Mechanistic and structural insights into the quality control of the MHC I antigen processing pathway

DISSERTATION

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich 14

Biochemie, Chemie und Pharmazie

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Christian Winter

aus Schwerte

Frankfurt am Main, 2022

(D30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Robert Tampé

Zweitgutachter: Prof. Dr. Klaas Martinus Pos

Datum der Disputation: 06.03.2023

I.   List of publications

1. Semisynthetic viral inhibitor for light control of the MHC I
    peptide loading complex
    C. Winter[‡], A. Domnick[‡], D. Cernova, R. Tampé
    Angewandte Chemie **2022**, e202211826

2. Molecular basis of MHC I quality control in the peptide loading complex
    A. Domnick[‡], **C. Winter**[‡], L. Sušac, L. Hennecke, M. Hensen, N. Zitzmann,
    S. Trowitzsch, C. Thomas, R. Tampé
    Nature Communications **2022**, 13, 4701

3. Light-guided intrabodies for on-demand *in situ* target recognition
    in human cells
    E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
    Chemical Science **2021**, 12, 5787-5795

4. Efficient amber suppression via ribosomal skipping for *in situ* synthesis
    of photoconditional nanobodies
    E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
    ACS Synthetic Biology **2022**, 11, 1466-1476 b

5. Structure of an MHC I–tapasin–ERp57 editing complex defines
    chaperone promiscuity
    I. Müller, **C. Winter**, C. Thomas, R. Spaapen, S. Trowitzsch, R. Tampé
    Nature Communications **2022**, 13, 5383

6. TAPBPR is necessary and sufficient for UGGT1-mediated quality control
    of MHC I
    L. Sagert, **C. Winter**, M. Zehetmaier, I. Rupert, C. Thomas, R. Tampé
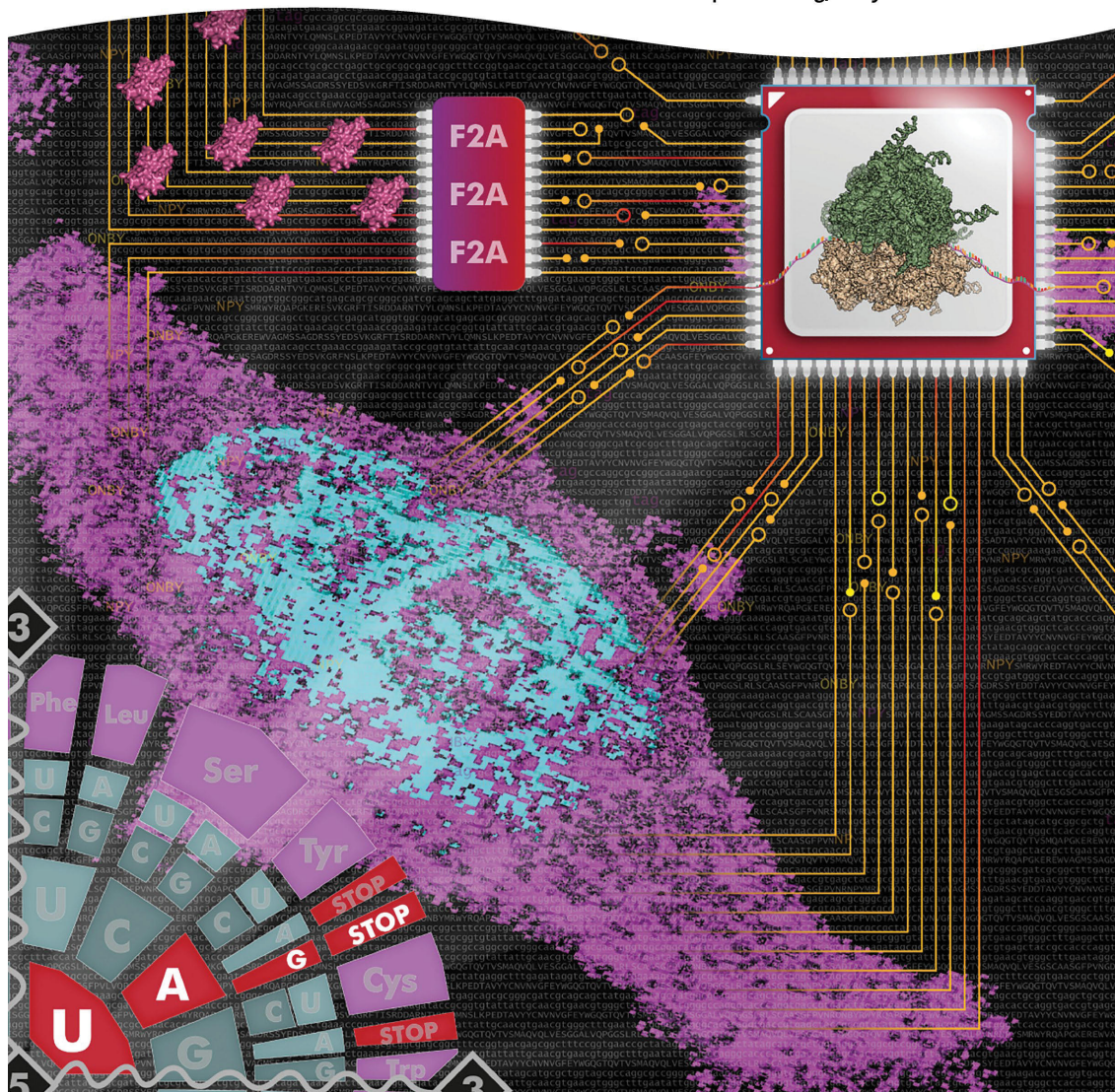   manuscript in preparation

[‡] Contributed equally

II. Cover pages

ACS
SyntheticBiology

pubs.acs.org/acssynthbio

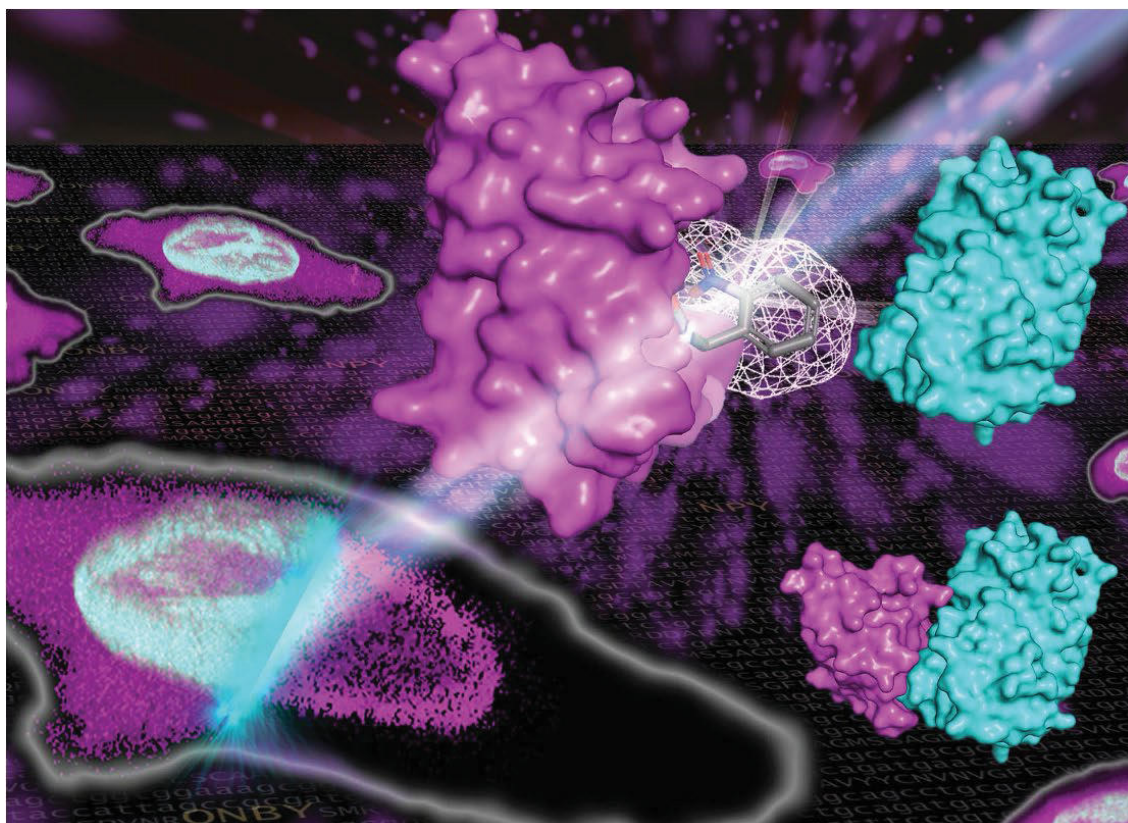ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

Front Cover Page of ACS Synthetic Biology

ACS Synthetic Biology **2022**, 11

Designed by E. F. Joest & C. Winter

Efficient amber suppression via ribosomal skipping
for *in situ* synthesis of photoconditional nanobodies

Showcasing research from Professor Robert Tampé's laboratory, Institute of Biochemistry, Goethe University Frankfurt, Germany, in collaboration with Professor Alexander Deiters, Department of Chemistry, University of Pittsburgh, PA, USA.

Light-guided intrabodies for on-demand *in situ* target recognition in human cells

Nanobodies are ideal probes for visualizing intracellular proteins. Inside living cells, unrestrained binding can cause interference with target function or localization. Here, we report a strategy to circumvent interference through the development of photo-conditional intrabodies. To regulate the interaction, we combine optochemical biology and genetic code expansion in stable cell lines. By equipping the paratope with photocaged amino acids, we control target binding with high spatiotemporal precision inside living cells. Due to the highly stable binding, light-guided intrabodies offer a versatile platform for comprehensive imaging and modulation of target proteins.

As featured in:

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

Registered charity number: 207890

Back Cover Page of Chemical Science

Chemical Science **2021**, 12

Designed by C. Winter & E. F. Joest

Light-guided intrabodies for on-demand in situ target recognition in human cells

III.   Index of abbreviations

Abbreviation	Description
2A site	2A self-cleaving peptide site
3D	three-dimensional
aaRS	tRNA synthetase
ABC	ATP-binding cassette
ABCB2	ATP-binding cassette transporter B family protein 2
ABCB3	ATP-binding cassette transporter B family protein 3
AF647	Alexa Fluor 647
B cell	B lymphocytes
β_2m	β_2 microglobulin
BHV	bovine herpesvirus
CD4$^+$	cluster of differentiation 4 positive
CD8$^+$	cluster of differentiation 8 positive
C$_H$	constant heavy chain
C$_L$	constant light chain
CLSM	confocal laser-scanning microscopy
CMC	critical micelle concentration
CMV	cytomegalovirus
CPXV	cowpox virus
crt	calreticulin
cryo-EM	cryogenic electron microscopy
cryo-ET	electron cryotomography
CSN	high affinity MHC I HLA-A*03:01 binder AIFCSNMTK
CT	C terminus
DIBMA	diisobutylene-maleic acid
DTT	1,4-dithiothreitol
E. coli	*Escherichia coli*
EBV	Epstein-Barr virus
EF1	strong EF-1 alpha promoter
EGFP	enhanced green fluorescent protein
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ERAD	ER associated degradation pathway
ERAP1/2	ER aminopeptidases associated with antigen processing 1 and 2
ERp57	protein disulfide-isomerase A3
ERQC	ER quality control
F2A	F2A self-cleaving peptide sequence
FACS	fluorescence activated cell sorting
Fmoc	fluorenylmethyloxycarbonyl protecting group
FSC	forward scatter
GdmCl	guanidinium chloride
GFP	green fluorescent protein
Glc	glucose
GlcNAc	N-acetyl glucosamine

GluI	ER-resident glucosidases I
GluII	ER-resident glucosidases II
H2B	histone 2B
hc	heavy chain
HCMV	human cytomegalovirus
HFIP	hexafluoroisopropanol
HLA	human leukocyte antigen
Hmp	2-hydroxy-3-mercaptopropionic acid
HPLC	high-performance liquid chromatography
HSV	herpes simplex virus
hυ	Planck relation; photon energy proportional to its frequency
ICP47	infected cell protein 47
IgG	immunoglobulin G
laser	light amplification by stimulated emission of radiation
LC	liquid chromatography
Man	mannose
MaxEnt1	maximum entropy-based MS deconvoluted algorithm
MC80	molluscum contagiosum protein 80
mCherry	DsRed derived red fluorescent protein
MCV	molluscum contagiosum virus
MHC I	major histocompatibility complex class I
MPAA	4-mercaptophenylacetic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MSP	membrane scaffold protein
Nb	nanobody
NBD	nucleotide binding domain
NCL	native chemical ligation
NK	natural killer
NMR	nuclear magnetic resonance spectroscopy
NPY	nitropiperonyl caged tyrosine
NPYRS	NPY optimized tRNA synthetase
NT	N terminus
ONB	*ortho*-nitrobenzene
ONBY	*ortho*-nitrobenzyl caged tyrosine
p2A-Amber	nanobody construct with upstream 2A site
PBS	phosphate-buffered saline
PLC	peptide loading complex
PLC::MSP2N2	PLC reconstituted in large MSP2N2 nanodiscs
pMHC I	peptide loaded MHC class I
PNGase F	peptide:N-glycosidase F
PylT	NPY optimized tRNA
PylT/PylRS	tRNA/aminoacyl tRNA synthetase
R9L	low affinity MHC I HLA-A*03:01 binder RRYCKSTEL
RMSD	root mean square deviation
SBP	streptavidin-binding peptide
SEC	size exclusion chromatography
SEC-MS	size exclusion chromatography-coupled mass spectrometry

SMA	styrene maleic anhydride
SPPS	solid-phase peptide synthesis
SSC	forward scatter
SUMO	small ubiquitin-related modifier
SUMO-tag	small ubiquitin-related modifier protein tag
T cell	T lymphocytes
TAP	transporter associated with antigen processing
TAPBPR	TAP-binding protein related
TCEP	tris(2-carboxyethyl)-phosphine
TCR	T-cell receptor
TMD	transmembrane domain
TMD0	transmembrane domain 0 of TAP
tRNA	transfer ribonucleic acid
Trt	tritylamine protecting group
UDP-Glc	uridine diphosphate glucose
UGGT1	UDP-glucose glycoprotein glucosyltransferase 1
Ulp1	Ubl-specific protease 1
UPLC	ultra-performance liquid chromatography
UV	ultraviolet
VA-044	2,2'-azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride
V_H	variable heavy chain
$V_H H$	camelid derived nanobody
V_L	variable light chain
V_{NAR}	cartilaginous fish derived nanobody

IV.   Table of contents

1. Abstract

The human body is permanently exposed to its environment and thus to viruses and other pathogens, which require a flexible response and defense. Alongside to the innate immune system, the adaptive immune system provides highly specialized protection against these threats. The major histocompatibility complex class I (MHC I) antigen presentation system is a cornerstone of the adaptive immune system and a major constituent of cellular immunity. Pathogens such as viruses that invade a cell will leave traces in the form of proteins and peptides which are degraded and loaded onto MHC I molecules. MHC I peptide loading is performed by peptide loading complex (PLC) in the membrane of the endoplasmic reticulum as part of a multifaceted and comprehensive quality control machinery. Monitored by multiple layers of quality assurance, the MHC I molecules consequently display the immune status of the cell on its surface. In this context, the captured fragment of the virus serves as a call for help issued by the cell, alerting the adaptive immune system to the infection to mount an appropriate immune response.

The three-dimensional structure as well as the mechanistic details of parts of this complex machinery were characterized in the context of this dissertation. Among other tools, light-modulable nanotools were developed in this thesis, which permit external regulation of cellular processes in temporal and spatial resolution. Furthermore, methods and model systems for the biochemical characterization of cellular signaling cascades, proteins, as well as entire cell organelles were developed, which are likely to influence the field of cellular immunity and protein biochemistry in the future.

This cumulative work comprises a total of six publications whose scientific key advances will be briefly outlined in this abstract. In the introduction, the scientific background as well as the current state of research and methodological background knowledge are conveyed. The results section condenses the main aspects of the publications and links them to each other. Further details can be retrieved from the attached original publications.

In "Semisynthetic viral inhibitor for light control of the MHC I peptide loading complex, Winter, Domnick *et al.*, Angew Chem Int Ed 2022" a photocleavable viral inhibitor of the peptide loading complex was produced by semi-synthesis. This nanotool was shown to be suitable for both purifying the PLC from human Raji cells as well as reactivating it in a light-controlled manner. Thus, this tool establishes the isolation of a fully intact and functional peptide loading complex for biochemical characterization. In addition, a novel flow cytometric analysis pipeline for microsomes was developed, allowing cellular vesicles to be characterized with single organelle resolution, similar to cells.

In "*Molecular basis of MHC I quality control in the peptide loading complex*, Domnick, Winter *et al.*, Nat Commun 2022" the peptide loading complex was reconstituted into large nanodiscs, and a cryo-EM structural model of the editing module at 3.7 Å resolution was generated. By combining the structural model with *in vitro* glycan editing assays, an allosteric coupling between peptide-MHC I assembly and glycan processing was revealed, extending the known model of MHC I loading and dissociation from the PLC. These mechanisms provide a prototypical example for endoplasmic reticulum quality control.

In a related context, in "*Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity*, Müller, Winter *et al.*, Nat Commun 2022" a recombinantly assembled editing module comprised of MHC I-tapasin-ERp57 was crystallized for X-ray structural biology. The resulting crystal structure at a resolution of 2.7 Å permitted the precise identification of characteristic features of the editing module and particularly of the peptide proofreading mechanism of tapasin. This study provided pivotal insights into the tapasin-mediated peptide editing of different MHC I allomorphs as well as similarities to TAPBPR-based MHC I peptide proofreading.

In "*TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I*, Sagert, Winter *et al.* (in preparation)" novel insights concerning the peptide proofreader TAPBPR and its close interplay with the folding sensor and glucosyltransferase UGGT1 were obtained. It was shown that TAPBPR is an integral part of the second level of endoplasmic quality control and is indispensable for effective MHC I coordination by UGGT1.

In "*Light-guided intrabodies for on-demand in situ target recognition in human cells*, Joest, Winter *et al.,* Chem Sci 2021" intracellular nanobodies were equipped with a photocaged target recognition domain by genetic code expansion via amber suppression. These intrabodies, acting as high-affinity binding partners endowed with a fluorophore, could be used in a light-triggered approach to instantaneously visualize their target molecule. This system provided a universal platform for light-activated manipulation of target molecules in living human cells.

This approach was extended to a model system transferable to any cell type in "*Efficient amber suppression via ribosomal skipping for in situ synthesis of photoconditional nanobodies*, Joest, Winter *et al.,* ACS Synth Biol 2022". Recurrent challenges in genetic code expansion were systematically screened, and a new suppression concept based on ribosomal skipping was established and demonstrated in living cells.

With the COVID-19 pandemic serving as a prime example, it is of utmost relevance for humankind to develop a comprehensive understanding of the precise mechanisms of

adaptive immune defense. Through detailed knowledge of these closely interconnected machineries, it may prospectively be feasible to reconstruct the effects of viral infections in cells and to develop vaccines that always generate optimal epitopes in patients to stimulate enhanced antibody production.

2. Zusammenfassung

Der menschliche Körper ist permanent seiner Umwelt und damit auch Viren und anderen Krankheitserregern ausgesetzt, die eine flexible Reaktion und Abwehr erfordern. Er muss jedoch auch intelligent auf endogene Bedrohungen wie entartete Zellen reagieren, während eine klar definierte Unterscheidung zwischen eigenen und fremden Entitäten jederzeit aufrechterhalten werden muss, um Autoimmunerkrankungen zu vermeiden. Das Immunsystem hat daher zahlreiche Barrieren wie die Schleimhäute entwickelt, um das Eindringen von Krankheitserregern zu verhindern, aber auch humorale und zelluläre Abwehrstrategien, um deren systemische Ausbreitung einzudämmen. Neben dem angeborenen Immunsystem, bietet das adaptive Immunsystem einen hochspezialisierten Schutz gegen diese Bedrohungen. Das MHC I-Antigenpräsentations-system ist ein Eckpfeiler des adaptiven Immunsystems und ein Hauptbestandteil der zellulären Immunität. Krankheitserreger wie Viren, die in eine Zelle eindringen, hinterlassen dort Spuren in Form von Peptiden und Proteinen, die proteasomal abgebaut und auf MHC I-Moleküle geladen werden können. Dies geschieht durch den *Peptide Loading Complex* in der Membran des endoplasmatischen Retikulums als Teil einer vielschichtigen und umfassenden Qualitätskontrollmaschinerie. Ein wesentlicher Aspekt dieser Qualitätsüberwachung ist die Visualisierung des Faltungsstatus von MHC I über das angelagerte Glykan. Die MHC I-Moleküle, die durch mehrere Schichten der Qualitätssicherung, sowohl im endoplasmatischen Retikulum als auch im Golgi-Apparat kontrolliert werden, zeigen daraufhin den Immunstatus der Zelle auf ihrer Oberfläche an. In diesem Zusammenhang dient das eingefangene und präsentierte Proteinbruchstück als Hilferuf der Zelle, um das adaptive Immunsystem auf die Infektion aufmerksam zu machen und eine Immunreaktion einzuleiten. Zu einer solchen Immunreaktion zählen der programmierte Zelltod und die Rekrutierung von Effektorzellen des adaptiven und des angeborenen Immunsystems.

Sowohl die dreidimensionale Struktur des *Peptide Loading Complex* als auch die mechanistischen Details dieser komplexen Maschinerie wurden im Rahmen dieser Dissertation charakterisiert. Zu diesem Zweck wurden die Komponenten des Komplexes entweder aus menschlichen Immunzellen extrahiert und in ein geeignetes Membranmimetikum rekonstituiert oder rekombinant exprimiert und *in vitro* assembliert. Mit modernsten strukturbiologischen Techniken wie kryogene Elektronenmikroskopie und Röntgenkristallographie konnte die räumliche Anordnung der Moleküle mit atomistischer Auflösung entschlüsselt werden. Weiterhin wurden Biomoleküle künstlich durch lichtempfindliche Bestandteile erweitert. Anschließend konnte Licht als bioorthogonaler

Impulsgeber genutzt werden, um diese Moleküle entweder zu zerstören und damit ihre biologische Funktion aufzuheben oder sie zu aktivieren und damit an ihren Zielort zu dirigieren. Dadurch konnten komplexe zelluläre Prozesse mit räumlicher und zeitlicher Auflösung durch Laserpulse gesteuert werden. Darüber hinaus wurden Methoden und Modellsysteme zur biochemischen Charakterisierung von zellulären Signalkaskaden, Proteinen sowie ganzen Zellorganellen entwickelt, die auch in Zukunft auf dem Gebiet der zellulären Immunität und der Proteinbiochemie ihren Nutzen finden werden.

Diese kumulative Arbeit umfasst insgesamt sechs Publikationen, deren wissenschaftlicher Beitrag in dieser Zusammenfassung kurz skizziert werden soll. In der Einleitung werden der wissenschaftliche Hintergrund, der aktuelle Stand der Forschung und methodisches Hintergrundwissen vermittelt. Der Ergebnisteil fasst die wesentlichen Aspekte der Publikationen zusammen und verknüpft sie miteinander. Weitere Details können den beigefügten Originalpublikationen am Ende der Arbeit entnommen werden.

In „*Semisynthetic viral inhibitor for light control of the MHC I peptide loading complex*, Winter, Domnick *et al.*, Angew Chem Int Ed 2022" wurde ein lichtspaltbarer viraler Inhibitor des *Peptide Loading Complex* durch Semisynthese hergestellt. Der virale Inhibitor bindet spezifisch an die Peptidbindungstasche der Transporteinheit des *Peptide Loading Complex* und blockiert diese, wodurch der Komplex irreversibel in seiner Konformation fixiert wird und nicht mehr in der Lage ist, dem endoplasmatischen Retikulum weiterhin Peptide zuzuführen. Dieser Mechanismus ist Teil moderner Isolationsstrategien, bei denen der PLC mittels Affinitätschromatographie mit diesem modifizierten viralen Inhibitor aus menschlichen Zellen präpariert wird. Ein großes Problem bei diesem Ansatz ist die irreversible Hemmung der Transportfunktion, die eine weitere biochemische Charakterisierung einschränkt. Da der Einbau von unnatürlichen Aminosäuren, dessen Peptidrückgrat durch Licht spaltbar ist, *in vivo* nicht möglich ist, wurde der virale Inhibitor in zwei Teile zerlegt und teil-synthetisch und sowie teil-rekombinant hergestellt. Die beiden Komponenten wurden dann in einer nativen chemischen Ligation wieder vereint. Das vollständige Nanotool erwies sich als geeignet, den *Peptide Loading Complex* sowohl aus menschlichen Raji-Zellen zu reinigen als auch lichtgesteuert wieder freizusetzen. Damit konnte erstmals ein vollständig intakter *Peptide Loading Complex* für die biochemische Charakterisierung isoliert werden. Die Analysen in dieser Studie haben gezeigt, dass die lichtgetriebene Reaktivierung mit hoher Effizienz abläuft und dass alle quantifizierbaren Funktionen vollständig wiederhergestellt werden konnten. Darüber hinaus wurde eine neuartige durchflusszytometrische Analysemethode für Mikrosomen entwickelt, die es ermöglicht, zelluläre Vesikel mit Einzelorganellenauflösung ähnlich wie bei Zellen zu untersuchen. Perspektivisch werden die hier erzielten Fortschritte eine weitere

Charakterisierung des vollständig aktiven PLC ermöglichen. Dadurch wird es zukünftig möglich sein, sowohl die neu zugänglichen Konformationen strukturbiologisch zu beschreiben als auch den PLC in Proteoliposomen zu rekonstituierten, die eine Abgrenzung von innen und außen bieten und dabei ideal für peptid-transport abhängige Studien in einer definierten Umgebung sind.

In „*Molecular basis of MHC I quality control in the peptide loading complex*, Domnick, Winter *et al.*, Nat Commun 2022" wurde der *Peptide Loading Complex* erstmals in großen *Nanodiscs* rekonstituiert. Das hat es ermöglicht, ein Kryoelektronenmikroskopisches Strukturmodell des Editiermoduls des mit einer Auflösung von 3.7 Å zu erstellen. Mit diesem räumlichen Modell konnten zuvor unaufgelöste Regionen des *Peptide Loading Complex* strukturell charakterisiert werden, wobei durch die gut aufgelösten Gylkane essenzielle Informationen über die Interprozesskommunikation der Peptidbeladung auf MHC I-Moleküle gewonnen werden konnte. Es wurde nachgewiesen, dass der in Nanodiscs rekonstituierte *Peptide Loading Complex* in der Lage ist, Peptide spezifisch aus der Bindetasche der MHC I-Moleküle auszutauschen. Durch die Kombination des Strukturmodells mit *in vitro* Glykan-Editierungstests wurde eine allosterische Kopplung zwischen Peptid-MHC I-Assemblierung und Glykanmodulation durch die α-Glukosidase GluII ermittelt, die das bekannte Modell der MHC I-Beladung und der Dissoziation vom *Peptide Loading Complex* erweitert. Es wurde festgestellt, dass die Glykanmodulation innerhalb des PLC stattfinden kann, sobald MHC I optimal beladen ist, ohne dass zuvor eine Dissoziation des Komplexes oder dessen Untereinheiten erforderlich ist. Diese Mechanismen sind ein prototypisches Beispiel für die Qualitätskontrolle im endoplasmatischen Retikulum.

In einem ähnlichen Zusammenhang wurde in „*Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity*, Müller, Winter *et al.*, Nat Commun 2022" ein rekombinant zusammengesetztes MHC I Editiermodul, bestehend aus den *Peptide Loading Complex* Komponenten MHC I, Tapasin und ERp57 für die Röntgenstrukturuntersuchung kristallisiert. Auch hier wurde eine lichtgesteuerte Herangehensweise gewählt, um MHC I in einem Peptid-empfänglichen Zustand zu fixieren. Die resultierende Kristallstruktur mit einer Auflösung von 2.7 Å ermöglichte die genaue Identifizierung charakteristischer Merkmale des Editiermoduls und insbesondere des Peptid-Austausch Mechanismus von Tapasin. Es wurde festgestellt, dass Tapasin eine Editierdomäne besitzt, die als hydrophober Zapfen in die Peptidbindungstasche von MHC I eingeschoben ist und über die Tapasin den Austausch von Peptiden innerhalb der Peptidbindungstasche von MHC I katalysiert. Dies lieferte entscheidende Einblicke in den

Tapasin-vermittelten Peptid-Austausch verschiedener MHC I-Allomorphe sowie in die Ähnlichkeiten mit der TAPBPR-basierten Peptid-Austausch Katalyse.

In „*TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I*, Sagert, Winter *et al.* (in Vorbereitung)" wurden neue Erkenntnisse über den Peptidkorrekturleser TAPBPR und seine enge Wechselwirkung mit dem Faltungssensor und Glukosyltransferase UGGT1 gewonnen. TAPBPR und UGGT1 sind Bestandteil der zweiten Ebene der Qualitätskontrollmaschinerie und kontrollieren die MHC I-Moleküle hinsichtlich ihrer Beladung in der peptidärmeren Umgebung des *cis*-Golgi und des endoplasmatischen Retikulums. Erst kürzlich gab es Hinweise darauf, dass TAPBPR und UGGT1 ihre Aktionen gemeinsam koordinieren und dass TAPBPR möglicherweise Substrate an UGGT1 vermittelt oder durch Interaktion weitere Substrate für UGGT1 zugänglich macht. Um diese Hypothese zu überprüfen, wurden zunächst geeignete Substrate für UGGT1 in Form eines peptidlosen MHC I-TAPBPR Komplexes mit nativer Glykanstruktur hergestellt. In einem lichtgesteuerten Ansatz wurden zusätzlich frustrierte MHC I-Moleküle erzeugt, die kein gebundenes Peptid enthielten oder durch TAPBPR stabilisiert wurden. *In vitro* Glukosylierungsexperimente und eine auf Massenspektrometrie basierende Auswertung der Glykanmuster lieferten Informationen über die Aktivität von UGGT1 und den Einfluss von TAPBPR. Es konnte gezeigt werden, dass TAPBPR ein integraler Bestandteil der zweiten Ebene der endoplasmatischen Qualitätskontrolle ist und für eine effektive Koordination von MHC I durch UGGT1 unerlässlich ist.

In „*Light-guided intrabodies for on-demand in situ target recognition in human cells*, Joest, Winter *et al.*, Chem Sci 2021" wurden intrazelluläre *Nanobodies* durch Erweiterung des genetischen Codes mittels *Amber*-Suppression mit einer lichtsensitiven Zielerkennungsdomäne ausgestattet. *Nanobodies* sind von Antikörpern abgeleitete Bindungsfragmente, die aufgrund ihres einkettigen Designs ohne Disulfidbrücken eine ausgezeichnete intrazelluläre Stabilität aufweisen. Im zellulären Umfeld können sie ihr Zielmolekül mit nanomolarer Affinität binden. Diese Bindung kann jedoch in der Regel nicht kontrolliert werden und findet sofort nach Produktion des *Nanobodies* statt, wodurch zelluläre Prozesse bereits unbeobachtet modifiziert werden. In dieser Arbeit wurden die lichtaktivierbaren *Intrabodies* (*Nanobodies* innerhalb von lebenden Zellen) mit einem Fluorophor ausgestattet (sogenannte *Chromobodies*) und konnten daraufhin in einem lichtaktivierten Ansatz zur sofortigen Visualisierung ihres Zielmoleküls eingesetzt werden. Zur Implementierung dieses Zielerkennungs- und Markierungskonzepts wurden stabile Humanzelllinien erzeugt, in denen die Expression der lichtaktivierbaren *Intrabodies* auf endogenen Expressionsniveau induziert werden konnte. Dieses System bietet eine

universelle Plattform für die lichtaktivierte Manipulation von austauschbaren Zielmolekülen in lebenden Zellen.

Dieses Konzept wurde in „*Efficient amber suppression via ribosomal skipping for in situ synthesis of photoconditional nanobodies*, Joest, Winter *et al.,* ACS Synth Biol 2022" zu einem Modellsystem erweitert, das auf jeden Zelltyp übertragbar ist und nicht auf einer einzigen stabilen Zelllinie beruht. Häufig wiederkehrende Probleme bei der Erweiterung des genetischen Codes wurden systematisch untersucht, wobei herausgefunden wurde, dass die Ungenauigkeit des Ribosoms zu Anfang der Protein-Translation Probleme bei einer effizienten *Amber*-Suppression auslösen kann. Daher wurde ein neues *Amber*-Suppression Konzept auf der Grundlage einer 2A-Sequenz etabliert und in lebenden Zellen demonstriert. Bei diesem Konzept wird während der ribosomalen Translation eine Peptidbindung nicht ausgebildet, wodurch es innerhalb desselben Genprodukts zu zwei individuellen Polypeptidsträngen kommt. Dadurch liegt ein potenziell frühes *Amber*-Codon weiter hinten im Gensegment und ist folglich nicht mehr so häufig durch ribosomale Ungenauigkeiten betroffen. Zukünftig kann dieses System durch einfache Transfektion auf beliebige humane Zelllinien übertragen werden, wobei die zu modulierenden Zielmoleküle relativ frei ausgewählt werden können. Dieses Modellsystem bietet eine leistungsstarke Plattform, um komplexe Signalwege gezielt zu modulieren oder beispielsweise Zielmoleküle lichtgesteuert der Degradation zuzuführen.

Nicht zuletzt vor dem Hintergrund der COVID-19-Pandemie ist es für die Menschheit von größter Bedeutung, ein umfassendes Verständnis der genauen Mechanismen der adaptiven Immunabwehr zu entwickeln. Durch die detaillierte Kenntnis dieser eng miteinander verknüpften Mechanismen könnte es in Zukunft möglich sein, die Auswirkungen von Infektionen in Zellen genau zu rekonstruieren und so beispielsweise Impfstoffe zu entwickeln, die bei Patienten stets optimale Epitope erzeugen. Dies würde Impfstoffe einerseits potenter machen und es andererseits erlauben gezielt die Produktion von optimalen Antikörpern anzuregen.

3. Introduction

This work utilizes a comprehensive range of methods from different disciplines of chemistry, biochemistry, and structural biology, for the understanding of which the fundamentals are to be conveyed in this introduction. The basic principles of major histocompatibility complex class I (MHC I)-mediated antigen presentation, especially in the context of the peptide loading complex (PLC), are discussed. The structural composition of the fully assembled peptide loading complex as well as the mechanistic sequences of its disassembly are analyzed in more detail in this thesis. In this context, both levels of MHC I quality control, on the one hand by the peptide loading complex and on the other hand via TAP-binding protein related (TAPBPR) and UDP-glucose glycoprotein glucosyltransferase 1 (UGGT1), are examined biochemically. In particular, the mechanistic interplay with the conserved MHC I glycan is analyzed. As the PLC is a highly sensitive assembly, new ways of purification and reconstitution using light-controllable nanotools are explored and light is introduced as a mild, contact-less and orthogonal trigger to control biological processes. The implementation of light-sensitive molecules into biological systems via semisynthesis and amber suppression will be discussed. In addition, the reader will be introduced to novel analysis pipelines for glycan analysis using intact protein mass spectrometry and to the analysis of cell organelles by flow cytometry, established in this work.

3.1 The immune system

Our body is constantly exposed to its environment and therefore exposed to exogenous pathogens like bacteria, fungi, and viruses. But it also has to react flexibly to endogenous threats like malignantly transformed cells, while a well-defined distinction between self and non-self entities needs to be maintained at all times to prevent autoimmune diseases. The immune system consequently evolved several physical barriers like the mucous membranes to prevent pathogens from entering, but also humoral and cellular defense strategies to prevent them from propagating systemically.

The immune system can be subdivided into two parts: the general innate immune system and the specialized adaptive immune system.[1] The innate immune system is the first line of defense against pathogens and is specialized in recognizing patterns that are foreign to the host organism, causing inflammatory responses that trigger recruitment of immune cells to remove the threat. The main effector mechanisms in innate immunity are epithelial and phagocytic enzymes, inflammation related serum proteins, lectins, antimicrobial peptides from phagocyte granules, phagocytes, natural killer (NK) cells, and dendritic

cells.[2,3] The immune response of the innate immune system is not specific to a particular pathogen. In the event of infection, it is in close contact with the adaptive immune system, which is able to mount an immune response against a specific pathogen. It is able to do so as it is composed of a repertoire of highly variable antigen-specific receptors which is capable of memorizing previous contacts with a pathogen in its immunological memory and reacting much faster in the event of renewed contact.[4] However, the adaptive immune response is still significantly slower than the immediate response of the innate immune system. The adaptive immune response can be further subdivided into the humoral and cell-mediated immune responses, which is primarily dependent on which lymphocyte has triggered the adaptive immune response. Lymphocytes are the main effector cells of the adaptive immune system comprising T cells (initiating cytotoxic cell-mediated immunity) and B cells (initiating humoral-driven immunity). The humoral-mediated immune response is based on B cells that differentiate into plasma B cells to produce antibodies upon pathogen contact. Antibodies can recognize molecular patterns of specific pathogens. This leads to inactivation, neutralization, lysis or phagocytosis.[5] The cell-mediated immune response is driven by T cell conveyed recognition of antigens by T cell receptors (TCRs) on the cell surface, in particular cytotoxic T cells (CD8+) and helper T cells (CD4+). TCRs exhibit a highly variable recognition interface, similar to those of antibodies. While CD4+ T cells await stimulation by peptide-loaded major histocompatibility complex (MHC) class II complexes, CD8+ T cells require peptide loaded MHC class I (pMHC I) stimulation. CD8+ T cell-TCRs generally need further co-stimulatory signals like CD80 or CD86 interaction with the antigen-presenting cell. Upon activation, CD8+ T cells proliferate to cytotoxic T cells which release cytotoxins (perforin, granzymes and granulysin) after stimulation, triggering the caspase cascade which can lead to cell death of the infected or malignantly transformed cell by apoptosis.[6]

In summary, the innate immune system provides the rapid first layer of defense against pathogens entering the organism, without being specialized to specific threats. It activates the adaptive immune system which deals with the high demands of specialized immune defense. A major part of the adaptive immune defense is based on the recognition of xenobiotic molecules via presentation on MHC I molecules, the emergence of which is explained in the next section.

3.2   The MHC I antigen presentation pathway

All vertebrates report on their cellular and extracellular immune status via MHC molecules on their cell surface. MHC molecules are present in all nucleated cell types and constitute carrier proteins to display antigenic peptides on the cell surface. MHC II molecules are

mostly used to present phagocytosed exogenous peptides on the cell surface and are not further described here. MHC I molecules, in contrast, present endogenous, cellular antigens to NK and CD8[+] T cells. In humans, three classical MHC I genes are present, which are named human leukocyte antigen-A (HLA-A), HLB-B and HLA-C allomorphs. With more than 24,000 human sequence variants known to date[7,8], MHC I molecules exhibit an extremely high amount of polymorphism, which is focused on regions that are in contact with antigens that can bind to MHC I.[9] These polymorphisms are crucial for the survival of humankind as they increase the variety of peptides that can be presented on the cell surface. A vast HLA variability equally reduces the chance that a specific pathogen generating inefficiently presented antigens exterminates the entire population. Every human can have up to six different HLA alleles which ensures that available MHC I molecules can display a comprehensive set of peptides on each cell.[10] Endogenous peptides are also presented on MHC I. However, these peptides of legitimate source do not trigger CD8[+] T cell activation and do not lead to cell death. This is achieved by selection and elimination of autoreactive T cells within the thymus.
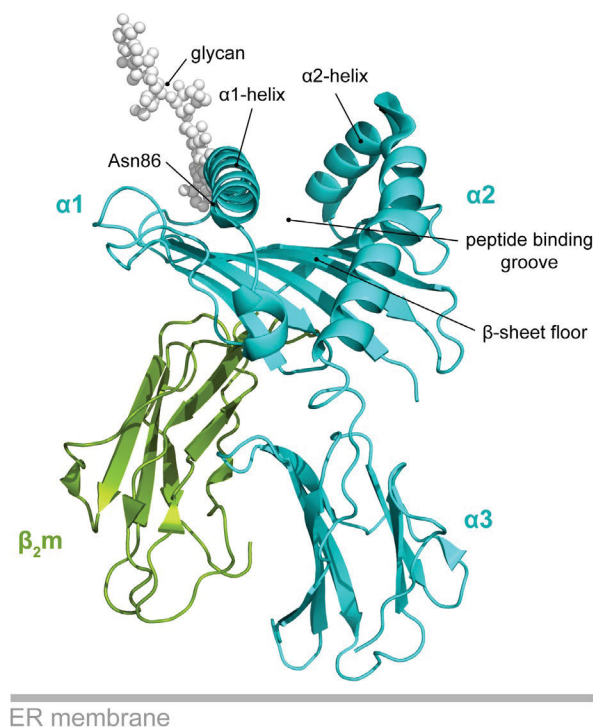

Figure 1: *Structural assembly of MHC I.* MHC I hc consists of three domains (α1-α3, teal) and includes the peptide binding groove. Here, the empty state in the context of the peptide loading complex is shown. Domain α3 is the main interactor for the β_2m light chain (green). At Asn86 an N-linked glycan (partially shown, white) is present, indicating the loading status of MHC I. The ER membrane (grey) denotes the native orientation of the complex. PDB: 7QPD[11], partly shown.

MHC I is a heterodimer formed by the soluble light chain β_2-microglobulin (β_2m) and the membrane-anchored MHC I heavy chain (hc, Figure 1).[12,13] The small β_2m is a single Ig-like

fold domain attached to the three (α1-α3) domains of MHC I hc. The peptide binding groove is shaped by the α1 and α2 domain, which is confined by two alpha-helices at each side and a β-sheet floor. The Ig-like domain α3 is opposite to β_2m and bears the membrane anchor. An N-linked glycan ($Glc_3Man_9GlcNAc_2$) is co-translationally transferred to MHC I hc Asn86 by oligosaccharyltransferase (OST) and subsequently remodeled by ER-resident glucosidases I (GluI)[14,15] and II (GluII) to $Glc_1Man_9GlcNAc_2$.[15] A single terminal glucose indicates an unloaded or not optimal loaded MHC I.[11]
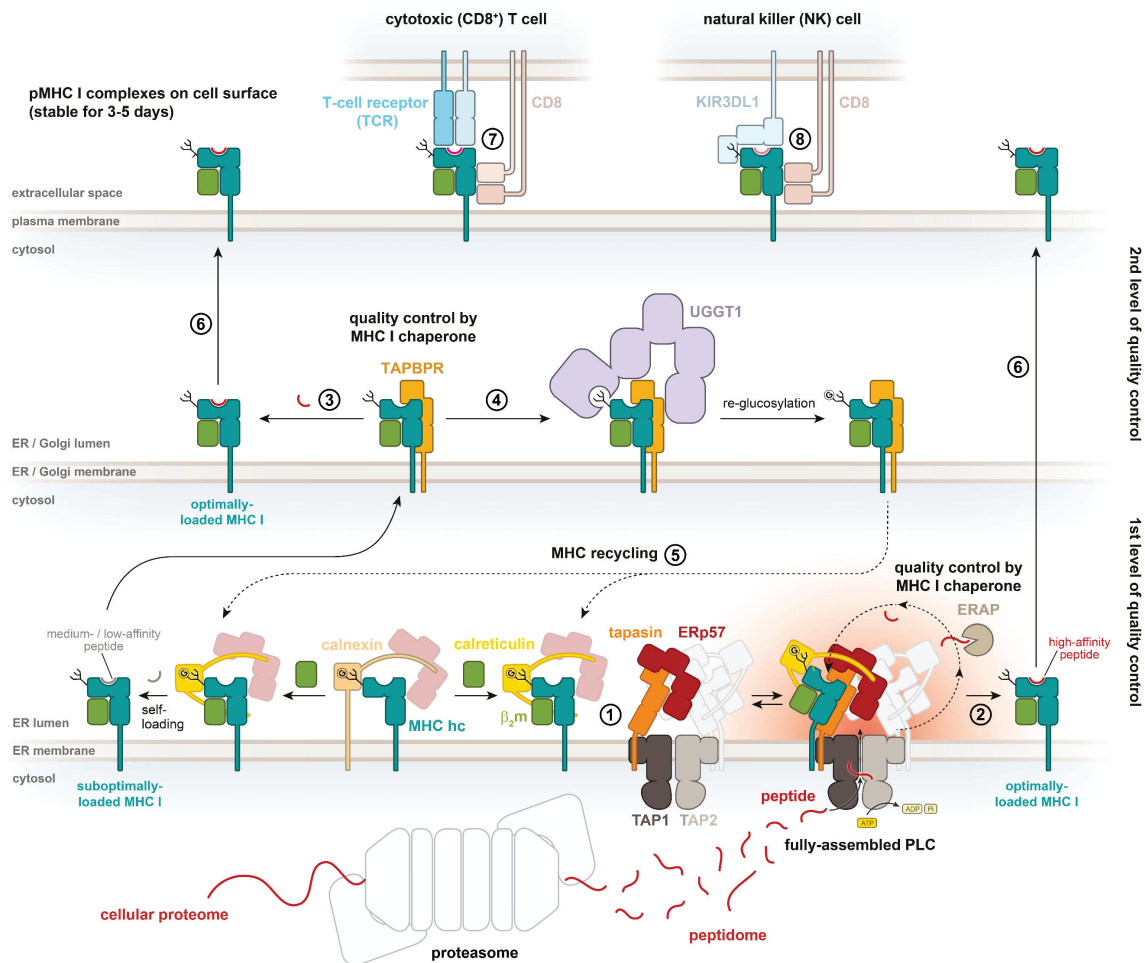

Figure 2: *Overview of the MHC I antigen presentation and quality control pathway.* (1) Calnexin and ERp57 bind to freshly produced mono-glucosylated MHC I hc. Calnexin is replaced by calreticulin that recruits the MHC I heterodimer to the peptide loading complex (PLC). In the PLC, tapasin chaperones MHC I peptide loading and ensures that MHC I is charged with high-affinity peptides (red cloud). (2) The PLC is centered on TAP, which transports proteasomal antigenic peptides to the ER lumen and provides a locally enriched amount of peptides, that can be loaded onto MHC I. (3) TAPBPR and UGGT1 constitute a second level of MHC I quality assurance, reglucosylating sub-optimally loaded MHC I for renewed peptide loading in the context of the PLC (5). Stable pMHC I complexes are released and trafficked (6) via the Golgi apparatus to the cell surface, where they present their antigenic cargo to cytotoxic T- (7) and NK-cells (8). Adapted from: Thomas and Tampé, Curr Opin Immunol, 2021.[15]

After biogenesis, MHC I hc is intrinsically instable and requires constant chaperoning, which leads to the rapid and potentially co-translational[16] recruitment of the scaffolding chaperone calnexin via the monoglucosylated glycan which subsequently recruits the thiol oxidoreductase ERp57 (Figure 2).[14,15] Once β_2m is successfully bound, calnexin is

replaced by calreticulin (crt) and the assembly of the peptide loading complex (PLC) is initiated.

The PLC is constituted from the endoplasmic reticulum (ER) resident heterodimeric ATP-binding cassette (ABC) transporter associated with antigen processing 1 and 2 (TAP), MHC I hc, β_2m, tapasin, protein disulfide-isomerase A3 (ERp57) and calreticulin. Stoichiometric numbers of MHC I, tapasin, ERp57 and calreticulin compose to an editing module, two of which can be assembled simultaneously centered on TAP. The PLC is the first and most important level of MHC I quality control. It is a highly dynamic complex that ensures that the bound MHC I molecules are supplied with a selection of suitable antigens. At the same time, it ensures that only high-affinity peptides bind to MHC I. Optimally loaded pMHC I molecules leaving the PLC are, in contrast to their unloaded counterparts, very stable and can remain on the cell surface for several days to reflect the immune status of the cell.
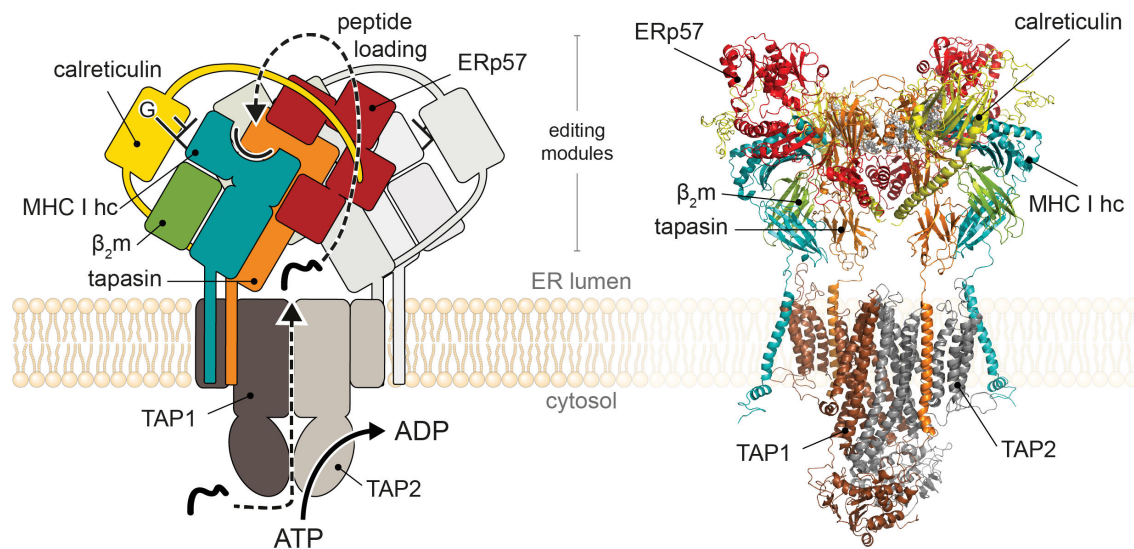

Figure 3: Schematic and modelled assembly of the peptide loading complex. The PLC is resident in the ER membrane and consists of the heterodimeric ABC transporter TAP1/2 (brown, grey) and two identical editing modules (one depicted in light grey in the scheme). Each editing module consists of the heterodimeric MHC I hc (teal), β_2m (green), tapasin (orange), ERp57 (red) and calreticulin (yellow). TAP actively transports proteasomal peptides into the ER lumen where they are loaded onto the highly chaperoned MHC I molecules. The structure model has been simulated[17] based on the cryo-EM structure of the human PLC[18] and colored accordingly.

Structurally, TAP anchors the PLC in the ER membrane and binds tapasin, while tapasin itself interconnects all other components of the editing module (Figure 3). In addition to the structuring role of tapasin, it assures the exchange of low affinity peptides in the MHC I peptide binding groove via its editing loop. Next to tapasin, ERp57 is another structuring component contacting calreticulin and tapasin, whereas its inherent oxidoreductase function is not active in the PLC.[18,19] Calreticulin serves various functions within the cell:

in the context of the PLC it serves as a chaperone recruiting MHC I into the PLC and keeping unloaded MHC I molecules stable.

Antigenic peptides translocated by TAP can either derive form the ubiquitin proteasome system or from defective ribosomal products. The former are derived from misfolded or short-lived proteins, and are marked for proteasomal degradation by the ubiquitination system.[20] Defective ribosomal products are termination products of failed ribosomal protein synthesis.[21] TAP preferentially transports peptides with a length of 8-16 amino acids, but the transport of longer (up to 40-mer)[22] peptides is also possible.[23-25] In comparison, MHC I molecules require to be loaded with peptides of a length of 8-12 amino acids, while most cargo peptides have a length of 8-9 amino acids.[26] Therefore, peptides transported into the ER can be further N-terminally trimmed by aminopeptidases such as ER aminopeptidases associated with antigen processing 1 and 2 (ERAP1/2) to potentially generate suitable antigens (Figure 2, step (2)).[27-29] The bound antigens can adopt different conformations within the peptide binding pocket of MHC I, depending on the allomorph and peptide sequence.[30] Within the PLC, tapasin aids in peptide exchange by extending the distance between the two helices of the binding groove through interaction with the $\alpha 2$ domain (Figure 2, step (2)).[31]

Tapasin is structurally associated with ERp57 and calreticulin and is considered to be the primary MHC I peptide editor in the peptide loading complex.[11] Peptide exchange within the MHC I binding pocket is facilitated by acceleration of peptide association and dissociation, in particular by the editing loop (scoop loop) of tapasin.[32,33] ERp57[Cys33] contacts tapasin[Cys95] via a disulfide bridge. ERp57 can switch between an intermolecular or intramolecular (to ERp57[Cys36]) disulfide bridge. This inherent mechanism of the oxidoreductase is referred to as the escape pathway to keep ERp57 non-covalently coupled to its substrates.[34] Different MHC I allomorphs have variations in their structure and peptide binding preferences. However, all of them could still be clients of tapasin with varying degrees of tapasin-dependence in their peptide exchange.[8] The mechanistic and conformational nature of the MHC I-tapasin-ERp57 interaction has not yet been fully elucidated and was investigated in parts of this work.[8]

Once an optimal peptide is loaded, the terminal glucose is removed by GluII from the Asn86 N-linked glycan, and MHC I is released from the PLC. However, the exact sequence of deglucosylation and PLC disassembly events is still elusive but was investigated in depth in this thesis.[35] Released and glucose-trimmed MHC I travel along the secretory pathway through the *Golgi* compartment where they can encounter a second layer of quality control mediated by the tapasin homologue TAPBPR (TAP-binding protein related), which is present in both the ER and *Golgi* apparatus. TAPBPR can independently catalyze

the exchange of low affinity peptides in a similar manner as tapasin in the context of the PLC.[32,33] TAPBPR is thought to collaborate with UDP-glucose:glycoprotein-glucosyltransferase 1 (UGGT1)[15] as a folding sensor[36] and UGGT1 senses incompletely folded MHC I after pMHC I-TAPBPR interaction (Figure 2, step (4)).[35] UGGT1 thereupon transfers a terminal glucose to the Asn86 N-glycan again, marking it as improperly folded. With this flag, MHC I cannot leave the ER or travel to the cell surface, and it is once again recruited to the PLC by calreticulin and eventually calnexin where it is subjected to a further cycle of quality control (Figure 2, section (5)). This complex interplay of several layers of quality control prevents non-optimal pMHC I from reaching the cell surface. The exact interplay between TAPBPR and UGGT1, as well as the sequence of events and the involvement of other components for the reglucosylation of MHC I are not yet known but were elaborated on in this thesis.[37]

If pMHC I passes all stages of quality control (and there may be others as yet unknown along the secretory pathway) and is thus optimally loaded, it will present its antigenic cargo on the cell surface, where it can remain stably assembled for several days.[38] MHC I hcs that have lost either β_2m or their cargo peptide are just transiently stable on the cell surface and cannot elicit T-cell or NK activation. These "lonely" MHC I hcs are internalized and can be recycled for another round of peptide presentation.[39]

3.3   TAP as a prime target for viral immune evasion

Viruses and other pathogens leave traces of their own proteome in the cellular peptidome. The MHC I antigen presentation system constantly screens this cellular waste for antigens of exogenous source - and is thus dependent on a constant and rapid supply of cellular peptides, as they are degraded within seconds in the cytosol.[40] With constant antigen supply as the central bottleneck for the entire MHC I antigen presentation system, it is not surprising that many viruses have developed their own disguise strategies in the form of proteins inhibiting TAP function to evade an immune response and proliferate undetected throughout the host organism.

