process of protein biogenesis is aided by a variety of factors, such as cellular chaperones, which bind to and protect exposed hydrophobic surfaces on nascent proteins to facilitate their maturation, translocases which enable passage of proteins into different cellular compartments, and insertases which allows insertion of hydrophobic segments of transmembrane proteins, termed transmembrane domains (TMDs), into a lipid bilayer.

The biogenesis of transmembrane proteins represents are particularly complex but important area of study for protein biogenesis. Transmembrane proteins carry out a variety of essential processes, such as transporting nutrients between cellular compartments and responding to

external cellular signals. Transmembrane proteins encode one or more TMDs, all of which need to be inserted into a lipid bilayer in the correct orientation and at the correct cellular location. In multi-TMD proteins, all TMDs must assemble together after insertion into the bilayer to form the correct 3-dimensional fold. There are ~5000 different membrane proteins predicted to be in the human genome (Dobson et al., 2015), each which can have different numbers of TMDs, are localized to different cellular membranes, and have unique TMD sequence features. Given this massive diversity of transmembrane proteins, it is necessary for human cells to encode several dedicated factors to ensure their correct biogenesis and function.

To date, substantial progress has been made in understanding the factors involved in membrane protein biogenesis. A subset of proteins destined for the endoplasmic reticulum (ER) encode signal sequences that are recognized at the initial stages of translation by the chaperone SRP, and then subsequently targeted to the Sec61 translocon for co-translational insertion (Shao & Hegde, 2011). A class of proteins which encode a single TMD at their C-terminus, called tail-anchored (TA) proteins, are bound by cytosolic chaperones such as TRC40 or calmodulin, which then deliver them to the ER insertase WRB/CAML or the ER membrane complex (EMC) respectively (Schuldiner et al., 2008; Guna et al., 2018). Despite our advancing knowledge in this field, there are many pathways which have not been well characterized. A majority of recent advances in our understanding of membrane protein biogenesis focus on ER membrane proteins, but we know substantially less about membrane protein biogenesis in mitochondria.

Human mitochondrial membrane protein biogenesis is complex and poorly understood

Mitochondria represent a particularly complex area for the study of membrane protein biogenesis. Mitochondria are eukaryotic organelles derived from an ancestral bacteria that was engulfed by the archaeal ancestor to all eukaryotes (Roger et al., 2017). Mitochondria have retained their bacterial morphology, with in outer membrane that contacts the cytoplasm, and an inner membrane which encloses a compartment termed the matrix. While mitochondria have still retained their own genome, nearly all of their genes have relocated

to the host nuclear genome. Nuclear-encoded mitochondrial proteins are therefore translated in the cytoplasm and then need to be transported to their respective compartment within mitochondria. Interestingly, mitochondria have retained some protein biogenesis machinery from bacteria, for instance, proteins targeted to the outer mitochondrial membrane (OMM) with a β-barrel topology are inserted by the SAM complex which is homologous to the bacterial BamA. Some proteins targeted to the inner mitochondrial membrane (IMM) are inserted by the Oxa1 insertase, which is homologous to bacterial YidC. One notable difference between bacteria and mitochondria is that there are over 100 transmembrane proteins with α-helical TMDs in the OMM, and only few or none in bacteria (Weiner & Li, 2008; Rath et al., 2021). In eukaryotes, 2 distinct OMM insertases for α-helical TMDs have been identified: the MIM complex in fungi, and ATOM36 in kinetoplastids (Doan et al., 2020; Vitali et al., 2018). Given that no such pathways have been identified in bacteria, each of these factors have most likely independently evolved in eukaryotes.

While the insertion pathway for α-helical TMDs is known in yeast and kinetoplastids, there is still no information for how this process occurs in humans. No prior studies have identified homologues of ATOM36 or the MIM complex in humans, raising the possibility that humans or animals as a whole have evolved a distinct OMM insertase. OMM proteins in humans are involved in a variety of physiologically important pathways such as mitochondrial import, apoptosis, fatty acid metabolism, and mitochondrial fusion and fission (Wang et al., 2020; Bose & Beal, 2016; Vyas et al., 2016). Since many if not all of these factors must be localized to the OMM in order to carry out their function, the process of insertion represents a potential point for regulation. Therefore, by understanding how human OMM proteins are inserted, we may gain a better understanding of the regulation of many important physiological processes.

<u>New cell biology and biochemistry methods aid the study human mitochondrial membrane protein biogenesis</u>

Our knowledge of mitochondrial membrane protein biogenesis is more extensive in other organisms such as the budding yeast *S. cerevisiae*. Reasons behind this include the increased

skills and resources needed to culture human cells, as well as lack of experimental tools to study membrane protein biogenesis in humans. In the study of membrane protein biogenesis, it is often necessary to purify individual proteins for detailed mechanistic studies or functional reconstitution. The expression and purification of human membrane proteins is particularly challenging and often inefficient, necessitating expression culture volumes on the scale of hundreds of milliliters or even liters of expensive mammalian cell culture medium. Additionally, in experiments aimed at identifying protein interaction partners, contaminants which non-specifically bind to affinity resin can also obscure results. Another large difference between humans and budding yeast is that is that the yeast genome is much easier to manipulate, and it only contains ~6000 genes compared to >20,000 for humans. Given this difference, combined with the overall increased difficulty of handling cultured human cell lines, it is much more challenging to identify novel gene functions in high-throughput screens. Existing methods to study mitochondrial membrane protein biogenesis require the tedious isolation of purified mitochondria from cells followed by additional analysis such as import of *in vitro* translated protein or western blot analysis, which is impractical to perform with more than a few dozen samples.

In Chapter 2 of this thesis, I describe the use of proteins from the thermophilic worm *A. pompejana* for structural and biophysical studies. Proteins from thermophilic organisms usually have increased solubility and thermostability, leading to higher yield during recombinant expression, and better results in biophysical experiments due to lack of aggregation. Due the closer evolutionary relationship with humans, *A. pompejana* proteins are expected to have much higher homology to their human counterparts than fungal proteins. The use of *A. pomejana* homologues to human proteins is an excellent choice when the human homolog cannot be expressed in large enough quantities, or when it is not stable enough for use in biophysical assays or structure determination.

In Chapter 3 of this thesis, I describe the development of a 28-plasmid toolkit for rapid affinity-based purification of proteins from human cells with higher yield and purity than previously described affinity-based methods. This method combines the advantages of using

high-affinity tag-binding nanobodies together with an engineered SUMO protease which can be used to perform protease-mediated elution much more efficiently than commonly used proteases. This toolkit enables the high yield isolation of notoriously difficult-to-purify soluble and membrane protein complexes, greatly facilitating their structural and biochemical characterization. This method also provides excellent results when used to identify protein binding partners via mass-spectrometry.

In Chapter 4 of this thesis, I describe the use of a newly developed split-fluorescent reporter for OMM protein insertion in a high-throughput genetic screen using human tissue culture cells to identify MTCH2 as an OMM insertase in human cells. Biochemical characterization of the OMM insertase activity is greatly facilitated by use of the previously developed nanobody toolkit for interrogating interactions between MTCH2 and its OMM clients and for purifying MTCH2 from human cells to test its function in liposome reconstitution assays. In Chapter 5 of this thesis, I present a continuation of the work described in chapter 4, where the same tools are used to characterize the function of metazoan and protist homologues of MTCH2 as well as plant homologues of ATOM36, in addition to uncovering a nucleotide-based mechanism for regulation of MTCH2 activity.

Chapter 2

CRYSTAL STRUCTURE AND BIOCHEMICAL ANALYSIS OF THE A. POMPEJANA CAMKII HUB DOMAIN

2.1 Abstract

Ca^{2+}/Calmodulin-dependent protein kinase II (CaMKII) is a neuronal signaling factor involved in memory formation. The hub domain of CaMKII forms a ring-shaped oligomer that positions its kinase domains in close proximity, leading to autophosphorylation upon stimulation by calcium signaling. Here we identify the sequence of a CaMKII homologue in the thermophilic worm *Alvinella pomoejana*. We then solved a crystal structure of the *Ap*CaMKII hub domain to 2.05 Å. Biochemical and structural analysis of *Ap*CaMKII shows that this protein exhibits increased thermostability over mammalian homologues but is otherwise highly similar. This work suggests that the highly thermostable protein sequences from *Alvinella pompejana* are an excellent choice for structure determination of challenging metazoan-specific targets.

2.2 Introduction

Protein thermostability is often a bottleneck in isolation and analysis during in vitro studies. Proteins derived from organisms adapted to high temperatures display favorable properties such as improved recombinant expression, higher solubility, and propensity to crystallize. Numerous crystal structures have been determined with recombinantly expressed proteins using sequences from the thermophilic fungus *Chaetomium thermophillum* (Amlacher et al., 2011), including the structures of many proteins from the nuclear pore complex, which ultimately enabled the determination of a composite structure of the entire symmetric core of the nuclear pore complex when combined with cryo-electron tomography (Lin et al., 2016).

In order to study neuron-specific molecular machinery with thermophilic proteins, it would be necessary to establish a thermophilic model organism where this machinery is conserved. *Alvinella pompejana*, a polychaete worm found in hydrothermal vents of the Pacific Ocean, is an organism that lives in extreme temperatures and has a nervous system (Holder et al., 2013). In this study, we use Ca^{2+}/Calmodulin-dependent protein kinase II (CaMKII), known for its role in learning and memory, as a model neuronal protein to investigate the utility of *A. pompejana* genes in facilitating the structural and biochemical characterization of neuron-specific proteins.

CaMKII is composed of an N-terminal Ser/Thr kinase domain, a flexible linker, and a C-terminal hub domain which allows CaMKII to oligomerize into a ring containing 12 or 14 subunits. Transient spiking of intracellular calcium during neuronal activity triggers calmodulin binding to the CaMKII flexible linker, which activates the kinase domain (Lisman et al., 2002; Kandel et al., 2014; Lisman & Raghavachari, 2015; Hook & Means, 2001; Hoelz et al., 2003). Once activated, a kinase domain phosphorylates its neighbors within the same oligomeric assembly, however, phosphorylation of kinase domains in separate oligomers only occurs at a slow rate (Rich & Schulman, 1998; Hanson et al., 1994). In order to effectively propagate a signaling event, CaMKII oligomers dynamically exchange subunits in a mechanism that is thought to involve toggling between dodecameric and tetradecameric states via gain or loss of a CaMKII dimer (Bhattacharyya et al., 2016).

Here, we identify the sequence of a CaMKII homologue encoded by the *A. pompejana* genome. Thermostability analysis reveals that *Ap*CaMKII indeed displays increased thermostability over mammalian isoforms, and further biochemical analysis and structure determination demonstrates that this protein is otherwise highly similar to mammalian CaMKII homologues. We additionally observe density in the crystal structure of *Ap*CaMKII which suggest the presence of an unknown ligand, and a striking level of heterogeneity in a loop conformation which may have functional implications.

2.3 Results

Thermostability and stoichiometry analysis of the hub domain from a putative CaMKII gene in *A. pompejana*

A candidate for *A. pompejana* CaMKII based on a partial genome reconstruction was obtained from Richard Copley (National Centre for Scientific Research, Villefranche-sur-Mer, France; data unpublished). The obtained sequence most closely matches the mouse CaMKII-α homolog, with 77.5% identity to mouse CaMKII-α. Residues 1-272 align to the mouse CaMKII-α kinase domain (1-271) with 84.9 % identity. Residues 341-477 align to the mouse CaMKII-α hub domain (342-478) with 64.5 % identity. Residue D136 in the kinase domain aligns with catalytic site residue D135 from mouse CaMKII-α and 3 residues known to be phosphorylated are conserved at sites T287, T306, and T307 (**Figure 2.1**).

To investigate the oligomeric state of the *A. pompejana* CaMKII candidate, the hub domain (341-477) was cloned and purified via recombinant expression in *E. coli* together with the equivalent domain in mouse CaMKII (336-478). When analyzed via size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), both proteins eluted as a single species with an observed molecular mass of 211.2 kDa for *A. pompejana* and 219.3 kDa for mouse, which corresponds to a stoichiometry of 13.1 and 13.0 respectively (**Figure 2.2A**). These measurements agree with previous studies which show that the CaMKII hub domain exists as a mixture of dodecameric and tetradecameric assemblies (Bhattacharyya et al., 2016). Since the hub domain of mammalian CaMKII homologues is already >90 °C, we analyzed thermostability of the kinase domain (Torres-Ocampo et al., 2020). Residues 8-275 were cloned and purified together with the equivalent domain in mouse (7-274). Thermostability analysis was performed by incubating purified protein samples at a series of increase temperature points and then quantifying the fraction of protein which remained soluble by centrifuging at high-speed to pellet aggregated protein and analyzing the resulting samples using SDS-PAGE. Thermostability analysis shows that an unfolding transition occurs around 28 °C for the mouse CaMKII kinase domain, while the unfolding transition occurs at a much higher temperature of 42 °C for the *Ap*CaMKII kinase

domain (**Figure 2.2**). The thermostability of each homologue increases slightly in the presence of magnesium and ADP (**Figure S2.4**).

Crystallographic analysis of the *A. pompejana* CaMKII hub domain reveals a tetradecameric ring with high similarity to other species

The *A. pompejana* CaMKII hub domain was crystallized and diffracted to 2.05 Å. A solution was readily obtained via molecular replacement using the mouse CaMKII hub domain (PDB ID 1HKX). The dataset was indexed in the space group P1, and the asymmetric unit contained 2 separate ring-shaped oligomers with 14 chains each, for a total of 28 chains. The two rings were offset by a roto-translation operation. The conformations of both rings were highly similar to one another, as well as to the tetradecameric ring formed by the mouse CaMKII hub domain. Individual chains of the mouse and *A. pompejana* hub domains can be aligned with an RMSD of 0.434 Å, and the complete tetradecameric rings can be aligned with an RMSD of 1.208 Å (**Figure 2.3**).

Analysis of the residues located at the 2-fold and 7-fold symmetric interfaces reveals a high degree of similarity between the *A. pompejana* and mouse hub domains, particularly for the 2-fold interface. 15 out of 22 sidechains contacting at the 7-fold symmetric interface are identical, while 17 out of 21 sidechains contacting at the 2-fold symmetric interface are identical. The mutations observed involve minimal changes to side chain properties (**Figure S2.1**).

Density found in the hub domain cavity suggests the presence of a bound molecule with a carboxyl group

While building the protein model, a strong electron density peak was observed which was coordinated by residues Y368 and R452 and which cannot be readily explained by any other region of the protein. The density appears to be planar with two lobes projecting towards the nitrogen atoms of R452 (**Figure 2.4A**). Density at a homologous site is also found in other CaMKII hub domain structures, where it is modelled as a Cl^- ion (PDB 1HKX) (Hoelz et

al., 2003), water (PDB 6OF8, 5IG3) (Bhattacharyya et al., 2016; McSpadden et al., 2019), glycine (PDB 2UX0) (Rellos et al., 2010), acetate (PDB 2W2C) (Rellos et al., 2010), and sulphate (PDB ID 5IG0) (Bhattacharyya et al., 2016). In each of these cases, molecules other than water appear to have been placed when they could be justified by the crystallization condition, however, most of these structures are in a resolution range where distinguishing features are not apparent. The highest resolution CaMKII hub domain structure at 1.75 Å is from *S. rosetta* (PDB ID 5IG0) (Bhattacharyya et al., 2016) and includes a sulphate ion at the equivalent position, however the presence of strong negative difference density peaks around this group in a F_o-F_c map suggest that the presence of a sulphate ion cannot explain this density.

There are only 5 small molecules and ions in our protein buffer and crystallization condition (MPD [2-methyl-2,4-pentanediol], tris [tris(hydroxymethyl)aminomethane], Na^+, Cl^-, and DTT [dithiothreitol]), none of which could readily explain the density that we observed. It is possible that this density corresponds to a small molecule that was present in the recombinant *E. coli* expression host and bound with high enough affinity to withstand the multiple chromatography steps used in the purification. The shape and chemical environment of this density suggests the presence of a carboxyl-group which forms an ionic bond with R452. When an acetate ion was modelled into this density and refined, we could not see any major difference density peaks in a F_o-F_c map, and the resulting B-factors were comparable to the surrounding atoms. Additionally, the refined positions of the oxygen atoms were in good hydrogen bonding distance with the nitrogen and oxygen atoms from the sidechains of R452 and Y368. These results support the possibility this site can be occupied by a small molecule containing a carboxyl group.

Multiple distinct conformations observed for the α4-β3 loop

When comparing subunits, we noticed substantial differences in the conformation of the α4-β3 loop, which was observed in four distinct conformations, referred to as I-IV (**Figure 2.4B**). In conformation I, the side chain of residue L402 is facing the solvent, whereas it is flipped towards the interior of the protein in conformations II-IV. In conformation III and

IV, helix α4 is followed by an additional helix-like turn in which a 2.8 Å hydrogen bond is formed between the carboxyl oxygen of Y396 and the amine nitrogen of L402 and a 3.1 Å (conformation III) or 3.6 Å (conformation IV) hydrogen bond is formed between the carboxyl oxygen of F398 and the amine nitrogen of A403. Conformation IV is only found in one chain (chain O) and has a few differences in the main chain dihedral angles compared to conformation III. The full C-terminus is resolved in chains with conformation II, while most other conformations show little to no density beyond G470, suggesting that conformation II enables an ordered positioning of the C-terminus (**Figure 2.4C**).

For specific chains, the α4-β3 loop electron density appeared to be an average of two conformations. In such cases, we modelled this loop with two alternative conformations. When analyzing the distribution of each loop conformation within oligomeric rings, we noticed that in many cases, 'lateral dimers' of CaMKII chains in contact via the 2-fold symmetric interface of the tetradecameric ring had matching loop conformations (**Figure 2.4D**). In many other CaMKII hub domain structures, the α4-β3 loop is either not modelled, or modelled into weak electron density. Of the 4 CaMKII hub domain structures with resolution below 2.5 Å, we notice that 3 are captured with an α4-β3 loop conformation highly similar to conformation III (PDB ID 5IG4, 2UX0, 6OF8) (Bhattacharyya et al., 2016; McSpadden et al., 2019; Rellos et al., 2010), with one remaining structure (PDB ID 5IG0) (Bhattacharyya et al., 2016) exhibiting an α4-β3 loop conformation different from all 4 seen in *A. pompejana*.

2.4 Discussion

In this work, we identify a gene homologous to mammalian CaMKII from the thermophilic worm *A. pompejana*. We are able to demonstrate that this protein exhibits substantially increased thermostability over its mammalian counterparts while retaining identical oligomerization properties in solution and exhibiting high structural similarity. These results suggest that other neuronal proteins from *A. pompejana* should also exhibit increase thermostability, facilitating their use for many biochemical applications, while also having sufficient homology to their mammalian counterparts to justify translating the results of such

work into an increased understanding of human biology. We additionally observe some noteworthy features in our crystal structure of *Ap*CaMKII that hint at previously unknown conformational changes and ligand binding properties.

The possibility of ligand biding to the CaMKII hub domain represents an exciting yet unexplored facet of CaMKII function. The CaMKII hub domain shares structural homology with a larger group of proteins found in both eukaryotes and bacteria including ketosteroid isomerase (KSI). KSI is an enzyme with an active sites at the equivalent position as the observed ligand density in CaMKII, although the catalytic residues are not present in CaMKII (Kim et al., 1997). It is possible that the ancestral gene of CaMKII was originally an enzyme that lost its catalytic activity but retained ligand binding properties for either structural or regulatory purposes. Additionally, a previous report considered the possibility of an unstructured protein C-terminus fitting into this pocket (Hoelz et al., 2003), and a docking analysis identified compounds with fused rings similar KSI ligands that could fit in this site (McSpadden, 2018).

We are able to resolve several distinct conformations in the α4-β3 loop of the CaMKII hub domain. We observe that the C-terminus becomes ordered and forms contacts with this loop only for one of its conformations, and is positioned to form a cap over the putative ligand binding pocket (**Figure 2.4C**). It has been previously shown that the N-terminal flanking residues of the CaMKII hub domain can form an additional β-strand which facilitates subunit exchange (Bhattacharyya et al., 2016), and structural analysis shows that loop conformations I, II, and to a lesser extent IV are in a position that would prevent the formation of this β-strand, while conformation III is in a position that would allow its formation. Additionally, the α4-β3 loop is sufficiently close to the putative CaMKII ligand binding pocket that certain bulky ligands bound to this site would likely bias its conformation. This analysis therefore presents the possibility ligand binding or other external factors such as post-translational modification could regulate CaMKII subunit exchange by biasing the conformation of this loop. While the sequence of this loop is not perfectly conserved, the length is conserved in many species, and residues properties are generally conserved within the same position,

raising the possibility that such conformational flexibility may occur for other CaMKII homologues.

2.5 Acknowledgements

Cloning and protein purification was assisted by Leah Soldner-Garcia. The research project was advised by André Hoelz. This research was supported by a Human Frontiers Science Program grant. We also acknowledge Jens Kaiser and the scientific staffs of SSRL Beamline 12-2. The operation at SSRL, and NSLS-II is primarily supported by the Department of Energy. We acknowledge the Gordon and Betty Moore Foundation, the Beckman Institute, and the Sanofi-Aventis Bioengineering Research Program for their support of the Molecular Observatory at the California Institute of Technology. T.A.S was supported by an NIH research service award (5 T32 GM07616). A.H. is a Faculty Scholar of the Howard Hughes Medical Institute and an Investigator of the Heritage Medical Research Institute. A.H. was also supported by NIH grant R01-GM117360, Caltech startup funds, and a Teacher-Scholar Award of the Camille & Henry Dreyfus Foundation.

2.6 Methods

Construct generation

cDNA encoding the putative *A. pompejana* CaMKII homologue was synthesized as a gene fragment (Twist Biosciences). All constructs used for expressing *A. pomejana* or mouse CaMKII fragments were generated by PCR amplification of the desired sequence with a forward primer flanked with a 5' BamHI restriction site and a reverse primer flanked with a 3' NotI restriction site. Residues 341-477 from *Ap*CaMKII and 336-478 from *Mm*CaMKII were used for the hub domain. Residues 8-275 from *Ap*CaMKII and 7-274 from *Mm*CaMKII were used for the kinase domain. The resulting PCR product was then cloned in between BamHI and NotI restriction sites in a pET28a vector containing an N-terminal SUMO tag (Mossessova & Lima, 2000). Catalytic aspartates (D136 for *Ap*CaMKII, D135 for

*Mm*CaMKII) were mutated to asparagines in all kinase domain fragments via quikchange mutagenesis (Agilent Technologies) to aid in protein expression.

Protein expression and purification

Each protein was expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene) in Luria-Bertani media and induced at an optical density of 0.6 with 0.5 mM IPTG (Gold Biotechnology). All hub domain constructs were expressed for 2 hours at 37 °C. All kinase domain constructs were expressed overnight at 18 °C. Both hub domain constructs were purified as follows: after expression, cells were pelleted by centrifugation, resuspended in Buffer A (20 mM Tris pH8.0, 500 mM NaCl, 5 mM 2-mercaptoethanol (β-ME), and 20 mM imidazole), and then flash frozen in liquid nitrogen. Cells were later thawed, mixed with 10 μg/ml deoxyribonuclease I (Roche) and 0.8 mM phenylmethylsulfonyl fluoride, lysed with a cell disruptor (Avestin), and clarified by centrifugation at 4 °C and 30,000 × g for 30 minutes. Clarified supernatant was loaded onto Ni-NTA affinity resin (GE Healthcare), washed extensively with Buffer A, and then eluted with a linear gradient of 20 - 500 mM imidazole. The eluted protein was cleaved with ULP1 to remove the His_6-SUMO tag and then buffer exchanged into Buffer B (20 mM Tris pH 8.0, 100 mM NaCl, 5 mM β-ME, and 20 mM imidazole). The protein was then flowed over Ni-NTA resin to remove the cleaved tags. The protein was then concentrated and loaded onto a HiLoad Superdex 200 16/60 PG size exclusion column (GE Healthcare) equilibrated in Buffer C (20 mM Tris pH 8.0, 100 mM NaCl, 5 mM DTT). Protein containing fractions were pooled and concentrated prior to use in biophysical experiments or crystallization trials. For purifying the kinase domain, the same protocol was used with one additional step: prior to size exclusion chromatography protein was loaded onto a HiTrap Q HP anion exchange column (GE Healthcare), washed with Buffer C, and then eluted with a linear gradient of 0.1 – 2 M NaCl. Protein containing fractions were then pooled and concentrated.

Pelleting thermostability assay

3.6 µg samples of purified protein were incubated for 10 min at each indicated temperatures between 20 and 52 °C. Soluble and pellet fractions were separated by centrifugation at 30,000 × g for 10 min at 4 °C. All samples were analyzed with SDS-PAGE and Coomassie staining. Gel images were quantified using the FIJI software package (Schindelin et al., 2012) in order to calculate the % of protein in the soluble fraction at each temperature. Experiments were performed in triplicate.

Multiangle light scattering coupled to analytical size-exclusion chromatography

All experiments were conducted using the Superdex 200 10/300 GL column equilibrated in a buffer containing 25 mM TRIS (pH 7.5), 100 mM sodium chloride, and 5 mM DTT. The chromatography system was connected in series with an 18-angle light-scattering detector (DAWN HELEOS II, Wyatt Technology), a dynamic light-scattering detector (DynaPro Nanostar, Wyatt Technology), and a refractive index detector (Optilab t-rEX, Wyatt Technology). Data were collected at 21°C every 1 s at a flow rate of 0.4 ml/min and analysed using ASTRA 6 software, thereby generating molar mass and mass distribution (polydispersity) measurements of the samples (Wyatt, 1997). Protein-containing fractions were analysed by SDS-PAGE followed by Coomassie brilliant blue staining.

Crystallization and structure determination

A. pompejana CaMKII was crystallized at 23 °C with the hanging drop method using 1 µl of protein solution (10 mg/ml) and 1 µl of reservoir solution, containing 0.1 M TRIS (pH 8.0) and 40% (v/v) MPD. X-ray diffraction data were collected at beamline 12-2 at the Stanford Synchotron Radiation Lightsource (SSRL) with a wavelength of 1.033 Å and processed with the XDS package (Kabsch, 2010). The structure was solved by molecular replacement with Phaser, using the structure of the mouse CaMKII hub domain (PDB ID IHKX) as a search model (Hoelz et al., 2003; McCoy et al., 2007). The structure was refined using the PHENIX package with TLS refinement (Adams et al., 2010). The final structure was refined to R_{work} and R_{free} values of 17.2 and 20.6%, respectively, with excellent geometry (97.72% of

residues in favored region of the Ramachandran plot, 2.28% in the allowed region, and 0% outliers).

Illustrations and figures

All structural figures and movies were ray traced in PyMol (Schrödinger). Sequence alignments were generated using ClustalW and visualized with ALSCRIPT (Thompson et al., 1994; Barton, 1993).

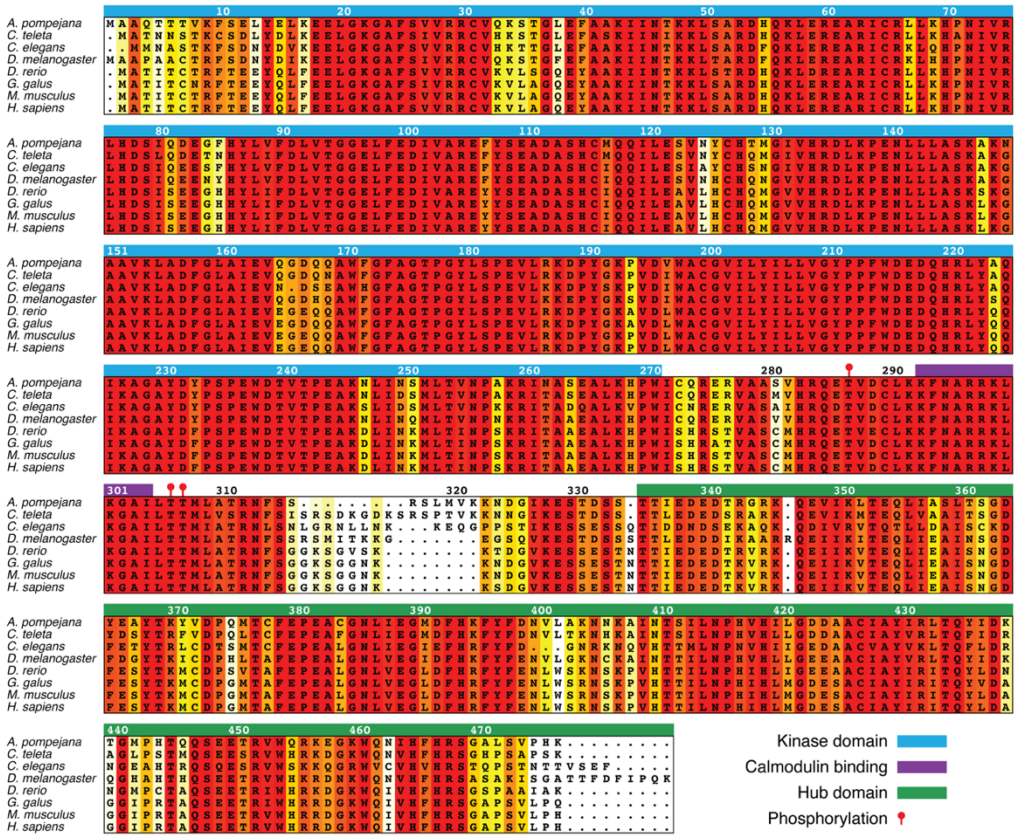


Figure 2.1. ApCaMKII encodes sequences features typical of other CaMKII homologues. ClustalW generated sequence alignment of various CaMKII isoforms including a putative CaMKII sequence from *A. pompejana*. The bottom 4 sequences all correspond to the vertabrate α isoform, and the sequences from *D. melanogaster* and *C. elegans* correspond to the C and D isoform respectively. Sequences are coloured by sequence similarity according to the BLOSUM62 matrix. <40% similar is indicated by white, 40-55% similar is indicated by yellow, and >55% similar is indicated by red. Phosphorylated sites are marked with red circles the calmodulin binding region is labelled with a purple rectangle.

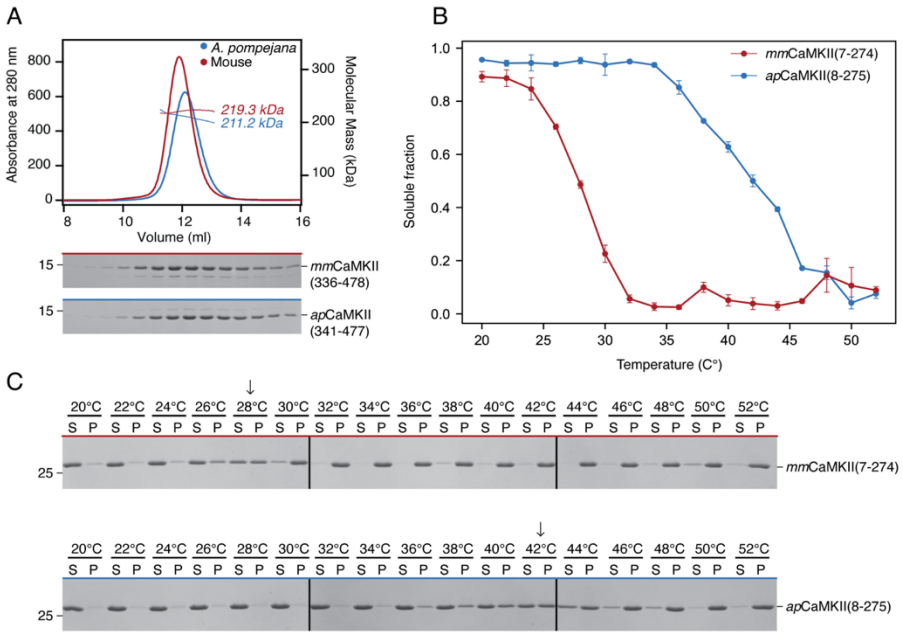


Figure 2.2. *Ap*CamKII oligomerizes similar to its mouse homologue and exhibits increased thermostability. (A) Size-exclusion chromatography coupled to multiangle light scattering profiles of purified *A. pompejana* and mouse CaMKII hub domains. (B) Purified *A. pompejana* and mouse CaMKII kinase domains were incubated at specific temperatures for 1 hour and then centrifuged to pellet aggregated protein. Samples of supernatant and pellet were then analyzed by SDS-PAGE and Coomassie staining followed by quantification to determine the fraction of soluble protein at each temperature. Error bars represent standard deviations from 3 replicates of each sample analyzed.(C) Coomassie stained SDS-PAGE gel from a single replicate of samples used to generate data in (B).

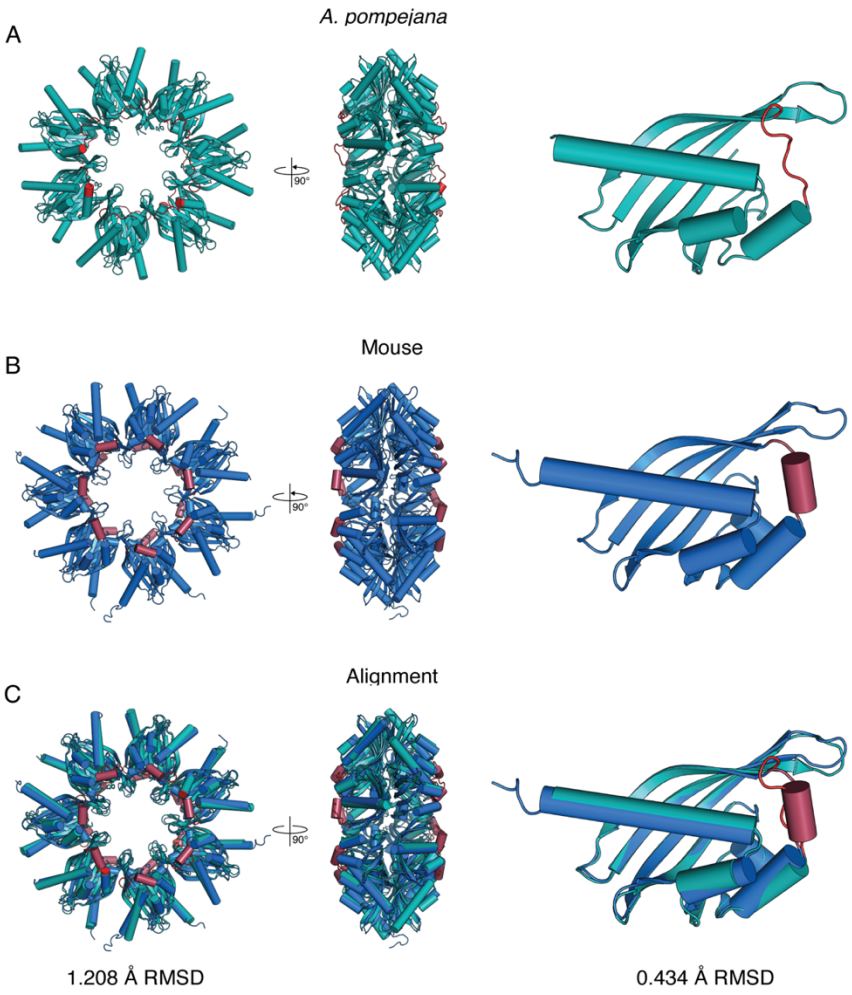


Figure 2.3. Geometry of CaMKII hub domain oligomeric assembly is highly conserved between *A. pompejana* and Mouse. Front and side views of the CaMKII hub domain tetradecameric ring (left and middle column) and a single hub domain protomer (right column) shown for (A) *A. pompejana* (this study), (B) mouse (b; PDB ID: 1HKX), and (C) a structural alignment of both isoforms.

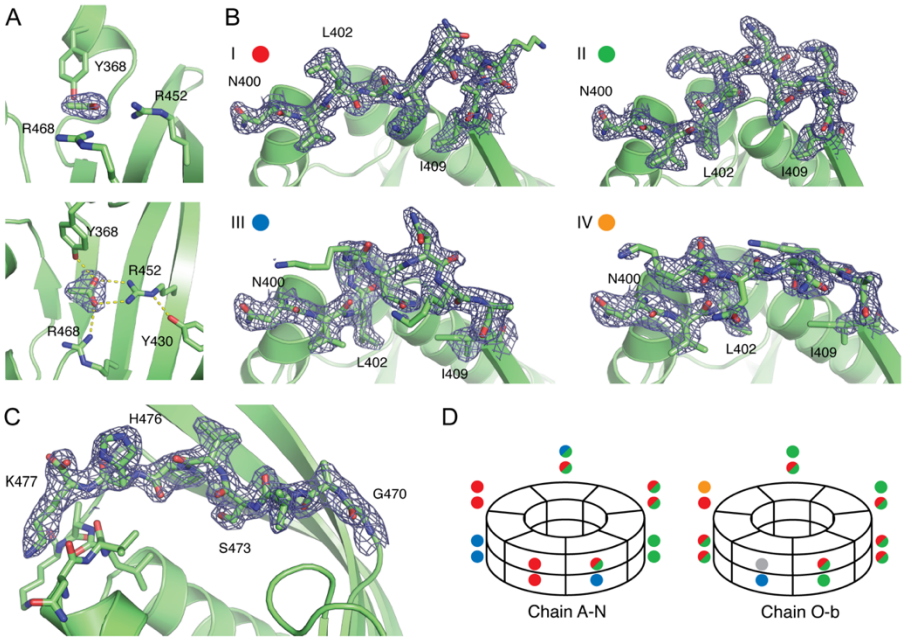


Figure 2.4. X-ray crystal structure of *Ap*CaMKII reveals unexpected ligand density and variable conformation of the α4-β3 loop. (A) The electron density from a $2F_o$-F_c map contoured at 1σ is shown around an acetate ion modelled in the CaMKII hub domain core along with interacting side chains displayed from the top and side. (B) The electron density from a $2F_o$-F_c map contoured at 1σ is shown around conformations I-IV for the α4-β3 loop. For each conformation, the following chain was used: chain A (I), chain V (II), chain D (III), chain O (IV). (C) The electron density from a $2F_o$-F_c map contoured at 1σ is shown around the ordered C-terminus in a chain where the α4-β3 loop adopts conformation II (chain V). (D) Diagram of ring 1 (chains A-N) and ring 2 (chains O-b) from the crystal structure labelled with colored dots corresponding to the specific α4-β3 loop observed in each chain. Dots with two colors represent instances where loops could be modelled as 2 alternative conformations. The loop could not be modelled in one chain which is labelled with a grey dot.

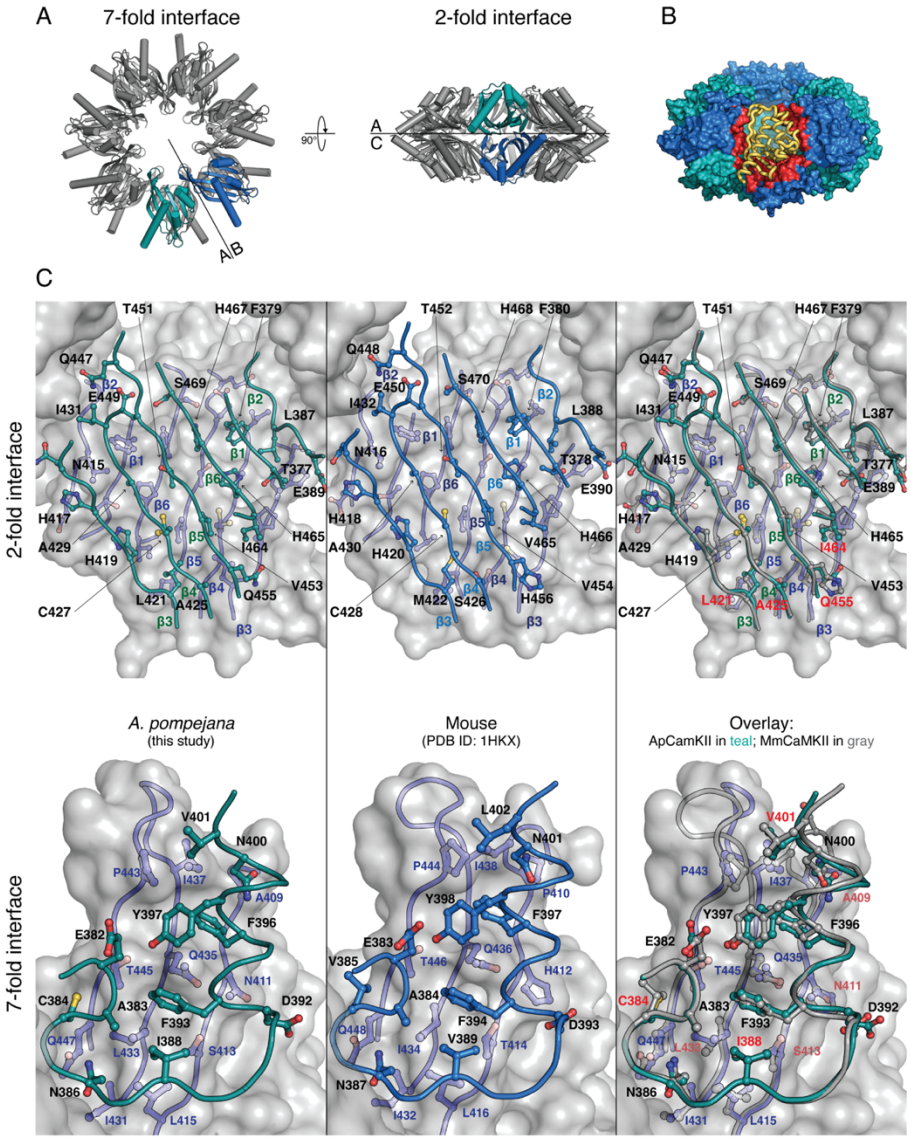


Figure S2.1. Detailed analysis of *Ap*CaMKII oligomerization interface. (A) Cartoon rendering of the CaMKII hub domain highlighting monomer interactions along the 7-fold symmetry interface (left) and the 2-fold symmetry interface (right). (B) Protomer A is shown in ribbon representation surrounded by a surface representation of the remaining protomers (interacting surfaces are highlighted in red). (C) Detailed view of 2-fold

interface (top) and the 7-fold interface (bottom). Each protomer is shown in ribbon representation with interacting side chains shown in ball and stick representation.

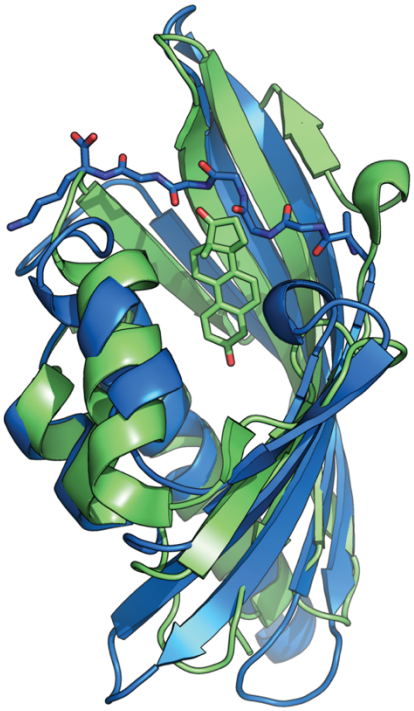


Figure S2.2. Comparison of *Ap*CaMKII hub domain fold with a ligand binding NTF2 superfamily member. Cartoon rendering of *A. pompejana* CaMKII hub domain in conformation II aligned to equilenin bound KSI (PDB ID: 1OH0) with an RMSD of 2.267 Å. The bound ligand shown in stick representation.

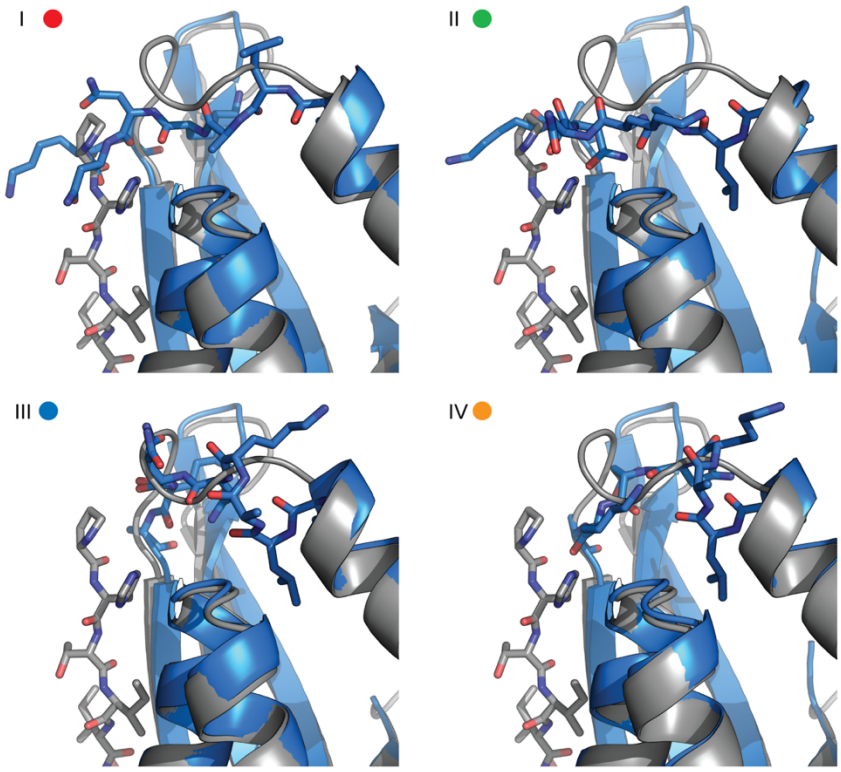


Figure S2.3. Some hub domain loop conformations are not compatible with the presence of an ordered hub domain N-terminal extension. Cartoon rendering of *A. pompejana* CaMKII hub domain subunits (blue) with conformations I-IV aligned to the *N. vectensis* CaMKII hub domain (grey) (PDB ID: 5IG5). The additional β-strand from the N-terminal extension of *N. vectensis* CaMKII and the α4-β3 in *A. pompejana* CaMKII are shown in stick representation.

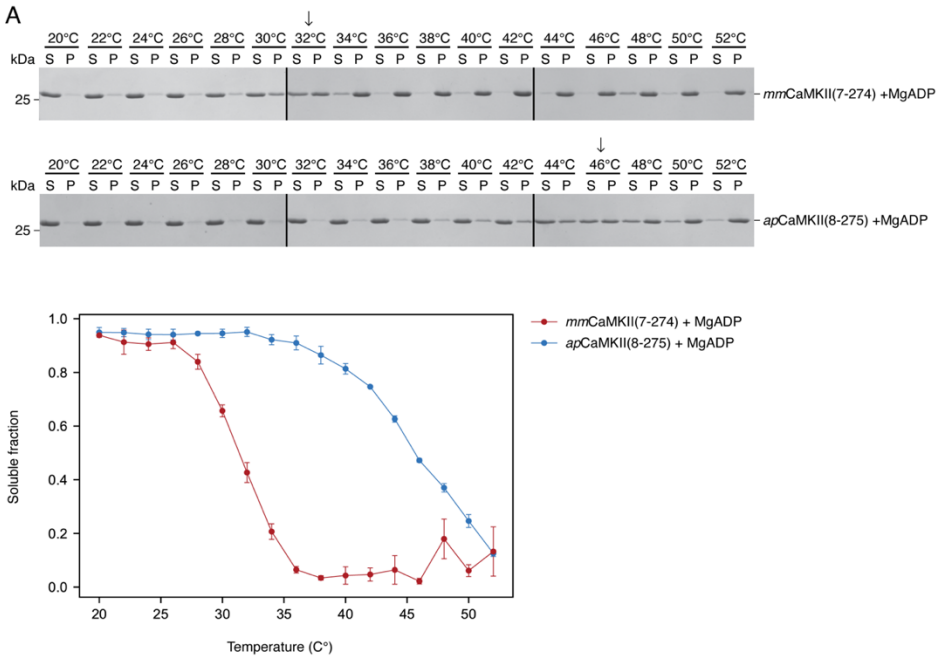


Figure S2.4. *Ap*CamKII exhibits increased thermostability in its ADP-bound state. (A) Purified *A. pompejana* and mouse CaMKII kinase domains were incubated at specific temperatures for 1 hour in the presence of 10 mM $MgCl_2$ and 1 mM ADP and then centrifuged to pellet aggregated protein. Samples of supernatant and pellet were then analyzed by SDS-PAGE and Coomassie staining followed by quantification to determine the fraction of soluble protein at each temperature. Error bars represent standard deviations from 3 replicates of each sample analyzed.(B) Coomassie stained SDS-PAGE gel from a single replicate of samples used to generate data in (A).

Chapter 3

A NANOBODY-BASED STRATEGY FOR RAPID AND SCALABLE PURIFICATION OF HUMAN PROTEIN COMPLEXES

The following chapter is adapted from Stevens et al., 2023 and modified according to the Caltech thesis format.

Taylor Anthony Stevens, Giovani Pinton Tomaleri, Masami Hazu, Sophia Wei, Vy N. Nguyen, Charlene DeKalb, Rebecca M. Voorhees, and Tino Pleiner. (2023). A nanobody-based strategy for rapid and scalable purification of human protein complexes. In press: *Nature Protocols*, bioRxiv.

3.1 Abstract

Native isolation of proteins in high yield and purity is a major bottleneck for analysis of their three-dimensional structure, function, and interactome. Here, we present a streamlined workflow for the rapid production of proteins or protein complexes using lentiviral transduction of human suspension cells, combined with highly-specific nanobody-mediated purification and proteolytic elution. (1) First, generation of a plasmid coding for a protein of interest fused to an N- or C-terminal GFP or ALFA peptide tag is rapidly achieved using the lentiviral plasmid toolkit we have designed. (2) Human suspension cell lines stably expressing the tagged fusion protein can be generated in <5 days using lentiviral transduction. (3) Leveraging the picomolar affinity of the GFP and ALFA nanobodies for their respective tags, proteins expressed even at low levels can be specifically captured from the resulting cell lysate in a variety of conditions, including detergents and mild denaturants. (4) Finally, rapid and specific elution of tagged or untagged proteins under native conditions is achieved within minutes at 4°C using an engineered SUMO protease. We demonstrate the wide applicability of the method by purifying multiple challenging soluble and membrane protein complexes to high purity from human cells. Our strategy is also directly compatible with many widely used GFP expression plasmids, cell lines and transgenic model organisms;

is faster than alternative approaches, requiring ~8 days from cloning to purification; and results in substantially improved yields and purity.

3.2 Introduction

A key prerequisite for many basic and pharmacological applications is the preparation of highly pure protein samples. Sample quality underlies the success of diverse experimental techniques including structural approaches (e.g. cryo-EM and X-ray crystallography), proteomics (e.g. interactome studies), and high-throughput studies of protein function (e.g. biophysical assays or *in vitro* drug screens).

Human proteins can be notoriously difficult to express and purify at the scale and purity needed for their structural and functional characterization. Their heterologous expression, e.g. in bacteria, frequently results in extensive degradation, insoluble aggregates or lack of essential post-translational modifications. These problems are exacerbated for multi-subunit protein complexes and membrane proteins, which may rely on mammalian specific factors for their biogenesis. Mammalian expression systems are thus in many cases the only alternative. However, culturing mammalian cells can be costly, making it essential to streamline both expression and protein purification workflows to maximize yield and purity, while saving valuable time and resources.

Here, we present a rapid and easily scalable workflow to purify soluble and membrane-spanning human proteins and protein complexes from human suspension cells for structural and functional analysis. Our workflow combines advances in both expression cell line generation (Elegheert et al., 2018), as well as purification strategy (Pleiner et al., 2020), and employs two highly versatile and complementary affinity tags – GFP and the small ALFA peptide tag (Götzke et al., 2019).

First, we describe how to rapidly generate stable polyclonal human suspension cell lines expressing a GFP- or ALFA-tagged protein of interest using lentiviral transduction. The use of lentivirus combines multiple advantages of both transient transfection and monoclonal cell

line generation(Elegheert et al., 2018). (1) Lentivirus is fast and cost-efficient, stable cell lines can be generated within <5 days using lentivirus produced from a single well of a 6-well plate. (2) It enables highly tunable expression, where transduction efficiencies can be titrated to either >80-90% to maximize yield, or to <30% to ensure a single copy of cDNA per cell. When optionally coupled with FACS, polyclonal cell lines can be generated that contain uniform expression across all cells. Finally (3), lentivirus ensures reproducible and scalable expression, because once generated, cell lines can be stored or expanded indefinitely. To streamline the cloning of suitable expression plasmids, we provide an extensive plasmid toolbox, an accompanying cloning guide, as well as general recommendations for GFP/ALFA tag placement.

Second, we present a nanobody-based strategy for the native isolation of human proteins that routinely results in higher yield and purity than canonical epitope tag-based approaches. This is achieved through the combination of high-affinity binding to anti-GFP or ALFA nanobodies with selective elution by an engineered SUMO protease ($SENP^{EuB}$) (Vera Rodriguez et al., 2019). $SENP^{EuB}$ allows nearly 1,000-fold faster release of resin-bound proteins than TEV or 3C proteases, and elution can therefore be performed quickly under gentle conditions on ice. Rapid and selective elution of nanobody-captured proteins ensures high sample purity and preserves delicate protein complexes. We provide bacterial expression plasmids and protocols to inexpensively produce both nanobodies and protease in amounts sufficient for hundreds of purifications.

Through the combination of lentivirus-based cell line generation and nanobody-mediated affinity purification our protocol reduces the time and cost required to prepare high-quality protein samples. Together these strategies can therefore significantly accelerate the structural and functional analysis of otherwise difficult to express human protein complexes.

Applications

One advantage of this protocol is that the nanobody-based purification strategy is highly flexible and modular. It can be used for purification of proteins or protein complexes under

a variety of conditions and from lysates prepared from various eukaryotic or prokaryotic sources. In particular, utilizing the anti-GFP nanobody for protein capture can leverage the vast number of existing sources for GFP-tagged proteins (e.g. yeast, flies, worms and mice). Publicly available and commercial sources are enumerated in **Extended Data Figure 3.1**. Further, the nanobody•affinity tag pair can easily be exchanged to other natural, laboratory-evolved or computationally designed binding proteins, as well as protein-specific binders(Pleiner et al., 2015; Aksu et al., 2018), making the system highly adaptable. A selection of potential options is listed in **Extended Data Figure 3.2**. We therefore envision that this strategy will be suitable for any structural or functional application requiring protein purification from a cell, extract, or tissue.

One particularly powerful application of this protocol is generating samples for structural analysis, which requires large scale production of highly pure samples (Pleiner et al., 2020; Chen et al., 2022). For example, we have used this strategy to purify the nine-subunit human ER membrane protein complex (EMC), for structure determination using single particle cryo-EM (Pleiner et al., 2020). For the EMC, and many protein complexes, we have found that expression of a single tagged subunit results in its incorporation into the intact complex in place of the endogenous subunit. Any excess unassembled subunits are typically degraded by the ubiquitin-proteasome pathway, as has been previously described (Juszkiewicz & Hegde, 2018; Pleiner et al., 2021). Therefore, by exploiting endogenous protein quality control machinery in human cells, it is often possible to introduce a tag into a protein complex without expressing multiple subunits or first generating knock out cell lines.

Similarly, our protocol can be used to generate highly pure samples for functional assays (e.g. to characterize a protein's enzymatic activity, stability or DNA/RNA/lipid-binding properties). We recently used it to isolate the 33 kDa mitochondrial outer membrane insertase MTCH2 to reconstitute its protein-insertion activity *in vitro* (Guna et al., 2022).

Further, because of the high specificity, efficient capture, and rapid elution, this protocol is also well-suited for analysis of physical interaction partners by mass spectrometry.

Overview of the procedure

Our protocol comprises the following steps that can be carried out in parallel. Additional details are outlined in the experimental design section (**Fig. 3.1**):

(i) Cloning of the protein of interest into a lentiviral transfer plasmid as a GFP- or ALFA tag fusion (for ease of use a plasmid toolbox is provided).
(ii) Generation of lentivirus and transduction of human suspension cells to stably integrate the desired open reading frame.
(iii) Recombinant expression of anti-GFP or ALFA nanobodies and protease in *E. coli* followed by Ni^{2+}-chelate affinity purification.
(iv) Nanobody-mediated purification from the expanded suspension cell culture.

Two different purification strategies – TagON and TagOFF

We present two different strategies to purify a GFP- or ALFA-tagged protein in either a tagged (TagON) or scarless untagged form (TagOFF) (**Fig. 3.2**). Both strategies employ distinct biotinylated anti-GFP or ALFA nanobody fusion proteins that are immobilized onto streptavidin beads to capture GFP- or ALFA-tagged proteins from cell lysate.

In the TagON strategy (**Fig. 3.2a**), a $SUMO^{Eu}$ module is inserted between the biotin acceptor peptide (Avi) tag and nanobody, so that $SENP^{EuB}$ cleavage releases the nanobody along with its bound protein. The eluted protein therefore retains the fluorescent GFP tag, which may be useful for detection during downstream applications such as Fluorescence detection size exclusion chromatography (FSEC) (Kawate & Gouaux, 2006). In the TagOFF strategy (**Fig. 3.2b**), a $SUMO^{Eu}$ module is placed in between GFP/ALFA tag and the protein so that cleavage by $SENP^{EuB}$ only releases the protein, while both nanobody and tag are retained on the resin. This strategy is useful if the presence of the tag interferes with downstream applications.

Purification of soluble and membrane-bound multi-subunit protein complexes

To demonstrate the utility of both TagON and TagOFF, we used the nine-subunit EMC as a model substrate. GFP- or ALFA-tagged EMC2 or EMC5 are incorporated into the EMC in place of their respective endogenous subunits, allowing the intact complex to be isolated in high purity (**Fig. 3.3a-b**). Using the TagOFF strategy, capture of GFP-SUMO^{Eu}-EMC2 with a non-cleavable anti-GFP nanobody allowed native isolation of completely untagged EMC following SENP^{EuB} cleavage, while the affinity tag and nanobody were retained on the beads.

As a proof-of-concept, we successfully isolated other challenging soluble and membrane protein complexes by tagging a single subunit (**Fig. 3.3c**). For example, GFP-SRP72 efficiently incorporated into the ribonucleoprotein signal recognition particle (SRP), which could be isolated in its native form. Purified SRP was fully functional and bound to stalled ribosome nascent chains exposing a transmembrane domain in a hydrophobicity sensitive manner as observed before for native SRP (Voorhees & Hegde, 2015) (**Fig. 3.3d**). Similarly, we used this strategy to isolate the entire 26S proteasome from cells expressing GFP-tagged RPN11, the ER-resident membrane protein complex NOMO-NCLN-TMEM147 via TMEM147-GFP, and the plasma membrane localized anion exchanger SLC4A2/AE2.

Comparison with other methods

The advantages of this purification protocol over existing methods stem primarily from combining two innovations.

First, we achieve higher sample purity by using highly specific nanobodies and affinity beads. Our approach exploits the picomolar affinity between the nanobodies and their epitope tags to selectively capture tagged proteins from cell lysate. Additionally, both the nanobody fusion proteins and affinity resin are fully orthogonal to eukaryotic and prokaryotic host proteins. For example, the nanobody-decorated streptavidin beads are passivated by blocking any excess binding sites with free biotin or charged PEGylated biotin derivatives. Therefore,

in contrast to the TwinStrep:Streptactin system, our strategy does not capture endogenous biotinylated proteins.

Second, rapid and selective protease elution of nanobody-captured proteins additionally ensures high sample purity and quality. In contrast, commonly used epitope tag-binding monoclonal antibodies (e.g. FLAG, HA, Myc, or V5), rely on native elution using excess epitope peptide. This is often highly inefficient and requires prolonged incubation at room temperature, which releases background binders and can cause aggregation of fragile proteins or dissociation of weaker binding partners. A side-by-side comparison shows that EMC isolated from an EMC5-GFP cell line using $SENP^{EuB}$ cleavage is substantially purer than EMC isolated from an EMC5-3xFLAG cell line using traditional anti-FLAG purification (**Fig. 3.4a**). The combination of specific nanobodies, passivated affinity resin and protease elution therefore yields high final sample purity.

Protease cleavage is thus a powerful alternative over inefficient competitive peptide elution. However, common proteases like human rhinovirus (HRV) 3C protease or Tobacco etch virus (TEV) protease inefficiently cleave resin-bound proteins, especially on ice. Often large amounts of contaminating protease and lengthy overnight incubations (>13h) are needed to circumvent these issues. In contrast, low nanomolar concentrations of $SENP^{EuB}$ are sufficient to gently release resin-bound proteins on ice within a few minutes (**Fig. 3.4b**).

Through both innovations, our protocol thus improves the signal-to-noise ratio of affinity purifications and better preserves sensitive protein complexes and their transient interaction partners. This yields less false interactors for affinity purification mass spectrometry (AP-MS) experiments and should also reduce contaminants that may interfere with functional experiments.

3.3 Experimental design

Optimal introduction of a tag on a protein subunit or complex

The first critical consideration for the purification of a multi-subunit protein complex is to choose a subunit to tag. We frequently found that ectopic expression of a single tagged subunit of a protein complex in human cells can result in the replacement of its endogenous counterpart. This strategy yields stoichiometric complexes in human cells when efficient cellular quality control mechanisms exist that degrade excess tagged and endogenous subunits of the chosen complex. Existing functional and structural data can often be used to inform subunit choice. For example, scaffold subunits of protein complexes typically contain multiple hydrophobic interfaces and are especially short-lived in the unassembled state, ensuring rapid elimination of excess subunits. Such subunits might thus represent prime candidates for first tagging trials (Pleiner et al., 2021) and can be identified from mass spectrometry studies that analyzed proteome-wide degradation kinetics (McShane et al., 2016). More stable subunits can accumulate after overexpression and may need to be removed by additional purification steps.

The next important consideration is to identify the optimal location of the affinity tag to minimize its interference with protein function. For protein complexes, the tag must be compatible with complex assembly and integrity. Compatibility can often be inferred from prior published studies or large-scale GFP-tagging efforts (**Extended Data Figure 3.1**), provided a careful analysis of localization and function was performed. Alternatively, existing structures or structural models of the proteins can be analyzed. A flexible unstructured terminus that is not part of a folded domain can usually tolerate a protein tag. Conversely, regions with high sequence conservation might indicate a functional requirement or binding site. In general, smaller peptide tags (like the ALFA tag) tend to be less disruptive than larger globular tags. The ALFA tag can also be placed internally into exposed loops and thus provides an alternative for proteins that cannot be tagged at either terminus.

If the presence of a tag will interfere with downstream applications, the TagOFF strategy should be used to completely remove the tag during purification. For this, the appropriate non-cleavable nanobody fusion protein needs to be generated and combined with a protein

expression vector that contains a protease cleavage site. As described below, N-terminal tags can be removed completely (scarless), yet C-terminal tag removal will always generate a cleavage scar.

Cloning of a lentiviral transfer plasmid encoding the tagged protein of interest

Once a tagging and purification strategy is identified, the coding sequence of the protein of interest needs to be cloned into a lentiviral transfer plasmid. We provide an extensive toolbox of 2nd generation lentiviral transfer plasmids to facilitate fusion of a protein of interest to a GFP or ALFA tag. All plasmids are available via Addgene (see **Table 3.1** for IDs). Each tag is either N-terminal (pTS93-pTS102) or C-terminal (pTS103-pTS116) and either non-cleavable for TagON or cleavable for TagOFF purification. A detailed cloning guide is provided in the supplements (**Supplementary Data 1**).

All constructs contain a CMV promoter fused to two Tet operator sequences (CMV-TetO2), which are Tet repressor (TetR) binding sites. The CMV-TetO2 promoter is thus doxycycline-inducible only in a TetR+ cell line, but will otherwise express constitutively.

We strongly recommend homemade $SENP^{EuB}$ protease for tag removal, but we also provide plasmids encoding the cleavage sites of the commercially available 3C, TEV and SUMOstar proteases. Because these proteases all cleave at the C-terminus of their recognition site, N-terminal tags can thus be nearly completely removed. For C-terminal tags, however, this results in a cleavage scar. The scars of 3C and TEV protease (both 6 amino acids [aa]) are smaller than those of the $SENP^{EuB}$ (96 aa) and SUMOstar proteases (99 aa). If desired these larger C-terminal scars can be removed after elution by an additional 3C or TEV protease cleavage step in solution. Our SUMO protease-cleavable C-terminal GFP or ALFA-tag encoding plasmids thus contain additional 3C or TEV cleavage sites between multiple cloning site and SUMO module.

One consideration in construct selection is that to maximize protein yield it can be beneficial to isolate successfully transduced cells via fluorescence-activated cell sorting (FACS) to

obtain a homogenous population. While cells expressing GFP-tagged proteins can easily be sorted using GFP fluorescence, we additionally equipped expression vectors encoding ALFA-tagged proteins with a TagBFP expression cassette. TagBFP is separated from the ALFA-tagged protein using a porcine teschovirus P2A sequence, which mediates peptide bond skipping by the ribosome (de Felipe et al., 2006) and thus results in the efficient synthesis of two separate proteins from a single mRNA – TagBFP and the ALFA-tagged protein.

Generation of stable human suspension cell lines using lentiviral transduction

Lentiviral particles can be generated by co-transfecting adherent Lenti-X 293T cells with a 2nd generation lentiviral transfer plasmid encoding a GFP- or ALFA-tagged protein of interest and 2nd generation packaging (e.g. psPAX2, Addgene #12260) and envelope plasmids (e.g. pMD2.G Addgene #12259). Transfected cells secrete lentiviral particles into the culture medium that are pseudotyped with the VSV-G envelope protein, which binds the widely expressed LDL receptor on the target cell surface and thus confers broad host cell tropism.

Lentiviral particles can then be used to transduce multiple human suspension and adherent cell lines that are compatible with our purification protocol. We favor a very rapid lentiviral transduction approach (Elegheert et al., 2018) to generate Expi293F cell lines expressing a GFP or ALFA-tagged protein of interest. This approach is fast, easy to use and generates stable cell lines that robustly express a protein of interest. We prefer Expi293F (TetR-) and Expi293F inducible cell lines (TetR+) (Thermo Fisher Scientific, USA), because they maintain high viability even at very high cell density and thus maximize the yield of cells per liter of culture medium. To express secreted or membrane proteins without complex N-glycans, the N-acetylglucosaminyltransferase I knockout (GnTI-) derivatives of these cell lines should be used.

During infection, lentiviral particle-containing culture medium is simply mixed with human Expi293F suspension cells. This leads to the stable integration of the open-reading frame

encoded on the transfer plasmid into a random genomic location. For first small-scale purification trials, we recommend expanding transduced cells to medium density in ~50 ml. Once optimal conditions are found, purifications from ~1-2 L high density cultures often provide enough material for structural and functional assays.

If lentiviral work is not feasible, multiple alternative approaches can be used to instead generate human suspension cell lines expressing a GFP- or ALFA-tagged protein of interest. These include for example transient transfection using polyethylenimine or Expifectamine (Thermo Fisher Scientific, USA), baculovirus transduction(Goehring et al., 2014) and adaptation of stable adherent cell lines to suspension growth(Chaudhary et al., 2012). Such stable adherent cell lines can be generated by Flp-In recombination(O'Gorman et al., 1991), antibiotic selection(Chaudhary et al., 2012) or CRISPR knock-in (Leonetti et al., 2016; Koch et al., 2018; Cho et al., 2022).

Cell lysis method and purification conditions

The subcellular localization and stability of the protein of interest dictates the appropriate lysis method to prepare cell extracts for affinity purification. Mechanical lysis is appropriate to purify soluble cytosolic or nuclear proteins. Nuclear DNA-bound proteins additionally need to be dissociated from DNA using a high salt lysis buffer (Nilsen, 2013). Membrane proteins must be solubilized from whole cells or enriched membrane fractions using detergents or other membrane mimetics (Seddon et al., 2004; Lee et al., 2016). The choice of detergent for solubilization can affect membrane protein extraction efficiency and stability.

Another critical factor to consider is the protein's intrinsic stability. In general, lysis buffers with slightly elevated salt concentration (> ˜150 mM) and detergent can strongly reduce non-specific binding of lysate proteins to affinity resins. For example, the widespread RIPA lysis buffer contains high concentrations of ionic (deoxycholate and SDS) and non-ionic detergents (Triton-X-100 or NP-40) and is used for both soluble and membrane protein purification. However, some proteins and protein complexes can be very sensitive to high

ionic strength or detergents. To determine the optimal conditions for a protein or complex of interest, we recommend optimizing lysis buffer composition in pilot small-scale purification trials.

Protein purification by nanobody capture and protease release

The major components of our purification strategy are highly stable proteins that are compatible with most commonly used lysis buffer compositions.

For example, we use the anti-GFP nanobody Enhancer (12.8 kDa), which binds GFP with high affinity (590 pM) and specificity in many different systems(Rothbauer et al., 2006; Kirchhofer et al., 2010). Various widespread GFP variants (including wildtype GFP, EGFP and superfolder GFP (Pédelacq et al., 2006)), as well as closely related fluorescent proteins are compatible (**Extended Data Figure 3.3**). The ALFA nanobody (13.4 kDa) binds a rationally designed, 15 amino acid α-helical peptide with very high affinity (26 pM) (Götzke et al., 2019). Both nanobodies, as well as GFP and ALFA tags are stable and bind efficiently even under harsh conditions, including high concentrations of most detergents, salt and even urea (**Extended Data Figure 3.4**).

For our purification strategies, both nanobodies are immobilized onto Streptavidin beads using a fused biotin acceptor peptide (Avi tag), which is modified with a single biotin by the biotin ligase BirA either during expression in *E. coli* or after purification *in vitro* (Schatz, 1993; Beckett et al., 1999; Fairhead & Howarth, 2015). As one of the strongest non-covalent interactions in nature, the biotin-streptavidin interaction withstands harsh purification conditions.

Finally, our protocol makes use of the $SUMO^{Eu}$ protein (10.5 kDa), which was engineered to be resistant against cleavage by endogenous eukaryotic SUMO proteases (Vera Rodriguez et al., 2019). $SUMO^{Eu}$-tagged nanobodies or proteins are thus stable in a variety of eukaryotic cells and cell extracts. The more commonly used yeast SUMO is useful for expression in prokaryotic systems, but is very rapidly cleaved in eukaryotic cells and cell lysate (Frey &

Görlich, 2015; Vera Rodriguez et al., 2019). The engineered $SUMO^{Eu}$•$SENP^{EuB}$ pair therefore extends the use of the SUMO protease technology to eukaryotes. A similar, but completely orthogonal SUMOstar system (Liu et al., 2008) is commercially available (LifeSensors, USA), but slightly less resistant against host SUMO protease cleavage (Vera Rodriguez et al., 2019).

Preparation of the nanobody fusion proteins and $SENP^{EuB}$ protease

All recombinant proteins required for this protocol can easily be purified in very high yield and purity from single 1 L *E. coli* cultures. Such preparations generate sufficient reagents for hundreds of purifications from eukaryotic cells. We provide expression plasmids and protocols for the generation of $SENP^{EuB}$ protease, as well as biotinylated anti-GFP and anti-ALFA nanobodies with or without the $SUMO^{Eu}$ module in between the Avi-tag and nanobody (see Table 3.1 for Addgene IDs).

The anti-GFP nanobody can easily be biotinylated during expression in the *E. coli* strain CVB101 (Avidity LLC, USA), which allows IPTG-inducible co-expression of the biotin ligase BirA. The anti-ALFA nanobody, however, expresses better in the *E. coli* Rosetta-gami2 strain (Millipore-Sigma, USA), which does not express BirA. Purified ALFA nanobody therefore needs to be biotinylated with purified BirA in solution. For this, we provide a BirA expression plasmid (pTP264) and purification protocol. BirA can, however, also be obtained commercially (Avidity LLC, USA).

Affinity resins containing immobilized (non-cleavable) anti-GFP or ALFA nanobodies are commercially available (ChromoTek, Germany; NanoTag Biotechnologies, Germany) and are directly compatible with the TagOFF strategy.

Controls

If our protocol is to be used for the identification of interaction partners via mass spectrometry, it is important to include controls to account for trace contaminants that non-specifically bind to the affinity tag, nanobody or affinity resin. A good control is a matched

cell lysate expressing solely the affinity tag without a fused protein i.e. only GFP (e.g. generated using plasmid pTP924). Less optimally, the cell lysate containing the protein of interest can be split into two even fractions and additionally incubated with beads that either do not contain any immobilized nanobody or that contain a non-specific nanobody, e.g. anti-ALFA nanobody for a GFP-tagged protein.

Similarly, for functional assays, it is important to control that the observed activity of the purified protein is not caused by a contaminant. If possible, a catalytically dead mutant should be purified in parallel under identical conditions.

The sorting of cell lines transduced with multiple different lentiviruses, encoding compatible fluorescent proteins, might require multicolor compensation to correct for cross-channel fluorescence bleed-through. Non-transduced, as well as single color control cells are typically required for compensation and also allow efficient sort gate placement. Plasmids pTP341 and pTP924, which express TagBFP and GFP, respectively, from a constitutive CMV promoter, can be used for this purpose.

Limitations

One limitation of the lentiviral transduction approach is that the ectopic expression of a protein complex subunit can in some cases result in its purification in excess over other endogenous complex members. This especially affects subunits that are stable in the unassembled state. Frequently, these excess subunits can be separated by an additional size-exclusion chromatography purification step. However, multiple preventative measures can improve protein complex yield and stoichiometry. 1) Freestyle 293-F suspension cells can be used instead of Expi293F cells (Thermo Fisher Scientific, USA). We observed that overexpression of unassembled subunits is more limited in these cells, likely due to tighter regulation by the ubiquitin-proteasome system. However, Freestyle 293-F cells cannot be grown to as high density as Expi293F cells, decreasing yield per liter of expression. 2) Because our lentiviral vectors are regulated by doxycycline-inducible promoters, the expression level can also be fine-tuned by varying induction time in Expi293F TetR+ cells.

3) The expression level can be reduced by exchanging the strong CMV promoter for a weaker one (Qin et al., 2010) and/or by removing the expression-enhancing WPRE element. 4) For membrane proteins it can be beneficial to first enrich a membrane fraction to remove excess non-incorporated and aggregated subunits in the cytosol. Purification from the solubilized membrane fraction can therefore yield higher sample quality and better complex stoichiometry.

If these approaches do not yield the expected outcome, it is possible to overexpress multiple subunits in parallel. This can be achieved in three different ways: 1) via co-transduction with multiple lentiviral particles, 2) via fusion of multiple subunits into a single expression plasmid using P2A sites or 3) via a combination of 1-2. The efficiency of dual or triple transduction is typically lower than transduction with a single lentivirus. Multi-color sorting via FACS is therefore required to obtain a fully transduced cell line. Our lentiviral transfer plasmid toolbox already allows generating compatible GFP and BFP expression plasmids. A third compatible plasmid can easily be generated by introducing mCherry.

Constitutive ectopic expression of a protein can be toxic or reduce cellular fitness. In a heterogenous cell population containing untransduced cells, this can lead to a gradual loss of more slowly growing transduced cells. In such cases, using the inducible expression system is likely to alleviate the negative selection pressure resulting from the constitutive expression of a toxic protein. Alternatively, successfully transduced cells can be sorted via FACS to remove faster growing untransduced cells.

The yield of lentiviral particles decreases sharply with increasing transfer plasmid size, resulting in strongly reduced titers for transfer plasmids encoding large proteins e.g. inserts between 5’ and 3’ Long-terminal repeats (LTRs) above 8 kbp. After subtraction of essential elements this leaves around 5 kbp for the GFP- or ALFA-tagged protein of interest (˜180 kDa). If necessary, removal of the WPRE element can yield an extra ˜600 bp.

Expertise needed to implement the protocol:

Our purification protocol requires only basic skills in recombinant protein production in *E. coli*, as well as cell culture handling and can even be carried out by qualified undergraduate students with prior training. The generation of stable human suspension cell lines using lentiviral transduction, however, requires special biosafety training and conditions in compliance with the relevant institutional and governmental biosafety regulations. Elegheert and colleagues expertly summarized common safety practices associated with lentiviral work (Elegheert et al., 2018). However, if use of lentivirus is not feasible, multiple other strategies for plasmid delivery can be used instead (see above).

3.4 Material

Cell lines

- Lenti-X 293T cell line (Takara Bio Inc., Japan; cat. no. 632180)
- Gibco Expi293F cells (TetR-; Thermo Fisher Scientific, USA; cat. no. A14527)
- Gibco Expi293F GnTI- cells (TetR-; Thermo Fisher Scientific, USA; cat. no. A39240)
- Gibco Expi293F inducible cells (TetR+; Thermo Fisher Scientific, USA; cat. no. A39241)
- Gibco Expi293F inducible GnTI- cells (TetR+; Thermo Fisher Scientific, USA; cat. no. A39242)
- Gibco Freestyle 293-F cell line (Thermo Fisher Scientific, USA; cat. no. R79007)

E. coli strains

- NEBExpress I^q chemically competent cells (New England Biolabs, USA; cat. no. C3037I)
- CVB101 chemically competent cells (Avidity, USA)
- Stellar competent cells (Takara Bio Inc., Japan; cat. no. 636763)

  CRITICAL Cloning DNA with repetitive elements, like our lentiviral transfer plasmid toolbox, requires a recA- *E. coli* strain like *E. coli* Stellar.
- Rosetta-gami 2 competent cells (Novagen / Millipore-Sigma, USA; cat. no. 71350-3)

Reagents

- Gibco Expi293 Expression (Expi) Medium (Thermo Fisher Scientific, USA; cat. no. A14351-01)
- Gibco FreeStyle 293 expression medium (Thermo Fisher Scientific, USA; cat. no. 12338018)
- Gibco DMEM, high glucose, no glutamine (Thermo Fisher Scientific, USA; cat. no. 11960051)
- Gibco L-glutamine 200 mM (100x Gln) (Thermo Fisher, cat. no. 25030081)
- Gibco Penicillin-Streptomycin 5000 U/mL (100x Pen-Strep) (Thermo Fisher Scientific, USA; cat. no. 15070063)
- HyClone fetal bovine serum (FBS) (Cytiva, USA; cat. no. SH30071.03)
- Gibco DPBS, no calcium, no magnesium (Thermo Fisher Scientific, USA; cat. no. 14190136)
- Gibco Trypsin-EDTA (0.25%), phenol red (Thermo Fisher Scientific, USA; cat. no. 25200056)
- Dimethyl Sulfoxide (DMSO) (Thermo Fisher Scientific, USA; cat. no. A13280.36)
- Gibco Opti-MEM I (1x) Reduced Serum Medium (Thermo Fisher Scientific, USA; cat. no. 31985-062)
- TransIT-293 Transfection Reagent (Mirus Bio, USA; cat. no. MIR 2704)
- LB Broth, Miller (Fisher Scientific, USA; cat. no. BP1426-2)
- LB Agar, Miller (Powder) (Fisher Scientific, USA; cat. no. BP1425-500)
- Tryptone (Fisher Scientific, USA; cat. no. BP9726-5)
- Yeast extract, granulated (Fisher Scientific, USA; cat. no. BP9727-5)
- SOC Recovery Medium (Thermo Fisher Scientific, USA; cat. no. 15544034)
- Carbenicillin (Disodium Salt) (Fisher Scientific, USA; cat. no. BP26485)
- Kanamycin Sulfate (Fisher Scientific, USA; cat. no. BP906-5)
- Chloramphenicol (Fisher Scientific, USA; cat. no. BP904-100)
- IPTG (Millipore-Sigma, USA; cat. no. I6758-5G)
- Imidazole (Millipore-Sigma, USA; cat. no. I2399-500G)

- NaCl (Fisher Scientific, USA; cat. no. S9888-5KG)
- Tris Base, Trizma (Millipore-Sigma, USA; cat. no. T6066-5KG)
- HEPES (Millipore-Sigma, USA; cat. no. H3375-500G)
- 1 M Magnesium Acetate solution (Millipore-Sigma, USA; cat. no. 63052-100ML)
- Magnesium Chloride (VWR, USA; cat. no. MK5958-04)
- Potassium Acetate (Millipore-Sigma, USA; cat. no. 60035-1KG)
- Hydrochloric acid, concentrated 37%/12N (Millipore-Sigma, USA; cat. no. HX0603-3)
- Potassium hydroxide pellets (VWR, USA; cat. no. 6984-06)
- Glycerol (Fisher Scientific, USA; cat. no. BP229-4)
- Sucrose (Millipore-Sigma, USA; cat. no. 50389-5KG)
- DTT (Millipore-Sigma, USA; cat. no. D9163-25G)
- Ethanol 200 Proof (VWR, USA; cat. no. TX-89125172CAL)
- Methanol (VWR, USA; cat. no. BDH1135-4LG)
- PMSF (Thermo Fisher Scientific, USA; cat. no. 36978)
- Ribonuclease A from bovine pancreas (Millipore-Sigma, USA; cat. no. R6513-50MG)
- Adenosine 5′-triphosphate dipotassium salt hydrate (Millipore-Sigma, USA; cat. no. A8937-1G)
- D-Biotin (Millipore-Sigma, USA; cat. no. B4501-1G)
- $dPEG_{24}$-biotin acid (Quanta Biodesign, USA; cat. no. 10773)
- Triton-X-100 (Millipore-Sigma, USA; cat. no. X100-500)
- Tween-20 (Millipore-Sigma, USA; cat. no. P1379-1L)
- GDN (Anatrace, USA; cat. no. GDN101 25 GM)
- Doxycycline hyclate (Millipore-Sigma, USA; cat. no. D9891-1G)
- 20 % (w/v) SDS (VWR, USA; cat. no. 97062-442)
- Bromophenol Blue (Millipore-Sigma, USA; cat no. B0126-25G)
- Urea (Millipore-Sigma, USA; cat. no. U0631-500G)
- Roche cOmplete, Mini, EDTA-free protease inhibitor cocktail (Millipore Sigma, USA; cat. no. 11836170001)
- Ni-NTA agarose resin 25 ml (Qiagen, Germany; cat. no. 30210)

- Pierce Streptavidin magnetic beads (Thermo Fisher Scientific, USA; cat. no. 88816)

Plasmids

- pMD2.G lentivirus envelope plasmid (Addgene ID #12259)
- psPAX2 lentivirus packaging plasmid (Addgene ID #12260)

Sequencing primers

- CMV forward (5'– CGCAAATGGGCGGTAGGCGTG –3')
- WPRE reverse (5'– GTTGCCTGACAACGGGCC –3')
- GFP C-terminus forward (5'– GGAGACGGTCCCGTCCTC –3')
- BFP C-terminus forward (5'– GATACTGCGACCTCCCTAGC –3')
- GFP N-terminus reverse (5'– TGGCCATTCACGTCTCCGTC –3')
- BFP N-terminus reverse (5'– CTTGAAGTGATGGTTGTCCACGGTGC –3')

Equipment

- 125 mL Erlenmeyer flask, vent cap, plain bottom, PETG (Celltreat, cat. no. 229801)
- 490 cm^2 tissue culture treated roller bottle, vented cap (1 L roller bottle; Celltreat, cat. no. 229383)
- 850 cm^2 tissue culture treated roller bottle, vented cap (2 L roller bottle; Celltreat, cat. no. 229385)
- Tissue culture 6-well plates (Genesee Scientific, USA; cat. no. 25-105)
- Tissue culture 150 mm plates (Fisher Scientific, USA; cat. no. FB012925)
- 5 ml serological (Genesee Scientific, USA; cat. no. 12-102)
- 10 ml serological (Genesee Scientific, USA; cat. no. 12-104)
- 25 ml serological (Genesee Scientific, USA; cat. no. 12-106)
- 50 ml serological (Genesee Scientific, USA; cat. no. 12-107)
- Steriflip-GP Sterile centrifuge tube top filter unit (Millipore Sigma, cat. no. SCGP00525)
- 500 ml sterile filter units (Genesee Scientific, USA; cat. no. 25-227)
- Corning cryogenic vials, external thread (2.0 mL; Millipore Sigma, cat. no. CLS430661)

- 5 ml microcentrifuge tube, sterile (Eppendorf, Germany; cat. no. 0030119487)
- 15 ml disposable centrifuge tube, sterile (Fisher Scientific, USA; cat. no. 05-539-12)
- 50 ml disposable centrifuge tube, sterile (Fisher Scientific, USA; cat. no. 05-539-8)
- FlowTubes with strainer cap (VWR, USA; cat. no. 76449-658)
- Tissue culture CO_2 incubator (Thermo Fisher Scientific, USA; model no. Heracell 240i)
- Tissue culture S41i CO_2 Incubator Shaker (Eppendorf, Germany; cat. no. S41I120010)
- Class II, type A2 biosafety cabinet (Baker, USA; model: SterilGard III SG403)
- Cell sorter (Sony Biotechnology, USA; model no. SH800S)
- Automated cell counter (Thermo Fisher Scientific, USA; model no. Countess 3)
- Cell counting slides (Bulldog-bio, USA; cat. no. DHC-N01)
- Tissue culture Floid cell imaging station (Thermo Fisher Scientific, USA; cat. no. 4471136)
- Benchtop centrifuge model no. 5810 with swing bucket rotor S-4-104 (Eppendorf, Germany; cat. no. 022627110) including 4x 750 ml swing buckets + 15 ml conical tube adapters, as well as matching aerosol-tight caps (cat. no. 022638661) for lentiviral harvest.
- Refrigerated floor centrifuge (Thermo Fisher Scientific, USA; model no. Sorvall RC6+) with Fiberlite F9-4x1000y rotor
- Corning CoolCell cell freezing vial container (Fisher Scientific, USA; cat. no. 07-210-002)
- Fast performance liquid chromatography system (Cytiva, USA; model no. Äkta Pure 25 M)
- Size exclusion chromatography column (Cytiva, USA; model no. Superose 6. Increase 3.2/300)
- Hamilton gas-tight syringe 50 µl (Millipore-Sigma, USA; cat. no. 26280-U)
- Ultra centrifugal filters for protein concentration (Millipore-Sigma, USA; model no. Amicon Ultra 0.5 or 4 with protein-specific molecular weight cut-off)
- Disposable chromatography column, 20 ml (Bio-Rad, USA; cat. no. 7321010)
- PD-10 desalting columns (Cytiva, USA; cat. no. 17085101)

- Wheaton Dounce tissue grinder, 15 mL, tight pestle (DWK Life Sciences, USA; cat. no. 357544)
- Branson Ultrasonics sonifier SFX250 cell disruptor (Fisher Scientific, USA; cat. no. 15-345-138) with disruptor horn (cat. no. 22-020860)
- Invitrogen DynaMag-2 Magnet (Thermo Fisher Scientific, USA; cat. no. 12321D)
- Invitrogen DynaMag-15 Magnet (Thermo Fisher Scientific, USA; cat. no. 12301D) (cost-efficient alternatives are available here: https://sergilabsupplies.com)

Reagent setup

LB Broth (LB): To make 1 L of LB weigh in 25 g of LB Broth and dissolve in 1 L ddH_2O. Sterilize by autoclaving for 45 min. at 121˚C using liquid cycle.

Super Broth (SB): To prepare 5 L of SB weigh in 175 g tryptone, 100 g yeast extract, 25 g NaCl and dissolve in 4.5 L ddH_2O. Adjust pH to 7-7.5 with 1M NaOH and top up to 5 L. Aliquot 1 L into 2 L Erlenmeyer flasks and sterilize by autoclaving for 45 min. at 121˚C using liquid cycle.

LB agar plates: Dissolve 25 g LB Agar in 1 L ddH_2O. Sterilize by autoclaving for 45 min. at 121˚C using liquid cycle. When cooled to 45-50˚C add 1 ml of a 1000x stock of antibiotic, mix, and pour into 100x15 mm petri dishes. Allow to solidify and dry at room temperature.

50 mg/ml Kanamycin (Kan) (1000x stock): To prepare 10 ml of a 50 mg/mL stock solution of kanamycin, dissolve 0.5 g of kanamycin sulfate in 9.5 mL of H_2O. Top up the volume to 10 mL and filter-sterilize using a 0.22-µm filter. Store in 1 mL aliquots at −20°C.

100 mg/ml Carbenicillin (Carb) (1000x stock): To prepare 10 ml of a 100 mg/mL stock solution of carbenicillin, dissolve 1 g of carbenicillin (disodium salt) in 9.5 mL of H_2O. Top up the volume to 10 mL and filter-sterilize using a 0.22-µm filter. Store in 1 mL aliquots at −20°C.

50 mg/ml Chloramphenicol (Cam): To prepare 10 ml of a 50 mg/mL stock solution of chloramphenicol, dissolve 0.5 g of chloramphenicol in 9.5 mL 100 % ethanol. Top up the volume to 10 mL and filter-sterilize using a 0.22-µm filter. Store in 1 mL aliquots at −20°C.

1 M IPTG: To make up 10 ml of 1 M IPTG, weigh in 2.38 g of IPTG and dissolve in 8 ml ddH_2O. Top up the volume to 10 mL and filter-sterilize using a 0.22-µm filter. Store in 1 mL aliquots at −20°C.

2 M Tris/HCL pH 7.5: To make 1 L of 2 M Tris/HCL at pH 7.5, weigh in 242.2 g Tris base and dissolve in 800 ml ddH2O. Add 134.3 ml concentrated 37%/12N HCL. Top up the volume to 1 L with ddH_2O and filter-sterilize. Store at RT.

1 M HEPES/KOH pH 7.5: To make 0.5 L of 1 M HEPES/KOH at pH 7.5, weigh in 119.2 g HEPES and dissolve in 350 ml ddH_2O. Adjust pH to 7.5 with potassium hydroxide pellets. Top up the volume to 0.5 L with ddH_2O and filter-sterilize. Store at 4˚C.

5 M NaCl: To make 1 L of 5 M NaCl, dissolve 292 g of NaCl in 700 ml ddH_2O and add ddH_2O up to 1 L. Filter-sterilize and store at RT.

5 M KAcetate (KAc): To make 200 ml of 5 M KAc, dissolve 98.14 g of KAc in 180 ml ddH_2O and add 800 µl concentrated 37%/12 N HCl. Add ddH_2O up to 200 ml and filter-sterilize. Store at RT.

1 M $MgCl_2$: To make 100 ml of 1 M $MgCl_2$**,** dissolve 20.3 g of $MgCl_2$ in 80 ml ddH_2O and add ddH_2O up to 100 ml. Filter-sterilize and store at RT.

2 M Imidazole pH 7.5: To make 250 ml of 2 M imidazole pH 7.5, dissolve 34.04 g imidazole in 200 ml ddH_2O. Adjust pH to 7.5 with concentrated 37%/12N HCl and add ddH_2O up to 250 ml. Filter-sterilize and store at 4 °C, protected from light.

1 M DTT: To make 10 ml of 1 M DTT, dissolve 1.5 g of DTT in 8 ml ddH_2O and add ddH_2O up to 10 ml. Make 1 ml aliquots and store at -20˚C.

100 mM PMSF: To make 50 ml of 100 mM PMSF, weigh in 0.87 g PMSF and dissolve in 50 ml methanol. Store at -20˚C. **! CAUTION** Both PMSF and methanol are toxic.

100 mM ATP: To make 10 ml of 100 mM ATP, weigh in 551 mg of Adenosine 5′-triphosphate disodium salt hydrate and dissolve in 8 mL ddH_2O. Adjust to pH 7.0 with potassium hydroxide pellets. Top up the volume to 10 mL with ddH_2O and make 0.5 ml aliquots. Store at -20˚C.

10 mg/ml RNase A: To make 5 ml of a 10 mg/ml stock of RNaseA, reconstitute 50 mg lyophilized RNase A in 5 ml 1x PBS and freeze in 1 ml aliquots, store at -20˚C

50 mM biotin stock: To make 10 ml of a 50 mM biotin stock in 50 mM HEPES/KOH pH 7.5, weigh in 122 mg biotin and dissolve in 10 ml 50 mM HEPES/KOH pH 7.5. Make 1 ml aliquots and store at -20˚C.

50 mM dPEG$_{24}$-biotin acid stock: To make 1 ml of a 50 mM dPEG24-biotin acid stock in 50 mM HEPES/KOH pH 7.5, weigh in 68.6 mg dPEG24-biotin acid and dissolve in 1 ml 50 mM HEPES/KOH pH 7.5. Make 100 µl aliquots and store at -20˚C.

8 M Urea: To prepare 50 ml of 8 M Urea, weigh in 24 g Urea and dissolve in 20 ml ddH_2O. Add ddH_2O up to 50 ml, filter-sterilize and store at RT.

1 mg/ml doxycycline: To prepare a 50 ml of 1 mg/ml doxycycline, weigh in 50 mg doxycycline and dissolve in 50 ml ddH_2O. Filter-sterilize using a 0.22-µm filter in a tissue culture hood. Make 1 ml aliquots and store at -20˚C. For the induction of larger culture volumes a separate 10 mg/ml stock is useful.

25x Protease inhibitor cocktail: Dissolve 1x Roche cOmplete, Mini, EDTA-free protease inhibitor cocktail tablet in 1 ml ddH_2O. Use immediately or store at -20˚C.

10 % (w/v) GDN: To prepare 10 ml 10 % (w/v) GDN, weigh in 1 g GDN and add 8 ml ddH_2O. Rotate head-over-tail at RT until GDN is fully dissolved. Top up to 10 ml with ddH_2O and make 1 ml aliquots. Store at -20˚C.

20 % (v/v) Triton-X-100: To prepare 100 ml of 20 % (v/v) Triton-X-100, transfer 20 ml 100 % (v/v) Triton-X-100 into a 100 ml glass bottle and add 80 ml ddH_2O. Incubate tumbling or rotating until the solution is well mixed. Store at RT.

5x SDS-PAGE sample buffer: 250 mM Tris/HCl pH 6.8, 5 % (w/v) SDS, 50 % (v/v) glycerol, 500 mM DTT. Prepare 40 ml by mixing the following: 1.211 g Tris base, 10 ml 20% (w/v) SDS, 20 ml 100 % (v/v) glycerol and 3.084 g DTT. Adjust pH to 6.8 by adding 0.79 ml 37%/12N HCl and then add ddH_2O to 40 ml. Mix by inverting gently until all DTT has dissolved and then add 40 mg Bromophenol Blue. Mix as above and make 1 ml aliquots. Store at -20˚C.

Resuspension buffer: 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF. Prepare 1 L by mixing the following: 25 ml 2 M Tris/HCl pH 7.5, 60 ml 5 M NaCl, 10 ml 2 M imidazole and 905 ml ddH_2O. Store at 4˚C. Add 1 mM DTT and 1 mM PMSF directly before use.

Low-salt-biotin-ATP-RNase A (LSBAR) buffer: 50 mM HEPES/KOH pH 7.5, 100 mM KAc, 2 mM MgAc, 1 mM DTT, 1 mM ATP, 50 µM biotin and 100 µg/ml RNAse A. Prepare 50 ml by mixing the following: 2.5 ml HEPES/KOH pH 7.5, 1 ml 5 M KAc, 0.5 ml 100 mM ATP, 50 µl 50 mM biotin, 0.5 ml 10 mg/ml RNase A, 50 µl 1 M DTT and 45.4 ml ddH_2O. Prepare fresh shortly before use.

Low-salt-ATP (LSA) buffer: 50 mM HEPES/KOH pH 7.5, 100 mM KAc, 2 mM MgAc, 1 mM DTT, 1 mM ATP. Prepare 50 ml by mixing the following: 2.5 ml HEPES/KOH pH 7.5, 1 ml 5 M KAc, 0.5 ml 100 mM ATP, 50 µl 1 M DTT and 45.95 ml ddH_2O. Prepare fresh shortly before use.

High-salt buffer: 50 mM Tris/HCl pH 7.5, 1 M NaCl, 20 mM imidazole, 1 mM DTT. Prepare 50 ml by mixing the following: 1.25 ml 2 M Tris/HCl pH 7.5, 10 ml 5 M NaCl, 0.5 ml 2 M imidazole and 38.2 ml ddH_2O. Store at 4˚C. Add 50 µl 1 M DTT directly before use.

Imidazole elution buffer: 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM DTT. Prepare 50 ml by mixing the following: 1.25 ml Tris/HCl pH 7.5, 3 ml 5 M NaCl, 12.5 ml 2 M imidazole pH 7.5, 5 ml 100 % glycerol and 28.2 ml ddH_2O. Store at 4˚C. Add 50 µl 1 M DTT directly before use.

BirA storage buffer: 50 mM Tris/HCl pH 7.5, 200 mM NaCl, 1 mM DTT, 250 mM sucrose. Prepare 50 ml by mixing the following: 1.25 ml Tris/HCl pH 7.5, 2 ml 5 M NaCl, 6.58 ml 1.89 M/65% (w/v) sucrose and 40.12 ml ddH_2O. Store at 4˚C. Add 50 µl 1 M DTT directly before use.

ALFA Nb storage buffer: 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 250 mM sucrose. Prepare 50 ml by mixing the following: 1.25 ml Tris/HCl pH 7.5, 3 ml 5 M NaCl, 6.58 ml 1.89 M/65% (w/v) sucrose and 39.17 ml ddH_2O. Store at 4˚C.

5x biotinylation buffer: 250 mM HEPES/KOH pH 7.5, 500 mM NaCl, 50 mM ATP, 62,5 mM $MgCl_2$ and 50 mM biotin. Prepare 10 ml by mixing the following: 1.25 ml 2 M Tris/HCl pH 7.5, 1 ml 5 M NaCl, 5 ml 100 mM ATP, 0.625 ml 1 M $MgCl_2$ and 2.125 ml ddH_2O. Weigh in 122 mg biotin and dissolve in 10 ml buffer as prepared above. Make 1 ml aliquots and store at -20˚C.

SA test binding (STB) buffer: 50 mM Tris/HCl pH 7.5, 200 mM NaCl, 0.1 % (v/v) Triton-X-100, 1 mM DTT. Prepare 20 ml by mixing the following: 0.5 ml Tris/HCl pH 7.5, 0.8 ml 5 M NaCl, 0.1 ml 20 % (v/v) Triton-X-100, 20 µl 1 M DTT and 18.58 ml ddH_2O. Prepare freshly before use.

Solubilization buffer: 50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM MgAc, 1 mM ATP, 1 % (w/v) GDN, 1x Roche cOmplete protease-inhibitor cocktail, 1 mM DTT. Prepare 10 ml by mixing the following: 0.5 ml 1 M HEPES/KOH pH 7.5, 0.4 ml 5 M NaCl, 20 µl 1

M MgAc, 0.1 ml 100 mM ATP, 1 ml 10% (w/v) GDN, 0.4 ml 25x Roche cOmplete protease-inhibitor cocktail, 10 µl 1 M DTT and 7.57 ml ddH_2O. Prepare freshly before use.

Wash buffer: 50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM MgAc, 0.5 mM ATP, 0.01 % (w/v) GDN, 1x Roche cOmplete protease-inhibitor cocktail, 1 mM DTT. Prepare 10 ml by mixing the following: 0.5 ml 1 M HEPES/KOH pH 7.5, 0.4 ml 5 M NaCl, 20 µl 1 M MgAc, 0.05 ml 100 mM ATP, 10 µl 10% (w/v) GDN, 0.4 ml 25x Roche cOmplete protease-inhibitor cocktail, 10 µl 1 M DTT and 8.61 ml ddH_2O. Prepare freshly before use.

Wash buffer (-ATP): 50 mM HEPES/KOH pH 7.5, 200 mM NaCl, 2 mM MgAc, 0.01 % (w/v) GDN, 1x Roche cOmplete protease-inhibitor cocktail, 1 mM DTT. Prepare 1 ml by mixing the following: 0.05 ml 1 M HEPES/KOH pH 7.5, 0.04 ml 5 M NaCl, 2 µl 1 M MgAc, 1 µl 10% (w/v) GDN, 0.04 ml 25x Roche cOmplete protease-inhibitor cocktail, 1 µl 1 M DTT and 0.866 ml ddH_2O. Prepare freshly before use.

DMEM/10%FBS/1xGln medium (DMEM): To 500 mL Gibco DMEM, high glucose, no glutamine add 50 mL FBS and 5.5 mL 100x Gln. (Optional) Add 5.5 mL 100x Pen-Strep.

3.5 Procedure

Generation of biotinylated nanobodies and $SENP^{EuB}$

Preparation of biotinylated GFP nanobodies ● Timing ~4d

Steps 1-12 describe the expression of biotinylated anti-GFP nanobodies from plasmids pTP396 and pTS117 in *E. coli* CVB101. (Optional) Take samples for SDS-PAGE to help troubleshoot protein expression and purification as outlined in **Supplementary Data 2**.

CRITICAL STEP The ALFA nanobody does not express well in *E. coli* CVB101 and thus needs to be expressed in an alternative strain and biotinylated in vitro (see below).

1 **Transformation via heat-shock**: Thaw one vial of *E. coli* CVB101 cells on ice for 10 min. and then add 1 µl of ~100 ng/µl plasmid. Mix and incubate for 30 min on ice.

Heat-shock vial at 42˚C for 30 sec and then quickly remove and incubate for 1 min on ice. Resuspend heat-shocked cells in 300 µl SOC recovery medium and shake at 37˚C for 1 h. Plate 50 µl of transformed cells onto an LB agar plate containing 50 µg/ml Kan and 10 µg/ml Cam using the dilution streak method. Incubate plate at 37˚C overnight.

CRITICAL STEP Chloramphenicol is required to maintain the BirA expression plasmid.

2 **Starting a pre-culture**: Pick a single colony from the plate and transfer into a 1.5 ml Eppendorf tube containing 200 µl SB supplemented with 50 µg/ml Kan and 10 µg/ml Cam (SB-Kan-Cam). Incubate shaking at 37˚C for 4-5 h and then transfer to 100 ml SB-Kan-Cam in a 2-5 L baffled flask. Incubate this pre-culture at 37˚C overnight with shaking at ˜220 rpm.

3 **Inducing the main culture**: The next morning, measure the optical density of a 1:10 dilution of the pre-culture at 600 nm (OD_{600}) using a spectrophotometer. Dilute the pre-culture with SB-Kan-Cam to an OD_{600} of ˜1 to make up a 1 L main culture. Weigh in ˜12.2 mg biotin from powder, resuspend in 1 ml SB-Kan and add to main culture (˜50 µM final). Incubate main culture at 18˚C with shaking for around 1 h. Induce main culture by addition of 0.2 mM IPTG and incubate shaking overnight (˜20 h) at 18˚C.

CRITICAL STEP Biotin supplementation is crucial for efficient biotinylation of the expressed nanobodies.

4 **Harvest main culture**: Harvest cells by centrifugation at 9,220 g for 10 min. at 4˚C. Discard supernatant and resuspend cell pellet in 120 ml resuspension buffer. Aliquot 4x 30 ml into 50 ml tubes and freeze in liquid nitrogen.

! CAUTION Wear safety goggles when handling liquid nitrogen.

PAUSE POINT Frozen *E. coli* cell pellets can be stored at -80˚C for many months.

5 **Cell lysis**: Rapidly thaw cell pellets in lukewarm water and place on ice as soon as the last frozen clumps are thawed. Place cells into a ~250 ml thin-walled metal beaker and transfer into an ice-water bath. Lyse cells by sonication on a Branson sonifier using the flat tip and 4x 1 min. sonication cycles at 100% amplitude. Each cycle should have roughly 1 sec pulses followed by 2 sec breaks. In between the four cycles mix lysate and allow to cool down for 30 sec - 1 min.

CRITICAL STEP Sonication creates heat that can activate *E. coli* proteases and denature proteins, causing degradation, aggregation or excessive chaperone binding. Keep thawed cells in an ice-water bath at all times. Ideally perform sonication in a 4˚C cold room. Keep sonication probe properly submerged in the cell lysate to prevent foam formation (aggregated protein).

6 **Centrifugation:** Centrifuge the lysate for 30 min. at 35,000 g and 4˚C. Take off and pool supernatant.

PAUSE POINT Lysate can be supplemented with either 250 mM sucrose or 10% glycerol and frozen in liquid nitrogen for storage at -80˚C for many months depending on protein stability.

7 **Binding to Ni^{2+}-resin:** Equilibrate around 2 ml settled Ni-NTA agarose resin with 20 ml resuspension buffer in a disposable 20 ml gravity flow column. Drip speed can be enhanced by attaching a long and wide, ideally blunt-ended needle to the column. Transfer equilibrated beads to a 100 ml glass bottle containing ~60 ml lysate and incubate for 1 h at 4˚C with constant mixing. We suggest either purifying 2x 60 ml lysate in two separate columns or just freezing 60 ml lysate for future purification.

8 **Washing the Ni^{2+}-resin:** Transfer the suspension back to the column and discard the flow-through. Wash the resin with 20 ml of the following buffers in this order: 1) resuspension buffer, 2) LSBAR buffer, 3) high-salt buffer and finally 4) resuspension buffer.

9 **Elution:** Detach needle if used, apply 1 ml imidazole elution buffer and discard flow-through. Elute stepwise by adding 0.5 ml elution buffer to the resin. Collect a total of 6x 0.5 ml fractions in separate 1.5 ml tubes.

10 **Determine concentration:** Measure UV absorbance of each fraction at 280 nm and pool fractions with the highest protein content. Estimate the concentration of the pooled purified nanobody using its specific extinction coefficient at 280 nm (ε_{280}) (see **Table 3.2**). Make sure to use elution buffer as a blank control. A typical yield for pTP396 is 60 mg per 1 L of bacterial culture. pTP396 can be stored in highly concentrated form, e.g. ~600 µM stocks.

11 **SDS-PAGE**: Analyze expression and purification using SDS-PAGE and Coomassie staining. Compare purified protein to **Fig. 3.5a**.

? TROUBLESHOOTING

12 **Freeze**: Aliquot purified nanobodies into multiple 10 and 50 µl aliquots in thin-walled 200 µl PCR tubes and flash freeze in liquid nitrogen. Store in small boxes or 50 ml tubes at -80˚C.

PAUSE POINT Frozen protein aliquots are stable at -80˚C for years.

CRITICAL STEP Although aliquots typically tolerate multiple freeze-thaw cycles, repeatedfreeze-thawing or prolonged incubation at elevated temperatures will result in nanobodyaggregation. Mark freeze-thaw cycles on the tubes and keep aliquots on ice immediately after thawing.

Preparation of $SENP^{EuB}$ protease and biotin ligase BirA ● Timing ~4 d

$SENP^{EuB}$ and BirA are expressed from pAV0286 and pTP264, respectively, by following steps 1-12 with the following modifications:

- Step 1: Use *E. coli* strain NEBExpress instead, as no biotinylation is required. Leave out Cam from the LB-Kan agar plate, pre-culture and main culture.
- Step 3: Induce protein expression for 6 hours at 18˚C. There is no need to supplement the medium with biotin.
- Step 7: Use 2 ml Ni-NTA agarose beads and incubate with 120 ml lysate.
- Step 8: Replace LSABR buffer with LSA buffer (without biotin and RNase A).
- Steps 11-12: For $SENP^{EuB}$ follow steps 11-12, since it can be stored in elution buffer. For BirA it is necessary to exchange the elution buffer to BirA storage buffer using a PD-10 desalting column. This is necessary because BirA activity is sensitive to ionic strength and glycerol concentration. Closely follow the manufacturer's instructions. Pool peak fractions based on absorbance at 280 nm, measure the final concentration of the pooled fractions and analyze an aliquot by SDS-PAGE. Flash freeze aliquots of the stock in liquid nitrogen as described above.

Preparation of ALFA nanobodies ● Timing ~4 d

ALFA nanobodies are expressed from pTP298 and pTS118 by following steps 1-12 with the following modifications:

- Step 1: Use *E. coli* strain Rosetta-gami 2 instead and supplement LB-Kan agar plates, pre- and main cultures additionally with 34 µg/ml Cam to maintain the pRARE2 plasmid.

CRITICAL STEP This strain does not support biotinylation during expression. The resulting anti-ALFA tag nanobodies therefore need to be biotinylated with purified biotin ligase BirA *in vitro* after Ni^{2+}-chelate affinity purification and buffer exchange as described below.

- Step 3: Induce protein expression for 6 hours at 18˚C. There is no need to supplement the medium with biotin.
- Step 7: Use 2 ml Ni-NTA agarose beads and incubate with 120 ml lysate.

- Steps 11-12: Exchange the imidazole elution buffer after Ni^{2+}-purification to ALFA Nb storage buffer using a PD-10 desalting column. Closely follow the manufacturer's instructions, pool peak fractions based on absorbance at 280 nm, measure the final concentration and 260/280 nm absorbance ratio of the pooled fractions.

  PAUSE POINT Either proceed straight to *in vitro* biotinylation or flash freeze 1 ml aliquots in liquid nitrogen.

In vitro biotinylation of ALFA nanobodies with purified BirA ● Timing ~5 h

Ni^{2+}-purified and buffer exchanged ALFA nanobodies expressed from pTP298 and pTS118 are biotinylated with BirA in solution. Excess biotin then needs to be removed by buffer exchange.

13 **Set up reaction:** Mix 100 nmoles of pTP298 or pTS118 with 1.5 nmoles of BirA enzyme, 300 µl 5x biotinylation buffer and add ddH_2O to a final reaction volume of 1.5 ml.

CRITICAL STEP *In vitro* biotinylation is most efficient when the substrate is provided at a concentration of at least 40 µM. Reaction volumes can be scaled up or down.

14 **Incubation:** Incubate *in vitro* biotinylation reaction for 3 h at 25˚C in a thermomixer with only occasional mixing every 30 minutes.