The heterodimeric ABC transporter TAP is composed of TAP1 (ABCB2) and TAP2 (ABCB3), which each contain one transmembrane domain (TMD) and one nucleotide binding domain (NBD)[41]. Each TMD is composed of six core transmembrane helices and a TMD0 (four transmembrane helices) representing the main interaction hub for tapasin and thus for the entire PLC (Figure 3).[42-45] The cytosolic NBDs provide the energy for active transport[46] by ATP hydrolysis but binding of antigenic peptides is independent from ATP binding.[24,47,48] TAP is composed of multiple cytosolic loops that interconnect the NBDs and TMDs. By

contacting the NBDs, their nucleotide charging status can be sensed and energetically transferred to the TMD during the active cycle.[45,49,50] The peptide binding pocket of TAP has a width of about 2.1 nm and allows an "extended kink" conformation of the peptides, with the middle part of the sequence protruding from the binding pocket[51,52], which also explains the binding specificity of the transporter: The peptide substrates of a length of 8-40 amino acids are recognized mainly by their N- and C-terminal residues, bearing specific binding patterns that strongly increase binding affinity.[53] Charged or hydrophobic residues at position 2 as well as hydrophobic residues at position 3 are advantageous, while negatively charged residues at position 1 have an adverse effect. The C terminus should be hydrophobic for effective binding. TAP reporter peptides utilized in the studies presented here were designed according to these aspects. The recognition patterns for peptide transport are in line with the requirement of MHC I binding[53], providing an energetically advantageous pre-selection of peptides by TAP. Sterically demanding modifications in the middle of the peptide sequence are accommodated without problems.[11,54]
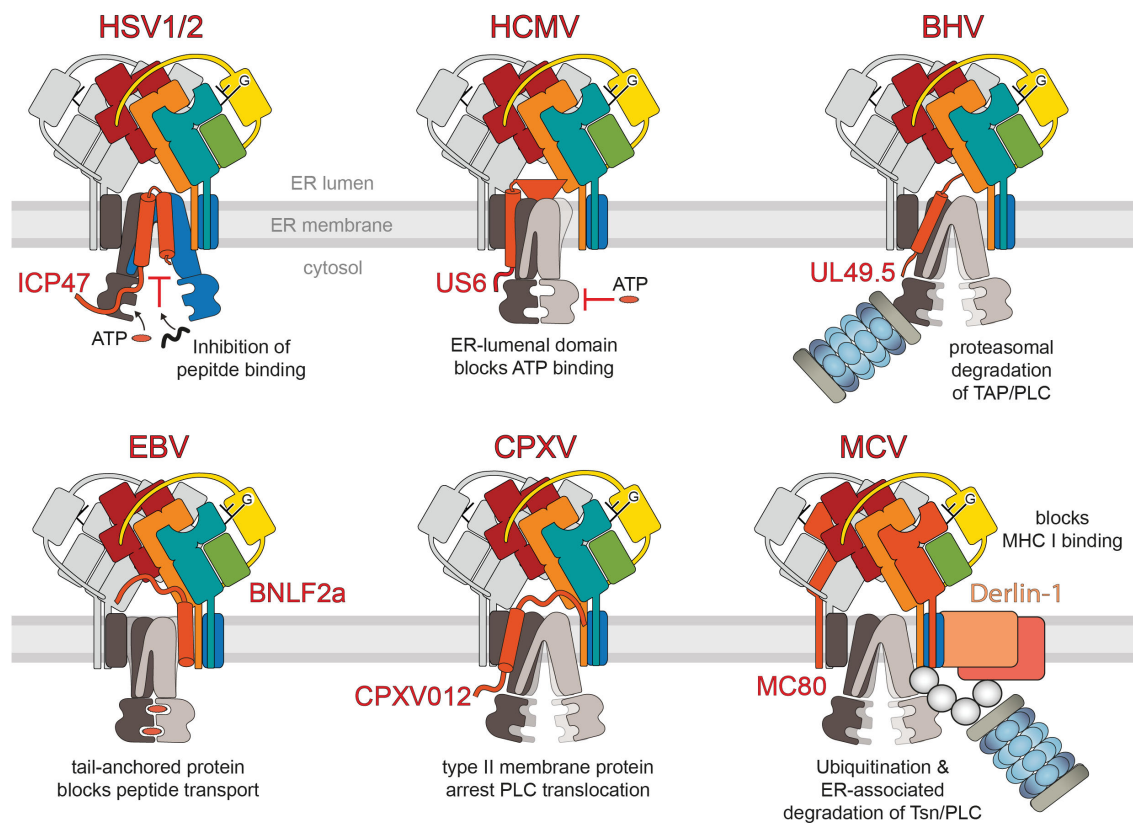

Figure 4: *Viral immune evasins to inhibit TAP function*. The herpes simplex virus (HSV) derived ICP47 is the only soluble viral immune evasin targeting TAP by blocking the peptide binding pocket. The ER membrane resident, human cytomegalovirus (HCMV) derived US6 inhibits ATP binding, while UL49.5 (bovine herpesvirus, BHV) forces TAP into a confirmation inhibiting peptide translocation, simultaneously initiating the proteasomal degradation of TAP. BNLF2a (Epstein-Barr virus, EBV) blocks both peptide and ATP binding with its cytosolic tail, while CPXV12 (cowpox virus, CPXV) does so via its ER luminal tail. MC80 in cooperation with Derlin-1 (molluscum contagiosum virus, MCV) sabotages MHC I antigen presentation by targeting tapasin for degradation. Figure was adapted from collaborative research center 1507 establishment & funding proposal, 2021.[55]

Viral evasion strategies were identified at virtually every point of the antigen presentation system, of which a total of five direct TAP inhibitors arised from viral origin.[52] Four of these inhibitors originate from the *herpesviridae* family while two derived from the *orthopoxviridae* respectively *molluscipoxviridae* (Figure 4). Except for the herpesviral infected cell protein 47 (ICP47), all viral inhibitors are membrane-bound proteins. They interact with TAP and block ATP binding (US6[56], BNLF2a[57], CPXV12[58]), initiate proteasomal degradation of conformationally trapped TAP or tapasin (UL49.5[59], MC80[60]) or block the peptide binding pocket (ICP47[61-65], BNFL2a, CPXV12). The type of inhibitor that interacts with TAP also appears to provide allele-specific trapping of MHC I molecules within the PLC.[66] As a soluble protein, ICP47 has the advantage that it is recombinantly expressible and can be added externally to the PLC, which greatly facilitates biochemical analysis. In addition, ICP47 can be extensively modified with affinity tags[11,18], fluorophores[11] or photocleavable amino acids[67,68]. Since it has been used extensively in the present work, the structural design, and features of ICP47 will be discussed in detail below.

ICP47 is an intrinsically unstructured protein comprising 88 amino acids which specifically blocks the peptide binding pocket of TAP with an affinity of K_d = 50 nM.[64] In close proximity to membranes[69] its active binding site (residues 2-34) adapts a helix-loop-helix conformation that inserts into the peptide binding pocket of TAP.[45] This region is succeeded by a section (residues 35-55) that stabilizes the interaction with TAP.[70] When ICP47 is bound to TAP, ATP binding is unimpaired[54,64] while hydrolysis requires a conformational change that is inhibited by TAP fixation in an inward facing conformation.[45] However, this conformational fixation can also be of advantage: elegant PLC purification strategies[11,18] are based on affinity chromatography using ICP47-trapped PLC. It is assumed that the conformational fixation by ICP47 enhances the stability of the PLC during purification and limits the flexibility of this highly dynamic complex to such an extent that structure elucidation is possible at all. Since there has been no way to remove ICP47 from the TAP binding pocket while keeping the highly sensitive PLC intact, TAP and due to the close-meshed interconnection of all PLC components, the entire PLC remain irreversibly conformationally trapped and transport deficient. A part of the present work is therefore devoted to generate a cleavable ICP47 variant by the introduction of photosensitive amino acids, using light as an extremely mild and orthogonal trigger to remove ICP47 from the PLC.[68]

3.4 PLC reconstitution in membrane mimetics

In addition to membrane proteins, cellular membranes contain various phospholipids, glycolipids, and cholesterol. Biochemically, working on membrane proteins is a particular challenge: Membrane proteins are not stable in aqueous solution and their function and activity may already change with variation of the lipid composition of the membrane.[71] Usually, membrane proteins are extracted from their native lipid environment using detergents. When their critical micelle concentration (CMC) is exceeded, detergents form micelles (Figure 5A) that can stabilize the hydrophobic regions of membrane proteins and can thus keep them in solution.[72] However, detergents rarely meet the complex requirements of a native lipid membrane and frequently large libraries must be screened to find suitable detergents that do not irreversibly damage the target protein.[73] With regard to downstream biochemical and especially structural analyses, a palette of alternatives was created to not interfere with these and to mimic the properties of the native lipid membrane more accurately.[74] On the one hand, there are amphipols, short polymers with a strongly hydrophilic backbone complemented by hydrophobic side chains that bind to transmembrane regions.[75-77]
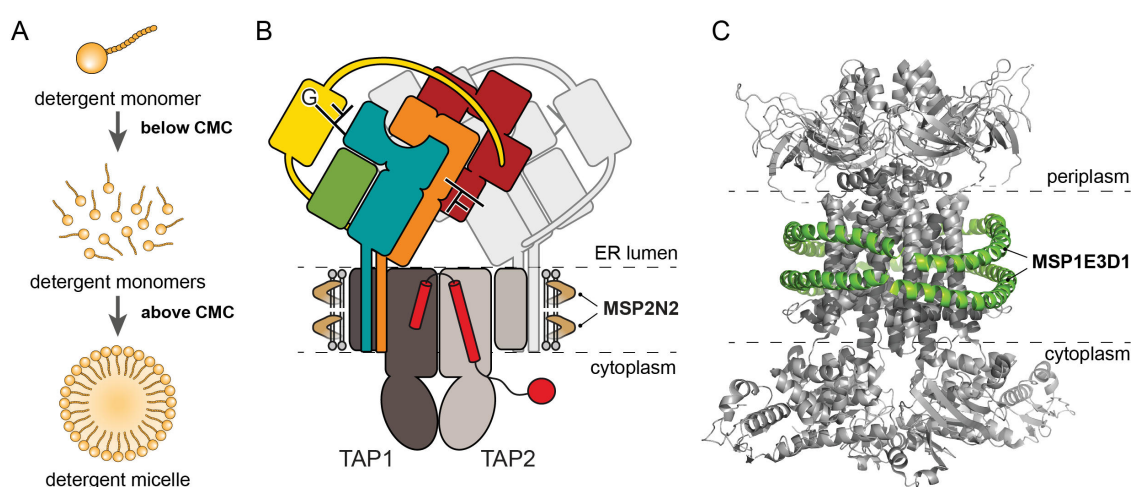

Figure 5: *Nanodiscs to reconstitute ABC transporters for structural investigations.* **A)** Detergent solubilization of membrane proteins is a prerequisite for most membrane mimetic reconstitutions as it is for nanodiscs. **B)** Schematic representation of MSP2N2 nanodiscs stabilizing a large (15-17 nm[78]) lipid belt around the transmembrane region of the PLC. The native orientation is marked with dashed lines. Figure adapted from Domnick, Winter *et al*. Nat Commun, 2022.[11] **C)** Cryo-electron microscopy structure model of the ABC transporter complex MlaFEDB (grey, PDB 6XBD)[79], reconstituted in MSP1E3D1 (~12 nm[80], green) nanodiscs. The native orientation is marked with dashed lines.

Among membrane mimetics, these are the closest to classic detergents. Bicelle forming systems incorporate lipids into their micelles and thus approach native membranes more closely. They form a disc shape and usually consist of a mixture of long, bilayer-forming phospholipids and detergent molecules or short phospholipids.[72,81] On the other hand, there is the field of nano-sized self-assembled lipid bilayers termed nanodiscs.[82] These lipid structures are held together in nanodisc-shape by saposin[83], copolymers (SMA[84],

DIBMA[85]), peptides[86] or membrane scaffold proteins (MSP).[78,87,88] In particular, the latter two will be briefly discussed here, since MSP2N2[11] (Figure 5B), saposin(not published) and peptidiscs(not published) are the only successful PLC reconstitution strategies to date, besides detergent reconstitution in digitonin[73] or glyco-diosgenin (GDN)[18].

The underlying concept of nanodiscs is to provide the protein with a lipid membrane that is as native as possible. This is achieved by one or more domains of the high-density lipoprotein apolipoprotein A1 wrapping around a part of a phospholipid membrane of the target protein (Figure 5C). These protein domains are amphipathic, helical, and are self-assembling but must be complemented with additional lipids to fill the belt. They have been further developed for different belt sizes, so that there are now a variety of different MSPs available in the range of 8-17 nm, which is also dependent on the ration of MSP to lipid. The number after the MSP indicates the number of MSP repeats, while an 'E' indicates an extension, and a 'D' indicates a deletion - respectively to exactly match the desired belt size. For structural and mechanistic studies in this work, the PLC was reconstituted into MSP2N2 nanodiscs consisting of two MSP repeat regions, each consisting of ten helices, forming large nanodiscs with 15-17 nm in diameter.

Peptidiscs, unlike nanodiscs, are flexible in size and do not necessarily require additional lipids to form a disc. In this "one size fits all" approach, the discs are self-assembled from amphipathic peptides arranged in an antiparallel pattern around the hydrophobic domains of the target protein-lipid complex.[86] The peptide belt wraps around the remaining lipids on the target protein, but the size can be adjusted by addition of further lipids.[72] In initial unpublished screening experiments, this approach was able to generate structural and mechanistic results comparable to those obtained with MSP2N2 reconstituted PLC[11].

The reconstitution methods described so far all have a focus on structural analysis (e.g., cryogenic electron microscopy (cryo-EM))[74], as they stabilize the target complex as an individual particle for which the resolution in the Z plane is less important and that can be picked and analyzed in an automated manner. In terms of biochemical and mechanistic characterization, however, proteins reconstituted in this way are usually suboptimal – especially in the case of membrane transporters. Transport activities cannot be adequately mapped without a physical barrier, creating an inside and outside. Another disadvantage is the absence of the natural membrane curvature, which can have an influence on the overall complex conformation and thus also on the activity.

For these analyses, unilamellar liposomes are the reconstitution method of choice.[72,89] Liposomes can be produced in sizes between 20 nm and 10 µm with the advantage of incorporating cellular membrane extracts, defined lipids or artificial phospholipids. The

specific properties of the liposomes used can thus be adapted very precisely to the native membrane[90] or to the specific requirements of the analytical method. For reconstitution, membrane proteins are usually solubilized in detergent and transferred to detergent-destabilized liposomes (Figure 6). The protein-containing liposomes are termed proteoliposomes. Since more than a single target protein can be reconstituted per liposome, it is even possible to analyze complex-formation and clustering of target proteins. The large surface area allows free diffusion within the membrane, which is restricted within disc-forming reconstitution methods.[91] Thus, proteoliposomes allow the analysis of proteins in a setup that already closely resembles a cell, with the advantage that the influence exerted by other cellular components can be excluded. Even effects of different lipid compositions or membrane curvatures can be investigated.[92,93] In structural analysis, proteoliposomes impose much greater challenges than disc-forming reconstitution methods. Although there are examples of proteoliposome analyses using cryo-EM[94,95], they are exceptionally challenging and partly require two-dimensional pre-crystallization[96,97]. Electron cryotomography (Cryo-ET) could represent the ultimate solution for the structural analysis of membrane proteins in proteoliposomes.[98,99] However, the technical barriers are currently still very high[95] - and if they are conquered, structural analysis of membrane proteins could potentially be performed *in situ* in their native cell environment.[100,101]
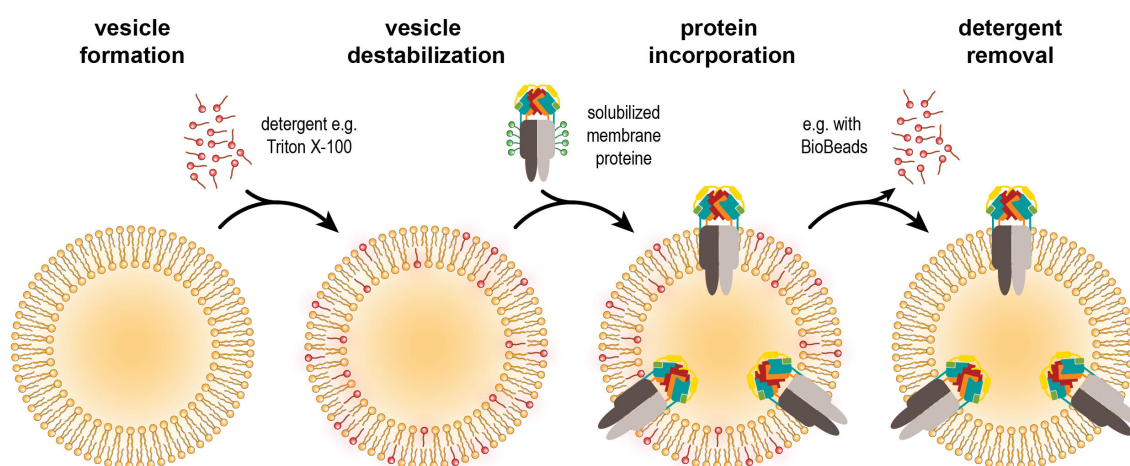

Figure 6: *Generation of proteoliposomes*. Liposomes are destabilized with detergent (e.g., Triton X-100). Detergent-solubilized membrane proteins are mixed with destabilized membrane proteins and the detergent is removed from the solution (e.g., using BioBeads). The membrane protein is reconstituted into liposomes as it cannot be stabilized outside the membrane after detergent removal. The figure was inspired from Johnson and Lee, Methods Enzymol, 2015[102].

In the work reported herein, the PLC was reconstituted in nanodiscs and peptidiscs. Since a light-controlled nanotool for purification and subsequent reactivation of the PLC has also been established, a feasible reconstitution of the active PLC in liposomes has now been achieved for the first time. In combination with novel analytical methods for liposomes

presented below, this approach may allow to study PLC assembly, peptide loading, and decay in a realistic membrane environment with free diffusion. In the future, this approach will provide opportunities to analyze clustering of PLCs[103,104], which can be structurally resolved by cryo-ET. The active NBDs could further allow trapping in different conformations, which may further unravel the mechanistic details of the peptide loading complex.

3.5 Determination of MHC I glycan status in ERQC

The Asn86 N-linked glycan is an essential marker of MHC I quality control and determines the fate of each MHC I molecule. Therefore, an extensive quality control machinery ensures that the correct state of MHC I glycan is maintained. The main actors involved are tapasin/GluII in the PLC, and TAPBPR[32]/UGGT1[105] in the second level of quality control of the ER and Golgi (Figure 7). If the correct loading and folding of the intrinsically unstable MHC I fails or stops, e.g. because no suitable antigen can be obtained, the ER resident mannosidase I (ERmanI) will tag the protein for degradation by removal of a terminal mannose at the B or C branch[106]. MHC I degradation is launched via the ER associated degradation pathway (ERAD).
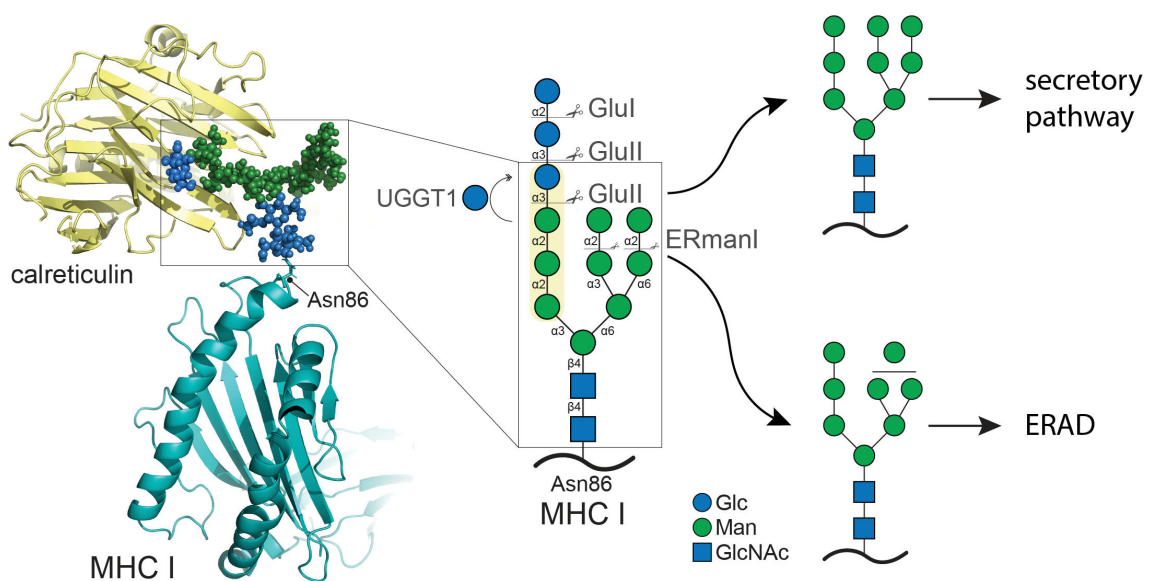

Figure 7: *The MHC I glycan-mediated ER quality control process*. The $Glc_3Man_9GlcNAc_2$ glycan initially transferred to MHC I by OST is subsequently trimmed to $Glc_1Man_9GlcNAc$ by ER resident GluI and GluII, signaling incomplete or suboptimal folding. Upon peptide loading, the remaining glucose is trimmed by GluII, signaling an optimally loaded MHC I and releasing the complex to the secretory pathway. If the folding process of MHC I stalls or fails, mannoses at the B and C branches will be trimmed by ERmanI, assigning MHC I to degradation along the ERAD pathway. In the second level of quality control, the terminal glucose can be reinstalled by UGGT1, which triggers ER-retention or ER-reimport, as appropriate, and triggers a new cycle of quality control in the PLC. The structure model was simulated[17] based on the cryo-EM structure of the human PLC[18]. Parts of the glycan model were adapted from Domnick, Winter *et al.*, Nat Commun, 2022.[11] Glc, glucose; Man, mannose; GlcNAc, N-acetyl glucosamine.

GluII and UGGT1 are opposing components of a quality control cycle - with UGGT1 as the second level of quality control, which re-examines the former substrates of GluII before translocation to their final destination - and returns them in case of folding defects.[107] However, it is not yet clear how exactly ER quality control can distinguish between proteins that could still be correctly folded by further quality control cycles mediated by calnexin/calreticulin interaction and irreparably misfolded proteins.

GluII is composed of a catalytic α-subunit and an auxiliary β-subunit.[36,108] The soluble heterodimer trims the first α3-bound glucose from nascent glycoproteins and also the second glucose as part of the ER quality control (ERQC).[109] The first cleavage step is performed considerably faster than the removal of the second glucose, so that calnexin/calreticulin can incorporate the target proteins into the ERQC pathway.[110] In this way, misfolded or incompletely folded glycoproteins are prevented from exiting the ER prematurely.[107] The β-subunit contacts various client proteins and contains an ER retention sequence, which retains the heterodimer in the ER.[111] Various studies have investigated whether the mannose structure of the B and C branches has an influence on the recognition and deglucosylation activity of GluII.[112,113] In line with recent publications, enzymatic activity at lower mannose species was also verified in this work.[11]

UGGT1 as a counterpart to GluII is a large (170 kDa), monomeric, and soluble glycosyltransferase and present in both the ER and Golgi.[114] It comprises two sections with the large, N-terminal section likely representing the folding sensor and mediating interaction with target molecules.[115,116] The smaller C-terminal section contains the catalytic domain that binds uridine diphosphate glucose (UDP-Glc). UGGT1 utilizes the UDP-Glc to transfer glucose to the terminal A branch mannose of its target molecules, thereby requiring high concentrations of calcium.[107] UGGT1 is a highly sensitive folding sensor that can recognize and reglucosylate misfolded N-glycosylated proteins as its substrate.[105] In particular, hydrophobic sections of the client proteins appear to trigger recognition by UGGT1. UGGT1 is unique in its ability to combine both a chaperone function and a glycosyltransferase function. However, UGGT1 recognizes primarily advanced folded client proteins with folding defects and not early folding intermediates, which places UGGT1 at the end of the ERQC process.[117] UGGT1 has the highest glycosyltransferase activity towards $Man_9GlcNAc_2$ glycans, which readily decreases by 50% towards $Man_8GlcNAc_2$ and only a fraction of the original activity remains towards shorter mannose species.[118]

Recently, UGGT1 was shown to interact with TAPBPR to provide complementation to tapasin and GluII.[35] TAPBPR is understood to be a tapasin homolog with analogous function and structure.[119] However, the function of TAPBPR does not appear to be as

drastic as that of tapasin, but instead seems to be specialized for fine tuning the peptide-cargos on MHC I.[32] UGGT1 might also recognize misfolded MHC I molecules without the assistance of TAPBPR.[120] Putatively, TAPBPR could provide a bridge for fine-tuning pMHC I quality control and mediate additional substrate specificity through UGGT1 interaction. Thus, different conformations of suboptimal charged MHC I molecules might trigger UGGT1 activity when complexed with TAPBPR.[35]

Within the ER quality control cycle, many questions remain unanswered, especially in regard to the accurate sequence of events. With respect to the PLC, it remains unknown when and where exactly GluII removes the terminal glucose. It is not clear whether this process can be performed within the PLC and if there are any special preconditions for this process. Within the work performed herein, these unresolved issues will be addressed.[11] With respect to the second level of MHC I quality control, apart from the evidence of interaction between TAPBPR and UGGT1, the mechanism of interaction and the precise role that TAPBPR may play is unknown. These aspects are also tackled in the section of compiled publications.[37]

In order to make well-founded statements about the glycan status of MHC I molecules, it is a prerequisite to be able to identify the glycan structure and composition. Modern ways of glycan identification usually employ mass spectrometry (MS) to identify the glycan pattern. Information about the glycan pattern and their spatial arrangement can also be obtained without the assistance of mass spectrometry - for example, using normal phase high-performance liquid chromatography (HPLC) for which a prior glycan release from the target molecule is required.[121,122] However, these methods are typically considerably less sensitive and accurate compared to MS approaches. MS methods can basically be divided into three different areas: bottom-up, middle-down, and top-down glycoproteomics[123], of which the advantages and disadvantages will be briefly explained below:

The vast majority of protein-MS analyses and their post-translational modifications are based on bottom-up approaches. Bottom-up refers to the fact that the protein or protein mixture to be analyzed is first broken down chemically or enzymatically into smaller components before it is determined by MS.[124] Complex samples can be separated beforehand by gel electrophoresis, so that the proteins of interest can be cut out and processed individually. However, modern liquid chromatography (LC) separation methods and powerful analytical pipelines can also analyze complex mixtures with hundreds of peptides. A common enzyme for digesting protein samples is for instance trypsin, as it produces peptides with an average length of 14 amino acids, which is within the ideal length of 6-50 amino acids for computational analysis.[125] Using fragmentation in MS/MS measurements, in addition to the exact mass of the peptides, their amino acid sequence

and post-translational modifications can be determined. This information can then be used to infer the identity of the intact proteins, usually in conjunction with protein databases. The bottom-up approach has several clear advantages: The requirements on separation performance and resolving power of the mass spectrometer are reasonable: The generated peptides are easy to separate chromatographically and possess a low number of maximum charges due to their length, which makes bottom-up samples relatively homogeneous, independent of the parent proteins. The low molecular mass of the individual peptides also places low demands on the resolving power of the mass spectrometer to generate isotope-resolved raw spectra. In addition, positions of post-translational modifications can be readily determined. The lower limit of detection is in the range of pg/mL.[126] However, by digesting the target protein into smaller fragments, information about the biologically active entity is lost. Only part of the peptides produced are retained during sample preparation and analysis. In particular, peptides carrying hydrophilic posttranslational modifications such as glycans are underrepresented.[123] Glycans can frequently be investigated only in specialized assays, which require prior release from glycoproteins and derivatization with a variety of reagents for analysis.[127] Without prior purification of the individual glycoproteins, the released glycans can no longer be assigned to their source proteins.

Top-down approaches, in contrast, analyze undigested, active proteins, thus reflecting their biologically relevant form more accurately.[128] Measurements of native complexes may even allow conclusions to be drawn about protein-protein interactions.[129] All post-translational modifications and biotransformations are represented by the intact mass and cannot be lost during sample preparation. In addition to protein purification, further sample preparation is not necessary for basic measurements (Figure 8), and protein mixtures can be applied directly to the LC-MS system. However, there are complex requirements for LC: three-dimensional structures must be denatured at high temperatures. Protein samples are very inhomogeneous depending on the type of protein and sometimes elute at high concentrations of the organic solvent, which impedes effective ionization in the ion source.[123,130] Membrane proteins present a particular challenge in this respect. Coelution of proteins should be avoided, either by limiting the maximum number of proteins that can be analyzed simultaneously or by using more sophisticated separation methods such as ultra-performance liquid chromatography (UPLC). Due to their high intact mass, these samples also place higher demands on the resolving power and the measuring range of the mass spectrometer. Due to the broad charge envelopes of larger proteins, the peak intensity is distributed among many individual signals, which reduces the sensitivity compared to smaller molecules, so that

the sensitivity can range from ng/mL to µg/mL, depending on the analyte.[126] For non-isotopically resolved spectra, as is common for proteins with high molecular weights, reconstruction of the total uncharged mass from the charge envelope using deconvolution is required (Figure 8). Modern deconvolution strategies yield identical results, with differences being only subtle.[131] For a meaningful analysis of post-translational modifications, however, the total masses of the proteins must first be properly assigned to the protein sequence, since the top-down approach does not allow sequencing by fragmentation. To determine the total mass of glycoproteins, a Peptide:N-glycosidase F (PNGase F) digest can be helpful, which removes all glycans tracelessly.[132] A further limitation is apparent in the analysis of glycan patterns: Here, the exact spatial structure, as well as the composition, cannot be readily determined and must be combined with other analytical methods. For example, mannose and glucose are constitutional isomers and therefore cannot be directly distinguished by their mass.
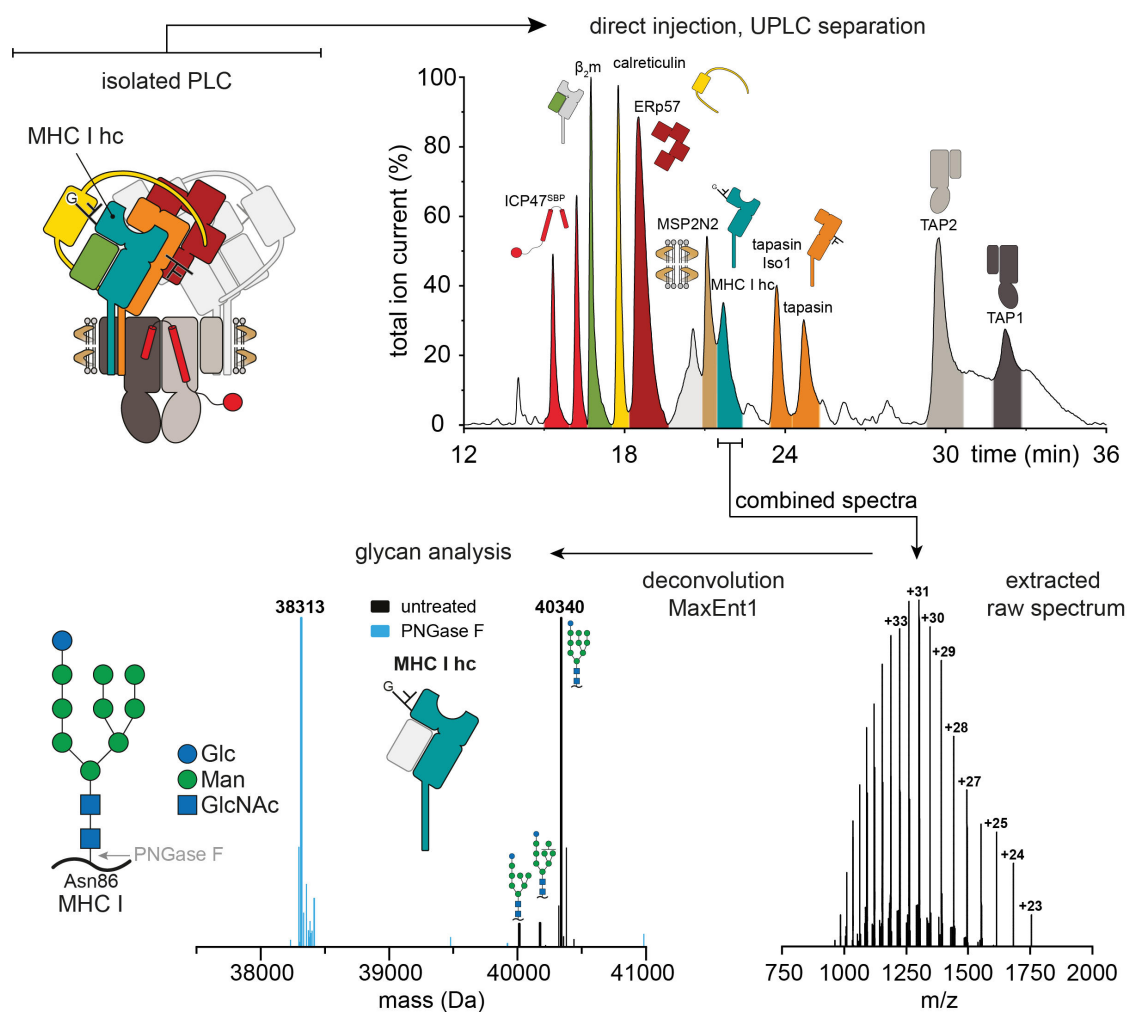
Figure 8: *Intact protein mass spectrometry workflow*. A protein mixture or protein-containing reaction mixture is injected into the chromatography system at denaturing temperature without preliminary treatment. Reactions are stopped by heat and spatial segregation of the components on the reverse phase column. Combined raw spectra are extracted for individual components. Charge envelopes are deconvoluted by MaxEnt1 algorithm to yield the uncharged mass. Glycans can now be identified by glycan mass summed with the total mass of the protein. Figure adapted from Domnick, Winter *et al.*, Nat Commun, 2022.[11]

Middle-down approaches and try to take advantage of the strengths of both procedures. The approach is particularly applied to analyze antibodies where glycan modifications are also of high scientific and therapeutic interest. The approach relies on a limited digestion that preserves large portions of the target molecule. For example, IgG antibodies can be enzymatically cleaved into parts with a molecular weight of 25 kDa[133], which can readily enable isotope-resolved mass analysis.[126]

In this work mass spectrometry was employed to mechanistically trace the GluII and UGGT1-mediated quality control machineries. The top-down approach was chosen for the analysis of intact proteins, but in particular also for the analysis of glycan patterns on MHC I and tapasin (Figure 8).[11,37,68] The protein complex is denatured and separated by UPLC without requiring pre-treatment. By combining a series of individual raw spectra, a charge envelope for the individual target protein is extracted and subsequently deconvoluted using a quantitative, entropy-based algorithm (MaxEnt1).[134] In MaxEnt1 the combined raw spectra are decomposed into vectors and a probability map of the peak distances of the charge envelope is generated. This map is then iteratively compared with models of charge distributions of the parent mass. In all analyses of this work these models were iterated until the fit converged.

An advantage of this top-down approach was the direct applicability after *in vitro* glycan processing by GluII or UGGT1 without further preparation of the sensitive sample, which excluded additional sources of error, interference, or alteration of the reaction conditions. Furthermore, the PLC contains tapasin, an additional glycoprotein with very unusual glycan modifications[11], which would interfere with MHC I glycans in an assay relying on glycan release. Since the general structure of glycans at MHC I or tapasin was already known and could be correlated to recent structural biology findings[11], it was not necessary to determine the precise spatial linkage of mannose or glucose molecules. The inability to analytically distinguish between glucose and mannose entities is a shortcoming of this approach. However, indirect glucose detection allows sufficiently accurate identification, rendering the top-down approach for the determination of MHC I glycans the fastest and most reliable approach overall.

3.6 Techniques in structural biology

The field of structural biology is currently dominated by three different techniques for 3D protein structure determination: X-ray crystallography, cryogenic electron microscopy (cryo-EM) and three-dimensional (3D) nuclear magnetic resonance spectroscopy (NMR), which are briefly discussed below.

With the beginning of the 1940s, X-ray crystallography initiated a revolution in structural biology or made it possible at all by today's standards.[135] Before the cryo-EM era dawned in 2014, nearly 90% (86% in 2022)[136] of all protein structures were obtained using X-ray crystallography, highlighting the tremendous importance of this method to science. In X-ray crystallography, either endogenous proteins are purified and concentrated or - since very high protein concentrations are usually required for crystallization - recombinant expression is used. For structure determination, the highly concentrated protein must be crystallized. However, in contrast to small molecules, proteins usually do not tend to crystallize, as in a cellular system this must be avoided urgently. Therefore, the target protein must be forced to crystallize. In almost all cases this requires extensive systematic screening of varying buffer and environmental conditions. However, even then not all proteins can be crystallized. Protein engineering is thus required to eliminate potential interfering factors in the crystallization process. These can be post-translational modifications such as glycosylation or the removal of flexible N and C termini or, with regard to membrane proteins, the deletion of the transmembrane region. In X-ray crystallography, sample preparation usually accounts for a substantial part of the total analysis time or fails when no suitable crystallization condition can be found. Protein crystal imaging requires extremely high X-ray energy at tunable wavelengths, which is usually provided by synchrotrons. The protein crystal is frozen at ~100 K prior imaging in order to limit the damage caused by the radiation and to be able to ideally record a complete a data set from a single crystal.[135] During data acquisition the oscillation/rotation protocol is still used today.[137] The crystalline structure of the sample causes the incident X-ray beam to diffract in specific directions (Figure 9A), which are recorded by charge-coupled or pixel array detectors.[138] In addition to the information of the diffraction spots intensities, the phase of these spots must be solved - additional information that can be obtained for example computationally by molecular replacement.[139] An electron density map can then be generated by a series of Fourier transformations on the basis of the recorded data. This electron density map ideally reveals the position of every atom in the structure so that a model of the protein can be fitted. The demanding sample preparation and the absolute requirement that the protein can be crystallized makes it especially difficult to keep up with the structure determination of membrane proteins. As a result, only about 1% of all published X-ray structures are of membrane proteins.[135]

NMR for structure elucidation relies on the principle that atomic nuclei with half-integer spin behave like tiny bar magnets. These nuclei include the ubiquitously present ^{1}H as well as the ^{13}C and ^{15}N atoms that are easily introduced into biological systems. In a strong

magnetic field, the nuclei can adopt a preferred orientation that has different energy contents. An electromagnetic pulse can thereupon induce an excited state in the atomic nuclei, which relaxes in an identical frequency. Since the field around each individual atom is influenced by its neighboring atoms, this resonance of the excited nucleus shows minimal deviations.[140] Based on these deviations (chemical shifts), the exact chemical environment of each resonant nucleus can be deduced. In 3D NMR, these resonance shifts are transferred to three different species (mostly C, N, and H). These can then be computationally resolved into peptide chains and ultimately three-dimensional interactions at atomic resolution.[141] In stark contrast to other imaging methods in structural biology, the analysis is indirect and does not damage the target molecule. However, the signal intensity strongly diminishes for larger molecules, which makes 3D-NMR mainly suitable for smaller proteins. This is also one reason why 3D-NMR, in contrast to the other methods, has been a niche application when comparing the number of solved structures.[136] Nevertheless, especially in the time before cryo-EM, it has been a valuable tool to study structures that cannot crystallize.[142]

In contrast to X-ray crystallography, transmission cryo-EM uses electrons to image the specimen. These high-energy electrons directly interact with the Coulomb potential of individual molecules. Focusing them on a back focal plane directly provides two-dimensional micrographs with structural information about the specimen without the need for crystallization (Figure 9B). Since the target molecules are ideally arranged in random orientations, a 3D reconstruction can be built from a large set of high-resolution 2D micrographs of individual orientations.[143] Given that the trajectories of the electrons are affected by all substances, the measurement is performed under extreme vacuum. In order to keep the biological sample intact and hydrated in this environment and to limit sample movement during the imaging process, it is embedded in an ultra-thin layer of amorphous ice and the micrographs are recorded using liquid nitrogen cooling. Proteins are plunge frozen on cryo-EM grids with a perforated carbon film. This method was developed by Dubochet *et al.* and still allows efficient sample preparation today.[144,145] The protein sample is applied to a cryo-EM grid and the excess liquid is removed with the aid of an absorbent (blotting) paper. This leaves just enough liquid for the sample to be retained by the surface tension of the water inside the holes of the carbon film. Rapid immersion by plunging in liquid ethane causes the sample to vitrify at -146 °C. Since the biological samples in cryo-EM are frozen without further pretreatment, the native conformation may be better preserved than in artificial crystallization environments of X-ray crystallography.[146] The "resolution revolution"[147] in 2014 driven by the development of direct electronic detectors, as well as the development of automated and user-friendly

data analysis toolsets[148-151] led to an exponential growth in the field of single particle cryo-EM.[136]

Overall, there is no single best method for structural elucidation of biomolecules. NMR, for example, suites well for smaller molecules but struggles to resolve large proteins and protein complexes, whereas in cryo-EM small molecules often yield low-contrast micrographs and large complexes are more amenable for analysis.
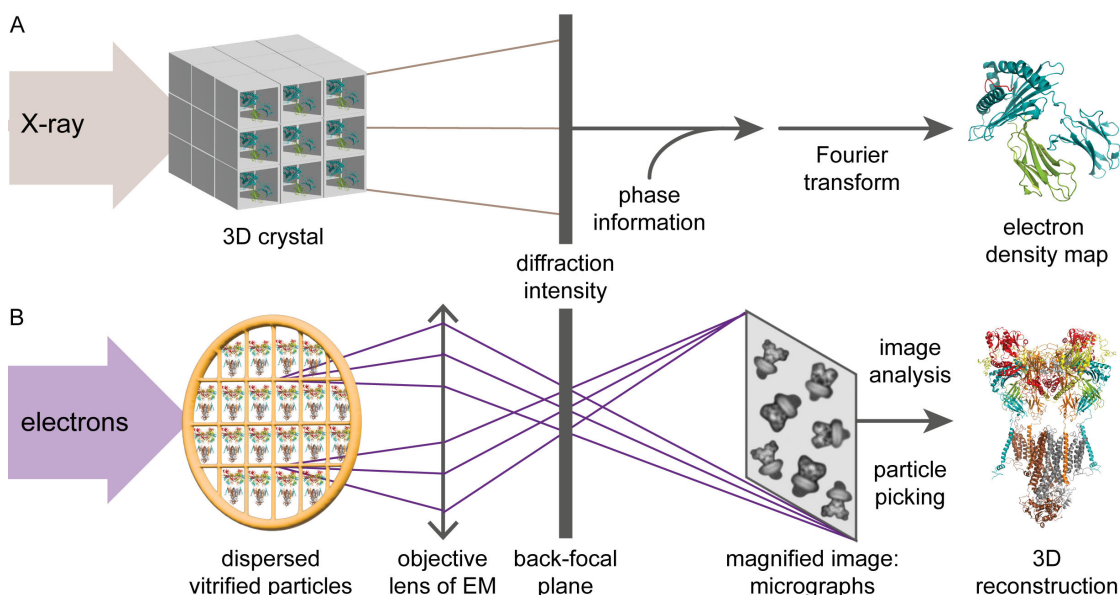

Figure 9: *Technical details of X-ray crystallography and cryo-EM.* **A)** Structure determination by X-ray crystallography. Incident X-rays are diffracted by a protein crystal in which molecules are arranged periodically in three dimensions. Since the X-rays are not focused and thus only the intensities are recorded, the diffraction pattern must be supplemented with additional information to solve the phase problem. A three-dimensional structure can subsequently be fitted computationally using a series of Fourier transforms. **B)** Structure determination by cryo-EM. Vitrified molecules of a protein sample diffract irradiated electrons and are thus directly imaged. 2D projections (micrographs) of hundreds of thousands of such molecules in different orientations can be reconstructed into a three-dimensional structural model. Figure was adapted from Wang *et al.*, Protein Sci, 2017.[152]

X-ray crystallography can be used to structurally determine both small and large molecules but only if they can be crystallized and diffract to the required resolution - in which case it is not clear whether the conformation is selected or forced, for example via crystal contacts. Thus, methods described for structure elucidation have their strengths and weaknesses and complement each other.[152] By now, all approaches of biological structure elucidation can be supported by the latest revolution in this field: AlphaFold as an artificial intelligence assisted protein structure predictor.[153] The sheer bulk of over 200 million different proteins known renders it impossible to solve all structures with the aid of established methods of structure elucidation.[154] AlphaFold, as a rapid tool freely available to scientists, can make highly accurate structural predictions in a remarkably short time solely based on the amino acid sequence. The evolution of these tools is proceeding at a

rapid pace, and what was recently considered paradoxical[155] can become reality at any time.

3.7 Photocontrolled processes in biological environments

Within biological environments, processes and signal cascades are precisely regulated in time and space. For the scientific characterization of these systems, an equally accurate control of these biological processes is necessary. Compared to chemical and physical intervention methods, light is a particularly elegant method for controlling biological processes, since modern optical systems offer a very high spatial and temporal resolution for the guidance of light pulses. The intensity of the irradiated light can similarly be accurately adjusted covering several orders of magnitude. Chemical and physical intervention methods have not yet achieved this precision and frequently struggle to pass through membranes and thus deliver signal cascades to the interior of a biological system.

The methods of photoactivation can be divided into three groups: Reversible modulation of molecules or processes by photoswitches, irreversible activation or inactivation by uncaging of photo-protected molecules, and genetically encoded, light-controlled elements, encompassing the discipline of optogenetics.[156]

The field of photoswitches is largely based on organic photochromic systems in which light-driven isomerization reactions are triggered. The subsequent conformational change, which also imposes altered steric demands on their environment, can trigger or inhibit cellular processes. The most commonly used representatives of this category are azobenzenes, in which light-induced *cis/trans* isomerization of their N=N bond can be triggered. The molecules are transposed from the planar *trans* configuration to the 55° inclined *cis* configuration, accompanied by an end-to-end distance change of 3.5 Å (in azobenzene-modified peptide backbones).[157] This conformational change can consequently induce functional changes in the target molecule. The great advantage of photoswitches of an actual on-off control of the target system is opposed by the general disadvantage that a complete conversion of all photoswitchable molecules into the active or inactive configuration cannot be achieved and the yields are usually limited to a maximum of 80%.[156] Thus, these forms of switches are particularly useful when a full conversion is not necessary, or when the switch is coupled to non-linear (transistor-like) behavior or processes, for whose activation just a certain signal amplitude has to be exceeded.[158]

In contrast to photoswitches, photocages can typically be activated or inactivated once, depending on the application. Photocaging is achieved by positioning a photolabile

protective group at a key position in the molecule. Due to chemical inertness and suppression of reactions with the protected entity or the high steric demands of the protecting group, binding or reaction is suppressed. Probably the best known example of a photocaged molecule is a light-activatable ATP, whereby theoretically any ATP-dependent process can be activated by uncaging ATP with light.[159] Since there is a vast number of photocaged molecules with diverse activation methods[160], the following section will focus on photocaged peptides and proteins which are modulated by *ortho*-nitrobenzyl (ONB) based moieties.

Photo(de)activatable peptides and proteins can exist in different forms: either native protein functions or binding motifs are specifically blocked by the introduction of photocages, or the structure of the biomolecule is modified by uncaging so that a biological function is activated or deactivated. In this context, photodegradable peptides will be discussed in detail (Figure 10), as they have been used in several of the studies conducted in this thesis.[8,37,68] In this process, an unnatural and photocleavable ONB-based β-amino acid is incorporated into the primary amino acid sequence of the peptide or protein. Exposure to ultraviolet (UV) light can subsequently break the primary structure of the polypeptide chain, usually causing the biomolecule to lose its function (Figure 10A).[8,67,68] However, it is equally possible to release bioactive peptides from larger structures. ONB-based photocages are among the most commonly used photocages[160], despite their comparatively very poor photochemical properties[161] and the excitation maximum in the UV range at 365 nm, which can induce adverse side effects in biological systems. Many ONB-based derivatives are commercially available, which avoids complex synthesis so that they can be directly incorporated into the target molecules. However, so far there is no method of incorporating backbone cleaving amino acids into biomolecules of living organisms by genetic modification. The method of choice is therefore to incorporate these amino acids by Fmoc solid-phase peptide synthesis (SPPS) and manually deliver them to the target molecule.[8,67,68,162,163] The applicability of these at least partially synthetic peptides and proteins is thus limited to *in vitro* assays in isolated systems - unless sophisticated biomolecule delivery methods are employed.[164,165]

The mechanism of ONB photocleavage (Figure 10B) has been extensively characterized, for example by femtosecond spectroscopy[166]. However, many of these characterizations have been performed in organic solvents rather than in aqueous environments, which may result in different reaction mechanisms occurring in aqueous biological systems. The ground state of the molecule is raised to an excited state after UV irradiation. In a Norrish type II reaction[167], an intramolecular hydrogen atom is subsequently abstracted by the excited nitro group and an *aci*-nitro (Z or E tautomer) intermediate is formed.[160] The

stability of this tautomer and the reaction back to the ground state are strongly dependent on the substitution and in particularly on the pH.[168] By means of LC-MS analyses, it was also confirmed in this work that the photoreaction benefits strongly from acidic pH values, which ideally should be below a pH of 6.5 (Figure 10C). The *aci*-nitro intermediate then forms an irreversible closed conformation and the ring opening leads to the abstraction of the leaving group via a hemiacetal. A transition state that can form the *aci*-nitro tautomer is a radical intermediate.[160] Consequently, ONB-substituted biomolecules are sensitive to radical environments, which similarly to excitation by light can induce the *aci*-nitro intermediate and thus the cleavage reaction.[68,169]
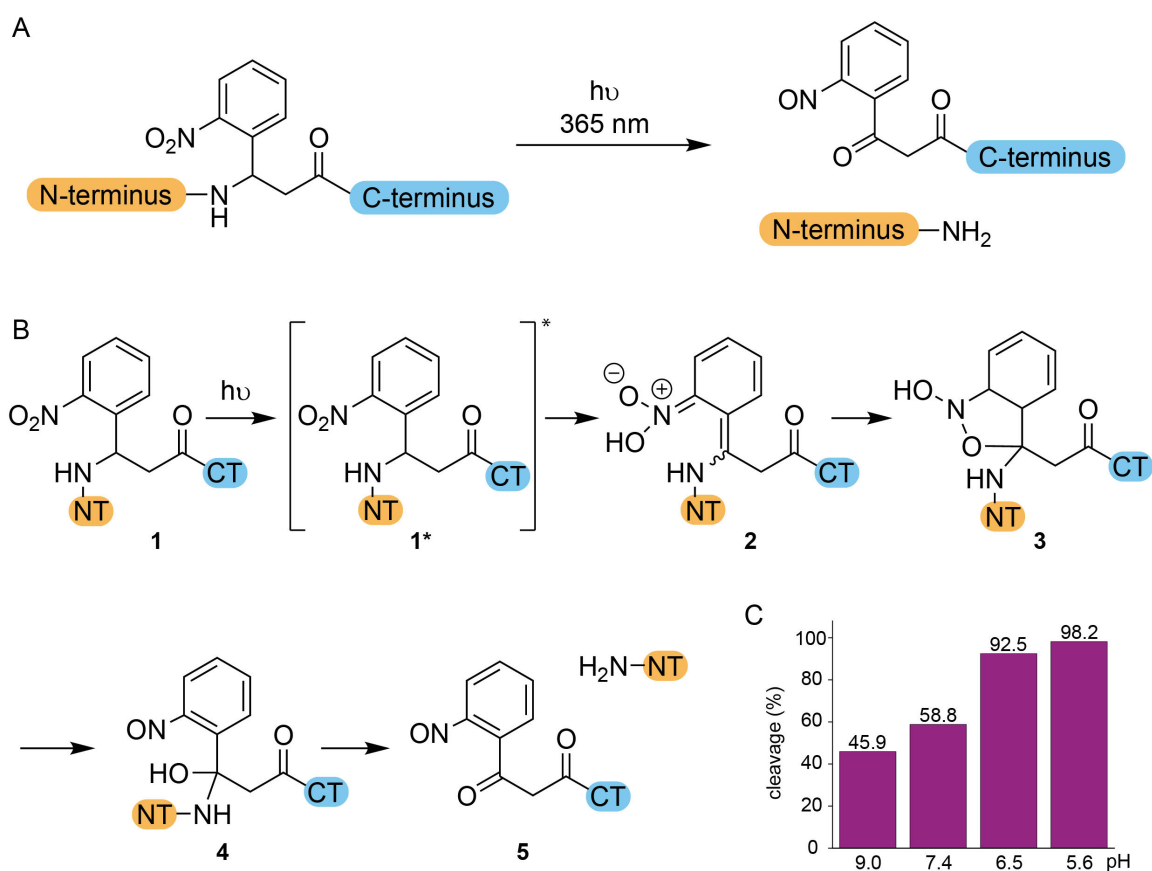

Figure 10: *ONB based photodegradable peptides and reaction mechanism.* **A)** A peptide in which the unnatural β-amino acid Anp (3-amino-3-(2-nitrophenyl)-propanoic acid) has been incorporated. Excitation with UV light (hυ) cleaves the peptide backbone at the position of the unnatural amino acid. Two peptide fragments are released. **B)** Detailed photocleavage mechanism of an ONB containing polypeptide. The ground state of ONB (1) is converted to an excited state (1*) by UV light. An *aci*-nitro intermediate (2) can then perform an irreversible ring closure (3) and via a hemiacetal (4) release the leaving group (NT, 5). CT, C terminus, NT, N terminus. Figure was adapted from Pengfei Wang, Asian J Org Chem, 2013.[168] **C)** Cleavage of a peptide as shown in A) analyzed at different pH values by LC-MS. The photocleavage is diminished at basic pH, while full cleavage can be achieved at acidic pH.

The third group of methods for photoactivation outlined here comprises genetically encoded, light-controlled elements. One particularly successful approach among all others is genetic code expansion via amber suppression. Genetic code expansion enables the synthesis of proteins with unnatural amino acids within living cells.[170,171] These amino acids can contain a modified reactive moiety in their side chain, e.g. for click chemistry as

a lock-and-key element.[172-174] Moreover, it is possible to introduce small fluorophores[175,176] or even caged amino acids[177,178] for light activation. The genetic code, which has near universal applicability in most domains of life, contains a total of 64 triplet codons to encode 20 canonical amino acids. During translation within the ribosome, the messenger ribonucleic acid (mRNA) triplet codons are complemented by transfer ribonucleic acids (tRNAs) which deliver amino acids for the nascent polypeptide chain. Each tRNA is initially loaded with an associated amino acid by its specific aminoacyl tRNA synthetase. Among the 64 triplet codons, 61 code for tRNAs. The remaining three codons are stop codons and are referred to as opal (UGA), ochre (UAA), and amber (UAG) codons. Stop codons initiate the binding of release factors by the absence of matching tRNAs and thus facilitate the release of the polypeptide chain. Amber suppression takes advantage of the fact that stop codons are comparatively rare in the genome and that the amber codon is by far the least abundant codon in the genome. However, the proportions can be different depending on the target species.
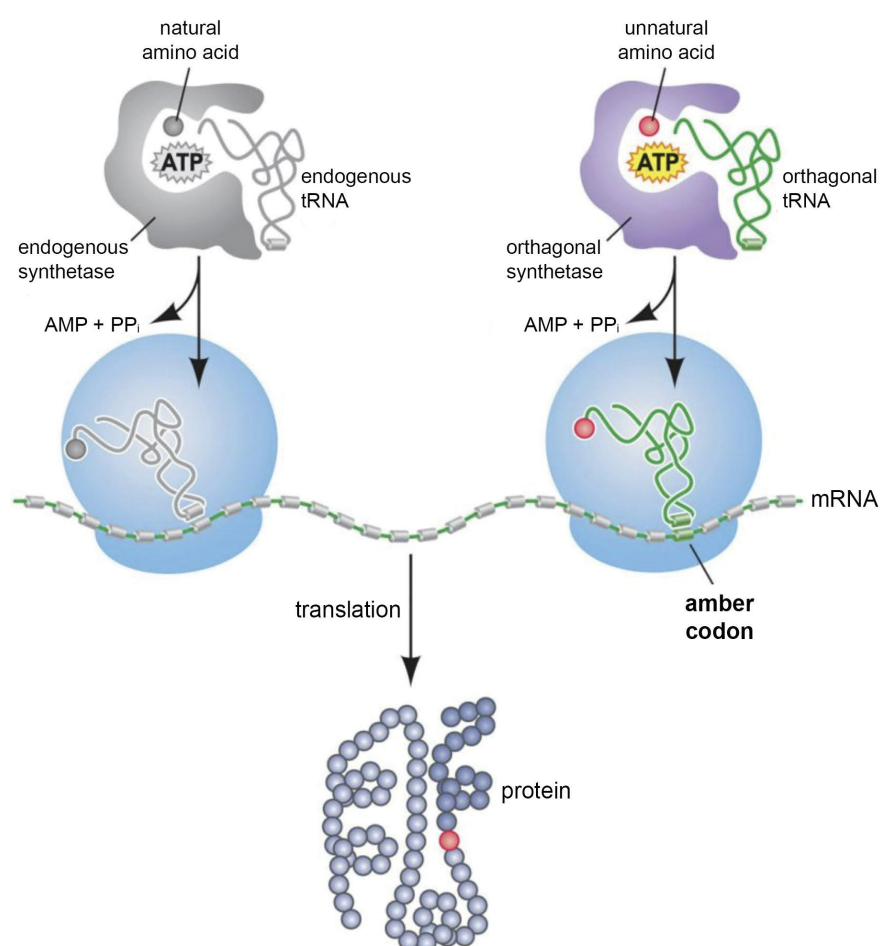
Figure 11: *Principle of genetic code expansion*. Natural amino acids are delivered to the ribosome during translation by loaded tRNAs via their specific aminoacyl tRNA synthetase and recognize their dedicated mRNA triplet codon. Utilizing amber suppression, unnatural amino acids can be loaded via modified tRNA/aminoacyl tRNA synthetase pairs and subsequently be incorporated into the nascent polypeptide chain via the amber codon (UAG). Figure adapted from Lei Wang, Acc Chem Res, 2017.[179]

Consequently, genetic code expansion via amber suppression is based on the introduction of an additional tRNA/aminoacyl tRNA synthetase pair into the target system, which has been modified by protein engineering to recognize amber as a complementary codon to load a specific unnatural amino acid onto the tRNA and ultimately introduce it into the growing polypeptide chain (Figure 11). The site-specific incorporation of the unnatural amino acid is achieved by genetic engineering of the corresponding gene segment. Additional methods of genetic code expansion, such as the introduction of orthogonal ribosomes[180,181] or the expansion of the genetic code to codons with four base pairs[182], will not be further discussed here.

Although amber suppression offers numerous options for the specific *in vivo* modification of proteins, the deep intervention into the translation machinery of the cells also causes negative effects.[170] By the presence of tRNAs for a stop codon, the mRNA continues to be translated and instead of terminating, an unnatural amino acid is incorporated with unknown consequences. While the impact on *E. coli* with an abundance of the amber codon of only about 0.03-9% may be moderate, the impact on mammalian cells with an abundance of about 23% is unpredictable.[182,183] Consequently, as a model system for amber suppression, *E. coli* strains have been generated in which all amber codons were replaced by ochre or opal codons.[184,185] This approach was further developed with extreme effort to the point that there is now an *E. coli* strain with an entirely artificial genome that only utilizes 61 codons and thus even provides three different unused codons for genetic code expansion and thus for the introduction of unnatural amino acids.[183]

The cell can potentially react to the presence of modified tRNAs - for example, an artificial tRNA could be recognized and loaded by endogenous aminoacyl tRNA synthetases. Also, the unnatural amino acid might be loaded onto cellular tRNAs by endogenous aminoacyl tRNA synthetases. To avoid this, tRNA/aminoacyl tRNA synthetase pairs are used that are as phylogenetically distant as possible from the target organism: Hence, orthogonal tRNA/aminoacyl tRNA synthetase pairs from archaea are commonly used for amber suppression in bacteria and eukaryotes.[186] These are then further adapted to the target system by directed evolution.[187] Among the most commonly used systems are modified pyrrolysine incorporating tRNA/aminoacyl tRNA synthetase (PylT/PylRS) pairs originating from the archaea *Methanosarcina barkeri* respectively *mazei*, which were also applied in the work described here.[188,189]

Dozens of different amino acid derivatives are currently available for incorporation by amber suppression.[190-192] The development of these molecules must be carried out within certain limits: On the one hand, the side chain is limited by steric demands, for example within the ribosome. On the other hand, the synthetase must be able to bind and transfer

the molecule effectively to the tRNA. If the transfer of the unnatural amino acid by the aminoacyl tRNA synthetase is slow, sufficient mature tRNAs may not be available. In both situations, the tRNAs compete with release factors, which can lead to a premature termination of translation and thus to inefficient amber suppression. So far, tyrosine[177,178], serine[193], lysine[194], and cysteine[195] derivatives are available for the incorporation of photocaged amino acids by means of amber suppression systems. Despite their poor optical performance, ONB cages are frequently used here, since photocages with better excitation properties are typically sterically more demanding.[177,178,187,193] The underlying concept of this type of modification is that binding motifs can be reactivated after photocleavage by restoring the canonical amino acid structure. Therefore, protein-protein interactions and conformations can be controlled by light in high spatial and temporal resolution in both bacteria and mammalian cells. The guidance of modified proteins by light to specific cellular sites e.g., to trigger signaling cascades or bind target molecules, is a sophisticated model system for amber suppression and has been realized in parts of this thesis.[188,189]

3.8   Nanobodies to modulate and track cellular processes

In this work, a model system of genetic code expansion was demonstrated using light guided antibody fragments. Conventional IgG antibodies are 150 kDa sized, Y-shaped glycoproteins with two symmetric polypeptide chains, each comprising one antigen binding site. Every chain consists of a heavy chain with three immunoglobulin-fold constant domains (C_H) and one variable domain (V_H) as well as a light chain composed of one constant domain (C_L) and one variable domain (V_L) as illustrated in Figure 12. The light and heavy chains are bridged by disulfides, as are the heavy chains by intrachain bonds. Antibodies are generated by the adaptive immune system to bind and neutralize their target molecule bearing an unique binding motif. For this purpose, an antibody binds its target with high affinity and specificity, which is mediated by the paratopic region of the variable domains. After binding, the constant region can trigger signaling cascades to other parts of the immune system.[196,197]

In biochemistry, antibodies are ubiquitously applied due to their unmatched flexibility combined with excellent specificity. By targeted immunization of host mammals, polyclonal antibodies with specific target molecules can be obtained upon affinity purification.[198,199] Monoclonal antibodies that recognize a single epitope and thus have identical paratope regions are generated via immortalized hybridoma cell lines.[200] While antibodies are used in a plethora of *ex vivo* experiments such as immunoblotting, ELISA assays or for affinity purification, they can also be used for *in vivo* assays such as flow

cytometry or fluorescence microscopy. This is possible since antibodies can be labeled with other molecules, such as fluorophores or affinity tags like biotin.[201] Ultimately, even when labeled with small molecule fluorophores, antibodies remain comparatively large molecules that, on the one hand, exhibit low spatial localization accuracy of the target molecule in high-resolution microscopy[104] and, on the other hand, do not penetrate well into tissue. Also, due to steric constraints, some epitopes may not be accessible. Therefore, research was conducted on options to reduce the size of antibodies. One method of antibody truncation involves the enzyme papain and generates Fab fragments (Figure 12).[202]
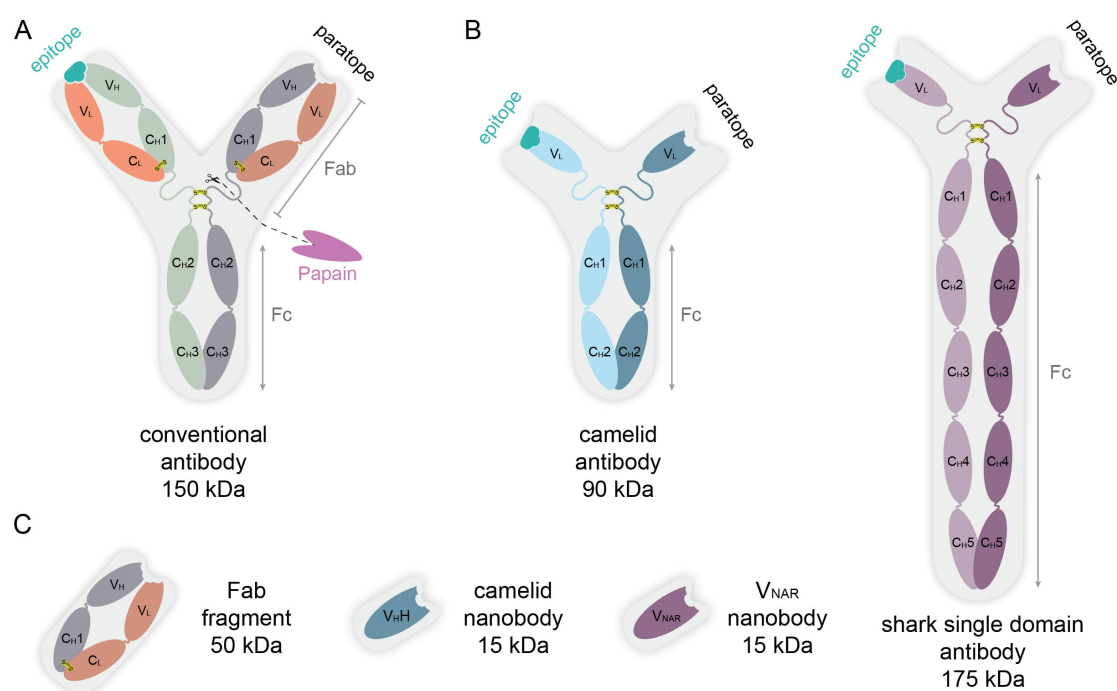

Figure 12: Schematic illustration of conventional and heavy chain only antibodies. **A)** Y-shaped, conventional, 150 kDa IgG antibodies consist of two heavy and two light chains, each composed of several constant (C_H, C_L) and single variable domains (V_H, V_L). The paratope region is composed of the variable regions of the heavy and light chain (V_H and V_L). **B)** Heavy chain only antibodies from camelids or cartilaginous fish, in contrast, have no light chain and consist exclusively of two heavy polypeptide chains. Therefore, the paratope region consists of only a single V_H element with full epitope binding capability. **C)** Papain can be used enzymatically generate truncated antibodies (Fab Fragments), which are only one-third the size of intact antibodies. In contrast, from heavy chain only antibodies, much smaller nanobodies can be generated recombinantly (V_HH for camelid nanobodies, V_{NAR} for cartilaginous fish derived nanobodies), which effectively bind their designated epitope despite their small size.

Fab fragments, with a molecular weight of 50 kDa, measure just one-third the size of antibodies and are thus sterically less demanding.[203] Although this approach can prove helpful for structural investigations, the molecule is still relatively large.[204] Both Fab chains are linked by interchain disulfide bridges. As with conventional antibodies, this is a problem when used in reducing environments such as the cytosol, since antibodies and Fab are not stable in the cytosolic redox environment. Single-domain antibody fragments, which have a molecular weight of only 12-15 kDa can provide a remedy.[205,206] Besides

single domain antibody fragments which are elaborately engineered from conventional antibodies[207-209], antibody fragments known as nanobodies can be obtained from heavy-chain only antibodies (Figure 12). These are found in cartilaginous fishes (e.g. nurse sharks) and camelids and possess only heavy chains.[205,210,211] Therefore, the paratope region is composed of only a single V_H, which can be isolated recombinantly – termed V_HH when camelid derived or V_{NAR} when cartilaginous fish derived.[200] These nanobodies possess excellent chemical and thermal stability and can thus be used as intrabodies, i.e. within the reducing environment of the cytosol.[212,213] They are suitable candidates for the manipulation of cellular processes by binding or obstruction of binding sites as well as for life cell imaging.[188,189,214]

In the present work, intrabodies were *in situ* recombinantly expressed in human cells as a model system for modulating the green fluorescent protein (GFP). In this process, the properties of the intrabodies were photoconditionally enhanced by genetic code expansion.[188,189]
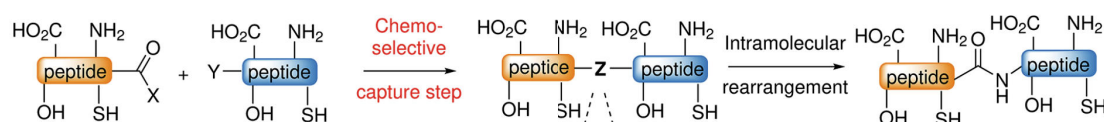
3.9 Semi-synthetic proteins produced by native chemical ligation

While the incorporation of modified amino acids via genetic code expansion using amber suppression elegantly integrates the incorporation pathway and even the delivery of the modified protein to the target site into its method, different approaches must be employed for the *in vitro* incorporation of unnatural amino acids. It is now possible to incorporate β–amino acids by means of amber suppression[215] - but backbone cleaving β-amino acids are not yet an option and thus remain to be incorporated synthetically. Fully synthetic approaches using SPPS are only reasonable peptides up to a length of approx. 50 amino acids[216], since the yields for longer peptides decrease rapidly. Even though small proteins of up to 164 amino acids are potentially available, the yield, processing and exclusion of unwanted by-products is laborious and error-prone.[217] Besides different ways of enzymatically linking polypeptide fragments, e.g. via sortase[218], split inteins[219-221], expressed protein ligation[222], trypsiligase[223], peptiligase[224], or combined approaches[225], the native chemical ligation (NCL) is the most widely studied approach for bond formation between biomolecules and is discussed below.[226,227]

The amide bond is ubiquitously present in life and is efficiently catalyzed in the endogenous synthesis of proteins. Chemical ligation of polypeptides therefore primarily focuses on the principle of chemoselective amide bond formation (Figure 13A). In all these NCL strategies, peptide ligation is based on the formation of a chemically selective bond between two
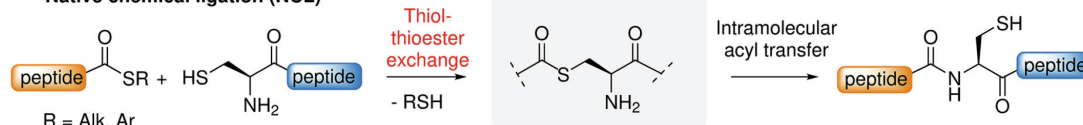
peptides (capture step), which can be converted to a peptide bond by an intramolecular rearrangement (rearrangement step). The details of both the "capture" and "rearrangement" steps can vary depending on the type of ligation method. Hereafter, the native chemical ligation via the side chain mediated by a thiol-thioester exchange is discussed (Type I, Figure 13B). Other methods involve either the α-nitrogen (Type II, Figure 13C) or a combination of the α-amino group and the side chain (Type III, Figure 13D).

Figure 13: *The different types of chemical ligation by amide bond formation.* **A)** The general procedure for chemoselective amine bond formation during a ligation reaction. Ligations are generally based on a capture step, in which a bond is initially formed between the two fragments, and a subsequent intramolecular rearrangement, in which the native structure of the peptide bond is restored. **B)** The most common ligation reaction is the native chemical ligation based on a thiol-thioester exchange. Ligation requires a C-terminal thioester at the N-terminal ligation fragment and an N-terminal cysteine at the C-terminal ligation fragment. **C)** Type II ligations proceed via an activated α-nitrogen atom and include Staudinger ligation and ketoacid-hydroxylamine ligation. **D)** Type III ligations are combined approaches that include both the α-amino group and a side chain in their reaction, forming an oxazolidine. The serine/threonine ligation is a representative of this reaction type. Figure adapted from Agouridas *et al.*, Chem Rev, 2019.[226]

With over 90% of all chemical protein syntheses[226], Type I ligation is the most commonly used, due to its amenable implementation, robustness, and its long-standing establishment since 1994.[228]

In order to conduct the thiol-thioester exchange of the native chemical ligation, on the one hand a thioester is required at the N-terminal fusion fragment and on the other hand a thiol i.e., a cysteine at the C-terminal fusion fragment is required. N-terminal cysteines can be introduced into synthetically by SPPS. However, for long C-terminal fragments, as used in the present work[68], the cysteine must be generated recombinantly. Translation of all proteins starts with a methionine as the initiation codon - and thus also as the first amino acid, which prevents direct production of recombinant proteins with N-terminal cysteines.

The endogenous methionine aminopeptidase can be used to remove the N-terminal methionine, but the efficiency of this method varies significantly with the amino acid sequence and the expression system, which limits its general applicability.[229] A reliable method is to use a protease with a "scar-free" cutting sequence, i.e. one that does not retain any other amino acids upstream of the cysteine target site. The small ubiquitin-related modifier (SUMO) Ubl-specific protease 1 (Ulp1) from *Saccharomyces cerevisiae* recognizes a short amino acid sequence (SUMO-tag) that can be recombinantly inserted in front of protein sequences and cuts specifically at the C terminus of this sequence according to a G-G-|-X motif.[230-232] For NCL approaches, a G-G-|-C motif can be designed which can be released after recombinant expression and purification by Ulp1 (Figure 14A). N-terminal cysteines are about 30-100 times more acidic than internal cysteines, thus at neutral pH of a native chemical ligation they more readily ionize and provide the driving force as a reactive group for bond formation to the thioester.[226] NCL is carried out under strongly denaturing conditions, both to maintain the reactivity of the cysteine, whose acidity is severely dependent on the direct chemical environment, and to exclude interfering secondary structures. Since these cysteines are highly reactive, they are also prone to oxidation (with a highest oxidation rate at pH 7.0[226]) and must therefore be kept in a reduced state by agents like glutathione, 1,4-dithiothreitol (DTT) or tris(2-carboxyethyl)-phosphine (TCEP).[233]

There are diverse options for the generation of thioesters with various derivatives and catalysts for NCL.[216,226] Thioesters exhibit moderate reactivity to nucleophilic attacks while being stable to hydrolysis. This mediocrity makes them ideal for NCL, since side reactions are limited, reaction conditions are mild, and reaction times are still within acceptable limits. While it is possible to generate thioesters for NCL directly within the expression system[222,234], chemical methods are easier to apply to full synthetic fragments generated in SPPS as performed here.[68] However, thioesters generally have the problem

of not being stable towards the harsh Fmoc-SPPS coupling conditions. Therefore, they can either be introduced retroactively or protected appropriately during synthesis. Among a variety of different protection methods, some are particularly suitable for NCL of highly hydrophobic peptides such as those encountered in this work. O to S acyl shift systems for thioester generation increase the solubility of these proteins due to their hydrophilicity and are based on an *S*-peptidyl-2-mercaptoethanol structure. This moiety withstands the Fmoc-SPPS conditions since the reactive thioester is generated after synthesis by an O to S acyl shift at neutral pH (Figure 14B). The Tietze lab has recently published a model system for native chemical ligation of highly hydrophobic proteins, which includes the described methods for thioester generation and protection.[235] In this approach, a 2-hydroxy-3-mercaptopropionic acid (Hmp) is prepared as a thioester-forming moiety using SPPS for the N-terminal peptide. After the O to S acyl shift, the thioester is initially activated by transesterification with the aryl thiol catalyst 4-mercaptophenylacetic acid (MPAA), which is thereafter attacked by the nucleophilic cysteine (Figure 14B). In order to suppress carboxyester formation as a thioester side reaction and to keep poorly soluble peptides in solution, this method also relies on larger amounts of hexafluoroisopropanol (HFIP) in the ligation mixture. After the capture step involving the side chain of the cysteine, the native peptide backbone is restored in an intramolecular rearrangement by an S to N acyl shift. An application of this model system to hydrophobic proteins for the control of biological processes had yet to be demonstrated and has been applied to hydrophobic viral inhibitors in parts of this work.[68]
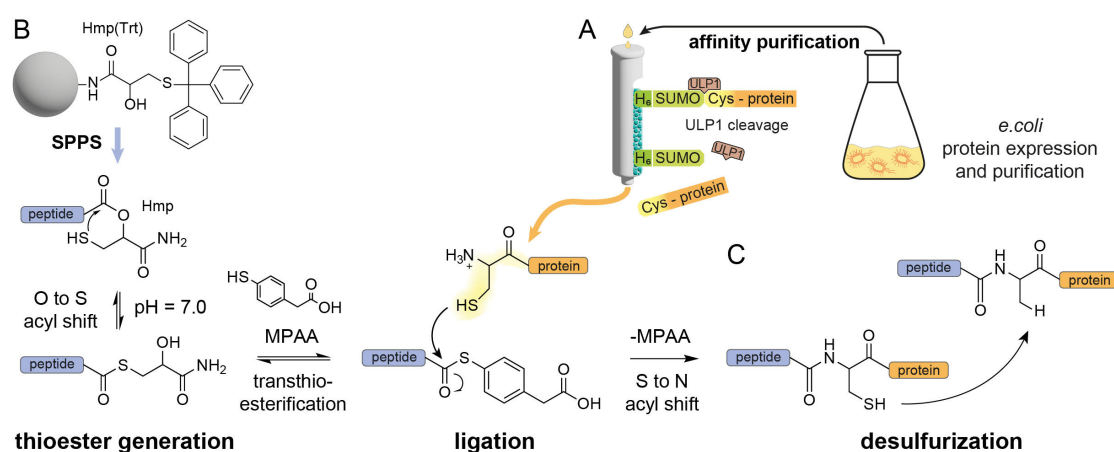
Figure 14: *Native chemical ligation and desulfurization using an O to S acyl shift system*. **A)** Generation of a C-terminal protein fragment containing an N-terminal cysteine for NCL. The protein containing an N-terminal SUMO-tag is purified by affinity chromatography and the tag is removed by ULP1 cleavage. **B)** Specific NCL approach as utilized in this thesis. A thioester surrogate is introduced into a synthetic peptide by SPPS. The reactive thioester is formed after a pH-induced O to S acyl shift. The thioester is activated by the catalyst MPAA for ligation, whereupon the molecule is ready for ligation with the cysteine-containing moiety. The native peptide bond conformation is consequently restored by an S to N acyl shift. **C)** For traceless ligation a radical based desulfurization is performed. Parts of the Figure were adapted from Winter, Domnick *et al.*, Angew Chem, 2022.[68]

As illustrated in Figures 13 and 14, type I ligation reactions are not "traceless". A cysteine invariably remains at the site of the ligation. Ideally, the ligated biomolecule can be designed in such a way that a cysteine is natively present at the ligation site. However, this is rarely possible. Artificial cysteines introduced into the sequence can lead to unexpected effects on binding sites or to undesired conformational changes due to the formation of inter- or intramolecular disulfide bonds. Therefore, it is advisable to remove superfluous cysteines after NCL. Ideally an abundant native alanine can be chosen as a ligation site since it is possible to convert a cysteine into an alanine by various protocols. Many published desulfurization procedures rely on a metal-based reaction via Raney nickel[233,236] or Pd/Al$_2$O$_3$[237]. However, these methods are quite harsh and may cause modification of biomolecules, such as the epimerization of secondary alcohols.[236] By now, there are also mild, metal-free reaction procedures described that rely on radical desulfurization by trialkylphosphines.[238] This approach mostly uses the water-soluble radical initiator VA-044 in combination with TCEP.[235,239]

In the present work, this metal-free desulfurization procedure was applied (Figure 14C). Commonly, the NCL reaction was worked up before desulfurization, which not only involved additional effort but also an inevitable loss of material. Here, the NCL and desulfurization approach was unified to such an extent that further processing is no longer necessary and the approach can be employed almost as a "one pot" reaction.[68]

3.10 Novel single organelle-based flow cytometry assays

Flow cytometry was developed as a highly quantitative method to record various physical parameters of individual cells of a cell suspension in high throughput to gain insight into the population and subpopulation characteristics of cells. Flow cytometers are essentially constructed similarly to a microscope. Nevertheless, the design is optimized for high throughput, so that modern flow cytometers can process more than 5,000 particles per second in real time. The physical parameters are derived based on the scattering properties of the cell and, if applicable, on fluorescence by fluorescent molecules introduced to the cells.

The flow system is based on the hydrodynamic transport of a cell suspension, which is focused by a laminar sheath stream (usually PBS) and individual particles are guided across a light source (Figure 15A). Lasers serve as light source, which is focused onto the sample. According to the physical properties of the analyte, the light is scattered by the sample at the point of optical measurement. The 488 nm laser line usually provides the basic information: Here, incident light is detected in the direction of the laser beam

(forward scatter, FSC). The signal of the FSC is proportional to the cell surface area and thus provides information on the cell size. In contrast, scattered light is detected at a 90-degree angle to the axis of the laser beam (side scatter, SSC) (Figure 15B). This parameter provides information about the granularity or internal complexity of the cell, which is determined by the number of cell organelles or membranes. In addition, state-of-the-art flow cytometers can also determine fluorescence intensities of different dyes, which are similarly monitored by detectors at a 90-degree angle to the direction of light. The number of parameters that can be measured simultaneously depends on the optical configuration, and presently dozens can be determined simultaneously.[240-242]

Flow cytometers are highly flexible instruments that can be used for a wide range of applications.
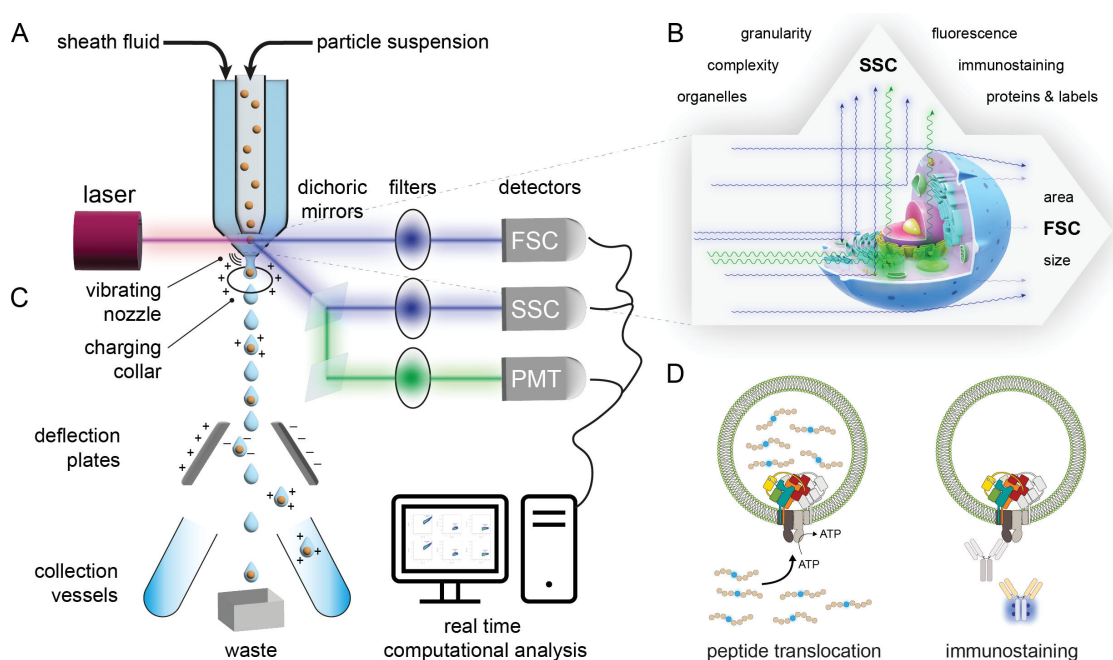
Figure 15: *Overview of a flow cytometry setup with sorting module (FACS).* **A)** Fluidics and detection system of a flow cytometer. A particle suspension is individualized and hydrodynamically focused by a sheath stream. Laser light is scattered at the point of optical detection and fluorophores are excited. Based on the scattered or emitted light, information about the physical properties of the measured particle can be obtained. **B)** Principle of light scattering and fluorophore excitation of measured particles. Incident light is scattered by components of the cell. Light scattering measured at the angle of incident light provides information about particle size (FSC), while scattered light measured at 90-degree angle provides information about granularity (SSC). Emitted fluorescence of excited fluorophores is also detected here. **C)** Sorting unit for fluorescence activated cell sorting (FACS). After detection of the particles, the liquid is passed through a vibrating nozzle, generating a stream of droplets. The droplets can be specifically charged by a charging collar, deflected by deflector plates, and collected in vessels. **D)** With the development of new analysis pipelines in flow cytometry, it is now possible to characterize cellular organelles such as microsomes or artificial vesicles such as liposomes. For this purpose, transport assays with fluorescent reporter peptides or immunostainings can be used.

One of the most common applications of flow cytometers is clinical diagnosis to characterize blood samples, sometimes simply to determine the cell count of individual subpopulations.[243] Other applications include the analysis of fluorescently tagged recombinantly expressed proteins[188,189] and cell surface staining by antibodies[8]. Analyses

of the cell cycle[244], cell proliferation, or apoptosis[245] are also among standard applications. However, modern flow cytometers are increasingly used to analyze particle suspensions that are not limited to cells. While cells have a diameter of about 10 µm, much smaller particles can be analyzed, such as bacteria[240,246-249], viruses[250], sperm[251], plankton[252], synaptosomes[253] or even artificial nanoparticles like liposomes[254]. For these small particles in the sub-micrometer range, it is mostly mandatory to perform additional fluorescence staining, e.g. with membrane-intercalating dyes, as these nanoparticles cannot otherwise efficiently be distinguished from the particulate background in the buffer based on FSC and SSC.[254-256]

Flow cytometers can be enhanced by a sorting module to perform fluorescence-activated cell sorting (FACS, Figure 15C).[257] In this case, the laminar particle flow is not only guided through the flow cell but is additionally passed through a vibrating nozzle. A uniform droplet stream with >10000 droplets/sec is generated. The drops contain either one, none, or multiple particles. By placing a charging collar directly below the nozzle, charges can be transferred to individual droplets as desired.[242] Whether a charge is transferred to a drop depends on the gating. Gating is the technique used to analyze flow cytometry data: Usually either a single parameter is plotted one-dimensionally as a histogram or two-dimensionally as a dot plot. By successively confining regions of plots with differing parameters, subpopulations of particles can be selected or excluded based on their physical properties. In addition, the information whether a generated droplet contains a single particle, none, or multiple particles can be obtained. If a particle in a droplet meets the criteria adjusted during gating, the charging collar deposits a charge on the droplet. The droplet can subsequently be deflected by an electrostatic deflection system and sorted into a container. Using single cell sorting, for example, monoclonal cell lines with desired properties can be generated very efficiently.[258]

Both in the context of antigen presentation as part of the ERQC and in relation to membrane proteins, flow cytometry is of great interest. While the effects of viral inhibitors and alterations on ERQC have been widely studied in various publications via the expression levels of pMHC I on the cell surface[8,66,259-261], this merely represents the endpoint of the whole process without being able to resolve intermediate steps. However, it is not possible to transfer such a complex system as the ERQC into artificial liposome environments. This can be addressed by microsomes, vesiculated fragments of the ER that form during cell disruption.[262] These organelles comprise the ER-resident membrane proteins, including the ER-luminal content, while preserving the native ER membrane topology.[68] In this PhD thesis, a novel method for the analysis of microsomes by flow cytometry was developed, which enables the investigation of antigen processing at the

single organelle level without the need for membrane staining.[68] Membrane transporters such as TAP can be investigated by transporting fluorescent reporter peptides, as can individual protein components provided specific antibodies are available (Figure 15D). In regards to the PLC, the uniformity of the microsomes has disadvantages: The hitherto weakly characterized assembly as well as dissociation of the complex are not readily accessed with antibodies as mechanistical key domains of MHC I, tapasin, and calreticulin are located inside the lumen of microsomes. Since the so far unpublished reconstitution of the PLC into liposomes has been achieved in this thesis, flow cytometry of nanoparticles can be presumed to provide further mechanistic insights into this multi-chaperone-client interaction network of the PLC and thus into the tightly coordinated events of ERQC in the near future.

It has not yet been possible to sort and enrich nanoparticles such as liposomes or microsomes by FACS for further experimentation. On the one hand, based on the reanalysis of sorted particles, it seems to be possible to isolate particles with desired properties. On the other hand, the particles are highly diluted after sorting and either the enrichment methods involve loss of material, or the throughput is too low to obtain substantial amounts of nanoparticles for further biochemical characterization. If in the future it is possible to effectively sort nanoparticles in the same way as cells, this could allow organelles such as microsomes to be specifically enriched for key properties such as PLC content. In combination with high-end imaging methods like cryo-ET, potentially field moving questions such as ER compartmentalization and clustering of protein complexes can be resolved.

4. Results and discussion

This chapter is intended to provide the reader with a summary of the cumulative publications reprinted in the next chapter. The aim is to highlight how the publications interlinked with each other. There are three general cornerstones that these scientific studies address: Firstly, light is used as an orthogonal trigger to control cellular processes from the exterior. Thereby, model systems and nanotools for spatiotemporally resolved manipulation of cellular systems were developed. Secondly, structural analyses of the whole PLC as well as of individual components have been performed and provide novel, high-resolution insights into the molecular interplay of the multivalent chaperone interaction system. The last cornerstone is based on the biochemical analysis of the ERQC system with MHC I as the key actor. Here, mechanistic aspects of the two closely interlaced levels of the MHC I quality control cycle were investigated to unravel the complex quality assurance mechanisms on the way to an optimally loaded MHC I on the cell surface.

4.1 Light as an orthogonal trigger to modulate cellular processes

4.1.1 Light-guided intrabodies for on-demand *in situ* target recognition in human cells

Nanobodies (Nb), as described in the introduction, are versatile binding fragments that have excellent intracellular stability owing to their single-chain design and can thus also be used as intrabodies. They can bind their target molecule with up to sub-nanomolar affinity. However, this binding typically cannot be controlled. In cells that express a nanobody recombinantly, the production of the nanobody can at most be controlled by a promoter - but due to the high affinity, the nanobody reaches its designated target site immediately after synthesis. This can be confounding for many assays since there is no defined time point zero and within several hours of ribosomal nanobody production, the effects desired to be resolved may have already passed. In the present publication, this challenge was addressed by combining recombinantly expressed intrabodies with genetic code expansion by amber suppression: Photoconditional intrabodies were generated in mammalian cells. These intrabodies carried sterically demanding, photocaged derivatives of their endogenous amino acid at a key position for target molecule engagement. This system was designed to serve as a proof of principle to provide a blueprint with universal applicability for photoconditional intrabodies in human cells that can be guided to their target motif by light.

For this purpose, an α-GFP nanobody well described in literature was used.[263,264] This nanobody was recombinantly fused with an mCherry (fusion construct Nb[mCherry]) as a monitoring tool for expression and for subsequent intracellular localization. As a result, the nanobody was no longer exclusively suited for the control of intracellular processes by target molecule binding, but also as a so-called chromobody for live-cell imaging by confocal laser-scanning microscopy (CLSM). Consequently, using crystal structures of the interaction partners Nb and GFP, the binding interface was visualized and screened for suitable positions for amino acid substitution (Figure 16A).

Here, the analysis revealed Tyr37 of the nanobody to be a promising position for the incorporation of a photocage. On the one hand it was located within the epitope binding region of the nanobody and on the other hand it established a direct contact to GFP[Arg168] within the binding interface. Tyrosine is one of the amino acids that provided a choice of various unnatural amino acids with corresponding tRNAs/tRNA synthetase pairings. Therefore, in cooperation with the Deiters Lab, the sterically demanding cages ONBY, as well as NPY were produced. ONBY and NPY should inevitably create a molecular clash within the binding pocket and thus render Nb binding impossible (Figure 16B).[177,178]
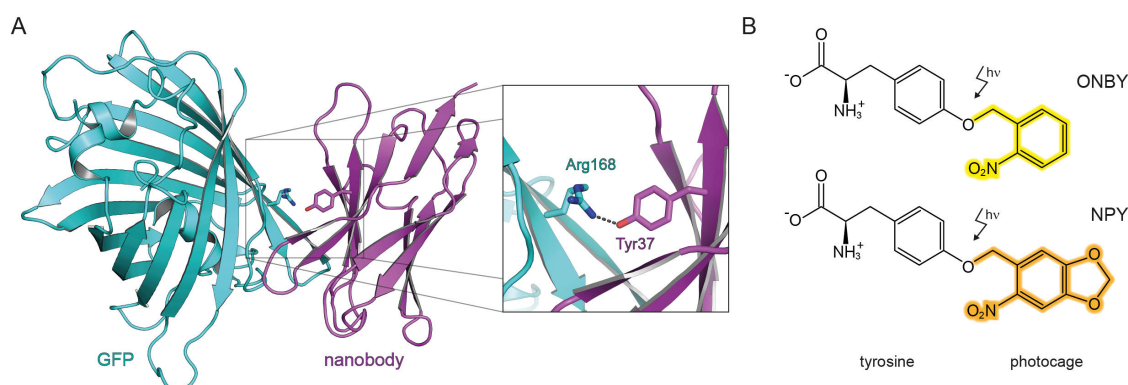
Figure 16: *Design concept of photoconditional nanobodies for on-demand tracing of cellular targets.* **A)** *X-ray structure of the GFP-nanobody complex. Tyr37 of Nb is located at a key position of the nanobody's paratope and directly interacts with Arg168 at the GFP binding interface.* **B)** *Unnatural tyrosine derivatives that can replace a natural tyrosine via genetic code expansion. Upon exposure to UV light, the endogenous tyrosine structure is restored. The moieties that can be removed by UV illumination are highlighted. Figure was adapted from Joest* et al., *Chem Sci, 2021.*[188]

The photocaged tyrosine derivative ONBY is an ONB-based photocage, like the other photocages in the publications reported in this thesis. Conversely, NPY with its nitropiperonyl cage is an evolved cage with improved photochemical properties. This covers an increased excitation wavelength and tuned extinction coefficient as well as quantum yield of the photoreaction.[178] However, it is sterically more demanding than its ONBY counterpart.

Finally, the target molecules had to be defined for the completed model system: EGFP with its Nb binding interface was recombinantly fused to the nuclear envelope protein LaminA

(^{EGFP}LaminA) on the one hand and to the histone H2B (H2B^{EGFP}) on the other hand. The basic concept of the model system is the light-guided recruitment of the chromobodies to the GFP binding interface of LaminA or H2B, upon its previous inactive presence in the cytosol (Figure 17A). Since a whole series of modified proteins had to be reproducibly delivered into the cell now, it was refrained from transfecting for each experiment either plasmids for Nb^{mCherry}, further plasmids for the NPYRS/PylT (NPY optimized tRNA synthetase/tRNA) pair and plasmids for ^{EGFP}LaminA or H2B^{EGFP}. Therefore, at least for the Nb^{mCherry} construct, stable cell lines were generated, each stably expressing photocaged Nb (^{Amb}Nb^{mCherry}) or, for control purposes, the otherwise unmodified Nb (^{WT}Nb^{mCherry}) in a tetracycline-inducible manner.
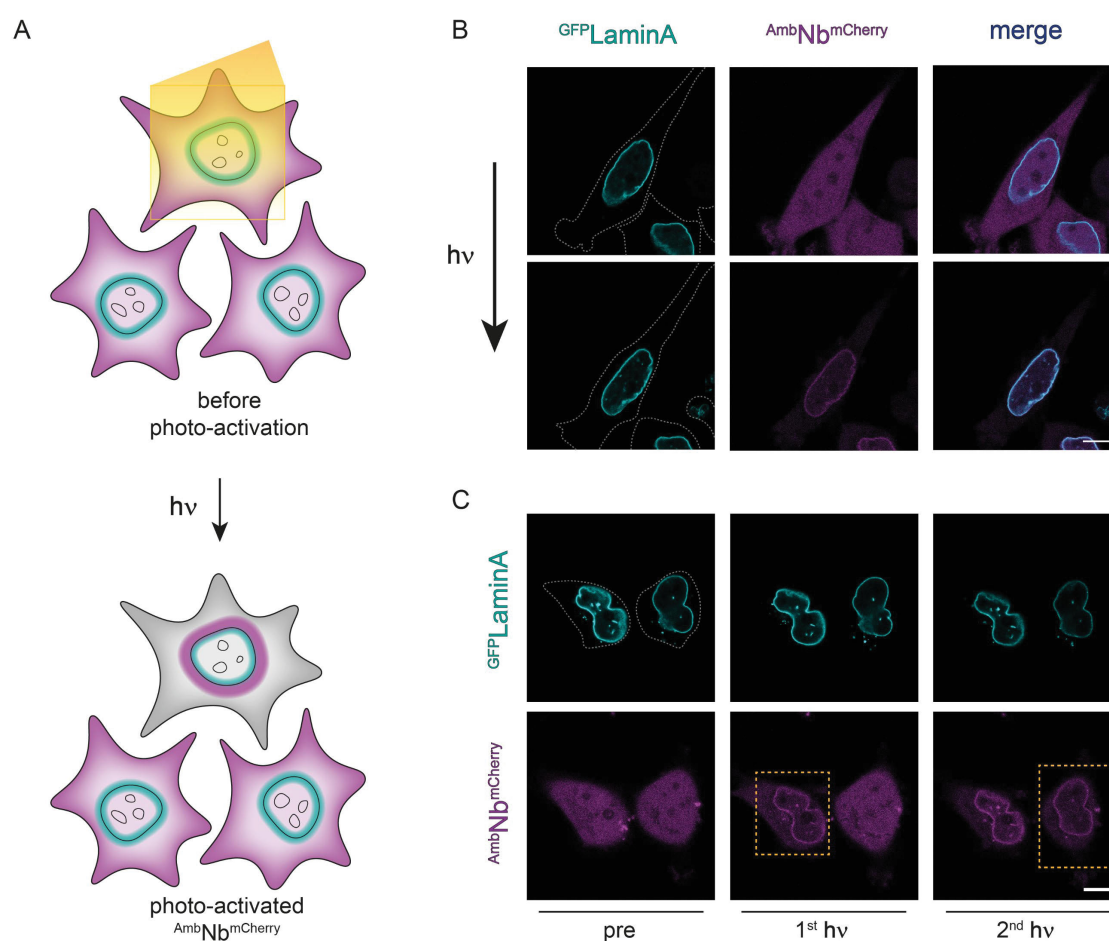
Figure 17: *Photoactivation of light-guided nanobodies*. **A)** The basic concept of nanobody photoactivation. ^{Amb}Nb^{mCherry} (purple), which is present inactively in the cytosol, is instantaneously recruited to its binding site on LaminA (teal) by a pulse of light. **B)** Light activation of ^{Amb}Nb^{mCherry}, which was inactive in the cytosol. After UV illumination, recruitment to the nuclear envelope protein LaminA was evident by colocalization of GFP and mCherry channels. Scale bar = 10 µm. **C)** Micrometer-scale single cell photoactivation of ^{Amb}Nb^{mCherry}. Adjacent cells were not affected by light activation of the target cell. Scale bar = 10 µm. Figure was adapted from Joest *et al.*, Chem Sci, 2021.[188]

Thereupon, the first step was to verify inducible expression of the two Nb constructs in the generated monoclonal cell lines. The two unnatural amino acids offered for incorporation were successfully inserted in amber suppression conditions, whereas expression was

prematurely aborted in the absence of NPYRS/PylT or ONBY/NPY and no mCherry fluorescence could be detected. For analytical purposes, a combination of CLSM and FACS was employed in this study: on the one hand, CLSM provided convincing, high-resolution images of single cells with distinctly visualized effects, but is always subject to the bias of individual, selected cells. FACS, on the other hand, does not visualize intracellular structures, but offers a statistical power unmatched by CLSM, and prevents conclusions from being drawn solely based on individual cells, rendering the methods ideally complementary. The functional system at that point could be extended by transient transfection to include the actual target molecules [EGFP]LaminA and H2B[EGFP]. The unmodified nanobody, after induction, colocalized with the target protein as expected. However, this was not the case for the photoactivatable nanobody under amber suppression conditions: an increased mCherry level was detected in the cytosol, but no colocalization with the target molecule was observed. Using UV-triggered uncaging, specific binding to the target pattern was induced for both [EGFP]LaminA and H2B[EGFP], and almost the entire cytosolic fluorescence (i.e., the nanobody) was recruited to the target position within a few minutes (Figure 17B). Moreover, micrometer-scale photoactivation demonstrated the high spatial resolution of the system, allowing it to trigger targeted nanobody interactions at the single-cell level without affecting adjacent cells (Figure 17C).

This model system was applied in mammalian cells to demonstrate that photoconditional intrabodies generated by genetic code expansion are a powerful tool to effectively suppress premature target binding and to guide molecules to their target position upon request. This blueprint has general applicability and would be transferable to any target protein that can be expressed with a fusion protein such as GFP. Besides GFP, a combination with much smaller protein tags would also be conceivable if specific nanobodies are available. Signal proteins in sophisticated protein networks such as transphosphorylation by receptor tyrosine kinases could potentially be modulated. In addition to targeting and tracking for microscopic visualization by chromobodies, the intrabodies could be fused with functional tags: For instance, with proteasomal degradation signals for the directed degradation of target molecules. In conclusion, this study has established an elegant platform using photoconditional intrabodies as a tool for the future investigation of human cellular proteins, networks, and signaling pathways.

4.1.2 Efficient amber suppression via ribosomal skipping for *in situ* synthesis of photoconditional nanobodies

In publications such as the one summarized above, applied genetic code expansion via amber suppression frequently appears to be rather straight forward: once the

corresponding genetic modifications have been implemented, all that remains to be done is to provide the modified tRNA synthetase (aaRS)/tRNA pair and the unnatural amino acid. Finally, the system will work. Practically, the process is usually more complicated: on the one hand, genetic code expansion can lead to unpredictable side effects, especially in mammalian cells. On the other hand, the efficiency of amber suppression is largely dependent on various factors: The target protein, its mRNA sequence, the position of the artificial amber codon within the sequence, the ratio of aaRS/tRNA to mRNA level, and the intrinsic properties of the unnatural amino acid. Some of these factors can be manipulated in the design of the specific amber suppression system, while others cannot: for example, the position of the amino acid to be transformed may be predetermined by the structure. There is no universal blueprint that can be followed to optimize a non-functioning amber suppression system and ultimately only an educated guess can be made to eliminate possible interferences.

In the work described here, an extensive screening of multiple factors affecting amber suppression was performed to tune a suboptimal system. Ultimately, a new approach for designing amber suppression systems was established, which can be used to remedy difficult amber suppression targets. The system consisted of the identical Nb^{mCherry} - GFP interaction pair as described in the previous section. However, a stable cell line was not established since transient transfection offers the advantage of high flexibility towards other cell lines and novel targets. Additionally, the expression level of tetracycline-induced stable cell lines is on an endogenous level, which impedes overexpression for biochemical purification of the engineered protein.

The Nb^{mCherry} system previously described in stable target lines was cloned into a vector for transient transfection under control of a CMV promoter and transfected into human cells (photoactivatable ^{PA}Nb^{mCherry}, Figure 18A, construct I). Instead of the expected absence of ^{PA}Nb^{mCherry} expression without co-transfection of NPYRS/PylT and provision of unnatural amino acid, flow cytometric experiments revealed fluorescence levels reaching up to 40% of the unmodified construct (^{UM}NB^{mCherry}, Figure 18B). Thus, a substantial portion of translations of the transfected construct I (Figure 18A, construct I), despite the absence of tRNAs, did not result in the termination of translation but in the incorporation of another amino acid or the skipping of the codon. Under complete amber suppression conditions, no significant increase in fluorescence was detected. The concept of targeting by light activation is undermined by high read-through which could not be accepted for further analysis.

Therefore, in a comprehensive approach, different constructs were consecutively engineered to screen for factors that may have an adverse impact on effective amber

suppression. Firstly, additional copies of the modified tRNA were inserted into the construct (Figure 18A, construct II), to prevent the codon from being skipped without incorporation of the unnatural amino acid by increasing the amount of available tRNA. Subsequently, the second start codon within construct III (mCherry[Met1]), was silenced by an M1A mutation. This aimed to prevent the ribosome from potentially attaching and translating mCherry individually.
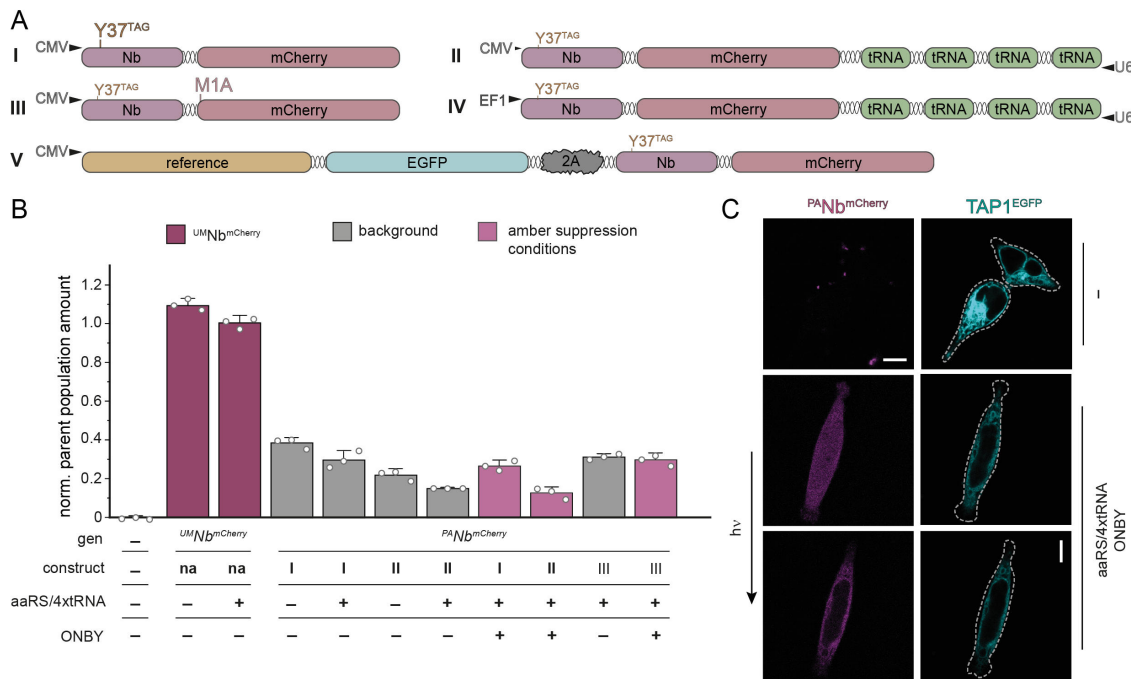

Figure 18*: Comprehensive screening and optimization of various factors affecting amber suppression.* **A)** Expression cassettes for [PA]Nb[mCherry] production to screen the effects of multiple factors affecting amber suppression. (I) represents the previously used nanobody under control of a CMV promoter. In (II), 4 additional copies of the modified tRNA were added to the construct. In (III), the second initiation methionine was removed by the M1A mutation. In (IV), the concept of (II) was shifted towards control of a strong EF-1α promoter (EF1). Construct (V) represents the successful p2A-Amber construct that prevents a read-through by using a 2A site with an upstream target protein. 2A, ribosomal skipping site. **B)** mCherry fluorescence was quantified by flow cytometry for the constructs in A under CMV control at different conditions. No significant difference between amber suppression conditions and background expression was detected. **C)** The [PA]Nb[mCherry] construct (V) with TAP1[EGFP] as the target protein transfected into human cells. The inactive nanobodies present in the cytosol were effectively recruited to their target site in the ER membrane after light activation. Figure was adapted from Joest *et al.*, ACS Synth Biol, 2022.[189]

However, subsequent in-gel fluorescence analyses revealed that the full-length construct was transcribed in each case. In construct IV, the ratio of mRNA to tRNAs was harmonized by placing the construct under the control of the strong EF-1α promoter (EF1). None of the constructs led to a considerable improvement of background expression or to a significantly increased level of intact nanobody generation in full amber suppression conditions (Figure 18B). In contrast, the strong EF1 promoter of construct IV even led to a drastic aggravation of background expression.

Since ribosome profiling in other publications recently implied that ribosome reliability is lower at the onset of translation[265,266], an indication was derived that the modification at

position 37 might be so early within the sequence that the codon is readily skipped by the ribosome. Consequently, the target protein of the nanobody was equipped with an F2A ribosomal skipping site and placed upstream of the $^{PA}Nb^{mCherry}$ construct (Figure 18A - construct V, p2A-Amber). A 2A site with the recognition sequence "DxExNPGP", derived from the foot-and-mouth disease virus, forces the ribosome to skip by perturbing the bond formation and thereby initiating the start of a new polypeptide chain within the same translation.[267] As a result, the amber site was much further down the construct and the ribosome should no longer be as error-prone at this site.

With this approach, it was finally possible to establish effective amber suppression conditions. The non-specific background caused by read-through was greatly diminished to a few percent in flow cytometric analysis. In full amber suppression conditions, effective amber suppression could be induced in up to 60% of the cells, which is a comparatively high value. The new construct allowed to co-express various targets such as TAP1^{EGFP} and H2B^{EGFP} simultaneously with the nanobody in a consistent ratio. Using CLSM, a homogeneous distribution of the inactive nanobody in the cytosol and a successful binding to the target proteins after light activation could be visualized (Figure 18C). In general, it would also be possible to directly fuse other proteins or protein tags N-terminally to a nanobody to achieve an identical effect. However, nanobodies are functionally susceptible to N-terminal modifications, rendering a 2A site the better choice.

Ultimately, systematic screening identified a deadlock in genetic code expansion and inefficient protein synthesis was surmounted by the new translation strategy p2A-Amber. In particular, when aiming for higher protein amounts, established methods can become inefficient. The p2A-Amber approach allows efficient amber suppression results even at high translational load. Finally, genetic code expansion by amber suppression is a system that cannot follow a fixed scheme in assay design and must be extensively optimized depending on the application. The field needs more approaches like these to develop effective and easy-to-implement protocols for universal applicability of amber suppression.

A comparison of these two publications raises a question: why did genetic code expansion require extensive optimization for transient transfection, while the same construct performed properly in stable cell lines? The reason could be the difference in the expression level of the amber suppressed protein: The stable cell line produces the nanobody at endogenous levels and thus in much lower abundance than in a system with transient transfection under a CMV promoter. A further indication towards this direction is given by the strong EF1 promoter used in construct IV, which induced even higher read-

through. Therefore, the effect seems to correlate with higher mRNA levels of the amber suppressed protein and thus with a high translational load.

4.1.3 Semisynthetic viral inhibitor for light control of the MHC I peptide loading complex

In addition to the photoactivatable nanobodies developed here, which represent an elegant platform for the *in vivo* modulation of cellular targets, a versatile viral inhibitor was evolved into a photocleavable nanotool to control the peptide loading complex by light.

Current strategies for biochemical characterization and structural analysis of the PLC rely on a pulldown by the viral TAP inhibitor ICP47 to isolate the complex. This strategy is hypothesized to be successful as ICP47 arrests the PLC in an outward-facing conformation, stabilizing it for purification and reconstitution. However, this approach has the inherent disadvantage that TAP, as the central component of the PLC, remains irreversibly functionally stalled and hence the conformation of TAP can no longer be altered. Current studies in structural biology have not yet been able to resolve the transmembrane and cytosolic domains of intact, ICP47-inhibited PLC.[11,18] Other conformations could potentially render this region more accessible to structural analyses by confining the intrinsic flexibility in this region. In terms of biochemical characterization, the function of TAP due to irreversible inhibition has never been studied in the context of the native PLC. For reconstitution into advanced membrane mimetics such as proteoliposomes that provide compartmentalization of inside and outside, it would be of little use employing intact but transport-incompetent PLC. Therefore, there is an urgent demand for a nanotool that can both stabilize the sensitive PLC during purification, but simultaneously does not interfere with the transport capabilities of TAP as well as the complex multi-chaperone-client interaction network of the editing modules.

This challenge was addressed in the publication outlined in this section: Using a semi-synthetic approach, the viral TAP inhibitor ICP47 was transformed into a multi-purpose and light-controlled nanotool that can both purify the PLC via its affinity tag and release it in a photo-responsive manner. In various assays and controls, the proper function of this approach was documented and functional characterization of active TAP in the context of the native PLC was achieved.

The helix-loop-helix motif in ICP47 was previously screened for positions to incorporate the unnatural and photocleavable amino acid Anp, with position 15 shown to be fully TAP binding-competent (Figure 19A). Here, this concept was extended to the full-length ICP47 which was further equipped with a streptavidin binding peptide (SBP) affinity tag. The

functional concept of pc-ICP47^{SBP} involves specific binding to the peptide binding pocket of TAP, with the SBP affinity tag protruding sufficiently to allow interactions, for example, with a streptavidin column material (Figure 19B). UV light in the 365 nm range can induce the *ortho*-nitrobenzene of Anp to cleave the peptide backbone in pc-ICP47^{SBP}, leading to photofragmentation and thus to irreversible ICP47 inactivation with concomitant loss of affinity. Here, the hypothesis was raised that the two resulting fragments would diffuse out of the binding pocket owing to their loss of affinity, thus restoring TAP transport-competence.
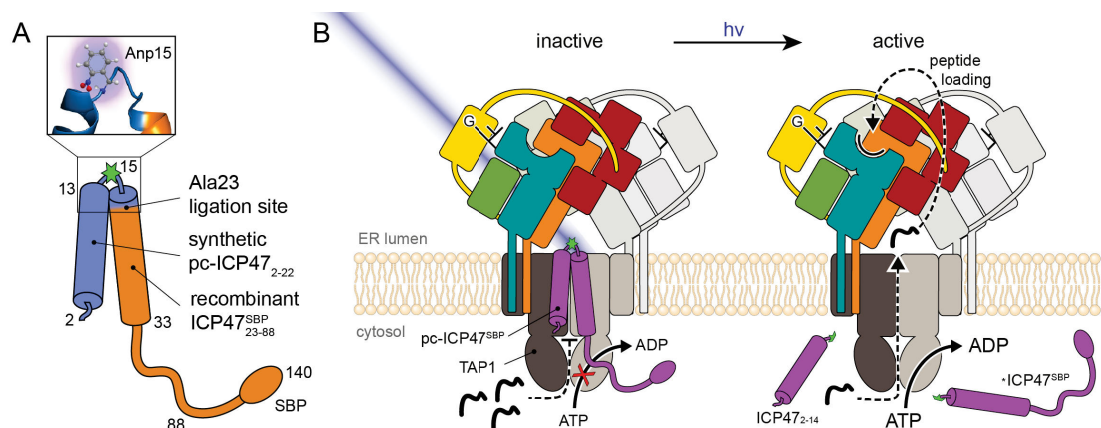

Figure 19: *Rational design and functional concept of pc-ICP47^{SBP}*. **A)** Amino acids 2-33 constitute a helix-loop-helix motif and shape the binding interface for the peptide-binding pocket in TAP. The 88 amino acid full-length ICP47 was enhanced with an SBP affinity tag. The semi-synthetic design was color-coded: The nanotool is composed of a synthetic part (blue) and a recombinant part (orange). At position 15 of the synthetic fragment the photocleavable amino acid Anp was incorporated. Both fragments are ligated at position 23, at which an alanine is present in the wildtype sequence. **B)** The functional concept of PLC activation by pc-ICP47^{SBP}. The peptide loading complex resides in the ER membrane and is inhibited by pc-ICP47^{SBP}. Upon light activation, pc-ICP47^{SBP} can be cleaved into two fragments, activating TAP, and allowing the PLC to regain its full function. Figure was adapted from Winter, Domnick *et al.*, Angew Chem, 2022.[68]

Since a construct comprising 140 amino acids could no longer be reasonably produced fully synthetically, a semi-synthetic approach was employed. In the absence of native cysteines in the wildtype sequence of ICP47, the closest possible alanine downstream of position 15, carrying the unnatural amino acid, was selected as the ligation site. Here an S to O acyl shift system was applied for native chemical ligation, since especially the helix-loop-helix motif and thus the synthetic N-terminal fragment of ICP47 is hydrophobic and has to be kept soluble during ligation.

Since the S to O acyl shift system is based on a capture step between thioester and cysteine, both moieties had to be installed at the N- and C-terminal fragments: N-terminally, a thioester surrogate was synthetically instituted as described in the introduction. To generate the C-terminal cysteine fragment, the alanine at position 23 was mutated into a cysteine and transformed into a vector with N-terminal SUMO fusion tag. After ULP1 cleavage, the C-terminal fragment with N-terminal cysteine was obtained (Cys-ICP47$_{24-88}$^{SBP}). The NCL reaction was completed within 5 h by MPAA catalysis at high

concentrations of guanidine hydrochloride (GdmCl) and hexafluoroisopropanol (HFIP) supplemented as chaotropic agents and to increase solubility. The reaction progress was monitored by LC-MS.

After reaching completion, the reaction solution was not purified as usual via HPLC, which inevitably involves loss of material, but merely stripped of excess MPAA catalyst by means of a spin SEC column usually applied for the desalting of protein samples. After NCL, the reaction had to be turned traceless by desulfurization, since it was not known whether a cysteine at this site might have an adverse effect on the binding affinity to TAP. A potential cleavage fragment might also be retained in the vicinity of TAP by disulfide bridging, which could impair the photorelease. Published protocols for radical desulfurization via VA-044 and TCEP were initially found to be ineffective, as the radical environment triggered radical-induced Anp fragmentation during the required reaction time, thus preventing an efficient conversion. However, this problem was overcome by careful temperature-dependent optimization of the reaction rates and adjustment of the buffer system. In exact compliance with the optimized reaction conditions, about 85% of intact pc-ICP47^{SBP} were retained after desulfurization.

The final semi-synthetic nanotool was then evaluated for proper photo-conditional operation. Pulldown experiments demonstrated comparable yields to established assays with non-photocleavable ICP47^{SBP}. The light-induced cleavage of pc-ICP47^{SBP} affected neither the hydrodynamic elution volume of the PLC in SEC experiments, nor the stoichiometric composition of the PLC subunits. In accordance with other *ortho*-nitrobenzene derivatives, the cleavage kinetics of pc-ICP47^{SBP} both in solution and bound to the PLC, proceeded monoexponentially and revealed almost complete ensemble photoconversion within 5 min. SDS-PAGE and immunoblotting of SEC peaks further demonstrated that the photofragments of pc-ICP47^{SBP} no longer coeluted with the PLC. It was therefore concluded that the fragments had been effectively removed from the binding pocket upon photocleavage. Since ICP47 release was accelerated in the presence of ATP in the buffer, as well as with peptides competing for the binding site, it was hypothesized that one or both fragments would be removed from the binding pocket by the inherent transport function of TAP.