CRITICAL STEP BirA activity is highly temperature-dependent. Biotinylation reactions on ice typically need at least 24 h to reach completion.

15 **Removal of excess biotin:** Load the entire biotinylation reaction on a PD-10 desalting column, pre-equilibrated in ALFA nanobody storage buffer according to the manufacturer's instructions. Let all 1.5 ml completely sink into the column and discard the flow-through. Then add 1 ml of ALFA nanobody storage buffer and discard the

flow-through. Now elute stepwise by adding 0.5 ml ALFA nanobody storage buffer to the PD-10 desalting column and collect 6x 0.5 ml fractions in separate 1.5 ml tubes.

16 **Pool biotinylated nanobody fractions:** Measure the absorbance at 280 nm, as well as the 260/280 nm absorbance ratio of the individual fractions. Pool only fractions with a 260/280 ratio close to the ratio of the purified nanobody before biotinylation. In most cases this ratio should be between 0.5-0.7. Measure the final concentration and freeze aliquots in liquid nitrogen.

CRITICAL STEP Including later fractions with increased 260/280 ratios (>0.7) risks contaminating the final biotinylated ALFA nanobody preparations with excess free biotin that will compete with nanobody binding to streptavidin beads.

PAUSE POINT Frozen protein aliquots are stable at -80˚C for years.

(Optional) **Assessing the biotinylation efficiency of the purified nanobodies** ● Timing ~**4 h**

The biotinylation efficiency of the purified nanobodies can be assessed by testing their binding to streptavidin beads. Use pre-chilled buffer and work on ice.

17 **Equilibrate beads:** Carefully resuspend magnetic Streptavidin beads until the slurry is homogenous and no clumps are left on the bottom or side of the bottle. Aliquot 25 µl slurry into a 1.5 ml tube and mix with 1 ml STB buffer. Retrieve beads by placing the tube into a magnetic rack, wait for 30 - 60 seconds to collect beads on the magnet and then aspirate all buffer.

18 **Prepare nanobodies:** Dilute 12 µg biotinylated nanobody in 60 µl STB buffer. Take an input sample by mixing 10 µl of the dilution with 2.5 µl 5x SDS-PAGE sample buffer.

CRITICAL STEP In order to assess biotinylation efficiency, it is essential to avoid oversaturating the beads. We found that the capacity of Pierce magnetic Streptavidin beads is around 0.5 µg of biotinylated nanobody fusion protein per 1 µl slurry of beads.

19 **Binding:** Remove tube from the magnetic rack and resuspend the beads with the remaining 50 µl diluted nanobody. Incubate for 30 min with occasional gentle flicking of the tube to prevent beads from settling.

20 Retrieve beads by placing the tube into a magnetic rack. Take off unbound fraction and take a sample by mixing 10 µl unbound fraction with 2.5 µl 5x SDS-PAGE sample buffer.

21 **Washing:** Wash beads by resuspending them in 1 ml STB buffer. Retrieve beads by placing the tube into a magnetic rack and then aspirate buffer. Repeat this step. For the last wash step, resuspend beads in 100 µl buffer to concentrate them in a smaller area. Retrieve beads and aspirate all buffer.

22 **Elution:** Resuspend beads in 60 µl 2x SDS-PAGE sample buffer containing 0.5 M urea and boil for 10 min. at 97˚C. Retrieve beads and take off the elution to a new tube.

23 **SDS-PAGE analysis:** Analyze the test binding by SDS-PAGE and Coomassie staining. Load 12.5 µl of input and unbound samples, as well as 10 µl of the elution per lane. The expected outcome is shown in **Fig. 3.5b**.

? TROUBLESHOOTING

(Optional) **Assessing the activity of purified SENP^{EuB}** ● Timing ~**3 h**

24 Make up 10 µM (0.32 µg/µl) purified pTP396_His-Avi-SUMO^{Eu}-anti-GFP nanobody in a final volume of 10 µl resuspension buffer with or without 250 nM purified SENP^{EuB}.

25 Incubate both reactions for 20 min on ice

26 Add 10 µl 2x SDS-PAGE sample buffer to each sample. Analyze the test cleavage by SDS-PAGE and Coomassie staining. Load 10 µl per lane (~1.6 µg of pTP396). The expected outcome is shown in **Fig. 3.5A**.

? TROUBLESHOOTING

Generation of a stable human suspension cell line ● Timing ~8 d

Production of high-titer lentiviral supernatants ● Timing ~5 d Step 27 can be started in parallel to step 1.

27 **Take Lenti-X 293T cells into culture:** Thaw a single frozen Lenti-X 293T cell aliquot, containing around ~8-10 x10^6 cells and add to 10 ml DMEM in a 15 ml tube. Spin for 3 min. at 300 g to pellet cells and remove DMSO. Aspirate medium and resuspend cell pellet in 1 ml DMEM. Add drop-wise into a 15 cm dish containing 19 ml DMEM. Distribute cells evenly and incubate in a 37˚C, 5% CO_2 cell culture incubator.

! CAUTION Cell cultures are a potential biological hazard. Make sure to work in an approved laminar flow hood and use proper sterile technique. Adhere to the relevant institutional and governmental guidelines for recommended protective personal equipment and proper disposal of waste.

CRITICAL STEP Prevent Lenti-X 293T cells from reaching confluency. We typically split cells 1:10 every two days. Treat these cells very gently to preserve high transfection efficiency, which is essential to reach high lentiviral titer. Freeze aliquots of very early passages and avoid prolonged culture (> 1 month).

28 **Seed Lenti-X 293T cells into 6-well plates**: Once the 15 cm dish reaches ~70% confluency, aspirate all culture medium, wash gently with 20 ml DPBS and detach cells by addition of 8 ml Trypsin-EDTA (0.25%). Incubate for 2 min at room temperature and then resuspend cells with 12 ml DMEM. Count resuspended cells and seed 1x10^6 cells per 6-well in 2.5 ml DMEM. Incubate 6-well plate for ~24 h.

29 (Optional) **Freeze aliquots**: Freeze excess resuspended HEK 293T cells from step 28. Pellet cells for 3 min at 300 g, aspirate medium and resuspend in DMEM containing 10 % (v/v) DMSO. Make 1 ml aliquots containing ~8-10 x10^6 cells. Place cell aliquots into a cell freezing container and store it at -80˚C for 24 h. Transfer frozen cell aliquots into a cryo-tank for long-term storage.

30 **Transfection:** Proceed with transfection once cells reach ~70-80% confluency to produce best lentiviral titers. Per 6-well mix 1.25 µg transfer plasmid, 0.94 µg psPAX2 and 0.32 µg pMD2.G in 250 µl Opti-MEM I Reduced-Serum Medium. Mix well and add 7.5 µl TransIT 293 Transfection Reagent. Mix again and incubate for 15 min at room temperature. Carefully add formed transfection complexes to 6-wells in drops and mix plate by swirling.

CRITICAL STEP Transfer plasmids contain repetitive sequences and should be prepared from recombination-deficient (recA$^-$) *E. coli* strains like Stellar.

! CAUTION Enhanced biosafety level measures are required for all following steps.

31 **Analyze transfection efficiency:** After 24-36 h analyze transfection efficiency of each 6-well by visually assessing the percentage of fluorescent cells on a tissue-culture microscope. Non-fluorescent wells usually indicate problems with transfection and will yield low lentiviral titers.

? TROUBLESHOOTING

32 **Harvest lentiviral supernatant:** 48 h after transfection harvest culture supernatant (~2.5 ml) and spin for 5 min at 500 g in sterile 5 ml tubes. Use rotor buckets with aerosol-tight caps and assemble and disassemble them in the laminar flow hood. Use supernatant immediately to transduce Expi293F suspension cells (step 34).

PAUSE POINT Single-use aliquots of lentiviral supernatants can be snap-frozen in liquid nitrogen for storage at -80˚C for multiple months with only slight loss of viral titer.

CRITICAL Repeated freeze-thaw cycles result in significant loss of lentiviral titer. It is best to not re-use once thawed aliquots.

Transduce and grow Expi293F suspension cells ● Timing ~8 d

Step 33 can be started in parallel to step 1 and 27 to allow the Expi293F cells to recover for a few days before transduction.

33 **Take Expi293F cells into culture**: Thaw a 1 ml aliquot containing ~20x10^6 Expi293F cells and add to 10 ml Expi medium in a 15 ml tube. Spin for 3 min at 300 g to remove DMSO. Resuspend cell pellet in 1 ml Expi medium and transfer to 49 ml of the same medium in a 125 ml Erlenmeyer flask with vented cap.

CRITICAL Grow Expi293F cells in a 37˚C orbital shaker with 8% CO2, >80% relative humidity and shaking at ~125 rpm. Maintain Expi293F stock between 0.5-2x10^6 cells/ml by diluting them 1:4 with fresh medium every two days. Freeze excess cells in Expi medium supplemented with 10% (v/v) DMSO to replenish cell aliquots.

! CAUTION Enhanced biosafety level measures are required for all following steps.

34 **Transduction with lentivirus**: In a new 125 ml Erlenmeyer flask with vented cap, mix 20 ml of Expi293F cells at 1x10^6 cells/ml with 2.5 ml freshly harvested lentiviral supernatant. Then transfer flask to shaking incubator and grow overnight. If needed this step can easily be scaled up or down.

CRITICAL STEP Transduction efficiency depends strongly on lentiviral titer. This can be optimized by mixing a constant amount of suspension cells with increasing volume of lentiviral supernatant.

35 **Exchange medium:** The next day transfer culture into a 50 ml tube, place into rotor buckets with aerosol-tight caps in the laminar flow hood and then spin at 300 g for 3 min. Take off supernatant and mix with 10% bleach to inactivate lentivirus. Resuspend cell pellet in 50 ml Expi medium and transfer to a new 125 ml Erlenmeyer flask with vented cap. Grow cells for 72 h.

36 **Analyze transduction efficiency**: Analyze the ratio of fluorescent cells 48 h after transduction either visually using a tissue culture microscope or more quantitatively via flow cytometry. Protein expression can also be analyzed by Western blotting. If using inducible cells, take off 2 ml of culture to a 6-well plate and induce for 24 h with 1 µg/ml DOX before analysis of transduction efficiency as above.

? TROUBLESHOOTING

37 **Growth phase**: For small-scale expression trials, 50 ml is a good initial culture volume and cells can be easily be grown to around $8x10^6$cells/ml without any drop in viability if the expressed protein is not toxic.

CRITICAL STEP If using inducible cells make sure to add 1 µg/ml DOX to the culture medium. Induction time can be varied, but 24-48 h before harvest is a good starting point. Prior to induction, make up a separate stock of non-induced cells at $0.5x10^6$ cells/ml in 50 mL in a new flask to keep them in culture.

38 **Cell harvest:** Remove flask from shaking incubator. If using non-inducible cells, count cells and make up a separate stock at around $0.5x10^6$cells/ml in 50 ml in a new flask to keep them in culture in case sorting is necessary or the culture needs to be expanded for large-scale preparations. Transfer remaining cells into a 50 ml tube and pellet cells as described above. Wash cells by resuspending them in 50 ml DPBS and repeating the spin. Pour off liquid, weigh dry cell pellet and proceed directly to cell lysis.

PAUSE POINT Alternatively, cell pellets can be frozen in liquid nitrogen and stored at -80˚C for multiple months.

(Optional) Sorting of transduced Expi293F cells ● Timing ~7-9 d

39 **Harvest cells:** Harvest 45x10^6 transduced and non-transduced Expi293F cells as described above. Resuspend the washed cell pellets in 3 ml DPBS supplemented with 20% (v/v) FBS to a concentration of ~15x10^6 cells/ml. Filter resuspended cells through strainer caps into 5 ml round bottom FACS tubes.

40 **Adjust gates:** Analyze non-transduced Expi293F cells on a SONY SH800S or equivalent cell sorter and adjust gates to select alive and single cells, as well as to determine background fluorescence level. Set gate for transduced fluorescent cells accordingly.

41 **Sort cells:** Aim to collect at least 2-4x10^6 cells in a 15 ml collection tube pre-filled with 3 ml Expi medium containing 0.5x Pen-Strep. Pellet collected cells to remove sheath fluid and resuspend in 10 ml Expi medium containing 0.5x Pen-Strep. Transfer to a new 125 ml Erlenmeyer flask with vented cap and grow overnight.

CRITICAL STEP Since most cell sorters are not operated under perfectly sterile conditions, it is essential to add Pen-Strep to prevent contamination. However, only 0.5x Pen-Strep is used since Expi293F cells are more sensitive to antibiotics than regular HEK 293T cells.

42 **Exchange medium:** 24 h after sorting, pellet cells and aspirate medium. Resuspend in 10 ml Expi medium without Pen-Strep and transfer back to the same flask.

CRITICAL STEP Prolonged exposure to Pen-Strep reduces cell viability.

43 **Recovery**: Let cells recover by growing them for 5-7 days. Assess cell density every day, replenish medium and dilute cells to grow up a new stock of sorted cells. Freeze aliquots of early passages.

44 Cell lysis ● Timing ~2 h

A) Mechanical lysis without detergent

CRITICAL STEP Lysis buffer composition is protein-specific and should be optimized in small-scale experiments.

i) **Resuspension**: Resuspend cell pellet in pre-chilled lysis buffer. Frozen cell pellets should be thawed quickly in a lukewarm water bath in the presence of lysis buffer and removed to ice immediately once thawing nears completion.

ii) **Lysis**: Lyse cells with multiple passes in dounce homogenizer using a tight fit pestle. Alternatively, you may also use a motor-driven potter-elvehjem homogenizer. Monitor the progression of cell lysis on a cell counter or tissue-culture microscope using trypan blue exclusion. Continue lysis until no more intact cells remain.

ii) **Remove cell debris:** Centrifuge cell lysate for 30 min at 35,000 g and 4˚C. Take off supernatant and proceed to affinity purification.

B) Lysis in detergent

CRITICAL STEP Optimal detergent type, salt concentration, as well as ratio of solubilization buffer to cell pellet weight is membrane protein, as well as application-specific and should be optimized in small-scale experiments. Gentle detergents like GDN are strongly preferred to maintain protein complex stability and to preserve transient protein interactions over harsher detergents like Triton-X-100 or NP-40, which if tolerated enable purifications with higher purity. We routinely solubilize whole cell pellets, but in certain cases it is advantageous to prepare a membrane fraction prior to solubilization. The steps below offer a good starting point.

i) **Solubilization:** Add 7 ml pre-chilled solubilization buffer, containing 1% (w/v) of a detergent of choice, per each gram of cell pellet and resuspend. Frozen cell pellets should be thawed quickly in a lukewarm water bath in the presence of solubilization

buffer and removed to ice immediately once thawing nears completion. Incubate mixing head-over-tail for 30 min at 4˚C.

ii) **Remove cell debris:** Centrifuge cell lysate for 30 min at 35,000 g and 4˚C. Take off supernatant and proceed to affinity purification.

Affinity purification ● Timing ~3 h

CRITICAL STEP The amount of beads and nanobody needed depend on the expression level of the protein of interest and should be optimized in small-scale experiments. The steps below offer a good starting point. Use pre-chilled buffer and work on ice.

45 **Equilibrate beads:** For every gram of cell pellet, equilibrate 60 µl of magnetic streptavidin beads slurry. Carefully resuspend beads until the slurry is homogenous and no clumps are left on the bottom or side of the bottle. Aliquot slurry into a 1.5 ml tube. Retrieve beads by placing the tube into a magnetic rack, wait for 30 sec to 1 min to collect all beads and then aspirate all liquid. Resuspend in 1 ml wash buffer, retrieve beads as above and aspirate wash buffer.

CRITICAL STEP For membrane proteins a detergent at a concentration above its critical micelle concentration (CMC) needs to be included in wash and elution buffers to keep it solubilized. We recommend using a concentration that corresponds to 2.5x CMC.

46 **Immobilize nanobody:** For every 60 µl of magnetic Streptavidin beads slurry, immobilize 20 µg of biotinylated GFP or ALFA nanobody. Resuspend beads in 500 µl wash buffer containing pre-diluted biotinylated nanobody and incubate beads for 15 min rotating head-over-tail at 4˚C.

47 **Block beads**: Retrieve beads using a magnet and aspirate wash buffer. Remove from the magnet and resuspend in 500 µl wash buffer containing 100 µM biotin or $dPEG_{24}$-biotin acid. Incubate for 5 min on ice to block unoccupied biotin binding sites.

CRITICAL STEP Blocking with biotin strongly reduces background binding of endogenous biotinylated proteins. Blocking with $dPEG_{24}$-biotin acid adds additional negative charge and further reduces non-specific background binding.

48 **Incubate with lysate:** Retrieve beads using magnetic rack and aspirate blocking buffer. Resuspend beads with cleared cell lysate and mix rapidly. Incubate the mixture for 1 h at 4˚C rotating.

49 **Washing:** Retrieve beads on magnet. Depending on the final scale a magnetic rack that can hold 15 ml or even 50 ml tubes should be used at this step. Aspirate cell lysate and transfer all beads into a 1.5 ml tube using 1 ml wash buffer. Wash beads three times with 1 ml wash buffer. For the fourth wash, resuspend beads with 100 µl wash buffer (-ATP) and transfer to a new tube. Use slightly more volume if significantly more beads were used.

50 **Elution:** Resuspend beads in wash buffer containing 250 nM purified $SENP^{EuB}$. Use a volume corresponding to the original volume of bead slurry used. To keep the eluate more concentrated, up to 1/3rd of the original volume can be used. Incubate for 20 min on ice.

CRITICAL STEP If a soluble protein is stable in the presence of detergents (like 0.05% [v/v] Tween-20 or Triton-X-100), these can be included during elution to improve recovery by preventing non-specific binding of the cleaved protein to the beads.

51 (Optional) Spin eluate for 5 min at 15,000 g at 4˚C to pellet magnetic beads that sometimes get carried over. Take off supernatant.

52 **Post-Elution**: Resuspend beads in 2x SDS-PAGE sample buffer containing 0.5 M urea and boil for 10 min at 97˚C. Retrieve beads on magnet and take off post-elution sample.

53 **SDS-PAGE analysis:** Analyze samples of the eluate and post-elution by SDS-PAGE. We frequently load 1, 2 and 4 µl per lane and stain the gel with either Sypro Ruby or Coomassie.

? TROUBLESHOOTING

(Optional) Size-exclusion chromatography ● Timing 1 d

An additional size-exclusion chromatography run might be useful to remove excess tagged subunit and nanobody. We prefer the Superose 6 Increase 3.2/300 column due its small column volume of only 2.4 ml, which limits sample dilution during the run.

54 **Column equilibration**: Equilibrate HPLC, Superose 6 Increase 3.2/300 column and 50 µl sample loop in filtered and degassed wash buffer without protease inhibitor cocktail.

55 **Sample loading**: Load sample into a 50 µL hamilton syringe and inject into sample loop.

CRITICAL STEP Spin eluate for 10 min at 15,000 g at 4˚C before loading to pellet magnetic beads and aggregated protein. Take off supernatant. If eluate volume is above 50 µl concentrate using ultrafiltration spin columns with the appropriate molecular weight cut-off.

56 **Elution**: Inject sample loop content onto column and elute in 100 µl fractions over 1.2x column volume. Measure absorbance at 280 nm, optionally at 490 nm to detect GFP fluorescence or 260 nm to detect non-protein contaminants.

57 **SDS-PAGE analysis**: Choose fractions spanning sample peak and prepare SDS-PAGE sample by mixing 10 µl of each fraction with 2.5 µl 5x SDS-PAGE sample buffer. If this is a high yield purification you may instead prefer to dilute a smaller volume of

each fraction to 10 µL with wash buffer to avoid overloading the gel. Analyze samples by SDS-PAGE and stain with either Sypro Ruby or Coomassie.

58 Wash HPLC, Superose 6 Increase 3.2/300 column and 50 µl sample loop with degassed ddH_2O.

59 Pool protein-containing fractions. If necessary, concentrate protein using ultrafiltration spin columns with the appropriate molecular weight cut-off.

Timing

Generation of biotinylated nanobodies and $SENP^{EuB}$: ~4-5 d

Steps 1-12, Preparation of biotinylated GFP nanobodies: ~4 d

Steps 1-12, Preparation of $SENP^{EuB}$ protease and biotin ligase BirA: ~4 d

Steps 1-12, Preparation of ALFA nanobodies: ~4 d

Steps 13-16, *In vitro* biotinylation of ALFA nanobodies with purified BirA: ~5 h

Steps 17-23, (Optional) Assessing the biotinylation efficiency of the purified nanobodies: ~4 h

Steps 24-26, (Optional) Assessing the activity of purified $SENP^{EuB}$: ~3 h

Generation of a stable human suspension cell line: ~8 d

Steps 27-32, Production of high-titer lentiviral supernatants: ~5 d

Steps 33-38, Transduce and grow Expi293F suspension cells: ~8 d

Steps 39-43, (Optional) Sorting of transduced Expi293F cells: ~7-9 d

Step 44, Cell lysis: ~2 h

Steps 45-53, Affinity purification: ~3 h

Steps 54-59, (Optional) Size-exclusion chromatography: ~1 d

Troubleshooting

Troubleshooting advice is listed in Table 3.3.

Anticipated results

We have used this protocol as outlined in **Fig. 3.1** to isolate numerous soluble and membrane-bound proteins for structural, functional, and mass spectrometry analysis (**Fig. 3.3**). For proteins that are not part of a protein complex, high quality preparations can easily be achieved with often only minimal prior optimization. In certain cases, lysis buffer composition needs to be optimized in small-scale purification trials to obtain best results. In cases where little prior knowledge exists about a protein complex of interest, it is first necessary to optimize which subunit of the complex to tag and also at which terminus to tag this subunit.

Another critical consideration for the success of the protocol is to take great care to achieve high-titer lentiviral production as outlined above. Only with good lentiviral preparations can high transduction efficiencies of suspension cells be reached. In many cases, transduction efficiencies are readily between 80-90%, e.g. for EMC3-GFP (**Fig. 3.6**). In certain cases, e.g. for much larger proteins, lower transduction efficiencies necessitate either larger-scale cultures to achieve a comparable yield or an additional sorting step to obtain a fully transduced polyclonal cell line. Sorting typically postpones first purification trials by one extra week, which is required for recovery and expansion of the sorted cells.

Using this protocol, we have achieved improved yields for challenging multi-subunit membrane protein complexes. For example, we could isolate 0.45 mg of EMC or ˜0.8 mg of MTCH2 per 1 L of suspension cell culture. Much greater yields can be achieved for soluble protein complexes and in particular for soluble monomeric proteins.

3.6 Acknowledgements

We thank Pamela Bjorkman for access to her lab's cell sorter, as well as the Caltech Flow Cytometry facility. This work was supported by: the Heritage Medical Research Institute (RMV), the NIH's National Institute Of General Medical Sciences DP2GM137412 (RMV),

the Deutsche Forschungsgemeinschaft (TP), and the Tianqiao and Chrissy Chen Institute (TP, MH).

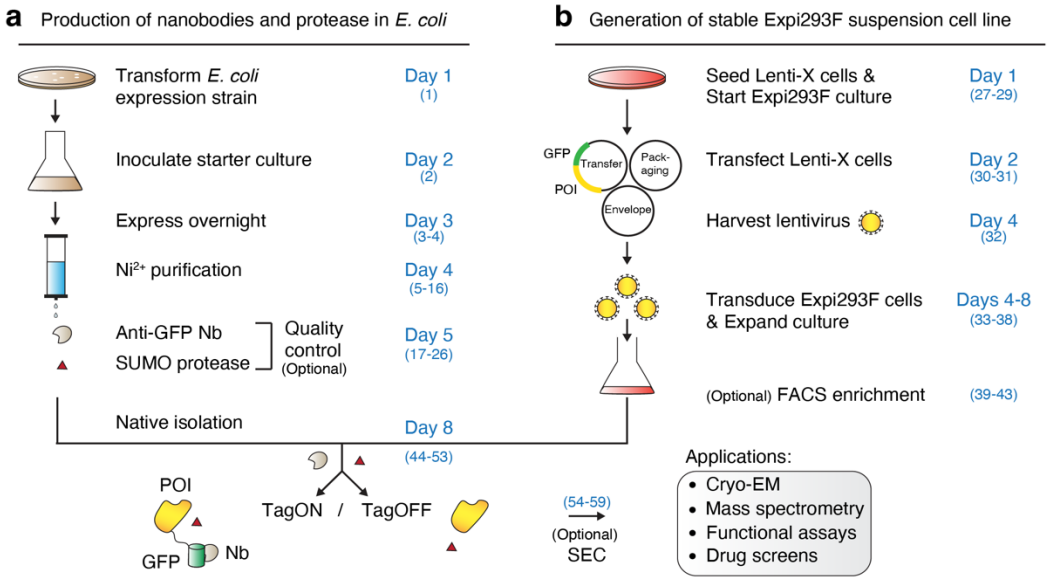


Figure 3.1. Schematic overview of the protocol. (a) Anti-GFP/ALFA nanobody (Nb)-based capture agents and SUMO protease are separately expressed in *E. coli* and purified via Ni^{2+}-chelate affinity chromatography. (b) Lentivirus encoding a GFP/ALFA-tagged protein of interest (POI) is generated and used to transduce human Expi293F suspension cells. The resulting stable cell line is expanded and either used directly for first small-scale purification trials or optionally first sorted using fluorescence-activated cell sorting (FACS). In its fastest format our protocol can be completed in only 8 days, going from DNA prep to purified protein. Individual protocol step numbers are listed in parentheses. SEC = Size-exclusion chromatography.

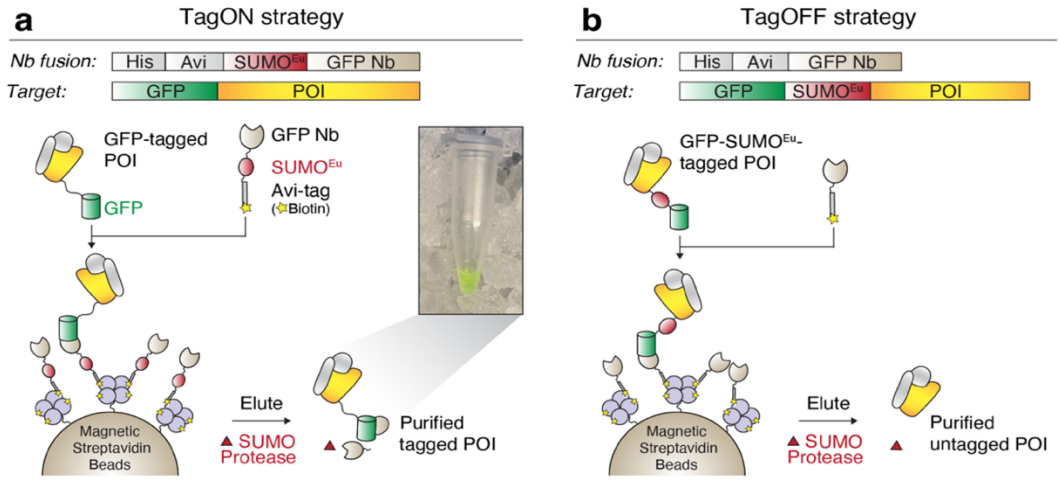


Figure 3.2. Native isolation of GFP/ALFA-fused proteins in tagged (TagON) or untagged form (TagOFF). (a) Schematic of the TagON strategy. A biotinylated SUMO^{Eu}-fused anti-GFP nanobody (Nb) is immobilized onto magnetic streptavidin beads. Nb-decorated beads are then incubated with cell lysate containing an expressed GFP-tagged protein-of-interest (POI). Following wash steps, GFP-tagged POI and bound interaction partners (grey) are rapidly eluted by cleavage with the engineered SUMO protease SENP^{EuB}. The inset depicts the result of a high-yield TagON purification (GFP-AE2, see Fig. 3.3c). (b) Schematic of the TagOFF strategy. SUMO^{Eu} is placed between GFP and POI and the POI is captured by a non-cleavable nanobody. This allows scarless tag-free elution of the POI by SENP^{EuB}.

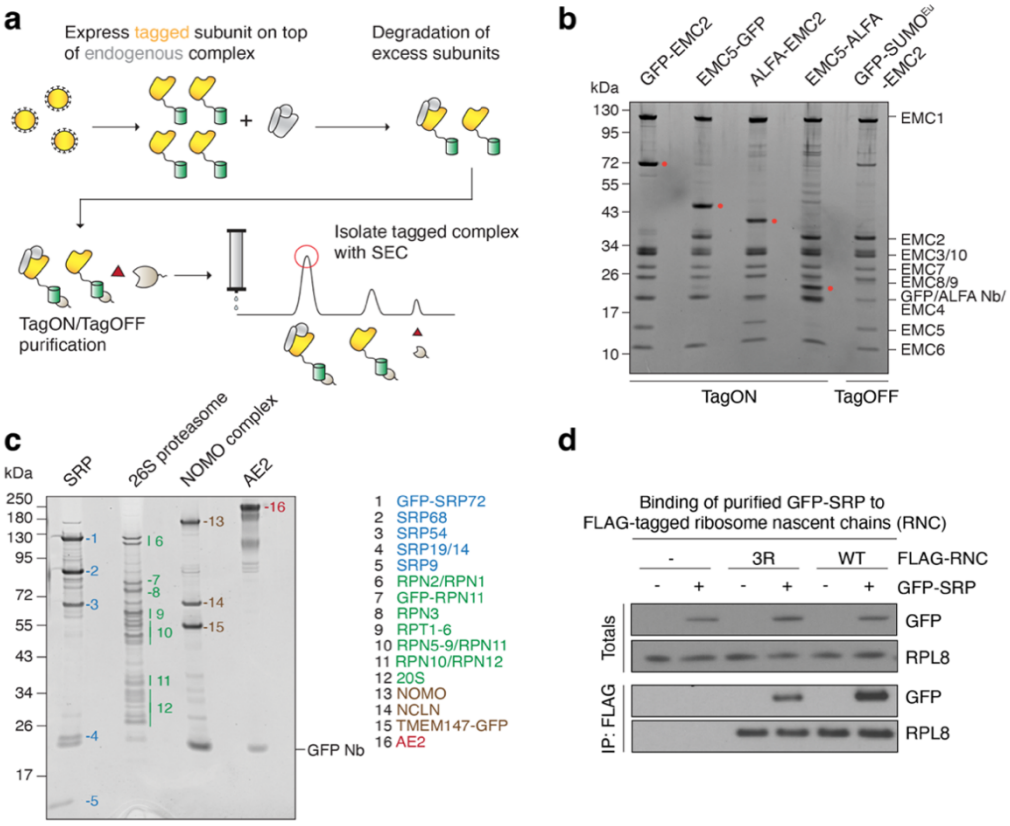


Figure 3.3. Purification of soluble and membrane protein complexes from human suspension cells. (a) Ectopically expressed tagged subunits of a protein complex replace their endogenous untagged counterparts (grey) through proteasomal degradation of excess subunits. After purification, any remaining excess subunit and nanobody can be removed via size-exclusion chromatography (SEC). (b) Peak fractions of SEC runs of the ER membrane protein complex (EMC) purified via GFP- or ALFA-tags fused to either EMC2 or EMC5 subunits using the TagON strategy. Following the TagOFF strategy, the GFP-SUMO^{Eu}-EMC2 cell line allowed purification of completely untagged EMC. Tagged subunits are marked with a red dot. (c) SEC peaks of various samples purified via GFP-tags, including the signal recognition particle (SRP), 26S proteasome, NOMO-NCLN-TMEM147 complex, as well as SLC4A2/AE2. (d) Purified GFP-SRP is functional. Stalled ribosome nascent chain complexes (RNC) exposing either a wildtype (WT) or triple arginine mutant (3R) transferrin transmembrane domain (TMD) with 3x FLAG tag were produced by *in vitro* translation in rabbit reticulocyte extract supplemented with purified

GFP-SRP complex where indicated. Total and FLAG-IP samples were analyzed by SDS-PAGE and Western blotting. GFP-SRP co-purified strongest with WT TMD RNCs as shown before for native SRP (Voorhees & Hegde, 2015).

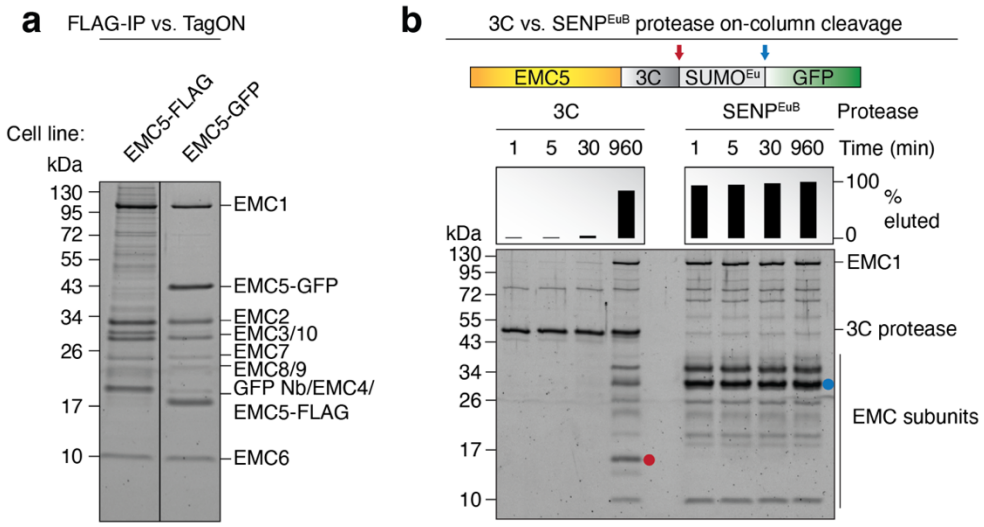


Figure 3.4. Comparison to other methods. (a) Comparison of FLAG and TagON purification of the ER membrane protein complex (EMC). Elutions from either EMC5-FLAG (using M2 FLAG affinity resin [Millipore-Sigma, USA]) or EMC5-GFP (using TagON) purifications were analyzed by SDS-PAGE and Sypro Ruby staining. (b) Comparison of 3C and SENP^{EuB} protease on-column cleavage efficiency. Purification of EMC via EMC5-3C-SUMO-GFP (using the TagOFF strategy). Cell lysate was incubated with magnetic Streptavidin beads, containing immobilized non-cleavable biotinylated anti-GFP nanobody (expressed from pTS117). After washing, beads were split in half and incubated either with 2 µM 3C or 250 nM SENP^{EuB} protease at 4˚C for the indicated time frames. Under these conditions, SENP^{EuB} cleaves nearly 1,000-fold faster, allowing for rapid elution in just 1 min as opposed to lengthy overnight incubations typically required for elution with 3C or TEV protease. The products of 3C and SENP^{EuB} cleavage are marked with red and blue dots, respectively.

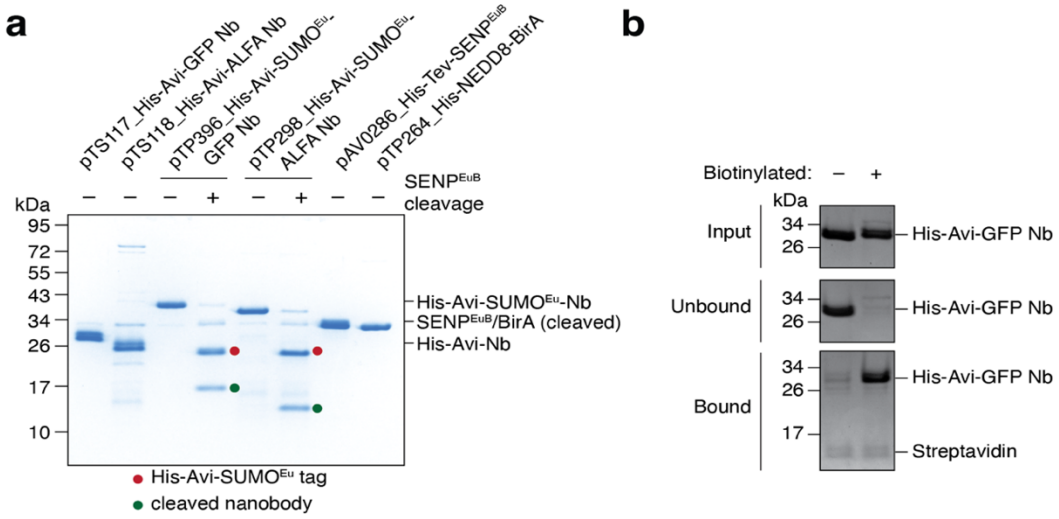


Figure 3.5. Quality control of purified proteins (a) 1 µg of each of the protein reagents required for the protocol was analyzed by SDS-PAGE and Coomassie staining. pTP396 and pTP298 were additionally cleaved with $SENP^{EuB}$ in solution to assess its activity. His-NEDD8-tagged BirA was purified by Ni^{2+}-chelate affinity purification and eluted by cleavage with NEDP1 protease (Frey & Görlich, 2014; Pleiner et al., 2018), resulting in untagged BirA. The removal of the His-NEDD8 tag is not required. Nb = nanobody (b) Coomassie-stained SDS-PAGE gel showing the quantitative streptavidin test binding of purified, biotinylated His-Avi-anti-GFP Nb (pTS117; steps 17-23). The protein was expressed and purified from *E. coli* NEBExpress (-biotinylation) or *E. coli* CVB101 (+biotinylation).

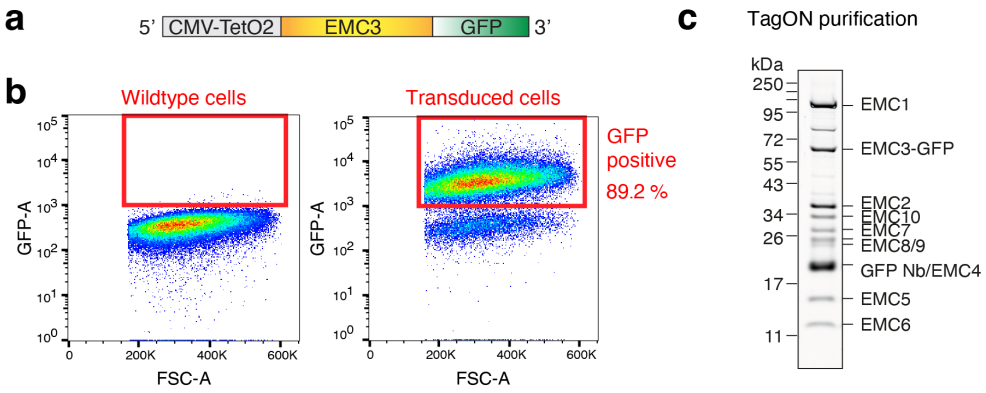


Figure 3.6. TagON purification of the EMC from a stable Expi293F EMC3-GFP suspension cell line. (a) Schematic overview of the EMC3-GFP expression cassette encoded on the lentiviral transfer plasmid. (b) Flow cytometry analysis of Expi293F cells 48 h after transduction as described in steps 33-38. Cells were gated first for alive cells (SSC-A vs. FSC-A), second for single cells (FSC-A vs. FSC-H) and third for GFP-fluorescent cells (GFP-A vs. FSC-A). Nearly 90% of cells were GFP positive. (c) TagON purification of the EMC via EMC3-GFP. An aliquot of the elution was analyzed by SDS-PAGE and Sypro Ruby staining.

Plasmid	Open-reading frame	Addgene ID
pTS093	GFP-22xGS-MCS	199346
pTS094	GFP-22xGS-3C-MCS	199347
pTS095	GFP-22xGS-TEV-MCS	199348
pTS096	GFP-22xGS-SUMO^{Eu}-MCS	199349
pTS097	GFP-22xGS-SUMOstar-MCS	199350
pTS098	TagBFP-P2A-ALFA-22xGS-MCS	199351
pTS099	TagBFP-P2A-ALFA-22xGS-3C-MCS	199352
pTS100	TagBFP-P2A-ALFA-22xGS-TEV-MCS	199353
pTS101	TagBFP-P2A-ALFA-22xGS-SUMO^{Eu}-MCS	199354
pTS102	TagBFP-P2A-ALFA-22xGS-SUMOstar-MCS	199355
pTS103	MCS-22xGS-GFP	199356
pTS104	MCS-3C-22xGS-GFP	199357
pTS105	MCS-TEV-22xGS-GFP	199358
pTS106	MCS-3C- SUMO^{Eu}-22xGS-GFP	199359
pTS107	MCS-TEV- SUMO^{Eu}-22xGS-GFP	199360
pTS108	MCS-3C-SUMOstar-22xGS-GFP	199361
pTS109	MCS- TEV-SUMOstar-22xGS-GFP	199362
pTS110	MCS-22xGS-ALFA-P2A-TagBFP	199363
pTS111	MCS-3C-22xGS-ALFA-P2A-TagBFP	199364
pTS112	MCS-TEV-22xGS-ALFA-P2A-TagBFP	199365
pTS113	MCS-3C-SUMO^{Eu}-22xGS-ALFA-P2A-TagBFP	199366
pTS114	MCS-TEV- SUMO^{Eu}-22xGS-ALFA-P2A-TagBFP	199367
pTS115	MCS-3C-SUMOstar-22xGS-ALFA-P2A-TagBFP	199368
pTS116	MCS-TEV-SUMOstar-22xGS-ALFA-P2A-TagBFP	199369
pTP341	TagBFP-3xFLAG (constitutive)	199391
pTP924	GFP (constitutive)	199392
pTP396	His$_{14}$-Avi-27xGS-SUMO^{Eu}-19xGS-anti GFP nanobody	149336
pTS117	His$_{14}$-Avi-45xGS-anti GFP nanobody	199370
pTP298	His$_{14}$-Avi-27xGS-SUMO^{Eu}-2xGS-anti ALFA tag nanobody	199390
pTS118	His$_{14}$-Avi-28xGS-anti ALFA tag nanobody	199371
pAV0286	His$_{14}$-TEV-SENP^{EuB} protease	149333
pTP264	His$_{14}$-*bd*NEDD8-BirA	149334

Table 3.1. TagON and TagOFF-compatible lentiviral transfer plasmid toolbox and *E. coli* expression plasmids. MCS = multiple cloning site; GS = Glycine/Serine-rich spacer of indicated length in amino acids; 3C = HRV 3C protease recognition site; TEV = Tobacco etch virus protease recognition site.

Construct	Protein	Mw (kDa)	ε_{280} ($M^{-1}cm^{-1}$)
pTP396	His_{14}-Avi-27xGS-$SUMO^{Eu}$-19xGS-anti GFP Nb	31.8	33920
pTS117	His_{14}-Avi-45xGS-anti GFP Nb	21.2	32430
pTP298	His_{14}-Avi-27xGS-$SUMO^{Eu}$-2xGS-anti ALFA Nb	31.2	25440
pTS118	His_{14}-Avi-45xGS-anti ALFA Nb	20.6	23950
pAV0286	His_{14}-TEV-$SENP^{EuB}$ protease	32.4	45380
pTP264	His_{14}-*bd*NEDD8-BirA	48.9	50420

Table 3.2. Properties of all purified proteins. Mw = Molecular weight; ε_{280} = extinction coefficient at 280 nm; Nb = nanobody

Step	Problem	Possible reason	Solution
11	No protein after Ni^{2+}-purification	Forgot to add IPTG to induce protein expression	Add IPTG to 0.2 mM
11	No protein after Ni^{2+}-purification	Forgot to add imidazole to elution buffer	Use elution buffer containing imidazole at 500 mM
11	Low protein yield after Ni^{2+}-purification	Protein might have aggregated during expression due to elevated expression temperature	Make sure to adapt diluted main culture to 18˚C for 1 h before induction and maintain induced main culture at 18˚C.
11	Low protein yield after Ni^{2+}-purification	Protein might have aggregated during purification	Keep *E. coli* cells/ lysate/ purified protein on ice at all times and prevent foaming during sonication
11	*E. coli* purified protein is partially degraded	Activation of *E. coli* proteases during cell lysis	Make sure to add 1 mM PMSF to resuspension buffer and keep *E. coli* cells/ lysate/ purified protein on ice at all times
23	GFP Nanobodies produced in *E. coli* CVB101 do not bind to Streptavidin beads	Forgot to supplement CVB101 main culture with biotin or forgot to use chloramphenicol to maintain BirA plasmid	Make sure to propagate CVB101 with 10 µg/ml chloramphenicol and add 50 µM biotin to main culture
23	*In vitro* biotinylated ALFA nanobodies do not bind to Streptavidin beads	Poor quality BirA preparation or wrong biotinylation buffer composition	Check BirA by SDS-PAGE and check 5x biotinylation buffer composition. Repeat BirA preparation.
26	$SENP^{EuB}$ does not cleave nanobodies in solution	Poor quality $SENP^{EuB}$ preparation	Check $SENP^{EuB}$ by SDS-PAGE. Repeat $SENP^{EuB}$ preparation.
31	Transfected Lenti-X 293T cells non-fluorescent	Poor DNA quality	Check plasmid DNA integrity and prepare new stock
36	Low transduction efficiency	Lenti-X 293T cells passaged for prolonged time and/or overgrown prior to seeding (bad cell health)	Thaw a new Lenti-X 293T cell aliquot and avoid overgrowing and passaging them > 1 month

36	Low transduction efficiency	Transfected Lenti-X 293T cells at too high or too low confluency	Make sure to transfect at around 70-80% confluency
36	Low transduction efficiency	Forgot or added wrong amounts or bad quality of packaging and envelope plasmids	Make sure to add correct amounts of packaging and envelope plasmids and consider preparing new stocks
36	Low transduction efficiency	DNA insert size comes close to or exceeds packaging limit (˜8 kbp between 5'-3' LTRs)	Sort transduced cells and grow up a fully transduced population
36	Low transduction efficiency	Protein of interest is toxic to cells	Try expressing an inactive mutant, or use an alternative expression method such as chemical transfection or recombinase-mediated cell line generation
53	Low protein yield from Expi293F cells	Poor lentiviral titer or transduction efficiency	Repeat lentiviral preparation or sort transduced cells and grow up a fully transduced population
53	Low protein yield from Expi293F cells	Inefficient protease elution	Analyze post-elution sample by SDS-PAGE and check for uncleaved bait (TagON) or bait-prey complex (TagOFF). Check protease activity in solution (see steps 24-26) and repeat protease preparation if inactive. In some cases the SUMO tag may be less accessible, requiring increased protease concentration and incubation time to achieve efficient elution.
53	Low protein yield from Expi293F cells	Eluted protein non-specifically sticks to beads (sometimes for soluble proteins in buffer without detergent)	Include low concentration of detergent (e.g. 0.05 % [v/v] Tween-20 or Triton-X-100) in elution buffer to reduce sticking. Block Streptavidin beads with biotin-PEG-COOH instead of biotin to increase charge repulsion between beads and protein. The cleaved protein may be released in additional wash steps.

53	Low protein yield from Expi293F cells	Protein aggregated due to suboptimal lysis buffer composition	Optimize lysis buffer composition e.g. salt or detergent concentration in small-scale pre-tests.
53	Low recovery of the prey's endogenous protein complex subunits	Tagged protein is stable on its own after overexpression	Consider tagging a different, less stable subunit of the complex. Alternatively: 1) Use Freestyle 293-F cells instead. 2) Use inducible Expi293F cells line and vary DOX induction time. 3) Exchange CMV to weaker PGK or UbC promoter or remove WPRE element. 4) Purify membrane proteins from the membrane fraction rather than whole cells.

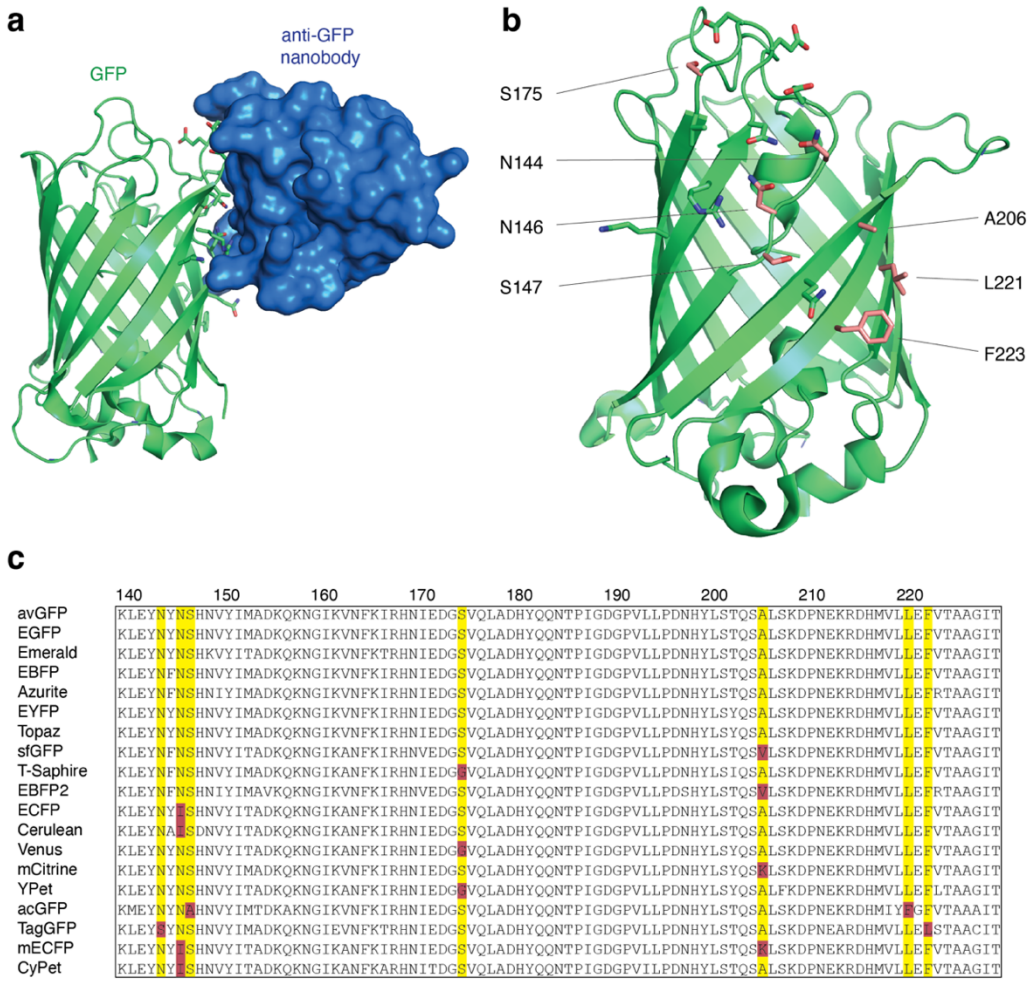
Table 3.3. Troubleshooting table

Category	Organism	Kind	Reference	Source
Fungi	*S. cerevisiae*	4,159 strains	Huh et al., 2003(Huh et al., 2003)	Thermo Fisher http://bit.ly/3EYKoHl
	S. pombe	1,058 strains	Hayashi et al., 2009(Hayashi et al., 2009)	NBRP http://bit.ly/3SJIK24
Invertebrates	*C. elegans*	14,637 constructs	Sarov et al., 2012(Sarov et al., 2012)	TransgenOme http://bit.ly/3KYDSEi
		> 100 strains		CGC http://bit.ly/3IQYpIn
	D. melanogaster	23,169 constructs	Sarov et al., 2016(Sarov et al., 2016)	TransgenOme http://bit.ly/3KYDSEi
		880 lines	Sarov et al., 2016(Sarov et al., 2016)	VDRC http://bit.ly/3KTmNvo
Mammals	*Mouse*	ES cell lines	Poser et al., 2008(Poser et al., 2008)	
		58,000 ORFs		origene http://bit.ly/3ZAEDrg
	Rat	18,000 ORFs		origene http://bit.ly/3ZAEDrg
	Humans	78,000 ORFs		origene http://bit.ly/3ZAEDrg
		HeLa cell lines	Poser et al., 2008(Poser et al., 2008)	
		1,125 HeLa cell lines	Hein et al., 2015(Hein et al., 2015)	

Extended Data Figure 3.1. Publicly or commercially available GFP-tagged plasmids, cell lines or transgenic organisms. Thousands of plasmids encoding GFP- and ALFA-tagged proteins from various organisms can also be obtained from Addgene. ORF = open-reading frame. ES = embryonic stem cell.

Tag	Binder	Reference
GFP variants	13x 'LaG' nanobodies that all bind to *A. victoria* GFP and its derivatives CFP, BFP, and YFP. 11x of those also bind to *A. macrodactyla* CFP.	Fridy et al., 2014
RFP variants	6x 'LaM' nanobodies that all bind to mCherry, 1x also binds to DsRed.	
C-tag	Nanobody that binds C-terminal 'EPEA' peptide tag, but also binds to endogenous human α-synuclein	De Genst et al., 2010
Spot/BC2-tag	Nanobody that binds 'PDRVRAVSHWSS' peptide, also called BC2 tag, but also binds to endogenous human β-catenin	Braun et al., 2016 ; Virant et al., 2018
PepTag	PepTagNB a nanobody that binds the α-helical peptide 'AVERYLKDQQLLGIW' derived from HIV glycoprotein 41 (gp41)	Strokappe et al., 2019; Traenkle et al., 2020
NanoTag VHH05	NbVHH05 nanobody that binds 'QADQAEKELARQIS' peptide tag	Xu et al., 2022
NanoTag 127D01	Nb127D01 nanobody that binds 'SFEDFWKGED' peptide tag	
Rabbit IgG	2x anti-rabbit IgG nanobodies, for purification of IgGs or IgG targets	Pleiner et al., 2018
Mouse IgG	29x anti-mouse IgG nanobodies, light chain and subclass specific binders, for purification of IgGs or IgG targets	
ED or ZZ domain dimer from Protein A	ZpA963 affibody dimer (~13.5 kDa) binds ED or ZZ domain dimer tags with 16 pM affinity	Lindborg et al., 2013; Vera Rodriguez et al., 2019
Colicin E7	Immunity protein 7 binds DNase deficient Colicin E7 with very high affinity (K_D ~10^{-14}-10^{-17}M)	Vassylyeva et al., 2017

Extended Data Figure 3.2. Selection of previously characterized affinity binder pairs that could be used for TagON/OFF purifications.



Extended Data Figure 3.3. Overview of anti-GFP nanobody compatible fluorescent protein variants. (a) Crystal structure of GFP bound to anti-GFP nanobody (Nb) (PDB ID: 3K1K) (Kirchhofer et al., 2010) with GFP shown in green with cartoon rendering, anti-GFP Nb shown in blue with surface rendering, and specific residues on GFP that make contact with the anti-GFP Nb shown in stick rendering. (b) Front-view of the anti-GFP nanobody binding surface of GFP with participating residues shown in stick rendering. Residues mutated in other fluorescent protein variants colored in salmon. (c) Multiple sequence alignment of various fluorescent protein variants. Columns corresponding to residues contacted by the anti-GFP Nb are highlighted in yellow, and any mutations to these positions are highlighted in red. Mutation of I146N was previously shown to restore anti-GFP Nb binding in CFP variants (Kubala et al., 2010).