It was further demonstrated that photoreleased PLC was competent in binding for fresh ICP47^{AF647}, as well as for native peptide substrate. After reactivation of the peptide-binding capacity, peptide binding kinetics in fluorescence anisotropy studies were comparable to previously published data in which TAP was measured individually.[73] Due to inhibition, the NBDs on TAP were ATP binding-competent, but any hydrolysis activity was disabled.[62,64]

Reactivated PLC, however, indicated hydrolysis activity that was dependent not solely on the light-induced removal of pc-ICP47^{SBP}, but also on the presence of peptide substrate.
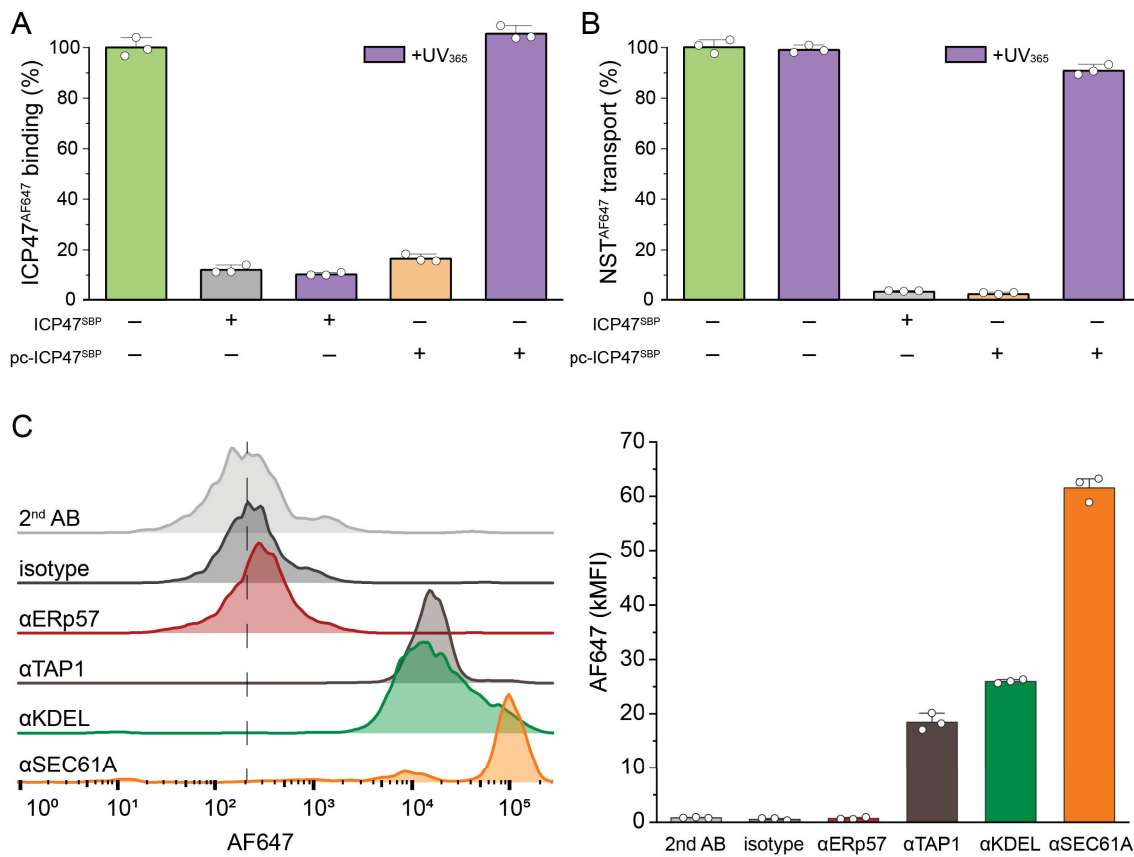

Figure 20: *pc-ICP47^{SBP} is functional in the native environment of the ER in microsomes, as demonstrated by a novel, ubiquitously applicable flow cytometric approach with single organelle resolution.* **A)** The re-binding of ICP47^{AF647} was examined in PLC-containing microsomes by flow cytometry. Non-inhibited microsomes (green), could bind fluorescent ICP47^{AF647} unhindered, whereas ICP47^{SBP} (grey) or pc-ICP47^{SBP} (orange) inhibited microsomes did not show substantial binding behavior. UV exposure restored the initial binding characteristics (violet). **B)** Transport of fluorescent reporter peptides in microsomes studied by flow cytometry. PLC in untreated microsomes (green) was unaffected by UV exposure (violet). ICP47^{SBP} (grey) or pc-ICP47^{SBP} (orange) effectively suppresses transport activity. Light-activated, the native transport rate was restored (violet). **C)** Flow cytometric traces and associated bar diagram of immunostainings against representative ER motifs. While luminal components like ERp57 cannot be visualized by antibodies, indigenous cytosolic motifs can be readily visualized by flow cytometry of microsomes. Figure was adapted from Winter, Domnick *et al.*, Angew Chem, 2022.[68]

Due to the excellent applicability in purified PLC, we further evaluated the suitability in native ER membranes. A novel flow cytometric analysis pipeline was established, which allowed the detection of transport rates and protein levels in microsomes by fluorescence. Using a fine-tuned gating strategy, it was furthermore possible to visualize these particles with an average diameter of only 150 nm without additional membrane staining. First, as already done in purified PLC, we tested whether photoreleased PLC is binding competent for fluorescent ICP47^{AF647}. Despite the consumption of the minimal sample amounts, the flow cytometric measurements demonstrated that pc-ICP47^{SBP} is quantitatively displaced from the binding pocket and can bind fresh ICP47^{AF647} (Figure 20A).

Finally, the transport competence of PLC was determined after photomodulation in the native ER membrane. Owing to the transport-induced accumulation and the associated amplification of fluorescent signal, transport studies in microsomes are ideal analytes for this type of flow cytometric approach. Again, photoactivated PLC in microsomes was shown to have similar transport capabilities as uninhibited PLC, whereas pc-ICP47[SBP]-inhibited PLC did not (Figure 20B).

Ultimately, it could be demonstrated that this new flow cytometric analysis pipeline was not only suitable for the analysis of membrane transporters but was rather of general applicability for ER membrane proteins for which cytosolically binding antibodies are available (Figure 20C). The procedure was comparable to immunostaining with cells: Proteins could be visualized both by an antibody sandwich approach as well as by directly labeled primary antibodies. Flow cytometric detection of subcellular organelles can therefore provide analytical access to membrane-bound enzymes, channels, and transporters. The analysis benefits from general advantages of flow cytometry analysis including high statistical applicability due to large numbers of detected particles.

In conclusion, the photoconditional semisynthetic viral inhibitor can be used for purification of solubilized PLC as well as for highly efficient synchronous activation in native membrane environments. Thus, this approach provides an urgently desired nanotool, which is a prerequisite for numerous further analyses: pc-ICP47[SBP] may open new avenues in structural biology of the PLC by exploring new native or forced (e.g., via vanadate trapping) conformations using cryo-EM. In a context of a fully active PLC, functional analyses now permit to investigate the close interplay between the editing module and the transporter. Eventually, this light-guided approach may allow high-resolution structural analyses and visualization of PLCs clustering in proteoliposomes by cryo-ET measurements, which would otherwise potentially be suppressed by ICP47.[103,104]

4.1.4 Further applications of light-modulated molecules

In two further publications discussed here, light-triggered approaches were used to fitfully modulate target molecules. In "Structure of an MHC I-tapasin-ERp57 editing complex defines chaperone promiscuity" outlined in the next chapter, a light-degradable peptide was likewise used to revert a molecule to its parent condition: For crystallization intents, an MHC I-tapasin-ERp57 editing complex was to be assembled from individually produced proteins. Since peptide deficient MHC I is intrinsically unstable, during recombinant expression it accumulates in inclusion bodies and must be refolded in the presence of an optimal peptide. This approach therefore employs a high-affinity MHC I H2-D^b binder (RGPGRAFJ*TI), which at position 8 (J*) carries the photocleavable Anp. With the assistance

of this high-affinity binder, it was possible to stabilize MHC I for refolding and further handling. Conversely, as long as a high-affinity binder is bound, MHC I will not tightly interact with tapasin, since it already is optimally loaded. Therefore, once again the photocleavable approach was applied, in which the high affinity peptide is cleaved into two low affinity peptide fragments. After dissociation of the peptide fragments, MHC I requires immediate chaperoning by tapasin. To verify this approach, the complexes were analyzed for their peptide content by SEC-MS before and after UV exposure (Figure 21A). It was shown that about 96% of the refolded MHC I molecules released their peptide (Figure 21B). In consequence, this was an efficient approach to generate peptide deficient MHC I molecules in the presence of the peptide editor tapasin-ERp57 for crystallization and subsequent structure determination.
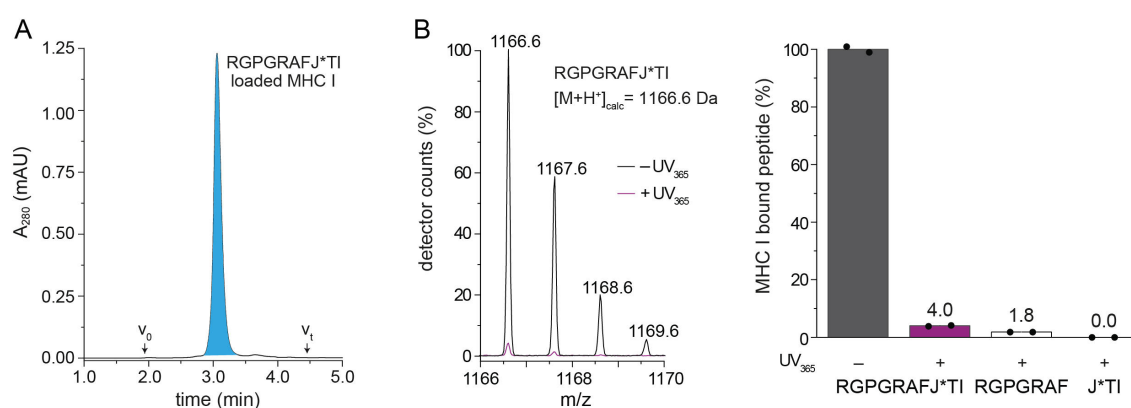
Figure 21: *Release of light-degradable MHC I high affinity binders analyzed by SEC-MS.* **A)** SEC profile of the MHC I hc-β_2m-RGPGRAFJ*TI complex. The chromatograms were identical before and after UV illumination. Combined spectra of the region highlighted in blue were utilized for MS analysis in B. The intact pMHC I complex decomposed into its individual components during ionization. **B)** Isotope resolved MS raw spectrum of the single charged high affinity MHC I binder. After UV exposure, the amount of peptide present in the sample is reduced to a minimum (purple). The bar diagram illustrates the ratios between intact peptide and residual peptide fragments between unexposed and exposed samples. The C- and N-terminal fragments are effectively released from the MHC I peptide binding groove. Figure adapted from Müller *et al.*, Nat Commun, 2022.[8]

In a similar assay, the UGGT1-mediated activity towards the MHC I glucosylation status was investigated in "TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I". Complementary to the MHC I-RGPGRAFJ*TI complex facing tapasin, MHC I was dissociated from TAPBPR with the high affinity binder ETVSKQSJ*V to generate vacant pMHC I molecules (Figure 22A). Since MHC I was already engaged with a high affinity binder, UGGT1 had no rationale to intervene (Figure 22B). In a UV-triggered approach converting the high-affinity binder into a fragmented non-binder, MHC I was destabilized ("frustrated") and the so that a reglucosylation mediated by UGGT1 could be investigated (Figure 22C). Likewise, the effective release of the light-modulated peptide was detected by SEC-MS. Using this approach, the performance of UGGT1-mediated glucosylation in the presence and absence of TAPBBR was examined by LC-MS glycan analysis. Results of this assay are presented in chapter 4.3.2.

Overall, the majority of the publications discussed here utilized photoconditional biomolecules combined with light-modulation as an orthogonal trigger. However, the modified nanotools themselves were mostly not in focus of the analysis but were applied as a tool to modulate the proteins or complexes of interest for biochemical investigation.
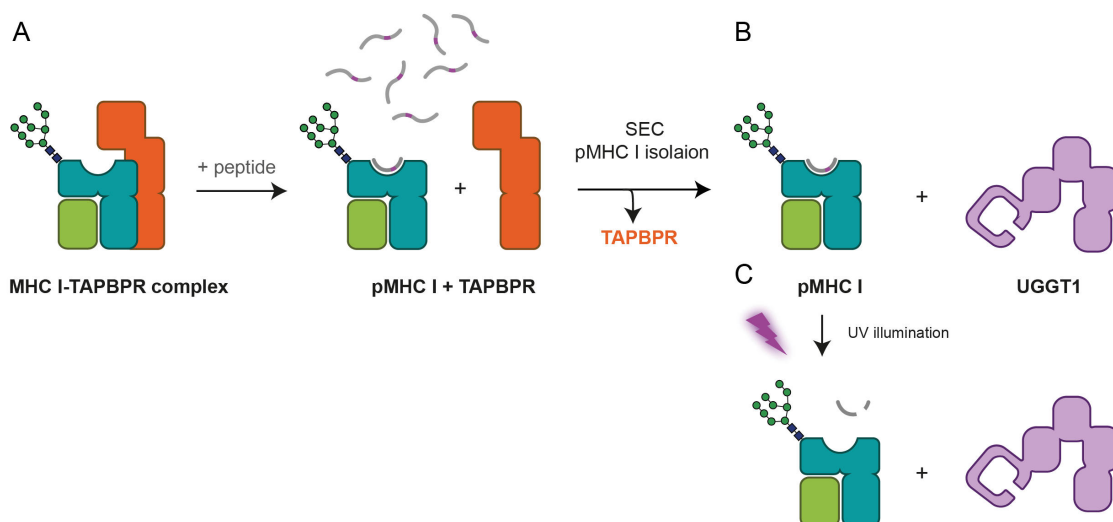
Figure 22: *Schematic representation of the UV-triggered assay for the analysis of the individual glucosylation activity of UGGT1.* **A)** The MHC I-TAPBPR complex is dissociated by addition of the high-affinity MHC I-binding peptide ETVSKQSJ*V, permitting removal of TAPBPR by SEC. **B)** The MHC I glycan is subsequently probed for the presence of the terminal glucose by LC-MS after UGGT1 processing. **C)** A destabilized ("frustrated") MHC I, devoid of peptide and TAPBPR was produced by UV-mediated peptide degradation. Here as well, the MHC I glycan is subsequently probed for the presence of the terminal glucose by LC-MS after UGGT1 processing. Figure adapted from Sagert *et al*. (manuscript in preparation).[37]

In summary, this chapter reveals that light is an extremely versatile and elegant approach to manipulate biochemical processes. In this context, the use of light in the UV range had no obvious detrimental effects. In the future, the field of light-driven processes in applied biochemistry will hopefully be influenced to a greater extent by considerably superior photocages such as coumarins, rather than ONB-based cages.

4.2   Structural insights to the mechanisms of MHC I peptide loading

4.2.1   Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity

As previously described, tapasin is the main peptide editor for MHC I clients in the context of the PLC. However, it is still elusive how exactly the chaperoning process mediated by tapasin-ERp57 proceeds within the editing module and which molecular rearrangements occur within MHC I-tapasin-ERp57 upon complex formation. In this study, a crystal structure of the MHC I-tapasin-ERp57 complex in a peptide-receptive state was generated at atomistic resolution (2.7 Å, Figure 23A). This high-resolution structure in combination with systematic substitution of key residues has revealed novel insights into the interaction and selection mechanisms of tapasin on chaperoned MHC I molecules.

To assemble the complex for crystallography, the cytosolic and transmembrane regions of tapasin, ERp57 and MHC I were removed and consequently only the ectodomains were produced. Applying the photo-triggered approach described above (4.1.4), complex formation between MHC I and tapasin-ERp57 was induced and immediately subjected to microseed matrix screening for crystallization. The atomistic structure model of the MHC I-tapasin-ERp57 complex generated from the diffraction data indicated the characteristic conformations of the individual components: ERp57 with a twisted U-shape and tapasin with an L-shaped structure embracing the MHC I peptide binding pocket (Figure 23A). On the one hand, the so-called editing loop of tapasin is positioned atop the F-pocket of the peptide binding pocket of MHC I, on the other hand, the β-hairpin touches the bottom of the peptide binding pocket from underneath.

Figure 23B depicts the structure of unloaded MHC I (teal) in direct superposition with peptide-loaded MHC I (grey). It is evident that the interaction with tapasin remodels the α1 and α2 helices of the MHC I binding pocket, which particularly widens the F-pocket and thus increases the accessibility for peptide clients. This expansion of the binding pocket is further stabilized by an external network of hydrogen bonds with tapasin. The opposite interaction with the β-hairpin realigns the floor of the peptide-binding pocket, which is equally accompanied by a reorientation of the β-hairpin itself in comparison with tapasin in the absence of a client molecule (Figure 23C, grey). The structure model uncovers a variety of crucial domain movements of the outer contact regions in the interface with

MHC I. In the course of the study, multiple residues involved in intermolecular contacts were mutated and their influence on MHC I surface expression was investigated by flow cytometry. When mutated, almost all amino acid residues at key positions of the interaction surface between tapasin and MHC I lead to a significant reduction of MHC I surface expression of about 30-70% and underline the importance of conserved molecular contacts between tapasin and MHC I. However, for a detailed description of all contacts and their effects in the binding interface, the original document should be consulted.
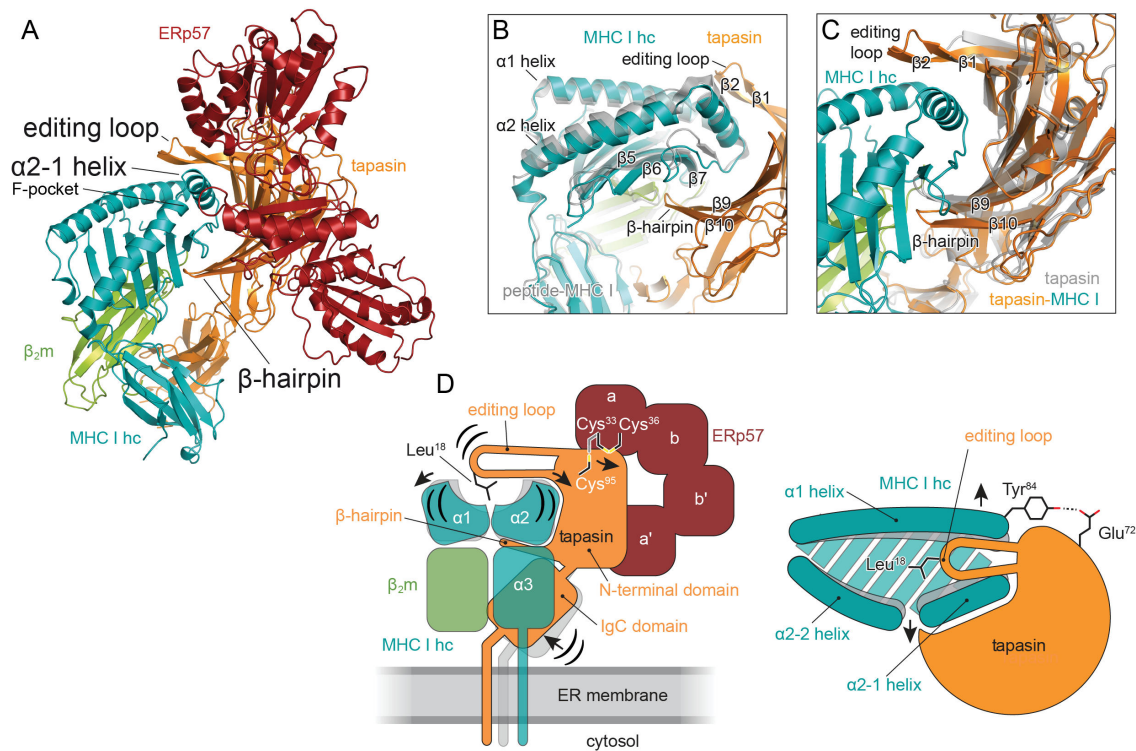
Figure 23: *Structural and mechanistic details of the MHC I-tapasin-ERp57 editing complex.* **A)** Cartoon representation of the MHC I-tapasin-ERp57 complex. The crystal structure assembled by a photo-triggered approach diffracted to 2.7 Å resolution (PDB 7QNG). **B)** Zoom-in to the binding interface between tapasin and MHC I. Tapasin clamps the MHC I binding pocket from above with its editing loop and from beneath with the β-hairpin. Compared to peptide-bound complex (PDB 2F74, grey), chaperoned MHC I features an expanded α1 as well as α2-1 helix, and its underside of the peptide-binding groove reshaped by the β-hairpin. **C)** Identical zoom-in as shown in B, but with superposition of client-free tapasin. The intrinsically unstructured editing loop rigidifies upon client recognition, and the β-hairpin is rearranged. **D)** Schematic model of all major conformational rearrangements through the tapasin-MHC I interaction in side view, as well as top view. Additionally, to the described rearrangements in the context of the MHC I peptide binding pocket, the IgC domain of tapasin is also reoriented for an enlarged interaction surface with MHC I. The editing loop shields the F-pocket with Leu18 acting as a hydrophobic plug. Figure adapted from Müller *et al.*, Nat Commun, 2022.[8]

The high resolution of the crystal structure further revealed with atomistic resolution that the editing loop, which is intrinsically disordered in the client-unbound state, adopts a rigid configuration that stabilizes unloaded MHC I clients by mimicking the hydrophobic C terminus of a potentially bound antigen. Again, targeted amino acid substitutions in conjunction with flow cytometric analyses of MHC I surface expression were able to highlight the conformational contribution of individual amino acids. Thus, evidence was obtained that the editing loop of tapasin mainly acts as a hydrophobic plug within the

binding pocket. In addition to pulling apart the α1 and α2 helices of MHC I and rearranging the bottom of the binding pocket, F-pocket widening via the hydrophobic plug constitutes to the main modes of chaperoning of MHC I clients by tapasin. The crucial molecular rearrangements and motions are schematically depicted in the model in Figure 23D. It is possible that the editing loop shields the F-pocket from C termini of cargo peptides. Yet, it remains unclear how, for example, C-terminally extended peptides may then be accommodated by tapasin-mediated peptide editing. The detailed contribution of ERp57 to the chaperoning process is still unknown. It is possible that, in addition to its function as a structural protein and for docking of calreticulin, it conformationally prepares tapasin for engagement with MHC I. Further studies in this area may shed more light on the complementary modes of action between tapasin and TAPBPR. These studies might clarify why, for example, largely tapasin-independent MHC I allomorphs exhibit the same conserved residues in the interaction interface with tapasin.

4.2.2 Molecular basis of MHC I quality control in the peptide loading complex

As opposed to the previously presented approach, the editing module was not crystallized here, but instead resolved by cryo-EM. Since the structure-giving sample in this case was not artificially truncated, expressed and assembled, it contained all stably associated components of the PLC, membrane components of the proteins, and post-translational modifications. In consequence, the PLC was initially required to be vitrified onto grids as a high-quality sample for cryo-EM data acquisition. Previous studies indicated that when solubilized in detergent, the PLC was not stable enough to yield a suitable cryo-EM sample without artificial crosslinking. Hence, new reconstitution routes for the PLC had to be explored in this work. Among other tested candidates, reconstitution in large (16 nm diameter) MSP2N2 nanodiscs as well as reconstitution via peptidiscs revealed promising preliminary results. Since the initial screenings with nanodisc reconstituted samples produced slightly superior micrographs, we proceeded with nanodisc reconstituted PLC (PLC::MSP2N2). The nanodisc-belt seemed to be crucial for the stabilization of the complex, as it was possible for the first time to image the fully assembled PLC without crosslinking for stabilization. The acquired 2D and 3D classes could be categorized according to different stages of editing module assembly: one editing module was always fully assembled, whereas different components were missing in the adjacent editing module. Since essential and structure-shaping components such as ERp57 and calreticulin were missing in the presence of MHC I, it was assumed that these must have been artifacts of sample preparation, vitrification or contact with the air/water interface. Focused refinement of approximately 53,000 fully assembled particles (initially about

90,000 particles) resulted in a three-dimensional map with an average resolution of 3.7 Å (Figure 24A). Despite various attempts of a focused reconstruction of the membrane part, as well as the cytosolic domains, it was not possible to resolve these sections of the PLC, as the editing modules were too flexible in relation to TAP. Since the entire editing module was addressed here, which naturally included the MHC I-tapasin-ERp57 complex, some structural findings are close to those presented in the last section, albeit not with the same degree of atomistic resolution.
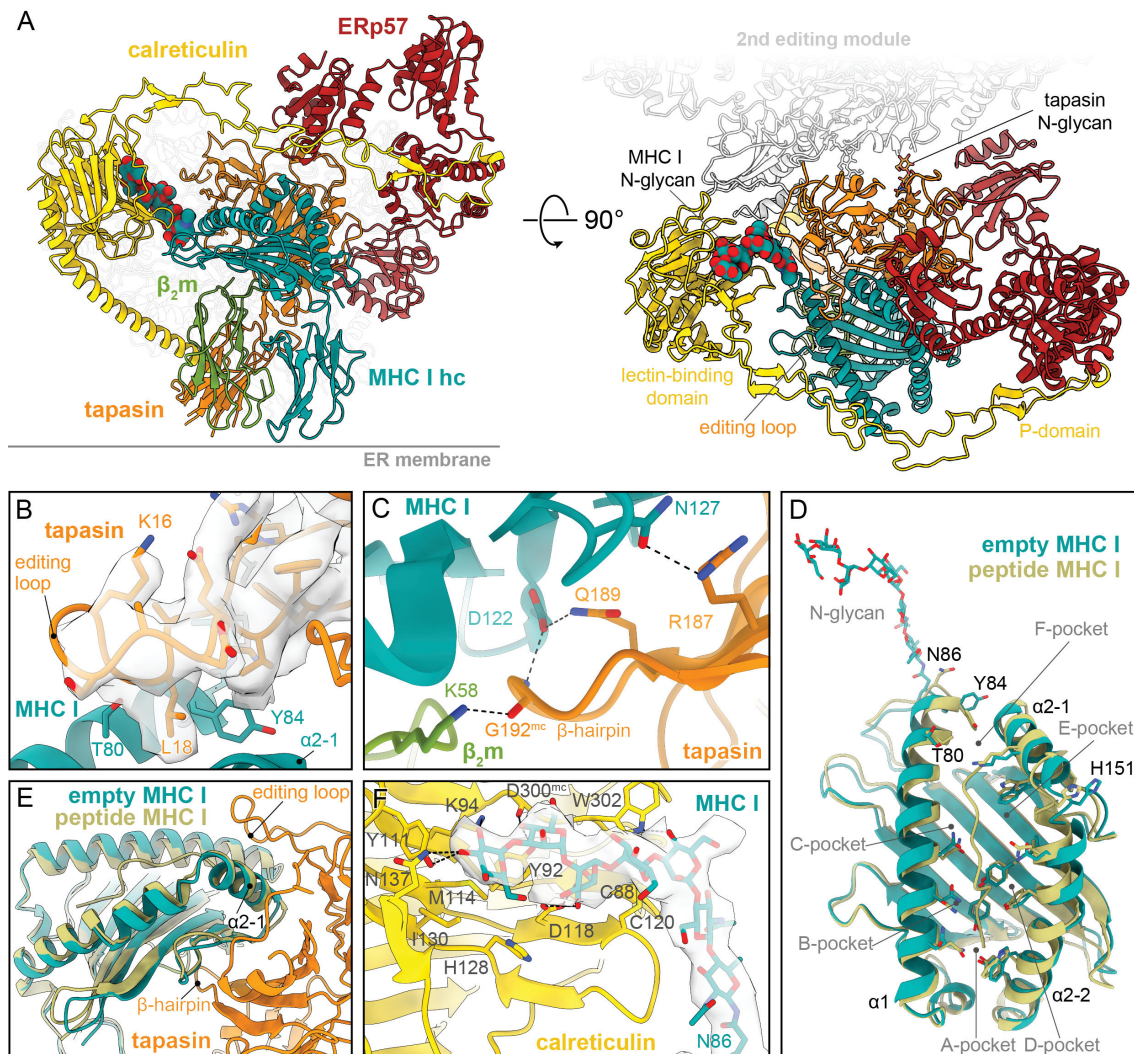
Figure 24: *The structure of the multivalent MHC I chaperone network revealed by the 3.7 Å cryo-EM structure model of the PLC.* **A)** Side and top view of PLC editing modules with MHC I bound in peptide receptive state. For a better overview, one of the editing modules is shaded in grey. The native orientation is implicated by the ER membrane. **B)** The editing loop of tapasin sits atop the F-pocket of the peptide-binding groove of MHC I, with Leu18 plugging into the F-pocket. The cryo-EM map of the editing loop is depicted as isosurface (light grey). **C)** The β-hairpin of tapasin in contact with the bottom of the peptide binding pocket of MHC I. The β-hairpin contacts both MHC I and β₂m. **D)** Superposition of peptide loaded MHC I (HLA-A*03:01, PDB 3RL1, khaki) and empty MHC I (teal), top view. The peptide-clamping conformation of the MHC I binding pocket in contrast to widened structure in peptide receptive MHC I is visible. The N-glycan is depicted as a stick model and is attached to the end of the α1-helix. **E)** Superposition of the structures as in D. Molecular rearrangements are particularly evident in the positioning of the α1, α2-1, α2-2 helices as well as the bottom of the binding pocket. **F)** Zoom-in to the MHC I Asn86 N-glycan as a stick model in contact with the lectin domain of calreticulin. The cryo-EM map is shown as isosurface (light grey). Figure adapted from Domnick, Winter *et al.*, Nat Commun, 2022.[11]

Similarly, the rigid tapasin editing loop is located above the F-pocket of the MHC I binding pocket, with the side chain of Leu18 plugged into the F-pocket (Figure 24B). The β-hairpin contacts and rearranges the bottom of the peptide binding groove, most notably via Arg187, while also contacting β_2m (Figure 24C). By rearranging the MHC I peptide groove bottom, tapasin could feasibly mediate its substrate specificity. Conversely, since the Arg187-interacting MHC I Asn127 is not a conserved among MHC I allomorphs, an influence on the varying tapasin-dependence of MHC I allomorphs could be partially explained by the engagement with the β-hairpin. Direct superposition with peptide-loaded MHC I (Figure 24D, E) revealed that the MHC I binding pocket is significantly widened in the context of the PLC, with a root mean square deviation (RMSD) of 1.8 Å for the α1 helix and 2.0 Å for the hinge region of the α2 helix. Here, the Leu18 of the editing loop of tapasin is positioned in such a manner that it would collide with MHC I Tyr84 in the peptide-loaded state of MHC I (Figure 24D). In addition to Tyr84, further amino acid residues are displaced, which contributes to the extension of the F-pocket along the α1 and α2-1 helices, and thus facilitates peptide exchange.

Most strikingly, in this structural model, it was possible to simultaneously resolve the glycans of tapasin and MHC I. While only the basic glycan stem ($GlcNAc_2Man_1$) was rigid enough to be modelled at high resolution, the complete A branch of the MHC I glycan could be modelled. The function and flexibility of the tapasin glycan will be addressed in the next chapter. The Asn86-linked glycan bridges the end of the α1 helix of the MHC I peptide binding pocket (Figure 24D) with the lectin domain of calreticulin (Figure 24F). The terminal glucose is crucial for calreticulin binding and is contacted by several amino acid residues of calreticulin that shape the glucose binding pocket. However, the mannose residues of the B and C branches of the glycan could not be modulated at high resolution since they remained highly flexible in contrast to the A branch, which is fixed at both ends. It can thus be hypothesized that the MHC I N-glycan establishes a direct, rigid link to the lectin domain of calreticulin and upon peptide binding, is capable of relaying the α1 helix shift directly to calreticulin. Consequently, the MHC I glycan would serve as a loading sensor for direct inter-process communication between MHC I and calreticulin, whereupon calreticulin can release the glycan to serve as a folding marker to be flagged by GluII. This process will be examined in the next chapter.

In summary, two intersecting structural analyses were conducted employing X-ray crystallography and cryo-EM. On the one hand, as expected, the structural findings considerably overlap, which supports the anticipated accuracy of the structure models obtained. On the other hand, the approaches focused on substantially different details: while the PLC as a native sample with full post-translational modification and membrane

domains probably reflects the actual state within the cell more accurately, the MHC I-tapasin-ERp57 structural model allows much higher resolutions that can describe interactions, movements, and consequent mechanisms more precisely. Moreover, amino acids in key positions of the interaction surface were biochemically assessed for their structural relevance. The model of the native PLC has not yet revealed its cytosolic and transmembrane domain, but important insights into the inter-subunit communication by glycan-mediated information transmission have been gained, which will be biochemically characterized in the next chapter concerning mechanistic insights into MHC I quality control.

4.3   Mechanistic insights into MHC I quality control

4.3.1 Molecular basis of MHC I quality control in the peptide loading complex

The new, high-resolution details of the editing module, among other features, have uncovered a key positioning of the A branch of the MHC I glycan as a possible information bridge to calreticulin. Therefore, we biochemically characterized the glycoproteins of the native PLC in more detail. Since the structure of the MHC I glycan only permitted to model the A branch, whereas for tapasin just the basic glycan stem was resolved, we wanted to draw further conclusions about their overall structure and biological function. To this end, we subjected the entire PLC in its MSP2N2 reconstituted form to top-down mass spectrometric characterization.

It was possible to determine and assign all primary sequences as well as the posttranslational modifications of the components in the native PLC as well as ICP47 (Figure 8). PLC generated from Raji cells featured an additional, previously rudimentarily characterized tapasin variant. Bottom-up LC-MS analysis revealed that this variant did not contain a divergent glycan pattern as initially suspected, but instead is splice variant with a modified and elongated C terminus, incorporated into the PLC at a ratio of about one-third compared to isoform 1.

Based on the structural model, the intact mass spectra of the glycoproteins MHC I (Figure 25A) and tapasin (Figure 25B) were of particular interest. MHC I exhibited a significant mass peak of 40,340 Da and minor species at lower molecular weight in the native PLC. Therefore, to determine the MHC I allomorphs detectable in the PLC, PNGase F digestion was performed to completely remove all glycans. This revealed a single peak for MHC I at 38,313 Da, attributed to the allomorph HLA-A*03:01 which was also the predominant allomorph in PLC LC-MS analyses. After assignment of the MHC I sequence, the expected $Glc_1Man_9GlcNAc_2$ glycan tree could also be mapped by mass spectrometry based on the intact mass spectra (Figure 25A). Glycans that lacked one or two mannoses, either at the B/C branch or both, were present in low abundances. This is typically a signal for ubiquitination and subsequent degradation in the ER associated protein degradation pathway (ERAD).

Like MHC I, tapasin also showed a single species in MS after PNGase F digestion (Figure 25B). However, the assigned glycan pattern after successful determination of the primary sequence was exceptionally complex, as tapasin revealed almost all possible glycan structures ranging from an intact $Man_9GlcNAc_2$ down to a single GlcNAc. Such a severe trimming of the glycan tree is unusual, since typically already the removal of the first

mannoses leads to degradation via the ERAD pathway. The tapasin glycan is presumed to be involved in shaping a peptide tunnel of the lateral window between the editing modules.[18] On the one hand, the irregular glycan pattern offers an explanation for the poor resolution in the structural model. On the other hand, it also suggests that an intact tapasin glycan is not essential for PLC function. Possibly, a shorter glycan may even be beneficial since the lateral window might be opened for longer peptides that require N-terminal trimming by ERAD1/2. It seems that tapasin undergoes several rounds of glycan processing while being protected from ERAD-mediated degradation.
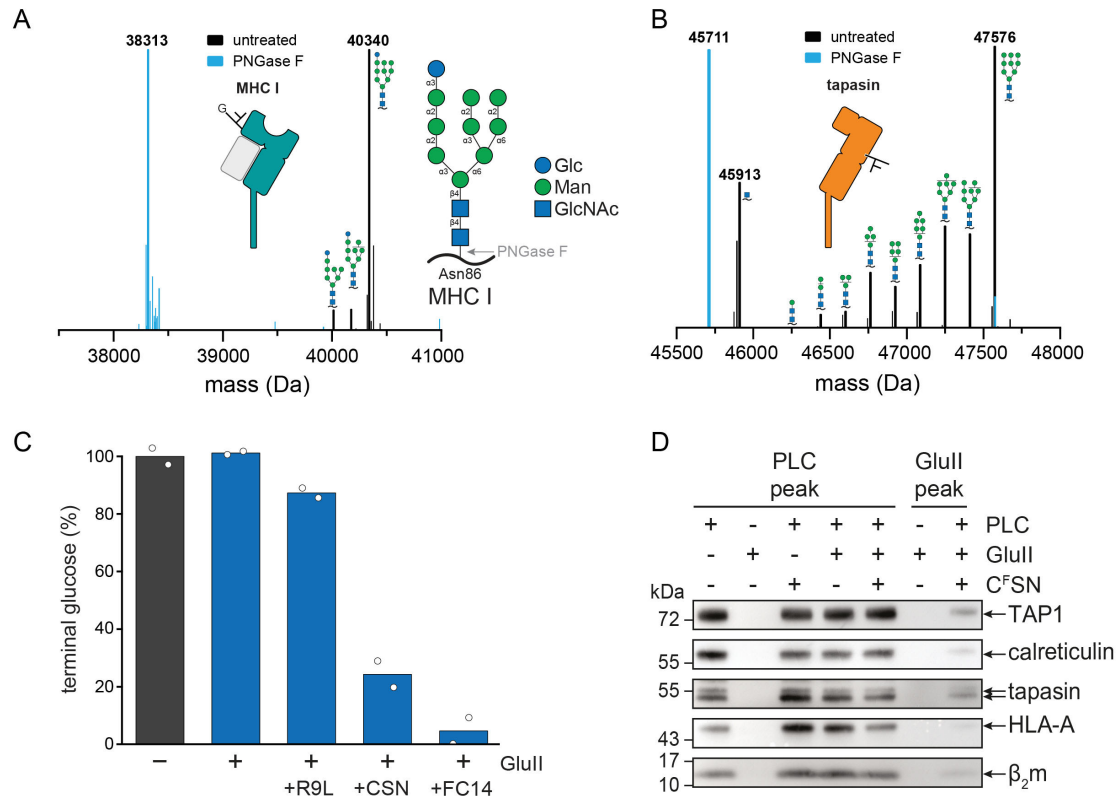
Figure 25: *Mass spectrometric and mechanistic characterization of PLC glycoproteins*. **A)** Top-down LC-MS analysis of MHC I in the context of the native PLC. PNGase F deglycosylated MHC I HLA-A*03:01 (cyan) exhibited a homogeneous peak at 38,313 Da without further posttranslational modifications. Native MHC I almost exclusively carried a $Glc_1Man_9GlcNAc_2$ glycan (40,340 Da), while glycans with one or two missing mannoses at the B or C branch were also present in minor amounts (black). **B)** Identical analysis as in A of tapasin. Deglycosylated tapasin featured a homogeneous peak without posttranslational modifications (cyan). Native tapasin reveals an exceptionally complex glycan pattern with $Man_{9-0}GlcNAc_{2-1}$ (black). **C)** Proportions of terminal glucose of the MHC I glycan at different conditions analyzed by top-down LC-MS. GluII by itself failed to remove terminal glucose compared to untreated native PLC (black). Low affinity HLA-A*03:01 binding peptide (R9L) had minimal effect, whereas the high affinity binder (CSN) was able to induce GluII-mediated deglucosylation. Complete deglucosylation was achieved by fos-choline-14 (FC14) induced complex dissociation. **D)** Immunoblot against PLC components after SEC purification. The native PLC was incubated with GluII and peptides in the same conditions as in C, and the subunit composition was determined. Calreticulin as a soluble component is still part of the PLC after GluII-mediated deglucosylation. Figure adapted from Domnick, Winter *et al.*, Nat Commun, 2022.[11]

Within the PLC, tapasin as the main peptide editor ensures that MHC I molecules are loaded with optimal peptides while catalyzing the exchange of poorly associated peptides. At a previously unknown time point after MHC I peptide loading, GluII removes the terminal glucose of the MHC I-Asn86 glycan, thus flagging MHC I as optimally loaded. In this work,

an *in vitro* glycan processing assay demonstrated that GluII can deglucosylate MHC I molecules immediately after loading while still confined within the PLC.

Figure 25C shows the results of intact MS analyses, which indicate that GluII cannot induce deglucosylation on its own. However, when high-affinity binding peptides were offered, the glycan readily became accessible for GluII allowing to elicit deglucosylation. In general, GluII was also able act on unloaded MHC I molecules, as demonstrated by complete deglucosylation upon Fos-choline-14 induced complex disassembly. Thus, it was concluded that a conformational change must have occurred within the PLC for the terminal glucose, which is intrinsically shielded by calreticulin, to be accessible by GluII. Initial assumptions that calreticulin as a soluble component dissociates after loss of contact with the terminal glucose of the MHC I glycan could not be corroborated. SEC purification of the PLC in the presence of GluII and high-affinity peptide demonstrated by SDS-PAGE and immunoblot (Figure 25D) that the PLC retained an unaltered amount of calreticulin after glycan processing.
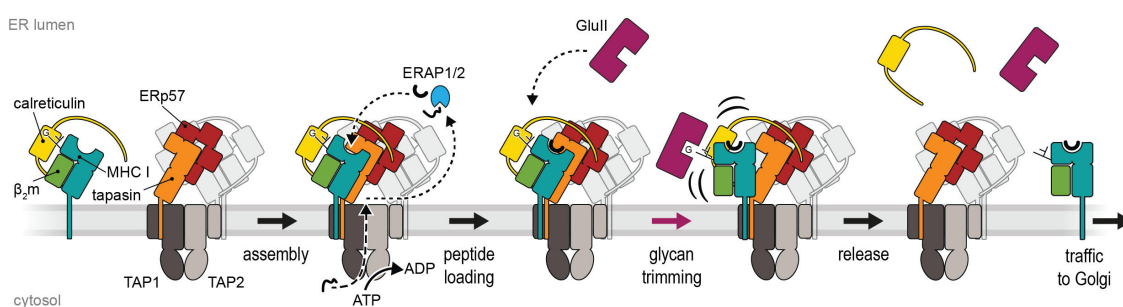
Figure 26: *Model of MHC I assembly, peptide editing, and quality control*. Calreticulin recruits MHC I to the PLC. Peptides are delivered into the ER by TAP and, if necessary, trimmed to the appropriate length by ERAP1/2. After an optimal peptide for MHC I has been found by tapasin catalysis, the glycan becomes accessible for GluII-mediated deglucosylation. This process can occur within the PLC with no need for calreticulin to leave the PLC if complex dissociation is constrained. Subsequently, MHC I can leave the PLC and reach the cell surface via Golgi trafficking. Figure adapted from Domnick, Winter *et al.*, Nat Commun, 2022.[11]

Therefore, an allosteric coupling between glycan processing and peptide proofreading within the PLC was evidenced. Based on the data obtained from the structural model as well as the *in vitro* deglucosylation assays, we proposed an updated model for MHC I peptide loading and successive quality control (Figure 26). A profound change in this model involves GluII acting directly within the PLC. Under normal circumstances, the positions of calreticulin and GluII with regard to their binding sites for the terminal glucose of the MHC I glycan would clash. Therefore, calreticulin is required to gain a high degree of flexibility by releasing the interaction of its lectin domain to the terminal glucose so that it can hinge away upon GluII engagement to such an extent that GluII can bind unimpeded. However, it must be kept in mind that this was concluded from a nanodisc-constrained system and MHC I cannot leave the PLC by diffusion. Therefore, the sequence of events within a system that offers unlimited diffusion might be different.

In conclusion, by combining *in vitro* glycosylation assays with high-resolution structural models of the PLC editing modules, which also revealed post-translational modifications such as glycans, we were able to demonstrate the PLC as a sophisticated model system of the ERQC. In the future, these complex mechanics of this multi-chaperone network employing the PLC as a major interaction hub could be further explored by model systems that allow unlimited diffusion such as proteoliposomes.

4.3.2 TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I

Whereas the PLC quality control mechanisms employed on MHC I represent the first level of ERQC in the context of antigen presentation, the second level of ER quality control mediated by TAPBPR and UGGT1 were investigated in this study.

UGGT1 has previously been characterized as a highly sensitive protein folding sensor that can engage a variety of substrates.[107,268] One of these substrates is MHC I, which is reglucosylated when present in a suboptimal folding or loading state and thereby primed for a new round within the ERQC cycle.[120,269] However, previous studies employed artificial substrates for corresponding experiments, which poorly mimic conditions within the cell.[270,271] In the context of MHC I-based antigen presentation, it is still enigmatic whether UGGT1 can independently probe the charge state and reglucosylate MHC I if necessary. Earlier approaches investigated MHC I molecules lacking a defined glycan structure.[120] Since the MHC I glycan is tightly regulated within the antigen presentation system, a defined $Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ glycan is essential for valid analysis. Recently, evidence has been provided indicating that UGGT1 interacts with the MHC I peptide-proofreader TAPBPR.[35]

Consequently, it was necessary to verify the hypothesis that TAPBPR assists in, or is even compulsory for, the reglucosylation of suboptimally folded MHC I by extending the interactome of UGGT1 to MHC I through TAPBPR-UGGT1 interaction.

First, a native-like substrate for UGGT1 had to be engineered: On the one hand, MHC I hc (HLA-A*68:02) was tethered to β_2m and a Fos-zipper, while on the other hand, TAPBPR was expressed with a Jun-zipper. *In vivo*, both components compose a Jun-Fos complex, producing a linked, peptide-deficient β_2m-MHC I hc-TAPBPR complex. Here, expression proceeded in the presence of kifunensine to generate a defined $Man_9GlcNac_2$ glycan. The artificial Jun-Fos zipper was removed by TEV cleavage after purification, ultimately generating an MHC I-TAPBPR complex with a defined $Man_9GlcNac_2$ glycan serving as an ideal substrate for UGGT1.

The MHC I-TAPBPR complex was subsequently subjected to a novel *in vitro* glucosylation assay in the presence of recombinantly expressed UGGT1 (Figure 27A). In the absence of UDP-glucose (UDP-Glc) substrate for UGGT1, an unaltered $Man_9GlcNAc_2$ MHC I glycan pattern was identified (Figure 27B). In mass spectrometry, the glycan pattern on MHC I was comparable to native MHC I present in the PLC, apart from the desired absence of glucose (Figure 25A). Species with reduced amounts of B/C branch mannoses were also present in this case, albeit in different proportions. In the presence of UDP-Glc, successful glucosylation of MHC I in the MHC I-TAPBPR complex was observed, as apparent by the shift of the $Man_9GlcNac_2$ peak to higher molecular weight by 162 Da ($Glc_1Man_9GlcNac_2$, Figure 27C). It was therefore concluded that UGGT1 recognizes the MHC I-TAPBPR complex with $Man_9GlcNac_2$ as a native substrate and, in the absence of a loaded peptide, flags it as insufficiently folded by transferring a terminal glucose. The effect of conversion of $Man_9GlcNAc_2$ to a $Glc_1Man_9GlcNAc_2$ glycan observed by mass spectrometry was restricted to a maximum achievable conversion of about 80% of the peak intensity. With respect to time, temperature and concentration, this fixed value could not be exceeded.

Since the newly established UGGT1 glucosylation assay operated reproducibly with the MHC I-TAPBPR complex, the initial hypothesis of whether UGGT1 can perform glucosylation independently could now be addressed. To this end, high affinity MHC I binding peptide was used to dissociate TAPBPR, followed by isolation of pMHC I from TAPBPR by SEC. In the glucosylation assay, UGGT1 did not exert any measurable effect on isolated pMHC I (Figure 27D, bar diagram). However, it was anticipated that UGGT1 would not further process previously optimally loaded MHC I molecules. Therefore, a light-triggered approach illustrated in Figure 22 was used to produce frustrated MHC I molecules devoid of peptide and TAPBPR. UGGT1 was equally unable to glucosylate frustrated MHC I molecules in the absence of TAPBPR (Figure 27D). In a rescue experiment, it was demonstrated that once TAPBPR was supplemented to these frustrated MHC I molecules, UGGT1 resumed its glycosyltransferase activity and modified the glycan. However, as expected, optimally loaded MHC I were still disregarded by UGGT1 even in the presence of TAPBPR (Figure 27D).

Consequently, the initial hypothesis was confirmed to the extent that TAPBPR not only assists but is an essential prerequisite for the glucosylation of suboptimally folded MHC I molecules. Based on the findings of this study, two possible modes of action of TAPBPR and UGGT1 emerge: Either TAPBPR binds to MHC I in its role as a peptide proofreader and thereby independently reads out the folding state of MHC I, prompting UGGT1 to be recruited for reglucosylation. Alternatively, it is possible that by binding to UGGT1, TAPBPR

expands the interactome of UGGT1. By its inherent folding sensor function, UGGT1 thereupon recognizes the folding state of its client protein MHC I and acts upon it.
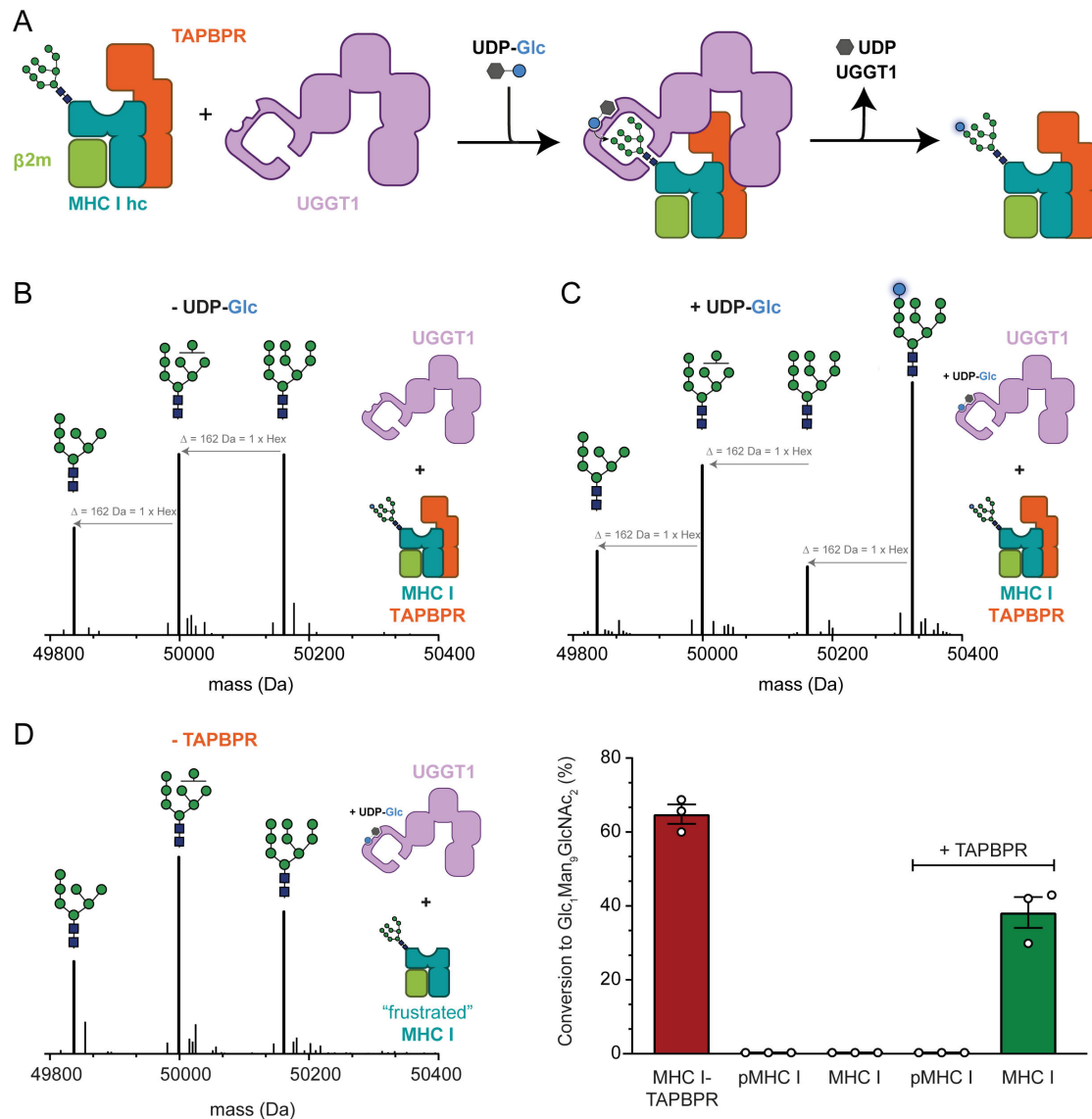
Figure 27: *Analysis of TAPBPR-mediated effects on UGGT1 based glucosylation of peptide-deficient MHC I.* **A)** The recombinantly assembled MHC I-TAPBPR complex bearing a $Man_9GlcNAc_2$ glycan is provided as a substrate in a novel UGGT1 glucosylation assay. UGGT1 transfers the glucose from UDP-Glc to the MHC I glycan in a calcium-dependent manner, thereby tagging MHC I as not fully folded. **B)** The native glycan pattern of MHC I largely features the required $Man_9GlcNAc_2$ pattern, which is not altered by UGGT1 in the absence of UDP-Glc substrate. **C)** In the presence of UGGT1 and UDP-Glc, a glucose is loaded onto the $Man_9GlcNAc_2$ glycan, marking MHC I as not correctly folded. **D)** In a light-triggered approach (Figure 22), frustrated MHC I molecules were generated devoid of peptide and TAPBPR. UGGT1 cannot induce glucosylation without TAPBPR. This unaltered whether MHC I possesses a cargo peptide or not (bar diagram). If TABPBR is added back to these MHC I molecules (+TAPBPR), frustrated MHC I molecules can be rescued and subjected to UGGT1 glucosylation, while optimally loaded MHC I are unaffected. Figure adapted from Sagert *et al.* (manuscript in preparation).[37]

This study substantially expanded our understanding of the second level of the ERQC with TAPBPR emerging as a crucial factor for the MHC I quality control cycle. In the future, there will be an urgent demand for structural analyses of UGGT1 in complex with client proteins to unravel the numerous remaining questions concerning the function and binding interface of UGGT1 with TAPBPR. Research on the MHC I-TAPBPR-UGGT1 interconnection

in the second level of the MHC I quality control pathway is still relatively young, hence presently not even the domains through which an interaction between the binding partners could take place is known. Overall, this work demonstrated that the second level of the MHC I ERQC is critically reliant on TAPBPR as a peptide proofreader that mediates incompletely folded MHC I to the folding sensor and glucosyltransferase UGGT1. In the absence of TAPBPR, misfolded MHC I molecules might not be subjected to repeated quality control in the ER, as UGGT1 alone would not be able to tag them appropriately.

5. References

[1]   Chaplin, D. D. Overview of the immune response. *J Allergy Clin Immunol* **125**, S3-23 (2010).
        https://doi.org:10.1016/j.jaci.2009.12.980

[2]   Janeway, C. A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216 (2002).
        https://doi.org:10.1146/annurev.immunol.20.083001.084359

[3]   Anaya, J. M., Shoenfeld, Y., Rojas-Villarraga, A., Levy, R. A. & Cervera, R. in *Autoimmunity: From Bench to Bedside*   (eds J. M. Anaya *et al.*) (El Rosario University Press, 2013).

[4]   Medzhitov, R. & Janeway, C. A., Jr. Innate immune induction of the adaptive immune response. *Cold Spring Harb Symp Quant Biol* **64**, 429-435 (1999).
        https://doi.org:10.1101/sqb.1999.64.429

[5]   Kenneth Murphy, Casey Weaver & Charles A Janeway, J. *Immunobiology*. Vol. 9 (Garland Science, 2016).

[6]   Chang, H.-F. *et al.* Preparing the lethal hit: interplay between exo- and endocytic pathways in cytotoxic T lymphocytes. *Cell Mol Life Sci* **74**, 399-408 (2017).
        https://doi.org:10.1007/s00018-016-2350-7

[7]   Robinson, J. *et al.* Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genet* **13**, e1006862 (2017).
        https://doi.org:10.1371/journal.pgen.1006862

[8]   Müller, I. K. *et al.* Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity. *Nat Commun* **13**, 5383 (2022).
        https://doi.org:10.1038/s41467-022-32841-9

[9]   Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. *Annu Rev Immunol* **31**, 443-473 (2013).
        https://doi.org:10.1146/annurev-immunol-032712-095910

[10]   Trowitzsch, S. & Tampé, R. Multifunctional chaperone and quality control complexes in adaptive immunity. *Annu Rev Biophys* **49**, 135-161 (2020).
        https://doi.org:10.1146/annurev-biophys-121219-081643

[11]   Domnick, A. *et al.* Molecular basis of MHC I quality control in the peptide loading complex. *Nat Commun* **13**, 4701 (2022).
        https://doi.org:10.1038/s41467-022-32384-z

[12]   Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70**, 1035-1048 (1992).
        https://doi.org:10.1016/0092-8674(92)90252-8

[13]   Madden, D. R. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* **13**, 587-622 (1995).
        https://doi.org:10.1146/annurev.iy.13.040195.003103

[14]   Kozlov, G. & Gehring, K. Calnexin cycle - structural features of the ER chaperone system. *FEBS J* **287**, 4322-4340 (2020).
        https://doi.org:10.1111/febs.15330

[15]   Thomas, C. & Tampé, R. MHC I assembly and peptide editing - chaperones, clients, and molecular plasticity in immunity. *Curr Opin Immunol* **70**, 48-56 (2021).
        https://doi.org:10.1016/j.coi.2021.02.004

[16]   Chapman, D. C. & Williams, D. B. ER quality control in the biogenesis of MHC class I molecules. *Semin Cell Dev Biol* **21**, 512-519 (2010).
        https://doi.org:10.1016/j.semcdb.2009.12.013

[17]   Fisette, O., Schroder, G. F. & Schafer, L. V. Atomistic structure and dynamics of the human MHC-I peptide-loading complex. *Proc Natl Acad Sci U S A* **117**, 20597-20606 (2020).

https://doi.org:10.1073/pnas.2004445117

[18] Blees, A. *et al.* Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525-528 (2017).
https://doi.org:10.1038/nature24627

[19] Peaper, D. R. & Cresswell, P. Regulation of MHC class I assembly and peptide binding. *Annu Rev Cell Dev Biol* **24**, 343-368 (2008).
https://doi.org:10.1146/annurev.cellbio.24.110707.175347

[20] Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895-899 (2003).
https://doi.org:10.1038/nature02263

[21] Dolan, B. P., Li, L., Takeda, K., Bennink, J. R. & Yewdell, J. W. Defective ribosomal products are the major source of antigenic peptides endogenously generated from influenza A virus neuraminidase. *J Immunol* **184**, 1419-1424 (2010).
https://doi.org:10.4049/jimmunol.0901907

[22] Koopmann, J. O., Post, M., Neefjes, J. J., Hammerling, G. J. & Momburg, F. Translocation of long peptides by transporters associated with antigen processing (TAP). *Eur J Immunol* **26**, 1720-1728 (1996).
https://doi.org:10.1002/eji.1830260809

[23] Neefjes, J. J., Momburg, F. & Hammerling, G. J. Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science* **261**, 769-771 (1993).
https://doi.org:10.1126/science.8342042

[24] Uebel, S. *et al.* Requirements for peptide binding to the human transporter associated with antigen processing revealed by peptide scans and complex peptide libraries. *J Biol Chem* **270**, 18512-18516 (1995).
https://doi.org:10.1074/jbc.270.31.18512

[25] Kisselev, A. F., Akopian, T. N., Woo, K. M. & Goldberg, A. L. The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J Biol Chem* **274**, 3363-3371 (1999).
https://doi.org:10.1074/jbc.274.6.3363

[26] Evnouchidou, I. & van Endert, P. Peptide trimming by endoplasmic reticulum aminopeptidases: Role of MHC class I binding and ERAP dimerization. *Hum Immunol* **80**, 290-295 (2019).
https://doi.org:10.1016/j.humimm.2019.01.003

[27] Serwold, T., Gonzalez, F., Kim, J., Jacob, R. & Shastri, N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419**, 480-483 (2002).
https://doi.org:10.1038/nature01074

[28] York, I. A. *et al.* The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. *Nat Immunol* **3**, 1177-1184 (2002).
https://doi.org:10.1038/ni860

[29] Chang, S. C., Momburg, F., Bhutani, N. & Goldberg, A. L. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc Natl Acad Sci U S A* **102**, 17107-17112 (2005).
https://doi.org:10.1073/pnas.0500721102

[30] Garrett, T. P., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. Specificity pockets for the side chains of peptide antigens in HLA-Aw68. *Nature* **342**, 692-696 (1989).
https://doi.org:10.1038/342692a0

[31] Thomas, C. & Tampé, R. Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing. *Science* **358**, 1060-1064 (2017).
https://doi.org:10.1126/science.aao6001

[32]   Hermann, C. *et al.* TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst. *eLife* **4** (2015).
https://doi.org/10.7554/eLife.09617

[33]   Morozov, G. I. *et al.* Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *Proc Natl Acad Sci U S A* **113**, E1006-1015 (2016).
https://doi.org/10.1073/pnas.1519894113

[34]   Dong, G., Wearsch, P. A., Peaper, D. R., Cresswell, P. & Reinisch, K. M. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* **30**, 21-32 (2009).
https://doi.org/10.1016/j.immuni.2008.10.018

[35]   Neerincx, A. *et al.* TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway. *eLife* **6** (2017).
https://doi.org/10.7554/eLife.23049

[36]   Strous, G. J., Van Kerkhof, P., Brok, R., Roth, J. & Brada, D. Glucosidase II, a protein of the endoplasmic reticulum with high mannose oligosaccharide chains and a rapid turnover. *J Biol Chem* **262**, 3620-3625 (1987).
https://pubmed.ncbi.nlm.nih.gov/3546312/

[37]   Sagert, L. *et al.* TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I. *manuscript in preparation* (2022).

[38]   Harndahl, M. *et al.* Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur J Immunol* **42**, 1405-1416 (2012).
https://doi.org/10.1002/eji.201141774

[39]   Silva, Z. *et al.* MHC Class I Stability is Modulated by Cell Surface Sialylation in Human Dendritic Cells. *Pharmaceutics* **12** (2020).
https://doi.org/10.3390/pharmaceutics12030249

[40]   Reits, E. *et al.* Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. *Immunity* **18**, 97-108 (2003).
https://doi.org/10.1016/s1074-7613(02)00511-3

[41]   Gaudet, R. & Wiley, D. C. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J* **20**, 4964-4972 (2001).
https://doi.org/10.1093/emboj/20.17.4964

[42]   Van Kaer, L., Ashton-Rickardt, P. G., Ploegh, H. L. & Tonegawa, S. TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4-8+ T cells. *Cell* **71**, 1205-1214 (1992).
https://doi.org/10.1016/s0092-8674(05)80068-6

[43]   Koch, J., Guntrum, R., Heintke, S., Kyritsis, C. & Tampé, R. Functional dissection of the transmembrane domains of the transporter associated with antigen processing (TAP). *J Biol Chem* **279**, 10142-10147 (2004).
https://doi.org/10.1074/jbc.M312816200

[44]   Seyffer, F. & Tampé, R. ABC transporters in adaptive immunity. *Biochim Biophys Acta* **1850**, 449-460 (2015).
https://doi.org/10.1016/j.bbagen.2014.05.022

[45]   Oldham, M. L., Grigorieff, N. & Chen, J. Structure of the transporter associated with antigen processing trapped by herpes simplex virus. *eLife* **5** (2016).
https://doi.org/10.7554/eLife.21829

[46]   Androlewicz, M. J., Anderson, K. S. & Cresswell, P. Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. *Proc Natl Acad Sci U S A* **90**, 9130-9134 (1993).
https://doi.org/10.1073/pnas.90.19.9130

[47]   Androlewicz, M. J. & Cresswell, P. Human transporters associated with antigen processing possess a promiscuous peptide-binding site. *Immunity* **1**, 7-14 (1994). https://doi.org/10.1016/1074-7613(94)90004-3

[48]   van Endert, P. M. *et al*. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity* **1**, 491-500 (1994). https://doi.org/10.1016/1074-7613(94)90091-4

[49]   Dalmas, O. *et al*. The Q-loop disengages from the first intracellular loop during the catalytic cycle of the multidrug ABC transporter BmrA. *J Biol Chem* **280**, 36857-36864 (2005). https://doi.org/10.1074/jbc.M503266200

[50]   Parcej, D. & Tampé, R. ABC proteins in antigen translocation and viral inhibition. *Nat Chem Biol* **6**, 572-580 (2010). https://doi.org/10.1038/nchembio.410

[51]   Herget, M. *et al*. Conformation of peptides bound to the transporter associated with antigen processing (TAP). *Proc Natl Acad Sci U S A* **108**, 1349-1354 (2011). https://doi.org/10.1073/pnas.1012355108

[52]   Mayerhofer, P. U. & Tampé, R. Antigen translocation machineries in adaptive immunity and viral immune evasion. *J Mol Biol* **427**, 1102-1118 (2015). https://doi.org/10.1016/j.jmb.2014.09.006

[53]   Uebel, S. *et al*. Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proc Natl Acad Sci U S A* **94**, 8976-8981 (1997). https://doi.org/10.1073/pnas.94.17.8976

[54]   Gorbulev, S., Abele, R. & Tampé, R. Allosteric crosstalk between peptide-binding, transport, and ATP hydrolysis of the ABC transporter TAP. *Proc Natl Acad Sci U S A* **98**, 3732-3737 (2001). https://doi.org/10.1073/pnas.061467898

[55]   Tampé, R. *Proposal for the Establishment and Funding of Collaborative Research Center 1507: Membrane-associated Protein Assemblies, Machineries and Supercomplexes*. (Goethe University Frankfurt, 2021).

[56]   Ahn, K. *et al*. The ER-luminal domain of the HCMV glycoprotein US6 inhibits peptide translocation by TAP. *Immunity* **6**, 613-621 (1997). https://doi.org/10.1016/s1074-7613(00)80349-0

[57]   Horst, D. *et al*. Specific targeting of the EBV lytic phase protein BNLF2a to the transporter associated with antigen processing results in impairment of HLA class I-restricted antigen presentation. *J Immunol* **182**, 2313-2324 (2009). https://doi.org/10.4049/jimmunol.0803218

[58]   Alzhanova, D. & Fruh, K. Modulation of the host immune response by cowpox virus. *Microbes Infect* **12**, 900-909 (2010). https://doi.org/10.1016/j.micinf.2010.07.007

[59]   Koppers-Lalic, D. *et al*. Varicelloviruses avoid T cell recognition by UL49.5-mediated inactivation of the transporter associated with antigen processing. *Proc Natl Acad Sci U S A* **102**, 5144-5149 (2005). https://doi.org/10.1073/pnas.0501463102

[60]   Harvey, I. B., Wang, X. & Fremont, D. H. Molluscum contagiosum virus MC80 sabotages MHC-I antigen presentation by targeting tapasin for ER-associated degradation. *PLoS Pathog* **15**, e1007711 (2019). https://doi.org/10.1371/journal.ppat.1007711

[61]   York, I. A. *et al*. A cytosolic herpes simplex virus protein inhibits antigen presentation to CD8+ T lymphocytes. *Cell* **77**, 525-535 (1994). https://doi.org/10.1016/0092-8674(94)90215-1

[62]   Früh, K. *et al*. A viral inhibitor of peptide transporters for antigen presentation. *Nature* **375**, 415-418 (1995). https://doi.org/10.1038/375415a0

[63]   Hill, A. *et al*. Herpes simplex virus turns off the TAP to evade host immunity. *Nature* **375**, 411-415 (1995).

https://doi.org:10.1038/375411a0

[64] Ahn, K. *et al*. Molecular mechanism and species specificity of TAP inhibition by herpes simplex virus ICP47. *EMBO J* **15**, 3247-3255 (1996).
https://pubmed.ncbi.nlm.nih.gov/8670825/

[65] Tomazin, R. *et al*. Stable binding of the herpes simplex virus ICP47 protein to the peptide binding site of TAP. *EMBO J* **15**, 3256-3266 (1996).
https://pubmed.ncbi.nlm.nih.gov/8670826/

[66] Sethumadhavan, S. *et al*. Viral immune evasins impact antigen presentation by allele-specific trapping of MHC I at the peptide-loading complex. *Sci Rep* **12**, 1516 (2022).
https://doi.org:10.1038/s41598-022-05000-9

[67] Braner, M. *et al*. Optical control of the antigen translocation by synthetic photo-conditional viral inhibitors. *Chem Sci* **10**, 2001-2005 (2019).
https://doi.org:10.1039/c8sc04863k

[68] Winter, C., Domnick, A., Cernova, D. & Tampé, R. Semisynthetic Viral Inhibitor for Light Control of the MHC I Peptide Loading Complex. *Angew Chem Int Ed*, e202211826 (2022).
https://doi.org:10.1002/anie.202211826

[69] Beinert, D., Neumann, L., Uebel, S. & Tampé, R. Structure of the viral TAP-inhibitor ICP47 induced by membrane association. *Biochemistry* **36**, 4694-4700 (1997).
https://doi.org:10.1021/bi962940v

[70] Herbring, V., Baucker, A., Trowitzsch, S. & Tampé, R. A dual inhibition mechanism of herpesviral ICP47 arresting a conformationally thermostable TAP complex. *Sci Rep* **6**, 36907 (2016).
https://doi.org:10.1038/srep36907

[71] Frenkel, E. J., Roelofsen, B., Brodbeck, U., van Deenen, L. L. & Ott, P. Lipid-protein interactions in human erythrocyte-membrane acetylcholinesterase. Modulation of enzyme activity by lipids. *Eur J Biochem* **109**, 377-382 (1980).
https://doi.org:10.1111/j.1432-1033.1980.tb04804.x

[72] Majeed, S., Ahmad, A. B., Sehar, U. & Georgieva, E. R. Lipid Membrane Mimetics in Functional and Structural Studies of Integral Membrane Proteins. *Membranes* **11** (2021).
https://doi.org:10.3390/membranes11090685

[73] Herget, M. *et al*. Purification and reconstitution of the antigen transport complex TAP: a prerequisite for determination of peptide stoichiometry and ATP hydrolysis. *J Biol Chem* **284**, 33740-33749 (2009).
https://doi.org:10.1074/jbc.M109.047779

[74] Autzen, H. E., Julius, D. & Cheng, Y. Membrane mimetic systems in CryoEM: keeping membrane proteins in their native environment. *Curr Opin Struct Biol* **58**, 259-268 (2019).
https://doi.org:10.1016/j.sbi.2019.05.022

[75] Tribet, C., Audebert, R. & Popot, J. L. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc Natl Acad Sci U S A* **93**, 15047-15050 (1996).
https://doi.org:10.1073/pnas.93.26.15047

[76] Popot, J. L. Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu Rev Biochem* **79**, 737-775 (2010).
https://doi.org:10.1146/annurev.biochem.052208.114057

[77] Popot, J. L. *et al*. Amphipols from A to Z. *Annu Rev Biophys* **40**, 379-408 (2011).
https://doi.org:10.1146/annurev-biophys-042910-155219

[78] Grinkova, Y. V., Denisov, I. G. & Sligar, S. G. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng Des Sel* **23**, 843-848 (2010).
https://doi.org:10.1093/protein/gzq060

[79] Coudray, N. *et al*. Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *eLife* **9** (2020). https://doi.org/10.7554/elife.62518

[80] Ritchie, T. K. *et al*. Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol*, 211-231 (2009). https://doi.org/10.1016/s0076-6879(09)64011-8

[81] Ram, P. & Prestegard, J. H. Magnetic field induced ordering of bile salt/phospholipid micelles: new media for NMR structural investigations. *Biochim Biophys Acta* **940**, 289-294 (1988). https://doi.org/10.1016/0005-2736(88)90203-9

[82] Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett* **2**, 853-856 (2002). https://doi.org/10.1021/nl025623k

[83] Frauenfeld, J. *et al*. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat Methods* **13**, 345-351 (2016). https://doi.org/10.1038/nmeth.3801

[84] Knowles, T. J. *et al*. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *J Am Chem Soc* **131**, 7484-7485 (2009). https://doi.org/10.1021/ja810046q

[85] Oluwole, A. O. *et al*. Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angew Chem Int Ed* **56**, 1919-1924 (2017). https://doi.org/10.1002/anie.201610778

[86] Carlson, M. L. *et al*. The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *eLife* **7** (2018). https://doi.org/10.7554/elife.34085

[87] Schuler, M. A., Denisov, I. G. & Sligar, S. G. Nanodiscs as a New Tool to Examine Lipid–Protein Interactions. *Methods Mol Biol*, 415-433 (2013). https://doi.org/10.1007/978-1-62703-275-9_18

[88] Denisov, I. G. & Sligar, S. G. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem Rev* **117**, 4669-4713 (2017). https://doi.org/10.1021/acs.chemrev.6b00690

[89] Akbarzadeh, A. *et al*. Liposome: classification, preparation, and applications. *Nanoscale Res Lett* **8**, 102 (2013). https://doi.org/10.1186/1556-276x-8-102

[90] Smirnova, I. A., Ädelroth, P. & Brzezinski, P. Extraction and liposome reconstitution of membrane proteins with their native lipids without the use of detergents. *Sci Rep* **8** (2018). https://doi.org/10.1038/s41598-018-33208-1

[91] Mouritsen, O. G. Model Answers to Lipid Membrane Questions. *Cold Spring Harb Perspect Biol* **3**, a004622-a004622 (2011). https://doi.org/10.1101/cshperspect.a004622

[92] Hickey, K. D. & Buhr, M. M. Lipid Bilayer Composition Affects Transmembrane Protein Orientation and Function. *J Lipids* **2011**, 1-9 (2011). https://doi.org/10.1155/2011/208457

[93] Hresko, R. C., Kraft, T. E., Quigley, A., Carpenter, E. P. & Hruz, P. W. Mammalian Glucose Transporter Activity Is Dependent upon Anionic and Conical Phospholipids. *J Biol Chem* **291**, 17271-17282 (2016). https://doi.org/10.1074/jbc.m116.730168

[94] Helvig, S., Azmi, I. D. M., Moghimi, S. M. & Yaghmur, A. Recent Advances in Cryo-TEM Imaging of Soft Lipid Nanoparticles. *AIMS Biophys*. **2**, 116-130 (2015). https://doi.org/10.3934/biophy.2015.2.116

[95] Yao, X., Fan, X. & Yan, N. Cryo-EM analysis of a membrane protein embedded in the liposome. *Proc Natl Acad Sci U S A* **117**, 18497-18503 (2020). https://doi.org/10.1073/pnas.2009385117

[96] Stokes, D. L., Rice, W. J., Hu, M., Kim, C. & Ubarretxena-Belandia, I. Two-Dimensional Crystallization of Integral Membrane Proteins for Electron Crystallography. *Methods Mol Biol*, 187-205 (2010). https://doi.org/10.1007/978-1-60761-762-4_10

[97] Sejwal, K. *et al.* Proteoliposomes – a system to study membrane proteins under buffer gradients by cryo-EM. *Nanotechnol Rev* **6**, 57-74 (2017). https://doi.org/10.1515/ntrev-2016-0081

[98] Sharp, T. H., Koster, A. J. & Gros, P. Heterogeneous MAC Initiator and Pore Structures in a Lipid Bilayer by Phase-Plate Cryo-electron Tomography. *Cell Rep* **15**, 1-8 (2016). https://doi.org/10.1016/j.celrep.2016.03.002

[99] Dunstone, M. A. & de Marco, A. Cryo-electron tomography: an ideal method to study membrane-associated proteins. *Philos Trans R Soc Lond B Biol Sci* **372** (2017). https://doi.org/10.1098/rstb.2016.0210

[100] Lucic, V., Leis, A. & Baumeister, W. Cryo-electron tomography of cells: connecting structure and function. *Histochem Cell Biol* **130**, 185-196 (2008). https://doi.org/10.1007/s00418-008-0459-y

[101] Baumeister, W. Cryo-electron tomography: A long journey to the inner space of cells. *Cell* **185**, 2649-2652 (2022). https://doi.org/10.1016/j.cell.2022.06.034

[102] Johnson, Z. L. & Lee, S. Y. Liposome reconstitution and transport assay for recombinant transporters. *Methods Enzymol* **556**, 373-383 (2015). https://doi.org/10.1016/bs.mie.2014.11.048

[103] Wieneke, R., Raulf, A., Kollmannsperger, A., Heilemann, M. & Tampé, R. SLAP: Small Labeling Pair for Single-Molecule Super-Resolution Imaging. *Angew Chem Int Ed Engl* **54**, 10216-10219 (2015). https://doi.org/10.1002/anie.201503215

[104] Koller, N. *et al.* Nanoscale organization of the MHC I peptide-loading complex in human dendritic cells. *Cell Mol Life Sci* **79** (2022). https://doi.org/10.1007/s00018-022-04472-2

[105] Ritter, C., Quirin, K., Kowarik, M. & Helenius, A. Minor folding defects trigger local modification of glycoproteins by the ER folding sensor GT. *EMBO J* **24**, 1730-1738 (2005). https://doi.org/10.1038/sj.emboj.7600645

[106] Avezov, E., Frenkel, Z., Ehrlich, M., Herscovics, A. & Lederkremer, G. Z. Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man5-6GlcNAc2 in glycoprotein ER-associated degradation. *Mol Biol Cell* **19**, 216-225 (2008). https://doi.org/10.1091/mbc.e07-05-0505

[107] D'Alessio, C., Caramelo, J. J. & Parodi, A. J. UDP-GlC:glycoprotein glucosyltransferase-glucosidase II, the ying-yang of the ER quality control. *Semin Cell Dev Biol* **21**, 491-499 (2010). https://doi.org/10.1016/j.semcdb.2009.12.014

[108] Pelletier, M. F. *et al.* The heterodimeric structure of glucosidase II is required for its activity, solubility, and localization in vivo. *Glycobiology* **10**, 815-827 (2000). https://doi.org/10.1093/glycob/10.8.815

[109] Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate specificity analysis of endoplasmic reticulum glucosidase II using synthetic high mannose-type glycans. *J Biol Chem* **281**, 31502-31508 (2006). https://doi.org/10.1074/jbc.M605457200

[110] Alonso, J. M., Santa-Cecilia, A. & Calvo, P. Glucosidase II from rat liver microsomes. Kinetic model for binding and hydrolysis. *Biochem.* **278**, 721-727 (1991).
https://doi.org:10.1042/bj2780721

[111] D'Alessio, C., Fernandez, F., Trombetta, E. S. & Parodi, A. J. Genetic evidence for the heterodimeric structure of glucosidase II. The effect of disrupting the subunit-encoding genes on glycoprotein folding. *J Biol Chem* **274**, 25899-25905 (1999).
https://doi.org:10.1074/jbc.274.36.25899

[112] Grinna, L. S. & Robbins, P. W. Substrate specificities of rat liver microsomal glucosidases which process glycoproteins. *J Biol Chem* **255**, 2255-2258 (1980).
https://pubmed.ncbi.nlm.nih.gov/7358666/

[113] Totani, K., Ihara, Y., Matsuo, I. & Ito, Y. Substrate Specificity Analysis of Endoplasmic Reticulum Glucosidase II Using Synthetic High Mannose-type Glycans. *J Biol Chem* **281**, 31502-31508 (2006).
https://doi.org:10.1074/jbc.m605457200

[114] Zuber, C. *et al.* Immunolocalization of UDP-glucose:glycoprotein glucosyltransferase indicates involvement of pre-Golgi intermediates in protein quality control. *Proc Natl Acad Sci U S A* **98**, 10710-10715 (2001).
https://doi.org:10.1073/pnas.191359198

[115] Guerin, M. & Parodi, A. J. The UDP-glucose:Glycoprotein Glucosyltransferase Is Organized in at Least Two Tightly Bound Domains from Yeast to Mammals. *J Biol Chem* **278**, 20540-20546 (2003).
https://doi.org:10.1074/jbc.m300891200

[116] Modenutti, C. P. *et al.* Clamping, bending, and twisting inter-domain motions in the misfold-recognizing portion of UDP-glucose: Glycoprotein glucosyltransferase. *Structure* **29**, 357-370 e359 (2021).
https://doi.org:10.1016/j.str.2020.11.017

[117] Labriola, C., Cazzulo, J. J. & Parodi, A. J. <i>Trypanosoma cruzi</i>Calreticulin Is a Lectin That Binds Monoglucosylated Oligosaccharides but Not Protein Moieties of Glycoproteins. *Mol Biol Cell* **10**, 1381-1394 (1999).
https://doi.org:10.1091/mbc.10.5.1381

[118] Sousa, M. C., Ferrero-Garcia, M. A. & Parodi, A. J. Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* **31**, 97-105 (1992).
https://doi.org:10.1021/bi00116a015

[119] Sagert, L., Hennig, F., Thomas, C. & Tampé, R. A loop structure allows TAPBPR to exert its dual function as MHC I chaperone and peptide editor. *eLife* **9** (2020).
https://doi.org:10.7554/elife.55326

[120] Zhang, W., Wearsch, P. A., Zhu, Y., Leonhardt, R. M. & Cresswell, P. A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules. *Proc Natl Acad Sci U S A* **108**, 4956-4961 (2011).
https://doi.org:10.1073/pnas.1102527108

[121] Guile, G. R., Rudd, P. M., Wing, D. R., Prime, S. B. & Dwek, R. A. A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal Biochem* **240**, 210-226 (1996).
https://doi.org:10.1006/abio.1996.0351

[122] Radcliffe, C. M. *et al.* Identification of Specific Glycoforms of Major Histocompatibility Complex Class I Heavy Chains Suggests That Class I Peptide Loading Is an Adaptation of the Quality Control Pathway Involving Calreticulin and ERp57. *J Biol Chem* **277**, 46415-46423 (2002).
https://doi.org:10.1074/jbc.m202466200

[123] Li, J. *et al.* Advances in glycopeptide enrichment methods for the analysis of protein glycosylation over the past decade. *J Sep Sci* (2022).
https://doi.org:10.1002/jssc.202200292

[124] Nakayasu, E. S. *et al.* Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat Protoc* **16**, 3737-3760 (2021).
https://doi.org:10.1038/s41596-021-00566-6

[125] Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A. & Zahedi, R. P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J Proteomics* **75**, 1454-1462 (2012).
https://doi.org:10.1016/j.jprot.2011.11.016

[126] Kang, L., Weng, N. & Jian, W. LC-MS bioanalysis of intact proteins and peptides. *Biomed Chromatogr* **34**, e4633 (2020).
https://doi.org:10.1002/bmc.4633

[127] Yang, X. & Bartlett, M. G. Glycan analysis for protein therapeutics. *J Chromatogr B Analyt Technol Biomed Life Sci* **1120**, 29-40 (2019).
https://doi.org:10.1016/j.jchromb.2019.04.031

[128] Cupp-Sutton, K. A. & Wu, S. High-throughput quantitative top-down proteomics. *Mol Omics* **16**, 91-99 (2020).
https://doi.org:10.1039/c9mo00154a

[129] Fussl, F., Strasser, L., Carillo, S. & Bones, J. Native LC-MS for capturing quality attributes of biopharmaceuticals on the intact protein level. *Curr Opin Biotechnol* **71**, 32-40 (2021).
https://doi.org:10.1016/j.copbio.2021.05.008

[130] Astefanei, A., Dapic, I. & Camenzuli, M. Different Stationary Phase Selectivities and Morphologies for Intact Protein Separations. *Chromatographia* **80**, 665-687 (2017).
https://doi.org:10.1007/s10337-016-3168-z

[131] Qiu, X., Kang, L., Case, M., Weng, N. & Jian, W. Quantitation of intact monoclonal antibody in biological samples: comparison of different data processing strategies. *Bioanalysis* **10**, 1055-1067 (2018).
https://doi.org:10.4155/bio-2018-0016

[132] Tarentino, A. L. & Plummer, T. H., Jr. Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. *Methods Enzymol* **230**, 44-57 (1994).
https://doi.org:10.1016/0076-6879(94)30006-2

[133] Kellie, J. F., Kehler, J. R. & Szapacs, M. E. Application of high-resolution MS for development of peptide and large-molecule drug candidates. *Bioanalysis* **8**, 169-177 (2016).
https://doi.org:10.4155/bio.15.249

[134] Reinhold, B. B. & Reinhold, V. N. Electrospray ionization mass spectrometry: Deconvolution by an Entropy-Based algorithm. *J Am Soc Mass Spectrom* **3**, 207-215 (1992).
https://doi.org:10.1016/1044-0305(92)87004-i

[135] Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **159**, 995-1014 (2014).
https://doi.org:10.1016/j.cell.2014.10.051

[136] Berman, H. M. *et al. The Protein Data Bank - Number of Released PDB Structures per Year*, <https://www.rcsb.org/stats/all-released-structures> (2022).

[137] Kabsch, W. The rotation method in crystallography edited by U. W. Arndt and A. J. Wonacott. *Acta Crystallogr B* **34**, 1049-1050 (1978).
https://doi.org:https://doi.org/10.1107/S0567740878004896

[138] Kraft, P. *et al.* Performance of single-photon-counting PILATUS detector modules. *J Synchrotron Radiat* **16**, 368-375 (2009).
https://doi.org:10.1107/S0909049509009911

[139] Taylor, G. L. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66**, 325-338 (2010).

https://doi.org:10.1107/S0907444910006694

[140] Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. & Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J* **278**, 687-703 (2011). https://doi.org:10.1111/j.1742-4658.2011.08004.x

[141] Marion, D. An introduction to biological NMR spectroscopy. *Mol Cell Proteomics* **12**, 3006-3025 (2013). https://doi.org:10.1074/mcp.O113.030239

[142] Curry, S. Structural Biology: A Century-long Journey into an Unseen World. *Interdiscip Sci Rev* **40**, 308-328 (2015). https://doi.org:10.1179/0308018815z.000000000120

[143] De Rosier, D. J. & Klug, A. Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 130-134 (1968). https://doi.org:10.1038/217130a0

[144] Dubochet, J. & McDowall, A. W. VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *J Microsc* **124**, 3-4 (1981). https://doi.org:10.1111/j.1365-2818.1981.tb02483.x

[145] Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**, 129-228 (1988). https://doi.org:10.1017/s0033583500004297

[146] Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876-880 (2018). https://doi.org:10.1126/science.aat4346

[147] Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **343**, 1443-1444 (2014). https://doi.org:10.1126/science.1251652

[148] Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017). https://doi.org:10.1038/nmeth.4169

[149] Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160 (2019). https://doi.org:10.1038/s41592-019-0575-8

[150] Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat Commun* **11**, 5208 (2020). https://doi.org:10.1038/s41467-020-18952-1

[151] Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185 (2021). https://doi.org:10.1038/s41592-020-01049-4

[152] Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci* **26**, 32-39 (2017). https://doi.org:10.1002/pro.3022

[153] Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706-710 (2020). https://doi.org:10.1038/s41586-019-1923-7

[154] Service, R. F. STRUCTURAL BIOLOGY 'The game has changed.' AI triumphs at protein folding. *Science* **370**, 1144-1145 (2020). https://doi.org:10.1126/science.370.6521.1144

[155] Levinthal, C. How to fold graciously. *Mossb Spect Biol Proc* **67**, 22-26 (1969).

[156] Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-controlled tools. *Angew Chem Int Ed Engl* **51**, 8446-8476 (2012). https://doi.org:10.1002/anie.201202134

[157] Beharry, A. A. & Woolley, G. A. Azobenzene photoswitches for biomolecules. *Chem Soc Rev* **40**, 4422-4437 (2011).

https://doi.org:10.1039/c1cs15023e

[158] Gorostiza, P. & Isacoff, E. Y. Optical switches for remote and noninvasive control of cell signaling. *Science* **322**, 395-399 (2008).
https://doi.org:10.1126/science.1166022

[159] Kaplan, J. H., Forbush, B. & Hoffman, J. F. RAPID PHOTOLYTIC RELEASE OF ADENOSINE 5'-TRIPHOSPHATE FROM A PROTECTED ANALOG - UTILIZATION BY NA-K PUMP OF HUMAN RED BLOOD-CELL GHOSTS. *Biochemistry* **17**, 1929-1935 (1978).
https://doi.org:10.1021/bi00603a020

[160] Klan, P. *et al.* Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy. *Chem Rev* **113**, 119-191 (2013).
https://doi.org:10.1021/cr300177k

[161] Hasan, A. *et al.* Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates. *Tetrahedron* **53**, 4247-4264 (1997).
https://doi.org:10.1016/s0040-4020(97)00154-3

[162] Parker, L. L., Kurutz, J. W., Kent, S. B. H. & Kron, S. J. Control of the Yeast Cell Cycle with a Photocleavable α -Factor Analogue. *Angew Chem Int Ed* **45**, 6322-6325 (2006).
https://doi.org:10.1002/anie.200602439

[163] Toebes, M. *et al.* Design and use of conditional MHC class I ligands. *Nat Med* **12**, 246-251 (2006).
https://doi.org:10.1038/nm1360

[164] Doerner, J. F., Febvay, S. & Clapham, D. E. Controlled delivery of bioactive molecules into live cells using the bacterial mechanosensitive channel MscL. *Nat Commun* **3**, 990 (2012).
https://doi.org:10.1038/ncomms1999

[165] Sharei, A. *et al.* Cell Squeezing as a Robust, Microfluidic Intracellular Delivery Platform. *J Vis Exp* (2013).
https://doi.org:10.3791/50980

[166] Schmierer, T. *et al.* Femtosecond spectroscopy on the photochemistry of ortho-nitrotoluene. *Phys Chem Chem Phys* **12**, 15653-15664 (2010).
https://doi.org:10.1039/c004025h

[167] Norrish, R. G. W. & Kirkbride, F. W. 204. Primary photochemical processes. Part I. The decomposition of formaldehyde. *J. Chem. Soc.*, 1518-1530 (1932).
https://doi.org:10.1039/JR9320001518

[168] Wang, P. Photolabile Protecting Groups: Structure and Reactivity. *Asian J Org Chem* **2**, 452-464 (2013).
https://doi.org:10.1002/ajoc.201200197

[169] Burai, R., Ait-Bouziad, N., Chiki, A. & Lashuel, H. A. Elucidating the Role of Site-Specific Nitration of alpha-Synuclein in the Pathogenesis of Parkinson's Disease via Protein Semisynthesis and Mutagenesis. *J Am Chem Soc* **137**, 5041-5052 (2015).
https://doi.org:10.1021/ja5131726

[170] Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301**, 964-967 (2003).
https://doi.org:10.1126/science.1084772

[171] Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53-60 (2017).
https://doi.org:10.1038/nature24031

[172] Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew Chem Int Ed Engl* **50**, 3878-3881 (2011).
https://doi.org:10.1002/anie.201008178

[173] Nikic, I., Kang, J. H., Girona, G. E., Aramburu, I. V. & Lemke, E. A. Labeling proteins on live mammalian cells using click chemistry. *Nat Protoc* **10**, 780-791 (2015).
https://doi.org:10.1038/nprot.2015.045

[174] Saal, K.-A., Richter, F., Rehling, P. & Rizzoli, S. O. Combined Use of Unnatural Amino Acids Enables Dual-Color Super-Resolution Imaging of Proteins *via* Click Chemistry. *ACS Nano* **12**, 12247-12254 (2018). https://doi.org:10.1021/acsnano.8b06047

[175] Charbon, G. *et al.* Subcellular Protein Localization by Using a Genetically Encoded Fluorescent Amino Acid. *ChemBioChem* **12**, 1818-1821 (2011). https://doi.org:10.1002/cbic.201100282

[176] Lang, K. *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* **4**, 298-304 (2012). https://doi.org:10.1038/nchem.1250

[177] Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* **45**, 2728-2731 (2006). https://doi.org:10.1002/anie.200600264

[178] Luo, J., Torres-Kolbus, J., Liu, J. & Deiters, A. Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. *ChemBioChem* **18**, 1442-1447 (2017). https://doi.org:10.1002/cbic.201700147

[179] Wang, L. Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. *Acc Chem Res* **50**, 2767-2775 (2017). https://doi.org:10.1021/acs.accounts.7b00376

[180] Rackham, O. & Chin, J. W. A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol* **1**, 159-166 (2005). https://doi.org:10.1038/nchembio719

[181] Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr Opin Chem Biol* **11**, 537-542 (2007). https://doi.org:10.1016/j.cbpa.2007.08.011

[182] Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441-444 (2010). https://doi.org:10.1038/nature08817

[183] Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514-518 (2019). https://doi.org:10.1038/s41586-019-1192-5

[184] Mukai, T. *et al.* Codon reassignment in the Escherichia coli genetic code. *Nucleic Acids Res* **38**, 8188-8195 (2010). https://doi.org:10.1093/nar/gkq707

[185] Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357-360 (2013). https://doi.org:10.1126/science.1241459

[186] Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001). https://doi.org:10.1126/science.1060077

[187] Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J Am Chem Soc* **126**, 14306-14307 (2004). https://doi.org:10.1021/ja040175z

[188] Joest, E. F., Winter, C., Wesalo, J. S., Deiters, A. & Tampé, R. Light-guided intrabodies for on-demand in situ target recognition in human cells. *Chem Sci* **12**, 5787-5795 (2021). https://doi.org:10.1039/d1sc01331a

[189] Joest, E. F., Winter, C., Wesalo, J. S., Deiters, A. & Tampé, R. Efficient Amber Suppression via Ribosomal Skipping for In Situ Synthesis of Photoconditional Nanobodies. *ACS Synth Biol* **11**, 1466-1476 (2022). https://doi.org:10.1021/acssynbio.1c00471

[190] Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413-444 (2010). https://doi.org/10.1146/annurev.biochem.052308.105824

[191] Shandell, M. A., Tan, Z. & Cornish, V. W. Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* **60**, 3455-3469 (2021). https://doi.org/10.1021/acs.biochem.1c00286

[192] Chen, J. & Tsai, Y. H. Applications of Genetic Code Expansion in Studying Protein Post-translational Modification. *J Mol Biol* **434**, 167424 (2022). https://doi.org/10.1016/j.jmb.2021.167424

[193] Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* **3**, 769-772 (2007). https://doi.org/10.1038/nchembio.2007.44

[194] Wang, Y. S. *et al.* A genetically encoded photocaged Nepsilon-methyl-L-lysine. *Mol Biosyst* **6**, 1557-1560 (2010). https://doi.org/10.1039/c002155e

[195] Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J Am Chem Soc* **136**, 2240-2243 (2014). https://doi.org/10.1021/ja412191m

[196] Pier, G. B., Lyczak, J. B. & Wetzler, L. M. *Immunology, Infection, and Immunity*. (ASM Press, 2004).

[197] Murphy, K. M. & Weaver, C. *Janeway's Immunobiology: Ninth International Student Edition*. (Garland Science, 2016).

[198] Tang, D.-C., Devit, M. & Johnston, S. A. Genetic immunization is a simple method for eliciting an immune response. *Nature* **356**, 152-154 (1992). https://doi.org/10.1038/356152a0

[199] Chambers, R. S. & Johnston, S. A. High-level generation of polyclonal antibodies by genetic immunization. *Nat Biotechnol* **21**, 1088-1092 (2003). https://doi.org/10.1038/nbt858

[200] Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497 (1975). https://doi.org/10.1038/256495a0

[201] Hudson, P. J. & Souriau, C. Engineered antibodies. *Nat Med* **9**, 129-134 (2003). https://doi.org/10.1038/nm0103-129

[202] Andrew, S. M. & Titus, J. A. Fragmentation of immunoglobulin G. *Curr Protoc Immunol* **Chapter 2**, Unit 2 8 (2001). https://doi.org/10.1002/0471142735.im0208s21

[203] Strong, R. K. *et al.* 3-DIMENSIONAL STRUCTURE OF MURINE ANTI-PARA-AZOPHENYLARSONATE FAB-36-71 .1. X-RAY CRYSTALLOGRAPHY, SITE-DIRECTED MUTAGENESIS, AND MODELING OF THE COMPLEX WITH HAPTEN. *Biochemistry* **30**, 3739-3748 (1991). https://doi.org/10.1021/bi00229a022

[204] Kovari, L. C., Momany, C. & Rossmann, M. G. The use of antibody fragments for crystallization and structure determinations. *Structure* **3**, 1291-1293 (1995). https://doi.org/10.1016/s0969-2126(01)00266-0

[205] De Groof, T. W. M., Bobkov, V., Heukers, R. & Smit, M. J. Nanobodies: New avenues for imaging, stabilizing and modulating GPCRs. *Mol Cell Endocrinol* **484**, 15-24 (2019). https://doi.org/10.1016/j.mce.2019.01.021

[206] Hansmeier, N. R. *et al.* 3D Visualization of Human Blood Vascular Networks Using Single-Domain Antibodies Directed against Endothelial Cell-Selective Adhesion Molecule (ESAM). *Int J Mol Sci* **23** (2022). https://doi.org/10.3390/ijms23084369

[207]   Tang, Z. *et al*. A Human Single-Domain Antibody Elicits Potent Antitumor Activity by Targeting an Epitope in Mesothelin Close to the Cancer Cell Surface. *Mol Cancer Ther* **12**, 416-426 (2013).
https://doi.org/10.1158/1535-7163.mct-12-0731

[208]   Li, N., Fu, H., Hewitt, S. M., Dimitrov, D. S. & Ho, M. Therapeutically targeting glypican-2 via single-domain antibody-based chimeric antigen receptors and immunotoxins in neuroblastoma. *Proc Natl Acad Sci U S A* **114**, E6623-E6631 (2017).
https://doi.org/10.1073/pnas.1706055114

[209]   Gebauer, M. & Skerra, A. Engineered Protein Scaffolds as Next-Generation Therapeutics. *Annu Rev Pharmacol* **60**, 391-415 (2020).
https://doi.org/10.1146/annurev-pharmtox-010818-021118

[210]   Hamers-Casterman, C. *et al*. Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446-448 (1993).
https://doi.org/10.1038/363446a0

[211]   Dooley, H. & Flajnik, M. F. Antibody repertoire development in cartilaginous fish. *Dev Comp Immunol* **30**, 43-56 (2006).
https://doi.org/10.1016/j.dci.2005.06.022

[212]   van der Linden, R. H. *et al*. Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies. *Biochim Biophys Acta* **1431**, 37-46 (1999).
https://doi.org/10.1016/s0167-4838(99)00030-8

[213]   Kunz, P. *et al*. The structural basis of nanobody unfolding reversibility and thermoresistance. *Sci Rep* **8** (2018).
https://doi.org/10.1038/s41598-018-26338-z

[214]   Rothbauer, U. *et al*. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* **3**, 887-889 (2006).
https://doi.org/10.1038/nmeth953

[215]   Maini, R. *et al*. Incorporation of beta-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg Med Chem* **21**, 1088-1096 (2013).
https://doi.org/10.1016/j.bmc.2013.01.002

[216]   Conibear, A. C., Watson, E. E., Payne, R. J. & Becker, C. F. W. Native chemical ligation in protein synthesis and semi-synthesis. *Chem Soc Rev* **47**, 9046-9068 (2018).
https://doi.org/10.1039/c8cs00573g

[217]   Hartrampf, N. *et al*. Synthesis of proteins by automated flow chemistry. *Science* **368**, 980-987 (2020).
https://doi.org/10.1126/science.abb2491

[218]   Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* **126**, 2670-2671 (2004).
https://doi.org/10.1021/ja039915e

[219]   Volkmann, G. & Liu, X. Q. Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein. *PLoS One* **4**, e8381 (2009).
https://doi.org/10.1371/journal.pone.0008381

[220]   Shah, N. H. & Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem Sci* **5**, 446-461 (2014).
https://doi.org/10.1039/C3SC52951G

[221]   Yao, Z. *et al*. Split Intein-Mediated Protein Ligation for detecting protein-protein interactions and their inhibition. *Nat Commun* **11**, 2440 (2020).
https://doi.org/10.1038/s41467-020-16299-1

[222]   Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **95**, 6705-6710 (1998).
https://doi.org/10.1073/pnas.95.12.6705

[223] Liebscher, S. *et al.* N-terminal protein modification by substrate-activated reverse proteolysis. *Angew Chem Int Ed Engl* **53**, 3024-3028 (2014). https://doi.org:10.1002/anie.201307736

[224] Toplak, A., Nuijens, T., Quaedflieg, P., Wu, B. & Janssen, D. B. Peptiligase, an Enzyme for Efficient Chemoenzymatic Peptide Synthesis and Cyclization in Water. *Adv Synth Catal* **358**, 2140-2147 (2016). https://doi.org:10.1002/adsc.201600017

[225] Thompson, R. E., Stevens, A. J. & Muir, T. W. Protein engineering through tandem transamidation. *Nat Chem* **11**, 737-743 (2019). https://doi.org:10.1038/s41557-019-0281-2

[226] Agouridas, V. *et al.* Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem Rev* **119**, 7328-7443 (2019). https://doi.org:10.1021/acs.chemrev.8b00712

[227] Thompson, R. E. & Muir, T. W. Chemoenzymatic Semisynthesis of Proteins. *Chem Rev* **120**, 3051-3126 (2020). https://doi.org:10.1021/acs.chemrev.9b00450

[228] Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994). https://doi.org:10.1126/science.7973629

[229] Gentle, I. E., De Souza, D. P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug Chem* **15**, 658-663 (2004). https://doi.org:10.1021/bc049965o

[230] Li, S.-J. & Hochstrasser, M. A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246-251 (1999). https://doi.org:10.1038/18457

[231] Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **5**, 75-86 (2004). https://doi.org:10.1023/B:JSFG.0000029237.70316.52

[232] Mukhopadhyay, D. & Dasso, M. Modification in reverse: the SUMO proteases. *Trends Biochem Sci* **32**, 286-295 (2007). https://doi.org:10.1016/j.tibs.2007.05.002

[233] Goldman, R., Stoyanovsky, D. A., Day, B. W. & Kagan, V. E. Reduction of phenoxyl radicals by thioredoxin results in selective oxidation of its SH-groups to disulfides. An antioxidant function of thioredoxin. *Biochemistry* **34**, 4765-4772 (1995). https://doi.org:10.1021/bi00014a034

[234] Evans, T. C., Jr., Benner, J. & Xu, M. Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256-2264 (1998). https://doi.org:10.1002/pro.5560071103

[235] Baumruck, A. C., Tietze, D., Steinacker, L. K. & Tietze, A. A. Chemical synthesis of membrane proteins: a model study on the influenza virus B proton channel. *Chem Sci* **9**, 2365-2375 (2018). https://doi.org:10.1039/c8sc00004b

[236] Nishide, K., Shigeta, Y., Obata, K., Inoue, T. & Node, M. Reductive desulfurization using the Raney nickel sodium hypophosphite combination system without racemization of a secondary alcohol. *Tetrahedron Lett* **37**, 2271-2274 (1996). https://doi.org:10.1016/0040-4039(96)00273-0

[237] Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* **123**, 526-533 (2001). https://doi.org:10.1021/ja003265m

[238] Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* **46**, 9248-9252 (2007). https://doi.org:10.1002/anie.200704195

[239] Bondalapati, S., Jbara, M. & Brik, A. Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. *Nat Chem* **8**, 407-418 (2016). https://doi.org/10.1038/nchem.2476

[240] Macey, M. G. *Flow Cytometry: Principles and Applications*. (Humana Press, 2007).

[241] Wilkerson, M. J. Principles and Applications of Flow Cytometry and Cell Sorting in Companion Animal Medicine. *Vet Clin North Am Small Anim Pract* **42**, 53-71 (2012). https://doi.org:https://doi.org/10.1016/j.cvsm.2011.09.012

[242] Adan, A., Alizada, G., Kiraz, Y., Baran, Y. & Nalbant, A. Flow cytometry: basic principles and applications. *Crit Rev Biotechnol* **37**, 163-176 (2017). https://doi.org/10.3109/07388551.2015.1128876

[243] Leach, M., Drummond, M. & Doig, A. *Practical Flow Cytometry in Haematology Diagnosis*. (Wiley, 2013).

[244] Gray, J. W., Dolbeare, F., Pallavicini, M. G., Beisker, W. & Waldman, F. Cell Cycle Analysis Using Flow Cytometry. *Int J Radiat Biol* **49**, 237-255 (1986). https://doi.org/10.1080/09553008514552531

[245] Darzynkiewicz, Z. *et al.* Cytometry in cell necrobiology: analysis of apoptosis and accidental cell death (necrosis). *Cytometry* **27**, 1-20 (1997). https://pubmed.ncbi.nlm.nih.gov/9000580/

[246] Shapiro, H. M. Microbial analysis at the single-cell level: tasks and techniques. *J Microbiol Methods* **42**, 3-16 (2000). https://doi.org:https://doi.org/10.1016/S0167-7012(00)00167-6

[247] Lloyd, D. *Flow Cytometry in Microbiology*. (Springer London, 2013).

[248] Batani, G., Bayer, K., Boge, J., Hentschel, U. & Thomas, T. Fluorescence in situ hybridization (FISH) and cell sorting of living bacteria. *Sci Rep* **9**, 18618 (2019). https://doi.org/10.1038/s41598-019-55049-2

[249] Marcos-Fernandez, R., Sanchez, B., Ruiz, L. & Margolles, A. Convergence of flow cytometry and bacteriology. Current and future applications: a focus on food and clinical microbiology. *Crit Rev Microbiol*, 1-22 (2022). https://doi.org/10.1080/1040841X.2022.2086035

[250] Marie, D., Brussaard, C. P. D., Thyrhaug, R., Bratbak, G. & Vaulot, D. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* **65**, 45-52 (1999). https://doi.org/10.1128/aem.65.1.45-52.1999

[251] Simard, O. *et al.* Step-specific Sorting of Mouse Spermatids by Flow Cytometry. *J Vis Exp*, e53379 (2015). https://doi.org/10.3791/53379

[252] Veldhuis, M. & Kraay, G. Application of flow cytometry in marine phytoplankton research: Current applications and future perspectives. *Sci Mar* **64**, 121-134 (2000). https://doi.org/10.3989/scimar.2000.64n2121

[253] Poulopoulos, A. *et al.* Subcellular transcriptomes and proteomes of developing axon projections in the cerebral cortex. *Nature* **565**, 356-360 (2019). https://doi.org/10.1038/s41586-018-0847-y

[254] Stefan, E., Hofmann, S. & Tampé, R. A single power stroke by ATP binding drives substrate translocation in a heterodimeric ABC transporter. *eLife* **9** (2020). https://doi.org/10.7554/elife.55943

[255] Givan, A. L., Hawley, T. S. & Hawley, R. G. Flow Cytometry: An Introduction. *Methods Mol Biol* **699**, 1-29 (2011). https://doi.org/10.1007/978-1-61737-950-5_1

[256] Luquet, E., Biesemann, C., Munier, A. & Herzog, E. Purification of Synaptosome Populations Using Fluorescence-Activated Synaptosome Sorting. *Methods Mol Biol* **1538**, 121-134 (2017). https://doi.org/10.1007/978-1-4939-6688-2_10

[257] Julius, M. H., Masuda, T. & Herzenberg, L. A. Demonstration That Antigen-Binding Cells Are Precursors of Antibody-Producing Cells After Purification with a Fluorescence-Activated Cell Sorter. *Proc Natl Acad Sci U S A* **69**, 1934-1938 (1972). https://doi.org:10.1073/pnas.69.7.1934

[258] Battye, F. L., Light, A. & Tarlinton, D. M. Single cell sorting and cloning. *J Immunol Methods* **243**, 25-32 (2000). https://doi.org:https://doi.org/10.1016/S0022-1759(00)00225-8

[259] Howarth, M., Williams, A., Tolstrup, A. B. & Elliott, T. Tapasin enhances MHC class I peptide presentation according to peptide half-life. *Proc Natl Acad Sci U S A* **101**, 11737-11742 (2004). https://doi.org:doi:10.1073/pnas.0306294101

[260] Matschulla, T. *et al.* A highly conserved sequence of the viral TAP inhibitor ICP47 is required for freezing of the peptide transport cycle. *Sci Rep* **7**, 2933 (2017). https://doi.org:10.1038/s41598-017-02994-5

[261] Seidel, E. *et al.* A slowly cleaved viral signal peptide acts as a protein-integral immune evasion domain. *Nat Commun* **12** (2021). https://doi.org:10.1038/s41467-021-21983-x

[262] Sabatini, D. D. Preparation of crude rough microsomes from tissue culture cells. *Cold Spring Harb Protoc* **2014**, 980-987 (2014). https://doi.org:10.1101/pdb.prot079996

[263] Kirchhofer, A. *et al.* Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* **17**, 133-138 (2010). https://doi.org:10.1038/nsmb.1727

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

[265] Dana, A. & Tuller, T. Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. *PLoS Comput Biol* **8**, e1002755 (2012). https://doi.org:10.1371/journal.pcbi.1002755

[266] Riba, A. *et al.* Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc Natl Acad Sci U S A* **116**, 15023-15032 (2019). https://doi.org:10.1073/pnas.1817299116

[267] Donnelly, M. L. L. *et al.* Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J Gen Virol* **82**, 1013-1025 (2001). https://doi.org:10.1099/0022-1317-82-5-1013

[268] Wang, N., Seko, A., Takeda, Y. & Ito, Y. Glycan dependent refolding activity of ER glucosyltransferase (UGGT). *Biochim Biophys Acta Gen Subj* **1864**, 129709 (2020). https://doi.org:10.1016/j.bbagen.2020.129709

[269] Wearsch, P. A., Peaper, D. R. & Cresswell, P. Essential glycan-dependent interactions optimize MHC class I peptide loading. *Proc Natl Acad Sci U S A* **108**, 4950-4955 (2011). https://doi.org:10.1073/pnas.1102524108

[270] Hachisu, M. & Ito, Y. Chemical Approaches to Elucidate Enzymatic Profiles of UDP-Glucose: Glycoprotein Glucosyltransferase. *Chem Pharm Bull* **64**, 687-690 (2016). https://doi.org:10.1248/cpb.c16-00209

[271] Adams, B. M., Canniff, N. P., Guay, K. P., Larsen, I. S. B. & Hebert, D. N. Quantitative glycoproteomics reveals cellular substrate selectivity of the ER protein quality control sensors UGGT1 and UGGT2. *eLife* **9** (2020). https://doi.org:10.7554/eLife.63997

6.   Publications

Hereinafter, the reader will encounter the original documents associated with the previously summarized publications, as listed below:

1. Semisynthetic viral inhibitor for light control of the MHC I peptide loading complex
    C. Winter‡, A. Domnick‡, D. Cernova, R. Tampé
    Angewandte Chemie **2022**, e202211826

2. Molecular basis of MHC I quality control in the peptide loading complex
    A. Domnick‡, **C. Winter‡**, L. Sušac, L. Hennecke, M. Hensen, N. Zitzmann, S. Trowitzsch, C. Thomas, R. Tampé
    Nature Communications **2022**, 13, 4701

3. Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity
    I. Müller, **C. Winter**, C. Thomas, R. Spaapen, S. Trowitzsch, R. Tampé
    Nature Communications **2022**, 13, 5383

4. Light-guided intrabodies for on-demand in situ target recognition in human cells
    E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
    Chemical Science **2021**, 12, 5787-5795

5. Efficient amber suppression via ribosomal skipping for in situ synthesis of photoconditional nanobodies
    E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
    ACS Synthetic Biology **2022**, 11, 1466-1476 b

Since this manuscript has not yet been published, the following original document cannot be included here in order to avoid complications with the submission and potential copyright infringements:

6. TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I
    L. Sagert, **C. Winter**, M. Zehetmaier, I. Rupert, C. Thomas, R. Tampé
    manuscript in preparation

‡ Contributed equally

Check for updates

Accepted Article

WILEY-VCH

RESEARCH ARTICLE

Semisynthetic Viral Inhibitor for Light Control of the MHC I Peptide Loading Complex

Christian Winter[a],#, Alexander Domnick[a],#, Darja Cernova[a], Robert Tampé[a]*

[a]   Institute of Biochemistry, Biocenter, Goethe University Frankfurt
        Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
        Prof. Dr. Robert Tampé, Corresponding Author
        E-mail: tampe@em.uni-frankfurt.de

        Supporting information for this article is given via a link at the end of the document.
        #Contributed equally, *To whom correspondence may be addressed

Abstract: The immune system detects virally or malignantly transformed cells via peptide-loaded major histocompatibility complex class I (pMHC I) molecules on the cell surface. MHC I molecules are loaded with cargo peptides in the endoplasmic reticulum (ER) by the highly dynamic multiprotein peptide loading complex (PLC). Here, we developed a semisynthetic approach to generate a photocleavable immune modulator ICP47 of Herpes simplex virus. Using this nanotool, we revealed key mechanistic events of the purified PLC, such as peptide binding and translocation coupled to ATP hydrolysis, triggered by light. We established a single-organelle flow cytometry assay to monitor light-controlled activation of the antigen processing machinery in native ER membranes. This photochemical modulation opens new opportunities for a comprehensive mechanistic analysis of the antigen processing machinery in vitro and native membrane environment.

Introduction

The identification and elimination of diseased cells is a major task of the adaptive immune system. It is based on the recognition of pMHC I molecules on the cell surface, presenting their antigenic cargo to cytotoxic T cells.[1] MHC I sample peptides within the context of the macromolecular peptide loading complex (PLC) in the ER membrane.[2] Several chaperones, comprising tapasin, ERp57, and calreticulin, coordinate MHC I peptide loading centered on the ATP-binding cassette (ABC) transporter associated with antigen presentation (TAP).[3] As major regulator of antigen processing and presentation, the PLC is a prime target for viral immune evasions.[4] Infected cell protein 47 (ICP47) encoded by Herpes simplex virus is a soluble immune modulator, which inhibits the antigen translocation by blocking peptide binding[5] and ATP hydrolysis of TAP.[6] The 88 amino acid (aa) ICP47 comprises an active N-terminal region (residues 2-34)[7] followed by a stabilizing region (residues 35-88).[8] A flexible loop region (residues 13-15) between the two α-helices was identified as an optimal position for insertion of photocleavable β-amino acids (Figure 1A).[9] Light is an orthogonal and contactless trigger to modulate molecular and cellular processes with minimal side effects. While a 54 aa synthetic photoconditional (pc-)ICP47$_{2-55}$ allowed for spatiotemporal control of the PLC in vitro[9], it cannot be used as a bait to isolate the native PLC from human cells for mechanistic and structural studies.[10] On the contrary, previous isolation strategies based on full-length viral inhibitors irreversibly inactivated the purified PLC, prohibiting downstream activity assays[3, 10-11].

Here, we addressed this crucial aspect by engineering a full-length ICP47 with a C-terminal flexible glycine-serine linker and a streptavidin-binding peptide (SBP) tag for affinity purification. Linker and affinity tag increase the solubility of the hydrophobic full-length ICP47, leading to an overall length of 140 aa for pc-ICP47^{SBP}. While peptides with a length of up to 50 residues are frequently synthesized by solid phase peptide synthesis (SPPS), the yield of longer polypeptides decreases sharply and is accompanied by an increasing number of side products.[12] Site-specific incorporation of photocaged amino acids has been accomplished in a cellular context via amber suppression.[13] However, the incorporation of backbone-cleaving β-amino acids has not been achieved so far.