Condition	GFP Nb	ALFA Nb
Urea	8 M	6 M
Triton X-100	1%	1%
SDS	1%	0.1%
GuHCl	4 M	2 M
NaCl	2 M	3 M
DTT	1 mM	100 mM
Deoxycholate	N/A	1%
Glycerol	30%	N/A
NP-40	2%	N/A
TCEP	0.2 mM	N/A
$MgSO_4$	N/A	1 M
DDM	N/A	1%

Extended Data Figure 3.4. Anti-GFP and anti-ALFA nanobodies withstand harsh buffer conditions. Comparison of buffer conditions tolerated by the anti-GFP nanobody Enhancer (https://bit.ly/3kLbpHr) and the anti-ALFAST nanobody (Götzke et al., 2019). Nb = nanobody; N/A = data not available.

Chapter 4

MTCH2 IS A MITOCHONDRIAL OUTER MEMBRANE PROTEIN INSERTASE

The following chapter is adapted from Stevens et al., 2023 and modified according to the Caltech thesis format.

Alina Guna*, Taylor A. Stevens*, Alison J. Inglis*, Joseph M. Replogle, Theodore K. Esantsi, Gayathri Muthukumar, Kelly C.L. Shaffer, Maxine L. Wang, Angela N. Pogson, Jeff J. Jones, Brett Lomenick, Tsui-Fen Chou, Jonathan S. Weissman, and Rebecca M. Voorhees. (2022). "MTCH2 is a mitochondrial outer membrane protein insertase". In: *Science* 6617, pp. 317-233. doi: 10.1126/science.add1856 * denotes equal contribution

4.1 Abstract

In the mitochondrial outer membrane, a-helical transmembrane proteins play critical roles in cytoplasmic-mitochondrial communication. Using genome-wide CRISPR screens, we identified MTCH2, and its paralog MTCH1, and showed that it is required for insertion of biophysically diverse tail-anchored (TA), signal-anchored, and multipass proteins, but not outer membrane b-barrel proteins. Purified MTCH2 was sufficient to mediate insertion into reconstituted proteoliposomes. Functional and mutational studies suggested that MTCH2 has evolved from a solute carrier transporter. MTCH2 uses membrane-embedded hydrophilic residues to function as a gatekeeper for the outer membrane, controlling mislocalization of TAs into the endoplasmic reticulum and modulating the sensitivity of leukemia cells to apoptosis. Our identification of MTCH2 as an insertase provided a mechanistic explanation for the diverse phenotypes and disease states associated with MTCH2 dysfunction.

4.2 Introduction

Mitochondria are organelles of endosymbiotic origin that have evolved to play a central role in eukaryotic cell metabolism and signaling (Friedman & Nunnari, 2014). Mitochondrial function and their ability to communicate with the cytosol depend on proteins embedded in the outer mitochondrial membrane. As a result, dysregulation of outer membrane protein function is associated with ageing and the pathogenesis of a variety of human diseases including Alzheimer's, Parkinson's, and many cancers (Wang et al., 2020; Bose & Beal, 2016; Vyas et al., 2016). In mammals, the insertion of a-helical proteins into the outer membrane, a function that would not have been required in the original endosymbiont, remains poorly understood (N & N, 2017). In yeast and trypanosomes the mitochondrial import protein 1 (Mim1) and pATOM36, respectively, have been implicated in this process (Doan et al., 2020; Vitali et al., 2018), but no clear homologs exist in mammalian mitochondria. One important class of α-helical outer membrane proteins are tail-anchored proteins (TAs), which are characterized by a single C-terminal transmembrane domain (TMD) and mediate diverse functions including apoptosis, innate immunity, and mitochondrial turnover and dynamics. Therefore, we set out to systematically identify and characterize the factors required for mitochondrial TA biogenesis in human cells.

4.3 Results

Using an in vitro competition assay we first showed that TA insertion does not strictly require the TOM complex, the major outer membrane translocase (Fig. 4.1A, fig. S4.1-4.2, (Setoguchi et al., 2006)). Therefore, to enable CRISPR-based screens (Gilbert et al., 2014), we adapted and validated (Fig. 4.1B, fig. S4.3) a split-GFP reporter (Le Vasseur et al., 2021) to measure insertion of the model TA, OMP25, into mitochondria. Amongst hits that increased mitochondrial integration of OMP25 were the ER membrane protein complex (EMC) and the ubiquilin (UBQLN) chaperone family (Fig. 4.1C; fig. S4.4A). These results are consistent with the EMC serving as the major insertase for mislocalized mitochondrial TAs into the ER (fig. S4.4; (Coukos et al., 2021)), and the UBQLNs' role in degrading mislocalized mitochondrial TAs (Itakura et al., 2016), leading to their accumulation in the cytosol (fig. S4.5).

Conversely, depletion of the outer membrane resident mitochondrial carrier homologue 2 (MTCH2) resulted in the most pronounced loss of OMP25 integration (Fig. 4.1D, fig. S4.6A). MTCH2 is a member of the solute carrier 25 (SLC25) family, integral membrane proteins best known for their role in transporting metabolites into the mitochondrial matrix, but its localization and sequence suggests its function has potentially diverged, and it has no known substrates or transporter activity (Ruprecht & Kunji, 2020). Further, loss of MTCH2 is associated with a variety of pleotropic phenotypes including defects in mitochondrial fusion, lipid homeostasis, and apoptosis (Labbé et al., 2021; Rottiers et al., 2017; Zaltsman et al., 2010). However, the underlying biochemical activity of MTCH2 is not known.

Because of the diverse phenotypes attributed to MTCH2, we excluded the possibility that dysregulation of lipogenesis (fig. S4.6B,C; (Labbé et al., 2021)), the outer membrane, or general mitochondrial protein biogenesis (Fig. 4.1E) could explain the observed biogenesis defect on OMP25. We next sought to determine if MTCH2 could be playing a more general role in the biogenesis of other mitochondrial outer membrane proteins. Using a quantitative proteomics strategy, we compared the steady-state levels of endogenous proteins in mitochondria isolated from wildtype or MTCH2 depleted cells (Fig. 4.2A, fig. S4.7A, Tables S2,3). We identified several outer membrane a-helical TA, signal anchored, and multipass proteins, that were reproducibly decreased upon loss of MTCH2 (Fig. 4.2B). Because MTCH2 levels do not appreciably alter the mRNA levels for these proteins (fig. S4.7C; (Replogle et al., 2022)), we concluded that the effects of MTCH2 on the mitochondrial outer membrane proteome must be occurring post-transcriptionally. To determine if MTCH2 exerts these effects specifically on biogenesis of nascent substrates, we tested a panel of mitochondrial proteins using our fluorescent reporter strategy (Fig. 4.1B). Consistent with the proteomics, MTCH2 affected the biogenesis of a functionally and biophysically diverse set of TA (Fry et al., 2021), signal anchored, and multipass proteins (Fig. 4.2C, D, fig. S4.8).

Based on these experiments, we reasoned that MTCH2 may have evolved the ability to insert a-helical proteins into the outer membrane. To test this hypothesis, we focused on TA proteins, because they are the largest class of a-helical outer membrane proteins and adopt a

uniform topology. Using an in vitro insertion assay with purified mitochondria (Fig. 4.3A, fig. S4.1C, fig. S4.9), we found that loss of MTCH2 affected the insertion of several mitochondrial TA proteins, but not unrelated intermembrane- or matrix-targeted controls (Fig. 4.3B, C and fig. S4.10, fig. S4.11). Further, using site-specific crosslinking (Fig. 4.3D, Table S4; (McKenna et al., 2020)), we demonstrated that MTCH2 physically associated with nascent substrates during their insertion (Fig. 4.3E, fig. S4.12).

Finally, to determine whether MTCH2 is sufficient for TA insertion, we purified MTCH2 (Fig. 4.3F) and optimized conditions for its reconstitution into liposomes (fig. S4.13). Using a panel of a-helical substrates, we show that purified MTCH2 specifically stimulated insertion of MTCH2 dependent, but not MTCH2 independent TAs and signal anchored proteins (Fig. 4.3G, fig. S4.14). To reconcile these results with earlier observations that trypsin-treated mitochondria remain competent for TA insertion, we found that in contrast to several subunits of the TOM complex, MTCH2 is largely trypsin resistant (Fig. 4.3H). Cumulatively, the requirement for MTCH2 in vivo and in vitro for TA insertion, together with its reconstituted insertase activity and physical association with substrates, rigorously establish MTCH2 as an insertase for a-helical mitochondrial outer membrane proteins.

Bioinformatic analysis reveals that in addition to MTCH2, other examples of SLC25 family members lacking canonical sequence motifs are found in both mitochondria and peroxisomes (Fig. 4.4A, fig. S4.15). Indeed, depletion of the close paralog MTCH1 (Palmieri, 2013) which is also localized to the mitochondrial outer membrane, had an additive effect to loss of MTCH2 on biogenesis of many mitochondrial TAs (Fig. 4.4B, fig. S4.17). This result is consistent with our genome-wide screen (fig. S4.17C) and the synthetic lethal relationship between MTCH1 and 2 (Shi et al., 2022). We therefore propose that MTCH1/2 are the founding members of a unique class of membrane protein insertases that exploit the SLC25 transporter fold (fig. S4.18).

In contrast to other solute carrier family members, in which the transmembrane helices close to form a pore that allows charged species to cross the membrane, the AlphaFold2 (Jumper et al., 2021) predicted model of MTCH2 contains a prominent groove that is accessible to

the membrane and lined with charged and polar residues (Fig. 4.4C, fig. S4.19). By introducing mutations at positions that altered the electrostatic potential of its intramembrane surfaces, we identified mutants that both diminish and enhance biogenesis of MTCH2-dependent but not independent substrates (Fig. 4.4C, fig. S4.19-4.20). We therefore concluded that MTCH2's role in TA insertion relies on a hydrophilic surface within the bilayer.

Given MTCH2's central role in mitochondrial TA biogenesis, we asked whether it may broadly affect cellular proteostasis. We find that indeed, depletion of MTCH2 leads to an increase in ER insertion of mitochondrial TAs, while MTCH2 overexpression leads to a commensurate decrease in their mistargeting to the ER (Fig. 4.4D, fig. S4.21-4.22). This effect was enhanced by further depleting ATP13A1 (McKenna et al., 2020), an ER dislocase for mislocalized mitochondrial TAs (fig. S4.21C). These data suggest that MTCH2 is the central 'gatekeeper' for the outer membrane: MTCH2 levels and activity dictate the cytosolic reservoir of mitochondrial TAs, which then can be re-routed to the ER if unable to successful integrate into mitochondria (Fig. 4.4D).

Finally, considering that insertion of several MTCH2-dependent TAs play a central role in apoptosis, we reasoned that MTCH2 activity may affect cellular sensitivity to apoptotic stimuli. To test this, we overexpressed MTCH2 in human K562 cells, which are derived from a myelogenous leukemia cell line known to upregulate the anti-apoptotic TA, BCL2L1 (Benito et al., 1995). We treated cells with imatinib, a leukemia treatment which targets the BCR-ABL oncogene, and measured apoptosis. We found that while knockout of MTCH2 did not appreciably alter apoptosis propensity in this system, overexpression of wild type MTCH2 markedly sensitizes K562 cells to imatinib treatment (Fig. 4.4E). Critically, by expressing a series of MTCH2 mutants we found that this sensitization depends on MTCH2's insertase activity.

4.4 Discussion

In summary, we have demonstrated that MTCH2 is a defining member of a family of membrane protein insertases that are necessary and sufficient for insertion of TAs into human mitochondria. MTCH2's insertase activity relies on a hydrophilic groove within the bilayer, an apparent example of convergent evolution of many membrane protein translocases including the EMC (Pleiner et al., 2020; Miller-Vedam et al., 2020; Bai et al., 2020), Hrd1 (Wu et al., 2020), and YidC (Kumazaki et al., 2014). A significant number of mitochondrial TAs are enriched in basic residues immediately C terminal to their TMDs (Rao et al., 2016), and may be particularly reliant on charged surfaces along their route into the membrane. MTCH2's role also appears to extend to the integration of a broader class of a-helical proteins into the outer membrane, including signal anchored and multipass proteins. Homologs of MTCH2 are present in metazoan peroxisomes and its orthologs are found throughout holozoa, suggesting that the MTCH2 family has co-opted the SLC25 transporter fold to function in diverse biological membranes. The transition from a solute carrier, which mediates transport of small molecules across the membrane, to an insertase appears to have been enabled by the evolution of a membrane accessible hydrophilic groove absent from MTCH2's SLC25 ancestors.

Previously, loss of MTCH2 has been reported to lead to a diverse range of phenotypes including dysregulation of mitophagy, mitochondrial fragmentation (Labbé et al., 2021), recruitment of tBID (Zaltsman et al., 2010), and altered lipid homeostasis (Rottiers et al., 2017), and was also identified in a recent genome-wide association study for Alzheimer's disease (Karch et al., 2016; Allen et al., 2015; Escott-Price et al., 2014). The identification of MTCH2 as a key gatekeeper for a-helical outer membrane proteins now provides a molecular explanation for its pleotropic phenotypes, many of which can be directly ascribed to defects in biogenesis of MTCH2 substrates.

4.5 Acknowledgements

We thank Jodi Nunnari and Maxence Le Vasseur for sharing the mitochondrial split GFP system. We thank T. Pleiner and Z. Levine for careful reading and input on the manuscript. We thank: the Whitehead Institute Flow Cytometry Core and Kathy Daniels for access to

FACS machines; the Whitehead Institute Genome Technology Core for support with sequencing of screen libraries; the Caltech Flow cytometry facility; and the Ting-Yu Wang and the Proteome Exploration Laboratory at Caltech for support for mass spectrometry. **Funding.** Research reported in this publication was supported by: Howard Hughes Medical Institute (JSW), Human Frontier Science Program 2019L/LT000858 (AG), the Heritage Medical Research Institute (RMV), and the Larry L. Hillblom Foundation (AJI). **Competing interests.** JMR consults for Maze Therapeutics and is a consultant for and equity holder in Waypoint Bio. JSW declares outside interest in 5 AM Venture, Amgen, Chroma Medicine, KSQ Therapeutics, Maze Therapeutics, Tenaya Therapeutics, Tessera Therapeutics and Third Rock Ventures. RMV is a consultant and equity holder in Gate Bioscience.

4.6 Methods

Plasmids and antibodies

Endogenous sequences used in this study for in vitro and in vivo analysis were sourced from UniProtKB/Swiss-Prot and included: squalene synthase isoform 1 (SQS/FDFT1; **Q6IAX1**), synaptojanin-2 binding protein (OMP25/SYNJBP; **P57105-1**), mitochondrial antiviral-signaling protein (MAVS; **Q7Z434-1**), mitochondrial import intermembrane translocase subunit Tim9 (TIM9; **Q9Y5J7-1**), vesicle associated membrane protein 2 (VAMP; **P51809-1**), FUN14 domain-containing protein 1 (FUNDC1; **Q8IVP5-1**), mitochondrial import receptor subunit TOM5 homolog (TOM5; **Q8N4H5-1**), mitochondrial import receptor subunit TOM22 homolog (TOM22; **Q9NS69-1**), ubiquitin carboxyl-terminal hydrolase 30 (USP30; **Q70CQ3-1**), apoptosis regulator BAX (BAX; **Q07812-1**), Bcl-2 homologous antagonist/killer (BAK1; **Q16611-1**), Bcl- 2-like protein 1 (BCL2L1; **Q07817-1**), Cytochrome b5 type B (CYB5B; **O43169-1**), peptidyl- prolyl cis-trans isomerase FKBP8 (FKBP8; **Q14318-1),** mitochondrial fission factor (MFF; **Q9GZY8-1),** inactive hydroxysteroid dehydrogenase-like protein 1 (HSDL1; **Q3SXM5-1**), amine oxidase [flavin-containing] A (MAOA; **P21397-1**), amine oxidase [flavin-containing] B (MAOB; **P27338-1**), mitochondrial Rho GTPase 1 (RHOT1; **Q8IXI2-1**), mitochondrial Rho GTPase 2 (RHOT2; **Q8IXI1-1**), CDGSH iron-sulfur domain-containing protein 1 (CISD1; **Q9NZ45**),

metaxin-1 (MTX1; **Q13505-3**), translocator protein (TSPO; **P30536)**, mitochondrial amidoxime-reducing component 1 (MTARC1; **Q5VT66-1**), mitochondrial cardiolipin hydrolase (PLD6; **Q8N2A8**), mitochondrial dynamics protein MID51 (MIEF1; **Q9NQG6-1**), mitochondrial import receptor subunit TOM20 homolog (TOM20; **Q15388**), mitochondrial import receptor subunit TOM7 homolog (TOM7; **Q9P0U1**), and mitochondrial carrier homolog 2 (MTCH2; **Q9Y6C9**).

Constructs for expression in rabbit reticulocyte lysate (RRL) used the SP64 vector as a backbone (Promega, Madison, WI). For in vitro insertion reactions, an N-terminal 3xFLAG tag and a C- terminal 6xHis tag were appended to the respective termini for affinity purification (see fig. S4.9A). Where noted, the transmembrane domains (TMDs) of TA proteins along with N- and C- terminal flanking sequences were instead conjugated to maltose binding protein (MBP) (Riggs, 2000) or villin headpiece (VHP) domains as illustrated in fig. S4.9B. When monitoring insertion into the endoplasmic reticulum (fig. S4.4 and S4.21), constructs were modified to replace the 6xHis tag with a C-terminal opsin tag containing a glycosylation acceptor site which can be used as a proxy for insertion (Brambillasca et al., 2005). TOM-dependent mitochondrial import substrates were designed by fusing dihydrofolate reductase (DHFR) to mitochondrial transit sequences from either *N. crassa* ATP synthase subunit 9 (residues 1–69) or *S. cerevisiae* cytochrome b2 (residues 1–167), which direct import to the mitochondrial matrix or intermembrane space (IMS), respectively.

For expression in K562 cells, the basis for all constructs was a mammalian expression lenti-viral backbone containing a UCOE-EF-1α promoter and a 3’ WPRE element ((Chen et al., 2019); Addgene #135448). The exception was the dual fluorescent reporter used for the CRISPRi screen (RFP-P2A-OMP25-GFP11) which was integrated into an SFFV-tet3G backbone (Jost et al., 2017). The dual color reporter system used for in vivo experiments has been previously described (Chitwood et al., 2018; Guna et al., 2018). Note that the mCherry variant of RFP was used in all instances, but the simpler nomenclature of RFP is used in the text and figures. For complementation with the GFP1-10 system, the GFP11 tag

(RDHMVLHEYVNAAGIT) was appended to the appropriate terminus of the indicated protein as determined by predicted topology (see Fig. 4.2C). In order to express GFP1-10 in the ER lumen, the human calreticulin signal sequence was appended preceding GFP1-10-KDEL (Cabantous et al., 2005; Kamiyama et al., 2016). For targeting to the intermembrane space of the mitochondria, either the targeting signal from MICU1 (aa.1-60) (Le Vasseur et al., 2021) or the targeting sequence from LACTB (aa 1-68) (Hung et al., 2014) was appended to the N terminus of GFP1-10. Expression of MTCH2 or indicated mutants was from a BFP-P2A-(MTCH2) cassette, allowing us to gate and sort for expressing cells.

The single sgRNA for MTCH2 (GACGGAGCCACCAAGCGACC) was generated by annealed oligo cloning of top and bottom oligonucleotides (Integrated DNA Technologies, Coralville, IA) into a lentiviral pU6-sgRNA EF-1α-Puro-T2A-BFP vector digested with BstXI/BlpI (Addgene, cat# 84832). BFP was excised in certain sgRNA variants when the color interfered. Though most experiments were done with a single guide that gave robust knock-down, key results were verified with an additional guide (GGGCTCACCGGGTCGCTTGG) to exclude possible off target effects. In many instances, we used a programmed dual sgRNA guide vector ((Replogle et al., 2020); Addgene #140096) to allow for multiple genes to be depleted at once or to increase efficiency of knock-down. Dual guide pairs included MTCH2-ATP13A1 (GACGGAGCCACCAAGCGACC, GGGTAAAGCAGCCCGGCGAA), MTCH1-MTCH1 (GCGGCACCGCCGCGAGCCCA, GAGCCCAGGGCGCCACTTCC), and MTCH2-EMC2 (GACGGAGCCACCAAGCGACC, GGAGTACGCGTCCGGGCCAA). Transient knock-out of MTCH1 was achieved by modifying a pLentiCRISPR backbone ((Kanarek et al., 2018), Addgene #102315) to express the following guides: GGACAACGCCCCGACCACTG and CTGCATCATCATCTCGTAGG.

Constructs for recombinant bacterial protein expression were all cloned in the pQE plasmid (Qiagen). Su9-DHFR and CaM-3C-Alfa-Sec61β(2-60)-OMP25(112-145)F128Amber were cloned downstream of a His_{14}-*bd*SUMO tag. A GFP Nb fusion protein used in this manuscript was modified from a previously established construct ((Pleiner et al., 2020);

Addgene ID #149336) to exclude the $SUMO^{Eu1}$ tag upstream of the GFP Nb. Constructs for the expression of $SENP^{EuB}$ ((Vera Rodriguez et al., 2019); Addgene ID #149333), *bd*SENP1 ((Frey & Görlich, 2014); Gift from Dirk Görlich; Addgene ID #104962), and BirA ((Pleiner et al., 2015); Gift from Dirk Görlich, Addgene ID #149334), and for site-specific incorporation of BpA at the Amber codon position ((Chin et al., 2002); Addgene #31190) have all been previously described.

Constructs for generating human stable cell lines for recombinant protein expression used either the pHAGE2 plasmid (gift from Magnus A. Hoffmann and Pamela Bjorkman) for lentiviral integration into the Expi293F cell line or the pcDNA5/FRT/TO plasmid (Thermo Scientific, #V652020) for recombinase-mediated integration into Flp-In 293 T-REx cell line. MTCH2 or HA-MTCH2 were N-terminally fused with a GFP-SUMOEu1 tag and downstream of a doxycycline-inducible CMV promoter. The EMC3-GFP-P2A-RFP expression vector has been previously described (Pleiner et al., 2020).

All plasmids are available on request.

The following antibodies were used in this study: MTCH2 (ab113707, Abcam, UK); FUNDC1 (OAAB12808, Aviva systems biology, USA); CYB5B (HPA007893, Atlas antibodies, USA); MIRO2 (RHOT2) (ab224089, Abcam, UK); CYC1 (4272, Cell signaling technology, USA); SYNJ2BP (OMP25) (15666-1-AP, Proteintech, USA); EMC3 (67205, Proteintech, USA); VDAC1 (sc-390996, Santa Cruz Biotech, USA); mitofilin (ab110329, Abcam, UK); SAMM50 (ab133709, Abcam, UK); ATP13A1 (16244-1-AP, Proteintech, USA); tubulin (T9026, Sigma-Aldrich, USA); UBQLN2 (WH0029978M3, Sigma-Aldrich, USA); TOM20 (sc-17764, Santa Cruz Biotech, USA); TOM40 (sc-365467, Santa Cruz Biotech, USA); TOM70 (14528-1-AP, Proteintech, USA). The ALFA tag was detected by coupling HRP to an ALFA nanobody (Götzke et al., 2019). The Sec61β antibody was a gift from Ramanujan Hegde. Secondary antibodies used for western blotting were: Goat anti-mouse- and anti-rabbit-HRP (#172-1011 and #170-6515, Bio-Rad, USA).

Cell culture and cell line generation

K562 cells were grown in RPMI-1640 with 25 mM HEPES, 2.0 g/L $NaHCO_3$, and 0.3 g/L L-glutamine supplemented with 10% FBS (or Tet System Approved FBS), 2 mM glutamine, 100 units/mL penicillin, and 100 μg/mL streptomycin. Cells were maintained at a confluency between 0.25×10^6 –1×10^6 cells/mL. HEK293T cells were grown in DMEM supplemented with 10% FBS, 100 units/mL penicillin and 100 μg/mL streptomycin. All cell lines were grown at 37 °C.

For the expression of EMC3-GFP, Flp-In 293 T-Rex cells were purchased from ThermoFisher Scientific (USA) (RRID: CVCL_U427) and grown in DMEM supplemented with 2 mM glutamine, 10% FBS, 15 μg/mL Blasticidine S, and 100 μg/mL Zeocin. The open reading frame to be integrated into the genomic FRT site was cloned into the pcDNA5/FRT/TO vector backbone and cell lines were generated according to the manufacturer's protocol. To allow for large scale growth, these cells were adapted to grow in suspension. Briefly, over the course of 10 days the FBS-supplemented DMEM was serially diluted with FreeStyle 293 Expression Medium (ThermoFisher Scientific). Once growing in 100% FreeStyle Medium, the cells were transferred to 1–2 L roller bottles (Celltreat, USA) and grown in a shaking incubator operating at 8% CO_2 and rotating at 125 rpm. For the expression of MTCH2, lenti-viral infected inducible Expi293F suspension cells were grown in Expi293 expression media (ThermoFisher Scientific) at 37 °C, 8% CO_2 and 125 rpm shaking in 1 L roller bottles with vented caps (Celltreat)

Three K562 cell lines were used as a basis for cell lines generated in this study: CRISPRi K562 cells expressing dCas9-BFP-KRAB (KOX1-derived) (Gilbert et al., 2014), K562-dCas9-BFP-KRAB Tet-On cells (Jost et al., 2017), and CRISPRi cells generated by stably expressing ZIM3 KRAB-dCas9-P2A-BFP from a UCOE-SFFV promoter (Replogle et al., 2022). To generate cell lines with GFP1-10 in the mitochondrial IMS or ER lumen, virus was made from the respective constructs: LACTB(GFP1-10), MICU1(GFP1- 10) or CalR(GFP1-10)-KDEL. CRISPRi K562 cells were infected with lenti-virus and sorted into 96-well plates as single cell clones using a Sony Cell Sorter (SH800S). After expansion, correct cell lines were confirmed by successful complementation with a construct targeted to

the respective compartment, appended to a GFP11. To generate the cell line used for screening, lenti-virus containing MICU1(GFP1-10) and RFP-P2A-OMP25-GFP11 under a tet-inducible promoter were co-infected to one copy per cell in CRISPRi K562 Tet-On cells, single cell sorted, and verified by induction with doxycycline (100 ng/uL) and microscopy in conjunction with MitoTracker staining to confirm correct localization.

To generate MTCH2 knock-out cell lines, K562 CRISPRi cells with or without MICU1(GFP1- 10) were nucleofected with a MTCH2 targeting guide in the pX458 backbone (Addgene #48138) using the Lonza SF Cell Line 96-well Nucleofector Kit (V4SC-2096). The pX458 backbone was adapted to express two sgRNAs targeting MTCH2 [AGCCGACATGTCTCTAGTGG], [GGCTTTGCGAGTCTGAACGT]. Two days following nucleofection, GFP-positive cells were single cell sorted into 96-well plates. After colonies from single cells grew out, loss of MTCH2 was confirmed by immunoblotting. CRISPR-Cas9-induced genome edits were identified using the computational pipeline described in (Canaj et al., 2019).

Lentivirus production

Lentivirus was generated by co-transfecting HEK293T cells with two packaging plasmids (pCMV-VSV-G and delta8.9, Addgene #8454) and the desired transfer plasmid using TransIT- 293 transfection reagent (Mirus). 48 hours after transfection, the supernatant was collected and flash frozen. In all instances, virus was rapidly thawed prior to transfection. Virus for the genome-wide CRISPRi screen was also generated using this method.

CRISPRi screen

The genome-scale CRISPRi screen was performed in duplicate as previously described (*9*, *50*). The hCRISPRi-v2 compact library (5 sgRNAs per gene, Addgene pooled library #83969) was transduced in duplicate into 330 million K562-CRISPRi-Tet-ON-((MICU1)-GFP1-10)-(tet-RFP- P2A-OMP25-GFP11) cells at multiplicity of infection (MOI) < 1 (percentage of transduced cells 48 hours after infection as measured by BFP positive cells:

30-35%). Cells were grown in 1 L of media in 1 L spinner flasks (Bellco, SKU: 1965-61010). 48 hours after spinfection with the genome-wide library, guide positive cells were selected with 1 μg/mL puromycin for three days. Following a 36 hour recovery, cells were induced with 100 ng/mL doxycycline for 36 hours and sorted on a FACS AriaII Fusion Cell Sorter. To ensure that the culture was maintained at an average coverage of more than 1000 cells per sgRNA, cells were diluted daily to 0.5×10^6 cells/mL.

During sorting, cells were gated for BFP (indicating a guide-positive cell), as well as GFP and RFP signal (successfully induced). Cells were sorted based on the GFP:RFP ratio of this final gated population. Roughly 40 million cells with either the highest (30%) or lowest (30%) RFP:GFP ratio were collected, pelleted and flash-frozen. Genomic DNA was purified using the Nucleospin Blood XL kit (Takara Bio, #740950.10) and amplified by index PCR with barcoded primers. The resulting guide library (~264 bp) was purified using SPRIbeads (SPRIselect Beckman Coulter #B23318). Sequencing was performed using an Illumina HiSeq2500 high throughput sequencer. Sequencing reads were aligned to the CRISPRi v2 library sequences, counted and quantified (Horlbeck et al., 2016). Generation of negative control genes and calculation of phenotype scores and Mann-Whitney p-values was performed as described previously (Gilbert et al., 2014; Horlbeck et al., 2016). Gene-level phenotypes and counts are available in Table S1.

Protein expression and purification

Su9-DHFR and CaM-3C-Alfa-Sec61β-OMP25(BpA)

The BL21(DE3) expression strain was used to express Su9-DHFR and CaM-3C-Alfa-Sec61β- OMP25(BpA) in LB media. Su9-DHFR expressing cultures were induced with 1 mM IPTG after an optical density of 0.6 was reached. After induction, Su9-DHFR was expressed at 37 °C for 3 hours. CaM-3C-Alfa-Sec61β-OMP25(BpA) was co-transformed with pEVOL-BpF and grown to an optical density of 0.2 followed by induction with 1% arabinose and the addition of 1 mM BpA (Bachem). Cells were then grown to an optical density of 0.6 followed by induction with 1 mM IPTG and expression at 25 °C for 6 hours.

Cells were pelleted by centrifugation and resuspended in a lysis buffer containing 500 mM NaCl, 50 mM Tris pH 7.5, 10 mM imidazole, 5 mM β-ME.

For purification, *E. coli* resuspensions were supplemented with EDTA-free protease inhibitor tablets (Roche) and lysozyme prior to lysis by sonication. Lysate was clarified by centrifugation at 18,000 rpm for 30 minutes in an SS-34 rotor. Clarified lysate was incubated with NiNTA resin for 30 minutes while rolling at 4 °C. NiNTA resin was washed extensively with resuspension buffer, and then equilibrated with a SENP elution buffer containing 150 mM NaCl, 50 mM Tris pH 7.5, 10 mM imidazole, 5 mM β-ME, 10% glycerol. NiNTA resin was then incubated with *bd*SENP1 for 2 hours at 4 °C to release SUMO-cleaved protein from the resin.

For CaM-3C-Alfa-Sec61β-OMP25(BpA) the resuspension buffer included 1 mM CaCl2, and SENP elution buffer contained 100 mM NaCl, 50 mM Tris pH 7.5, 10 mM imidazole, 5 mM β-ME, 1 mM $CaCl_2$. Additionally, protein was cleaved overnight with 3C protease following SENP1 elution. Cleaved protein was then concentrated to 250 µL and injected onto a Superdex 200 increase 10/300 GL size exclusion column equilibrated in a buffer containing 150 mM KoAc, 50 mM HEPES pH 7.4, 2 mM $MgAc_2$, 1 mM $CaCl_2$, 1 mM DTT, 10% glycerol. Protein-containing fractions were pooled, concentrated, aliquoted, and flash-frozen.

Biotinylated GFP-Nb and Alfa-NB

Expression and purification of all GFP and Alfa nanobody constructs as well as bdSENP1, BirA, and SENPEuB generally proceeded as follows: the NEB Express Iq expression strain was used with TB medium. Cultures were induced with 0.2 mM IPTG after an optical density of 2.0 was reached. Protein was expressed at 18 °C for 18–20 hours. Cells were pelleted by centrifugation and resuspended in a lysis buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 1 mM PMSF. Cells were lysed by sonication and lysate was clarified by centrifugation at 18,000 rpm for 30 minutes in an SS-34 rotor. Clarified lysate was incubated with NiNTA resin for 1 hour while rolling at 4 °C. NiNTA resin was washed extensively with resuspension buffer and the eluted with a buffer containing 50 mM

Tris pH 7.5, 300 mM NaCl, 500 mM imidazole, 250 mM sucrose. The imidazole was removed using a PD-10 desalting column (GE Healthcare, USA). The following protein-specific modifications were applied to the above protocol: the Rosetta-gami 2(DE3) expression strain was used to express His14-Avi- SUMOstar-AlfaNb and expression time was limited to 6 hours at 18 °C. SENPEuB was limited to an expression time of 6 hours at 18 °C. bdSENP1 was expressed as a His14-TEV- fusion and was cleaved with TEV protease overnight after buffer exchange and then run over NiNTA resin to remove the cleaved tag. The expression and purification of His14-Avi-SUMOEu1-anti GFP nanobody, bdSENP1, BirA, and SENPEuB have all (Pleiner et al., 2018) been previously described (Pleiner et al., 2020; Vera Rodriguez et al., 2019).

All nanobody constructs used in this publication were biotinylated in a buffer containing 50 mM Tris pH 7.5, 100 mM NaCl, 12.5 mM $MgCl_2$, 10 mM ATP, 10 mM biotin and BirA at a 1:50 molar ratio to the nanobody substrate. Biotinylation reactions were incubated at 25 °C for 3 hours and then buffer exchanged in a PD-10 column equilibrated with a buffer containing 50 mM Tris pH 7.5, 200 mM NaCl, 1 mM DTT, 250 mM sucrose. Biotinylated protein was aliquoted and flash-frozen.

MTCH2 and EMC

We isolated MTCH2, GFP-SUMOEu1-HA-MTCH2, and EMC (via EMC3-GFP) under native conditions from detergent-solubilized cells using a biotinylated anti-GFP nanobody, expressed and purified as previously described (Pleiner et al., 2020). Cells were grown in 1 L roller bottles. For the MTCH2 lines, expression was induced for at least 48 hours with 1 µg/mL doxycycline. Cells were then harvested by centrifugation, washed with 1×PBS, and then pellets were weighed.

Purification generally proceeded as follows: cell pellets were resuspended at a ratio of 1 g to 10 mL hypotonic lysis buffer containing 10 mM HEPES pH 7.5, 10 mM KoAc, 0.15 mM MgAc2, 0.5 mM DTT, supplemented with EDTA-free protease inhibitor tablets (Roche). The cell resuspension was incubated on ice for 10 minutes to allow cells to swell, and then

lysed in a Dounce homogenizer with 10× strokes. The NaCl concentration was adjusted to 180 mM immediately after Dounce homogenization. Cell membranes were pelleted by centrifugation at 18,000 rcf in an SS-34 (28020TS, Thermo Fisher) rotor for 10 minutes. Supernatant was discarded and cell membranes were washed by resuspending and pelleting 2× in membrane wash buffer containing 10 mM HEPES/KOH pH 7.5, 200 mM NaCl, 0.15 mM $MgAc_2$, 0.5 mM DTT. The resulting pellet was resuspended at a ratio of 1 g (original cell pellet weight) to 6.8 mL solubilization buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 1% deoxy-BigCHAP (DBC; Anatrace Cat # B310), 1 mM DTT, supplemented with EDTA-free protease inhibitor tablets (Roche). After 30 minutes of head-over-tail incubation with solubilization buffer, the lysate was cleared by centrifugation for 30 minutes at 4 °C and 18,000 rpm in a SS-34 rotor. The supernatant was then added to pre-equilibrated magnetic Streptavidin resin (Thermo Fisher) bound to biotinylated anti-GFP nanobody and blocked with free biotin. After 1 hour of binding while rolling at 4 °C, the resin was washed four times with wash buffer (solubilization buffer with 0.2% DBC). Resin was then incubated in wash buffer + 600 nM SENPEuB on ice for 2 hours to release SUMO-cleaved MTCH2 from the resin. Eluted samples were analyzed via SDS-PAGE with Sypro Ruby stain (Bio-Rad).

For EMC3-GFP, the whole cell pellet was solubilized in 1% DBC-containing buffer, without a hypotonic lysis and membrane washing step, and the GFP Nb was fused to a cleavable SUMOEu1 module. GFP-SUMOEu1-HA-MTCH2 was purified via a GFP Nb fused to a cleavable SUMOstar module.

Mitochondrial isolation, semi-permeabilized cells, and total membrane fractionation

Mitochondrial isolation for K562 cells was adapted from an established protocol (Richter-Dennerlein et al., 2014). K562 cells were centrifuged at 220*g* for 5 minutes. Pellets were washed once in PBS and pelleted by spinning at 500*g* for 5 minutes. Pellets were resuspended in a homogenization buffer containing 210 mM mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA, 1 mM PMSF, 2 mg/mL BSA. After incubating on ice for 10 minutes, cells were then lysed with a glass Dounce homogenizer with a tight-fitting pestle, or a Potter-

Elvehjem homogenizer motor-driven at 1600 rpm for large scale purifications. Homogenized cells were pelleted at 1300*g* for 5 minutes to remove nuclei and unbroken cells, then the supernatant was transferred to a clean tube. This step was repeated twice. Nuclei-free homogenized cells were then centrifuged at 11,000*g* for 10 minutes. The supernatant was removed and the mitochondria-containing pellet was then resuspended in an isolation buffer containing 210 mM D-mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA. Mitochondria were then pelleted and resuspended in fresh isolation buffer to wash away BSA and cytoplasmic proteins. After a final pelleting step, mitochondria were resuspended in a small volume (5–50 µL) of isolation buffer. To normalize mitochondrial samples, the protein concentration was measured using a Bradford assay.

For experiments using trypsin-treated mitochondria, the final two steps performed after pelleting mitochondria used import buffer containing 250 mM sucrose, 5 mM $MgAc_2$, 80 mM KOAc, and 20 mM HEPES. Mitochondria were then mixed with 0 or 50 µg/mL trypsin (Sigma-Aldrich #T1426) dissolved in import buffer and incubated on ice for 30 minutes. 50 µg/mL trypsin inhibitor (Sigma-Aldrich #T9128) and 1 mM PMSF were added to quench the reaction. Mitochondria were pelleted, washed once in import buffer + 5 µg/mL trypsin inhibitor, then resuspended in a concentrated volume prior to use in import experiments.

To further enrich mitochondrial samples for certain experiments, a percoll gradient was used. Isolation buffer density gradients were formed in 3 layers with 40%, 26%, and 12% percoll. Resuspended mitochondria were layered on top of the gradient. Gradients were centrifuged at 45,000*g* for 45 minutes in a TLS-55 rotor (Beckman Coulter). Pure mitochondria were retrieved from the 40%–26% percoll interface (see fig. S4.1B). Mitochondria were diluted 5-fold in isolation buffer, and then pelleted and washed in isolation buffer twice more.

HEK293T cells, either wild-type or an EMC5 knock-out background were semi-permeabilized using standard methods (Richter-Dennerlein et al., 2014). Briefly, $3 \times 10_6$ cells were collected and washed once with ice cold wash buffer containing 25 mM HEPES pH 7.4, 100 mM KOAc, 2 mM $MgAc_2$. The cells were resuspended in 1 mL SP buffer containing 25 mM Hepes pH 7.4, 100 mM KOAc, 2 mM $MgAc_2$, 50 µg/mL digitonin) and incubated

on ice for 5 minutes. The semi-permeabilized cells were collected by centrifugation at 500*g* for 5 minutes at 4 °C, with the digitonin removed by washing three times. Finally, the cells were pelleted at 12,000*g* for 15 seconds and resuspended in 10 µL wash buffer. To test the integrity of the outer mitochondria membrane as in fig. S4.1A, SP cells and purified mitochondria (as described above) were incubated with the indicated amount of proteinase K (PK). Following quenching, the resulting reaction was subjected to blotting against mitofilin, an intermembrane localized protein which should not be accessible by PK when the outer mitochondrial membrane is intact.

To fractionate total cellular membranes from cytosol, 4 million K562 cells were pelleted by spinning at 500*g* for 5 minutes, then washed 1× in PBS and pelleted again. The washed pellet was then resuspended in 150 µL of a digitonin lysis buffer containing 50 mM HEPES pH 7.5, 10 mM KAc, 2 mM $Mg(Ac)_2$, 1 mM DTT, 0.05% digitonin, and supplemented with EDTA-free protease inhibitor tablets (Roche). The resuspension was incubated on ice for 10 minutes to allow cells to swell, and then passed through a 27 G 1⁄4" needle attached to a syringe 12×. The sample was then adjusted to 0.19 M NaCl, and centrifuged for 15 minutes at 18000*g* in a cold benchtop centrifuge. The cytoplasm-containing supernatant was removed and multiple samples were normalized via UV absorbance at 280 nm. The pellet was washed in digitonin lysis buffer + 0.19 NaCl. The wash pellet was resuspended in a solubilization buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $Mg(Ac)_2$, 1 mM DTT, 1% GDN, and supplemented with EDTA- free protease inhibitor tablets (Roche). After incubation on ice for 30 minutes, the sample was centrifuged for 10 minutes at 18000g in a cold benchtop centrifuge. The solubilized membrane containing supernatant was removed and multiple samples were normalized via UV absorbance at 280 nm.

<u>In vitro translation and insertion</u>

In vitro translations were carried out in rabbit reticulocyte lysate (RRL) as previously described (Guna et al., 2018). Constructs for in vitro translation reactions were based on the SP64 vector (Promega, USA). Templates for transcription were generated by PCR, with primers binding and upstream of the SP64 promoter and roughly 200 bp downstream of the

stop codon (Sharma et al., 2010). Following transcription at 37 °C for 1.5 hours, reactions were used directly in a translation reaction. Substrates were translated for 15–30 minutes at 32 °C in the presence of radioactive 35S- methionine. Prior to addition of mitochondria or semi-permeabilized cells, 1 mM puromycin was added to prevent further synthesis.

Mitochondrial insertion reactions used isolated mitochondria, prepared as described above. Insertion reactions were performed by diluting 4 µL of a puromycin treated translation reaction in 50 µL of import buffer (250 mM sucrose, 5 mM $Mg(Ac)_2$, 80 mM KAc, 20 mM HEPES pH 7.4, 2.5 mM ATP, 15 mM succinate) with 15 µg of purified mitochondria and further incubating at 32 °C for 30 minutes. For competition experiments in Fig. 4.1A and fig. S4.2, insertion reactions were carried out in the presence of 1, 2, or 5 µM recombinant Su9-DHFR, and 5 µM methotrexate (BP266510, Fisher Chemicals, USA).

Protease digestions were initiated by the addition of proteinase K at 0.25 mg/mL, and reactions were then incubated on ice for 1 hour. Reactions were quenched by the addition of 5 mM PMSF in DMSO, followed by transfer to boiling 1% SDS (final concentration) in 0.1 M Tris/HCl pH 8.0. His-tagged protected fragments were enriched by incubating with NiNTA resin in IP buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 1% Triton). Proteinase K-digested reactions were diluted to 1 mL and mixed with 10 µL resin, then incubated with end-over-end mixing for 1.5 hours at 4 °C. The resin was further washed with 3 × 1 mL IP buffer, and the products eluted from the resin in sample buffer containing 50 mM EDTA pH 8.0.

Insertion reactions with semi-permeabilized (SP) cells used a ratio of 1 µL cells per 10 µL translation reaction. To verify insertion of substrates into the ER as indicated by glycosylation, a tripeptide competitor of glycosylation (Asn-Tyr-Thr) was added at 50 µM when indicated.

Mass spectrometry

For the crosslinking-IP samples, TCA-precipitated pellets were resuspended in a buffer containing 8 M Urea and 100 mM Tris pH 8.5. The sample was reduced by incubation with 3 mM TCEP for 20 minutes, then alkylated by incubation with 10 mM iodoacetamide for 15 minutes, all at room temperature. The sample was then digested with 2 ng/μL LysC for 4 hours at room temperature, diluted 4-fold with 100 mM Tris pH 8.5, and $CaCl_2$ was added to 1 mM. The sample was then digested with 4 ng/μL trypsin overnight at room temperature. Samples were acidified by adding trifluoroacetic acid to 0.5%, desalted using Pierce C18 spin columns (Pierce), lyophilized, and then resuspended in 2% acetonitrile, 0.2% formic acid.

For the mitochondrial proteomics experiments, the S-trap sample preparation kit (ProtiFi) was used according to the manufacturer's instructions. Sample was digested on the S-trap column with 1 μg Trypsin per 10 μg protein overnight at 37°C. In addition to the provided S-trap sample preparation protocol, a final elution step with 70% acetonitrile, 1% formic acid was added. Eluted peptides were lyophilized, and then resuspended in 2% acetonitrile, 0.2% formic acid. The peptide concentration was determined with a Quantitative Fluorometric Peptide Assay (Pierce) kit.

LC-MS/MS analysis for the crosslinking-IP experiment (Fig. 4.3D, Table S4) was performed with an EASY-nLC 1200 (ThermoFisher Scientific, San Jose, CA) coupled to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (ThermoFisher Scientific). Peptides were separated on an Aurora UHPLC Column (25 cm × 75 μm, 1.6 μm C18, AUR2-25075C18A, Ion Opticks) with a flow rate of 0.35 μL/min for a total duration of 75 min and ionized at 1.8 kV in the positive ion mode. The gradient was composed of 2%–6% solvent B (3.5 min), 6%–25% B (42 min), 25%– 40% B (14.5 min), and 40%–98% B (15 min); solvent A: 2% ACN and 0.2% formic acid; solvent B: 80% ACN and 0.2% formic acid. MS1 scans were acquired at the resolution of 60,000 from 375 to 1500 m/z, AGC target 3e6, and maximum injection time 15 ms. The 12 most abundant ions were targeted for MS2 scans acquired at a resolution of 30,000, AGC target 1e5, maximum injection time of 60 ms, and normalized collision energy of 28. Dynamic exclusion was set to 30 s and ions with charge +1, +7, +8

and >+8 were excluded. The temperature of ion transfer tube was 275 °C and the S-lens RF level was set to 60. MS2 fragmentation spectra were searched with SEQUEST running within Proteome Discoverer (version 2.5, Thermo Scientific) against the UniProt human reference proteome comprised of 79,052 proteins covering 20,577 genes (UP000005640). The maximum missed cleavages was set to 2. Dynamic modifications were set to oxidation on methionine (M, +15.995 Da), deamidation on asparagine and glutamine (N and Q, +0.984 Da), phosphorylation of serine and threonine (S and T, +79.966 Da), protein N-terminal acetylation (+42.011 Da), and protein N-terminal Met-loss (-131.040 Da). Carbamidomethylation on cysteine residues (C, +57.021 Da) was set as a fixed modification. The maximum parental mass error was set to 10 ppm, and the MS2 mass tolerance was set to 0.03 Da. The false discovery threshold was set strictly to 0.01 using the Percolator Node validated by q-value. The relative abundance of parental peptides was calculated by integration of the area under the curve of the MS1 peaks using the Minora LFQ node.

To identify enriched proteins, proteins that were detected in both samples were ranked by iBAQ intensity within each sample, and enrichment was assessed based on the difference between the +UV and the -UV iBAQ rank. The final results of this analysis are listed in Table S4.

LC-MS/MS analysis to assess differences in protein content between percoll gradient-enriched mitochondria derived from K562 wild-type or MTCH2-depleted cells (Fig. 4.2A, Table S2) was performed with an EASY-nLC 1200 (ThermoFisher Scientific) coupled to an Orbitrap Eclipse Tribrid mass spectrometer (ThermoFisher Scientific). Peptides were separated on an Aurora UHPLC Column (25 cm × 75 µm, 1.6 µm C18, AUR2-25075C18A, Ion Opticks) with a flow rate of 0.35 µL/min for a total duration of 75 min and ionized at 1.8 kV in the positive ion mode. The gradient was composed of 2%–6% solvent B (3.5 min), 6%–25% B (42 min), 25%–40% B (14.5 min), and 40%–98% B (15 min); solvent A: 2% ACN and 0.2% formic acid; solvent B: 80% ACN and 0.2% formic acid. MS1 scans were acquired at the resolution of 120,000 from 350 to 1,600 m/z, AGC target 1e6, and maximum injection time of 50 ms. MS2 scans were acquired in the ion trap using fast scan rate on

precursors with 2–7 charge states and quadrupole isolation mode (isolation window: 0.7 m/z) with higher-energy collisional dissociation (HCD, 30%) activation type. Dynamic exclusion was set to 30 s. The temperature of ion transfer tube was 300 °C and the S-lens RF level was set to 30. MS2 fragmentation spectra were searched with SEQUEST running within Proteome Discoverer (version 2.5, Thermo Scientific) against the reviewed sequences from the UniProt human reference proteome comprised of 20,383 proteins (UP000005640). The maximum missed cleavages were set to 2. Dynamic modifications were set to oxidation on methionine (M, +15.995 Da), deamidation (N and Q, +0.984 Da), protein N- terminal acetylation (+42.011 Da) and protein N-terminal Met-loss (-131.040 Da). Carbamidomethylation on cysteine residues (C, +57.021 Da) was set as a fixed modification. The maximum parental mass error was set to 10 ppm, and the MS2 mass tolerance was set to 0.6 Da. The false discovery threshold was set strictly to 0.01 using the Percolator Node validated by q- value. The relative abundance of parental peptides was calculated by integration of the area under the curve of the MS1 peaks using the Minora LFQ node.

LFQ was then performed with the Minora feature detector, feature mapper, and precursor ions quantifier nodes. Retention time alignment was performed with maximum RT shift of 5 min and a minimum S/N threshold of 10. Quantified peptides included unique + razor, protein groups were considered for peptide uniqueness, shared Quan results were not used, Quan results with missing values were not rejected, and precursor abundance was based on extracted ion intensity. Imputation was performed using a low abundance resampling method at the peptide level. The quantitative proteomics data was exported from ProteomeDiscoverer as an excel file and the MitoCoP database (Morgenstern et al., 2021) was used to remove all non-mitochondrial proteins from the dataset prior to statistical analysis using the R statistical computing environment (R version 4.0.2). Protein abundances were normalized between samples with a random forest (R randomForest version 4.6-14) regression (Breiman, 2001), then modeled for expression differences (R limma version 3.44.3) using the linear model fit, with fold changes calculated as a log2 difference. Proteins and peptides identified are listed in Table S2.

LC-MS/MS analysis to assess differences in protein content between crude isolated mitochondria derived from K562 wild-type or MTCH2-depleted cells (fig. S4.7A, Table S3) was performed similarly to the percoll-enriched mitochondrial samples with the following changes: peptides were separated on an Aurora UHPLC Column (25 cm × 75 μm, 1.7 μm C18, AUR3-25075C18, IonOpticks) with a flow rate of 0.35 μL/min for a total duration of 75 min and ionized at 1.6 kV in the positive ion mode. MS2 scans were acquired in the ion trap using fast scan rate on precursors with 2–7 charge states and quadrupole isolation mode (isolation window: 1.2 m/z) with higher-energy collisional dissociation (HCD, 30%) activation type. Dynamic exclusion was set to 15 s. For LFQ, Retention time alignment was performed with maximum RT shift of 2 min and a minimum S/N threshold of 5.

Volcano plot figures were generated in Python using the matplotlib package. Figures that specify submitochondrial localization are based on annotations from Mitocarta 3.0 (Rath et al., 2021), with the following modifications: subcellular localizations for TDRKH, HSDL1, and ARMC1 were added manually based on published data (Go et al., 2021; Wagner et al., 2019), TMEM11 is annotated as IM in Mitocarta 3.0; however, a recent preprint provides evidence that TMEM11 is largely in the OM, so this gene was included in analysis of OM proteins in our proteomic data (Gok & Friedman, 2022).

Photo-crosslinking of recombinant substrate with isolated mitochondria

An insertion reaction was prepared in a buffer containing 100 mM KAc, 50 mM HEPES pH 7.4, 2 mM $Mg(Ac)_2$, 0.5 mM $CaCl_2$. Percoll-gradient enriched mitochondria were added to a final protein concentration of 0.2 mg/mL and CaM-3C-Alfa-Sec61β-OMP25(BpA) was added to a final concentration of 1 μg/mL. Insertion was initiated by the addition of 2 mM EGTA. The insertion reaction was split into -UV and +UV samples. The +UV sample was transferred to a prechilled 6-well plate and left on top of an ice-cooled aluminum block ~8 cm under a UVP B- 100 lamp for 10 minutes.

Both samples were mixed with 10× volumes of IP buffer containing 100 mM KAc, 50 mM HEPES pH 7.4, 2 mM $Mg(Ac)_2$, and 1% Triton X-100. After incubating on ice for 10

minutes, samples were centrifuged at 18,000 rpm. Supernatant was mixed with 10 µL packed streptavidin agarose resin (Thermo Scientific Cat. #20357) which had been functionalized with biotinylated Alfa-Nb, blocked with free biotin, and equilibrated in IP buffer. After 1 hour of binding head- over-tail at 4 °C, the unbound fraction was removed, and the resin was then washed 2× with 1 mL IP wash buffer, 2× with 1 mL IP wash buffer + 0.5 M NaCl, and 2× with 1 mL IP wash buffer. Resin was then incubated with 10 µL IP wash buffer + 300 nM SUMOstar protease (LifeSensors Cat # SP4110) for 30 minutes on ice. 1 mL IP wash buffer was added to protease treated resin and incubated for 5 minutes on ice. Eluted protein was then removed from resin and TCA precipitated by adding 1:10 volume 100% TCA and incubating on ice for 10 minutes, followed by centrifuging at max speed in a chilled benchtop centrifuge for 10 minutes. The pellet was then washed 2× with ice-cold acetone and prepared for mass spectrometry as described below.

Samples used for immunoblotting were prepared with the following differences: isolated mitochondria were not purified on a percoll gradient; Pierce magnetic streptavidin resin was used (Thermo Scientific Cat. #88817), an IP buffer containing 200 mM NaCl, 50 mM HEPES pH 7.4, 1% Triton X-100 was used for solubilization; 4 × 1 mL wash steps; and an elution buffer containing 200 mM NaCl, 50 mM HEPES pH 7.4, and 0.05% Triton X-100 was used for 2× wash steps and then elution was performed in the same buffer + 300 nM SUMOstar in a 10 µL volume.

Proteoliposome reconstitutions and insertions

Reconstitutions of protein into liposomes were similar to previously described methods (Mariappan et al., 2011; Guna et al., 2018). The following phospholipids were obtained from Avanti Polar Lipids: phosphatidyl-choline (PC) and phosphatidyl-ethanolamine (PE) from egg, and synthetic 1,2-dioleoyl-sn-glycero-3- phosphoethanolamine-N-lissamine rhodamine B (Rh-PE). The liposome mixture contained PC:PE:Rh-PE at a mass ratio of 8:1.9:0.1. Rh-PE was used to monitor recovery throughout the reconstitutions and for quantification. Lipids were mixed at the indicated ratio as chloroform stocks, adjusted to 10 mM DTT and dried by centrifugation under vacuum overnight. The resulting lipid film was rehydrated to a final

concentration of 20 mg/mL in lipid buffer (15% glycerol, 50 mM HEPES pH 7.4) and mixed for 8 hours at 25 °C until a homogenous mixture was achieved. Lipids were then diluted with more lipid buffer and supplemented with DBC to produce a lipid/DBC mixture containing 2% DBC and 10 mg/mL lipids. BioBeads-SM2 (Bio- Rad) were prepared by activation with methanol, washing thoroughly with distilled water, and then were resuspended in water into a final slurry where they occupied 50% of the final volume. For reconstitutions, excess liquid was removed from BioBeads by aspiration just before use. Reconstitutions using purified EMC and MTCH2 in 0.25% DBC were obtained as described below. In initial experiments, we determined the relative concentration of purified MTCH2 compared with the amount in isolated mitochondria from K562 cells. Different dilutions of purified MTCH2 were mixed with constant amounts of lipids and adjusted to a final buffer concentration of 100 mM NaCl, 25 mM HEPES pH 7.4, 2 mM $MgCl_2$, 0.8% DBC. Liposomes were made using the same buffer and detergent conditions. A standard 100 µL reaction contained 10-40 µL purified MTCH2, 20 µL of the 10 mg/mL lipid/DBC mixture, and the remaining volume made up with buffer, salts, and detergent. The molar ratio of lipid:protein used in reconstitutions was 5500:1, 2500:1 and 1500:1 depending for the high, medium, or low MTCH2- PL conditions, respectively. This protein/lipid/detergent mixture was added to 120–140 µL BioBeads in 1.5 mL Eppendorf tubes. The slurry was mixed in a thermomixer for 18 hours at 4 °C. In some cases, the fluid phase was removed and added to another 1.5 mL Eppendorf tube containing 120 µL Biobeads, and mixed in a thermomixer for 2 hours at 23 °C. The fluid phase was then separated and diluted with ten volumes of ice-cold water. The proteoliposomes were sedimented in a TLA120.2 rotor at 70,000 rpm for 30 minutes, and resuspended in 16.7 µL liposome resuspension buffer (100 mM KAc, 50 mM HEPES pH 7.4, 2 mM $Mg(Ac)_2$, 250 mM sucrose, 1 mM DTT). Substrates for insertions were prepared by translating in RRL for 15 min, then the reactions were treated with 1 mM puromycin to prevent further synthesis. Insertion reactions consisted of 8 µL translation in RRL, 1 mM EGTA, and 2 µL buffer, liposomes, proteoliposomes, or isolated mitochondria. The reactions were incubated at 32 °C for 30 min, unless indicated otherwise, before being treated with 0.5 mg/mL PK for 1 hour on ice. The reactions were quenched and the protected fragments enriched with NiNTA resin as described above.

In order to determine the percentage of MTCH2 that was inserted in the correct orientation, we reconstituted GFP-SUMOEu1-HA-MTCH2 into proteoliposomes as described above. We then mixed 1 µL of the final resuspension with SENPEuB to a final concentration of 5 µM in the presence or absence of 1% Triton X-100. Cleavage reactions were incubated on ice for 30 minutes and analyzed by immunoblotting with a MTCH2 antibody.

Flow cytometry

For all reporter experiments, respective K562 cell lines were spinfected with lenti-virus for indicated constructs and analyzed by flow cytometry after 48–72 hours. All reporter experiments were performed at least twice. For the apoptosis experiment, wildtype K562 cells expressing either MTCH2-P2A-BFP or BFP alone were treated with 5 µM imatinib mesylate (461080010, Fisher Scientific, USA) for 72 hours. Cells were harvested, washed once with ice cold PBS, then resuspended in 100 µL staining buffer (10 mM HEPES, 140 mM NaCl, 2.5 mM $CaCl_2$ pH 7.4, 5% Annexin-FITC (Invitrogen, A13199), 50 µg/mL propidium iodide (P1304MP, Invitrogen, USA) before analysis by flow cytometry. For treatment with the GPAT inhibitor FSG67, K562 cells were treated for 16 hours with 75 µM FSG67 (Cedarlane Laboratories Cat. #10-4577) as previously described (Labbé et al., 2021) prior to analysis by flow cytometry. All samples were run on either an NXT Flow Cytometer (ThermoFisher) or a MACSQuant VYB (Mitenyi Biotec). Flow cytometry data was analyzed either in FlowJo v10.8 Software (BD Life Sciences) or Python using the FlowCytometryTools package.

Quantitative PCR

RNA levels were measured with quantitative PCR. Cells treated with a non-targeting guide or a guide targeting MTCH2 for eight days, and then RNA was extracted using an RNeasy kit (74104, Qiagen). The RNA was treated with amplification grade DNase I (18068015, Invitrogen), and then reversed transcribed using the SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (11752050, Invitrogen), according to the manufacturer's

instructions. The reactions were analyzed using a StepOnePlus Real-Time PCR system. The relative expression ratios were calculated compared to the housekeeping gene HPRT1, using primer efficiencies measured using a standard dilution curve. The data is in triplicate, and means and standard deviations are plotted. The primers used were as follows:

MTCH2 (fwd GTCAGCTTCCTGGTCTCTTTAG, rev CCTTGTCACTCTCCTGGTAATG); OMP25 (fwd GTGCATATCCTCCATCCCATAG, rev GGTCTTACCGGACCTCTTATT); CYB5B (fwd ACAGCCAGGTGAAAGCTAAA, rev CCAACACCCTCTATTTCGGTAG); RHOT2 (fwd AGCGTCTACAAGCACCATTAC, rev AGGACCCTGTAGAGTGAGAAG); VDAC (fwd CAATACACTAGGCACCGAGATT, rev TCTGTCCCGTCATTCACATTAG); FUND1 (fwd GTAGGTGGTGGCTTTCTTCTT, rev CTGCTTTGTTCGCTCGTTTC); CYC1 (fwd CAAGTGGTCAGTCCTGAAGAG, rev CTTCCAGATGAAGAGAGGCTTAG).

Microscopy

To visualize localization to the mitochondria, K562 cells were stained with 1 μM Mitotracker Deep Red FM (M22426, ThermoFisher) for 30 minutes. Cells were collected, spun down, washed in fresh media and plated onto a 96-well glass bottomed plate (160376, Thermo Fisher). K562 cells were then briefly spun down by centrifugation at 100*g* for 5 minutes and imaged on a Zeiss LSM710 NLO Laser Scanning confocal microscope.

Sequence alignments

An alignment of individual SLC25 repeats (Fig. 4.4A, S15) from various human transporter was generated using the hmmalign tool from the HMMER3.3 package (Eddy, 2011) and a precomputed hmm profile of the SLC25 family from pfam (pf00153) (Mistry et al., 2021). An alignment of MTCH2 homologues was generated using MUSCLE (Edgar, 2004) and the ESPRIPT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)) was used to display the alignment (fig. S4.16).

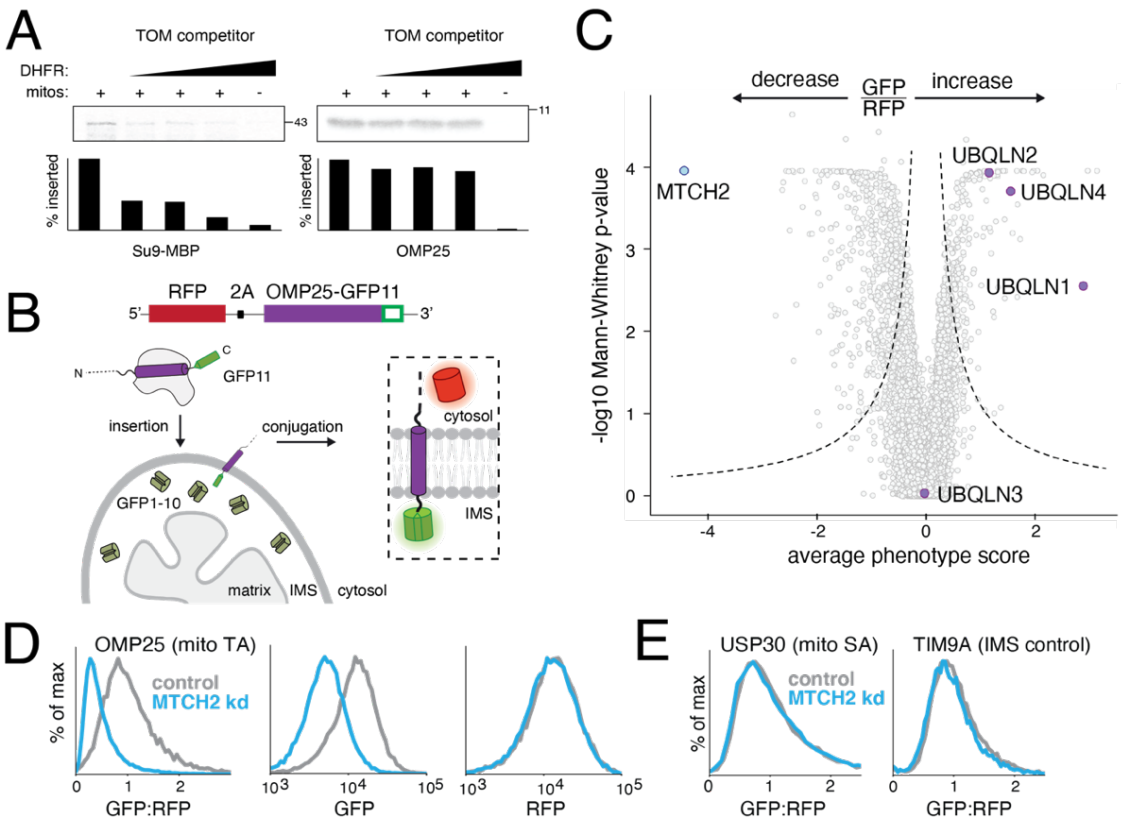


Figure 4.1. Systematic characterization of human mitochondrial TA biogenesis. (A) An 35S- methionine labelled TOM substrate (made from a fusion of the canonical TOM targeting sequence Su9 and the globular protein MBP) or OMP25 (a mitochondrial TA protein) were translated in rabbit reticulocyte lysate and released from the ribosome using puromycin. Competition assays were performed by incubation with purified mitochondria (see fig. S4.1) in the presence of increasing concentrations of a recombinant TOM competitor (Su9-DHFR). Mitochondrial insertion was assessed by protease protection and analyzed by SDS-PAGE and autoradiography. See also fig. S4.2. (B) Schematic of the split GFP reporter system used to specifically query integration of substrates into the outer mitochondrial membrane. A mitochondrial membrane protein fused to GFP11 is expressed in a cell constitutively expressing GFP1-10 in the intermembrane space (IMS) along with a translation normalization marker (RFP) Successful integration into the outer membrane results in complementation and GFP fluorescence. (C) Volcano plot of GFP:RFP stabilization phenotype for the three strongest sgRNAs versus Mann-Whitney p values

from two independent replicates of a genome-wide CRISPRi screen using OMP25-GFP11. Individual genes are displayed in grey, and specific factors that increase or decrease OMP25 mitochondrial integration are highlighted and labelled. (D) Integration into mitochondria of the OMP25-GFP11 reporter described in (B) was assessed in K562 cells expressing a non-targeting (control) or MTCH2 knock down sgRNA. GFP fluorescence relative to the normalization marker RFP was determined by flow cytometry and displayed as a histogram. Individual channels are also shown. (E) Biogenesis of USP30-GFP11, an outer membrane resident signal anchored protein, and TIM9A-GFP11, an IMS localized protein, were assessed as in (D).

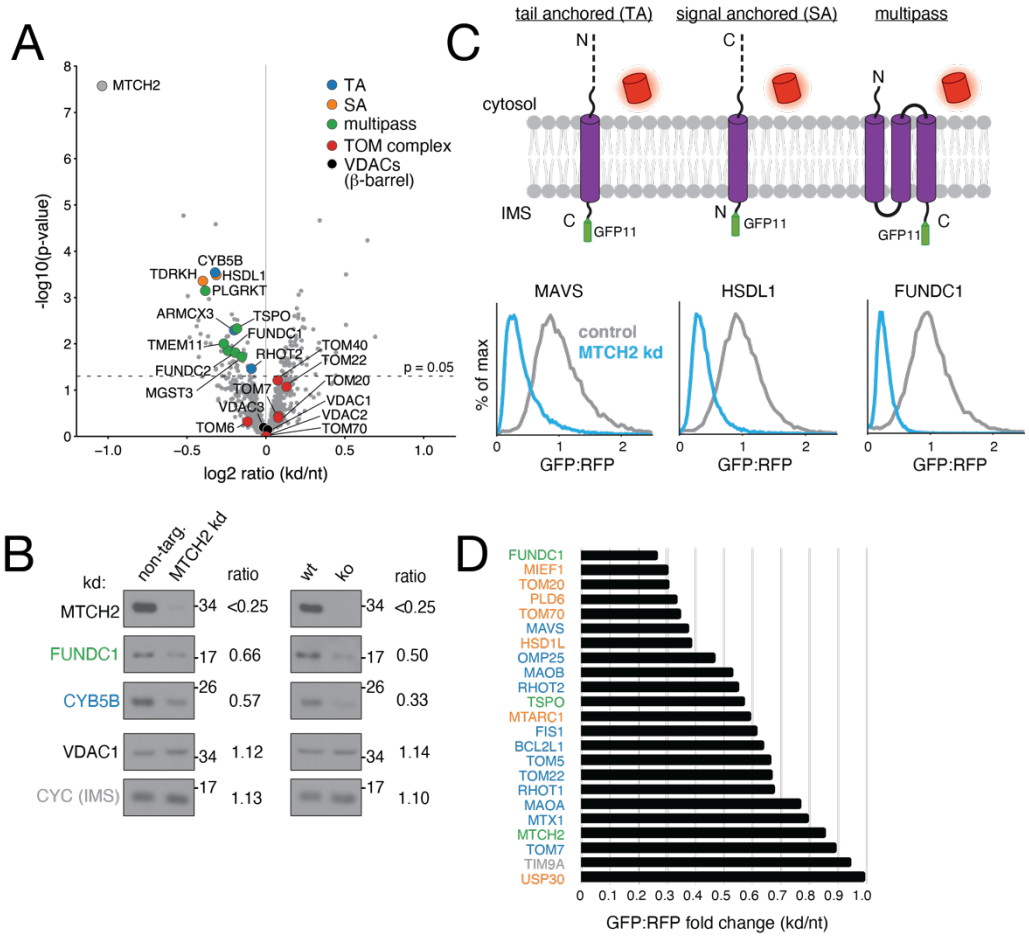


Figure 4.2. MTCH2 is required for mitochondrial outer membrane protein biogenesis. (A) Label-free mass spectrometry analysis of purified mitochondria isolated from K562 cells using a percoll gradient (fig. S4.1B) expressing a MTCH2 targeting sgRNA (kd) compared to a non- targeting control (nt). Displayed are proteins that across four biological replicates were statistically altered in MTCH2 depleted versus non-targeting guide expressing cells colored according to the indicated key (signal anchored: SA). (B) Immunoblotting of endogenous proteins in mitochondria isolated from MTCH2 depleted (kd) and control cells in (generated as in A; left), and wild type (wt) and MTCH2 knock out (ko) cells (right). Substrates are colored by topology based on the key shown in (A). Quantification of fold-change in depleted vs control cells is displayed as determined using a dilution series for each antibody. (C) Flow cytometry analysis of integration of outer membrane protein reporters using the split GFP system described in Fig. 4.1B. GFP fluorescence relative to an RFP expression control are displayed as histograms in MTCH2

knockdown versus non-targeting K562 CRISPRi cells. Displayed are representative examples of a TA, signal anchored (SA), and multipass membrane protein that have a MTCH2 dependent biogenesis defect. (D) Summary of dependence on MTCH2 for the indicated outer membrane substrates determined using the fluorescent reporter system shown in (C) and colored by topology based on the key in (A).

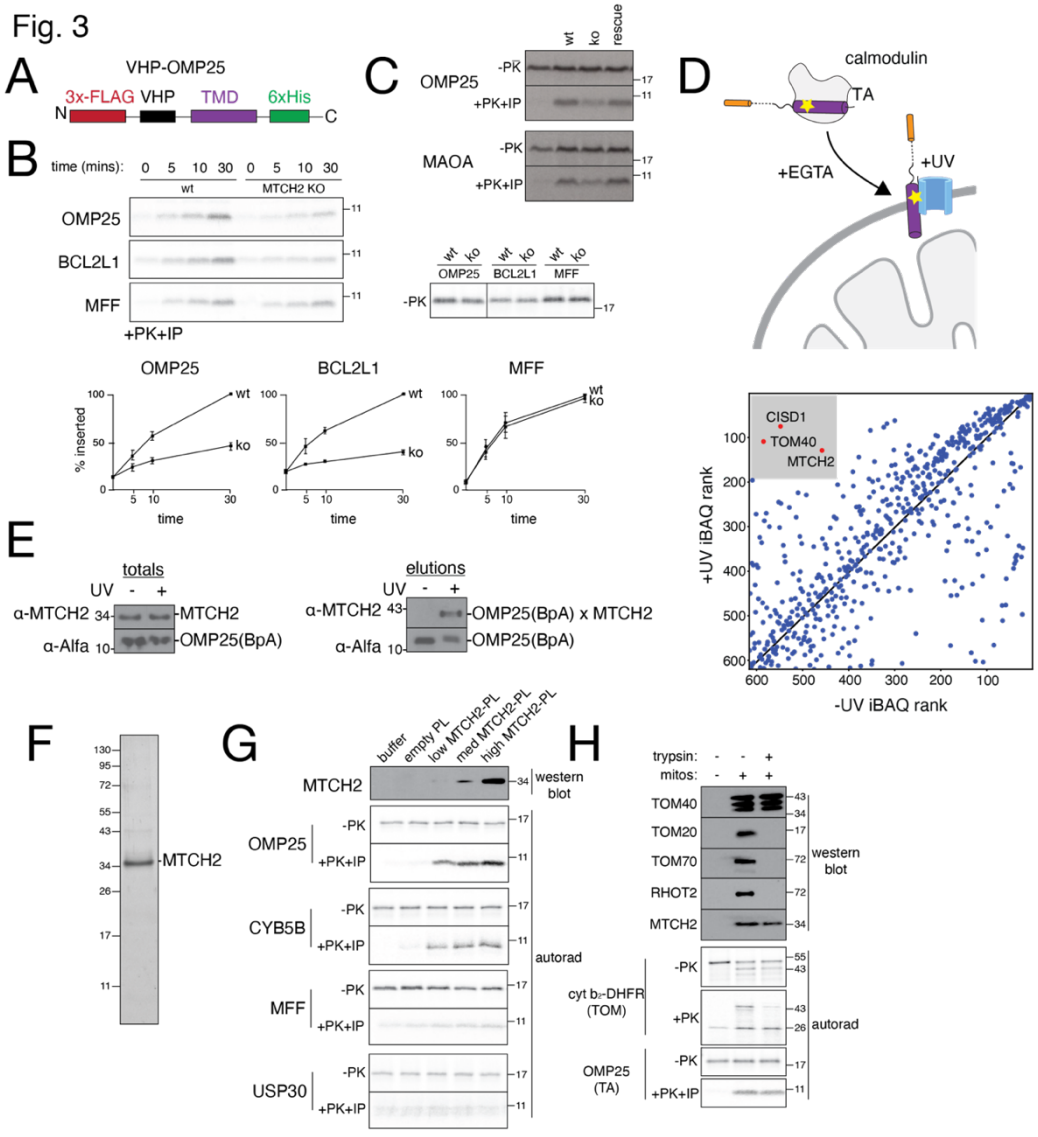


Figure 4.3. MTCH2 inserts diverse mitochondrial TAs into the outer membrane. (A) Schematic of the fusion between an inert N-terminal globular protein (VHP) and the TMDs of a panel of mitochondrial TAs (see also fig. S4.9) generated to probe TMD dependent insertion by MTCH2. (B) The indicated 35S-methionine labelled TA proteins were analyzed for in vitro insertion over time into mitochondria isolated from wild type (wt) or MTCH2 knockout (ko) K562 cells. Displayed are the samples prior to addition of protease (-PK; top right) and the protease protected fragment that has been affinity purified via a 6xHIS tag on the C-terminus of each substrate (+PK+IP; top left), ensuring insertion in the

correct topology. (Bottom) Quantification of three biological replicates are plotted with error bars indicating one standard deviation at each time point. (C) As in (B) comparing insertion of the indicated TA proteins into wild type, MTCH2 ko, and MTCH2 ko + MTCH2 rescue mitochondria. (D) (Top) Schematic showing the photocrosslinking strategy. OMP25 containing the photoactivatable amino acid BpA within its TMD was expressed and purified from *E. coli* as a complex with calmodulin. OMP25^{BpA} was released from calmodulin by addition of EGTA in the presence of mitochondria purified from K562 cells using a percoll gradient (fig. S4.1B). Crosslinking was activated by UV-irradiation, and the resulting crosslinked species were affinity purified via the Alfa-tag on the N-terminus of OMP25^{BpA} for identification by mass spectrometry. (Bottom) All proteins identified by mass spectrometry were ranked by iBAQ abundance, and those specifically enriched in the UV compared to the -UV control are highlighted. Though TOM40 and CISD1 were identified, they were not significant hits in our screen (fig. S4.12), while TOM40 was not required for biogenesis both in vitro (Fig. 4.1A) and in cells (fig. S4.12B) (E) As in (D) with the resulting elution analyzed by immunoblotting to assess levels of crosslinked OMP25 ^{BpA}-MTCH2. (F) MTCH2 was expressed and purified from human cells and analyzed by SDS-PAGE and Sypro-Ruby staining. (G) Following reconstitution (see fig. S4.13 for optimization of conditions), the recovered proteoliposomes were analyzed by immunoblotting for incorporation of MTCH2. Using a protease protection assay, the indicated MTCH2 dependent (OMP25, CYB5B) and independent (MFF, USP30) ^{35}S methionine labelled substrates synthesized in rabbit reticulocyte lysate were tested for insertion into liposomes reconstituted with increasing amounts of purified MTCH2 compared to an empty control. The resulting protease protected fragments were immunoprecipitated, imaged by autoradiography (autorad). (H) Mitochondria from wt K562 cells were treated with trypsin and their ability to insert TOM (Su9-DHFR) or TA substrates (OMP25) was assayed by protease protection as in (A). The indicated outer membrane proteins were confirmed to be degraded in a trypsin-dependent manner by immunoblot, while MTCH2 remained largely intact.

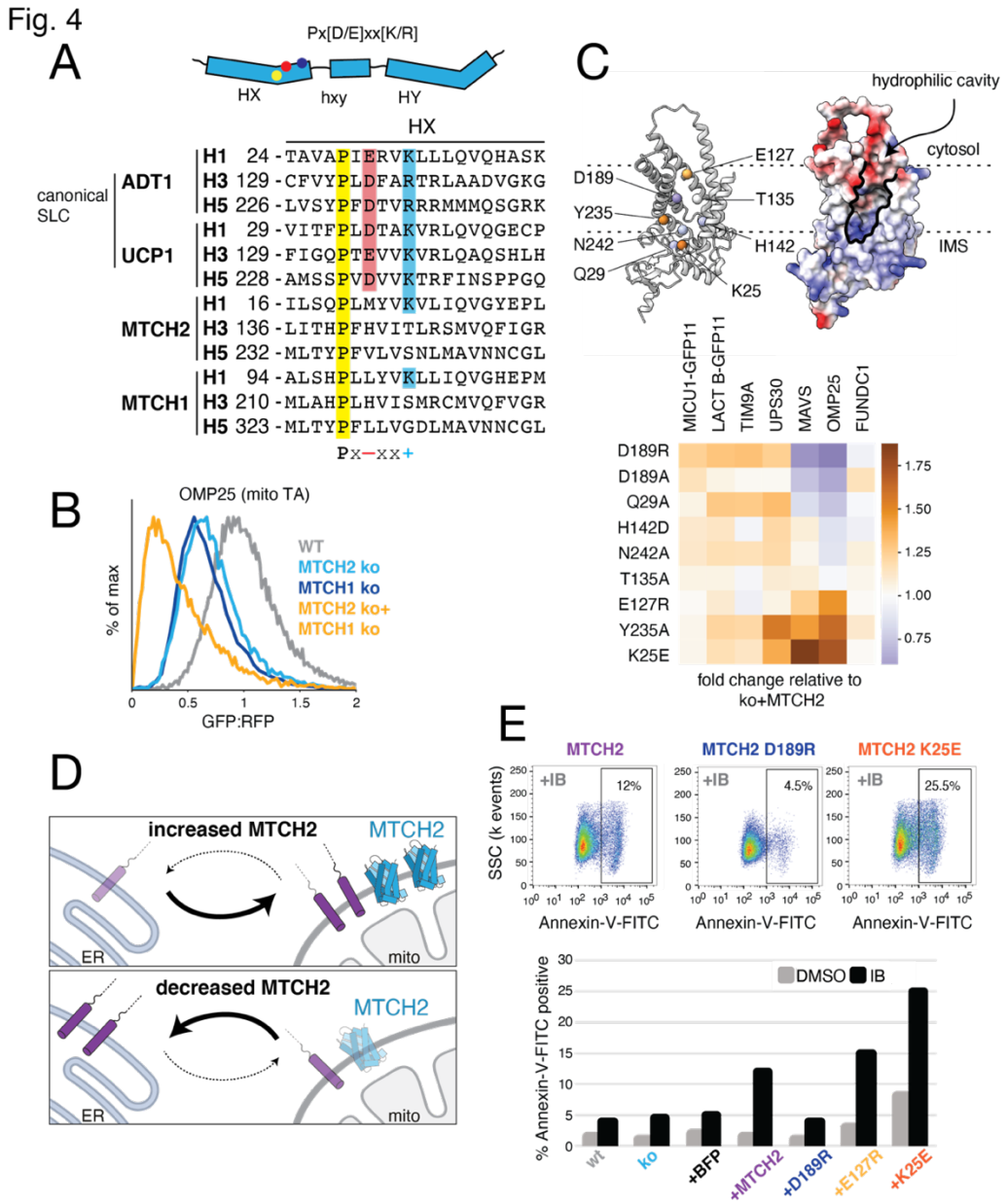


Figure 4.4. MTCH2 is a master regulator of outer membrane function. (A) (Top) SLC25 transporters are composed of three sets of two TMDs (six total). The location of the characteristic Px[D/E]xx[K/R] motif within a single SLC25 repeat is indicated. (Bottom) Sequence alignment of helices 1, 3, and 5 (with starting residues indicated) from two canonical inner membrane SLC25 transporters (ADT1, UCP1) and two diverged outer membrane SLC25 transporters (MTCH1, MTCH2), with residues from the Px[D/E]xx[K/R] motif highlighted. (B) Flow cytometry analysis of OMP25-GFP11

integration into the outer membrane using the reporter assay described in Fig. 4.1B. MTCH1 was depleted by transient knockout in either wild type (wt) or MTCH2 knock out (ko) cell lines. (C) (Top) AlphaFold2 predicted model of MTCH2 highlighting conserved polar and charged residues within the bilayer colored based on their effects on OMP25 shown below. (Bottom) using the reporter strategy shown in Fig. 4.1B, the indicated MTCH2 mutants, which alter the electrostatic potential of its TMDs, were tested for their effect on the indicated reporters (fig. S4.20). Depicted is a heat map summarizing the stimulation of each mutant relative to wild type MTCH2 on biogenesis of MTCH2 independent (MICU1, LACTB1, TIM9A, USP30) and dependent (MAVS, OMP25, FUNDC1) substrates. (D) Cell lines expressing GFP1-10 in the ER lumen were used to monitor mislocalization to the ER of mitochondrial TAs fused to a C-terminal GFP11. Table summarizing the analysis when either MTCH2 is depleted or overexpressed (data in fig. S4.20A, fig. S4.21, and fig. S4.22). (E) K562 cells expressing varying levels of MTCH2 or inactive (D189R) or hyperactive MTCH2 mutants (E127R or K25E; Fig. 4.4C) were treated with the chemotherapeutic imatinib mesylate (IB; 1 μM) or carrier (DMSO) for 72 hours. Apoptosis was assessed by staining with Annexin-V-FITC and analyzed by flow cytometry. Shown are representative dot plots displaying the fraction of apoptotic cells upon IB treatment expressing wt MTCH2 compared to in inactive (D189R) or hyperactive mutant (K25E) (Top) as well as a summary table for all MTCH2 constructs in IB vs carrier treated control.

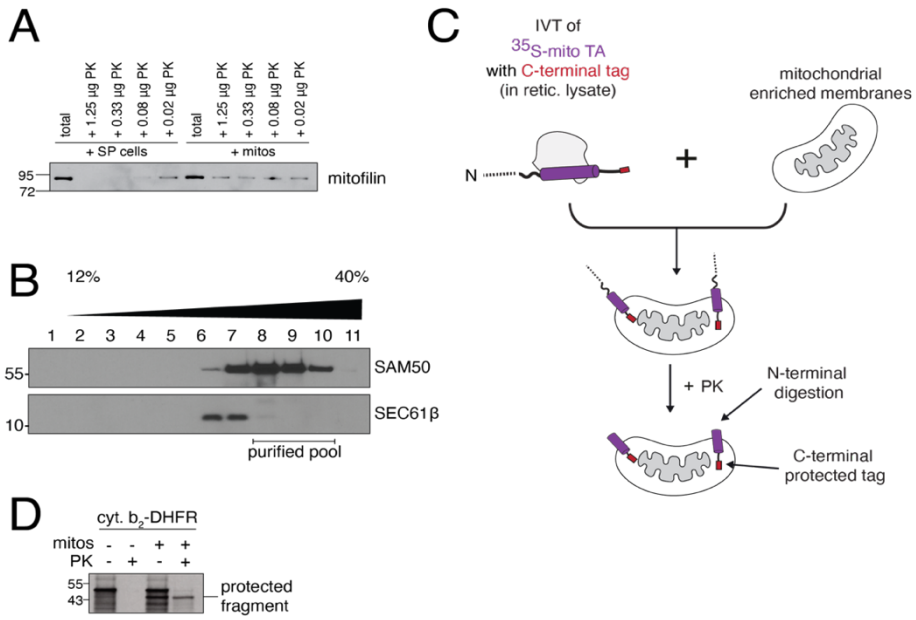


Figure S4.1. In vitro protease protection assays into human mitochondria. (A) The integrity of the outer membrane was tested using both semi-permeabilized K562 cells and isolated mitochondria by treating each with decreasing amounts of proteinase K (PK). The resulting reactions were analyzed by immunoblot for the mitochondrial intermembrane space (IMS) protein mitofilin, which should be inaccessible to PK when the outer membrane is intact. (B) Further purification of mitochondria from K562 cells using a percoll gradient and density centrifugation. Blots for a mitochondrial marker (SAM50) and an ER marker (SEC61β) demonstrate relative proportion of mitochondria and ER in the differential-centrifugation isolated mitochondria that are used for most in vitro experiments. Proteomics and crosslinking-mass spectrometry experiments were performed with percoll-gradient enriched mitochondria and used fractions 8-10 (Fig. 4.2A, Fig. 4.3D). (C) Schematic of the in vitro insertion assay, using protease protection of a C-terminal appended tag as a readout for insertion. In all insertion assays, substrates were translated using an in vitro translation reaction (IVT) in rabbit reticulocyte lysate supplemented with 35S-methionine, permitting detection by autoradiography. Translation was terminated by addition of puromycin, followed by incubation with isolated mitochondria derived from human K562 cells. After incubating with mitochondria, reactions were treated with PK. For TA proteins, the resulting protease protected band was

immunoprecipitated via a tag on its C-terminus, ensuring insertion into the outer membrane in the correct orientation, and analyzed by SDS-PAGE and autoradiography. (D) The IMS-localized TOM substrate composed of the cytochrome b targeting sequence fused to the inert protein DHFR (cyt. b2-DHFR) was translated and incubated with isolated mitochondria to confirm their activity. Insertion was detected by protease protection as described.

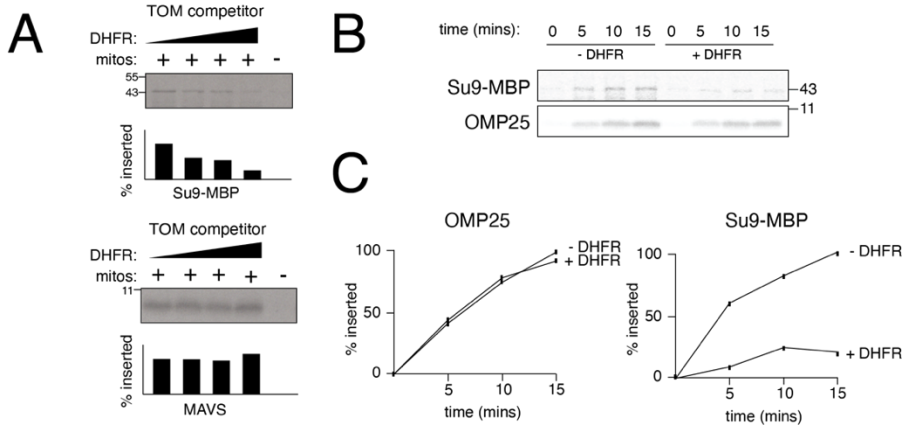


Figure S4.2. Insertion of mitochondrial TA proteins is not strictly dependent on the TOM40 translocase. (A) Competition assay as described in Fig. 4.1A with the mitochondrial TA protein, MAVS. Here, a fixed amount of mitochondria is incubated with increasing concentrations of purified, unlabeled Su9-DHFR (SU9-fused to the insert sequence DHFR; referred to as DHFR in the figure), a canonical TOM40 dependent substrate, and radiolabeled Su9-MBP or MAVS. As shown, there is a Su9-DHFR concentration dependent decrease in Su9-MBP insertion, assessed using the proteasome protection assay described in fig. S4.1C. This indicates that Su9-MBP relies on the same TOM40 mediated pathway for its insertion as Su9-DHFR. By contrast, MAVS, a mitochondrial outer membrane TA protein, inserts similarly into mitochondria independent of how much Su9-DHFR is present in the insertion reaction. This suggests that MAVS can utilize a TOM40 independent route for insertion into the outer membrane. (B) As in Fig. 4.1A, ^{35}S-methionine labelled TOM substrate Su9-MBP or the mitochondrial TA protein OMP25 were translated in rabbit reticulocyte lysate, released from the ribosome using puromycin and incubated with purified mitochondria either in the absence or presence of the recombinant TOM competitor Su9-DHFR (at a fixed concentration). Mitochondrial insertion was assessed at five minute intervals by protease protection, SDS-PAGE and autoradiography. (C) Time course of insertion from (B)

quantified. Error bars represent one standard deviation from three independent biological replicates.

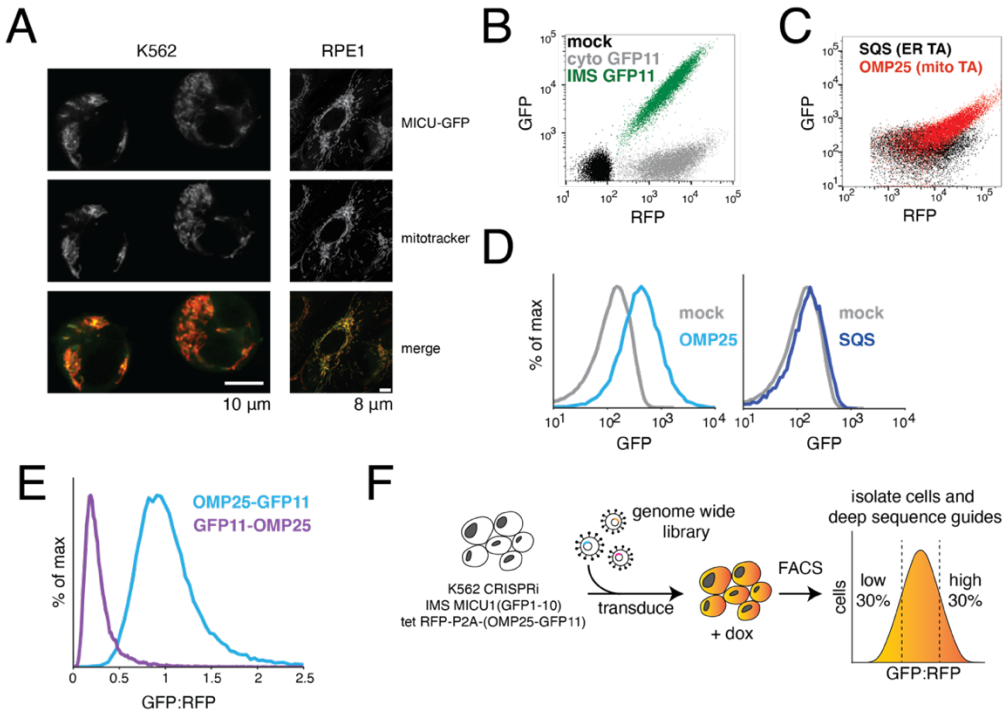


Figure S4.3. A CRISPRi screening platform to identify factors required for mitochondrial TA biogenesis in human cells. (A) Microscopy showing that the IMS-targeting sequence from MICU1 conjugated to full length GFP results in its localization to the mitochondria in mammalian K562 and RPE1 CRISPRi cells. (B) Either a cytosolic GFP11 or IMS targeted MICU-GFP11 were inserted in the backbone containing a translational control (RFP) separated by a viral 2A sequence (see Fig. 4.1B). These were independently expressed in K562 CRISPRi cells stably expressing MICU-GFP1-10 and analyzed by flow cytometry. Cytosolic GFP11 negligibly conjugated with IMS targeted GFP1-10, suggesting its correct localization to the IMS. (C) The endogenous sequences of two TA proteins: SQS, which is localized to the ER, and OMP25, which under these conditions is dual-localized to both the outer membrane and ER, were appended to a C-terminal GFP11 in the expression cassette described in Figure 4.1B. These constructs were independently introduced into cells expressing IMS localized GFP1-10 and analyzed by flow cytometry. (D) Histograms of (B) comparing GFP fluorescence for OMP25 and SQS compared to a mock transduced control. The marked increase in GFP fluorescence suggest that OMP25, but not SQS can successfully conjugate with GFP1-10 localized to the IMS.

(E) OMP25 conjugated to either an N- or C-terminal GFP11 were expressed in the expression cassette described in Fig. 4.1B in cells expressing IMS localized GFP1-10 and analyzed by flow cytometry. The majority of OMP25 is correctly integrated in an N-cytosolic, C-IMS orientation. (F) Workflow of the FACS-based CRISPRi screen. A K562 CRISPRi reporter cell line was constructed that constitutively expressed GFP1-10 in the IMS and the OMP25-GFP11 reporter under an inducible promoter. For the screen, these cells were transduced with a genome-scale CRISPR interference (CRISPRi) sgRNA library and then the OMP25-GFP11 reporter was induced with doxycycline for 24 hours prior to cell sorting. Cells were sorted based on ratiometric changes in GFP relative to RFP, and sgRNAs expressed in the isolated cells were identified using deep sequencing.

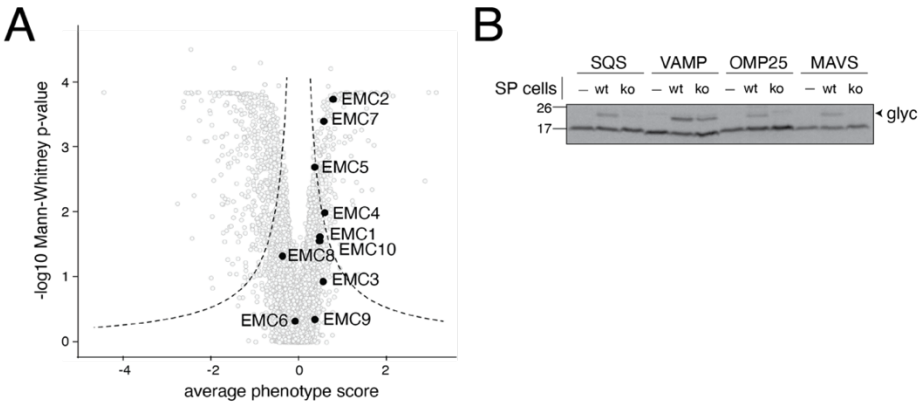


Figure S4.4. The ER membrane protein complex (EMC) is required for insertion of mislocalized mitochondrial TAs to the ER. (A) Volcano plot of the genome-wide CRISPRi screen with EMC subunits shown in black. (B) A panel of ER (SQS and VAMP) and mitochondrial (OMP25 and MAVS) TA proteins were conjugated to a C-terminal opsin epitope. The opsin epitope contains a consensus glycosylation sequence that is modified upon insertion into the ER lumen. These constructs were then translated in rabbit reticulocyte lysate in the presence of ^{35}S-methionine. The reactions were puromycin treated and incubated with either wild-type (wt) or EMC knockout (ko) semi-permeabilized (SP) cells. Insertion into the ER, as monitored by appearance of a glycosylated band ('glyc'), was dependent on the EMC for its canonical substrate SQS, and both mitochondrial TAs. In contrast, VAMP's insertion was unaffected by EMC knockout, consistent with its previously reported dependence on the GET pathway for insertion (Guna et al., 2018).

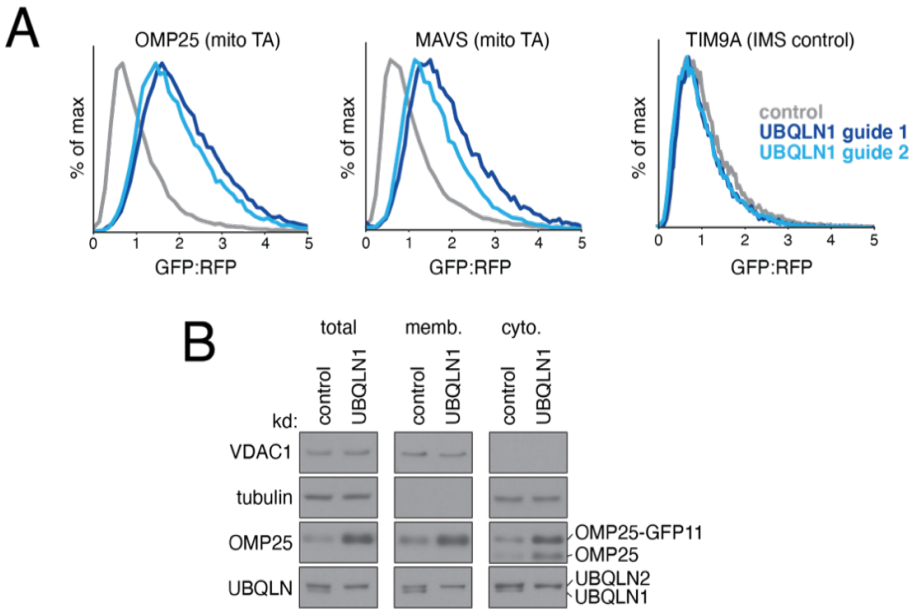


Figure S4.5. The UBQLNs are quality control factors for mitochondrial TAs. (A) K562 CRISPRi cells expressing IMS GFP1-10 were depleted of UBQLN1 using two different sgRNAs. Reporters were introduced for either two mitochondrial TAs (OMP25 and MAVS) or an IMS localized control (TIM9A) and cells were analyzed by flow cytometry. Lack of UBQLN1 results in a ratiometric increase in GFP:RFP fluorescence for mitochondrial TAs, consistent with its previously reported role in targeting mislocalized mitochondrial TAs for degradation by the ubiquitin proteasome pathway (Itakura et al., 2016). (B) K562 CRISPRi cells expressing IMS GFP1-10 and OMP25-GFP11 under an inducible promoter were depleted of UBQLN1 using sgRNA (above, guide 1). After doxycycline induced expression of OMP25-GFP11, cells were lysed and fractionated into cytosolic and membrane fractions using centrifugation. Levels of the indicated proteins were then assessed by SDS-PAGE and western blotting. In the absence of UBQLN1, more endogenous and GFP11 tagged OMP25 were found in both the cytosolic and membrane fractions.

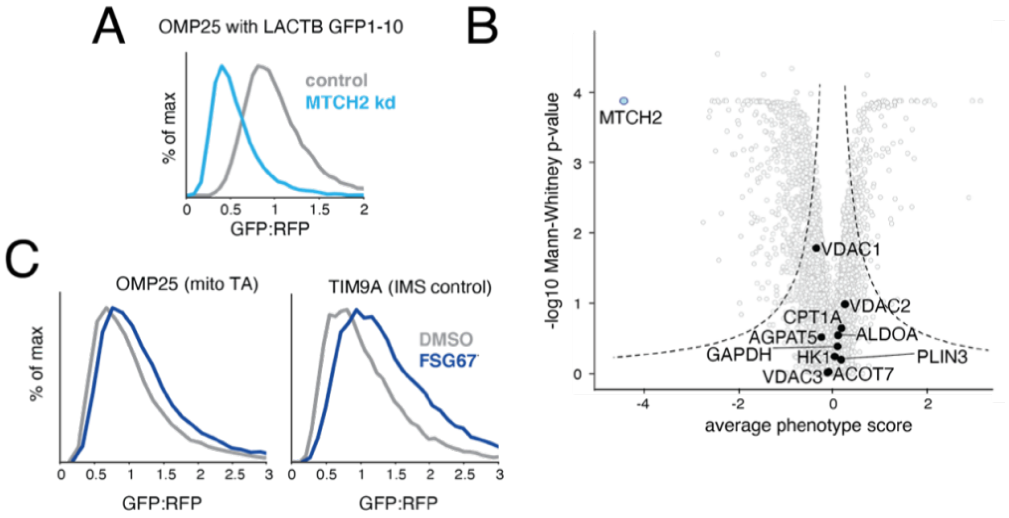


Figure S4.6. Assessing the effects of lipid biogenesis defects on mitochondrial TAs. (A) Flow cytometry analysis as in Fig. 4.1D but with an alternative IMS targeting sequence (Hung et al., 2014) derived from LACT appended to the GFP1-10. (B) Volcano plot of the genome-wide CRISPRi screen indicating MTCH2 (in light blue) and factors previously implicated as mediators of MTCH2-dependent outer membrane fatty acid synthesis or transport (in black). (C) K562 IMS GFP1-10 expressing cells were treated with the pan Glycerol 3-phosphate acyltransferase (GPAT) inhibitor FSG67 for 16 hours (75 μM) or a vehicle. MTCH2-dependent mitochondrial fusion has been shown to require GPAT catalyzed LPA synthesis. A reporter expressing either a mitochondrial TA (OMP25) or an IMS localized control (TIM9A) were expressed in GPATi cells and analyzed by flow cytometry.

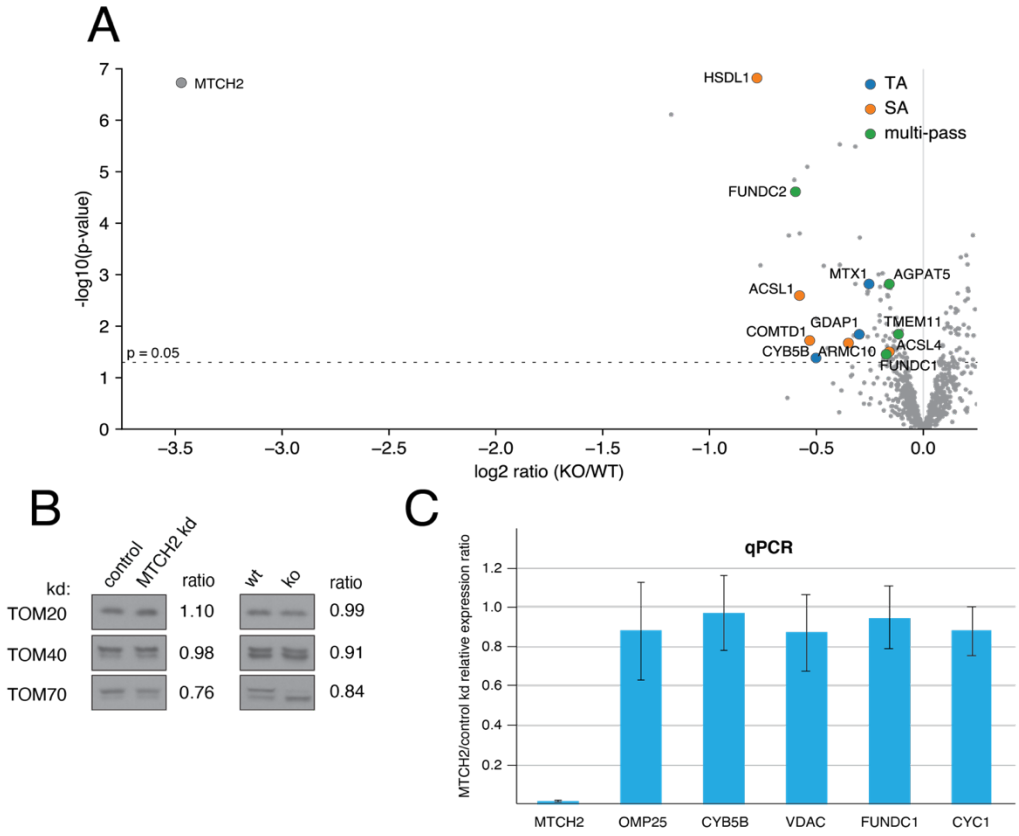


Figure S4.7. Analysis of proteomic and transcriptomic changes to mitochondrial outer membrane proteins in MTCH2 depleted cells (A) Label-free mass spectrometry analysis as in Fig. 4.2A of crude mitochondria isolated from K562 cells via differential centrifugation with MTCH2 knockout (ko) versus wild type (wt) cells. Highlighted are mitochondrial proteins that across four biological replicates were statistically altered by loss of MTCH2. (B) Western blotting of TOM complex subunits from MTCH2 depleted (kd) and non-targeting control cells (left), and wild type and MTCH2 knock out (ko) cells (right) as in Fig. 4.2B. Quantification of fold-change in depleted vs control cells is displayed as determined using a dilution series for each antibody. (C) qPCR of mitochondrial genes in MTCH2 versus control knock down in K562 cells (normalized to the housekeeping gene HRRT). See methods for specific details.

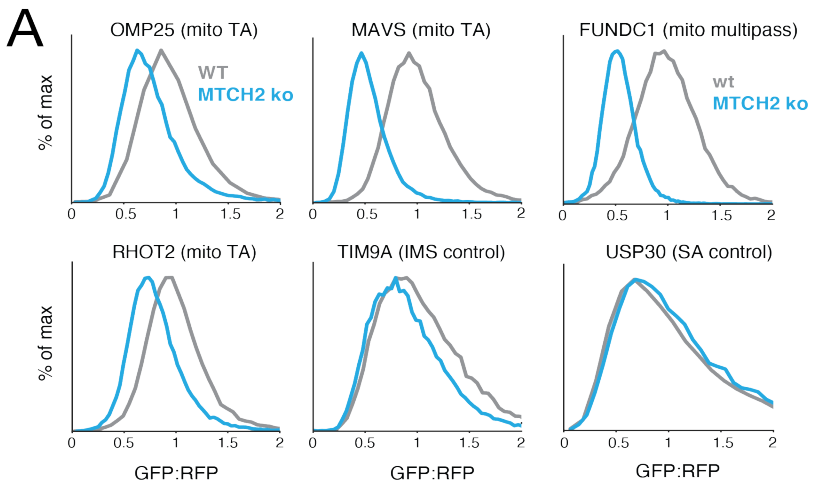


Figure S4.8. Analysis of TA proteins for MTCH2 dependent biogenesis in knockout cells. As in Fig. 4.2C but in wild type compared to MTCH2 knockout cells. Histograms summarizing flow cytometry analysis of the integration of the indicated mitochondrial proteins.

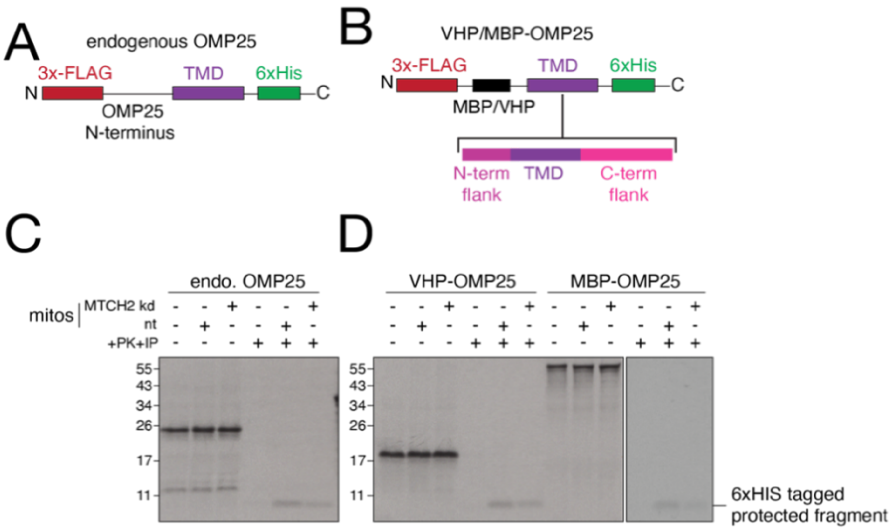


Figure S4.9. Establishing an in vitro system to test differential dependence on MTCH2 on insertion of a panel of mitochondrial TAs. (A, B) Schematic of OMP25 constructs used to test whether the dependence on MTCH2 for insertion that we observed with the full-length endogenous OMP25 (A) could be recapitulated with an artificial N-terminus (B). (C) Using the protease protection assay described in fig. S4.1, we compared insertion of the endogenous OMP25 containing a C-terminal 6xHIS tag into mitochondria isolated from K562 cells expressing either a non-targeting (nt) or MTCH2 sgRNA (kd). We observed that loss of MTCH2 specifically decreased the levels of a protease protected fragment consistent in size with the TMD and C-terminus of OMP25 (compare lanes 5 vs 6). In the absence of mitochondria, no protected fragment is observed (lane 4). Because this protease protected fragment could be immunoprecipitated using a C-terminal 6xHIS tag, we verified that OMP25 was inserted in the correct orientation, with its C-terminus in the IMS and its N-terminus facing the cytosol. (D) As in (A) but using a fusion of the OMP25 TMD and its flanking residues to both the unrelated globular proteins MBP and VHP. We concluded that the OMP25 TMD alone is sufficient to confer MTCH2 dependent insertion on both of the tested fusion proteins (compare lanes 5 vs 6 and 11 vs 12). Because the VHP-fusions were translated more efficiently, we generated a panel of mitochondrial TAs using the depicted VHP N-terminus as shown in (B). These were used for experiments in Fig. 4.3A, fig. S4.10-11.