In this study, we therefore aimed at establishing a semisynthetic approach to generate pc-ICP47^{SBP} from two fragments with a subsequent desulfurization[14] to achieve a traceless ligation. The short, synthetic fragment carries the photocleavable amino acid Anp (3-amino-3-(2-nitrophenyl)-propanoic acid), while the long recombinant fragment provides solubility and the SBP affinity tag. Only a minority of NCL approaches available to date[15] is compatible with hydrophobic peptides, such as the N-terminal fragment of pc-ICP47^{SBP}. The ligation method used in our study[16] relies on an O to S acyl shift by in situ generation of a thioester moiety, which reacts with a peptide fragment, containing an N-terminal cysteine.[17]

Over the past decades, flow cytometry has evolved into one of the most widely used methods in the fields of biotechnology and biomedicine. To quantify cell markers and ligand binding, a wide range of staining and analytical methods are available. Recently, applications have been extended to analyze particles much smaller than mammalian cells (~10 μm), ranging from bacteria (~2 μm)[18] to subcellular vesicles such as synaptosomes (~400 nm).[19] The latter are visualized by membrane-intercalating dyes.
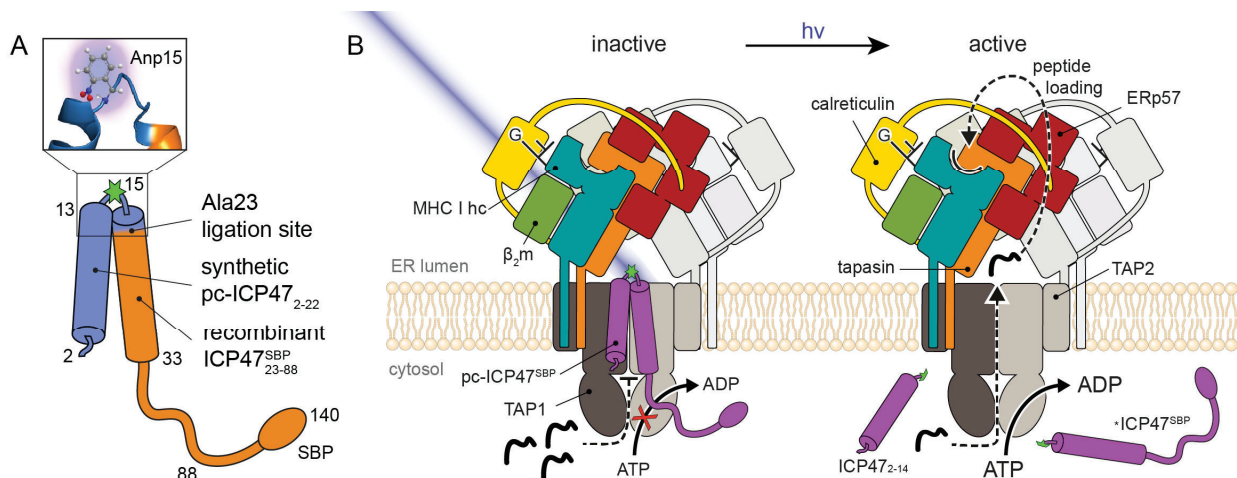
Figure 1. A) Rational design of semisynthetic photoconditional (pc-)ICP47. Residues 2-33 form a helix-loop-helix motif and the active binding region to the TAP complex. The N-terminal region of the semisynthetic inhibitor (blue) can be split by light at the photocleavable β-amino acid Anp. The recombinant part (aa 23-140) of the full-length pc-ICP47SBP harbors a C-terminal streptavidin-binding peptide (SBP) tag. Both parts are fused by native chemical ligation. Anp has been modeled into the 3D structure of the loop region for visualization (zoom-in, PDB: 5U1D).[20] B) The peptide loading complex in the ER membrane. Two editing modules are assembled around the TAP complex. If the PLC is arrested by pc-ICP47SBP (purple), peptide binding and translocation is blocked. Upon illumination and photorelease of ICP47₂₋₁₄ and *ICP47SBP, the PLC is activated.

Here, we established a flow cytometric approach on human microsomes (~150 nm) without the requirement of membrane staining, which permits the unrestricted usage of fluorophores and staining procedures (Figure S1). We utilized the semisynthetic full-length pc-ICP47SBP to isolate and to light control the native PLC in the ER membrane of human Burkitt's lymphoma cells. After photocleavage and release of pc-ICP47SBP, the PLC becomes active in peptide binding, peptide-coupled ATP hydrolysis, and peptide transport. The fully active PLC is a prerequisite for a comprehensive mechanistic and structural analysis of the antigen processing machinery.

Results and Discussion

Design of a semisynthetic photocleavable viral immune inhibitor controlling the PLC

Screening ideal positions for photocleavable backbone modifications in the active region of ICP47 revealed Met15 being the optimal site preserving full TAP inhibition.[9] Since synthesis of full-length ICP47 with an unnatural amino acid at Met15 is not accessible via SPPS in acceptable yields, we produced pc-ICP47SBP in two fragments for a native chemical ligation and subsequent desulfurization (Figure 1A). The N-terminal pc-ICP47₂₋₂₂-Hmp (ICP47₂₋₂₂15Anp-Hmp, Table S1) contained a SPPS compatible 2-hydroxy-3-mercaptopropionic acid (Hmp) moiety for in situ thioester generation and subsequent ligation to the C-terminal fragment of pc-ICP47SBP. The C-terminal CysICP47SBP (Cys-ICP47₂₄₋₈₈-G₄SGS(G₃S)₂-SBP) was recombinantly produced in *Escherichia coli* with an N-terminal SUMO tag to yield an N-terminal cysteine after proteolytic cleavage. N- and C-terminal fragments were ligated to pc-ICP47SBP (ICP47₂₋₈₈15Anp-linker-SBP). pc-ICP47SBP was designed to inhibit the PLC, blocking peptide translocation into the ER and subsequent peptide loading of MHC I (Figure 1B). In this arrested

state, the native PLC can be isolated by SBP affinity purification.[3] We hypothesized that upon UV illumination, the PLC is activated, restoring peptide binding, ATP hydrolysis, and transport.

Native chemical ligation, desulfurization, and photo-conversion of ICP47

Owing to its high hydrophobicity, the synthetic pc-ICP47₂₋₂₂-Hmp (1.2 mM) was dissolved in guanidine hydrochloride (5 M, GdmCl), 30% (v/v) hexafluoroisopropanol (HFIP), and TCEP (150 mM) in 200 mM Na₂HPO₄, pH 7.05 (Figure 2A). At pH 7.0, Hmp undergoes an O to S acyl shift, forming a thioester.[17] This rearrangement drives the 4-mercaptophenyl-acetic acid (MPAA)-catalyzed transthioesterification of pc-ICP47₂₋₂₂-Hmp to CysICP47SBP (1 mM), by which both fragments are joined together via the thiol sidechain of CysICP47SBP. In a subsequent S to N acyl shift, the amide backbone is restored. With an efficiency of ~90%, the ligation reaction completed within 5 h, while only a minor fraction of the thioester was hydrolyzed during the ligation (Figure 2B). Since MPAA interferes with the subsequent desulfurization, we removed MPAA below the detection limit by a rapid-spin size-exclusion chromatography (SEC). Native chemical ligation and desulfurization was therefore performed in a "one-pot" setup with a similar buffer system, containing 5 M GdmCl and 150 mM TCEP in 200 mM Na₂HPO₄, pH 7.0. The reaction was initiated by adding 15% (v/v) 2-methylpropane-2-thiol (tBuSH) and 10 mM 2,2'-azobis-[2-(2-imidazolin-2-yl)propane] dihydrochloride (VA-044). While free radical-mediated desulfurization is well described[14], its application to peptides containing the photocleavable o-nitrophenyl-based moiety is limited.[21] Since illumination of the o-nitrophenyl produces a radical intermediate in a type II Norrish reaction[22], backbone cleavage can be triggered by free radicals during desulfurization, which could drastically limit the yield in desulfurization reaction. As protection of the o-nitrophenyl from radicals[21] inhibited the reaction, we optimized the conditions towards a quantitative desulfurization.
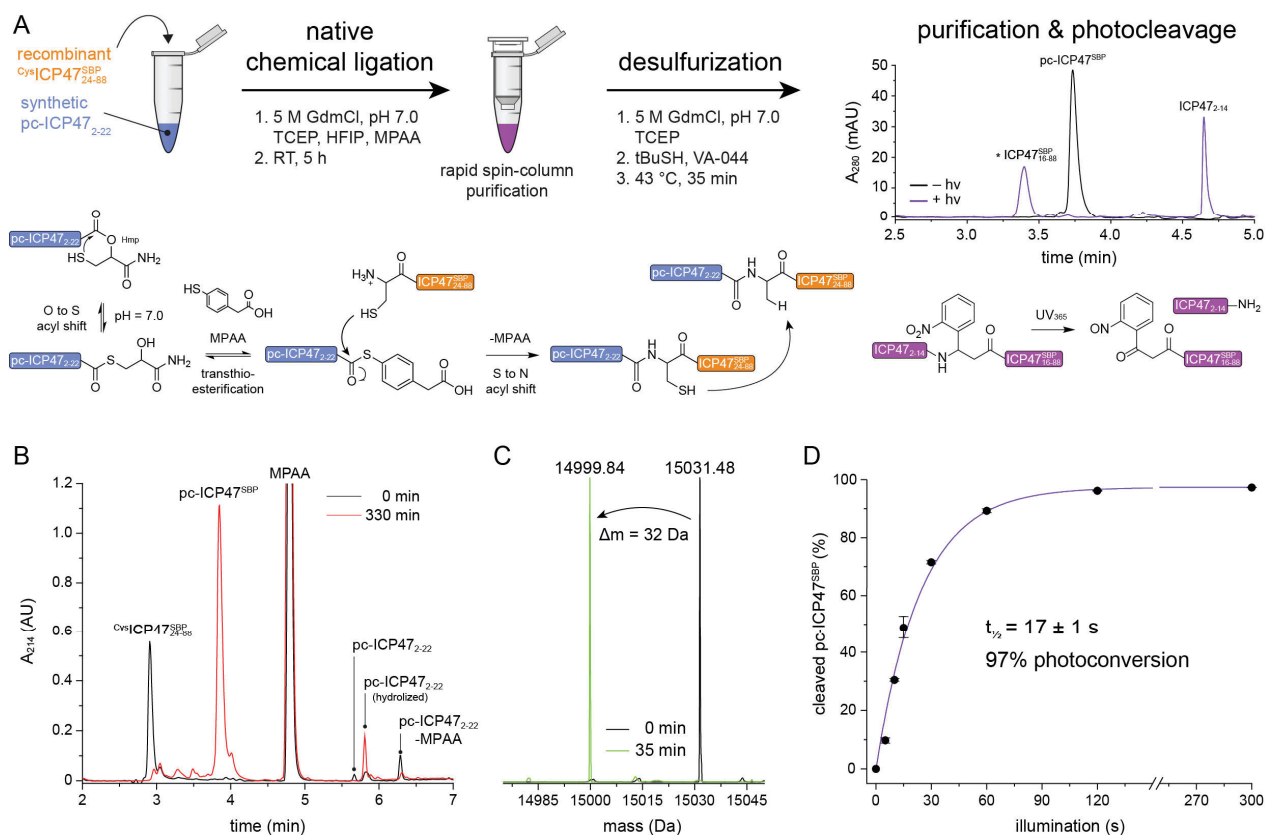
RESEARCH ARTICLE

Figure 2. A) Native chemical ligation and desulfurization strategy. A pH-induced S to O acyl shift converts the Hmp moiety of pc-ICP47$_{2-22}$ into a thioester, which is receptive for MPAA substitution. The thiol of ^{Cys}ICP47^{SBP} replaces MPAA, ligating the two fragments. Excess MPAA is removed by rapid-spin SEC. Free radical-based desulfurization is carried out at elevated temperatures and short reaction times to protect the photocleavable Anp. Pc-ICP47^{SBP} is obtained in high purity and cleaved into two fragments upon UV$_{365}$ illumination. B) NCL progress monitored by reverse-phase C$_{18}$ ultra performance liquid chromatography (UPLC). Pc-ICP47^{SBP} is formed from pc-ICP47$_{2-22}$-MPAA and ^{Cys}ICP47^{SBP} within 330 min. C) Desulfurization of pc-ICP47^{SBP} analyzed by LC-MS. Cys23 is quantitatively desulfurized to alanine as apparent by a mass shift of 32 Da (light green). D) Photocleavage of pc-ICP47^{SBP} analyzed by C$_{18}$-UPLC. Quantitative cleavage is reached after 120 s with $t_{1/2}$ = 17±1 s, yielding a maximal photoconversion of 97%, n=2.

After 35 min at 43 °C, full desulfurization was achieved as detected by LC-MS (Figure 2C), while 85% of the photoconditional peptide was preserved. After synthesis, pc-ICP47^{SBP} was isolated in high purity (Figure 2A).

We next investigated the photocleavage properties of pc-ICP47^{SBP} since a fast and complete conversion of the nanotool is essential for mechanistic studies. Upon UV$_{365}$ illumination (365 nm, 50 mW), a clean and complete cleavage of pc-ICP47^{SBP} into two fragments was observed by LC-MS. We observed the ensemble photoconversion with a half-life of $t_{1/2}$ = 17 ± 1 s in our illumination setup, whereby pc-ICP47^{SBP} was quantitatively cleaved after 2 min (Figure 2D).

Optical control of PLC activation

One of the key features of pc-ICP47^{SBP} is the ability to arrest and stabilize the native PLC for isolation and subsequent mechanistic and structural studies. We purified the detergent-solubilized human PLC from Burkitt's lymphoma cells using pc-ICP47^{SBP} as bait. The fully assembled PLC was monodisperse as demonstrated by SEC, and the hydrodynamic properties of the PLC did not change upon illumination, indicating that the macromolecular assembly of the PLC is unaffected by the photocleavage of the viral inhibitor (Figure 3A). The homogeneous composition of PLC samples before and after

photocleavage of pc-ICP47^{SBP} was demonstrated by SDS-PAGE (Figure 3B). Photocleavage of pc-ICP47^{SBP} resulted in a polypeptide fragment (*ICP47^{SBP}) with an apparent molecular weight of 17 kDa as shown by immunoblotting (Figure 3C, Figure S2A). Photorelease of PLC-bound pc-ICP47^{SBP} was completed after 180 s of UV$_{365}$ illumination Figure S2A). For optimal photocleavage, a slightly acidic pH was required as a pH above 7.0 diminished backbone cleavage as described.[23] PLC was found to be stable at pH 6.5, while aggregation occurred at pH ≤ 6.2 (Figure S3). Therefore, an illumination for 180 s at pH 6.5 was chosen in all experiments, ensuring both the integrity of the PLC and >90% cleavage efficiency of pc-ICP47^{SBP} (Figure S2B).

Peptide binding and coupled ATP hydrolysis of the PLC controlled by light

After photocleavage, the *ICP47^{SBP} fragment did not comigrate with the PLC as shown by SEC and subsequent immunoblotting (Figure 3C). We probed the unblocked binding site in the PLC (1 µM) by adding 10 µM of Alexa Fluor 647 (AF647)-labeled ICP47^{AF647} (Figure 3D, Figure S4). After photorelease of pc-ICP47^{SBP}, the PLC could bind ICP47^{AF647}, whereas no binding was detected before illumination.
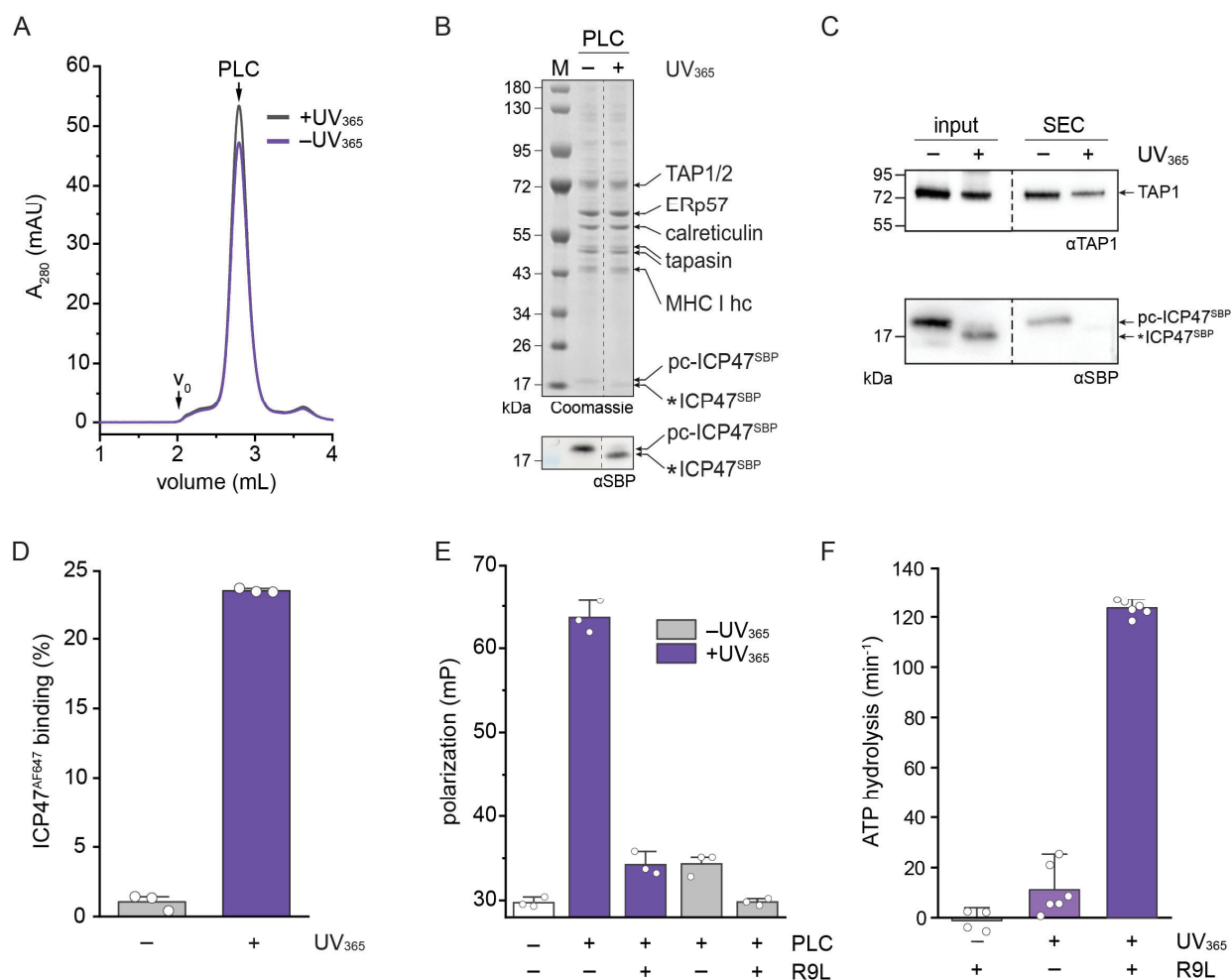
RESEARCH ARTICLE

Figure 3. Photorelease of pc-ICP47SBP and photoactivation of the PLC. A) The native PLC was purified via pc-ICP47SBP and analyzed by SEC. After UV365 illumination, the PLC remained fully assembled as the hydrodynamic shape was not altered (purple). B) The native PLC arrested by pc-ICP47SBP was analyzed by SDS-PAGE (4-12%, Coomassie-stained). After photocleavage of pc-ICP47SBP, the subunit composition of the PLC did not change. After illumination, PLC-bound pc-ICP47SBP was effectively photocleaved (>90%) as demonstrated by immunoblotting (αSBP). *ICP47SBP: cleaved C-terminal fragment of pc-ICP47SBP. C) The composition of the native PLC was analyzed by SEC and subsequent immunoblotting (αSBP) before and after photocleavage of pc-ICP47SBP. pc-ICP47SBP fragments were efficiently released after photocleavage. D) ICP47AF647 rebinding after photocleavage, analyzed by fluorescence-based SEC. After illumination, the TAP-binding pocket in the PLC (1 µM) is free for ICP47AF647 binding (10 µM, purple). n = 3 biological replicates. E) Peptide binding to TAP within the native PLC analyzed by fluorescence anisotropy. The PLC (150 nM final) was added to the fluorescent reporter peptide C4F (50 nM). C4F cannot bind to pc-ICP47SBP-arrested PLC (grey). After UV365 illumination, peptide binding was restored and can be competed with 1000-fold molar excess of unlabeled peptide (R9L, purple). n=3 replicates. F) ATP hydrolysis of the native PLC analyzed by radioactive phosphate release from [γ-32P]ATP. The PLC exhibits peptide-dependent ATP hydrolysis after UV365 illumination (purple). The vanadate-sensitive ATPase activity was plotted. n = 2 or 3, each spotted in duplicates.

Quantitative re-binding was not expected due to the lack of a membrane environment and the concomitant decrease in binding affinity to TAP.[24] We next examined binding of antigenic peptides to the PLC by fluorescence polarization using the fluorescein (F)-labeled peptide RRYC(F)KSTEL (C4F) before and after photorelease of pc-ICP47SBP (Figure 3E, Table S2). This reporter peptide binds with high affinity to human TAP[24b] but not to other components of the PLC.

In line with this notion, illumination triggered peptide binding to TAP while only a minimal binding was observed before photoactivation. Peptide binding was specific as demonstrated by competition to background level using an excess of the unlabeled peptide RRYQKSTEL (R9L). The binding kinetics were fast, reaching an association equilibrium and complete dissociation within 5 min upon peptide addition (Figure S5).

In addition, we detected a peptide-dependent ATP hydrolysis activity after the photocleavage and -release of pc-

ICP47SBP from the PLC (Figure 3F). Since peptide binding and ATP hydrolysis are tightly coupled, these results provide evidence of the peptide translocation activity of the isolated PLC and are consistent with data obtained from purified and reconstituted TAP complexes.[6b, 25] In summary, the native PLC, arrested and purified by pc-ICP47SBP, was activated by light for mechanistic studies.

Photocontrol of peptide transport followed by single-organelle recordings

To study the PLC in a near-native environment of the ER membrane, we developed a flow cytometric assay. We prepared microsomes from Burkitt's lymphoma cells that are tightly sealed and contain PLC in a homogeneous membrane orientation with the nucleotide-binding sites and the peptide-binding pocket of TAP exposed to the outside (Figure S1B).
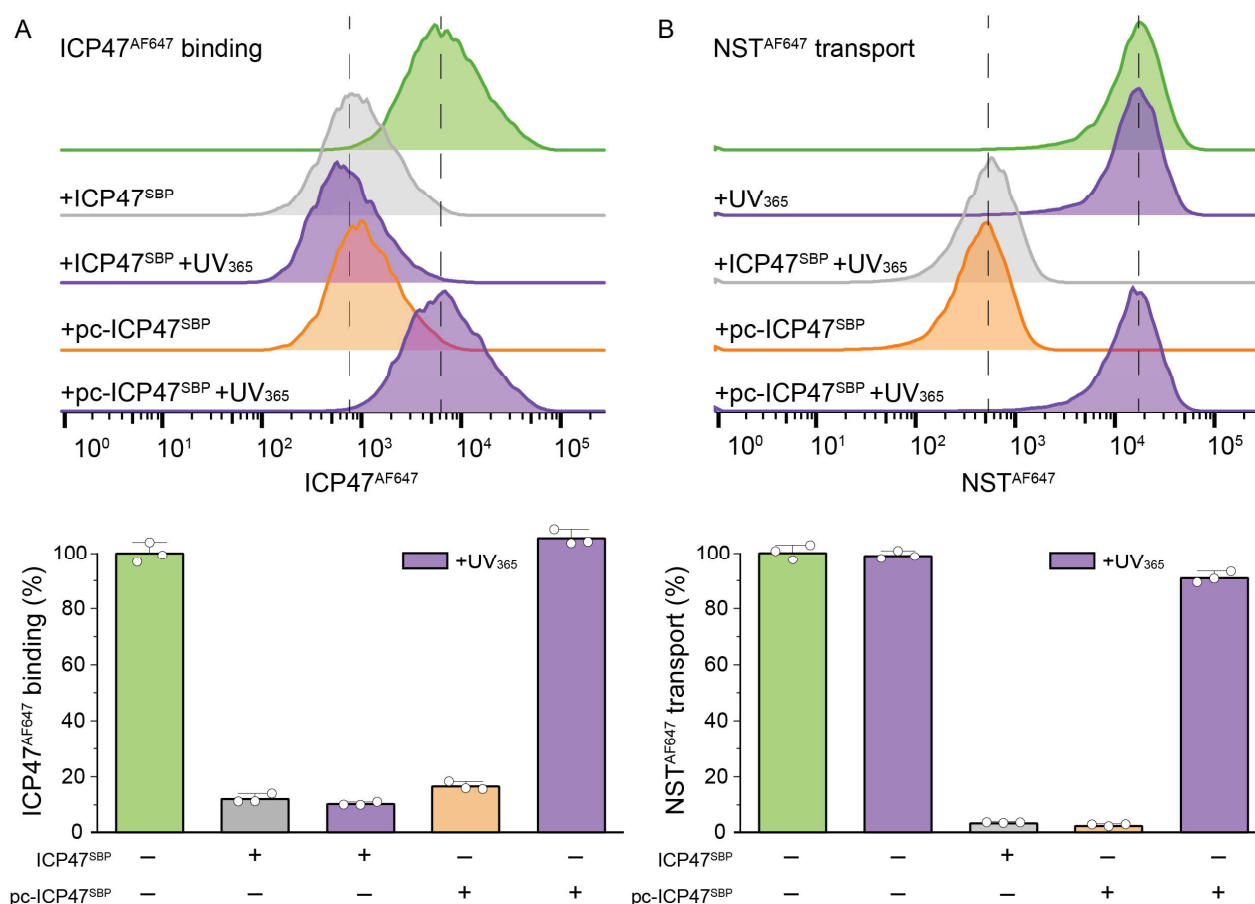
4

Figure 4. ICP47^{AF647} rebinding to PLC and NST^{AF647}peptide transport in microsomes after photocleavage pc-ICP47^{SBP} analyzed by single organelle recordings. A) Microsomes incubated with ICP47^{AF647}, representing 100% bound ICP47^{AF647} (green). Microsomes pre-incubated with ICP47^{SBP} (light grey) did not show any ICP47^{AF647} binding after UV$_{365}$ illumination, while pc-ICP47^{SBP} (orange) quantitatively restored ICP47^{AF647} binding after illumination (n = 3). B) PLC-mediated peptide translocation assay in microsomes analyzed by flow cytometry. Microsomal peptide transport was normalized to 100%. Peptide translocation in microsomes was not affected by UV$_{365}$ illumination. Binding pc-ICP47^{SBP} (10 µM) inhibited peptide translocation to the same extend as ICP47^{SBP} (10 µM). Upon photocleavage of pc-ICP47^{SBP}, PLC-mediated peptide translocation activity was fully restored (n = 3). Experimental setup and gating strategies are visualized in Figure S6.

pc-ICP47^{SBP} was quantitatively displaced from the binding pocket of TAP1/2 after photocleavage, as analyzed by rebinding of fluorescently labeled ICP47^{AF647} (Figure 4A). The non-photocleavable ICP47^{SBP} remained stably bound, preventing rebinding of the fluorescent competitor ICP47^{AF647}.

Next, we examined if the peptide transport properties of the PLC in the ER membrane can be restored upon photocleavage and -release of pc-ICP47^{SBP}. We observed an ATP-dependent transport of the Alexa Fluor 647-labeled peptide RRYNSTC$^{(AF647)}$L (NST^{AF647}) into microsomes, which was blocked to background level by ICP47^{SBP} (Figure 4B). In comparison to ICP47^{AF647} binding, an enhanced fluorescent signal was detected based on accumulation due to multiple rounds of transport. Photocleavage restored the transport properties to >90% of the microsomes containing the unblocked PLC. It is worth mentioning that illumination alone did not affect peptide transport. Furthermore, both proteins, ICP47^{SBP} and pc-ICP47^{SBP}, revealed the same degree of inhibition.

Conclusion

In this study we generated a photoconditional semisynthetic viral inhibitor by native chemical ligation, which allows precise control of the PLC for mechanistic analyses. We recovered the wildtype viral inhibitor and benefited from increased binding stability and enhanced water solubility through a flexible Gly-Ser linker and SBP tag. After photorelease of pc-ICP47^{SBP}, PLC remains stable and fully assembled with all chaperones and MHC I. The synchronously activated, native PLC will open new avenues for functional and structural analyses, which will provide important insights into the inner workings of peptide transport and the MHC I-loading cycle. By flow cytometry of microsomes, we demonstrated that pc-ICP47^{SBP} operates in the native environment of the ER membrane, whereas it was both efficiently removed from the TAP binding pocket as well as competent in reactivating transport of peptide epitopes. Flow cytometric analysis of subcellular organelles can provide analytical access to membrane transporters, channels, and enzymes (Figure S7) with studies benefiting from the high statistical usability and large sample size of the data. This approach might allow the reconstitution of activatable PLC into defined liposomal systems, allowing to analyze the PLC disassembly[10] and clustering by cryogenic electron microscopy and electron tomography. Thus, the photoconditional viral immune evasin is instrumental for both, purification and light-controlled mechanistic studies of arrested and active PLC in native membrane environments.

5

RESEARCH ARTICLE

Data and materials availability

References

[1] a) J. S. Blum, P. A. Wearsch, P. Cresswell, *Annu Rev Immunol* **2013**, *31*, 443-473; b) S. Trowitzsch, R. Tampé, *Annu Rev Biophys* **2020**, *49*, 135-161.

[2] a) C. Thomas, R. Tampé, *Curr Opin Immunol* **2021**, *70*, 48-56; b) N. Pishesha, T. J. Harmand, H. L. Ploegh, *Nat Rev Immunol* **2022**, online ahead of print.

[3] A. Blees, D. Januliene, T. Hofmann, N. Koller, C. Schmidt, S. Trowitzsch, A. Moeller, R. Tampé, *Nature* **2017**, *551*, 525-528.

[4] a) T. H. Hansen, M. Bouvier, *Nat Rev Immunol* **2009**, *9*, 503-513; b) P. U. Mayerhofer, R. Tampé, *J Mol Biol* **2015**, *427*, 1102-1118.

[5] a) K. Früh, K. Ahn, H. Djaballah, P. Sempe, P. M. van Endert, R. Tampé, P. A. Peterson, Y. Yang, *Nature* **1995**, *375*, 415-418; b) A. Hill, P. Jugovic, I. York, G. Russ, J. Bennink, J. Yewdell, H. Ploegh, D. Johnson, *Nature* **1995**, *375*, 411-415; c) R. Tomazin, A. B. Hill, P. Jugovic, I. York, P. van Endert, H. L. Ploegh, D. W. Andrews, D. C. Johnson, *EMBO J* **1996**, *15*, 3256-3266.

[6] a) K. Ahn, T. H. Meyer, S. Uebel, P. Sempe, H. Djaballah, Y. Yang, P. A. Peterson, K. Früh, R. Tampé, *EMBO J* **1996**, *15*, 3247-3255; b) S. Gorbulev, R. Abele, R. Tampé, *Proc Natl Acad Sci U S A* **2001**, *98*, 3732-3737.

[7] L. Neumann, W. Kraas, S. Uebel, G. Jung, R. Tampé, *J Mol Biol* **1997**, *272*, 484-492.

[8] a) B. Galocha, A. Hill, B. C. Barnett, A. Dolan, A. Raimondi, R. F. Cook, J. Brunner, D. J. McGeoch, H. L. Ploegh, *J Exp Med* **1997**, *185*, 1565-1572; b) V. Herbring, A. Baucker, S. Trowitzsch, R. Tampé, *Sci Rep* **2016**, *6*, 36907; c) T. Matschulla, R. Berry, C. Gerke, M. Doring, J. Busch, J. Paijo, U. Kalinke, F. Momburg, H. Hengel, A. Halenius, *Sci Rep* **2017**, *7*, 2933.

[9] M. Braner, N. Koller, J. Knauer, V. Herbring, S. Hank, R. Wieneke, R. Tampé, *Chem Sci* **2019**, *10*, 2001-2005.

[10] A. Domnick, C. Winter, L. Susac, L. Hennecke, M. Hensen, N. Zitzmann, S. Trowitzsch, C. Thomas, R. Tampé, *Nat Commun* **2022**, *13*, 4701.

[11] S. Sethumadhavan, M. Barth, R. M. Spaapen, C. Schmidt, S. Trowitzsch, R. Tampé, *Sci Rep* **2022**, *12*, 1516.

[12] N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, *Science* **2020**, *368*, 980-987.

[13] a) J. W. Chin, *Nature* **2017**, *550*, 53-60; b) N. Ankenbruck, T. Courtney, Y. Naro, A. Deiters, *Angew Chem Int Ed Engl* **2018**, *57*, 2768-2798; c) J. Brunnberg, V. Herbring, E. Gunther Castillo, H. Kruger, R. Wieneke, R. Tampé, *Commun Biol* **2021**, *4*, 430; d) E. F. Joest, C. Winter, J. S. Wesalo, A. Deiters, R. Tampé, *Chem Sci* **2021**, *12*, 5787-5795.

[14] Q. Wan, S. J. Danishefsky, *Angew Chem Int Ed Engl* **2007**, *46*, 9248-9252.

[15] V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoet, J. M. Monbaliu, O. Melnyk, *Chem Rev* **2019**, *119*, 7328-7443.

[16] A. C. Baumruck, D. Tietze, L. K. Steinacker, A. A. Tietze, *Chem Sci* **2018**, *9*, 2365-2375.

[17] F. Liu, J. P. Mayer, *J Org Chem* **2013**, *78*, 9848-9856.

[18] R. Marcos-Fernandez, B. Sanchez, L. Ruiz, A. Margolles, *Crit Rev Microbiol* **2022**, online ahead of print.

[19] E. Luquet, C. Biesemann, A. Munier, E. Herzog, *Methods Mol Biol* **2017**, *1538*, 121-134.

[20] M. L. Oldham, N. Grigorieff, J. Chen, *eLife* **2016**, *5*, e21829.

[21] R. Burai, N. Ait-Bouziad, A. Chiki, H. A. Lashuel, *J Am Chem Soc* **2015**, *137*, 5041-5052.

[22] T. Schmierer, S. Laimgruber, K. Haiser, K. Kiewisch, J. Neugebauer, P. Gilch, *Phys Chem Chem Phys* **2010**, *12*, 15653-15664.

[23] Y. V. Il'ichev, M. A. Schworer, J. Wirz, *J Am Chem Soc* **2004**, *126*, 4581-4595.

[24] a) C. Aisenbrey, C. Sizun, J. Koch, M. Herget, R. Abele, B. Bechinger, R. Tampé, *J Biol Chem* **2006**, *281*, 30365-30372; b) M. Herget, N. Kreissig, C. Kolbe, C. Scholz, R. Tampé, R. Abele, *J Biol Chem* **2009**, *284*, 33740-33749.

[25] N. Grossmann, A. S. Vakkasoglu, S. Hulpke, R. Abele, R. Gaudet, R. Tampé, *Nat Commun* **2014**, *5*, 5419.

Accepted Manuscript

WILEY-VCH

Entry for the Table of Contents
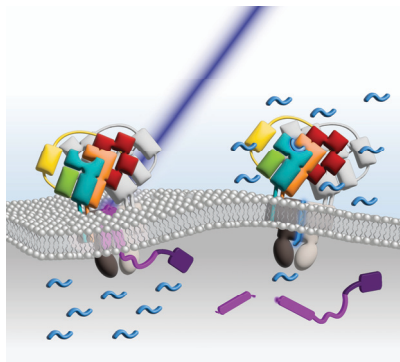
Text for the table of contents:

The immune system supervises every cell via MHC I molecules which are prepared in the peptide loading complex (PLC). Using a light-activated viral nanotool, the authors succeeded in spatiotemporally controlling antigen processing and reconstituting a completely active native PLC. The photo-conditional function of PLC was analyzed with a flow cytometry-based assay at single-organelle resolution.

Institute and/or researcher Twitter usernames: @Tampe_Lab; @SFB1507

Supporting Information
©Wiley-VCH 2021
69451 Weinheim, Germany

Semisynthetic Viral Inhibitors for Light Control of the MHC I Peptide Loading Complex

Christian Winter[a],#, Alexander Domnick[a],#, Darja Cernova[a], Robert Tampé[a]*

[a]   Institute of Biochemistry, Biocenter, Goethe University Frankfurt
        Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
        Prof. Dr. Robert Tampé, Corresponding Author
        E-mail: tampe@em.uni-frankfurt.de

        Supporting information for this article is given via a link at the end of the document.
        #Contributed equally, *To whom correspondence may be addressed

Abstract: The immune system detects virally or malignantly transformed cells via peptide-loaded major histocompatibility complex class I (pMHC I) molecules on the cell surface. MHC I molecules are loaded with cargo peptides in the endoplasmic reticulum (ER) by the highly dynamic multiprotein peptide loading complex (PLC). Here, we developed a semisynthetic approach to generate a photocleavable immune modulator ICP47 of Herpes simplex virus. Using this nanotool, we revealed key mechanistic events of the purified PLC, such as peptide binding and translocation coupled to ATP hydrolysis, triggered by light. We established a single-organelle flow cytometry assay to monitor light-controlled activation of the antigen processing machinery in native ER membranes. This photochemical modulation opens new opportunities for a comprehensive mechanistic analysis of the antigen processing machinery in vitro and native membrane environment.

SUPPORTING INFORMATION

Table of Contents

Experimental Procedures

Synthesis of 3-cholor-2-hydroxypropanoic acid

10 g of 3-chloro-1,2-propanediol (0.091 mol) were dissolved in 30 mL 80% nitric acid and heated to 80 °C. The temperature was raised to 80 °C until the reaction started. The reaction mixture was removed from the heater for 15 min after which the temperature was increased to 100 °C for 2 h. The reaction solution was then cooled to room temperature (RT) and the pH was adjusted to 4.0 with 8.8 g sodium bicarbonate. The product was extracted from the reaction solution with diethyl ether, after which the remaining solvent was evaporated until a viscous liquid was formed. The product was precipitated in cold chloroform, filtered and lyophilized (Lyovac GT2, Heraeus). Yield: 11.0 g (97%) 3-cholor-2-hydroxypropanoic acid. ^{1}H-NMR (acetone-d_6): δ3.88 (2H, CH$^\beta$), δ4.52 (1H CH$^\alpha$).

Synthesis of 2-hydroxy-3-(triphenylmethyl)thio-propanoic acid (Hmp(Trt)-OH)

5.85 g (0.047 mol) 3-cholor-2-hydroxypropanoic acid were dissolved in 47 mL dry dimethoxyethane and cooled to 0 °C on ice. Equimolar amounts of sodium hydride (60% w/w dispersion in mineral oil) were added slowly. After dropwise addition of a solution of 0.05 mol triphenylmethanethiol with 0.047 mol sodium hydride (60% w/w dispersion in mineral oil) in 47 mL dry dimethoxyethane, the mixture was stirred for 4 h. The solvent was evaporated, and the yellow solid was dissolved in 50% (v/v) diethyl ether/H_2O. The aqueous phase was collected, acidified with HCl, and the product extracted with ethyl acetate. The combined organic phase was evaporated until a viscous liquid formed. Residual solvent was removed by lyophilization. To increase purity, the product was recrystallized in 12 mL EtOH. H_2O was added dropwise while heating until the solution became cloudy. The suspension was cooled to RT and subsequently cooled at 4 °C for 30 min. Finally, the white precipitate was slurred with water and filtered. Yield: 11.0 g (97%) 2-hydroxy-3-(triphenylmethyl)thio-propanoic acid. ^{1}H-NMR (chloroform-d): δ7.19-7.37 (15H, Trt), δ2.63 (2H, CH$_2^\beta$), 3.81 (1H, CH$^\alpha$). M_{calc}: 364.1133 Da, M_{obs}: 364.1131 Da.

Synthesis of pc-ICP47$_{2-22}$-Hmp

0.2 mmol of Tentagel S RAM (0.22 mmol/g resin substitution; Rapp Polymere) were pre-swollen in dimethylformamide (DMF). The terminal fluorenylmethoxycarbonyl protecting group (Fmoc) was removed by treatment with 20% (v/v) piperidine in DMF for 15 min, at 22 °C with heavy agitation. Hmp(Trt)-OH was coupled to the resin two times, using a premixed solution of 6 mmol Hmp(Trt)-OH, 5 mmol N,N-diisopropylcarbodiimid (DIC, 0.5 M), and 5 mmol of hydroxybenzotriazole (HOBt, 0.5 M) in DMF for 60 min. The first amino acid (Tyr22) was coupled via Mitsunobu reaction dissolving 2 mmol Fmoc-Tyr(tBu)-OH in 2 mmol triphenylphosphine in 3.8 mL of THF. The solution was added to the resin and cooled to -20 °C. 2 mmol diisopropyl azodicarboxylate (DIAD, 40% (v/v) in toluene) were chilled to -20 °C and added slowly to the resin over a period of 20 min. After 2 h with heavy agitation at RT, the reaction solution was removed. The remaining amino acids (Carbolution, Table S1) were sequentially coupled to the Hmp(Trt)-Tyr(tBu)-loaded resin using standard Fmoc SPPS double coupling conditions on a Liberty Blue Peptide Synthesizer (CEM). (R)-3-(Fmoc-amino)-3-(2-nitrophenyl)propionic acid (Anp, J*, Alfa Aesar) was coupled twice for 30 min at 50 °C. The peptide was protected from light and cleaved in 87.5% trifluoroacetic acid (TFA), 5% thioanisole, 2.5% H_2O, 2.5% triisopropylsilane (TIPS), 2.5% 1,2-ethanethiol (EDT) for 2 h at RT. The product was precipitated in ether at -20 °C, dried over vacuum, and purified by C_{18} RP-HPLC. The identity was verified by LC-MS: pc-ICP47$_{2-22}$-Hmp: M_{calc}: 2636.1788 Da, M_{obs}: 2636.1753 Da.

ULP1 SUMO protease preparation

pFGET19_Ulp1 was a gift from Hideo Iwai (Addgene plasmid # 64697; RRID:Addgene_64697). The vector was transformed into competent *Escherichia coli* One Shot BL21(DE3) cells (Thermo Fisher) and grown to OD$_{600}$=0.6 at 37 °C. Expression was induced with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) for 15 h at 37 °C. The cells were lysed by sonication in Ni-NTA buffer (50 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$ pH 8.0, 150 mM NaCl, 20 mM imidazole, 0.2% Tween 20, 2 mM dithiothreitol (DTT)), supplemented with 2.5 mM phenylmethylsulfonyl fluoride (PMSF, Roth), 3.75 mg/mL lysozyme, 6.25 mM benzamidine, and 1 U/mL benzonase. ^{His6}ULP1 was isolated from the lysate by reverse-IMAC (Ni-NTA Agarose resin, QIAGEN). After washing thrice with Ni-NTA buffer, ^{His6}ULP1 was eluted supplementing the Ni-NTA buffer with 500 mM imidazole. The imidazole concentration was lowered to 20 mM by repetitive ultrafiltration (Amicon Ultra-15, 3 kDa NMWCO, Merck Millipore). ^{His6}ULP1 was supplemented with 15% (v/v) glycerol, aliquoted, snap frozen and stored at -80°C until use.

^{Cys}ICP47^{SBP} preparation

The ^{His6}SUMO-^{Cys}ICP47^{SBP} sequence (Table S1) was cloned into a pETM-11 (European Molecular Biology Laboratory, EMBL) using the SapI restriction site. The vector was transformed into competent *Escherichia coli* One Shot BL21(DE3) cells and grown to OD$_{600}$=0.7 at 37 °C. Expression was induced with 0.2 mM IPTG for 4 h at 22 °C. The cells were lysed by sonication in Ni-NTA buffer, supplemented with 2.5 mM PMSF, 3.75 mg/mL lysozyme, 6.25 mM benzamidine, and 1 U/mL benzonase. ^{His6}SUMO-^{Cys}ICP47^{SBP} was isolated from the lysate by reverse-IMAC. After washing thrice with Ni-NTA buffer, ^{His6}SUMO-^{Cys}ICP47^{SBP} was eluted supplementing the wash buffer with 500 mM imidazole. The imidazole concentration was lowered to 20 mM by repetitive ultrafiltration (Amicon Ultra-15, 3 kDa NMWCO). The ^{His6}SUMO-tag was removed by overnight ULP1 protease digestion. ^{Cys}ICP47^{SBP} was purified by reversed phase C_{18} HPLC (Agilent, 1200 Series System; PerfectSil 300 ODS C_{18} 5 µm 300x10 mm) applying a linear water/acetonitrile gradient from 5-50% (v/v) supplemented with 0.1% (v/v) TFA. Degassed buffers were always used to prevent oxidation. The identity of ^{Cys}ICP47^{SBP} was confirmed by LC-MS analysis. M_{calc}: 12514.57 Da, M_{obs}: 12514.61 Da.

ICP47^{SBP} preparation

The ICP47^{SBP} sequence was cloned into a pETM-11 using the NcoI and BamHI restriction sites. The vector was transformed into competent *Escherichia coli* One Shot BL21(DE3) cells and grown to OD$_{600}$=0.6 at 37 °C. Expression was induced with 0.2 mM IPTG for 6 h at 22 °C. The cells were lysed by sonication in Ni-NTA buffer, supplemented with 2.5 mM PMSF, 3.75 mg/mL lysozyme, 6.25 mM benzamidine, and 1 U/mL benzonase. ^{His6}TEV-ICP47^{SBP} was isolated from the lysate by reverse IMAC. After washing thrice with Ni-NTA buffer, ICP47^{SBP} was eluted supplementing the Ni-NTA buffer with 500 mM imidazole. The His$_6$-tag was removed by overnight TEV protease digestion. ICP47^{SBP} was purified by reversed phase C$_{18}$ HPLC applying a linear water/acetonitrile gradient from 5-60% (v/v) supplemented with 0.1% (v/v) TFA. The identity of ICP47^{SBP} was confirmed by LC-MS analysis. M$_{calc}$: 14694.06 Da, M$_{obs}$: 14694.68 Da.

ICP47^{AF647} preparation

ICP47$^{V86C-SBP}$ was produced in the same way as ICP47^{SBP}. ICP47$^{V86C-SBP}$ was labeled with 1.2 molar excess of Alexa Fluor 647 C$_2$ Maleimide (AF647, ThermoFisher) in phosphate-buffered saline (PBS) pH 7.2 with 20% DMF and 2.5 mM tris(2-carboxyethyl)phosphin (TCEP) for 4 h at RT. ICP47^{AF647} was purified by reversed phase C$_{18}$ HPLC applying a linear water/acetonitrile gradient from 5-60% (v/v) supplemented with 0.1% (v/v) TFA. The identity of ICP47^{AF647} was confirmed by LC-MS analysis. M$_{calc}$: 15938.55 Da, M$_{obs}$: 15938.77 Da.

Native chemical ligation of pc-ICP47$_{2-22}$-Hmp and ^{Cys}ICP47^{SBP}

1.0 molar equivalents (eq) of ^{Cys}ICP47^{SBP} (1 mM) were ligated to 1.2 eq pc-ICP47$_{2-22}$-Hmp. pc-ICP47$_{2-22}$-Hmp was dissolved in a ligation solution containing 5 M guanidine hydrochloride (GdmCl), 150 mM 4-mercaptophenyl-acetic acid (MPAA), 150 mM TCEP, 200 mM Na$_2$HPO$_4$ pH 7.0 and 20% (v/v) hexafluoroisopropanol (HFIP) were added to the solution. The pH was adjusted before dissolving pc-ICP47$_{2-22}$-Hmp, since high local concentrations of NaOH can destroy Hmp. The solution was flushed with argon and added to ^{Cys}ICP47^{SBP}. The ligation reaction was incubated for 5 h and monitored by LC-MS by acquiring 1 µL aliquots of the reaction mixture in an argon chamber. The samples were diluted with 19 µL 200 mM TCEP in H$_2$O + 0.1% (v/v) formic acid (FA), pH 2.0. pc-ICP47$^{23Cys-SBP}$: M$_{calc}$: 15031.31 Da, M$_{obs}$: 15031.58 Da.

Desulfurization of pc-ICP47$^{23Cys-SBP}$

The ligation cocktail containing pc-ICP47$^{23Cys-SBP}$ was prepared for desulfurization by removal of excess MPAA by rapid-spin size exclusion chromatography purification, using a Zeba Spin Desalting Column (7 K MWCO, Thermo Scientific), equilibrated with desulfurization buffer (150 mM TCEP in 5 M GdmCl, pH 7.05). The solution was flushed with argon and 15% (v/v) 2-methylpropane-2-thiol (tBuSH) were added. The reaction was started by addition of 10 mM 2,2'-azobis-[2-(2-imidazolin-2-yl)propane]dihydro-chloride (VA-044). Complete desulfurization was achieved after 35 min at 42 °C. The exact matching of reaction time and temperature is essential for quantitative desulfurization, since the radical desulfurization mechanism can also destroy the photolabile orthonitrobenzene on Anp. The reaction was monitored by LC-MS and purified by C$_{18}$ reverse phase HPLC, applying a linear water/acetonitrile gradient from 5-50% (v/v) supplemented with 0.1% (v/v) TFA. pc-ICP47^{SBP} yield: 85%; M$_{calc}$: 14999.24 Da, M$_{obs}$: 14999.37 Da.

LC-MS analysis

LC-MS measurements were performed on a BioAccord System (Waters) running Unify 1.9.4 (Waters). Peptides and small molecules were analyzed with an ACQUITY UPLC Peptide BEH C$_{18}$ Column, 130 Å, 1.7 µm, 2.1 mm x 100 mm (Waters), applying a linear water/acetonitrile gradient supplemented with 0.1% (v/v) FA at 60 °C, 30 V cone voltage, 0.8 kV capillary voltage, and 550 °C desolvation temperature. Mass spectra were recorded in positive polarity (peptides) or negative polarity (Hmp(Trt)-OH) at 5 Hz in MS^e mode at 50-2000 m/z. Masses of peptides and proteins were calculated and confirmed in Unify. Spectra of ICP47^{SBP}, pc-ICP47^{SBP} and ^{Cys}ICP47^{SBP} were deconvoluted in Unify using the integrated MaxEnt1 algorithm iterating to convergence with inclusion of the isotopic width. Deconvoluted spectra were centroidized based on peak height. Centroidized spectra were used for mass calculations. UV spectra were recorded at 214 nm and 280 nm with 2 Hz and plotted in OriginPro 2022 (OriginLabs).

pc-ICP47^{SBP} cleavage by UV illumination

UV illumination was performed in a self-assembled illumination setup. Samples were transferred into a 1.5 mL tube and placed in a chilled aluminum block on ice. A λ=365 nm diode (M365L2, ThorLabs) with attached adjustable SM2 lens tube (ThorLabs) was placed 1 cm above the opened tube and focused onto the bottom of the tube. The LED was powered and controlled by a DC2200 (ThorLabs). Samples were illuminated for 180 s, 25% lamp intensity (~50 mW) and analyzed by LC-MS. For analysis of cleavage kinetics, the peak height of MaxEnt1 deconvoluted and centered spectra were normalized to the 0 s sample. Data points were fitted in OriginPro 2022b (OriginLabs) using the ExpDecay1 function.

PLC purification

The native PLC was isolated from Burkitt's lymphoma cells (Raji ATCC CCL-86), cultured in RPMI 1640 medium (Gibco), supplemented with 10% Fetal Calf Serum (FCS, Capricorn), 3 mM HEPES-NaOH pH 7.5 (Gibco) at 37 °C and 8% CO$_2$ in a shaking incubator (Eppendorf). Cells were harvested by centrifugation, snap frozen in liquid nitrogen, and stored at -80 °C until use. Cell pellets were thawed and resuspended in 20 mM HEPES-NaOH pH 7.4,

150 mM NaCl, 10 mM MgCl$_2$, protease inhibitor mix (Serva) and incubated with either ICP47^{SBP}, pc-ICP47^{SBP} or ICP47^{AF647} for 15 min at 4 °C. Membranes were mixed with 2% (w/v) glyco-diosgenin (GDN, Anatrace) by douncing and 1 U/mL Benzonase was added. The suspension was incubated for 2 h at 4 °C under agitation. Insoluble material was removed by centrifugation (45 min, 100,000 x g). (pc-)ICP47^{SBP}-arrested PLC was bound to Streptavidin High-Capacity Agarose (Pierce) and washed extensively with buffer at pH 7.4, which was afterwards on-column exchanged to 20 mM MES pH6.8, 6.5 or 6.2, 150 mM NaCl, 0.05% GDN. PLC was eluted in 20 mM Hepes pH 7.4 or 20 mM MES pH6.8, 6.5 or 6.2, 150 mM NaCl, 0.05% (w/v) GDN, 2.5 mM biotin.

Biochemical analysis of the PLC

For SDS-PAGE, precast NuPAGE 4-12% gradient gels (Novex) were used. Gels were either stained by Instant Blue (Expedeon) or directly transferred to a PVDF membrane (Bio-Rad). The membranes were probed using the antibodies, anti-TAP1 (mAb 148.3[1], hybridoma supernatant dilution 1:20, produced in-house), anti-TAP2 (mAb 438.3[2], hybridoma supernatant dilution 1:20, produced in-house), anti-ERp57 (Abcam, catalogue number ab10287, dilution 1:2000).[2] (pc-)ICP47^{SBP} and *ICP47^{SBP} were detected by the anti-SBP antibody (Santa Cruz, catalogue number sc-101595, dilution 1:500). Integrity of the PLC was verified by SEC. SEC analysis was performed at 4 °C using a Shimadzu HPLC system, equipped with a semi-micro KW404-4F column (4.6×300 mm, Shodex) with the running buffer containing 20 mM MES pH 6.5, 150 mM NaCl 0.01% (w/v) GDN.

PLC pH tolerance test

PLC was purified as described above. After binding PLC to streptavidin agarose beads, the buffer was exchanged to either 20 mM HEPES pH 7.4, 150 mM NaCl, 0.05% (w/v) GDN or 20 mM MES buffer pH 6.8, 6.5, or 6.2, 150 mM NaCl, 0.05% (w/v) GDN. PLC stability was analyzed by SEC and the composition was verified by SDS-PAGE.

Analysis of PLC bound pc-ICP47^{SBP} cleavage kinetics

PLC purified via pc-ICP47^{SBP} has been used for this assay. Purified PLC was illuminated for 0, 15, 30, 60, and 180 s, respectively as described in the section for pc-ICP47^{SBP} cleavage by UV illumination. The pc-ICP47^{SBP} cleavage was analyzed by SDS-PAGE and anti-SBP immunoblotting. The anti-SBP immunoblot has been used for band quantification in ImageJ 1.53K. Data points were fitted in OriginPro 2022b (OriginLabs) using the ExpDecay1 function.

ICP47^{AF647} rebinding in the native PLC

The native PLC (1 µM final) purified with pc-ICP47^{SBP} or ICP47^{SBP} was treated with standard UV illumination conditions and supplemented with 10 mM ATP, 15 mM MgCl$_2$ in 150 mM NaCl, 0.05% (w/v) GDN, 20 mM MES pH 6.5. After 10 min at 32 °C, 10 µM of ICP47^{AF647} were added and incubated for 20 min at 4 °C. ICP47^{AF647} rebinding was verified by SEC comigration with the running buffer containing 20 mM MES pH 6.5, 150 mM NaCl 0.01% (w/v) GDN. ICP47^{AF647} comigration was recorded in the $\lambda_{ex/em}$ 650/665 nm channel, using PLC purified with ICP47^{AF647} as reference for 100% binding.

Peptide synthesis, labeling, and purification

Peptides used in this study are listed in Table S2. Peptides were synthesized by automated microwave-assisted solid-phase peptide synthesis by Fmoc chemistry (CEM, Liberty Blue). The C4 peptide (RRYCKSTEL) was labeled with 5-iodoacetamide fluorescein (5-IAF, Sigma-Aldrich) or AF647 in PBS/DMF buffer (8.1 mM Na$_2$HPO$_4$ pH 6.5, 137 mM NaCl, 2.7 mM KCl, 1.8 mM KH$_2$PO$_4$, 33% (v/v) DMF) with heavy agitation for 2 h at 20 °C, using a two-fold (5-IAF) or 1.2-fold (AF647) molar excess of dye. Samples were purified by reversed phase C$_{18}$ HPLC applying a linear water/acetonitrile gradient from 5-60% (v/v) supplemented with 0.1% (v/v) TFA. Purified peptides were snap frozen in liquid nitrogen and lyophilized. Peptide identity was confirmed by LC-MS analysis. RRYC^{AF647}KSTEL (NST^{AF647}): M$_{calc}$: 2121.790 Da, M$_{obs}$: 2121.782 Da, RRYC^{F}KSTEL (C4F): M$_{calc}$: 1543.677 Da, M$_{obs}$: 1543.671 Da, RRYQKSTEL (R9L): M$_{calc}$: 1180.644 Da, M$_{obs}$: 1180.650 Da.

Peptide binding to the PLC analyzed by fluorescence polarization

Binding of fluorescein (^{F})-labeled peptide C4F (RRYC^{F}KSTEL) was followed by fluorescence polarization at RT. For kinetic measurements the polarization of fluorescent peptides (50 nM) in 20 mM MES buffer pH 6.5, 150 mM NaCl, 0.05% (w/v) GDN was analyzed at $\lambda_{ex/em}$ of 485/520 nm using a microplate reader (CLARIOstar, BMG LABTECH). Afterwards, PLC was added to a final concentration of 150 nM, and the sample was mixed 5 s before polarization recording (500 rpm, double orbital mode). 150 µM of R9L (RRYQKSTEL) peptide was added for competition and the measurement continued for 5 min. For endpoint measurements the samples were premixed, incubated for 5 min at RT and resulting fluorescence polarization of C4F peptide was determined as stated above. The fluorescence anisotropy was calculated using:

$$r = \frac{I_\parallel - I_\perp}{I_\parallel + 2 \times I_\perp} \ (1)$$

Resulting curves were fitted using OriginPro 2021 by using linear fit for constant data (fitting function: $y = y_0 + b*x$), exponential fitting with two factors for C4F binding and R9L competition (fitting function: $y = y_0 + A1*exp(-k1x) + A2*exp(-k2x)$).

PLC ATP hydrolysis rate analyzed by radioactive phosphate release from [γ-^{32}P]ATP

500 nM (UV illuminated) PLC were mixed with ATPase buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 5 μM EGTA, 5 mM NaN$_3$, 1 mM Oubain, 0.01% (w/v) GDN) and supplemented with 10 mM MgCl$_2$ and 10 μM R9L and 0.5 mM Vanadate, depending on the stated sample condition. The reaction was started by addition of 5 mM ATP (containing trace amounts of [γ-^{32}P]ATP) for 20 min at 37 °C. Samples were set up in triplicates and spotted in duplicates on TLC PEI-Cellulose F (Merck) plates. TLC plates were run in 0.8 M LiCl in acetic acid and developed for 12 h on a Personal Molecular Imager (PMI) System (BioRad) running Quantity One 4.6.9 (BioRad). Band intensities were quantified in ImageJ 1.53K.