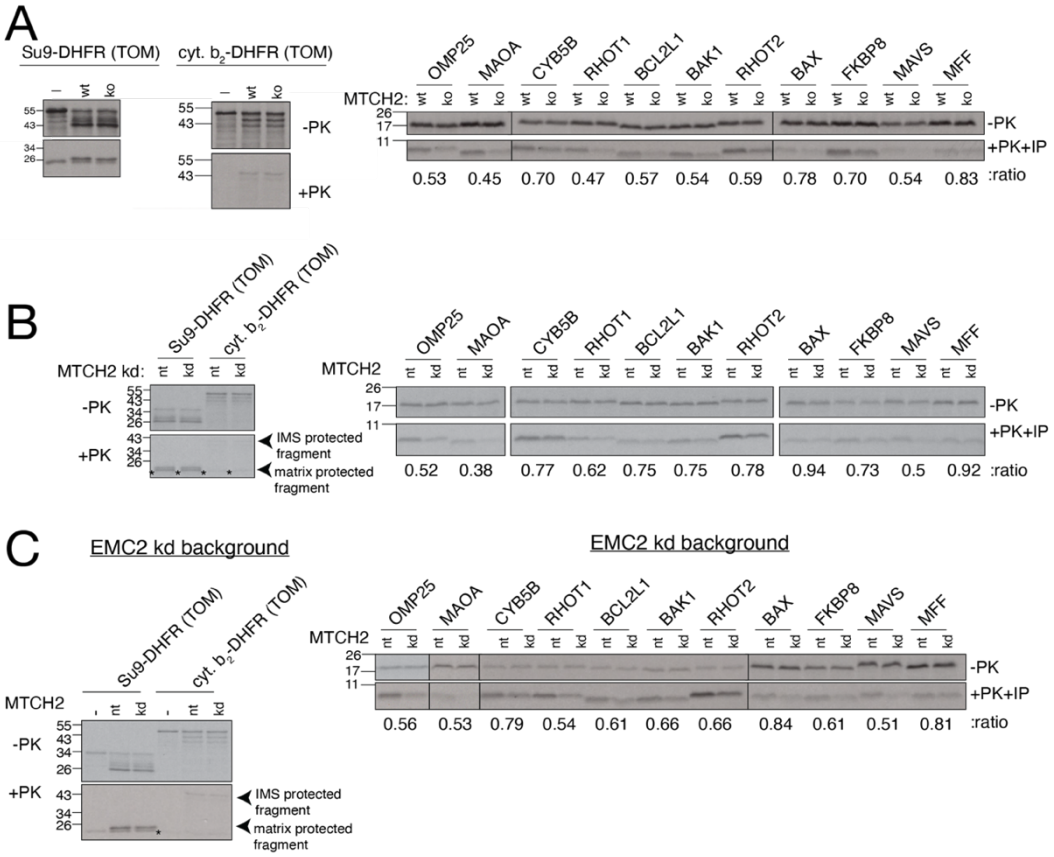


Figure S4.10 Insertion of mitochondrial TAs in vitro is affected by MTCH2 knockdown. (A) ^{35}S-methionine labelled TA proteins were analyzed for in vitro insertion into mitochondria isolated from wild type (wt) or MTCH2 knockout (ko) K562 cells. Here, the indicated TMDs and flanking residues were fused to VHP as described in fig. S4.9. Displayed are the samples prior to addition of protease (-PK) and the protease protected fragment that has been affinity purified via a 6xHIS tag on the C-terminus of each substrate (PK+IP), ensuring insertion in the correct topology. Canonical TOM substrate controls that are targeted to either the matrix (Su9-DHFR) or IMS (cyt. b2-DHFR) were tested in parallel. Relative insertion efficiency in MTCH2 depleted/wild type mitochondria has been quantitated and displayed for each substrate. (B) As in A except using mitochondria isolated from K562 CRISPRi cells expressing either a non-targeting (nt) or MTCH2 targeting (kd) sgRNA. A panel of mitochondrial TAs was tested in parallel with a matrix (Su9-DHFR) and IMS (cyt. b_2-DHFR) targeted control that rely on the TOM pathway,

which were unaffected. *Denotes the folded, protease resistant, DHFR domain that migrates immediately below the mature, matrix targeted control, and is visible in the absence of mitochondria. (C) Note that because some residual ER is present in the enriched mitochondrial membranes used in these insertion reactions (see fig. S4.1), for those substrates that are dual localized, we cannot formally differentiate between insertion into mitochondria vs ER in this assay alone. To address this, we have tested the complete TA panel in an EMC knockdown background, which we found eliminated mistargeting of mitochondrial TAs to the ER (fig. S4.4). Therefore, we perform insertion assays as in (A) except using mitochondria isolated from either EMC2 knock down or EMC2 and MTCH2 knock down K562 CRISPRi cells.

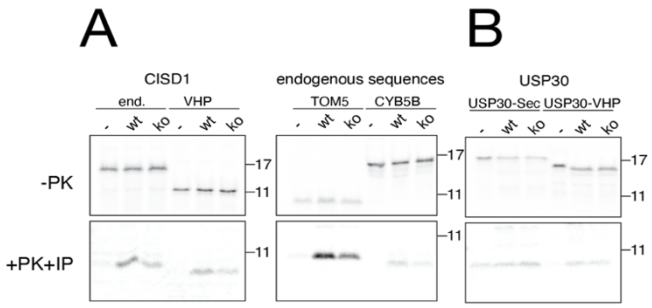


Figure S4.11. Insertion of a-helical outer membrane proteins display MTCH2 dependence in concordance with in vivo experiments. (A) The indicated ^{35}S-methionine labeled TA or signal anchored proteins were translated in rabbit reticulocyte lysate and following treatment with puromycin were incubated with wt or MTCH2 ko mitochondria. Successful insertion was determined by protease treatment and immunoprecipitation of the protected C-terminal 6xHIS tagged fragment. We find that insertion of the signal anchored protein CISD1, either when using the endogenous sequence or when its TMD is fused to the VHP cassette (described in fig. S4.9) is MTCH2 dependent. Similarly, endogenous TOM5 requires MTCH2 for insertion in vitro, consistent with our in vivo florescent reporter data (Fig. 4.2D). (B) To test insertion of the signal anchored protein USP30, we generated constructs in which the USP30 TMD and flanking residues was fused to an inert C-terminal linker along with two affinity tags on either termini (6xHis-USP30 TMD-sec/VHP-3xFLAG). We found that insertion of both of these USP30-fusion proteins is not affected by loss of MTCH2, consistent with our in vivo data (Fig. 4.1E). The USP30 TMD alone is sufficient to confer MTCH2 independent insertion, as insertion of fusions of the USP30 TMD to either the N-terminus of SEC61b or the inert globular protein VHP are both MTCH2 independent.

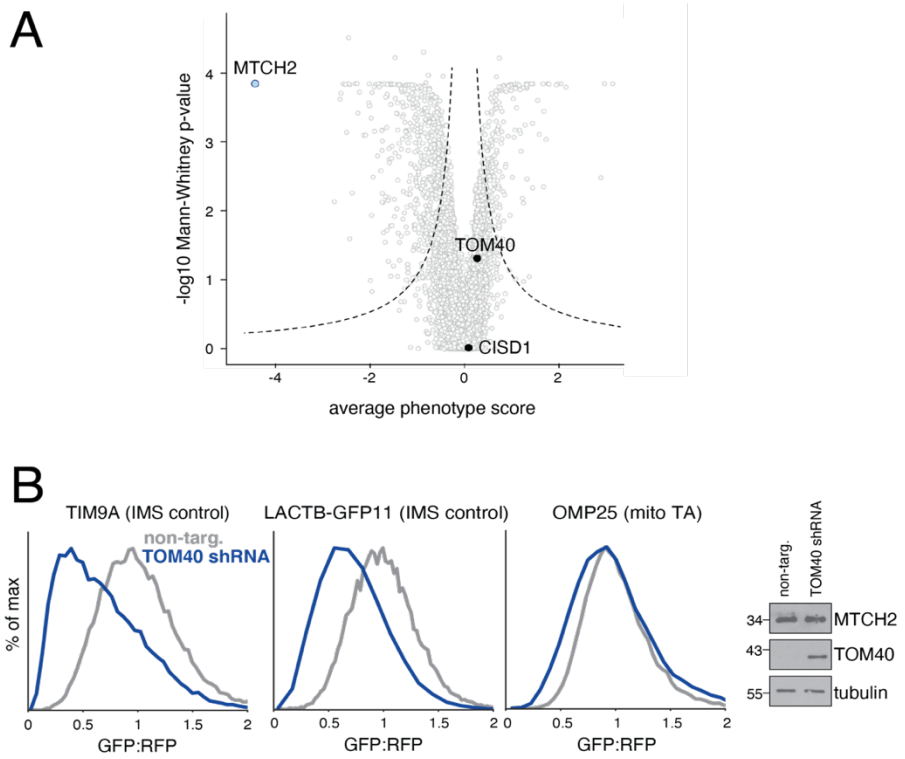


Figure S4.12. TOM40 and CISD1 are not strictly required for mitochondrial TA biogenesis. (A) Volcano plot of the genome-wide CRISPRi screen as in Fig. 4.1C indicating MTCH2 (in light blue) and the two other factors (TOM40 and CISD1, in black) that were enriched upon crosslinking OMP25 with purified mitochondria (Fig. 4.3D; see fig. S4.1C for mitochondrial purification). (B) Using our fluorescent reporter assay (Fig. 4.1B) we show that depletion of TOM40 using an shRNA (dark blue) compared to a scrambled control (grey) in wild type cells (wt) does not affect the biogenesis of the mitochondrial TA OMP25 under conditions where biogenesis of the canonical TOM substrates LACTB and TIM9A is diminished. These results are consistent with the TOM40 competition experiment (Fig. 4.1A, fig. S4.2) that shows that mitochondrial TA insertion in vitro is not affected by addition of a recombinant TOM40 substrate. Together these data suggest that mitochondrial TAs do not strictly require TOM40 for their biogenesis. However, we do not formally exclude a potential contribution of the TOM complex to either the recruitment or insertion of alpha-helical proteins into the outer membrane.

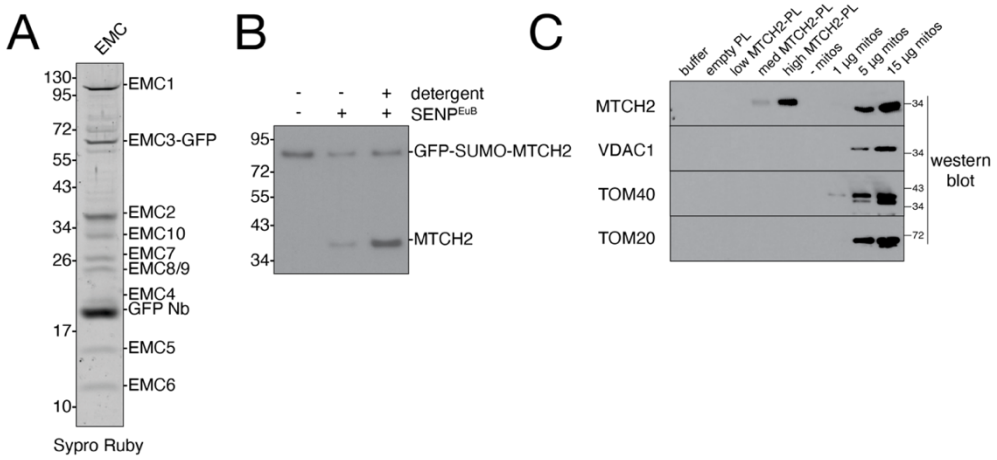


Figure S4.13. Establishing a reconstituted system to query MTCH2-dependent insertion into proteoliposomes. (A) The human EMC was expressed and purified as previously described using a GFP fused to the C-terminus of EMC3 in the detergent DBC (Pleiner et al., 2020) and visualized using Sypro Ruby staining. This was further used for generating EMC-proteoliposomes used in insertion reactions shown in fig. S4.14. (B) Following reconstitution of purified MTCH2 into proteoliposomes at lipid:protein ratios ranging from 6000:1 to 1000:1, we sought to determine the fraction that was correctly oriented relative to the cytosol (i.e. with the N-terminus facing the cytosol and the C-terminus within the lumen of the liposome). Because we found that MTCH2 is largely protease resistant (Fig. 4.3H), we could not use the classical strategy to determine its orientation after reconstitution. Therefore, we expressed and purified an N-terminally tagged GFP-SUMO-HA-MTCH2, such that the GFP tag and SUMO protease cleavage site would be exposed to the cytosol upon reconstitution in the correct orientation. Following reconstitution of GFP-SUMO-HA-MTCH2, intact proteoliposomes or matched controls that were solubilized by addition of 1% Triton, were treated with SUMO protease (SENP-EuB). Using this strategy, we determined that ~50% of GFP-SUMO-HA-MTCH2 is oriented correctly following reconstitution. (C) Analysis of the contents of reconstituted MTCH2 proteoliposomes as determined by western blotting for the indicated proteins. These results suggests that the TOM complex subunits, as well as the abundant outer membrane b-barrel protein, VDAC, are not present at appreciable amounts in the purified

proteoliposomes. Therefore, trace levels of the TOM complex are unlikely to be strictly required for any observed insertion activity of purified, reconstituted MTCH2.

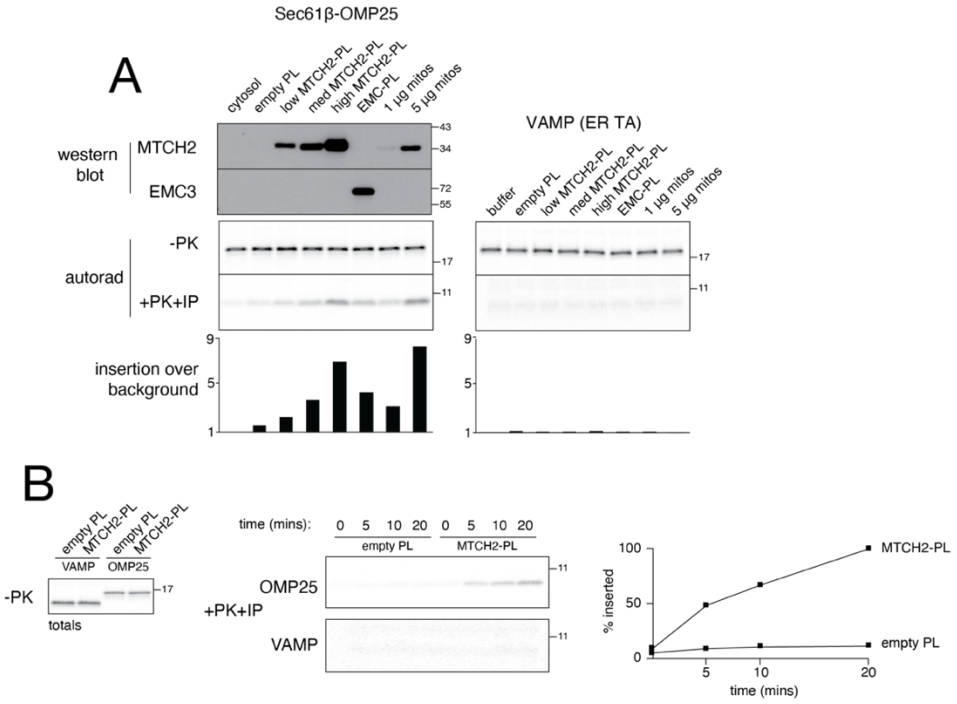


Figure S4.14. MTCH2 is sufficient to stimulate insertion of mitochondrial TAs into proteoliposomes at an efficiency similar to purified mitochondria and the EMC. (A) Following reconstitution, the recovered proteoliposomes were analyzed by western blotting for incorporation of the indicated proteins in comparison to purified mitochondria. Using a protease protection assay, Sec61β-OMP25 or the ER-TA, VAMP (a GET pathway substrate), synthesized in rabbit reticulocyte lysate was tested for insertion into liposomes reconstituted with increasing amounts of purified MTCH2, the human EMC, or an empty control. In parallel for comparison, we tested insertion into mitochondria isolated from K562s at two concentrations. The resulting protease protected fragments were immunoprecipitated, imaged by autoradiography (autorad), and their intensity quantified on a phosphoimager, and are displayed normalized to the cytosol only control. The presence of a small amount of protected fragment in cytosol alone may be indicative of chaperone binding. (B) As in (A), reconstituted proteoliposomes of MTCH2 or an empty control were incubated with OMP25 and VAMP for the indicated time points and the % inserted was quantitated using protease protection.

A

alignment		% ID
SLC25A46	MTCH1	16.5
SLC25A46	MTCH2	10.7
MTCH1	MTCH2	36.4

B

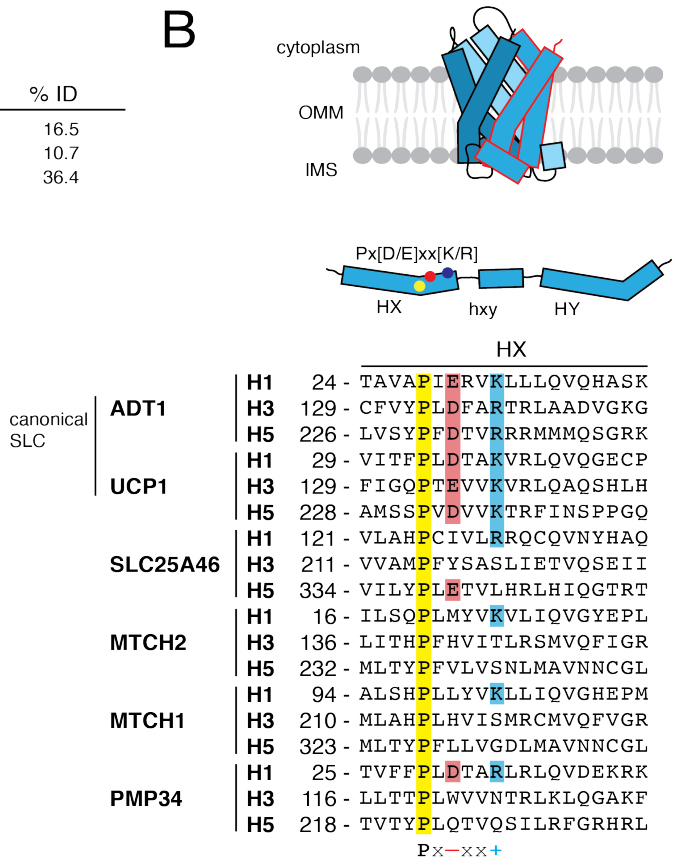


Figure S4.15. Shared features of outer membrane SLC25 transporters. (A) Sequence identity derived from pairwise alignment between SLC25A46, MTCH1, and MTCH2. (B) On top, a cartoon representation of SLC25 TMD arrangement showing the 3 SLC25 repeats in unique shades of blue, with a single repeat outlined in red. In the middle, a schematic showing the location of characteristic motifs within a single SLC25 repeat, which normally encodes 2 TM helices. On bottom, sequence alignment of all individual SLC25 repeats from 2 inner membrane SLC25 transporters (ADT1, UCP1) and 4 outer membrane SLC25 transporters, three mitochondrial (SLC25A46, MTCH1, MTCH2) and one peroxisomal (PMP34), with residues from the Px[D/E]xx[K/R] motif highlighted.

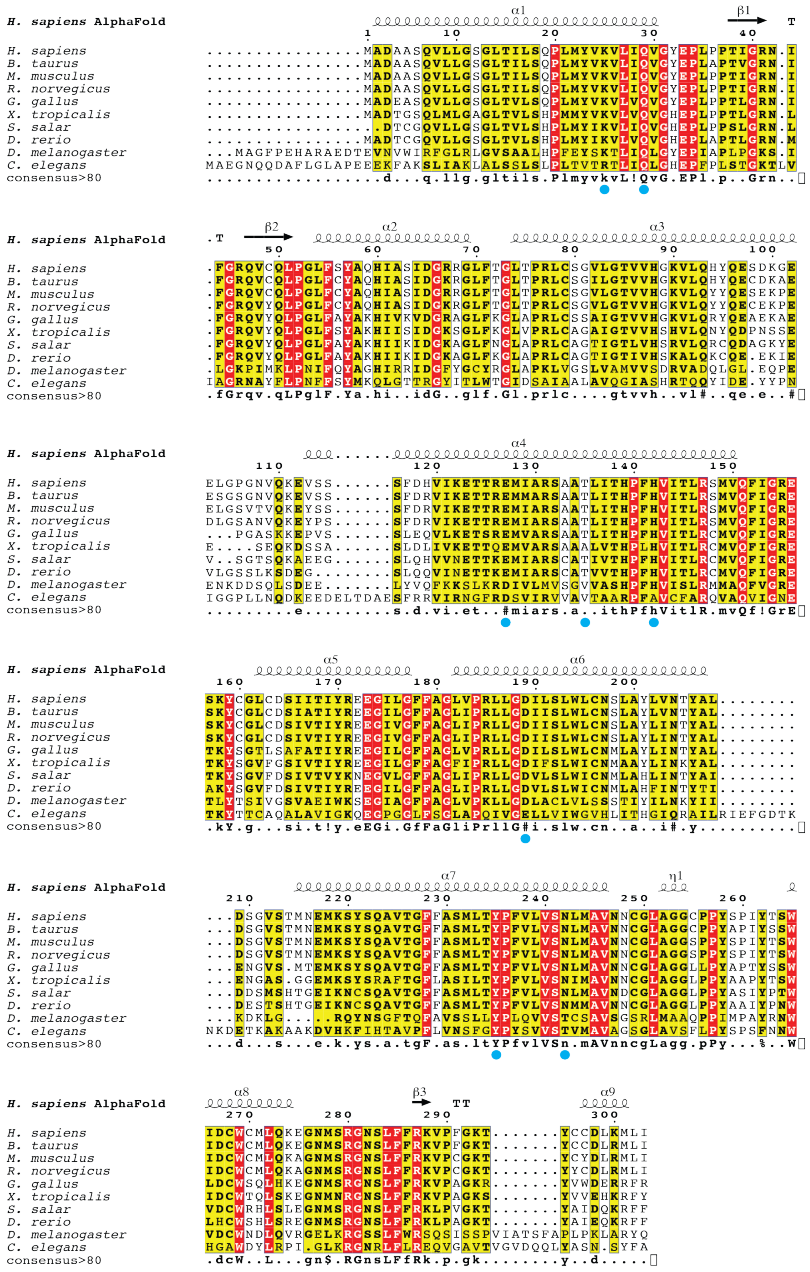


Figure S4.16. Multisequence alignment of homologs of MTCH2. MUSCLE (Edgar, 2004) generated sequence alignment of MTCH2 homologues from a variety of animal species, generated using the ESPRIPT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)). Red columns denote 100% sequence conservation, yellow columns denote

moderate conservations with similar residues bolded. Secondary structure derived from the human MTCH2 AlphaFold2-predicted model is displayed above the sequences.

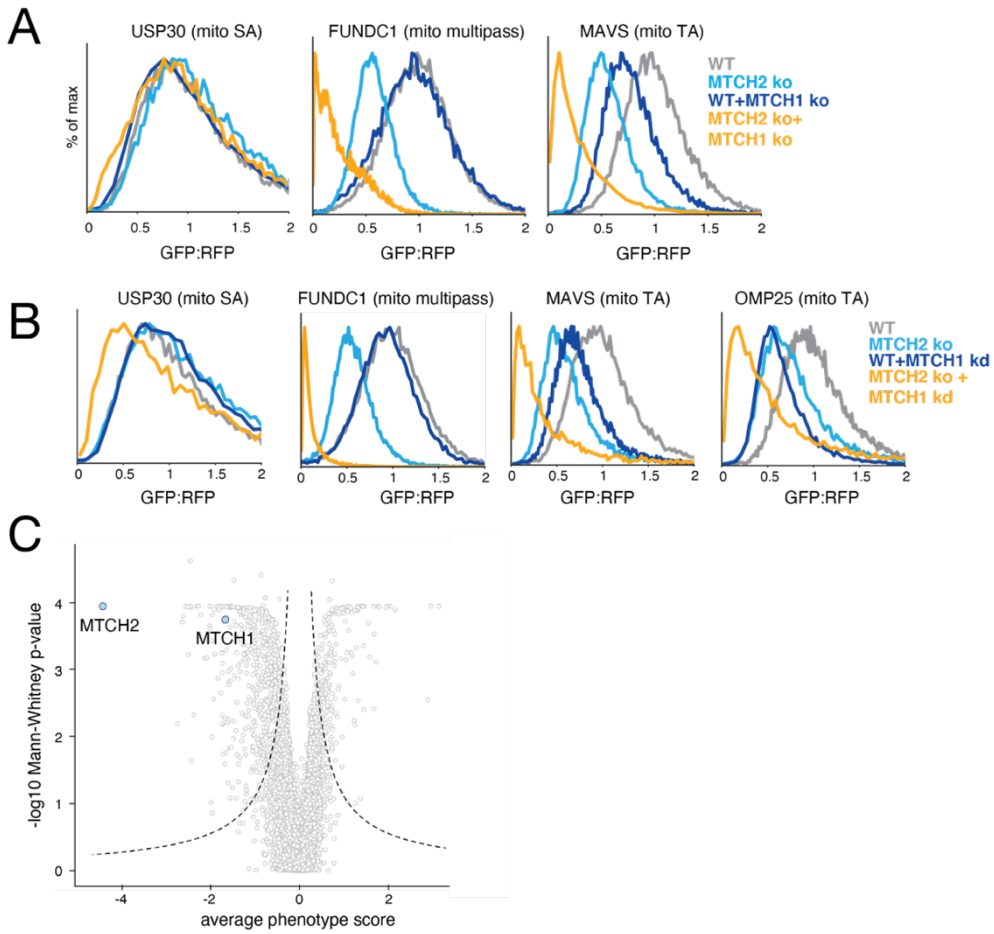


Figure S4.17. MTCH1 acts in a parallel, and partially redundant, pathway to mediate insertion of mitochondrial TAs. (A) As in Fig. 4.4B for a set of outer membrane reporters including a signal-anchored protein (USP30), a mitochondrial TA (MAVS) and a multipass protein (FUNDC1). (B) As in (A) but in either wild-type or MTCH2 knockout (ko) K562 CRISPRi cells expressing guides targeting MTCH1 for knock-down. (C) Volcano plot of the genome-wide CRISPRi screen with MTCH1 and MTCH2 highlighted.

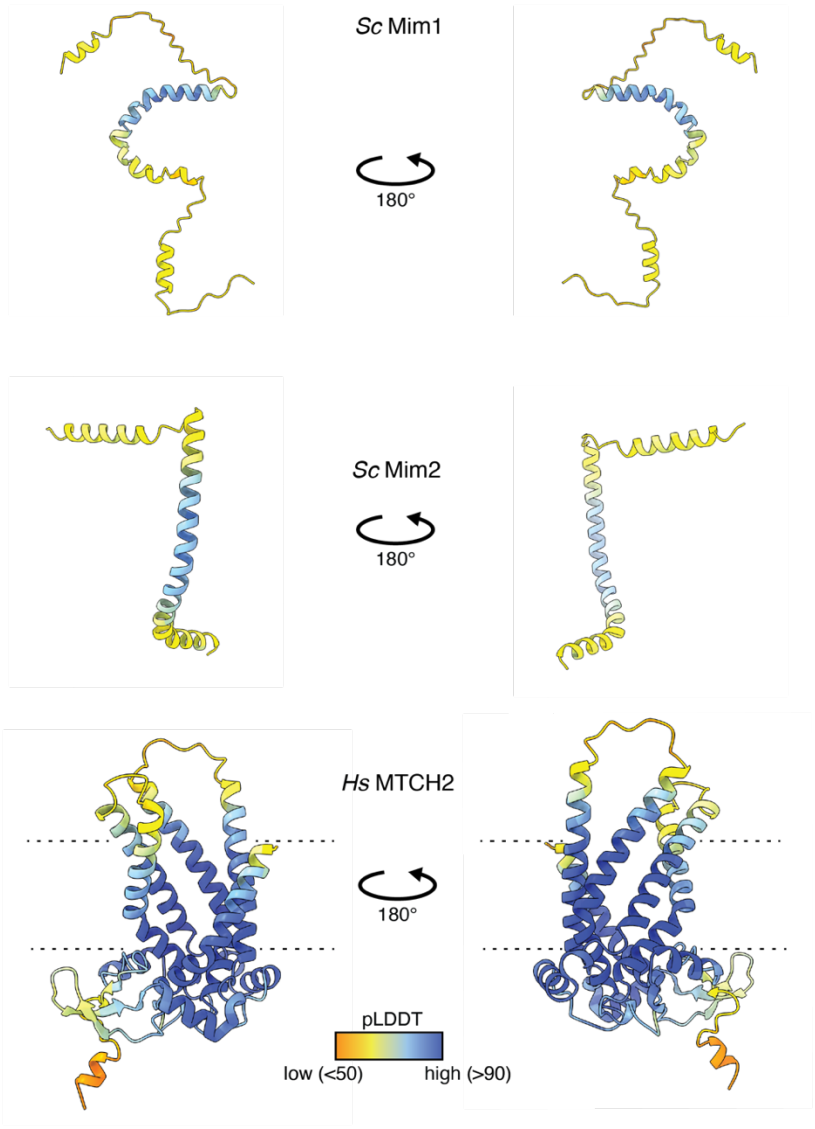


Figure S4.18. Comparison of the AlphaFold2 predicted structures of the *S. cerevisiae* Mim1 and 2 with human MTCH2. AlphaFold2 predicted models of the two subunits of the MIM complex from *S. cerevisiae*, Mim1 and Mim2. Neither bear significant resemblance to MTCH2 both at the sequence or topological level. Mim1 and Mim2 are 13 and 10 kDa proteins respectively, each containing a single transmembrane helix spanning the mitochondrial outer membrane. The MIM complex is estimated to be 150 kDa (via native PAGE analysis), presumably formed by oligimerization of multiple copies of each subunit (Dimmer et al., 2012). The MIM complex has been shown to be required for the insertion of many tail-anchored, signal-anchored, and multipass mitochondrial outer

membrane proteins in yeast (Doan et al., 2020; Becker et al., 2008; Waizenegger et al., 2005; Thornton et al., 2010; Hulett et al., 2008). The extensive prior characterization of the MIM complex, as well as the outer membrane insertase pATOM36 from trypanosomes (Vitali et al., 2018; Käser et al., 2016), is important for demonstrating the necessity of proteinaceous machinery in the insertion of transmembrane helices into the mitochondrial outer membrane. MTCH2 presumably evolved to perform a similar function via a completely unrelated protein fold, representing a striking case of convergent evolution.

Fig. S19

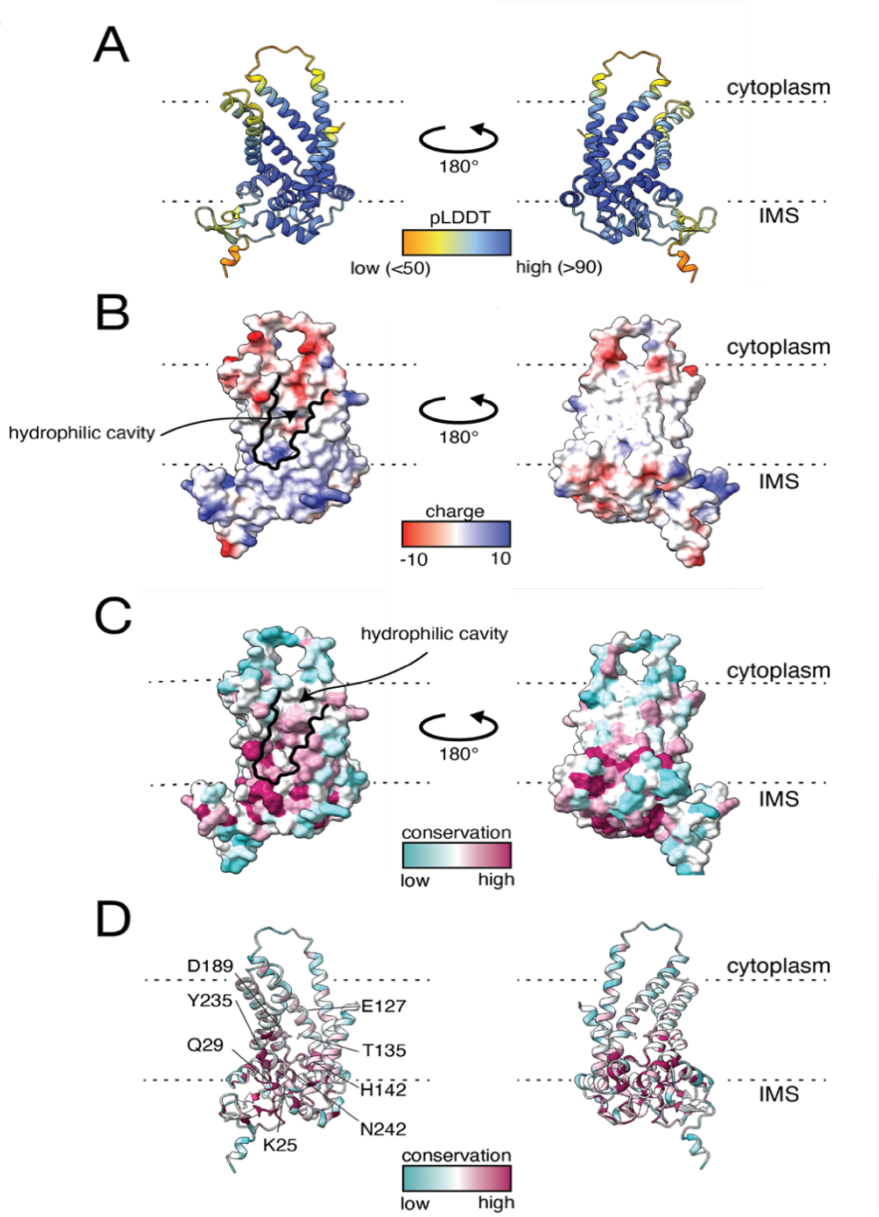


Figure S4.19. The AlphaFold2 model of human MTCH2 predicts the presence of a conserved hydrophilic groove within the membrane. (A) Two views of a cartoon rendering of the AlphaFold2 predicted MTCH2 model from the plane of the membrane, colored by model confidence (pLDDT). (B) As in (A) with surface representation of the predicted model, colored by coulombic electrostatic potential calculated by the ChimeraX structural visualization software package (Pettersen et al., 2021). (C) The same model

views shown in surface representation and colored by sequence conservation calculated from the same sequence alignment shown in fig. S4.16 in ChimeraX. (D) The same model views shown in cartoon representation, also colored by sequence conservation as above, and with residues used in mutational analysis shown in stick representation and labeled. Using these models enabled mutational analysis for a potential mechanism of MTCH2 mediated membrane protein insertion involving the conserved hydrophilic groove within the membrane. We reasoned that negatively charged residues lining the middle-to-upper portion of the hydrophilic cavity may be important for the docking of mitochondrial TAs which are enriched for flanking positive charges at their C-termini.

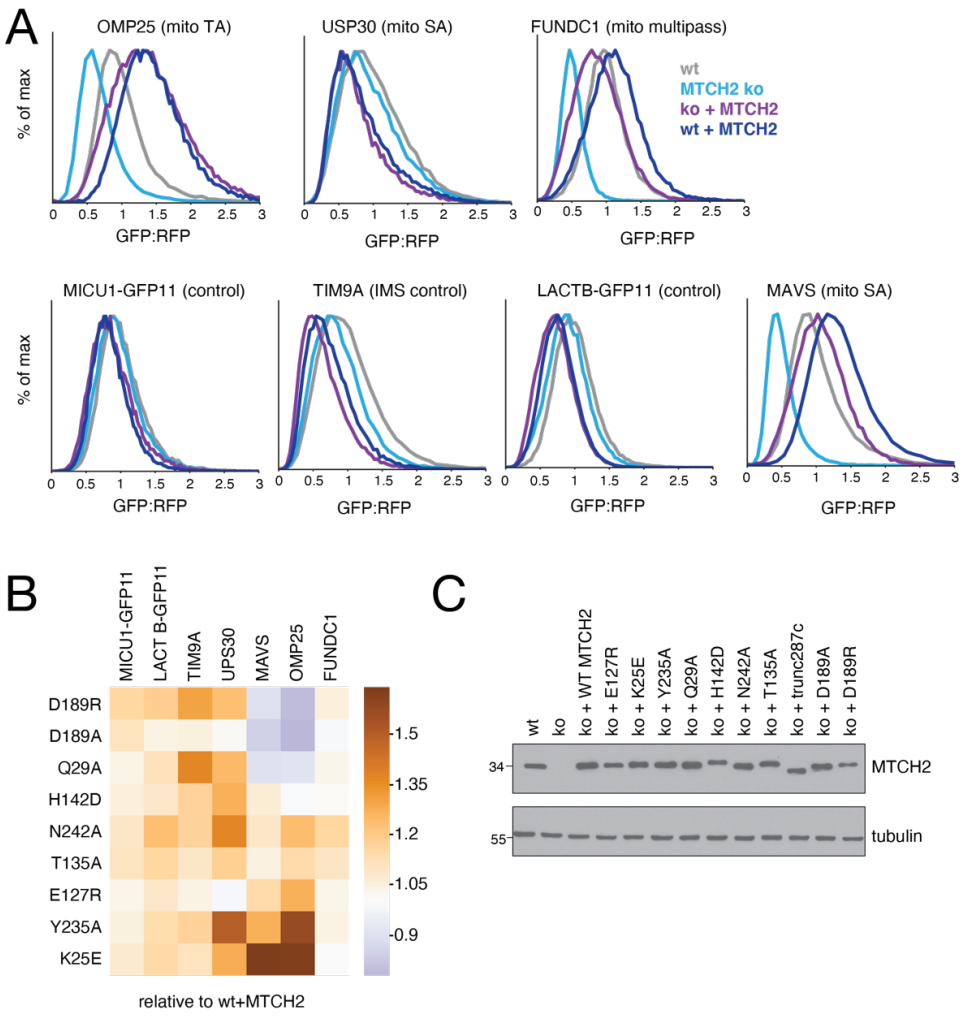


Figure S4.20. Point mutants to conserved polar and charged residues in MTCH2's TMDs affect integration of mitochondrial TAs. Using our fluorescent reporter assay (Fig. 4.1B), we tested the effects of expression of a series of MTCH2 mutants on the biogenesis of the indicated mitochondrial proteins. In (A) we depict the effects of ectopic expression of wild type MTCH2 in either wild type (wt) or MTCH2 knockout (ko). Because MTCH2 in this system is overexpressed compared to endogenous levels, it is sufficient to drive increased integration of MTCH2 dependent, but not independent, substrates into the mitochondrial outer membrane. (B) A heat map, summarizing the effects of expression of the indicated MTCH2 mutants on top of wild type cells, which are expressing a functional endogenous copy of MTCH2. Depicted is the fold-change in

insertion for the indicated substrates upon ectopic expression of each MTCH2 mutant compared to wild type MTCH2. Mutants were selected based on their conservation and localization within the hydrophilic groove of MTCH2 (Fig. 4.4C, fig. S4.18-19). A matched experiment performed in MTCH2 knockout cells is summarized in Fig. 4.4D. (C) Western blots showing the expression of wild-type MTCH2 and the indicated point mutants in a K562 MTCH2 ko background relative to wild-type cells. We note the differential effect of some mutants on TA substrates (MAVS, OMP25) compared to a multipass substrate (FUNDC1), which may indicate that FUNDC1 expression is itself regulated, and thus does not accumulate over endogenous levels. Alternatively, it may be that TAs and multipass proteins utilize different regions of MTCH2 for insertion, as was observed for the EMC, where certain mutations have differential effects on TA and multipass substrates (Miller-Vedam et al., 2020; Pleiner et al., 2020).

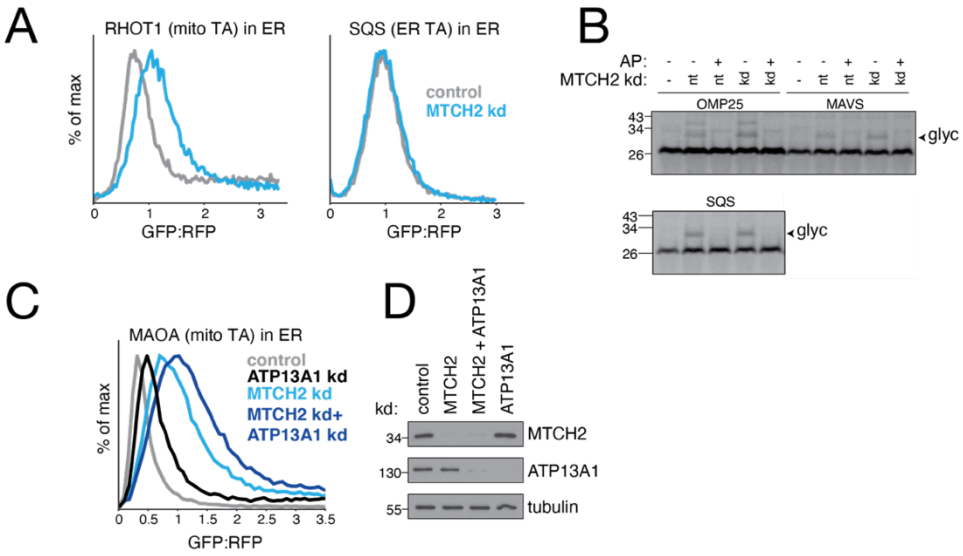


Figure S4.21. Depletion of MTCH2 causes increased mislocalization of mitochondrial TAs to the ER. (A) Cell lines expressing GFP1-10 in the ER lumen were used to monitor mislocalization to the ER of mitochondrial TAs fused to a C-terminal GFP11 using a similar strategy to that previously described (Inglis et al., 2020). Flow cytometry analysis of insertion into the ER of a mitochondrial TA (RHOT1) or an ER resident protein (SQS) in cells depleted of MTCH2. (B) Mislocalization of mitochondrial TAs in wildtype versus MTCH2 depleted cells was analyzed in vitro by appending a C-terminal opsin tag to their C-termini. The substrates were translation in reticulocyte lysate, puromycin treated and then mixed with semi-permeabilized cells. ER localization was detected as a glycosylated species ('glyc'). Acceptor peptide (AP) was used to prevent glycosylation, thereby confirming that the higher molecular weight band corresponded to a glycosylated species. Non-TA and ER resident TA controls were used to confirm the effect was specific to mislocalized mitochondrial TAs. (C) As in (A) but MTCH2 was depleted alongside the ER quality control factor ATP13A1. (D) Blots depicting MTCH2 and ATP13A1 levels in cells used for reporter assays in (C).

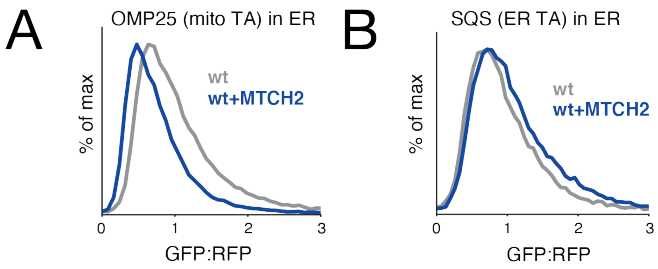


Figure S4.22. MTCH2 overexpression results in less mislocalization of mitochondrial TAs to the ER. As in fig. S4.21 but under conditions where MTCH2 is over-expressed. Note that there is a baseline of mislocalized OMP25-GFP11 reporter in the ER, likely due to its being a dual localized TA protein and over-expression of the construct.

Chapter 5

EVOLUTION AND REGULATION OF MTCH2

5.1 Abstract

MTCH2 is a membrane protein insertase in the human outer mitochondrial membrane (OMM). In this study we identify MTCH2 homologues distributed throughout the holozoan clade which includes metazoans and closely related protists. We also identify homologues of the trypanosome OMM insertase ATOM36 in plants and other protist lineages. We demonstrate that the identified homologues localize to the OMM in human cells, and are capable of inserting proteins into the OMM both in cells and *in vitro*. Lastly, we discovered that MTCH2 can bind to GTP, and provide evidence that GTP binding can regulate its OMM insertase activity.

5.2 Introduction

Eukaryotic cells are highly organized, with multiple organelles functioning together to coordinate processes essential for life. Information encoded in nuclear DNA is transcribed to RNA, which is transported to the cytoplasm where it is then translated into proteins. After translation, proteins must be targeted to their final destination, folded, and potentially assembled into quaternary complexes, a process collectively referred to as protein biogenesis. Due to the vast organizational complexity of eukaryotic cells, their protein biogenesis requires a large number of molecular machines which remain incompletely characterized.

Biogenesis of membrane proteins is a particularly challenging process. Integral membrane proteins encode one or more hydrophobic membrane spanning segments, termed transmembrane domains (TMDs). TMDs are inserted into the lipid bilayer by specialized proteins called insertases. TMDs must be inserted in the correct orientation, and all TMDs need to be assembled together into a correctly folded state, potentially with interaction

partners (Guna & Hegde, 2018). Membrane proteins residing in the outer mitochondrial membrane (OMM) represent an area of membrane protein biogenesis in need of further study, especially in mammalian cells. Mitochondria are organelles evolved from a bacterial ancestor which was incorporated into an archaeal host during the evolution of eukaryotes (Roger et al., 2017). In bacteria, nearly all characterized outer membrane proteins have β-barrel TMDs, which are inserted by the BamA/Omp85 insertase that is conserved in mitochondria as the SAM complex. However, in mitochondria, the OMM contains many more proteins with α-helical TMDs, and there is no known machinery in the bacterial outer membrane to insert α-helical TMDs (Weiner & Li, 2008; Rath et al., 2021; Busch et al., 2023).

Prior work has revealed three distinct OMM insertase for α-helical TMDs: the two-protein Mim complex in yeast, ATOM36 in trypanosomes, and more recently, MTCH2 in humans (Doan et al., 2020; Guna et al., 2022; Vitali et al., 2018). Each of these insertases appear to be members of distinct families, with no obvious homology to one another. Given the lack of α-helical TMDs in bacteria, it is expected that each of these families evolved after the transition of mitochondria from its bacterial ancestor into its present form as an organelle. In order to improve on our current understanding of mitochondrial evolution, it is imperative to know whether an OMM insertase was present in the last eukaryotic common ancestor (LECA) or if it evolved independently in later lineages.

MTCH2 is an atypical member of the SLC25 transporter family. SLC25 transporters typically contain six TMDs, and are localized to the inner mitochondrial membrane where they transport small molecules such as nucleotides or amino acids (Ruprecht & Kunji, 2020). MTCH2 resides in the OMM, is predicted to only contain five TMDs, and has no known transporter activity. The evolution of MTCH2 from an ancestor that likely resembled a more canonical SLC25 transporter represents a dramatic evolutionary change that warrants further study. MTCH2 can insert a variety of substrates into the OMM including apoptosis regulators such as Bax, Bak, and Bcl-xL. Apoptosis is a unique cell death pathway that evolved in metazoans (Bender et al., 2012). The process of apoptosis must be tightly regulated to avoid

catastrophic scenarios such as the widespread cell death or unchecked proliferation of mutated cells. Therefore, the evolution of MTCH2 could have been necessary to allow tighter regulation of OMM protein composition. While MTCH2 is not an essential gene on its own in humans, simultaneous knockdown of MTCH2 and its paralogue MTCH1 is lethal, suggesting that MTCH2 and MTCH1 together perform an essential role in human cells (Shi et al., 2022).

In this chapter, we describe the characterization of metazoan and protist MTCH2 homologues, as well as plant homologues of ATOM36. The study of these genes provides a better understanding of the distribution of OMM insertases within eukaryotes and their potential evolutionary landscape. Additionally, we demonstrate that MTCH2 is regulated by nucleotide binding, and hypothesize that this regulation was necessary to allow tight control of OMM insertion to avoid unregulated apoptosis in metazoans.

5.3 Results

MTCH2 homologues are conserved across metazoans

To search for homologues of OMM insertases, we took advantage of precomputed protein families in either the PANTHER of Pfam databases that each protein sequence in the UniProt database is annotated with (Thomas et al., 2022; UniProt, 2023; Mistry et al., 2021). A search for all sequences in UniProt annotated as belonging to the MTCH2-associated PANTHER family (PTHR10780) identified proteins distributed across all metazoan species and closely related protists within the holozoan clade (**Figure 5.1A**). Human MTCH2 possesses several key differences from other SLC25 family members, including lack of a charged-residue containing motif, and lack of a sixth transmembrane helix (Guna et al., 2022). Analysis of a sequence alignment with MTCH2 homologues from key metazoan and protist species shows that these features are also absent (**Figure S5.1**). Comparison of AlphaFold-based structural predictions of MTCH2 from humans and a homologue found in the protist *C. owczarzaki* shows that both proteins are predicted to contain a hydrophilic crevice and an extended β-

hairpin between the first and second transmembrane helices (Jumper et al., 2021) (**Figure 5.1D, left**).

Another major distinction between MTCH2 and most other SLC25 family members is its localization to the OMM, with its N-terminus facing the cytoplasm and C-terminus exposed to the inter-membrane space (IMS). To test whether the identified MTCH2 homologues are also OMM-localized with the same topology, we used a previously established split-fluorescent protein reporter assay based on the use of mitochondrial intermembrane-space localized GFP1-10 (Guna et al., 2022) (**Figure 5.1B**). We collected data of various GFP11-tagged MTCH2 homologues showing that like human MTCH2, other MTCH2 homologues insert into the OMM with their C-terminus in the IMS (**Figure 5.1C**). We therefore hypothesize that a MTCH2 ancestor with the same distinct structural features and OMM localization emerged during or slightly before the evolution of metazoans.

Outer mitochondrial membrane insertases with structural similarity to trypanosome ATOM36 are conserved throughout plants

A key feature of previously characterized membrane protein insertases is the presence of multiple transmembrane helices which enclose a hydrophilic crevice with a lateral opening to the lipid bilayer (Pleiner et al., 2020). To search for a potential OMM insertase in plants, we manually analyzed the AlphaFold-predicted structural models of all 64 genes identified as OMM localized in *A. thaliana* in a mitochondrial proteomics study (Duncan et al., 2011). We identified one particular gene, At5g55610.1 (UniProt accession Q9FM77), as a potential OMM insertase due to its predicted structural features. Upon further analysis, the AlphaFold-predicted structure of this protein was found to be highly similar to that of the trypanosomes OMM insertase ATOM36 (**Figure 5.1D, right**), therefore we consider this protein as a possible ATOM36 homologue. Interestingly, although ATOM36 is not an SLC25 family member, its AlphaFold-predicted structure bears a striking resemblance to that of MTCH2, with five transmembrane helices enclosing a hydrophilic crevice that opens to one side of the membrane (**Figure 5.1D**).

Previous studies of ATOM36 have confirmed that it is an integral membrane protein with its C-terminus in the cytoplasm, although the exact topology was not clear by sequence-based transmembrane helix prediction, likely due to the presence of many hydrophilic residues lining the described hydrophilic crevice (Pusnik et al., 2012). We carried out experiments with GFP11-tagged trypanosome ATOM36 showing that it inserts into human mitochondria with its N-terminus in the IMS and C-terminus in the cytoplasm (**Figure 5.1C, middle**). These findings agree with previous biochemical characterization, and suggest that the opening of the predicted hydrophilic crevice would face the cytoplasm, similarly to MTCH2. N-terminal GFP11-tagged *A. thaliana* ATOM36 also produces a moderate fluorescent signal despite much lower expression (**Figure 5.1C, left; Figure 5.2B**), indicating the same topology.

A search for all sequences in UniProt annotated as belonging to the plant ATOM36-associated PANTHER family (PTHR36074) identified proteins distributed across the viridiplantae clade, as well as other protist groups, primarily in the stramenopiles clade. For comparison, we also searched for sequences in UniProt annotated as belonging to the trypanosome ATOM36-associated Pfam family (PF19224) and identified proteins distributed across the euglenozoan clade. Sequence alignment of ATOM36 homologues reveals a relatively high degree of sequence conservation, particularly around regions of the protein predicted to be structured (**Figure S5.2-5.3**). The simplest explanation for the observed distribution of ATOM36 homologues is that it originally evolved in viridiplantae, and then was acquired in other groups by gene transfer from endosymbiotically acquired algal viridiplantae species, which is known to have happened in the ancestors of both stramenopiles and euglenozoans where the majority of non-viridiplantae ATOM36 homologues are found (**Figure S5.4**) (Hannaert et al., 2003; Qiu et al., 2013).

The identified MTCH2 and ATOM36 homologues are OMM insertases

To assess the activity of the MTCH2 and ATOM36 homologues, we first tested their ability to rescue the effect of a MTCH2 knockout (KO) on OMM protein insertion in live cells using our previously established split-fluorescent reporter (Guna et al., 2022). In this system, an N-

terminal fragment of GFP (GFP1-10) is targeted to the mitochondrial IMS, and we then query insertion of a second protein into the OMM by fusing a C-terminal fragment of GFP (GFP11). Once the target protein is successfully inserted into the OMM, GFP11 will be translocated to the IMS where it can complex with GFP1-10 to generate a fluorescent signal. We assessed the function of five non-human MTCH2 homologues and two ATOM36 homologues using seven OMM proteins as test substrates and two IMS-localized proteins as negative controls (**Figure 5.2A, C**). The rescue phenotype of human MTCH2 in MTCH2 KO cells is strongest for the MAVS, OMP25, RHOT2, and TOM70 substrates. With the same substrates, we see comparable levels of rescue using MTCH2 homologues from mouse, *X. tropicalis*, and *C. owczarzaki*, a diminished level of rescue from *D. melanogaster* MTCH2, and little to no rescue from *N. vectensis* MTCH2. Interestingly, for both ATOM36 homologue, we see much higher activity level in this reporter assay than any MTCH2 homologue we tested. This difference cannot be explained by expression levels, since all homologues are expressed at a similar level in human cells, except for *A. thaliana* ATOM36, which has a substantially lower expression level (**Figure 5.2D**).

To test whether the rescue we observe is due to a direct change in OMM insertase activity, we performed *in vitro* insertion assays with mitochondria purified from either WT or MTCH2 KO human cells overexpressing a MTCH2 homologue, which revealed that *H. sapiens and C. owczarzaki* both restore insertion of multiple OMM proteins to a similar degree, while having no effect on a matrix-targeted control (**Figure 5.2B**). Together, these data shows that MTCH2 and ATOM36 homologues from a wide variety of organisms are capable of inserting OMM proteins in human cells.

Nucleotide binds to MTCH2 and regulates its activity

Previous study of the SLC25 member UCP1 revealed that its proton leak activity is inhibited by nucleotide binding (Lee et al., 2015). Using a thermal stability assay based on the fluorescent dye CPM (7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin)), we tested whether MTCH2 could also bind nucleotides (Alexandrov et al., 2008). CPM reacts with regions of membrane proteins that become exposed upon misfolding and produces a

fluorescent signal. When incubated in the presence of increasing temperature, the derivative of fluorescence with respect to temperature (dF/dT) will peak at a temperature where a majority of the protein becomes misfolded. When measuring the thermal stability of MTCH2 alone, we observe two peaks at approximately 55 °C and 80 °C, suggesting a two-state unfolding process, perhaps with the higher temperature peak corresponding to the unfolding of individual SLC25 repeats. Previous studies of UCP1 found an unfolding transition of around 50 °C when it was solubilized in detergent (Lee et al., 2015). We then tested the effect of adding ATP, GTP, or GDP, and found that GTP specifically caused a shift in the lower temperature unfolding peak from ~55 °C to ~70 °C, suggesting a specific binding of GTP by MTCH2 (**Figure 5.3A**). Next, we performed the same experiment with different concentrations of GTP and found that the stabilization could only occur at GTP concentrations > 500 μM (**Figure 5.3B**). Since cellular GTP concentrations are typically around 700 μM, MTCH2 activity may be sensitive to physiological changes in GTP concentrations. Previous analysis of MTCH2 identified mutants that cause decreased or increased MTCH2 activity (Guna et al., 2022), and we reasoned that such changes could be due to differences in GTP binding properties caused by the mutations. To test this hypothesis, we purified two mutants, D189R which exhibits reduced insertase activity, and K25E which exhibits increased insertase activity, and tested their thermostabilities in the presence of increasing concentrations of GTP (**Figure 5.3C**). Interestingly, we found that D189R was largely stabilized by GTP at a concentration of 500 μM, while K25E was largely unaffected by GTP at a concentration of 1000 μM.

Based on these results, we hypothesize that nucleotide binding to MTCH2 inhibits its insertase function, and the altered insertase activity of the K25E and D189R mutants can be explained by decreased or increased nucleotide binding respectively. To test this hypothesis more rigorously, we performed insertion assays of substrates translated *in vitro* with MTCH2 reconstituted into proteoliposomes incubated with increasing concentrations of the non-hydrolysable GTP analogue GTPγS, since the translation extract used contains many enzymes that would hydrolyze GTP. Strikingly, we found that MTCH2 proteoliposomes incubated with either 2mM or 5 mM GTPγS showed a substantial decrease in insertion of

the *in vitro* translated OMM protein OMP25 at multiple time points, supporting the hypothesis that GTP inhibits MTCH2 insertase activity (**Figure 5.3D**). These results collectively provide convincing evidence that MTCH2 is bound to and regulated by nucleotide, which is likely to have important biological implications.

5.4 Discussion

Our work provides evidence that MTCH2 arose after the divergence of the holozoan clade, and that OMM protein insertion was its ancestral function. It is possible that MTCH2 evolved from a nucleotide transporting SLC25 family member, and that the nucleotide binding activity was conserved to enable regulation of MTCH2. The need for regulation of MTCH2 is not surprising given prior knowledge that MTCH2 activity can increase cellular sensitivity to apoptosis (Guna et al., 2022). It has been shown that MTCH2 modulation affects lipid accumulation in both mice and *C. elegans* (Rottiers et al., 2017). It would be interesting to investigate whether the nucleotide binding would enable changes to MTCH2 activity in starvation conditions to enable an appropriate metabolic response to starvation by remodeling the OMM proteome.

To fully understand the significance of nucleotide binding by MTCH2, further details on the kinetics and mechanism of action will be needed. It would be especially useful to know the dissociation constant (K_D) of nucleotide binding to MTCH2, to better understand how particular metabolic changes in cells may influence MTCH2 activity. Structural characterization of MTCH2 in its nucleotide bound state would enable more detailed studies of the mechanism behind nucleotide-dependent regulation of MTCH2 activity, for instance whether nucleotides function by steric hinderance of the insertase path within MTCH2, or whether they trigger a broader conformation change. More detailed study could provide insight into how other small molecules could be developed to more precisely or potently target MTCH2 activity. In our comparison of the OMM insertase activity of various MTCH2 and ATOM36 homologues, the ATOM36 homologues seem to be much more active than MTCH2 when considering their relative expression levels (**Figure 5.2C-D**). One possible explanation for this could be a lack of regulation of the ATOM36 family by nucleotides.

The demonstration that MTCH2 represents a metazoan OMM insertase class, and that extension of ATOM36 as a OMM insertase class found in plants and other protists represents a significant advance in our understanding of the evolutionary landscape of OMM proteins in eukaryotes. Published analyses suggest that multiple OMM proteins with α-helical TMDs were present in the LECA (Vlahou et al., 2011; Sinha & Manoj, 2019; Fukasawa et al., 2017). If we consider this in the context of the independent evolution of MTCH2 in metazoans, Mim1/2 in fungi, and potentially ATOM36 in plants, it suggests that membrane proteins in the LECA may have been capable of inserting into the OMM without the aid of an insertase. In this scenario, the basis for the evolution of OMM insertases in later diverging species could have been driven by a greater need to regulate the OMM protein composition. One interesting detail revealed by this analysis is that evolution of OMM insertases corresponds with the development of multicellularity in plants, animals, and fungi. Previous literature has argued that expanded mitochondrial protein diversity could support cell-type specific mitochondrial function in multicellular organisms, and in this case, the presence of a dedicated OMM insertase may become more important to regulate cell-type specific changes in OMM protein composition (Hammond et al., 2022), possibly explaining why OMM insertases are primarily found in multicellular organisms.

During the evolution of mitochondria from its bacterial ancestor, the OMM likely underwent a great degree of change due to its contact with the host cytoplasm. By studying OMM protein biogenesis in diverse eukaryotic lineages, we may develop a better understanding of the precise ancestral changes that occurred in the OMM during the early evolution of mitochondria. Further study of OMM protein composition and insertase activity in protist lineages lacking one of the three known OMM insertase families would provide further insight into the evolutionary constraints surrounding the evolution of the OMM proteome. Additionally, whole genome sequencing of underexplored protist lineages would also provide further insight by enabling more exhaustive queries for the distribution of OMM insertase families.

5.5 Acknowledgements

Alina Guna performed all FACS-based experiments. Erini Galatis assisted with protein expression and cloning. Alina Guna and Alison Inglis assisted with *in vitro* import experiments using purified mitochondria. Rebecca Voorhees aided in project conception and experimental design.

5.6 Methods

Plasmids and antibodies

All reporters and rescue constructs expressed in K562 were cloned into a previously described lentiviral backbone which includes a UCOE-EF-1α promoter and 3′ WPRE element (Guna et al., 2022). The design of reporters encoding RHOT2, OMP25, MICU, and MAVS was previously described (Guna et al., 2022). Additional sequences used in this study (sourced from the indicated UniProt accession) include *D. melanogaster* MTCH2 (Q9V3Y4), *X. tropicalis* MTCH2 (Q6P818), *C. owczarzaki* MTCH2 (A0A0D2WXZ5), *T. brucei brucei* ATOM36 (Q582I5), *A. thaliana* ATOM36 (Q9FM77). For complementation with GFP1-10 the sequence of GFP11 (RDHMVLHEYVNAAGIT) was fused to the N- or C-termini. To enable quantification of protein expression levels by western blotting, MTCH2 and ATOM36 homologues were fused to N- or C-terminal 3x FLAG (DYKDDDDK) tags.

Plasmids for expressing WT and mutant MTCH2 in mammalian cells were cloned into the pHAGE2 backbone with an N-terminal GFP-SUMOEu tag as previously described (Guna et al., 2022). Plasmids for expressing proteins in rabbit reticulocyte lysate (RRL) were cloned into the SP64 vector (Promega, Madison, WI), and all plasmids for RRL expression used here were described in a previous publication (Guna et al., 2022). All plasmids used for nanobody-based purification of MTCH2 have been previously described (Stevens et al., 2023).

Cell culture

K562 cells were grown in RPMI-1640 media supplemented with 25 mM HEPES, 2.0 g/L $NaHCO_3$, 10% FBS, 2 mM glutamine, 100 units/mL penicillin, and 100 µg/mL streptomycin.

Cells were maintained between $0.25 \times 10^6 – 1 \times 10^6$ cells/mL. Cells were grown at 37 °C in the presence of 5% CO_2. All experiments using K562 cells were done in a parental line with stable expression of dCas9-BFP-KRAB (KOX1-derived) and MICU1(GFP1- 10) which was described in a previous publication (Guna et al., 2022). Some K562 experiments were performed in a MTCH2 KO background using a cell line whose generation via CRISPR knockout was also previously described (Guna et al., 2022).

HEK293T cells were grown in DMEM supplemented with 10% FBS, 2 mM glutamine, 100 units/mL penicillin and 100 μg/mL streptomycin. Cells were grown at 37 °C in the presence of 5% CO_2. Expi293F cells were grown in Expi293 expression media (Gibco) at 37 °C while shaking at 125 RPM in the presence of 8% CO_2.

Lentivirus production

Lentivirus was produced by transfecting HEK293T cells with two packaging plasmids (pCMV-VSV-G and delta8.9, Addgene #8454) and the desired transfer plasmid with the use of TransIT- 293 transfection reagent (Mirus). 48 hours after transfection, the supernatant was collected and flash frozen. At the time of use, virus was rapidly thawed prior to transduction.

Protein production and purification

The generation of a stable cell line for overexpressing GFP-SUMOEu-MTCH2 in the Expi293F inducible background was previously described (Guna et al., 2022). Stable cell lines for expressing MTCH2 mutants were produced in an identical manner.

Purification proceeded as follows: cell pellets were resuspended in a hypotonic lysis buffer containing 10 mM HEPES pH 7.5, 10 mM KoAc, 0.15 mM $MgAc_2$, 0.5 mM DTT, supplemented with EDTA-free protease inhibitor tablets (Roche). After incubation on ice for 10 minutes to allow cells to swell, cells were lysed using a\\ 10× strokes in a Dounce homogenizer. After adjusting the NaCl concentration to 180 mM, cell membranes were pelleted by centrifugation at 18,000 rcf in an SS-34 (28020TS, Thermo Fisher) rotor for 10 minutes. Supernatant was discarded and cell membranes were washed by resuspending and

pelleting in membrane wash buffer containing 10 mM HEPES/KOH pH 7.5, 200 mM NaCl, 0.15 mM $MgAc_2$, 0.5 mM DTT. The resulting pellet was resuspended in a solubilization buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 1% deoxy-BigCHAP (DBC; Anatrace Cat # B310), 1 mM DTT, supplemented with EDTA-free protease inhibitor tablets (Roche). After 30 minutes of head-over-tail incubation with solubilization buffer, the lysate was cleared by centrifugation for 30 minutes at 4 °C and 18,000 rpm in a SS-34 rotor. The supernatant was then added to pre-equilibrated magnetic Streptavidin resin (Thermo Fisher) bound to biotinylated anti-GFP nanobody and blocked with free biotin. After 1 hour of binding while rolling at 4 °C, the resin was washed four times with wash buffer (solubilization buffer with 0.25 % DBC). Resin was then incubated in wash buffer + 1200 nM SENPEuB on ice for 2 hours to release SUMO-cleaved MTCH2 from the resin. Eluted samples were analyzed via SDS-PAGE with Sypro Ruby stain (Bio-Rad).

GFP Nb (pTS117) and SENPEuB protease were purified via NiNTA-affinity chromatography after recombinant expression in *E. coli*, following a previously described protocol (Stevens et al., 2023).

Thermal shift assay

An assay for measuring the thermal stability of membrane proteins was adapted for use here (Alexandrov et al., 2008). Prior to analysis, proteins were buffer exchanged using a NAP-5 desalting column (Cytiva) with a buffer containing 50 mM HEPES pH 7.5, 200 mM NaCl, 2 mM $MgAc_2$, 0.25% DBC to remove reducing agent. Protein concentrations were then measured with A280. For each sample analyzed, protein and CPM (7-Diethylamino-3-[4'-Maleimidylphenyl]-4-Methylcoumarin) were mixed with the desalting buffer at a final concentration of 0.1 mg/ml and 4µg/ml respectively with varying concentrations of nucleotide in a final volume of 30µl. Samples were then analyzed in a StepOnePlus RT-PCR system (Applied Biosystems) with the following protocol: temperature was held at 25 °C for 2 minutes, increased at a rate of 1% (~1 °C/min) to 99 °C and then held at 99 °C for 2 minutes.

Fluorescence measurements were made using the SYBR green filter. Data was plotted and analyzed as the derivative of Fluorescence with respect to temperature (dF/dT).

Mitochondrial isolation

To generate MTCH2 KO cell lines overexpressing either *H. sapiens* or *C. owczarzaki* MTCH2, K562 cells were transduced with lentiviral particles containing MTCH2 with an N-terminal BFP-P2A fusion to enable FACS isolation. As a control, cells were all transduced with lentivirus only expressing BFP. To sort MTCH2 expressing or control cells, a BD Biosciences FACS Aria Fusion Cell Sorter was used to sort 20 M cells for each sample. Cells were then expanded for no more than seven days prior to harvesting for use in mitochondrial isolation experiments.

Mitochondrial isolation was performed similarly to a previously described protocol (Guna et al., 2022). K562 cells were pelleted by centrifugation at 220*g* for 5 minutes. Pellets were washed once in in a homogenization buffer containing 210 mM mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA, 1 mM PMSF, 2 mg/mL BSA. Pellets were then resuspended in 2 ml homogenization buffer and incubated on ice for 10 minutes. Cells were then lysed with a 2 mL glass Dounce homogenizer with a tight-fitting pestle. Homogenized cells were pelleted at 1300*g* for 5 minutes to remove nuclei and unbroken cells, then the supernatant was transferred to a clean tube. This step was repeated twice. Nuclei-free homogenized cells were then centrifuged at 11,000*g* for 10 minutes. The supernatant was removed and the mitochondria-containing pellet was then resuspended in an isolation buffer containing 210 mM D-mannitol, 70 mM sucrose, 5 mM HEPES pH 7.4, 10 mM EDTA. The mitochondrial pellet was then washed 2x in isolation buffer. After a final pelleting step, mitochondria were resuspended in a small volume (5–50 μL) of isolation buffer. Prior to use in insertion experiments, mitochondrial concentration was normalized via Bradford assay.

Proteoliposome reconstitutions and insertions

Reconstitutions of MTCH2 into liposomes followed a protocol based on previously described methods (Guna & Hegde, 2018; Guna et al., 2022; Mariappan et al., 2011). Reconstitutions used egg-derived phosphatidyl-choline (PC) and phosphatidyl-ethanolamine (PE), and synthetic 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine-N-lissamine rhodamine B (Rh-PE), all obtained from Avanti polar lipids. Chloroform stocks of each lipid were mixed at a mass-based ratio of 8:1.9:0.1 PC:PE:Rh-PE, followed by addition of DTT to 10 mM. The mixture was then dried by centrifugation under vacuum overnight. The resulting lipid film was rehydrated in lipid buffer (15% glycerol, 50 mM HEPES pH 7.4) and mixed by shaking at 25 °C for 3-5 hours with until a homogenous mixture was achieved. Rh-PE absorbance at 573 nm was used to quantify the lipid concentration after rehydration. BioBeads-SM2 (Bio- Rad) were activated with methanol, washed thoroughly with distilled water, and then were resuspended in water into a final slurry where they occupied 50% of the final volume. For reconstitutions, excess liquid was removed from BioBeads by aspiration just before use. A standard 90 µL reaction contained 36 µL purified MTCH2, 180 µg lipids, and the remaining volume made up with buffer, salts, and detergent to achieve a final concentration of 0.8% DBC. The molar ratio of lipid:protein used in reconstitutions was 2000:1. This reconstitution mixture was then added to 90 µL BioBeads in a 1.5 mL Eppendorf tube. The slurry was mixed in a thermomixer for 18 hours at 4 °C. The reconstituted liposomes were then separated from the biobeads, diluted with five volumes of ice-cold water, and sedimented in a TLA120.2 rotor at 70,000 rpm for 30 minutes. After removing the supernatant, the pelleted liposomes were resuspended in 18 µL of a liposome resuspension buffer containing 100 mM KAc, 50 mM HEPES pH 7.4, 2 mM $Mg(Ac)_2$, 250 mM sucrose, 1 mM DTT.

In vitro translation and insertion

In vitro translations were carried out in rabbit reticulocyte lysate (RRL) as previously described (Guna & Hegde, 2018). Templates for transcription were generated by PCR, with primers binding and upstream of the SP64 promoter and atleast 12 bp downstream of the stop codon (Sharma et al., 2010). mRNA was produced by transcription of the PCR product

using SP6 polymerase at 37 °C for 1.5 hours. mRNA was then used directly in a translation reaction for 15–30 minutes at 32 °C in the presence of radioactive ^{35}S-methionine. Prior to addition of mitochondria or reconstituted proteoliposomes, 1 mM puromycin was added to prevent further synthesis.

Mitochondrial insertion reactions used isolated mitochondria, prepared as described above. Insertion reactions were performed by diluting 4 µL of a puromycin treated translation reaction in 50 µL of import buffer (250 mM sucrose, 5 mM $Mg(Ac)_2$, 80 mM KAc, 20 mM HEPES pH 7.4, 2.5 mM ATP, 15 mM succinate) with 15 µg of purified mitochondria and further incubating at 32 °C for 30 minutes.

To test the effect of nucleotide addition on insertion into proteoliposomes, pre-existing nucleotides needed to be depleted by buffer exchanging the translations into a RRL buffer containing 80 mM KAc, 2.5 mM $Mg(Ac)_2$, 20 mM HEPES pH 7.5, 0.5 mM spermidine, and 1.5 mM reduced glutathione using a NAP-5 desalting column (Cytiva). After buffer exchange, 9µl translation was then mixed with 1 µl resuspended liposomes and 1.5µl nucleotide diluted into RRL buffer. Insertion reactions were then incubated at 32 °C for 0-20 minutes.

Protease digestions were initiated by the addition of proteinase K at 0.25 mg/mL, and reactions were then incubated on ice for 1 hour. Reactions were quenched by the addition of 5 mM PMSF in DMSO, followed by transfer to boiling 1% SDS (final concentration) in 0.1 M Tris/HCl pH 8.0. His-tagged protected fragments were enriched by incubating with NiNTA resin in IP buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 1% Triton). Proteinase K-digested reactions were diluted to 1 mL and mixed with 10 µL resin, then incubated with end-over-end mixing for 1.5 hours at 4 °C. The resin was further washed with 3 × 1 mL IP buffer, and the products eluted from the resin in Laemmli SDS-PAGE sample buffer containing 50 mM EDTA pH 8.0.

Flow cytometry

For all experiments, K562 cell lines were spinfected with lentiviral particles packaged with the indicated construct and analyzed by flow cytometry after 72 hours. All samples were run on an NXT Flow Cytometer (ThermoFisher). Flow cytometry data was analyzed in Python using the FlowCytometryTools package.

Sequence analysis

MTCH2, ATOM36, and MIM1 homologue sequences were obtained by querying UniProt for sequences annotated with the respective PANTHER (PTHR10780 for MTCH2, PTHR36074 for plant ATOM36) or Pfam (PF19224 for trypanosome ATOM36, PF08219 for MIM1) families (Mistry et al., 2021; Thomas et al., 2022). All AlphaFold-predicted models were obtained from a published database (https://AlphaFold.ebi.ac.uk/) (Jumper et al., 2021). All sequence alignments were generated using MUSCLE (Edgar, 2004) and visualized using the ESPRIPT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)).

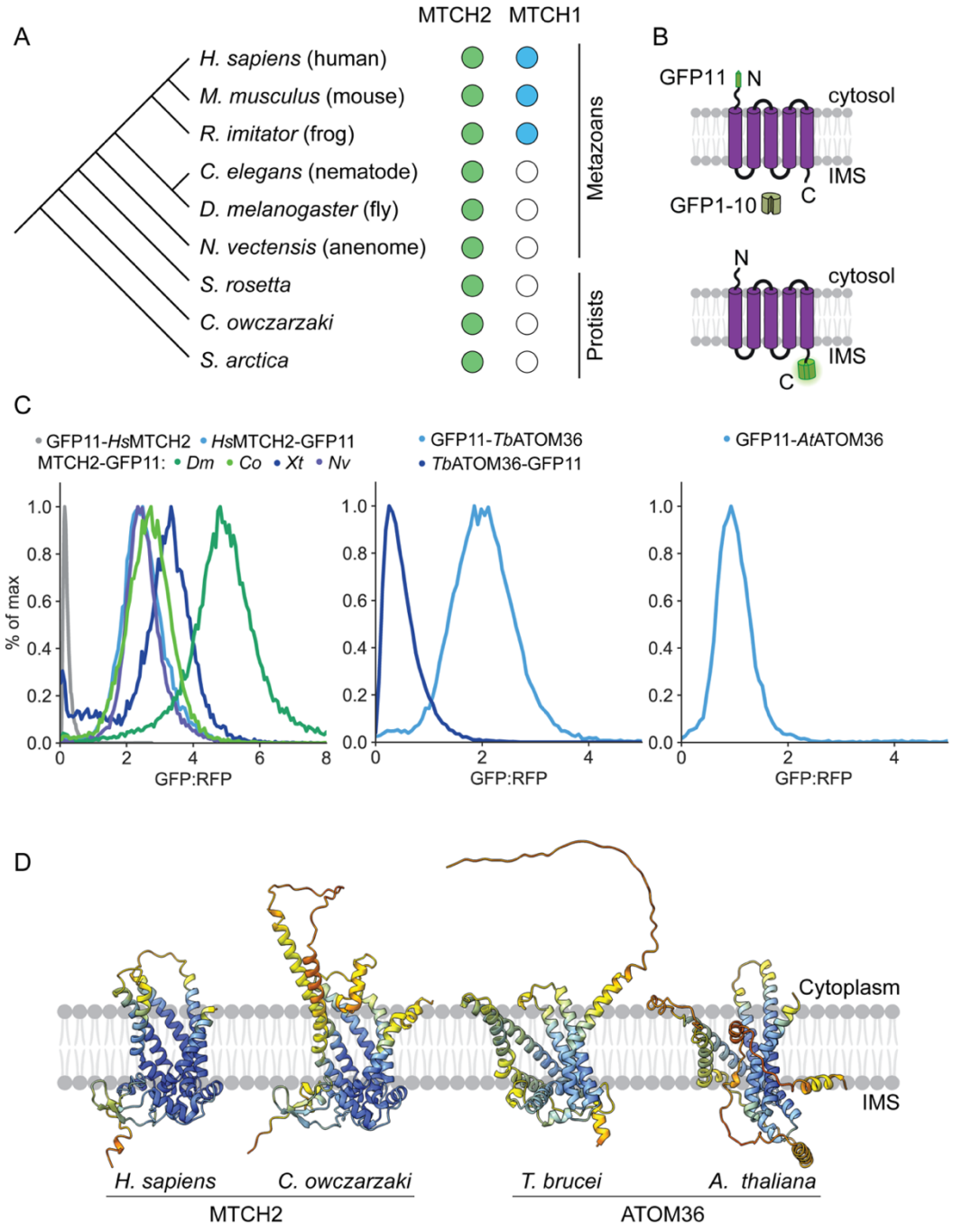


Figure 5.1. (A) Species tree of metazoans and closely related protists with graphical representation of MTCH2 homologue conservation. (B) Diagram of split GFP system with IMS GFP1-10 in the mitochondrial inter-membrane space (IMS) for testing outer mitochondrial membrane localization. (C) Left: analysis of GFP11-tagged MTCH2

homologues reveals conserved localization of C-terminus to the IMS. Middle: analysis of GFP11-tagged ATOM36 reveals localization of N-terminus to the IMS. Right: Analysis of GFP11-tagged plant ATOM36 suggests localization of N-terminus to the IMS (D) Comparison of MTCH2 and ATOM36 AlphaFold models suggest structural similarity. Sequences are colored according the standard AlphaFold coloring scheme where blue indicates a high per-residue model confidence score (pLDDT), yellow indicates a medium score, and orange indicates a low score.

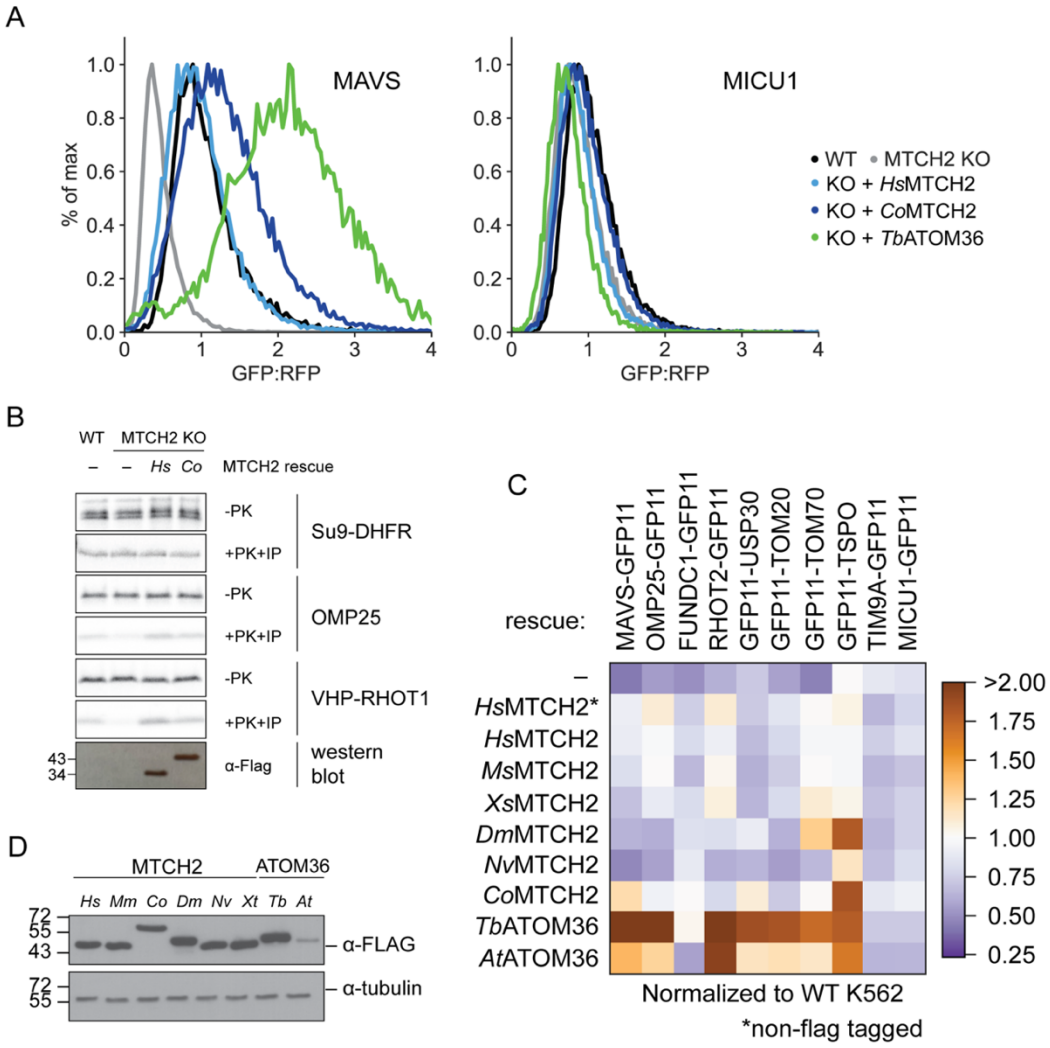


Figure 5.2. (A) Insertion of RFP-P2A-MAVS-GFP11 and RFP-P2A-MICU1-GFP11 into mitochondria of MICU1-GFP1-10 expressing K562 cells measured via FACS. Histograms are shown for WT, MTCH2 KO, and MTCH2 KO rescued with human MTCH2 or a protists homologue, and with ATOM36. (B) Rescue of the MTCH2 KO insertion phenotype was tested by purifying mitochondria from KO cells expressing either *H. sapiens* or *C. owczarzaki* MTCH2 and analyzing insertion of *in vitro* translated, ^{35}S-labelled OMM proteins. The top six panels are SDS-PAGE autoradiographs and the bottom panel is a western blot. The SDS-PAGE autoradiographs labeled (-PK) represent

the total protein that was input in the insertion reaction, and the panels labeled (+PK+IP) represent the protected fragment remaining after protease digestion which was affinity purified using NiNTA agarose. Relative intensity of bands in +PK+IP gels are a measure for membrane insertion of the translated proteins. Anti-FLAG western blot confirms comparable expression of either FLAG-tagged MTCH2 homologue in the rescue mitochondria. (C) Heat map of median GFP:RFP values for the indicated reporter (x-axis) expressed in MTCH2 KO K562 rescued with the indicated MTCH2 homologue (y-axis). All values are normalized to the median GFP:RFP value of each reporter expressed in WT K562 cells. (D) Western blot of FLAG-tagged MTCH2 and ATOM36 homologues demonstrates similar expression levels for all tagged proteins except *At*ATOM36, which rescues efficiently despite a lower expression level. Hs = *H. sapiens*, Co = *C. owczarzaki*, Mm = *M. musculus*, Dm = *D. melanogaster*, Nv = *N. vectensis*, Xt = *X. tropicalis*, Tb = *T. brucei brucei*, At = *A. thaliana*.

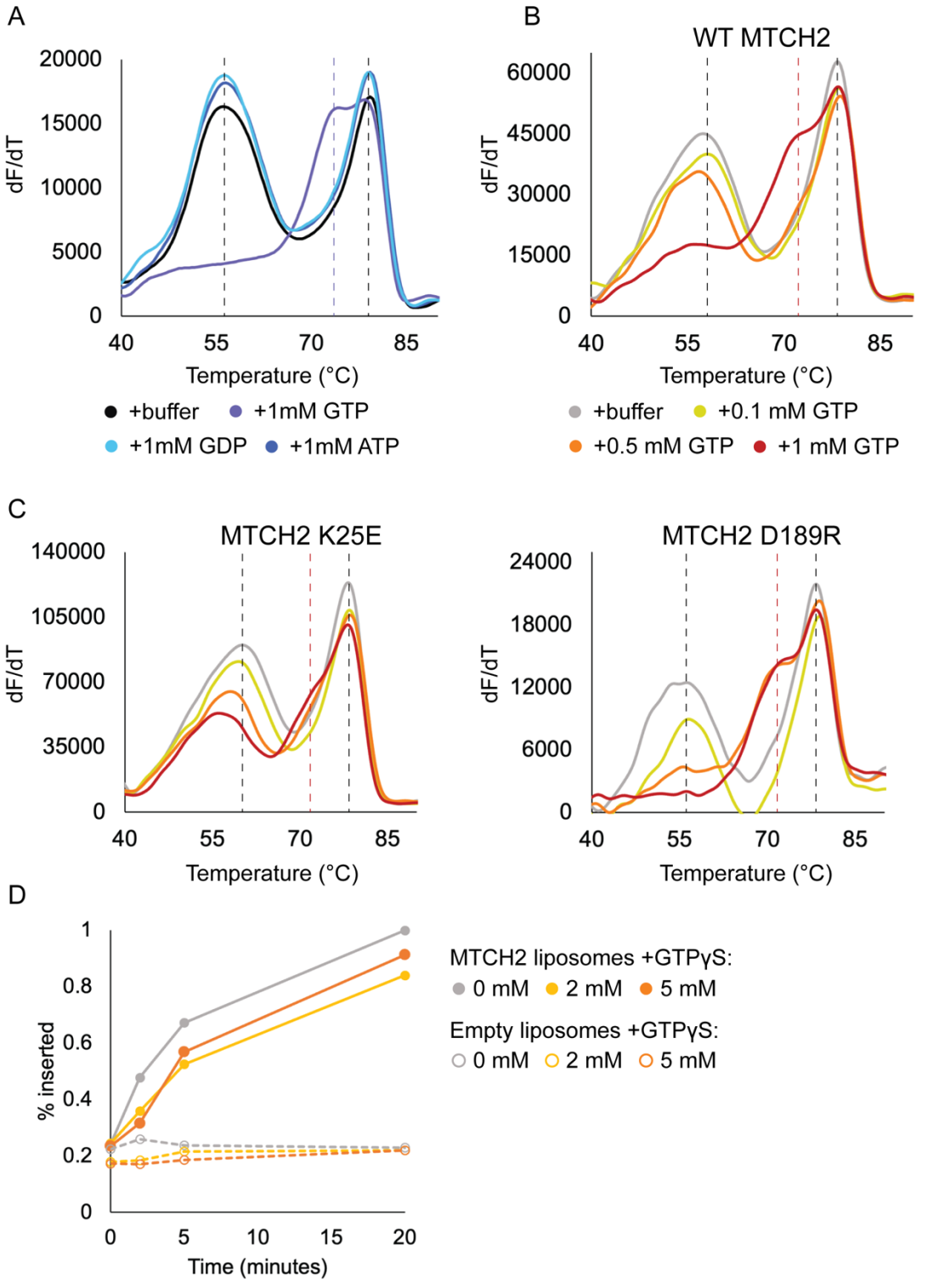
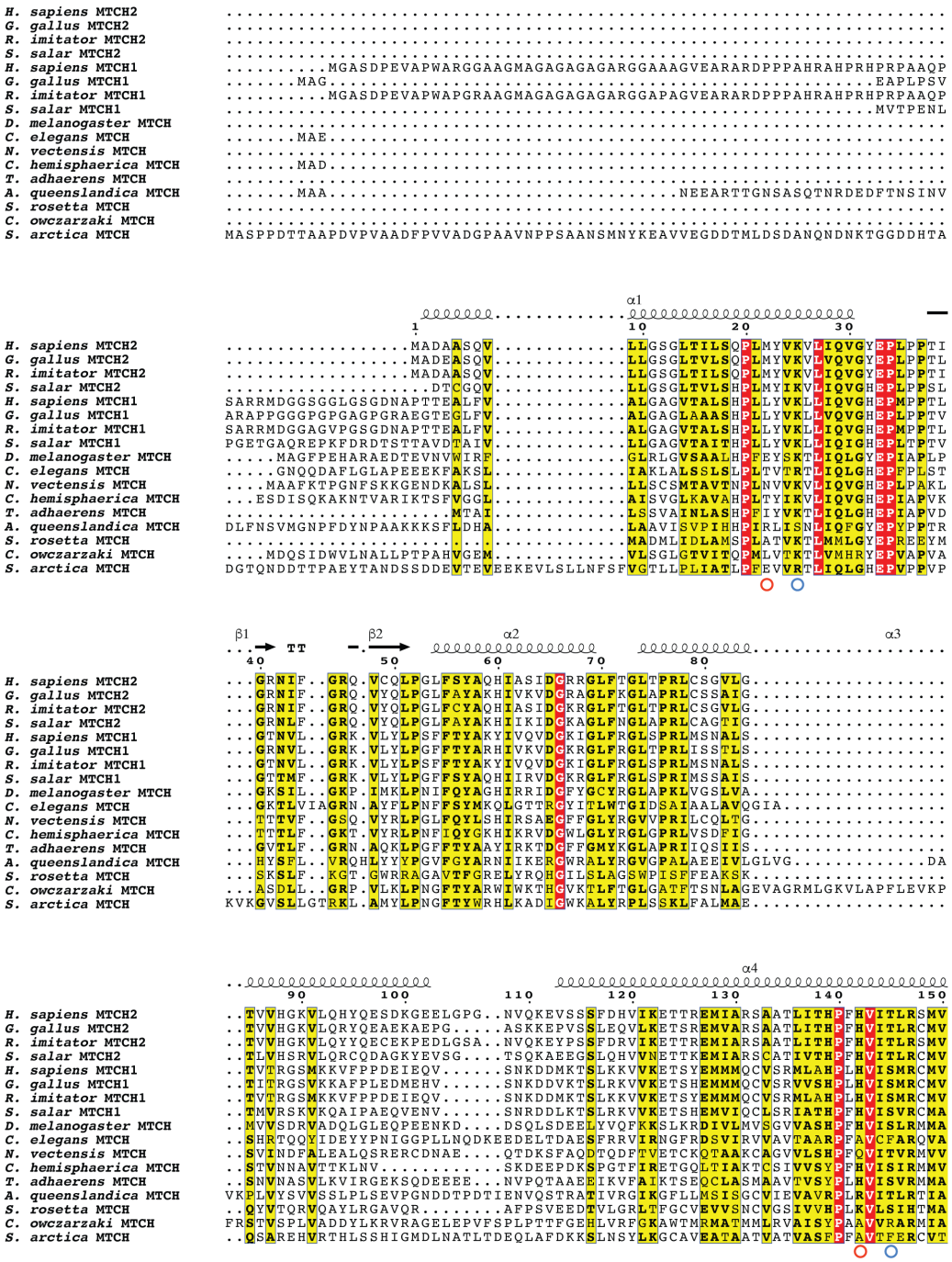

A
20000
15000
10000
5000
0
dF/dT
40
55
70
85
Temperature (°C)
+buffer
+1mM GTP
+1mM GDP
+1mM ATP
B
WT MTCH2
60000
45000
30000
15000
0
dF/dT
40
55
70
85
Temperature (°C)
+buffer
+0.1 mM GTP
+0.5 mM GTP
+1 mM GTP
C
MTCH2 K25E
140000
105000
70000
35000
0
dF/dT
40
55
70
85
Temperature (°C)
MTCH2 D189R
24000
18000
12000
6000
0
dF/dT
40
55
70
85
Temperature (°C)
D
1
0.8
0.6
0.4
0.2
0
% inserted
0
5
10
15
20
Time (minutes)
MTCH2 liposomes +GTPγS:
0 mM
2 mM
5 mM
Empty liposomes +GTPγS:
0 mM
2 mM
5 mM

Figure 5.3. (A) A thermal shift assay based on the fluorescent dye CPM shows a large increase in stability by addition of GTP, but not GDP or ATP, suggesting that MTCH2 is a GTP binding protein. The Y-axis is the derivative of the fluorescent signal with respect to temperature (dF/dT). Horizontal dashed lines represent two peaks of unstabilized MTCH2 (in black) or GTP-stabilized MTCH2 (in color). (B) Analyzing MTCH2 thermal stability with various concentrations of GTP suggests binding requires > 500 μM GTP. (C) Analyzing thermal stability of MTCH2 mutants shows altered binding properties. Color scale is the same as in (B). (D) Analyzing insertion of OMP25 into reconstituted MTCH2 liposomes with increasing concentrations of GTPγS shows a GTP-dependent reduction of insertion.


H. sapiens MTCH2
G. gallus MTCH2
R. imitator MTCH2
S. salar MTCH2
H. sapiens MTCH1
G. gallus MTCH1
R. imitator MTCH1
S. salar MTCH1
D. melanogaster MTCH
C. elegans MTCH
N. vectensis MTCH
C. hemisphaerica MTCH
T. adhaerens MTCH
A. queenslandica MTCH
S. rosetta MTCH
C. owczarzaki MTCH
S. arctica MTCH
α1
β1
β2
α2
α3
α4

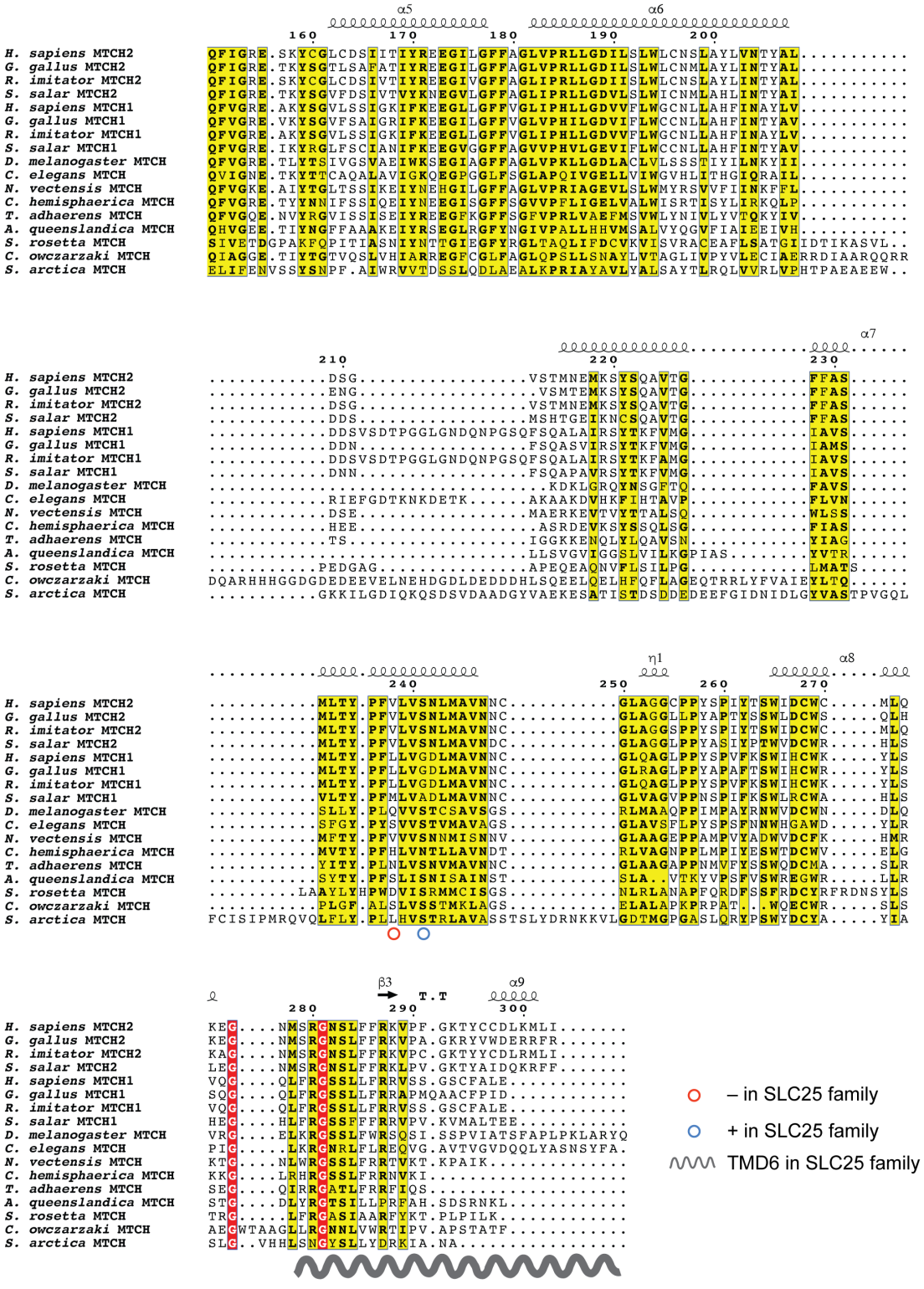
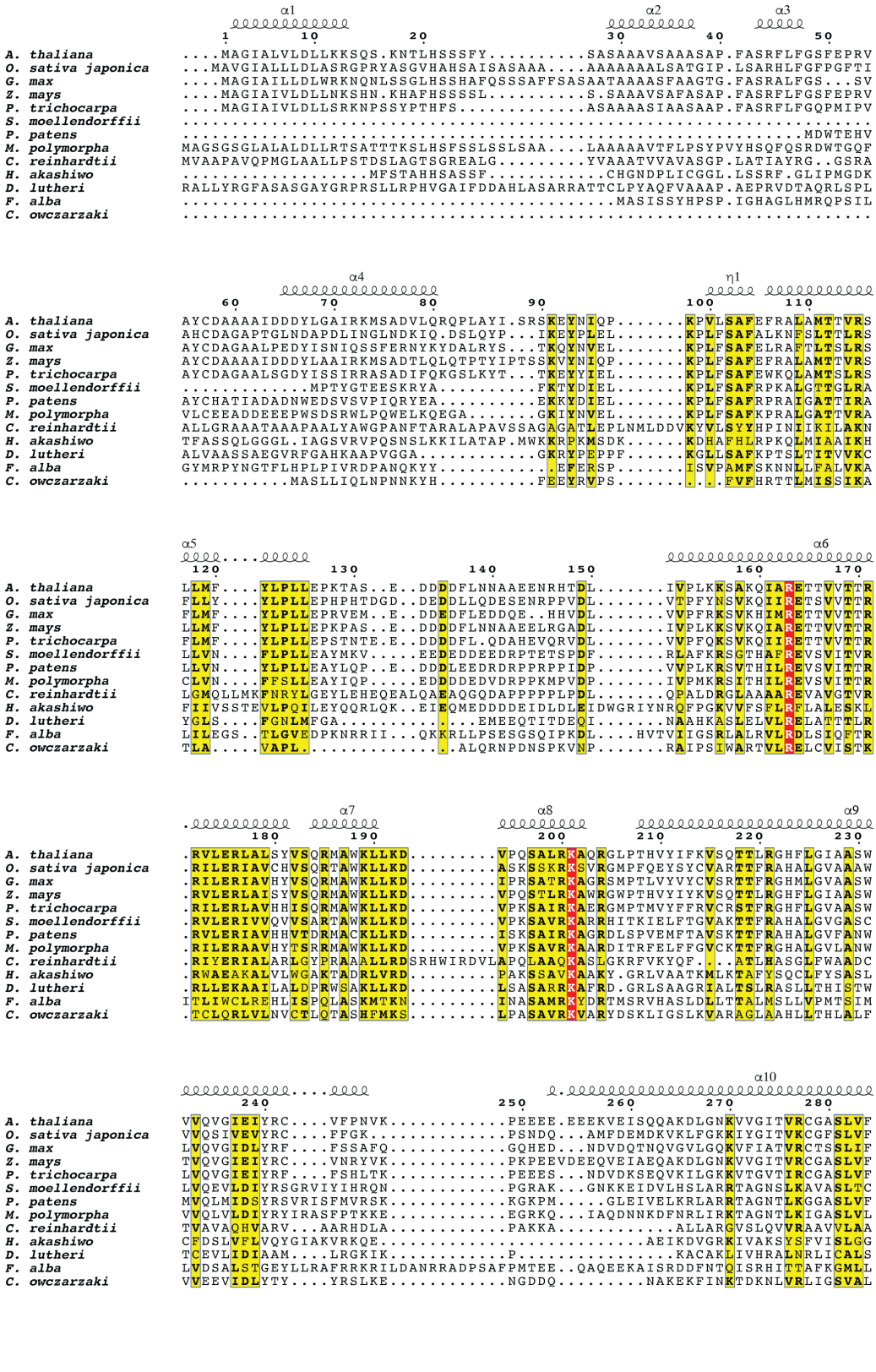

α5
α6
160
170
180
190
200
H. sapiens MTCH2
G. gallus MTCH2
R. imitator MTCH2
S. salar MTCH2
H. sapiens MTCH1
G. gallus MTCH1
R. imitator MTCH1
S. salar MTCH1
D. melanogaster MTCH
C. elegans MTCH
N. vectensis MTCH
C. hemisphaerica MTCH
T. adhaerens MTCH
A. queenslandica MTCH
S. rosetta MTCH
C. owczarzaki MTCH
S. arctica MTCH
α7
210
220
230
η1
α8
240
250
260
270
β3
T.T
α9
280
290
300
– in SLC25 family
+ in SLC25 family
TMD6 in SLC25 family

Figure S5.1. Sequence alignment of MTCH2 homologues throughout metazoans. Sequence alignment was generated using MUSCLE (Edgar, 2004) and visualized using the ESPRIPRT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)). Secondary structure representation shown on top based on *H. sapiens* MTCH2 AlphaFold-predicted model. Annotations for conserved charged residues and the typical location of TMD6 are below the sequences.


α1
α2
α3
α4
η1
α5
α6
α7
α8
α9
α10
A. thaliana
O. sativa japonica
G. max
Z. mays
P. trichocarpa
S. moellendorffii
P. patens
M. polymorpha
C. reinhardtii
H. akashiwo
D. lutheri
F. alba
C. owczarzaki

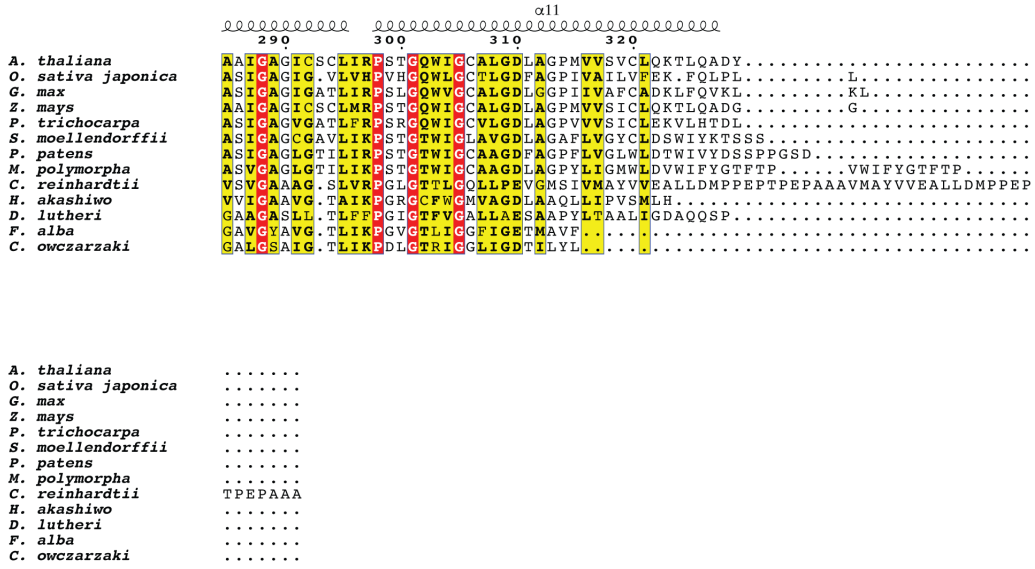
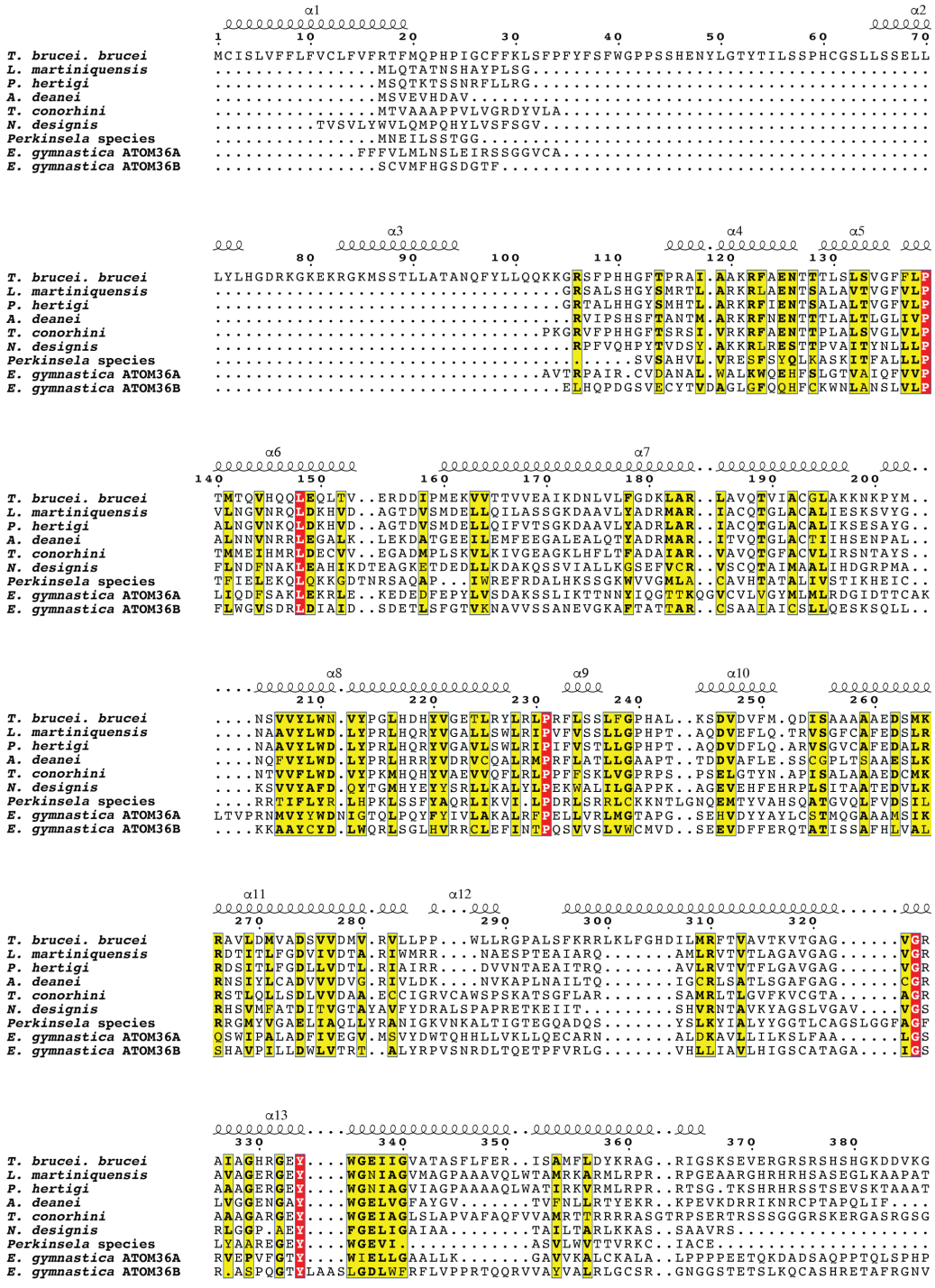


Figure S5.2. Sequence alignment of ATOM36 homologues throughout plants and protists. Sequence alignment was generated using MUSCLE (Edgar, 2004) and visualized using the ESPRIPRT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)). Secondary structure representation shown on top based on *A. thaliana* ATOM36 AlphaFold-predicted model.