Microsome preparation

Human microsomes were isolated from Burkitt's lymphoma cells cultured as stated above. 1.5 g cells were harvested by centrifugation and resuspended in a homogenization buffer (8 mL of 10 mM HEPES-KOH, 10 mM KCl, 6 mM MgCl$_2$, 1 mM DTT, 0.5 mM PMSF, 0.5 mM benzamidine, 4 units/mL RNAse inhibitor, pH 7.4) at 4 °C. The cell suspension was homogenized in a tight-fitting dounce homogenizer for 20 strokes on ice. After homogenization the solution was made isotonic by addition of 250 mM sucrose. Cell debris were removed by centrifugation at 800 x g for 5 min and 5500 x g for 15 min. The supernatant was collected and brought to 1.8 M sucrose by addition of homogenization buffer containing 2.5 M sucrose. Microsomes were separated in a floating assay by density gradient centrifugation in a SW41 tube (Beckman) containing (from bottom to top): 5 mL 1.8 M sucrose/cell homogenate and homogenization buffer containing 2 mL 1.5 M, 2 mL 1.3 M, 1 mL 1.0 M, 1 mL 0.6 M and 1 mL of 0.25 M sucrose. The density gradient centrifugation was performed in a SW41 swing-out rotor (Beckman) at 286,500 x g for 6 h at 4 °C. The cloudy microsome fraction in the interface between the 1.5 M and 1.8 M sucrose fractions was collected by puncturing the tube wall with a canula, diluted fivefold and centrifuged at 160,000 x g for 45 min at 4 °C. The pellet was dissolved in homogenization buffer and the crude protein concentration was estimated using a microBCA assay (Thermo Scientific). 250 mM sucrose was added and the microsomes were snap frozen in liquid nitrogen and stored at -80 °C until use.

Microsome size determination

Microsomes were analyzed by dynamic light scattering to confirm their diameter using a NanoSight LM10 (Malvern, UK). Data presented are the average of three scans, repeated two times, acquired at 25 °C after an initial 5 min thermal equilibration period. Microsomal size distribution was plotted in OriginPro 2022b and is shown in Figure S1.

PLC topology in microsomes analyzed by proteinase K protection

The proteinase K protection assay to analyze PLC membrane topology was performed as described elsewhere.[3] 60 μg (crude protein) microsomes were incubated for 15 or 30 min at 37 °C with and without 400 μg/mL proteinase K (Thermo Scientific) in the presence or absence of SDS. The reaction was stopped by addition of 20 mM PMSF for 5 min on ice. Samples were precipitated in 10% TCA for 30 min. After centrifugation (5 min, 12,000 x g), pellets were washed with acetone, dried and analyzed by SDS-PAGE and immunoblotting.

Flow cytometry with microsomes

Flow cytometric analyses were performed with a FACSMelody (BD Biosciences) running FACSChorus 1.1.19.0 (BD Biosciences). FITC fluorescence was obtained using the 488 nm laser line with a 527/32/507LP filter set. AF647 fluorescence was obtained using the 640 nm laser line with a 660/10/660/10 filter set. FSC threshold was carefully adjusted to the buffer background in double-logarithmic FSC/SSC view to a background event rate between 0 and 5 events/s. Gates for doublet discrimination were applied to all samples. The experimental setup and gating strategy are described in Figure S6. Flow cytometry data was evaluated with FlowJo 10.6.2 (BD Biosciences).

Peptide translocation assay in microsomes

The transport activity of the peptide loading complex was analyzed by measuring the transport activity of the fluorescent labeled reporter peptide NST^{AF647} (RRYQNSTC^{AF647}L). 20 μg (crude protein) of microsomes were added to 18 μL of transport buffer (MES pH 6.5, 150 mM NaCl) containing 10 mM ATP, 10 mM MgCl$_2$, 10 mM creatine phosphate, 1 mg/mL creatine kinase. ICP47^{SBP} inhibited samples were pre-incubated with 10 μM of ICP47^{SBP} or pc-ICP47^{SBP} for 20 min on ice. UV illuminated samples were exposed to standard UV illumination conditions. The transport was initiated by the addition of 10 μM NST^{AF647} and performed for 5 min at 32 °C in a shaking incubator (300 rpm, Eppendorf). The sample was chilled on ice and the mixture was layered on top of 180 μL transport buffer containing 0.5 M sucrose in a TLA-100 tube (Beckman). Non-transported NST^{AF647} was separated by centrifugation at 228,000 xg for 15 min at 4 °C. The supernatant was removed, and the microsomal pellet was resuspended in 200 μL of transport buffer. The sample was washed again at 228,000 x g for 15 min at 4 °C and resuspended in 200 μL of transport buffer. All samples were prepared in triplicates. Transport activity was analyzed by flow cytometry.

SUPPORTING INFORMATION

Immunostaining of microsomes for flow cytometry

10 µg of microsomes (crude protein) in microsome buffer (10 mM HEPES-KOH pH 7.4, 10 mM KCl, 6 mM MgCl$_2$, 1 mM DTT) were incubated with 650 nM of primary antibody (anti-TAP1, clone mAb 148.3), anti-ERp57 (Abcam, catalogue number ab13506), anti-KDEL (Abcam, catalogue number ab176333), anti-SEC61A (Abcam, catalogue number ab183046), Mouse IgG1 - isotype control (Abcam, catalogue number ab18443) for 30 min at 4 °C. Excess of primary antibody was removed by ultracentrifugation (180,000 x g, 15 min, 4 °C). The samples were resuspended in microsome buffer containing 80 nM of AF647-labeled secondary antibody (goat anti mouse-AF647 (ThermoFisher, catalogue number A-21235), goat anti rabbit-AF647, (ThermoFisher, catalogue number A-21244)) and incubated for 30 min at 4 °C. Excess of secondary antibody was removed by ultracentrifugation (180,000 x g, 15 min, 4 °C). After resuspension in microsome buffer, samples were subjected to flow cytometric analysis.

Supplemental Tables and Figures

Table S1. ICP47 constructs generated and used in this study. J* = photocleavable 3-amino-3-(2-nitrophenyl)-propanoic acid (Anp); Hmp = 2-hydroxy-3-(triphenylmethyl)thio-propanoic acid; (^{AF647}) = Alexa Fluor 647

Construct name	sequence					
pc-ICP47^{SBP}	SWALEMADTF APNASLGVAH GELEQLRARL	LDN**J***RVGPRT RRTGGTVTDS EHHPQGQREP	YADVRDEINK PRNPVTRGGG	RGREDREAAR GSGSGGGSGG	TAVHDPERPL GSMDEKTTGW	LRSPGLLPEI RGGHVVEGLA
^{His6}SUMO-^{Cys}ICP47^{SBP}	MGSSHHHHHH KKTTPLRRLM DVRDEINKRG NPVTRGGGGS	GSGLVPRGSA EAFAKRQGKE REDREAARTA GSGGGSGGGS	SMSDSEVNQE MDSLRFLYDG VHDPERPLLR MDEKTTGWRG	AKPEVKPEVK IRIQADQTPE SPGLLPEIAP GHVVEGLAGE	PETHINLKVS DLDMEDNDII NASLGVAHRR LEQLRARLEH	DGSSEIFFKI EAHREQIGGC TGGTVTDSPR HPQGQREP
^{Cys}ICP47^{SBP}	**C**DVRDEINKR RNPVTRGGGG	GREDREAART SGSGGGSGGG	AVHDPERPLL SMDEKTTGWR	RSPGLLPEIA GGHVVEGLAG	PNASLGVAHR ELEQLRARLE	RTGGTVTDSP HHPQGQREP
ICP47^{SBP}	SWALEMADTF APNASLGVAH GELEQLRARL	LDNMRVGPRT RRTGGTVTDS EHHPQGQREP	YADVRDEINK PRNPVTRGGG	RGREDREAAR GSGSGGGSGG	TAVHDPERPL GSMDEKTTGW	LRSPGLLPEI RGGHVVEGLA
pc-ICP47	SWALEMADTF	LDN**J***RVGPRT	Y–**Hmp**			
pc-ICP47$_{2-55}$	SWALEMADTF	LDN**J***RVGPRT	YADVRDEINK	RGREDREAAR	TAVHDPERPL	LRSP
ICP47^{AF647}	GAMASWALEM LPEIAPNASL ELEQLRARLE	ADTFLDNMRV GVAHRRTGGT HHPQGQREP	GPRTYADVRD VTDSPRNP**C^{AF647}**T	EINKRGREDR RGSGGGSGGG	EAARTAVHDP SMDEKTTGWR	ERPLLRSPGL GGHVVEGLAG

Table S2. Reporter peptides generated and used in this study. (^{F}) = Fluorescein; (^{AF647}) = Alexa Fluor 647

Peptide name	sequence	TAP affinity	HLA-A*03:01, B*15:10, C*04:01, C*03:04 affinity
C4F	RRYC^{F}KSTEL	high	non-binder
NST^{AF647}	RRYC^{AF647}KSTEL	high	non-binder
R9L	RRYQKSTEL	high	non-binder
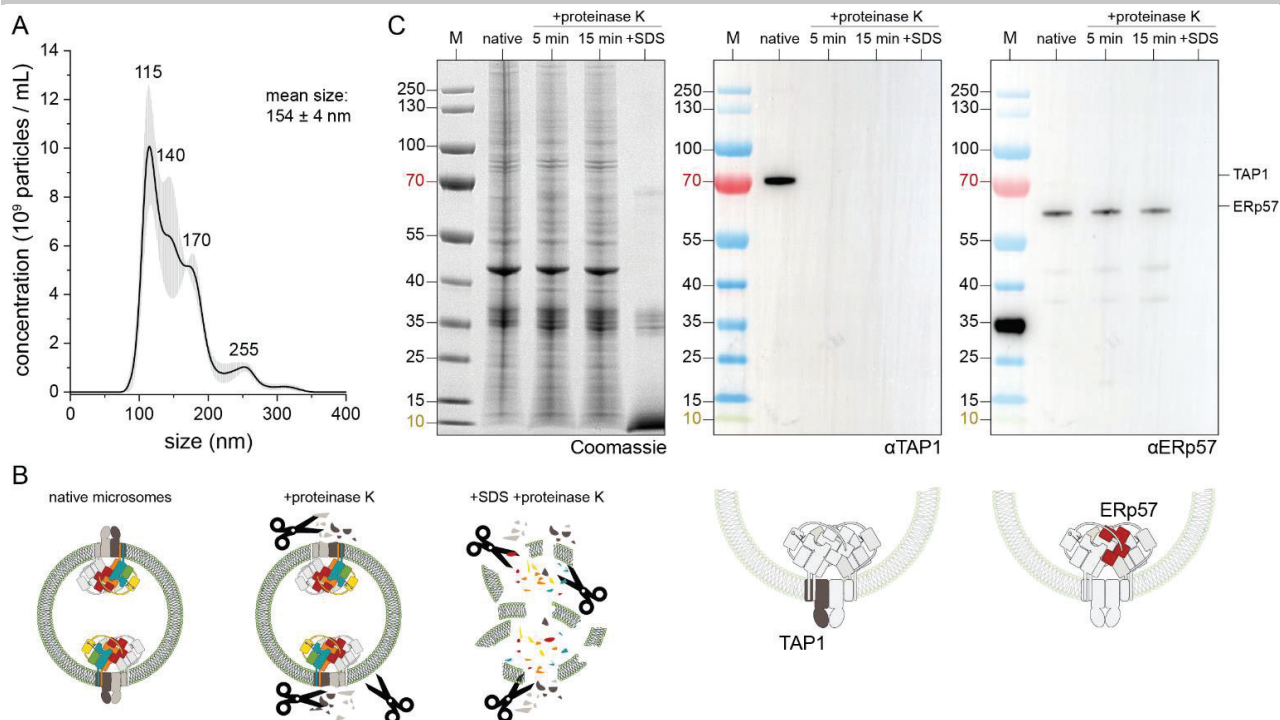
Figure S1. Analysis of microsome diameter and PLC membrane topology. A) Microsome size as determined by dynamic light scattering nanoparticle tracking. The mean microsome diameter is 154 ±4 nm. Error is given as ±SEM (light grey, n=3 in duplicate measurements). B) Schematic illustration of the proteinase K assay. While PLC components are intact in native microsomes, proteinase K digests all protein regions accessible from the exterior. SDS disintegrated microsomes serve as a control for complete digestion. C) Proteinase K protection assay to estimate PLC membrane topology, analyzed by SDS-PAGE (4-12%, reducing, Coomassie-stained) and αTAP1 and αERp57 immunoblots. The outward-facing TAP1 was completely degraded by proteinase K, while the inward-facing ERp57 was entirely protected from proteinase K-mediated degradation. PLC exhibited homogeneous membrane topology in Raji microsomes with the editing module facing to the microsomal lumen.
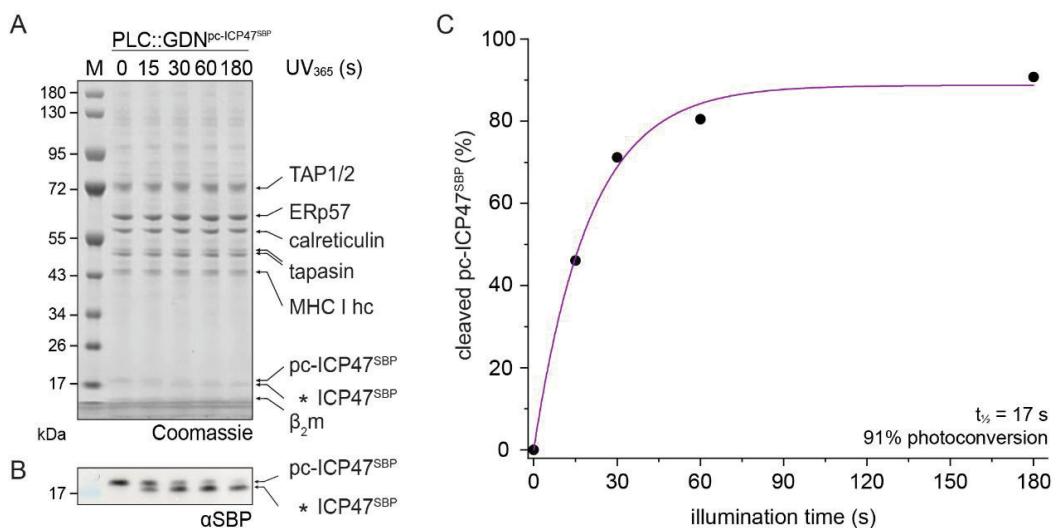
Figure S2. Photocleavage of PLC-bound pc-ICP47SBP. A) PLC purified by pc-ICP47SBP was analyzed by SDS-PAGE (4-12%, reducing, Coomassie-stained) before and after illumination. *ICP47SBP = C-terminal cleavage fragment. B) pc-ICP47SBP cleavage within the PLC analyzed by αSBP immunoblot. pc-ICP47SBP was cleaved within 180 s of UV365 illumination as apparent by the molecular weight shift to *ICP47SBP. C) pc-ICP47SBP cleavage kinetic based on αSBP immunoblot bands as shown in B. The cleavage kinetic of pc-ICP47SBP while bound to TAP1/2 was comparable to cleavage kinetics in solution as shown in Figure 1.
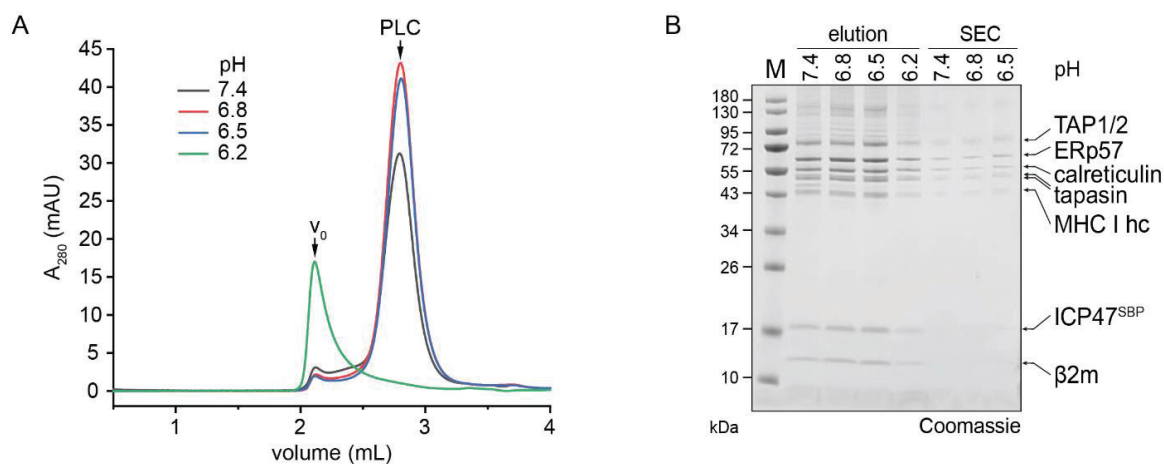
Figure S3. Analysis of PLC pH stability. A) pH stability of the PLC analyzed by SEC. For pH 6.8 and 6.5, elution volume and peak shape did not change in comparison to pH 7.4. At pH 6.2, the PLC aggregates as it elutes at the void volume (V_0). B) PLC preparations at different pH values analyzed by SDS-PAGE (4-12%, reducing, Commassie-stained), directly after preparation (elution) or after SEC purification (SEC, Figure S3A). Between pH 7.4 and 6.5, the PLC subunit composition did not change, while the yield was reduced at pH 6.2.
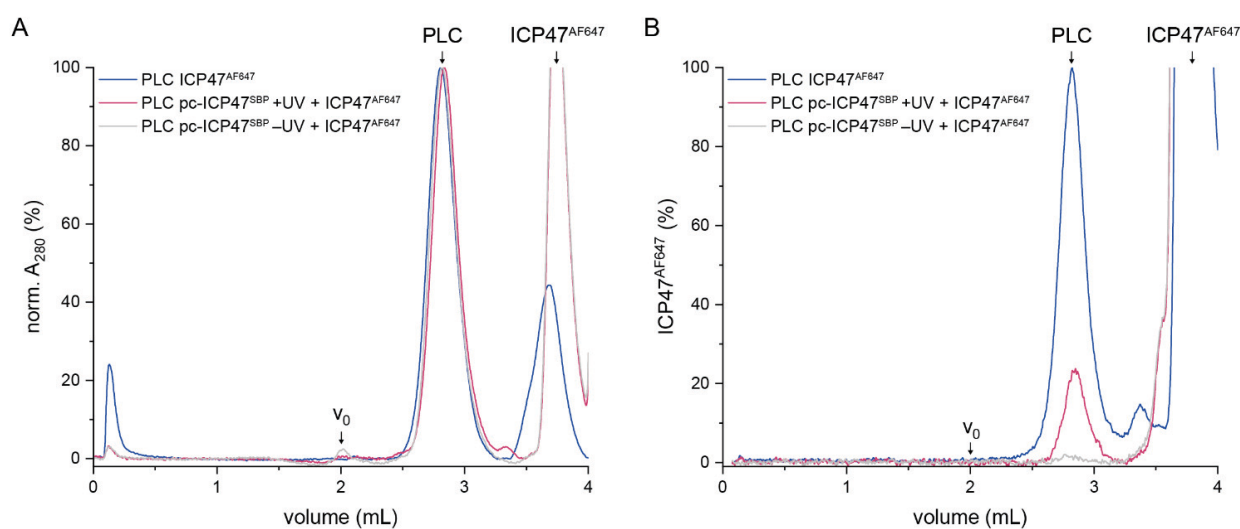
Figure S4. ICP47[AF647] rebinding to the GDN-solubilized PLC after photorelease of pc-ICP47[SBP]. A) PLC purified with ICP47[AF647] (blue) exhibited the same peak shape in SEC as PLC purified with pc-ICP47[SBP]. ICP47[AF647] rebinding was analyzed by addition of ICP47[AF647] to PLC purified with pc-ICP47[SBP]. PLC peak height was normalized to compensate for concentration differences after preparation. AF647 fluorescence in B was normalized with the corresponding factors. B) $\lambda_{ex/em}$ 650/665 nm chromatogram of SEC analyses shown in A. PLC purified with ICP47[AF647] was used as reference for full ICP47[AF647] occupancy (blue). PLC purified with pc-ICP47[SBP] did not show ICP47[AF647] fluorescence. After UV$_{365}$ illumination, 24% of the PLCs were binding competent for ICP47[AF647] (10 µM).
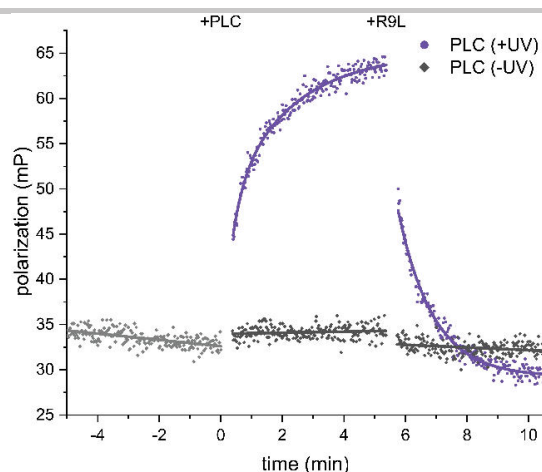
Figure S5. Unblocked PLC peptide binding kinetics. Peptide binding of the high affinity TAP binder C4F to the PLC monitored by fluorescence polarization. PLC (150 nM) was either untreated (black) or UV illuminated (purple) and added to C4F (50 nM). Specific binding was confirmed by addition of 3000-fold molar excess of competitor peptide (R9L).
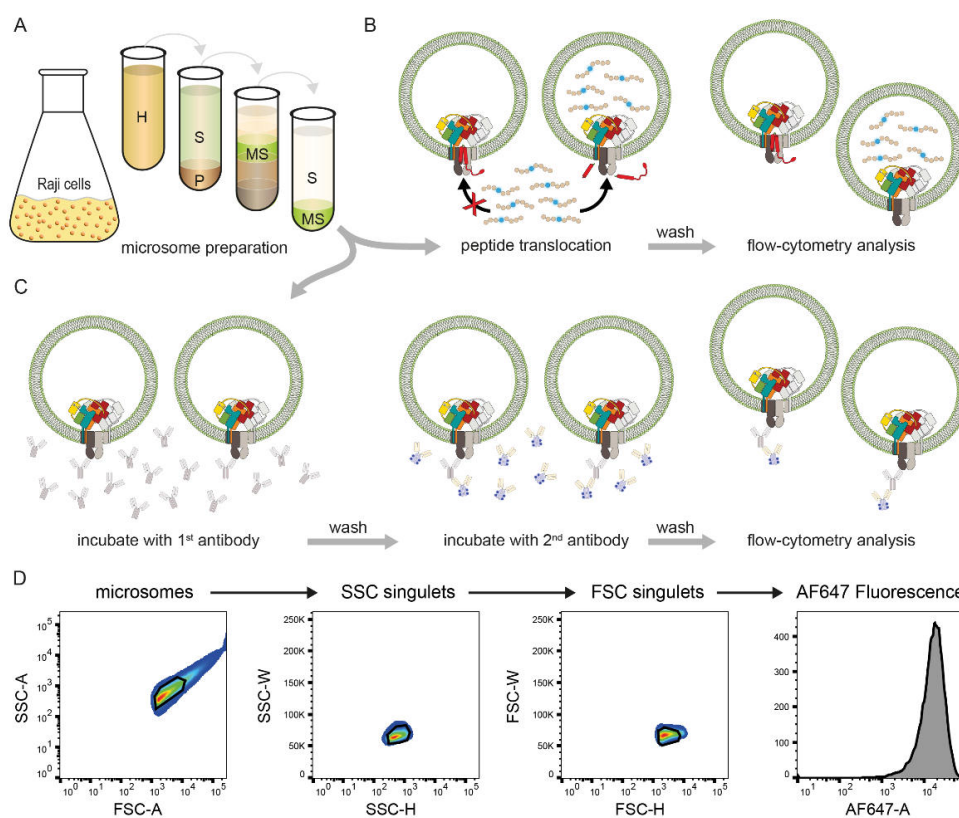
Figure S6. Analysis of microsomes by flow cytometry. A) PLC-containing microsomes were prepared from Raji cells using differential and sucrose gradient centrifugation. H, homogenate; S, supernatant; P, Pellet; MS, microsomes. Microsome size was determined in Figure S1A.PLC was homogeneously orientated with the editing module facing to the microsomal lumen, while the lumen was tightly sealed (Figure S1B). B) Fluorescent reporter peptides were transported into the microsomal lumen upon light induced removal of pc-ICP47^{SBP}, leading to peptide accumulation and signal amplification. After peptide transport, microsomes were analyzed by flow cytometry. C) Immunostaining of microsomes, analyzed by flow cytometry. Unlabeled primary antibody (650 nM) targeted against a membrane associated protein was incubated with microsomes. Unbound primary antibody was removed by ultracentrifugation. 80 nM secondary AF647 labeled antibody was added. Unbound secondary antibody was removed by ultracentrifugation. Microsomes were analyzed by flow cytometry. This assay can also be performed with labeled primary antibodies, omitting the second step. Exemplary staining of membrane-related targets is shown in Figure S7. D) Flow cytometric gating strategy for analysis of microsomes. The threshold of the FSC scatterer had to be configured in a double logarithmic view so that no background events were visible. The microsomes then revealed a characteristic population that is best gated with tight gates around the population center. Fluorescence was analyzed after SSC-H/SSC-W and FSC-H/SSC-W doublet discrimination.
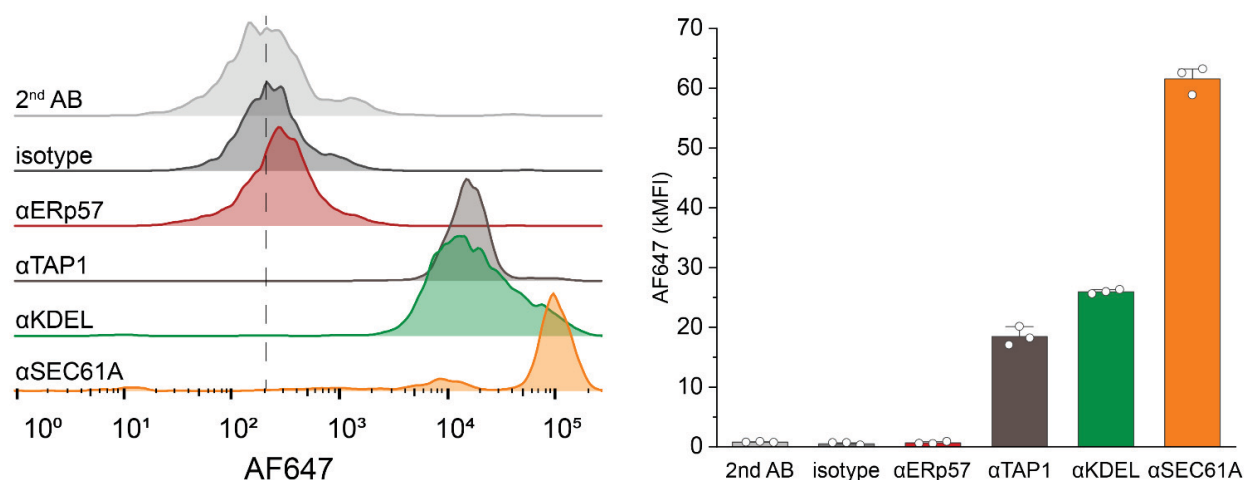
Figure S7. Immunostaining of microsomes against representative ER targets analyzed by flow cytometry. Treatment of microsomes with the secondary, AF647-labeled antibody (grey) or isotype control (dark grey) represented the background signal. Staining against ERp57 had no effect as the epitope is located at the luminal side of the membrane. Staining of the outward-facing epitopes of TAP1 (brown), KDEL (green, ER marker), and SEC61A (orange) lead to a significant shift of flow cytometry traces, indicating a general applicability of flow cytometric analyses for microsomes.

References

[1]   T. H. Meyer, P. M. van Endert, S. Uebel, B. Ehring, R. Tampé, *FEBS Lett* **1994**, *351*, 443-7.
[2]   P. M. van Endert, R. Tampé, T. H. Meyer, R. Tisch, J. F. Bach, H. O. McDevitt, *Immunity* **1994**, *1*, 491-500.
[3]   E. Luquet, C. Biesemann, A. Munier, E. Herzog, *Methods Mol Biol* **2017**, *1538*, 121-34.

Author contributions

C.W performed pc-ICP47[SBP] synthesis, ligation, and desulfurization. LC-MS analyses, flow cytometry and peptide synthesis were performed by C.W. Cell culture was performed by A.D., C.W. and D.C. PLC preparations were performed by A.D with assistance by D.C.. Biochemical analyses of the PLC were performed by A.D. and C.W.. Microsome preparation was performed by C.W. Microsome analysis subsequent biochemical assays were performed by C.W. SEC analyses were performed by A.D.. C.W., A.D. and R.T. prepared the figures and wrote the manuscript. Study and experiments were designed by C.W., A.D. and R.T.. The study was supervised and conceived by R.T..

Molecular basis of MHC I quality control in the peptide loading complex

Alexander Domnick [1,3], Christian Winter [1,3], Lukas Sušac[1], Leon Hennecke[1], Mario Hensen [2], Nicole Zitzmann [2], Simon Trowitzsch [1], Christoph Thomas [1] & Robert Tampé [1] ✉

Major histocompatibility complex class I (MHC I) molecules are central to adaptive immunity. Their assembly, epitope selection, and antigen presentation are controlled by the MHC I glycan through a sophisticated network of chaperones and modifying enzymes. However, the mechanistic integration of the corresponding processes remains poorly understood. Here, we determine the multi-chaperone-client interaction network of the peptide loading complex (PLC) and report the PLC editing module structure by cryogenic electron microscopy at 3.7 Å resolution. Combined with epitope-proofreading studies of the PLC in near-native lipid environment, these data show that peptide-receptive MHC I molecules are stabilized by multivalent chaperone interactions including the calreticulin-engulfed mono-glucosylated MHC I glycan, which only becomes accessible for processing by α-glucosidase II upon loading of optimal epitopes. Our work reveals allosteric coupling between peptide-MHC I assembly and glycan processing. This inter-process communication defines the onset of an adaptive immune response and provides a prototypical example of the tightly coordinated events in endoplasmic reticulum quality control.

The adaptive immune system crucially depends on cell surface presentation of the host proteome in the form of peptides on MHC I molecules[1–3]. These peptide-MHC I (pMHC I) complexes are sampled by cytotoxic T cells, which elicit apoptosis of target cells when non-self peptides are identified[4]. Kinetically stable pMHC I complexes presenting immunogenic peptides are necessary to mount an efficient and specific immune response[5,6]. Kinetically stable pMHC I complexes are ensured by stringent peptide proofreading and quality control processes[7–9], which are orchestrated by the peptide loading complex (PLC), a highly dynamic multi-chaperone machinery in the endoplasmic reticulum (ER)[10]. As central part of the PLC, the transporter associated with antigen processing (TAP) ensures a high local concentration of translocated peptides in the ER lumen proximal to the PLC[1,5,10,11]. The other PLC constituents, namely calreticulin, tapasin, and ERp57, function in recruiting and stabilizing peptide-receptive MHC I

clients, whereby tapasin serves an additional outstanding role as peptide loading and exchange catalyst[6,12–18]. These concerted actions ensure that only MHC I molecules loaded with optimal peptide epitopes are released from the PLC.

In addition to acquiring high-affinity peptides, glycosylation is immensely important for MHC I structure, stability, and quality control, and regulates pMHC I egress to the cell surface[1,4–6,19]. MHC I heavy chains (hc) are co-translationally N-linked glycosylated at a conserved asparagine (Asn86) with a branched $Glc_3Man_9GlcNAc_2$ glycan structure (Glc, glucose; Man, mannose; GlcNAc, N-acetylglucosamine). After trimming of the two outermost glucose residues by ER-resident glucosidases I (GluI) and II (GluII), mono-glucosylated MHC I molecules are specifically recognized by the lectin-like chaperone calnexin, which cooperates with the disulfide isomerase ERp57 in MHC I hc folding[1,14]. Upon binding of β_2-

microglobulin (β_2m) to the MHC I hc, calnexin is replaced by calreticulin, which recruits the mono-glucosylated MHC I to the PLC for peptide editing and quality control[1,6]. Removal of the final glucose of the $Glc_1Man_9GlcNAc_2$ glycan by the GluII heterodimer is a critical step in MHC I maturation and is obligatory for trafficking of MHC I to the cell surface via the secretory pathway[19–21]. However, the sequence of events during peptide editing and glycan processing, and the potential communication between these processes, has not yet been examined for MHC I or any other client-chaperone complexes in the endoplasmic reticulum quality control (ERQC) pathway.

Here, we reveal the molecular underpinnings of epitope proofreading and MHC I quality control within the PLC by structural and biochemical means. Although the overall architecture of the human PLC was reported recently[10], structural elements crucially involved in its catalytic mechanisms have remained unresolved due to the limited resolution. In this study, we determine the 3.7-Å single-particle cryogenic electron microscopy (cryo-EM) structure of the PLC editing module reconstituted in lipid nanodiscs (Nd), which enables us to identify pivotal features of this supramolecular chaperone assembly. Combining structural and functional studies, we discover that MHC I N-glycan processing is allosterically coupled to peptide editing, thus unraveling the central steps at the junction between pMHC I assembly and ERQC.

Results

MHC I molecules exhibit a homogeneous glycan structure in the PLC

To study the PLC in a near-native lipid environment, we isolated detergent-solubilized human PLC from Burkitt's lymphoma cells using the herpesviral protein ICP47 as bait[10]. After affinity purification, the PLC was reconstituted into large nanodiscs (~16 nm diameter) using the membrane scaffold protein MSP2N2. Size-exclusion chromatography (SEC) revealed a homogeneous population of PLC/MSP2N2/lipid particles (Fig. 1a). SDS-PAGE and immunoblotting showed identical stoichiometries of the reconstituted PLC subunits TAP1, TAP2, tapasin, ERp57, calreticulin, MHC I hc, β_2m, and the viral factor ICP47, when compared to detergent-solubilized PLC (Fig. 1b).

To determine the composition of N-linked glycans of the PLC glycoproteins, we analyzed the PLC by liquid chromatography-mass spectrometry (LC-MS) (Supplementary Fig. 1, 2, and Supplementary Table 1). Almost all MHC I molecules (>95%) carried a $Glc_1Man_9GlcNAc_2$ glycan, exemplified by a molecular weight (MW) of 40,340 Da for the most prevalent allomorph HLA-A*03:01 (Fig. 1c). PNGase F treatment led to a shift of 2027 Da (MW of $Glc_1Man_9GlcNAc_2$ minus water), consistent with the removal of the entire N-linked glycan (Fig. 1c). In contrast to MHC I, the glycan of tapasin linked to Asn233 was compositionally heterogeneous (Fig. 1d). While the main tapasin species carried a $Man_9GlcNAc_2$ glycan lacking the three terminal glucose
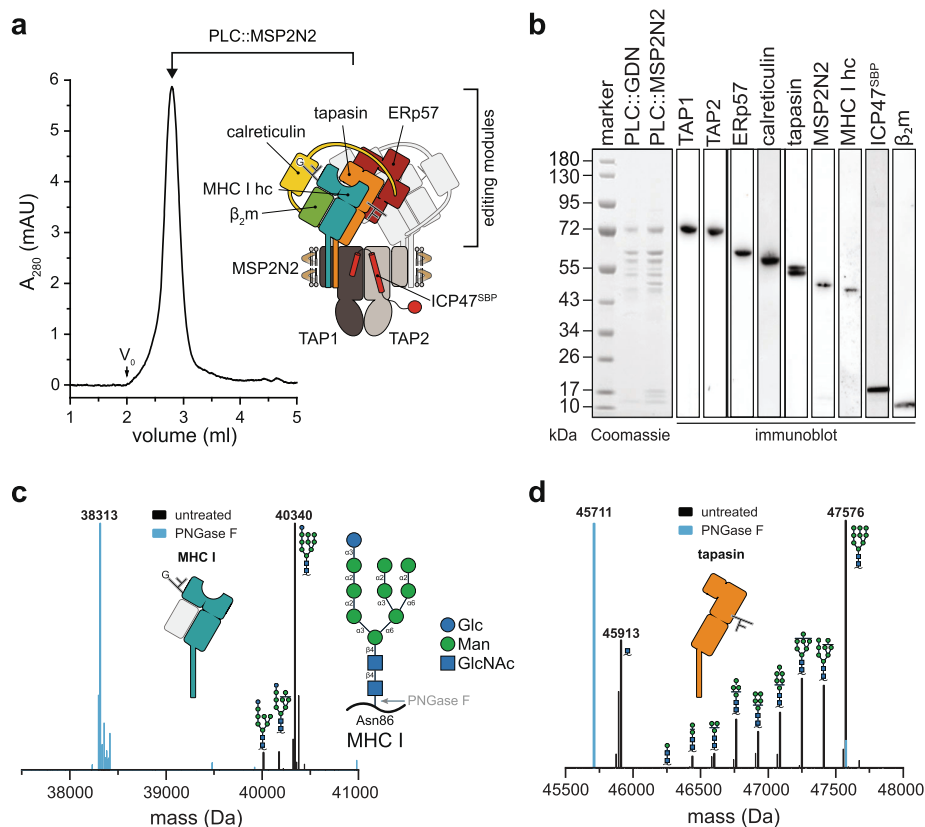
Fig. 1 | PLC reconstituted in large lipid nanodiscs displays distinct glycosylation modification of MHC I and tapasin. a PLC reconstituted in large lipid nanodiscs (PLC::MSP2N2) analyzed by size-exclusion chromatography (SEC). The PLC contains two editing modules, centered around the peptide transporter TAP1/2. For better visualization, one of the two editing modules is shown in light grey. **b** Compositional analysis of human PLC purified by ICP47^{SBP} in detergent (PLC::GDN) and reconstituted in lipid nanodiscs (PLC::MSP2N2) by SDS-PAGE, immunoblotting, and LC-MS (Supplementary Fig. 1 and 2). **c** Glycosylation pattern of MHC I allomorph HLA-A*03:01 associated with the PLC ($Glc_1Man_{9-7}GlcNAc_2$), analyzed by mass spectrometry (deconvoluted spectrum, black). PNGase F-treated sample is shown as reference in cyan. HLA-A*03:01 is the predominant allomorph in LC-MS analysis. **d** Glycosylation of tapasin in the PLC ($Man_{9-0}GlcNAc_{2-1}$), analyzed by mass spectrometry (deconvoluted spectrum, black). PNGase F-treated sample is shown as reference in cyan. The data in this figure are representative of two biological replicas. Source data are provided as a Source Data file (**a**, **b**) or in the Zenodo open access repository (**c**, **d**).
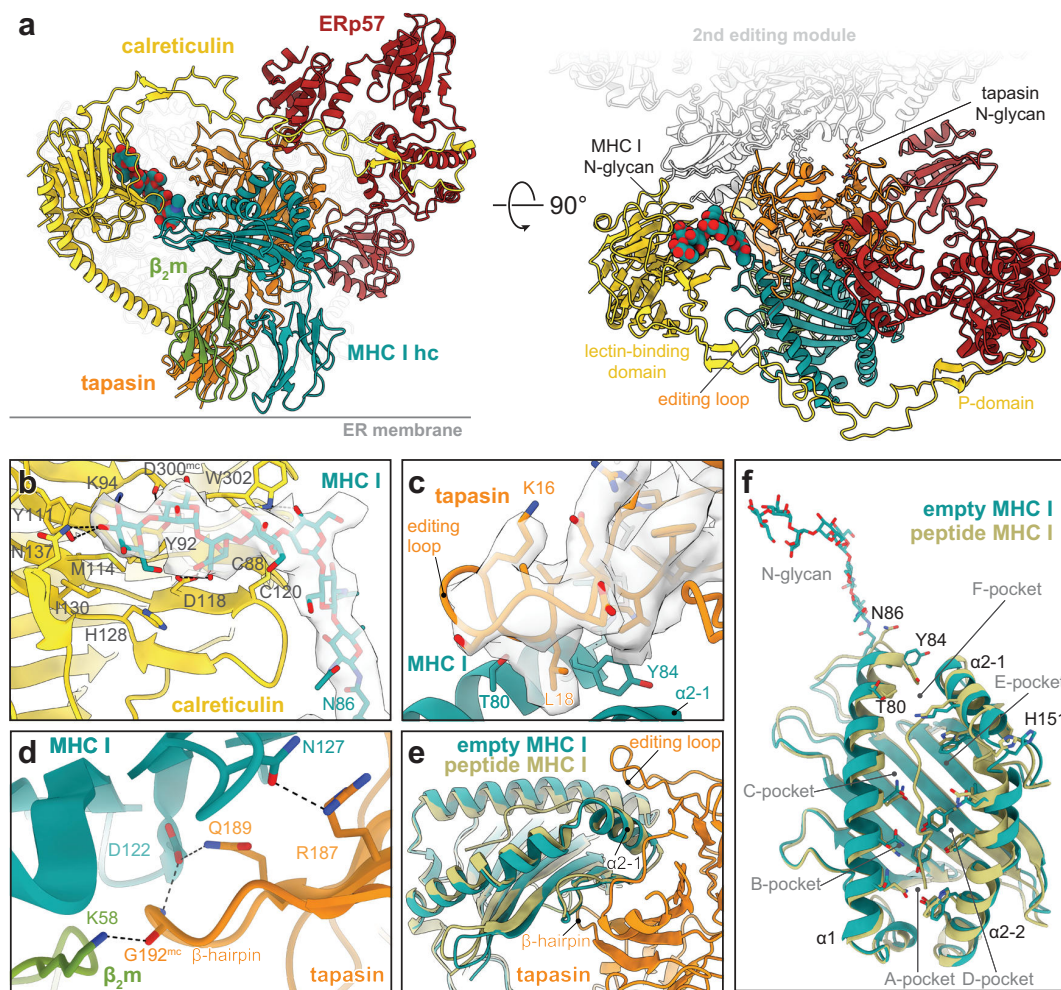
Fig. 2 | Structure of the multivalent MHC I chaperone network in the PLC.
a Cryo-EM structure of the peptide-receptive MHC I stabilized by a multivalent chaperone network in the PLC editing module (side and top view). The glycan trees of MHC I and tapasin are shown as space-filling model and stick model, respectively. Each subunit is colored separately (β_2m, green; calreticulin, yellow; ERp57, red; MHC I hc, teal; tapasin, orange). The position of the second editing module is indicated in light grey. **b** Interaction of the MHC I glycan with the lectin domain of calreticulin. The cryo-EM map of the glycan is depicted as a transparent isosurface

(contour level: 0.19, light grey). **c** Tapasin editing loop and interactions with the MHC I hc molecule. The cryo-EM map of the editing loop is depicted as transparent isosurface (contour level: 0.19, light grey). **d** β-hairpin of tapasin and interactions with β_2m and MHC I hc. **e** Superposition of peptide-loaded MHC I (khaki, HLA-A*03:01, PDB ID: 3RL1) with the empty MHC I of the PLC (side view, including the interface with tapasin). **f** Superposition of PLC-associated peptide-receptive MHC I with peptide-loaded MHC I (khaki, HLA-A*03:01, PDB ID: 3RL1). Source data is available at EMDB (EMD-14119) and PDB (7QPD).

residues, we observed the full range of trimmed glycans (Man$_{8-0}$-GlcNAc$_2$), even down to a single GlcNAc as minimal N-glycan modification. This unusual mannose trimming indicates that tapasin goes through multiple rounds of glycan processing while being protected against ER-associated degradation (ERAD).

Peptide-receptive MHC I are stabilized by a multi-chaperone network

To better understand the multi-chaperone network within reconstituted PLC, we determined the single-particle cryo-EM structure of its editing module to 3.7 Å resolution (Fig. 2). Reconstitution in lipid nanodiscs was crucial to stabilize the transient multi-chaperone machinery for cryo-EM analyses without chemical fixation[10]. As previously shown, 2D and 3D classifications revealed that the PLC is organized in two editing modules centered around the peptide transporter TAP1/2 (Supplementary Figs. 3–5). Despite the confinement in large lipid nanodiscs, the two editing modules and the transport complex TAP remained flexible relative to each other, limiting the

overall resolution of the entire PLC. 3D variation analysis depicts that one editing module was always fully assembled, showing tapasin, ERp57, calreticulin, MHC I hc, and β_2m, while the second one was displaced and lacked some of the PLC components (Supplementary Fig. 4). Focused refinement on the fully assembled editing module resulted in a cryo-EM map with an average resolution of 3.7 Å (Fig. 2a, Supplementary Fig. 5a–d and Supplementary Table 2).

As major interaction hub, tapasin contacts all other PLC subunits (Fig. 2a). In addition to the non-covalent interactions, tapasin is disulfide-linked via Cys95 to Cys33 of ERp57 (Supplementary Fig. 6a). The high-resolution structure allowed us to build an atomic model of the N-glycan bridging the lectin domain of calreticulin to Asn86 of MHC I hc (Fig. 2b). The terminal glucose is coordinated by Lys94, Asn137, and Tyr111 of calreticulin, consistent with the X-ray structure of the isolated lectin domain bound to a Glc$_1$Man$_3$ tetrasaccharide[22]. Whereas Met114 and Ile130 of calreticulin shape the glucose-binding pocket, further contacts to the N-glycan are established by Asp118, Asp300^{mc} (main chain), Tyr92, and Trp302. Additional calreticulin
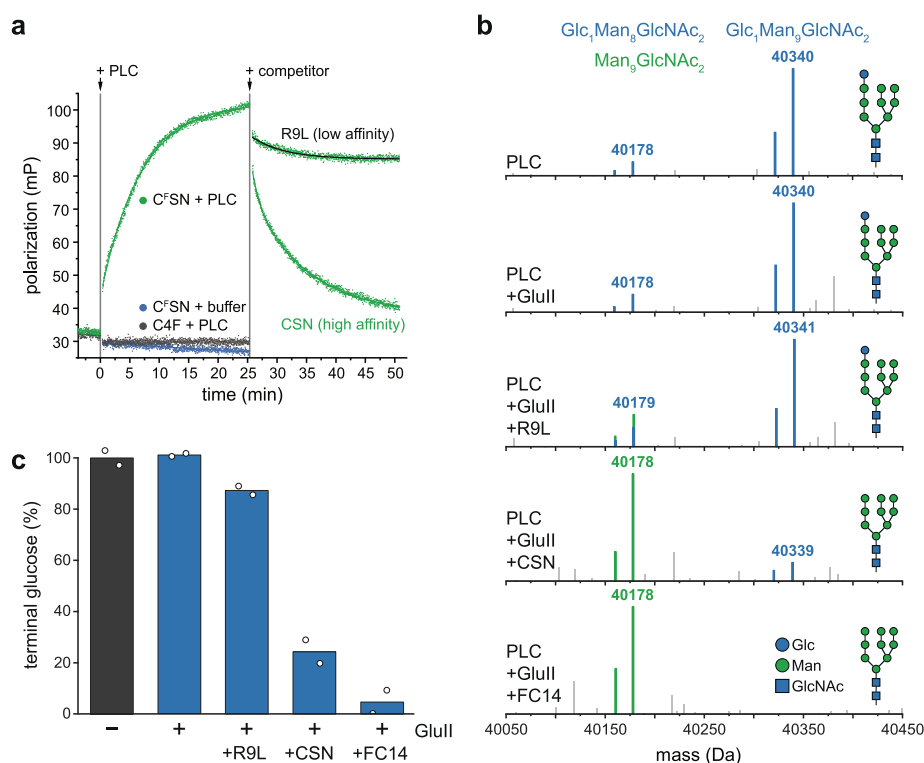
Fig. 3 | Peptide editing and allosterically coupled MHC I glycan processing by GluII in the PLC. a Peptide binding and proofreading of HLA-A*03:01 associated with the PLC reconstituted in lipid nanodiscs (75 nM of PLC) monitored by fluorescence polarization using equimolar concentrations of fluorescent reporter peptides: high-affinity epitope C[F]SN (AIFC[F]SNMTK) or low-affinity HLA-B*27:01-restricted epitope C4F (RRYC[F]KSTEL). For peptide editing, a 2000-fold molar excess of unlabeled high-affinity peptide (150 μM CSN, AIFCSNMTK) or low-affinity peptide (150 μM R9L, RRYQKSTEL, black trace) were used. **b** LC-MS analysis of HLA-A*03:01 glycan status during glycan processing. **c** The percentage of the Glc$_1$Man$_9$GlcNAc$_2$-modified HLA-A*03:01 is given as mean (n = 2). The data are representative of two biological replicas. Source data are provided as a Source Data file (**a**) or in the Zenodo open access repository (**b, c**).

residues involved in glycan binding are His128 and the disulfide bond between Cys88 and Cys120 (Fig. 2b). The mannoses in the B/C branch of the glycan were too flexible to be resolved at high resolution. In contrast to the N-glycan of MHC I, only the trisaccharide Man$_1$GlcNAc$_2$ stem of the Asn233-linked glycan in tapasin was defined by the cryo-EM map, suggesting that the remaining sugar moieties of the tapasin-linked glycan are flexible.

Besides the structurally defined N-linked glycan of MHC I, hereafter referred to simply as "MHC I glycan", our reconstruction revealed crucial features of the tapasin-MHC I multi-chaperone complex which have not been elucidated so far. A well-resolved loop of tapasin (residues 11–20), henceforth referred to as editing loop, is positioned on top of the F-pocket of the empty MHC I peptide-binding groove (Fig. 2c, Supplementary Fig. 6b). Leu18 of the editing loop contacts MHC I in a position that would clash with Tyr84 of peptide-bound MHC I (Fig. 2c). Thus, the structure suggests that Leu18 disrupts the contact of Tyr84 with the C terminus of cargo peptides and thereby contributes to peptide exchange catalysis[23]. This arrangement resembles the position of the scoop loop in TAPBPR, which is larger (16 amino acids vs. 10 in tapasin) but is very similarly positioned and also disturbs the interaction between Tyr84 in MHC I and the C terminus of the cargo peptide by inserting residues 34–36 into the MHC I peptide binding groove[24]. This creates a larger interface with the binding groove than the tapasin loop[25]. The displacement of Tyr84 is stabilized by the interaction with Glu72 of tapasin, consistent with TAPBPR-MHC I crystal structures[24,26]. Leu18 of the editing loop also supports the open MHC I conformation by pushing away Thr80 and Leu81 in the α1 helix, as well as Ala139 and Ile142 in the α2-1 helix of MHC I. HLA-A*03:01, the predominant allomorph in the isolated PLC,

harbors an acidic F-pocket, and it was hypothesized that this acidic F-pocket is stabilized by Lys16 of the editing loop[23]. However, this is not what we observed. On the contrary, the side chain of Lys16 is flexible in our structure and points away from the MHC I peptide-binding groove, which is consistent with previous simulations[25]. Notably, rat and mouse tapasin do not contain any basic residue in the editing loop, rendering the proposed F-pocket stabilization as a general accessory catalytic principle unlikely. Consistent with the structure of TAPBPR-MHC I chaperone complexes[24,26], the floor of the peptide-binding groove is acted upon by a β-hairpin of tapasin (Fig. 2d). Direct contacts in this region include Arg187 and Gln189 of tapasin, as well as Asn127 and Asp122 of MHC I. The latter residue also forms a hydrogen bond with the main-chain of Gly192 at the tip of the tapasin β-hairpin, which additionally interacts with Lys58 of β$_2$m. The contact of Arg187 in tapasin with the MHC I hc might contribute to substrate specificity of the peptide editor, as the interacting MHC I residue (Asn127 in HLA-A*03:01) is not conserved between allomorphs.

A superimposition of the chaperone-stabilized empty MHC I onto the X-ray structure of soluble, non-glycosylated peptide-loaded HLA-A*03:01 (ref. 27) highlights the open conformation of PLC-bound MHC I (Fig. 2e, f). The rearrangement of the α1 and α2 helices results in a widening of the peptide-binding groove, facilitating peptide exchange. The α1 helix (residue 57–85) displays Cα shifts which are most pronounced at the F-pocket (residues 79–85), with a root mean square deviation (RMSD) of 1.8 Å. Interaction with tapasin leads to a lateral shift of the α2 helix which is most pronounced in the hinge region (residues 149–153), with an RMSD of 2.0 Å. In contrast to TAPBPR-MHC I complexes[6,24,26], the empty peptide-binding groove of the PLC-bound MHC I is in its widened conformation stabilized on
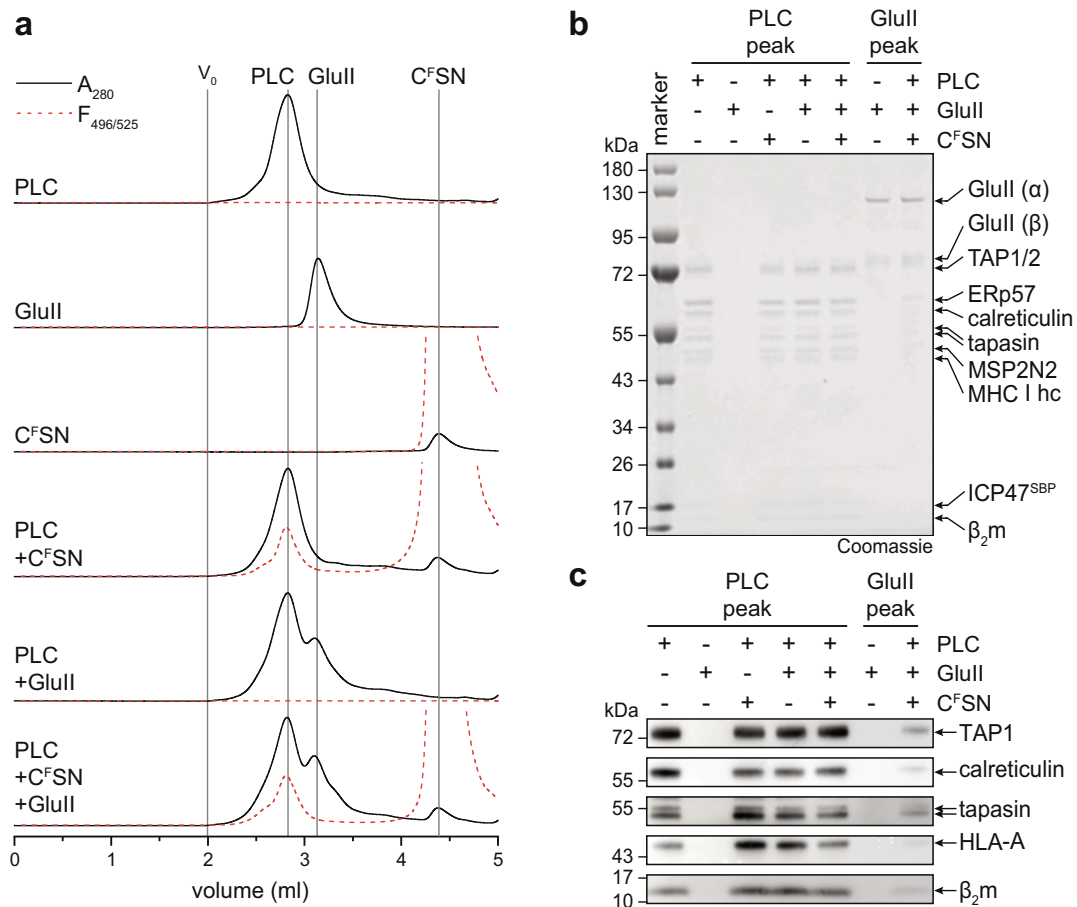
Fig. 4 | PLC remains fully assembled after peptide editing and glycan trimming by GluII. a The size of the Nd-reconstituted PLC remains unchanged after peptide binding and GluII trimming as demonstrated by SEC. Peptide loading and glycan trimming by GluII does not change the overall composition and stoichiometry of the PLC. **b, c** The PLC remains fully assembled after MHC I loading with high-affinity peptide epitope (CSN) and GluII-catalyzed deglucosylation of MHC I, as demonstrated by SDS-PAGE (**b**) and immunoblotting (**c**) of SEC fractions corresponding to the PLC and GluII peak. The data are representative of two biological replicas. Source data are provided as a Source Data file.

both helices, not only by multivalent protein-protein interactions with tapasin, but also by glycan-protein contacts via calreticulin. Based on the shift of the α1 helix, we hypothesized that the closing of the peptide-binding groove after successful peptide editing might result in a release of the glycan from the lectin domain of calreticulin.

The PLC reconstituted in near-native lipid environment is functional in peptide proofreading

Peptide proofreading by tapasin ensures that only kinetically stable pMHC I loaded with high-affinity peptides are released to the cell surface, whereas suboptimal, low-affinity peptides are discharged and eventually replaced by high-affinity epitopes[6,12–15,17,18]. Once MHC I molecules are loaded with an optimal epitope, they must leave the PLC to present the peptide cargo on the cell surface. To pass ERQC, the terminal glucose of the Asn86-linked MHC I glycan must be removed by GluII. We analyzed the peptide proofreading activity of lipid-embedded PLCs by fluorescence polarization using the fluorescein (^{f})-labeled peptide C^{F}SN (AIFC^{F}SNMTK). This HIV-Nef73-derived epitope binds with high affinity to HLA-A*03:01, the predominant MHC I allomorph in PLCs isolated from Burkitt's lymphoma cells. Within 15 min, binding of C^{F}SN was observed upon its addition to PLC-bound MHC I (Fig. 3a). In contrast, the suboptimal peptide C4F (RRYC^{F}KSTEL) did not interact with the PLC. Importantly, bound C^{F}SN peptides were rapidly displaced by an

excess of unlabeled CSN peptide, while a similar excess of the unfavored peptide R9L (RRYQKSTEL) did not trigger substantial peptide exchange (Fig. 3a). Thus, the reconstituted PLC was fully functional in catalyzing peptide proofreading.