α1
α2
1 10 20 30 40 50 60 70
T. brucei. brucei MCISLVFFLFVCLFVFRTFMQPHPIGCFFKLSFPFYFSFWGPPSSHENYLGTYTILSSPHCGSLLSSELL
L. martiniquensis .....MLQTATNSHAYPLSG
P. hertigi .....MSQTKTSSNRFLLRG
A. deanei .....MSVEVHDAV
T. conorhini .....MTVAAAPPVLVGRDYVLA
N. designis .....TVSVLYWVLQMPQHYLVSFSGV
Perkinsela species .....MNEILSSTGG
E. gymnastica ATOM36A .....FFFVLMLNSLEIRSSGGVCA
E. gymnastica ATOM36B .....SCVMFHGSDGTF
α3
α4
α5
80 90 100 110 120 130
T. brucei. brucei LYLHGDRKGKEKRGKMSSTLLATANQFYLLQQKKGRSFPHHGFTPRAI.AAKRFAENTTLSLSVGFFLP
L. martiniquensis .....GRSALSHGYSMRTL.ARKRLAENTSALAVTVGFVLP
P. hertigi .....GRTALHHGYSMHTL.ARKRFIENTSALALTVGFVLP
A. deanei .....RVIPSHSFTANTM.ARKRFNENTTLALTLGLIVP
T. conorhini .....PKGRVFPHHGFTSRSI.VRKRFAENTTPLALSVGLVLP
N. designis .....RPFVQHPYTVDSY.AKKRLRESTTPVAITYNLLLP
Perkinsela species .....SVSAHVL.VRESFSYQLKASKITFALLLP
E. gymnastica ATOM36A .....AVTRPAIR.CVDANAL.WALKWQEHFSLGTVAIQFVVP
E. gymnastica ATOM36B .....ELHQPDGSVECYTVDAGLGFQQHFCKWNLANSLVLP
α6
α7
140 150 160 170 180 190 200
T. brucei. brucei TMTQVHQQLEQLTV..ERDDIPMEKVVTTVVEAIKDNLVLFGDKLAR..LAVQTVIACGLAKKNKPYM..
L. martiniquensis VLNGVNRQLDKHVD..AGTDVSMDELLQILASSGKDAAVLYADRMAR..IACQTGLACALIESKSVYG..
P. hertigi ALNGVNKQLDKHVD..AGTDVSMDELLQIFVTSGKDAAVLYADRLAR..IACQTGLACALIKSESAYG..
A. deanei ALNNVNRRLEGALK..LEKDATGEEILEMFEEGALEALQTYADRMAR..ITVQTGLACTIIHSENPAL..
T. conorhini TMMEIHMRLDECVV..EGADMPLSKVLKIVGEAGKLHFLTFADAIAR..VAVQTGFACVLIRSNTAYS..
N. designis FLNDFNAKLEAHIKDTEAGKETDEDLLKDAKQSSVIALLKGSEFVCR..VSCQTAIMAALIHDGRPMA..
Perkinsela species TFIELEKQLQKKGDTNRSAQAP...IWREFRDALHKSSGKWVVGMLA..CAVHTATALIVSTIKHEIC..
E. gymnastica ATOM36A LIQDFSAKLEKRLE..KEDEDFEPYLVSDAKSSLIKTTNNYIQGTTKQGVCVLVGYMLMLRDGIDTTCAK
E. gymnastica ATOM36B FLWGVSDRLDIAID..SDETLSFGTVKNAVVSSANEVGKAFTATTAR..CSAAIAICSLLQESKSQLL..
α8
α9
α10
210 220 230 240 250 260
T. brucei. brucei ....NSVVYLWN.VYPGLHDHYVGETLRYLRLPRFLSSLFGPHAL..KSDVDVFM.QDISAAAAAEDSMK
L. martiniquensis ....NAAVYLWD.LYPRLHQRYVGALLSWLRIPVFVSSLLGPHPT..AQDVEFLQ.TRVSGFCAFEDSLR
P. hertigi ....NAAVYLWD.LYPRLHQRYVGAVLSWLRIPIFVSTLLGPHPT..AQDVDFLQ.ARVSGVCAFEDALR
A. deanei ....NQFVYLWD.LYPRLHRRYVDRVCQALRMPRFLATLLGAAPT..TDDVAFLE.SSCGPLTSAAESLK
T. conorhini ....NTVVFLWD.VYPKMHQHYVAEVVQFLRLPPFFSKLVGPRPS..PSELGTYN.APISALAAAEDCMK
N. designis ....KSVVYAFD.QYTGMHYEYYSRLLKALYLPEKWALILGAPPK..AGEVEHFEHRPLSITAATEDVLK
Perkinsela species ....RRTIFLYR.LHPKLSSFYAQRLIKVI.LPDRLSRRLCKKNTLGNQEMTYVAHSQATGVQLFVDSIL
E. gymnastica ATOM36A LTVPRNMVYYWDNIGTQLPQYFYIVLAKALRFPELLVRLMGTAPG..SEHVDYYAYLCSTMQGAAAMSIK
E. gymnastica ATOM36B ....KKAAYCYD.LWQRLSGLHVRRCLEFINTPQSVVSLVWCMVD..SEEVDFFERQTATISSAFHLVAL
α11
α12
270 280 290 300 310 320
T. brucei. brucei RAVLDMVADSVVDMV.RVLLPP...WLLRGPALSFKRRLKLFGHDILMRFTVAVTKVTGAG.....VGR
L. martiniquensis RDTITLFGDVIVDTA.RIWMRR....NAESPTEAIARQ.....AMLRVTVTLAGAVGAG.....VGR
P. hertigi RDSITLFGDLLVDTL.RIAIRR....DVVNTAEAITRQ.....AVLRVTVTFLGAVGAG.....VGR
A. deanei RNSIYLCADVVVDVG.RIVLDK....NVKAPLNAILTQ.....IGCRLSATLSGAFGAG.....CGR
T. conorhini RSTLQLLSDLVVDAA.ECCIGRVCAWSPSKATSGFLAR.....SAMRLTLGVFKVCGTA.....AGR
N. designis RHSVMFATDITVGTAYAVFYDRALSPAPRETKEIIT.....SHVRNTAVKYAGSLVGAV....VGS
Perkinsela species RRGMYVGAELIAQLLYRANIGKVNKALTIGTEGQADQS.....YSLKYIALYYGGTLCAGSLGGFAGF
E. gymnastica ATOM36A QSWIPALADFIVEGV.MSVYDWTQHHLLVKLLQECARN.....ALDKAVLLILKSLFAA.....LGS
E. gymnastica ATOM36B SHAVPILLDWLVTRT..ALYRPVSNRDLTQETPFVRLG.....VHLLIAVLHIGSCATAGA....IGS
α13
330 340 350 360 370 380
T. brucei. brucei AIAGHRGEY....WGEIIGVATASFLFER..ISAMFLDYKRAG..RIGSKSEVERGRSRSHSHGKDDVKG
L. martiniquensis AVAGERGEY....WGNIAGVMAGPAAAVQLWTAMRKARMLRPR..RPGEAARGHRHRHSASEGLKAAPAT
P. hertigi AAAGERGEY....WGNIAGVIAGPAAAAQLWATIRKVRMLRPR..RTSG.TKSHRHRSSTSEVSKTAAAT
A. deanei LVGGENGAY....WGELVGFAYGV.....TVFNLLRTYEKR..KPEVKDRRIKNRCPTAPQLIF....
T. conorhini AAAGARGEY....WGEIAGLSLAPVAFAQFVVAMRTTRRRRASGTRPSERTRSSSGGGRSKERGASRGSG
N. designis RLGGP.AEY....FGELIGAIAA.....TAILTARLKKAS..SAAVRS.....
Perkinsela species LYAAREGEY....WGEVI.....ASVLWVTTVRKC..IACE.....
E. gymnastica ATOM36A RVEPVFGTY....WIELLGAALLK.....GAVVKALCKALA..LPPPPEETQKDADSAQPPTQLSPHP
E. gymnastica ATOM36B R.ASPQGTYLAASLGDLWFRFLVPPRTQQRVVAYVALRLGCSR..GNGGSTETSLKQCASHRETAFRGNV

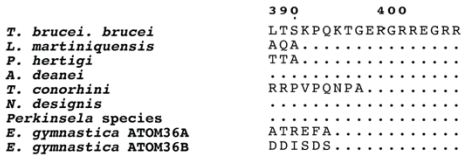


Figure S5.3. Sequence alignment of ATOM36 homologues throughout euglenozoa. Sequence alignment was generated using MUSCLE (Edgar, 2004) and visualized using the ESPRIPRT 3.0 server (https://espript.ibcp.fr, (Robert & Gouet, 2014)). Secondary structure representation shown on top based on *T. brucei brucei* ATOM36 AlphaFold-predicted model.

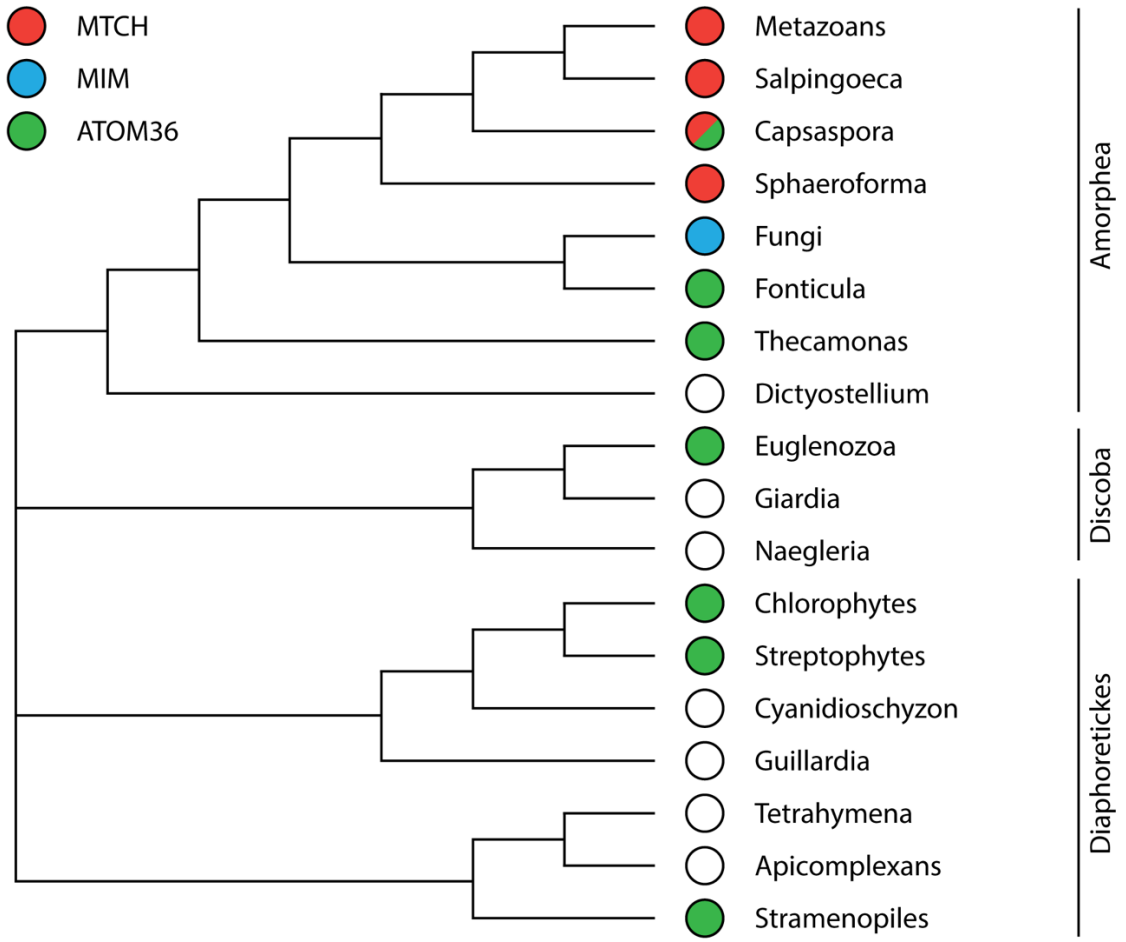


Figure S5.4. Analysis of OMM insertase distribution across a eukaryote wide species tree. A representative eukaryotic species tree is shown with the presence of OMM insertase families in each lineage indicated with a colored circle. The presence of OMM insertase is based on queries to UniProt for sequences annotated with the following PANTHER or Pfam families: PTHR10780 for MTCH2, PTHR36074 for plant ATOM36, PF19224 for trypanosome ATOM36, PF08219 for MIM1.

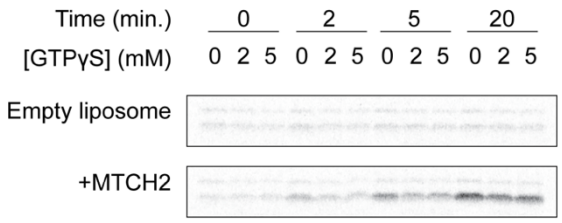


Figure S5.5. Gel quantified in Figure 5.3 panel D. For each lane, either empty or MTCH2-containing liposomes were indicated with *in vitro* translated, ^{35}S-labelled OMP25 in the presence of 2 or 5 mM GTPγS at 0, 2, 5, and 20 minute time points. The resulting sample was digested with proteinase K and the remaining 6xHis-tagged protected fragment was purified with NiNTA-agarose. To quantify the amount of inserted protein, samples were ran on SDS-PAGE and imaged via autoradiography.

BIBLIOGRAPHY

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 2), 213-221. https://doi.org/10.1107/S0907444909052925

Aksu, M., Pleiner, T., Karaca, S., Kappert, C., Dehne, H. J., Seibel, K., Urlaub, H., Bohnsack, M. T., & Görlich, D. (2018). Xpo7 is a broad-spectrum exportin and a nuclear import receptor. *J Cell Biol*, *217*(7), 2329-2340. https://rupress.org/jcb/article/217/7/2329/39132

Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A., & Stevens, R. C. (2008). Microscale fluorescent thermal stability assay for membrane proteins. *Structure*, *16*(3), 351-359. https://doi.org/10.1016/j.str.2008.02.004

Allen, M., Kachadoorian, M., Carrasquillo, M. M., Karhade, A., Manly, L., Burgess, J. D., Wang, C., Serie, D., Wang, X., Siuda, J., Zou, F., Chai, H. S., Younkin, C., Crook, J., Medway, C., Nguyen, T., Ma, L., Malphrus, K., Lincoln, S., . . . Ertekin-Taner, N. (2015). Late-onset Alzheimer disease risk variants mark brain regulatory loci. *Neurol Genet*, *1*(2), e15. https://doi.org/10.1212/NXG.0000000000000012

Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., Arumugam, M., Bork, P., & Hurt, E. (2011). Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, *146*(2), 277-289. https://doi.org/10.1016/j.cell.2011.06.039

Bai, L., You, Q., Feng, X., Kovach, A., & Li, H. (2020). Structure of the ER membrane complex, a transmembrane-domain insertase. *Nature*, *584*(7821), 475-478. https://doi.org/10.1038/s41586-020-2389-3

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng*, *6*(1), 37-40. https://doi.org/10.1093/protein/6.1.37

Becker, T., Pfannschmidt, S., Guiard, B., Stojanovski, D., Milenkovic, D., Kutik, S., Pfanner, N., Meisinger, C., & Wiedemann, N. (2008). Biogenesis of the mitochondrial TOM complex: Mim1 promotes insertion and assembly of signal-anchored receptors. *J Biol Chem*, *283*(1), 120-127. https://doi.org/10.1074/jbc.M706997200

Beckett, D., Kovaleva, E., & Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci*, *8*(4), 921-929. https://doi.org/10.1110/ps.8.4.921

Bender, C. E., Fitzgerald, P., Tait, S. W., Llambi, F., McStay, G. P., Tupper, D. O., Pellettieri, J., Sánchez Alvarado, A., Salvesen, G. S., & Green, D. R. (2012). Mitochondrial pathway of apoptosis is ancestral in metazoans. *Proc Natl Acad Sci U S A*, *109*(13), 4904-4909. https://doi.org/10.1073/pnas.1120680109

Benito, A., Grillot, D., Nuñez, G., & Fernández-Luna, J. L. (1995). Regulation and Function of Bcl-2 During Differentiation-Induced Cell Death in HL-60 Promyelocytic Cells. *American Journal of Pathology*, *146*, 481-490.

Bhattacharyya, M., Stratton, M. M., Going, C. C., McSpadden, E. D., Huang, Y., Susa, A. C., Elleman, A., Cao, Y. M., Pappireddi, N., Burkhardt, P., Gee, C. L., Barros, T., Schulman, H., Williams, E. R., & Kuriyan, J. (2016). Molecular mechanism of activation-triggered subunit exchange in Ca(2+)/calmodulin-dependent protein kinase II. *Elife*, *5*, e13405. https://doi.org/10.7554/eLife.13405

Bose, A., & Beal, M. F. (2016). Mitochondrial dysfunction in Parkinson's disease. *J Neurochem*, *139 Suppl 1*, 216-231. https://doi.org/10.1111/jnc.13731

Brambillasca, S., Yabal, M., Soffientini, P., Stefanovic, S., Makarow, M., Hegde, R. S., & Borgese, N. (2005). Transmembrane topogenesis of a tail-anchored protein is modulated by membrane lipid composition. *EMBO J*, *24*(14), 2533-2542. https://doi.org/10.1038/sj.emboj.7600730

Braun, M. B., Traenkle, B., Koch, P. A., Emele, F., Weiss, F., Poetz, O., Stehle, T., & Rothbauer, U. (2016). Peptides in headlock–a novel high-affinity and versatile peptide-binding nanobody for proteomics and microscopy. *Scientific reports*, *6*(1), 19211. https://www.nature.com/articles/srep19211

Breiman, L. (2001). Random Forests. *Machine Learning*.

Busch, J. D., Fielden, L. F., Pfanner, N., & Wiedemann, N. (2023). Mitochondrial protein transport: Versatility of translocases and mechanisms. *Mol Cell*, *83*(6), 890-910. https://doi.org/10.1016/j.molcel.2023.02.020

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol*, *23*(1), 102-107. https://doi.org/10.1038/nbt1044

Canaj, H., Hussmann, J. A., Li, H., Beckman, K. A., Weissman, J. S., & Leonetti, M. D. (2019). Deep profiling reveals substantial heterogeneity of integration outcomes in CRISPR knock-in experiments. *Biorxiv*.

Chaudhary, S., Pak, J. E., Gruswitz, F., Sharma, V., & Stroud, R. M. (2012). Overexpressing human membrane proteins in stably transfected and clonal human embryonic kidney 293S cells. *Nat Protoc*, *7*(3), 453-466. https://www.nature.com/articles/nprot.2011.453

Chen, J., Fruhauf, A., Fan, C., Ponce, J., Ueberheide, B., Bhabha, G., & Ekiert, D. (2022). Structure of an endogenous mycobacterial MCE lipid transporter. *bioRxiv*, *doi: 10.1101/2022.12.08.519548.*

Chen, J. J., Nathaniel, D. L., Raghavan, P., Nelson, M., Tian, R., Tse, E., Hong, J. Y., See, S. K., Mok, S. A., Hein, M. Y., Southworth, D. R., Grinberg, L. T., Gestwicki, J. E., Leonetti, M. D., & Kampmann, M. (2019). Compromised function of the ESCRT pathway promotes endolysosomal escape of tau seeds and propagation of tau aggregation. *J Biol Chem*, *294*(50), 18952-18966. https://doi.org/10.1074/jbc.RA119.009432

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc*, *124*(31), 9026-9027. https://doi.org/10.1021/ja027007w

Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., & Hegde, R. S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*, *175*(6), 1507-1519.e16. https://doi.org/10.1016/j.cell.2018.10.009

Cho, N. H., Cheveralls, K. C., Brunner, A.-D., Kim, K., Michaelis, A. C., Raghavan, P., Kobayashi, H., Savy, L., Li, J. Y., & Canaj, H. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science*, *375*(6585), eabi6983. https://www.science.org/doi/full/10.1126/science.abi6983

Coukos, R., Yao, D., Sanchez, M. I., Strand, E. T., Olive, M. E., Udeshi, N. D., Weissman, J. S., Carr, S. A., Bassik, M. C., & Ting, A. Y. (2021). An engineered transcriptional reporter of protein localization identifies regulators of mitochondrial and ER membrane protein trafficking in high-throughput CRISPRi screens. *Elife*, *10*, e69142. https://doi.org/10.7554/eLife.69142

de Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C., & Ryan, M. D. (2006). E unum pluribus: multiple proteins from a self-processing polyprotein. *Trends Biotechnol*, *24*(2), 68-75. https://doi.org/10.1016/j.tibtech.2005.12.006

De Genst, E. J., Guilliams, T., Wellens, J., O'day, E. M., Waudby, C. A., Meehan, S., Dumoulin, M., Hsu, S.-T. D., Cremades, N., & Verschueren, K. H. G. (2010). Structure and properties of a complex of α-synuclein and a single-domain camelid antibody. *Journal of Molecular Biology*, *402*(2), 326-343. https://www.sciencedirect.com/science/article/pii/S0022283610007424

Dimmer, K. S., Papić, D., Schumann, B., Sperl, D., Krumpe, K., Walther, D. M., & Rapaport, D. (2012). A crucial role for Mim2 in the biogenesis of mitochondrial outer membrane proteins. *J Cell Sci*, *125*(Pt 14), 3464-3473. https://doi.org/10.1242/jcs.103804

Doan, K. N., Grevel, A., Mårtensson, C. U., Ellenrieder, L., Thornton, N., Wenz, L. S., Opaliński, Ł., Guiard, B., Pfanner, N., & Becker, T. (2020). The Mitochondrial Import Complex MIM Functions as Main Translocase for α-Helical Outer Membrane Proteins. *Cell Rep*, *31*(4), 107567. https://doi.org/10.1016/j.celrep.2020.107567

Dobson, L., Reményi, I., & Tusnády, G. E. (2015). The human transmembrane proteome. *Biol Direct*, *10*, 31. https://doi.org/10.1186/s13062-015-0061-x

Duncan, O., Taylor, N. L., Carrie, C., Eubel, H., Kubiszewski-Jakubiak, S., Zhang, B., Narsai, R., Millar, A. H., & Whelan, J. (2011). Multiple lines of evidence localize signaling, morphology, and lipid biosynthesis machinery to the mitochondrial outer membrane of Arabidopsis. *Plant Physiol*, *157*(3), 1093-1113. https://doi.org/10.1104/pp.111.183160

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol*, *7*(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340

Elegheert, J., Behiels, E., Bishop, B., Scott, S., Woolley, R. E., Griffiths, S. C., Byrne, E. F. X., Chang, V. T., Stuart, D. I., & Jones, E. Y. (2018). Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat Protoc*, *13*(12), 2991-3017. https://www.nature.com/articles/s41596-018-0075-9

Escott-Price, V., Bellenguez, C., Wang, L. S., Choi, S. H., Harold, D., Jones, L., Holmans, P., Gerrish, A., Vedernikov, A., Richards, A., DeStefano, A. L., Lambert,

J. C., Ibrahim-Verbaas, C. A., Naj, A. C., Sims, R., Jun, G., Bis, J. C., Beecham, G. W., Grenier-Boley, B., . . . Cardiovascular Health Study, C. H. S. (2014). Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PLoS One*, *9*(6), e94661. https://doi.org/10.1371/journal.pone.0094661

Fairhead, M., & Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. *Methods Mol Biol*, *1266*, 171-184. https://doi.org/10.1007/978-1-4939-2272-7_12

Frey, S., & Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J Chromatogr A*, *1337*, 95-105. https://doi.org/10.1016/j.chroma.2014.02.029

Frey, S., & Görlich, D. (2015). The Xenopus laevis Atg4B protease: Insights into substrate recognition and application for tag removal from proteins expressed in pro-and eukaryotic hosts. *PLoS One*, *10*(4), e0125099. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0125099

Fridy, P. C., Li, Y., Keegan, S., Thompson, M. K., Nudelman, I., Scheid, J. F., Oeffinger, M., Nussenzweig, M. C., Fenyö, D., & Chait, B. T. (2014). A robust pipeline for rapid production of versatile nanobody repertoires. *Nature Methods*, *11*(12), 1253-1260. https://www.nature.com/articles/nmeth.3170

Friedman, J. R., & Nunnari, J. (2014). Mitochondrial form and function. *Nature*, *505*(7483), 335-343. https://doi.org/10.1038/nature12985

Fry, M. Y., Saladi, S. M., Cunha, A., & Clemons, W. M. (2021). Sequence-based features that are determinant for tail-anchored membrane protein sorting in eukaryotes. *Traffic*, *22*(9), 306-318. https://doi.org/10.1111/tra.12809

Fukasawa, Y., Oda, T., Tomii, K., & Imai, K. (2017). Origin and Evolutionary Alteration of the Mitochondrial Import System in Eukaryotic Lineages. *Mol Biol Evol*, *34*(7), 1574-1586. https://doi.org/10.1093/molbev/msx096

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., Guimaraes, C., Panning, B., Ploegh, H. L., Bassik, M. C., Qi, L. S., Kampmann, M., & Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647-661. https://doi.org/10.1016/j.cell.2014.09.029

Go, C. D., Knight, J. D. R., Rajasekharan, A., Rathod, B., Hesketh, G. G., Abe, K. T., Youn, J. Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L. Y., Popiel, E., Lambert, J. P., Coyaud, É., Cheung, S. W. T., Rajendran, D., Wong, C. J., Antonicka, H., Pelletier, L., Palazzo, A. F., . . . Gingras, A. C. (2021). A proximity-dependent biotinylation map of a human cell. *Nature*, *595*(7865), 120-124. https://doi.org/10.1038/s41586-021-03592-2

Goehring, A., Lee, C.-H., Wang, K. H., Michel, J. C., Claxton, D. P., Baconguis, I., Althoff, T., Fischer, S., Garcia, K. C., & Gouaux, E. (2014). Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat Protoc*, *9*(11), 2574-2585. https://www.nature.com/articles/nprot.2014.173

Gok, M. O., & Friedman, J. R. (2022). The outer mitochondrial membrane protein TMEM11 is a novel negative regulator of BNIP3/BNIP3L-dependent receptor-mediated mitophagy. *Biorxiv*.

Götzke, H., Kilisch, M., Martínez-Carranza, M., Sograte-Idrissi, S., Rajavel, A., Schlichthaerle, T., Engels, N., Jungmann, R., Stenmark, P., Opazo, F., & Frey, S. (2019). The ALFA-tag is a highly versatile tool for nanobody-based bioscience applications. *Nat Commun*, *10*(1), 4403. https://doi.org/10.1038/s41467-019-12301-7

Guna, A., & Hegde, R. S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Curr Biol*, *28*(8), R498-R511. https://doi.org/10.1016/j.cub.2018.02.004

Guna, A., Stevens, T. A., Inglis, A. J., Replogle, J. M., Esantsi, T. K., Muthukumar, G., Shaffer, K. C. L., Wang, M. L., Pogson, A. N., Jones, J. J., Lomenick, B., Chou, T. F., Weissman, J. S., & Voorhees, R. M. (2022). MTCH2 is a mitochondrial outer membrane protein insertase. *Science*, *378*(6617), 317-322. https://doi.org/10.1126/science.add1856

Guna, A., Volkmar, N., Christianson, J. C., & Hegde, R. S. (2018). The ER membrane protein complex is a transmembrane domain insertase. *Science*, *359*(6374), 470-473. https://doi.org/10.1126/science.aao3099

Hammond, M., Dorrell, R. G., Speijer, D., & Lukeš, J. (2022). Eukaryotic cellular intricacies shape mitochondrial proteomic complexity. *Bioessays*, *44*(5), e2100258. https://doi.org/10.1002/bies.202100258

Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A., & Opperdoes, F. R. (2003). Plant-like traits associated with metabolism of Trypanosoma parasites. *Proc Natl Acad Sci U S A*, *100*(3), 1067-1071. https://doi.org/10.1073/pnas.0335769100

Hanson, P. I., Meyer, T., Stryer, L., & Schulman, H. (1994). Dual role of calmodulin in autophosphorylation of multifunctional CaM kinase may underlie decoding of calcium signals. *Neuron*, *12*(5), 943-956. https://doi.org/10.1016/0896-6273(94)90306-9

Hayashi, A., Da-Qiao, D., Tsutsumi, C., Chikashige, Y., Masuda, H., Haraguchi, T., & Hiraoka, Y. (2009). Localization of gene products using a chromosomally tagged GFP-fusion library in the fission yeast Schizosaccharomyces pombe. *Genes to cells*, *14*(2), 217-225. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2443.2008.01264.x

Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., & Buchholz, F. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*, *163*(3), 712-723. https://www.sciencedirect.com/science/article/pii/S0092867415012702

Hoelz, A., Nairn, A. C., & Kuriyan, J. (2003). Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol Cell*, *11*(5), 1241-1251. https://doi.org/10.1016/s1097-2765(03)00171-0

Holder, T., Basquin, C., Ebert, J., Randel, N., Jollivet, D., Conti, E., Jékely, G., & Bono, F. (2013). Deep transcriptome-sequencing and proteome analysis of the hydrothermal vent annelid Alvinella pompejana identifies the CvP-bias as a robust measure of eukaryotic thermostability. *Biol Direct*, *8*, 2. https://doi.org/10.1186/1745-6150-8-2

Hook, S. S., & Means, A. R. (2001). Ca(2+)/CaM-dependent kinases: from activation to function. *Annu Rev Pharmacol Toxicol*, *41*, 471-505. https://doi.org/10.1146/annurev.pharmtox.41.1.471

Horlbeck, M. A., Gilbert, L. A., Villalta, J. E., Kampmann, M., & Weissman, J. S. (2016). Compact and highly active next- generation libraries for CRISPR-mediated gene repression and activation. *Elife*.

Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, *425*(6959), 686-691. https://www.nature.com/articles/nature02026

Hulett, J. M., Lueder, F., Chan, N. C., Perry, A. J., Wolynec, P., Likić, V. A., Gooley, P. R., & Lithgow, T. (2008). The transmembrane segment of Tom20 is recognized by Mim1 for docking to the mitochondrial TOM complex. *J Mol Biol*, *376*(3), 694-704. https://doi.org/10.1016/j.jmb.2007.12.021

Hung, V., Zou, P., Rhee, H. W., Udeshi, N. D., Cracan, V., Svinkina, T., Carr, S. A., Mootha, V. K., & Ting, A. Y. (2014). Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell*, *55*(2), 332-341. https://doi.org/10.1016/j.molcel.2014.06.003

Inglis, A. J., Page, K. R., Guna, A., & Voorhees, R. M. (2020). Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex. *Cell Rep*, *30*(11), 3691-3698.e5. https://doi.org/10.1016/j.celrep.2020.02.084

Itakura, E., Zavodszky, E., Shao, S., Wohlever, M. L., Keenan, R. J., & Hegde, R. S. (2016). Ubiquilins Chaperone and Triage Mitochondrial Membrane Proteins for Degradation. *Mol Cell*, *63*(1), 21-33. https://doi.org/10.1016/j.molcel.2016.05.020

Jost, M., Chen, Y., Gilbert, L. A., Horlbeck, M. A., Krenning, L., Menchon, G., Rai, A., Cho, M. Y., Stern, J. J., Prota, A. E., Kampmann, M., Akhmanova, A., Steinmetz, M. O., Tanenbaum, M. E., & Weissman, J. S. (2017). Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib Is a Microtubule-Destabilizing Agent. *Mol Cell*, *68*(1), 210-223.e6. https://doi.org/10.1016/j.molcel.2017.09.012

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Juszkiewicz, S., & Hegde, R. S. (2018). Quality Control of Orphaned Proteins. *Mol Cell*, *71*(3), 443-457. https://doi.org/10.1016/j.molcel.2018.07.001

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 2), 125-132. https://doi.org/10.1107/S0907444909047337

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., Ishikawa, H., Leonetti, M. D., Marshall, W. F., Weissman, J. S., & Huang, B. (2016). Versatile protein tagging in cells with split fluorescent protein. *Nat Commun*, *7*, 11046. https://doi.org/10.1038/ncomms11046

Kanarek, N., Keys, H. R., Cantor, J. R., Lewis, C. A., Chan, S. H., Kunchok, T., Abu-Remaileh, M., Freinkman, E., Schweitzer, L. D., & Sabatini, D. M. (2018).

Histidine catabolism is a major determinant of methotrexate sensitivity. *Nature*, *559*(7715), 632-636. https://doi.org/10.1038/s41586-018-0316-7

Kandel, E. R., Dudai, Y., & Mayford, M. R. (2014). The molecular and systems biology of memory. *Cell*, *157*(1), 163-186. https://doi.org/10.1016/j.cell.2014.03.001

Karch, C. M., Ezerskiy, L. A., Bertelsen, S., Alzheimer's Disease Genetics Consortium, A. D. G. C., & Goate, A. M. (2016). Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. *PLoS One*, *11*(2), e0148717. https://doi.org/10.1371/journal.pone.0148717

Käser, S., Oeljeklaus, S., Týč, J., Vaughan, S., Warscheid, B., & Schneider, A. (2016). Outer membrane protein functions as integrator of protein import and DNA inheritance in mitochondria. *Proc Natl Acad Sci U S A*, *113*(31), E4467-75. https://doi.org/10.1073/pnas.1605497113

Kawate, T., & Gouaux, E. (2006). Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure*, *14*(4), 673-681. https://www.sciencedirect.com/science/article/pii/S0969212606001092

Kim, S. W., Cha, S. S., Cho, H. S., Kim, J. S., Ha, N. C., Cho, M. J., Joo, S., Kim, K. K., Choi, K. Y., & Oh, B. H. (1997). High-resolution crystal structures of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue. *Biochemistry*, *36*(46), 14030-14036. https://doi.org/10.1021/bi971546+

Kirchhofer, A., Helma, J., Schmidthals, K., Frauer, C., Cui, S., Karcher, A., Pellis, M., Muyldermans, S., Casas-Delucchi, C. S., Cardoso, M. C., Leonhardt, H., Hopfner, K. P., & Rothbauer, U. (2010). Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol*, *17*(1), 133-138. https://doi.org/10.1038/nsmb.1727

Koch, B., Nijmeijer, B., Kueblbeck, M., Cai, Y., Walther, N., & Ellenberg, J. (2018). Generation and validation of homozygous fluorescent knock-in cells using CRISPR–Cas9 genome editing. *Nat Protoc*, *13*(6), 1465-1487. https://www.nature.com/articles/nprot.2018.042

Kubala, M. H., Kovtun, O., Alexandrov, K., & Collins, B. M. (2010). Structural and thermodynamic analysis of the GFP: GFP-nanobody complex. *Protein Sci*, *19*(12), 2389-2401. https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.519

Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., Sugano, Y., Mori, T., Dohmae, N., Hirata, K., Nakada-Nakura, Y., Maturana, A. D., Tanaka, Y., Mori, H., Sugita, Y., Arisaka, F., Ito, K., Ishitani, R., Tsukazaki, T., & Nureki, O. (2014). Structural basis of Sec-independent membrane protein insertion by YidC. *Nature*, *509*(7501), 516-520. https://doi.org/10.1038/nature13167

Labbé, K., Mookerjee, S., Le Vasseur, M., Gibbs, E., Lerner, C., & Nunnari, J. (2021). The modified mitochondrial outer membrane carrier MTCH2 links mitochondrial fusion to lipogenesis. *J Cell Biol*, *220*(11), e202103122. https://doi.org/10.1083/jcb.202103122

Le Vasseur, M., Friedman, J., Jost, M., Xu, J., Yamada, J., Kampmann, M., Horlbeck, M. A., Salemi, M. R., Phinney, B. S., Weissman, J. S., & Nunnari, J. (2021). Genome-wide CRISPRi screening identifies OCIAD1 as a prohibitin client and regulatory determinant of mitochondrial Complex III assembly in human cells. *Elife*, *10*, e67624. https://doi.org/10.7554/eLife.67624

Lee, S. C., Knowles, T. J., Postis, V. L. G., Jamshad, M., Parslow, R. A., Lin, Y.-P., Goldman, A., Sridhar, P., Overduin, M., & Muench, S. P. (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat Protoc*, *11*(7), 1149-1162. https://www.nature.com/articles/nprot.2016.070

Lee, Y., Willers, C., Kunji, E. R., & Crichton, P. G. (2015). Uncoupling protein 1 binds one nucleotide per monomer and is stabilized by tightly bound cardiolipin. *Proc Natl Acad Sci U S A*, *112*(22), 6973-6978. https://doi.org/10.1073/pnas.1503833112

Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S., & Huang, B. (2016). A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc Natl Acad Sci U S A*, *113*(25), E3501-E3508. https://www.pnas.org/doi/full/10.1073/pnas.1606731113

Lin, D. H., Stuwe, T., Schilbach, S., Rundlet, E. J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F. M., Collins, L. N., Davenport, A. M., Jeon, Y. E., & Hoelz, A. (2016). Architecture of the symmetric core of the nuclear pore. *Science*, *352*(6283), aaf1015. https://doi.org/10.1126/science.aaf1015

Lindborg, M., Dubnovitsky, A., Olesen, K., Björkman, T., Abrahmsén, L., Feldwisch, J., & Härd, T. (2013). High-affinity binding to staphylococcal protein A by an engineered dimeric Affibody molecule. *Protein Engineering, Design & Selection*, *26*(10), 635-644. https://academic.oup.com/peds/article/26/10/635/1511857

Lisman, J., & Raghavachari, S. (2015). Biochemical principles underlying the stable maintenance of LTP by the CaMKII/NMDAR complex. *Brain Res*, *1621*, 51-61. https://doi.org/10.1016/j.brainres.2014.12.010

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic and behavioural memory. *Nat Rev Neurosci*, *3*, 175-190. https://doi.org/10.1038/nrn753

Liu, L., Spurrier, J., Butt, T. R., & Strickler, J. E. (2008). Enhanced protein expression in the baculovirus/insect cell system using engineered SUMO fusions. *Protein Expr Purif*, *62*(1), 21-28. https://doi.org/10.1016/j.pep.2008.07.010

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R. S., & Keenan, R. J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature*, *477*(7362), 61-66. https://doi.org/10.1038/nature10362

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr*, *40*(Pt 4), 658-674. https://doi.org/10.1107/S0021889807021206

McKenna, M. J., Sim, S. I., Ordureau, A., Wei, L., Harper, J. W., Shao, S., & Park, E. (2020). The endoplasmic reticulum P5A-ATPase is a transmembrane helix dislocase. *Science*, *369*(6511), eabc5809. https://doi.org/10.1126/science.abc5809

McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J. A., Valleriani, A., & Selbach, M. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell*, *167*(3), 803-815.e21. https://doi.org/10.1016/j.cell.2016.09.015

McSpadden, E. D. (2018). *Structural and evolutionary studies on CaMKII oligomerization*. University of California, Berkeley].

McSpadden, E. D., Xia, Z., Chi, C. C., Susa, A. C., Shah, N. H., Gee, C. L., Williams, E. R., & Kuriyan, J. (2019). Variation in assembly stoichiometry in non-metazoan

homologs of the hub domain of Ca^{2+} /calmodulin-dependent protein kinase II. *Protein Sci*, *28*(6), 1071-1082. https://doi.org/10.1002/pro.3614

Miller-Vedam, L. E., Bräuning, B., Popova, K. D., Schirle Oakdale, N. T., Bonnar, J. L., Prabu, J. R., Boydston, E. A., Sevillano, N., Shurtleff, M. J., Stroud, R. M., Craik, C. S., Schulman, B. A., Frost, A., & Weissman, J. S. (2020). Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients. *Elife*, *9*, e62611. https://doi.org/10.7554/eLife.62611

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res*, *49*(D1), D412-D419. https://doi.org/10.1093/nar/gkaa913

Morgenstern, M., Peikert, C. D., Lübbert, P., Suppanz, I., Klemm, C., Alka, O., Steiert, C., Naumenko, N., Schendzielorz, A., Melchionda, L., Mühlhäuser, W. W. D., Knapp, B., Busch, J. D., Stiller, S. B., Dannenmaier, S., Lindau, C., Licheva, M., Eickhorst, C., Galbusera, R., . . . Warscheid, B. (2021). Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metab*, *33*(12), 2464-2483.e18. https://doi.org/10.1016/j.cmet.2021.11.001

Mossessova, E., & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell*, *5*, 865-876. https://doi.org/10.1016/s1097-2765(00)80326-3

Nilsen, T. W. (2013). Preparation of nuclear extracts from HeLa cells. *Cold Spring Harbor Protocols*, *2013*(6), pdb. prot075176. http://cshprotocols.cshlp.org/content/2013/6/pdb.prot075176.short

O'Gorman, S., Fox, D. T., & Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*, *251*(4999), 1351-1355. https://www.science.org/doi/pdf/10.1126/science.1900642

Palmieri, F. (2013). The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol Aspects Med*, *34*(2-3), 465-484. https://doi.org/10.1016/j.mam.2012.05.005

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, *24*(1), 79-88. https://www.nature.com/articles/nbt1172

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, *30*(1), 70-82. https://doi.org/10.1002/pro.3943

Pleiner, T., Bates, M., & Görlich, D. (2018). A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. *J Cell Biol*, *217*(3), 1143-1154. https://doi.org/10.1083/jcb.201709115

Pleiner, T., Bates, M., Trakhanov, S., Lee, C. T., Schliep, J. E., Chug, H., Böhning, M., Stark, H., Urlaub, H., & Görlich, D. (2015). Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *Elife*, *4*, e11349. https://doi.org/10.7554/eLife.11349

Pleiner, T., Hazu, M., Tomaleri, G. P., Januszyk, K., Oania, R. S., Sweredoski, M. J., Moradian, A., Guna, A., & Voorhees, R. M. (2021). WNK1 is an assembly factor for the human ER membrane protein complex. *Mol Cell*, *81*(13), 2693-2704.e12. https://doi.org/10.1016/j.molcel.2021.04.013

Pleiner, T., Tomaleri, G. P., Januszyk, K., Inglis, A. J., Hazu, M., & Voorhees, R. M. (2020). Structural basis for membrane insertion by the human ER membrane protein complex. *Science*, *369*(6502), 433-436. https://doi.org/10.1126/science.abb5008

Poser, I., Sarov, M., Hutchins, J. R. A., Hériché, J.-K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., & Bird, A. W. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods*, *5*(5), 409-415. https://www.nature.com/articles/nmeth.1199

Pusnik, M., Mani, J., Schmidt, O., Niemann, M., Oeljeklaus, S., Schnarwiler, F., Warscheid, B., Lithgow, T., Meisinger, C., & Schneider, A. (2012). An essential novel component of the noncanonical mitochondrial outer membrane protein import system of trypanosomatids. *Mol Biol Cell*, *23*(17), 3420-3428. https://doi.org/10.1091/mbc.E12-02-0107

Qin, J. Y., Zhang, L., Clift, K. L., Hulur, I., Xiang, A. P., Ren, B.-Z., & Lahn, B. T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One*, *5*(5), e10611. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010611

Qiu, H., Yoon, H. S., & Bhattacharya, D. (2013). Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front Plant Sci*, *4*, 366. https://doi.org/10.3389/fpls.2013.00366

Rao, M., Okreglak, V., Chio, U. S., Cho, H., Walter, P., & Shan, S. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *Elife*.

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T. J., Goodman, R. P., Grabarek, Z., Haas, M. E., Hung, W. H. W., Joshi, P. R., Jourdain, A. A., Kim, S. H., Kotrys, A. V., Lam, S. S., McCoy, J. G., Meisel, J. D., Miranda, M., Panda, A., . . . Mootha, V. K. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res*, *49*(D1), D1541-D1547. https://doi.org/10.1093/nar/gkaa1011

Rellos, P., Pike, A. C., Niesen, F. H., Salah, E., Lee, W. H., von Delft, F., & Knapp, S. (2010). Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol*, *8*(7), e1000426. https://doi.org/10.1371/journal.pbio.1000426

Replogle, J. M., Norman, T. M., Xu, A., Hussmann, J. A., Chen, J., Cogan, J. Z., Meer, E. J., Terry, J. M., Riordan, D. P., Srinivas, N., Fiddes, I. T., Arthur, J. G., Alvarado, L. J., Pfeiffer, K. A., Mikkelsen, T. S., Weissman, J. S., & Adamson, B. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol*, *38*(8), 954-961. https://doi.org/10.1038/s41587-020-0470-y

Replogle, J. M., Saunders, R. A., Pogson, A. N., Hussmann, J. A., Lenail, A., Guna, A., Mascibroda, L., Wagner, E. J., Adelman, K., Lithwick-Yanai, G., Iremadze, N.,

Oberstrass, F., Lipson, D., Bonnar, J. L., Jost, M., Norman, T. M., & Weissman, J. S. (2022). Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, *185*(14), 2559-2575.e28. https://doi.org/10.1016/j.cell.2022.05.013

Rich, R. C., & Schulman, H. (1998). Substrate-directed function of calmodulin in autophosphorylation of Ca2+/calmodulin-dependent protein kinase II. *J Biol Chem*, *273*(43), 28424-28429. https://doi.org/10.1074/jbc.273.43.28424

Richter-Dennerlein, R., Korwitz, A., Haag, M., Tatsuta, T., Dargazanli, S., Baker, M., Decker, T., Lamkemeyer, T., Rugarli, E. I., & Langer, T. (2014). DNAJC19, a mitochondrial cochaperone associated with cardiomyopathy, forms a complex with prohibitins to regulate cardiolipin remodeling. *Cell Metab*, *20*(1), 158-171. https://doi.org/10.1016/j.cmet.2014.04.016

Riggs, P. (2000). Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Molecular Biotechnology*.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*, *42*(Web Server issue), W320-4. https://doi.org/10.1093/nar/gku316

Roger, A. J., Muñoz-Gómez, S. A., & Kamikawa, R. (2017). The Origin and Diversification of Mitochondria. *Curr Biol*, *27*(21), R1177-R1192. https://doi.org/10.1016/j.cub.2017.09.015

Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., & Cardoso, M. C. (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods*, *3*(11), 887-889. https://www.nature.com/articles/nmeth953

Rottiers, V., Francisco, A., Platov, M., Zaltsman, Y., Ruggiero, A., Lee, S. S., Gross, A., & Libert, S. (2017). MTCH2 is a conserved regulator of lipid homeostasis. *Obesity (Silver Spring)*, *25*(3), 616-625. https://doi.org/10.1002/oby.21751

Ruprecht, J. J., & Kunji, E. R. S. (2020). The SLC25 Mitochondrial Carrier Family: Structure and Mechanism. *Trends Biochem Sci*, *45*(3), 244-258. https://doi.org/10.1016/j.tibs.2019.11.001

Sarov, M., Barz, C., Jambor, H., Hein, M. Y., Schmied, C., Suchold, D., Stender, B., Janosch, S., Kj, V. V., & Krishnan, R. T. (2016). A genome-wide resource for the analysis of protein localisation in Drosophila. *Elife*, *5*, e12068. https://elifesciences.org/articles/12068

Sarov, M., Murray, J. I., Schanze, K., Pozniakovski, A., Niu, W., Angermann, K., Hasse, S., Rupprecht, M., Vinis, E., & Tinney, M. (2012). A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. *Cell*, *150*(4), 855-866. https://www.sciencedirect.com/science/article/pii/S0092867412009464

Schatz, P. J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y)*, *11*(10), 1138-1143. https://www.nature.com/articles/nbt1093-1138

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-

source platform for biological-image analysis. *Nat Methods*, *9*(7), 676-682. https://doi.org/10.1038/nmeth.2019

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., Schwappach, B., & Weissman, J. S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell*, *134*(4), 634-645. https://doi.org/10.1016/j.cell.2008.06.025

Seddon, A. M., Curnow, P., & Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1666*(1-2), 105-117. https://www.sciencedirect.com/science/article/pii/S0005273604001610

Setoguchi, K., Otera, H., & Mihara, K. (2006). Cytosolic factor- and TOM-independent import of C-tail-anchored mitochondrial outer membrane proteins. *EMBO J*, *25*(24), 5635-5647. https://doi.org/10.1038/sj.emboj.7601438

Shao, S., & Hegde, R. S. (2011). Membrane protein insertion at the endoplasmic reticulum. *Annu Rev Cell Dev Biol*, *27*, 25-56. https://doi.org/10.1146/annurev-cellbio-092910-154125

Sharma, A., Mariappan, M., Appathurai, S., & Hegde, R. S. (2010). In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol Biol*.

Shi, X., Reinstadler, B., Shah, H., To, T. L., Byrne, K., Summer, L., Calvo, S. E., Goldberger, O., Doench, J. G., Mootha, V. K., & Shen, H. (2022). Combinatorial GxGxE CRISPR screen identifies SLC25A39 in mitochondrial glutathione transport linking iron homeostasis to OXPHOS. *Nat Commun*, *13*(1), 2483. https://doi.org/10.1038/s41467-022-30126-9

Sinha, S., & Manoj, N. (2019). Molecular evolution of proteins mediating mitochondrial fission-fusion dynamics. *FEBS Lett*, *593*(7), 703-718. https://doi.org/10.1002/1873-3468.13356

Stevens, T. A., Tomaleri, G. P., Hazu, M., Wei, S., Nguyen, V. N., DeKalb, C., Voorhees, R. M., & Pleiner, T. (2023). A nanobody-based strategy for rapid and scalable purification of native human protein complexes. *BioRxiv*.

Strokappe, N. M., Hock, M., Rutten, L., Mccoy, L. E., Back, J. W., Caillat, C., Haffke, M., Weiss, R. A., Weissenhorn, W., & Verrips, T. (2019). Super potent bispecific llama VHH antibodies neutralize HIV via a combination of gp41 and gp120 epitopes. *Antibodies*, *8*(2), 38. https://www.mdpi.com/481746

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci*, *31*(1), 8-22. https://doi.org/10.1002/pro.4218

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, *22*, 4673-4680. https://doi.org/10.1093/nar/22.22.4673

Thornton, N., Stroud, D. A., Milenkovic, D., Guiard, B., Pfanner, N., & Becker, T. (2010). Two modular forms of the mitochondrial sorting and assembly machinery are involved in biogenesis of alpha-helical outer membrane proteins. *J Mol Biol*, *396*(3), 540-549. https://doi.org/10.1016/j.jmb.2009.12.026

Torres-Ocampo, A. P., Özden, C., Hommer, A., Gardella, A., Lapinskas, E., Samkutty, A., Esposito, E., Garman, S. C., & Stratton, M. M. (2020). Characterization of CaMKIIα holoenzyme stability. *Protein Sci*, *29*(6), 1524-1534. https://doi.org/10.1002/pro.3869

Traenkle, B., Segan, S., Fagbadebo, F. O., Kaiser, P. D., & Rothbauer, U. (2020). A novel epitope tagging system to visualize and monitor antigens in live cells with chromobodies. *Scientific reports*, *10*(1), 14267. https://www.nature.com/articles/s41598-020-71091-x

UniProt, C. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, *51*(D1), D523-D531. https://doi.org/10.1093/nar/gkac1052

Vassylyeva, M. N., Klyuyev, S., Vassylyev, A. D., Wesson, H., Zhang, Z., Renfrow, M. B., Wang, H., Higgins, N. P., Chow, L. T., & Vassylyev, D. G. (2017). Efficient, ultra-high-affinity chromatography in a one-step purification of complex proteins. *Proc Natl Acad Sci U S A*, *114*(26), E5138-E5147. https://www.pnas.org/doi/full/10.1073/pnas.1704872114

Vera Rodriguez, A., Frey, S., & Görlich, D. (2019). Engineered SUMO/protease system identifies Pdr6 as a bidirectional nuclear transport receptor. *J Cell Biol*, *218*(6), 2006-2020. https://doi.org/10.1083/jcb.201812091

Virant, D., Traenkle, B., Maier, J., Kaiser, P. D., Bodenhöfer, M., Schmees, C., Vojnovic, I., Pisak-Lukáts, B., Endesfelder, U., & Rothbauer, U. (2018). A peptide tag-specific nanobody enables high-quality labeling for dSTORM imaging. *Nature communications*, *9*(1), 930. https://www.nature.com/articles/s41467-018-03191-2

Vitali, D. G., Käser, S., Kolb, A., Dimmer, K. S., Schneider, A., & Rapaport, D. (2018). Independent evolution of functionally exchangeable mitochondrial outer membrane import complexes. *eLife*, *7*.

Vlahou, G., Eliás, M., Kliest-Retzow, J., Wiesner, R. J., & Rivero, F. (2011). The Ras related GTPase Miro is not required for mitochondrial transport in Dictyostelium discoideum. *European Journal of Cell Biology*, *90*(4), 341-355.

Voorhees, R. M., & Hegde, R. S. (2015). Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. *Elife*, *4*, e07975. https://elifesciences.org/articles/7975

Vyas, S., Zaganjor, E., & Haigis, M. C. (2016). Mitochondria and Cancer. *Cell*, *166*(3), 555-566. https://doi.org/10.1016/j.cell.2016.07.002

Wagner, F., Kunz, T. C., Chowdhury, S. R., Thiede, B., Fraunholz, M., Eger, D., & Kozjak-Pavlovic, V. (2019). Armadillo repeat-containing protein 1 is a dual localization protein associated with mitochondrial intermembrane space bridging complex. *PLoS One*, *14*(10), e0218303. https://doi.org/10.1371/journal.pone.0218303

Waizenegger, T., Schmitt, S., Zivkovic, J., Neupert, W., & Rapaport, D. (2005). Mim1, a protein required for the assembly of the TOM complex of mitochondria. *EMBO Rep*, *6*(1), 57-62. https://doi.org/10.1038/sj.embor.7400318

Wang, W., Zhao, F., Ma, X., Perry, G., & Zhu, X. (2020). Mitochondria dysfunction in the pathogenesis of Alzheimer's disease: recent advances. *Mol Neurodegener*, *15*(1), 30. https://doi.org/10.1186/s13024-020-00376-6

Weiner, J. H., & Li, L. (2008). Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. *Biochim Biophys Acta*, *1778*(9), 1698-1713. https://doi.org/10.1016/j.bbamem.2007.07.020

N, W., & N, P. (2017). Mitochondrial Machineries for Protein Import and Assembly. *The Annual Review of Biochemistry*.

Wu, X., Siggel, M., Ovchinnikov, S., Mi, W., Svetlov, V., Nudler, E., Liao, M., Hummer, G., & Rapoport, T. A. (2020). Structural basis of ER-associated protein degradation mediated by the Hrd1 ubiquitin ligase complex. *Science*, *368*(6489), eaaz2449. https://doi.org/10.1126/science.aaz2449

Wyatt, P. J. (1997). Multiangle Light Scattering: The Basic Tool for Macromolecular Characterization. *Instrumentation Science & Technology*, *25*(1), 1-18.

Xu, J., Kim, A.-R., Cheloha, R. W., Fischer, F. A., Li, J. S. S., Feng, Y., Stoneburner, E., Binari, R., Mohr, S. E., Zirin, J., Ploegh, H. L., & Perrimon, N. (2022). Protein visualization and manipulation in Drosophila through the use of epitope tags recognized by nanobodies. *Elife*, *11*, e74326. https://elifesciences.org/articles/74326

Zaltsman, Y., Shachnai, L., Yivgi-Ohana, N., Schwarz, M., Maryanovich, M., Houtkooper, R. H., Vaz, F. M., De Leonardis, F., Fiermonte, G., Palmieri, F., Gillissen, B., Daniel, P. T., Jimenez, E., Walsh, S., Koehler, C. M., Roy, S. S., Walter, L., Hajnóczky, G., & Gross, A. (2010). MTCH2/MIMP is a major facilitator of tBID recruitment to mitochondria. *Nat Cell Biol*, *12*(6), 553-562. https://doi.org/10.1038/ncb2057