Glycan processing is allosterically coupled to peptide editing

To monitor the glycosylation status of HLA-A*03:01 during peptide proofreading, we established an in-vitro glycan processing assay. Reconstituted PLCs were subjected to GluII trimming and analyzed by LC-MS in the presence and absence of the high-affinity peptide CSN or in the presence of the non-binding peptide R9L (Fig. 3b, c). In the absence of peptides, the addition of GluII did not change the glycan of MHC I (Glc$_1$Man$_9$GlcNAc$_2$, 40,340 Da), demonstrating that the terminal glucose is protected from processing by GluII (Fig. 3b, c). However, in the presence of the high-affinity peptide CSN, a single glucose moiety was removed by GluII, reflected by a mass shift of 162 Da (Man$_9$-GlcNAc$_2$-MHC I, 40,178 Da) (Fig. 3b, c, Supplementary Fig. 7). In contrast, processing of the MHC I glycan by GluII was not induced by addition of the unfavored peptide R9L (Fig. 3b, c). Notably, the MHC I glycan was completely deglucosylated after disassembly of the PLC by a short incubation with the detergent Fos-choline-14 (Fig. 3b, c). These observations indicate that GluII-mediated glycosidic bond cleavage of the innermost glucose residue on the MHC I glycan is allosterically coupled to peptide editing.
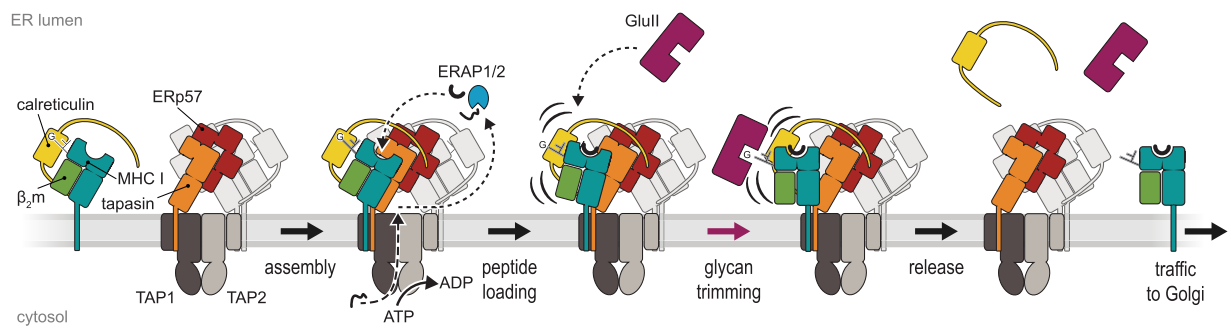
Fig. 5 | Mechanism of MHC I assembly, peptide editing, and quality control. Peptide-receptive MHC I heterodimers are recruited by calreticulin to the PLC and form the fully assembled PLC, which is composed of two editing modules. The peptide transporter TAP1/2 shuttles peptides from the cytosol into a molecular basket formed by the editing modules. Two lateral windows allow peptides to diffuse into the ER lumen to be edited by the ER-resident aminopeptidase ERAP1/2 and subsequently loaded onto MHC I molecules. After tapasin-facilitated peptide

loading and proofreading of pMHC I, the MHC I glycan becomes accessible for trimming by GluII in the presence of calreticulin. Thus, calreticulin is not required to leave the PLC for N-glycan editing when MHC I release is restricted. Glucose-trimmed pMHC I complexes are released from the PLC and traffic via the Golgi compartment to the cell surface. The remaining asymmetric PLC resides in the membrane awaiting the next calreticulin-associated MHC I heterodimers.

Glycan trimming and peptide proofreading are coordinated within the PLC

Based on the allosteric coupling of peptide proofreading and MHC I glycan processing, we finally addressed the question whether calreticulin must leave the PLC for steric reasons to allow glycan trimming by the GluII heterodimer, as suggested by current working models of ERQC[5,28,29]. To this end, we analyzed the PLC composition before and after peptide loading and glycan trimming. Using the high-affinity fluorescent peptide C^{F}SN as reporter, we followed peptide loading of MHC I in parallel by fluorescence-detection SEC. We did not observe any change in size and overall composition of the PLC upon peptide loading of MHC I and glycan trimming by GluII (Fig. 4), indicating that MHC I glycan processing by GluII occurs upon loading of optimal epitopes in the fully assembled PLC containing calreticulin (Fig. 4b, c). This implies that the supramolecular organization of the PLC is characterized by a significant degree of plasticity. Taken together, the finding of an allosteric coupling between peptide editing and glycan processing during MHC I quality control suggests that GluII is a transient component of the PLC providing additional layers of complexity in the organization and function of the PLC.

Discussion

MHC I quality control is of fundamental importance for adaptive immunity. Here we provide structural and biochemical evidence for an allosteric coupling between glycan processing and peptide proofreading coordinated within the PLC. The high resolution of our cryo-EM reconstruction allowed us to build a precise model of the multi-chaperone-client complex including tapasin, ERp57, calreticulin, MHC I hc, and β_2m and to define key mechanistic features. The structure identifies the editing loop of tapasin and the calreticulin-engulfed MHC I glycan as crucial elements during MHC I quality control. Our glycan processing studies demonstrated that, as far as the Nd-reconstituted PLC is concerned, the status of the MHC I glycan is coupled to the peptide-loading status. Based on our structural data and the identification of context-sensitive MHC I glycan processing, we propose the following model for peptide loading and quality control of MHC I (Fig. 5): After recruitment of peptide-receptive MHC I in complex with calreticulin into the PLC, MHC I complexes acquire high-affinity peptide epitopes through tapasin-catalyzed proofreading and peptide exchange. Binding of an optimal epitope weakens the interaction between MHC I and tapasin, but also leads to a rearrangement of its MHC I glycan. This rearrangement loosens the connection between the A branch of the glycan and the lectin domain of calreticulin and allows deglucosylation by GluII as a prerequisite for pMHC I

release from the PLC and subsequent trafficking to the cell surface. This model entails that peptide loading and proofreading of MHC I molecules are coupled to glycan trimming by GluII in a fully assembled PLC, establishing the role of the MHC I glycan as allosteric sensor in quality control. A limitation of the study is that the diffusion of membrane-bound PLC components, including the client (p)MHC I, was restricted by the membrane scaffold. Consequently, MHC I molecules are always in proximity to the PLC, and we could not determine whether pMHC I leaves the PLC before or after GluII trimming. For future experiments, it would be necessary to study the PLC in continuous membranes that allow free diffusion of the PLC and its components.

The structural and mechanistic data also provide the molecular basis of the specificity for glycosylated MHC I in the case of the PLC as opposed to the glycan-independent TAPBPR-MHC I interaction[30]. The observation that GluII-mediated glycan trimming of pMHC I takes place within the calreticulin-containing PLC establishes the notion that the PLC is a highly dynamic and malleable system able to accommodate large transient components. Thus, the PLC emerges as a supramolecular assembly that is more complex than previously anticipated, and which might function as an interaction hub for additional ERQC factors beyond GluII. Our study illustrates the core principles of how sophisticated multi-chaperone systems involving networks of both protein-protein and protein-glycan interactions orchestrate folding, assembly, and quality control of a client in the ER. In the case of MHC I, the multi-chaperone machinery ensures that only stable pMHC I complexes with optimal peptide epitopes are presented at the cell surface, thereby allowing the adaptive immune system to mount an efficient response against pathogens and cancer while avoiding autoimmune reactions. We expect that the prototypical example of tightly coordinated and dynamic quality control by the PLC provided here can be extended to ERQC processes in general.

Methods
ICP47^{SBP} preparation
The ICP47^{SBP} sequence was cloned into a pETM-11 (European Molecular Biology Laboratory, EMBL) using the NcoI and BamHI restriction sites. *Escherichia coli* One Shot BL21(DE3) cells (Thermo Fisher) were transformed and grown to $OD_{600} = 0.6$ at 37 °C. Expression was induced with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) for 6 h at 22 °C. The cells were lysed by sonication in 50 mM NaP$_i$ pH 8.0, 150 mM NaCl, 20 mM imidazole, 0.2% Tween 20, 2 mM DTT, supplemented with 2.5 mM PMSF, 3.75 mg/mL lysozyme, 6.25 mM benzamidine, and 1 U/mL benzonase. His$_6$-TEV-ICP47^{SBP} was isolated from the lysate via reverse IMAC (Ni-NTA Agarose resin, QIAGEN). After washing with

50 mM NaP$_i$ pH 8.0, 150 mM NaCl, 20 mM imidazole, 0.2% Tween 20, 10 mM DTT, ICP47^{SBP} was eluted supplementing the wash buffer with 500 mM imidazole. The His$_6$-tag was removed by overnight TEV protease digestion. ICP47^{SBP} was purified by reversed phase C$_{18}$ HPLC (Agilent, 1200 Series System; PerfectSil 300 ODS C$_{18}$ 5 μm 300x10 mm) applying a linear water/acetonitrile gradient from 5–60% supplemented with 0.1% (v/v) TFA. The identity of ICP47^{SBP} was confirmed by LC-MS analysis (M$_{calc}$: 14694.06 Da, M$_{obs}$: 14694.75 Da).

PLC purification
The native PLC was isolated from Burkitt's lymphoma cells (Raji ATCC® CCL-86), cultured in RPMI 1640 medium (Gibco), supplemented with 10% Fetal Calf Serum (FCS, Capricorn), 3 mM HEPES-NaOH pH 7.5 (Gibco) at 37 °C and 8% CO$_2$ in a shaking incubator (Eppendorf). Cells were harvested by centrifugation, snap frozen in liquid nitrogen, and stored at −80 °C until further use. Cell pellets were thawed and resuspended in 20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 10 mM MgCl$_2$, protease inhibitor mix (Serva) and incubated with ICP47^{SBP} for 15 min at 4 °C. Membranes were mixed with 2% (w/v) glyco-diosgenin (GDN, Anatrace) by douncing and incubated for 2 h at 4 °C under agitation. Insoluble material was removed by centrifugation (45 min, 100,000 × g). ICP47^{SBP}-arrested PLC was bound to Streptavidin High-Capacity Agarose (Pierce) and washed extensively. The PLC was either directly eluted in 20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.05% (w/v) GDN, 2.5 mM biotin (PLC::GDN) or reconstituted into MSP2N2 nanodiscs on the beads (PLC::MSP2N2).

MSP2N2 preparation
pMSP2N2 was a gift from Stephen Sligar (Addgene plasmid #29520). Membrane scaffold protein MSP2N2 was expressed and purified as described[31]. In brief, MSP2N2 was expressed in *E. coli* BL21(DE3) grown in LB media supplemented with 0.5% (w/v) glucose at 37 °C. At an OD$_{600}$ = 0.8, expression was induced with 1 mM IPTG, and cells were grown for 1 h at 37 °C. Subsequently, the temperature was lowered to 28 °C, and cells were grown for an additional 4 h. Cells were disrupted by sonication in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1% (w/v) Triton X-100, 4% Na-deoxycholate, protease inhibitor mix (Serva)). The protein was purified via its N-terminal His$_7$-tag.

PLC reconstitution in membrane scaffolds
PLC bound to Streptavidin High Capacity Agarose (Pierce) was reconstituted in MSP2N2 nanodiscs in a molar ratio of 1:40:500 (PLC:MSP2N2:lipids). First, PLC was incubated with bovine brain lipids (Sigma) for 15 min and 4 °C under constant agitation in buffer with reduced detergent concentration (20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.003% (w/v) GDN). Afterwards, MSP2N2 was added and incubated for further 30 min at 4 °C. To induce nanodisc formation, Biobeads SM-2 (Bio-Rad) were added and incubated for 1 h at 4 °C under constant agitation. Another batch of Biobeads was added for 1 h, 4 °C. The agarose resin was washed extensively with buffer without detergent (20 mM HEPES-NaOH pH 7.4, 150 mM NaCl) and reconstituted PLC::MSP2N2 eluted by biotin (20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 2.5 mM biotin). For subsequent experiments, PLC::MSP2N2 was concentrated using a 100 kDa MWCO filter concentrator (Amicon).

Biochemical analysis of the PLC
For SDS-PAGE, precast NuPAGE gradient gels (Novex) were used. Gels were either stained by Instant Blue (Expedeon) or directly transferred to a PVDF membrane (Bio-Rad). PLC composition was verified by using the antibodies, anti-TAP1 (mAb 148.3, hybridoma supernatant dilution 1:20)[32], anti-TAP2 (mAb 438.3, hybridoma supernatant dilution 1:20)[33], anti-tapasin (Abcam, catalogue number ab13518, dilution 1:1000), anti-HLA-A/B/C HC10 (Acris Antibodies, catalogue number AM33035PU-N, dilution 1:1000), anti-HLA-A (Abcam, catalogue number ab52922,

dilution 1:1000), anti-ERp57 (Abcam, catalogue number ab10287, dilution 1:2000), anti-calreticulin (Sigma, catalogue number C4606, dilution 1:1000), and anti-β$_2$m (Novo Antibodies, catalogue number HPA006361, dilution 1:1000). ICP47^{SBP} was detected by the anti-SBP antibody (Santa Cruz, catalogue number sc-101595, dilution 1:500), MSP2N2-His$_6$ was detected with the anti-His$_6$ antibody (Sigma, catalogue number H1029, dilution 1:1000). Integrity of the PLC was verified by size-exclusion chromatography (SEC). SEC analysis was performed at 4 °C using a Shimadzu HPLC system, equipped with a Shodex semi-micro KW404-4F (4.6 × 300 mm) column. A running buffer containing 20 mM HEPES-NaOH pH 7.4, 150 mM NaCl was used for PLC::MSP2N2 samples. For PLC::GDN samples, the buffer was supplemented with 0.01% (w/v) GDN. Uncropped gels and immunoblots are provided within the source data file.

Peptide design, labeling, and purification
The high-affinity HLA-A*03:01 binder AIFQSSMTK (K_D ≈ 10 nM, HIV-derived epitope) was chosen to create the CSN peptide (AIFCSNMTK, K_D ≈ 10 nM) for fluorescent labeling. Peptide affinities were predicted with NetMHC 4.1 (ref. 34). Peptides were synthesized by automated microwave-assisted solid-phase peptide synthesis by Fmoc chemistry (CEM, Liberty Blue). The CSN peptide and C4 peptide (RRYCKSTEL, K_D ≈ 20 μM) were labeled with 5-iodoacetamide fluorescein (5-IAF, Sigma−Aldrich) in PBS/DMF buffer (8.1 mM Na$_2$HPO$_4$ pH 6.5, 137 mM NaCl, 2.7 mM KCl, 1.8 mM KH$_2$PO$_4$, 33% (v/v) DMF) with heavy agitation for 2 h at 20 °C using a two-fold molar excess of 5-IAF. Samples were purified by reversed phase C$_{18}$ HPLC (Agilent, 1200 Series System; PerfectSil 300 ODS C$_{18}$ 5 μm 300 × 10 mm) applying a linear water/acetonitrile gradient from 5−60% supplemented with 0.1% (v/v) TFA. Purified peptides were snap frozen in liquid nitrogen and lyophilized (Lyovac GT2, Heraeus). Peptide identity was confirmed by LC-MS analysis (AIFCSNMTK: [M + H$^+$]$_{calc}$: 1014.475 Da, [M + H$^+$]$_{obs}$: 1014.475 Da; AIFC^{F}SNMTK: [M + H$^+$]$_{calc}$: 1401.550 Da, [M + H$^+$]$_{obs}$: 1401.549 Da; RRYQKSTEL: [M + H$^+$]$_{calc}$: 1180.636 Da, [M + H$^+$]$_{obs}$: 1180.635 Da; RRYC^{F}KSTEL: [M + H$^+$]$_{calc}$: 1542.669 Da, [M + H$^+$]$_{obs}$: 1542.663 Da).

Peptide binding and editing in the PLC
Binding of fluorescein (^{F})-labeled peptides C^{F}SN (AIFC^{F}SNMTK) or C4F (RRYC^{F}KSTEL) was followed by fluorescence polarization. The polarization of fluorescent peptides (75 nM) in buffer was analyzed at λ$_{ex/em}$ of 485/520 nm using a microplate reader (CLARIOstar, BMG LABTECH). Afterwards, PLC::MSP2N2 was added to a final concentration of 75 nM, and the sample was mixed 5 s before polarization recording (500 rpm, double orbital mode). 150 μM of CSN (AIFCSNMTK) or R9L (RRYQKSTEL) peptide was added for competition and the measurement continued. The fluorescence anisotropy was calculated using:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2 \times I_{\perp}} \tag{1}$$

Resulting curves were fitted using OriginPro 2020 by using linear fit for constant data (fitting function: y = y$_0$ + b*x), exponential fitting with either one factor for binding and R9L competition (fitting function: y = y$_0$ + A*exp(k*x)) or two factors for CSN competition (fitting function: y = y$_0$ + A$_1$*exp(-k_1x) + A$_2$*exp(-k_2x)).

LC-MS analysis
All LC-MS measurements were performed with a BioAccord System (Waters) running Unify 1.9.4 (Waters). Peptides were analyzed with an ACQUITY UPLC Peptide BEH C$_{18}$ Column, 130 Å, 1.7 μm, 2.1 mm × 100 mm (Waters), applying a linear water/acetonitrile gradient supplemented with 0.1% (v/v) formic acid at 60 °C, 30 V cone voltage, 0.8 kV capillary voltage, and 550 °C desolvation temperature. Mass spectra were recorded in positive polarity at 5 Hz in MS^e mode at 50−2000 m/z. Intact protein LC-MS measurements were acquired on an ACQUITY

UPLC Protein BEH C$_4$ Column, 300 Å, 1.7 μm, 2.1 mm × 150 mm (Waters) at 80 °C using a cone voltage of 60 V, 1.5 kV capillary voltage and 500 °C desolvation temperature. Mass spectra were recorded in positive polarity at 2 Hz in full scan mode at 400–7000 m/z. Masses of peptides and proteins were calculated and confirmed in Unify 1.9.4.053 (Waters). Intact mass spectra were deconvoluted in Unify using the integrated MaxEnt1 algorithm iterating to convergence. Spectra with high background noise were subjected to automatic baseline correction before deconvolution. Deconvoluted spectra were centroidized based on peak height, and mock spectra were extracted. Centroidized spectra were used for mass calculations. All intact protein mass spectra show the top 95% of signal intensity. UV spectra were recorded at 280 nm with 10 Hz.

Deglycosylation by PNGase F
20 μL 0.5 mg/ml PLC in 20 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.01% (w/v) GDN were heat-disintegrated for 15 min at 65 °C at 600 rpm. Subsequently, samples were incubated with 2 μL of PNGase F (NEB, 500 units/μL) at 37 °C. After 2 h, 2 μL of PNGase F were added. Samples were incubated overnight and directly subjected to LC-MS analysis.

GluII expression and purification
The *Mus musculus GANAB* (α-GluII α-subunit, UniProt accession no.: Q8BHN3-2 with the amino acid substitution F724G) and *M. musculus PRKCSH* (α-GluII β-subunit, UniProt accession no.: O08795 with the amino acid substitutions L88P and S90N) were cloned into the mammalian expression vector pHLsec carrying a C-terminal His$_6$- or an N-terminal FLAG-tag, respectively. Co-transfection into the FreeStyle 293 Expression System (Life Technologies) took place following the manufactures protocol. Cells were maintained for four days at 37 °C, 8% CO$_2$, shaking at 135 rpm. Cells were pelleted by centrifugation at 3000 × *g* for 45 min. The supernatant was adjusted to 1x PBS (pH 7.4) and sterile filtered. The whole sample was applied to a HisTrap excel column (GE LifeSciences) pre-equilibrated with 1x PBS and subsequently washed with 5 mM imidazole, 1x PBS. Elution took place using 10 column volumes of 350 mM imidazole, 5% (w/v) glycerol, 1x PBS. The imidazole containing buffer of the eluate was exchanged with 5% (w/v) glycerol, 1x PBS using Ultra-15 spin concentrator (Amicon, 30 kDa MWCO) and the concentrated sample applied to a Superdex 200 16/600 column (GE Lifesciences) pre-equilibrated with 20 mM Hepes (pH 7.5) and 150 mM NaCl.

Deglucosylation by GluII
1.2 μM of PLC::MSP2N2 in 20 mM HEPES-NaOH pH 7.4, 150 mM NaCl were preincubated with 20 μM of either high-affinity peptide (CSN), low-affinity peptide (R9L), or 1 mM FC14 for 30 min at 4 °C. Deglycosylation by GluII (0.6 μM final) was studied at 37 °C for 5 min. The reaction was stopped by 80 °C heat inactivation, and samples were analyzed by LC-MS.

Cryo-EM sample preparation and data collection
3 μL of PLC::MSP2N2 (1.8 mg/ml, 2.5 μM) sample were applied onto freshly glow-discharged copper grids (Quantifoil, Cu R1.2/1.3) and plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher). Micrographs were recorded automatically (SerialEM) on a 300-kV FEI Titan Krios in energy-filtered transmission electron microscopy mode with a K2 direct detector (Gatan) and a Gatan GIF Quantum SE post-column energy filter at 130,000x magnification and a pixel size of 1.05 Å. Dose-fractionated movies were acquired at an electron flux of 5.1 e⁻ per pixel per s over 14 s with 0.35 s exposures per frame (40 frames in total), corresponding to a total electron dose of ~65 e⁻ Å². Movies were recorded in the defocus range from −1.5 to −2.5 μm (Supplementary Table 2).

Cryo-EM data analysis
Cryo-EM data analysis was performed using CryoSPARC v2.15–3.2 (ref. 35). A total of 2341 movies were used for analysis. Motion correction was performed using patch motion correction implemented in CryoSPARC. The contrast transfer function (CTF) was estimated using patch CTF function of CryoSPARC. Initial particles were picked from a subset of data using a blob picker (minimum particle diameter 150 Å, maximum particle diameter 400 Å). The particles were extracted at a box size of 360 px and subjected to 2D classification to identify good particle picks. These picks were used to train Topaz (deep picker) on denoised images[36,37]. Subsequently, the trained picking model was utilized to pick particles in all images, yielding 613,746 particles picked and extracted at a box size of 360 px. 2D classification was used to identify good particles for ab initio model generation, which directly yielded maps for full PLC including the membrane region as well as a more focused map consisting of one editing module only. These maps were used as references to identify suitable particles for refinement in several rounds of 3D classification. The 3D classification resulted in two clean particle stacks for the full PLC (52,668 particles) and the editing module (97,952 particles). Each subset of particles was individually refined by homogeneous and non-uniform refinement. Optimal results were obtained using non-uniform refinement utilizing a mask for the region of interest, enforced non-negativity, optimization of per particle defocus, and optimized per group CTF parameters enabled. For the focused editing submodule map, a resolution of 3.73 Å was estimated using the 0.143 cut-off criteria (Supplementary Fig. 5). For the full PLC, the highest resolution obtained was 4.01 Å. However, this map seemed to represent a consensus map of different PLC compositions of mainly one well resolved editing module; the second tapasin and MHC I were less represented. Separating the underlying states using ab initio model generation and/or heterogeneous refinement failed. Heterogeneous reconstruction with the same set of particles by deep learning using the neural network-based algorithm of CryoDRGN[38] (https://github.com/zhonge/cryodrgn) (version 0.3.1) revealed different assembly states of the PLC. CryoDRGN was run for 50 epochs with input image poses and CTF parameters from a consensus homogeneous reconstruction in cryoSPARC, encoder and decoder network architectures of 3 hidden layers with 256 nodes per layer, 8-dimensional latent space, and pose refinement. Selected maps generated by cryoDRGN representative of the latent space, as judged from a principal component analysis (PCA) projection, were used as references in heterogeneous 3D refinement in CryoSPARC and resulted in five subsets of particles representing different assembly states of the PLC (Supplementary Fig. 4). Individual stacks were subjected to homogeneous and non-uniform refinements resulting in resolutions between 6 Å and 9 Å.

Model building
The structure of a PLC editing module (PDB ID: 6ENY) was initially fitted into the cryo-EM map using bulk flexible fitting by ISOLDE[39] implemented for ChimeraX. Final model building was carried out in COOT[40], and real-space refinement was performed using Phenix[41]. The editing loop of tapasin was modeled de novo. Carbohydrates were validated using Privateer[42] of the CCP4 software suite[43]. Refinement and validation statistics are summarized in Supplementary Table 2.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
LC-MS raw data were submitted to the open-access repository Zenodo with the Digital Object Identifier (DOI) https://doi.org/10.5281/zenodo.5793891. The cryo-EM density maps and the corresponding model were deposited in the Electron Microscopy Data Bank under accession

numbers EMD-14119 and PDB ID 7QPD. Source data are provided with this paper.

References

1. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. *Annu Rev. Immunol.* **31**, 443–473 (2013).
2. Madden, D. R. The three-dimensional structure of peptide-MHC complexes. *Annu Rev. Immunol.* **13**, 587–622 (1995).
3. Bjorkman, P. J & Parham, P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev. Biochem.* **59**, 253–288 (1990).
4. Cresswell, P., Ackerman, A. L., Giodini, A., Peaper, D. R. & Wearsch, P. A. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunol. Rev.* **207**, 145–157 (2005).
5. Trowitzsch, S. & Tampé, R. Multifunctional chaperone and quality control complexes in adaptive immunity. *Annu Rev. Biophys.* **49**, 135–161 (2020).
6. Thomas, C. & Tampé, R. MHC I assembly and peptide editing - chaperones, clients, and molecular plasticity in immunity. *Curr. Opin. Immunol.* **70**, 48–56 (2021).
7. Jappe, E. C. et al. Thermostability profiling of MHC-bound peptides: a new dimension in immunopeptidomics and aid for immunotherapy design. *Nat. Commun.* **11**, 6305 (2020).
8. Stronen, E. et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
9. Bashirova, A. A. et al. HLA tapasin independence: broader peptide repertoire and HIV control. *Proc. Natl Acad. Sci. USA* **117**, 28232–28238 (2020).
10. Blees, A. et al. Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525–528 (2017).
11. Neefjes, J., Jongsma, M. L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
12. Ortmann, B. et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* **277**, 1306–1309 (1997).
13. Chen, M. & Bouvier, M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. *EMBO J.* **26**, 1681–1690 (2007).
14. Wearsch, P. A. & Cresswell, P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat. Immunol.* **8**, 873–881 (2007).
15. Thirdborough, S. M. et al. Tapasin shapes immunodominance hierarchies according to the kinetic stability of peptide-MHC class I complexes. *Eur. J. Immunol.* **38**, 364–369 (2008).
16. Wearsch, P. A., Peaper, D. R. & Cresswell, P. Essential glycan-dependent interactions optimize MHC class I peptide loading. *Proc. Natl Acad. Sci. USA* **108**, 4950–4955 (2011).
17. Fleischmann, G. et al. Mechanistic basis for epitope proof-reading in the peptide-Loading complex. *J. Immunol.* **195**, 4503–4513 (2015).
18. Garbi, N. et al. Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. *Nat. Immunol.* **1**, 234–238 (2000).
19. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370–2376 (2001).
20. Ellgaard, L. & Helenius, A. Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* **4**, 181–191 (2003).
21. Caputo, A. T. et al. Structures of mammalian ER alpha-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals. *Proc. Natl Acad. Sci. USA* **113**, E4630–E4638 (2016).
22. Kozlov, G. et al. Structural basis of carbohydrate recognition by calreticulin. *J. Biol. Chem.* **285**, 38612–38620 (2010).
23. Lan, H. et al. Exchange catalysis by tapasin exploits conserved and allele-specific features of MHC-I molecules. *Nat. Commun.* **12**, 4236 (2021).
24. Thomas, C. & Tampé, R. Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing. *Science* **358**, 1060–1064 (2017).
25. McShan, A. C. et al. TAPBPR promotes antigen loading on MHC-I molecules using a peptide trap. *Nat. Commun.* **12**, 3174 (2021).
26. Jiang, J. et al. Crystal structure of a TAPBPR-MHC I complex reveals the mechanism of peptide editing in antigen presentation. *Science* **358**, 1064–1068 (2017).
27. Zhang, S. et al. Structural basis of cross-allele presentation by HLA-A*0301 and HLA-A*1101 revealed by two HIV-derived peptide complexes. *Mol. Immunol.* **49**, 395–401 (2011).
28. Chapman, D. C. & Williams, D. B. ER quality control in the biogenesis of MHC class I molecules. *Semin Cell Dev. Biol.* **21**, 512–519 (2010).
29. Peaper, D. R. & Cresswell, P. Regulation of MHC class I assembly and peptide binding. *Annu Rev. Cell Dev. Biol.* **24**, 343–368 (2008).
30. Neerincx, A. & Boyle, L. H. Preferential interaction of MHC class I with TAPBPR in the absence of glycosylation. *Mol. Immunol.* **113**, 58–66 (2019).
31. Grinkova, Y. V., Denisov, I. G. & Sligar, S. G. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng. Des. Sel.* **23**, 843–848 (2010).
32. Meyer, T. H., van Endert, P. M., Uebel, S., Ehring, B. & Tampe, R. Functional expression and purification of the ABC transporter complex associated with antigen processing (TAP) in insect cells. *FEBS Lett.* **351**, 443–447 (1994).
33. van Endert, P. M. et al. A sequential model for peptide binding and transport by the transporters associated with antigen processing. *Immunity* **1**, 491–500 (1994).
34. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
35. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
36. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
37. Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat. Commun.* **11**, 5208 (2020).
38. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
39. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D. Struct. Biol.* **74**, 519–530 (2018).
40. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D. Biol. Crystallogr* **60**, 2126–2132 (2004).
41. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
42. Agirre, J. et al. Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).
43. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D. Biol. Crystallogr* **67**, 235–242 (2011).

Acknowledgements

Author contributions

Cell culture and PLC preparations were devised by A.D. and L.H. A.D. prepared the cryo-EM samples and collected the fluorescence polarization data. C.W. carried out the LC-MS analyses, the GluII assays, and peptide synthesis. A.D., L.S., and C.T. analyzed the cryo-EM data and built the final PLC model. A.D., C.W., C.T., S.T., and R.T. prepared the figures and wrote the manuscript. M.H. and N.Z. provided the GluII. Study and experiments were designed by A.D., C.W., and R.T. The study was supervised and conceived by R.T.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at
https://doi.org/10.1038/s41467-022-32384-z.

Correspondence and requests for materials should be addressed to Robert Tampé.

Peer review information *Nature Communications* thanks Malini Raghavan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at
http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Molecular basis of MHC I quality control in the peptide loading complex

Alexander Domnick[1,#], Christian Winter[1,#], Lukas Sušac[1], Leon Hennecke[1], Mario Hensen[2],

Nicole Zitzmann[2], Simon Trowitzsch[1], Christoph Thomas[1], Robert Tampé[1*]

[1]Institute of Biochemistry, Biocenter, Goethe University Frankfurt, Max-von-Laue-Str. 9,

60438 Frankfurt am Main, Germany

[2]Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford,

OX1 3QU Oxford, UK

[#]Contributed equally, *To whom correspondence may be addressed

Email: tampe@em.uni-frankfurt.de

Supplementary Figure 1-7

Supplementary Table 1 and 2
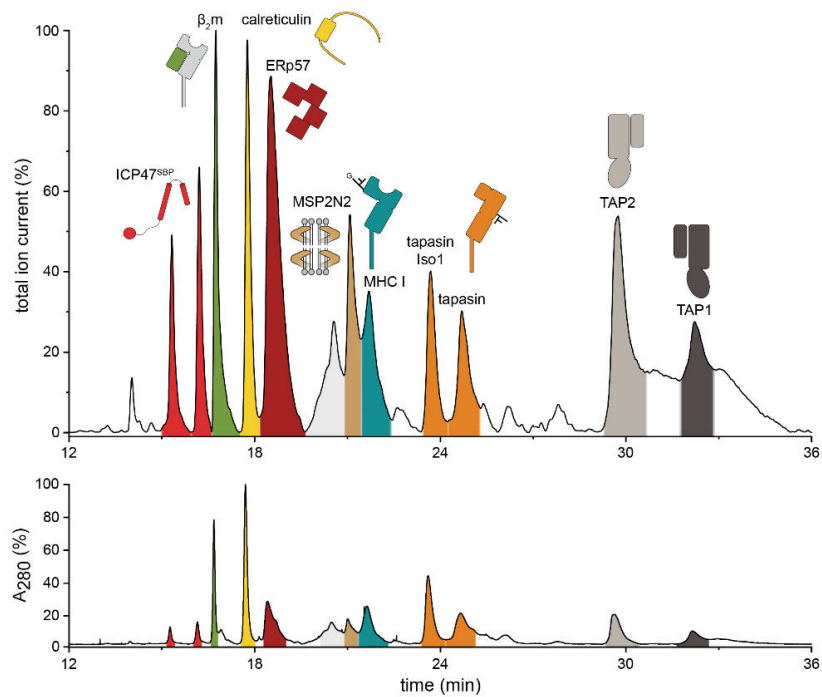
Supplementary Fig. 1 PLC subunits analyzed by LC-MS. The native, ICP47^{SBP}-bound PLC was reconstituted in large nanodiscs and subjected to LC-MS analysis. The separated PLC subunits were highlighted and assigned in their respective color in the TIC (total ion current) and A$_{280}$. All components of the PLC, including ICP47^{SBP} and the membrane scaffold protein MSP2N2, were successfully identified by mass spectrometry. The mass spectra and deconvoluted MS data for identified components are shown in Supplementary Fig. 2 and listed in Supplementary Table 1. ICP47^{SBP} is present in two variants, one of which contains the additional N-terminal sequence GAM. Tapasin appears in two isoforms in Raji cells.
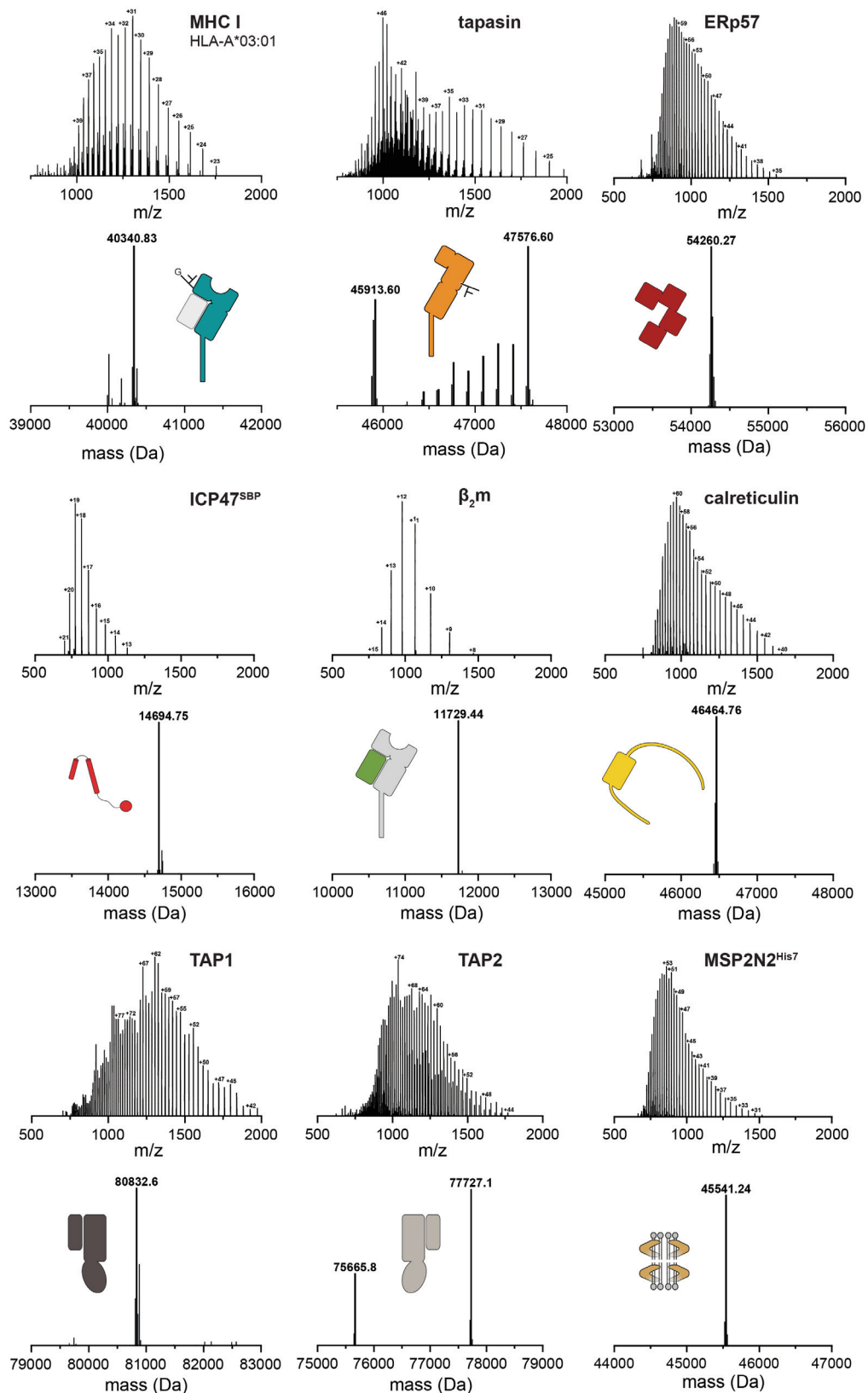
Supplementary Fig. 2 Identification of PLC subunits by intact protein LC-MS. ESI-MS and MaxEnt1 deconvoluted spectra of PLC components (Supplementary Fig. 1). The assignment of the sequences and theoretical masses of PLC components are shown in Supplementary Table 1.
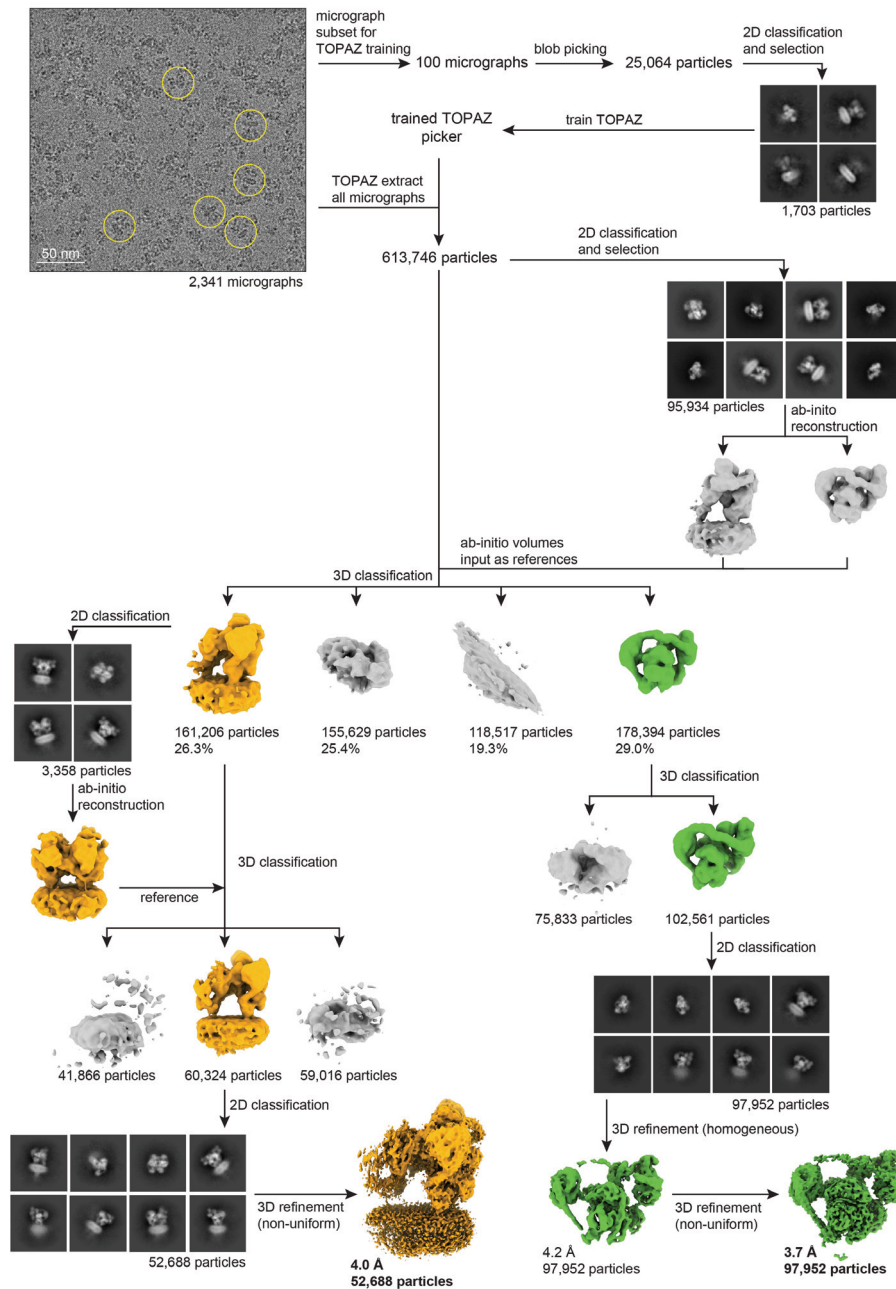
Supplementary Fig. 3 Cryo-EM data processing workflow. The full set of micrograph movies was corrected for motion and contrast transfer function in cryoSPARC. A subset of 100 micrographs was subjected to blob picking followed by 2D classification to obtain a training dataset of PLC particles for the deep picker TOPAZ. Subsequently, the trained TOPAZ instance was used to particle pick the entire set of micrographs. Initial 2D classification was then performed to select particles for ab-initio 3D reconstruction. The resulting ab-initio maps were used to classify the full dataset in 3D. 3D assemblies resembling PLC with density for the membrane region (orange, left) and a single editing module (green, right) were further sub-classified in 3D and 2D, and individually refined using homogenous and finally non-uniform refinement. All processing was performed using cryoSPARC.
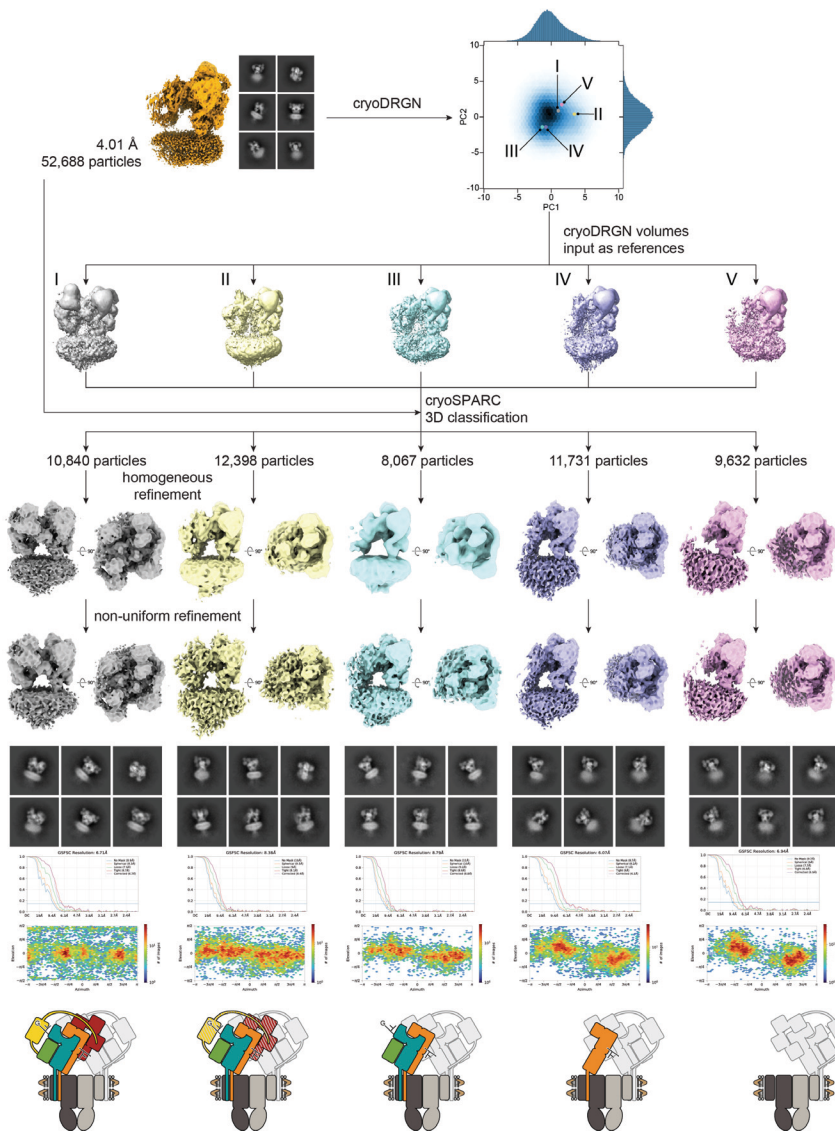
Supplementary Fig. 4 CryoDRGN analysis of different PLC assembly states. Particles of a single reconstruction from a homogeneous 3D refinement of the PLC (orange, "consensus map") were analyzed by deep learning using the neural network-based algorithm of cryoDRGN to reveal distinct PLC assemblies. Five different assembly states (I–V, colored tiles) were chosen from the principal component analysis (PCA) projection of the 8-dimensional latent space (top right) for 3D classification, followed by homogeneous and non-uniform refinement in cryoSPARC. The chosen 3D reconstructions generated by cryoDRGN showed distinct PLC assembly intermediates which could not be separated in cryoSPARC before. The fully assembled PLC (I) with two complete editing modules is shown in grey, whereas the reconstruction with a single editing module (V) is depicted in pink. Densities in yellow, blue, and purple (II – IV) represent reconstructions with intermediate assembly states of the second editing module. Depicted below each 3D map from non-uniform refinement are representative 2D class averages, Fourier shell correlation (FSC) curves of the 3D reconstructions from two independently refined half datasets, and angular assignments of particles from the final dataset with respect to the 3D structure above. Cartoons at the bottom depict the respective assembly state of the PLC by illustrating the subunits identified in each map.
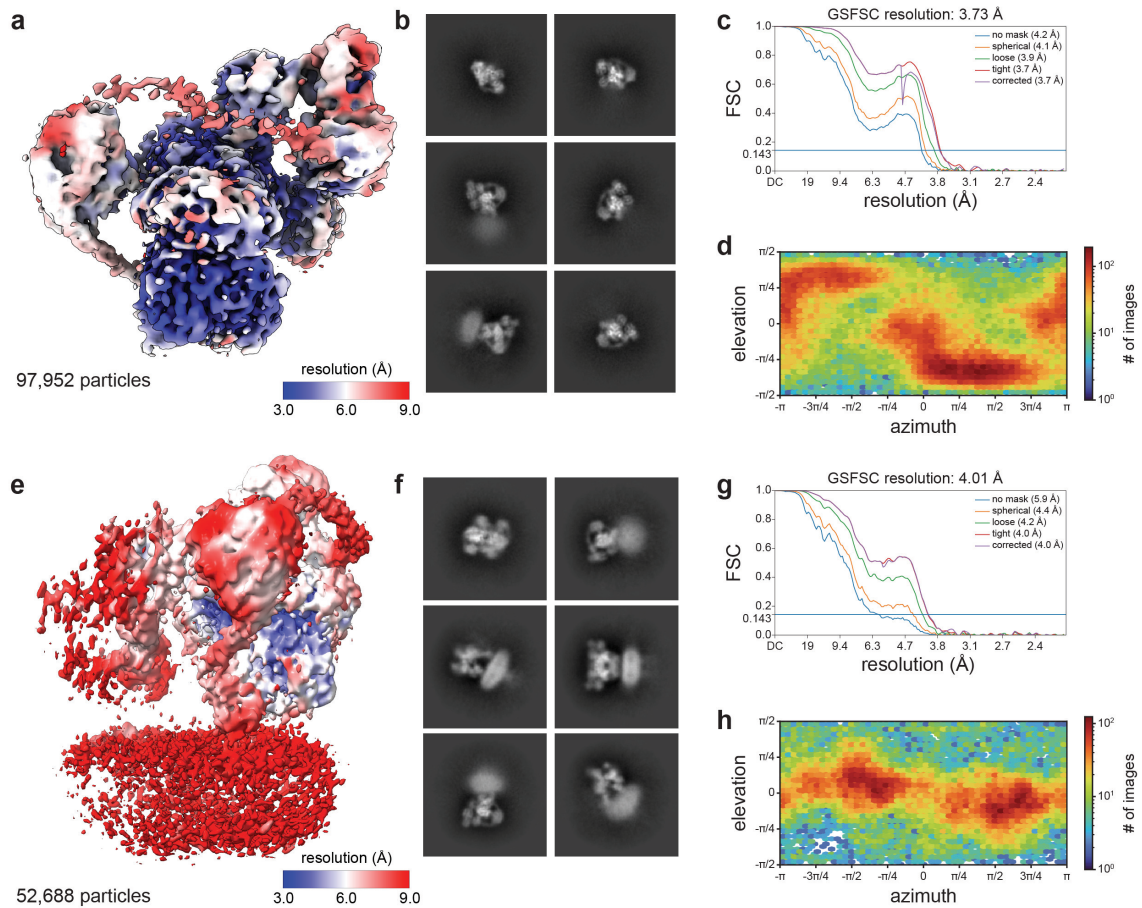
Supplementary Fig. 5 Cryo-EM analysis of the PLC reconstituted in lipid nanodiscs. a–d Single editing module. **e–f** Full PLC. **a, e** Local resolution distributions of the cryo-EM maps color-coded as indicated in the insets. **b**, **f** Representative 2D class averages from the final particle sets. Note the defined density for ER-lumenal editing module regions and blurred signal for all membrane-associated regions for the classes corresponding to the single editing module (**b**), while both editing modules are visible and the membrane region is more defined but still blurry in the classes corresponding to full PLC (**f**). **c, g** Fourier shell correlation (FSC) curves for the 3D reconstructions from two independently refined half datasets after non-uniform refinement in cryoSPARC. For the single editing module (**c**) and the full PLC (**g**), overall resolutions of 3.73 Å and 4.01 Å were achieved, respectively, as judged by the 0.143 threshold criterion. **d, h** Angular assignments for particles from the final datasets used for 3D reconstructions.
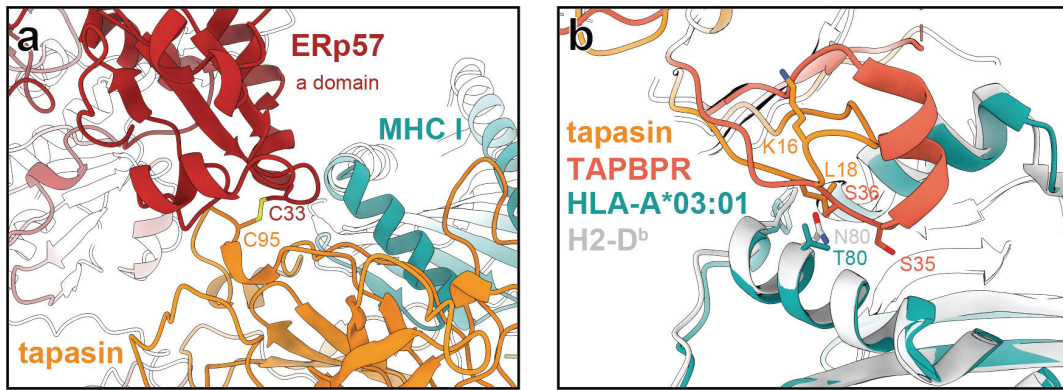
Supplementary Fig. 6 Structural features of the PLC editing. a ERp57 is linked to tapasin via an intermolecular disulfide bridge formed between Cys33 of ERp57 and Cys95 of tapasin. **b** Superposition of the tapasin editing loop in the PLC (orange/teal, PDB ID 7QPD) with the TAPBPR scoop loop in the TAPBPR-H2-Db chaperone complex (light red/light grey, PDB ID 5OPI).
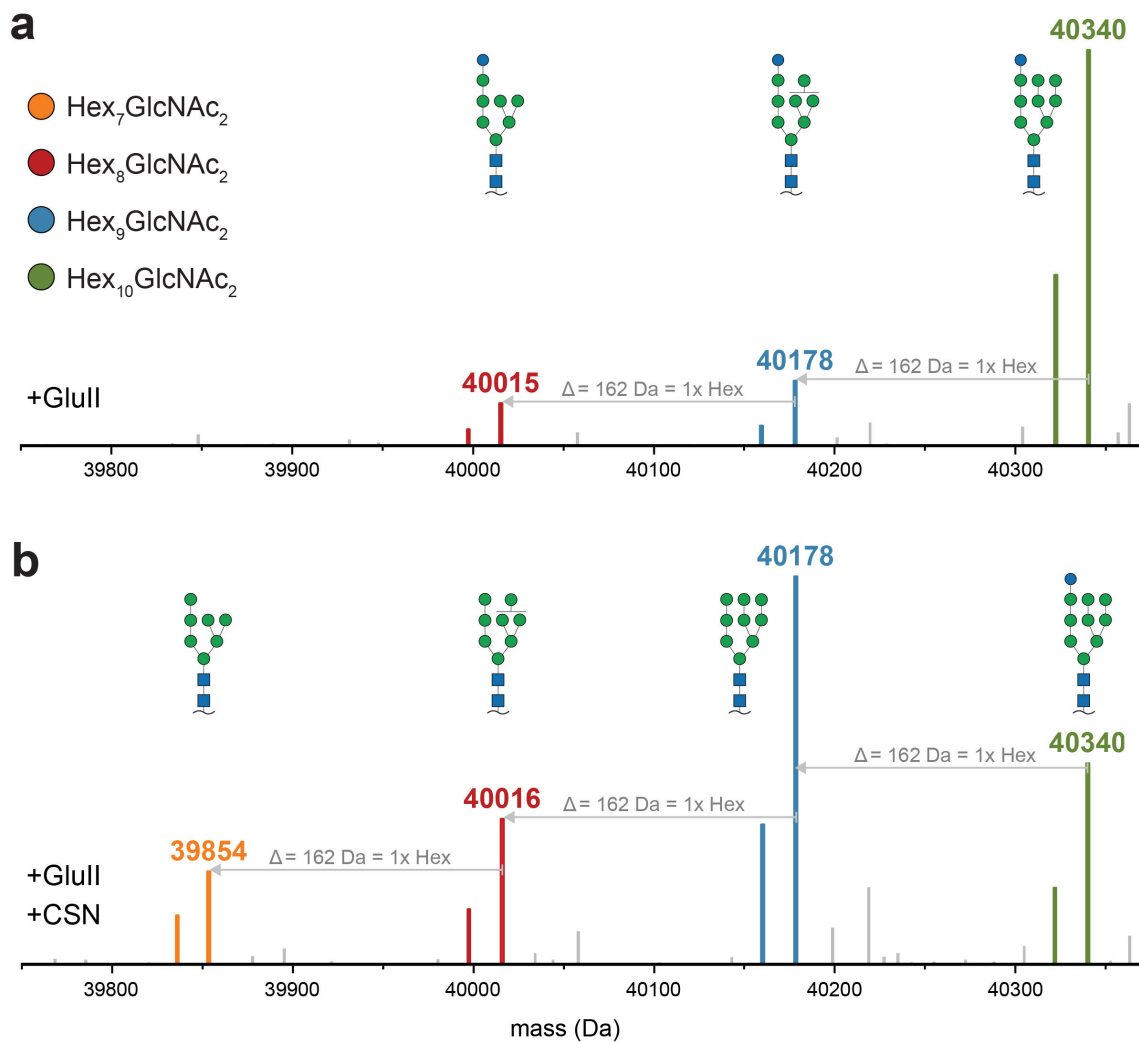
Supplementary Fig. 7 PLC-associated MHC I carries a terminal glucose in the A-branch. The minor MHC I species (Fig. 1 and Supplementary Fig. 2) carrying an N-linked glycan lack one or two mannose moieties at the B or C branches of the N-glycan tree, but still carry a terminal glucose on the A branch because GluII processing shifted the entire glycosylation pattern to lower molecular weight (MW). Deconvoluted MS spectra of HLA-A*03:01 are shown. **a** In the absence of peptides, no processing by GluII is detectable. The major HLA-A*03:01 fraction carries the $Glc_1Man_9GlcNAc_2$ glycan. A minor fraction of HLA-A*03:01 species harbor glycans with only nine (blue) or eight hexoses (red). **b** In the presence of high-affinity peptides, the pattern is shifted to lower MW and a new species with seven hexoses (orange) appears. The lower intensity mass peaks, lacking 18 Da in comparison to the labeled mass peaks, are a common artifact of rapid thermal decomposition inside the ESI-MS source,[46] which appeared to be more pronounced in glycoproteins.

Supplementary Table 1 | Subunits of the PLC identified by LC-MS

Component (reference)	AA Sequence (mature protein)				M_{cal} (Da)	M_{obs} (Da)
β₂m (Uniprot: P61769)	IQRTPKIQVY KNGERIEKVE RVNHVTLSQP	SRHPAENGKS HSDLSFSKDW KIVKWDRDM	NFLNCYVSGF SFYLLYYTEF	HPSDIEVDLL TPTEKDEYAC	11729.01	11729.44
calreticulin (Uniprot: P27797)	EPAVYFKEQF GDEEKDKGLQ KHEQNIDCGG GPGTKKVHVI PDNTYEVKID DWDERAKIDD DGEWEPPVIQ EYSPDPSIYA YAEEFGNETW EEEEAEDKED	LDGDGWTSRW TSQDARFYAL GYVKLFPNSL FNYKGKNVLI NSQVESGSLE PTDSKPEDWD NPEYKGEWKP YDNFGVLGLD GVTKAAEKQM DEDKDEDEED	IESKHKSDFG SASFEPFSNK DQTDMHGDSE NKDIRCKDDE DDWDFLPPKK KPEHIPDPDA RQIDNPDYKG LWQVKSGTIF KDKQDEEQRL EEDKEEDEEE	KFVLSSGKFY GQTLVVQFTV YNIMFGPDIC FTHLYTLIVR IKDPDASKPE KKPEDWDEEM TWIHPEIDNP DNFLITNDEA KEEEEDKKRK DVPGQAKDEL	46463.77	46464.76
ERp57 (Uniprot: P30101)	SDVLELTDDN PEYEAAATRL IFRDGEEAGA FKKFISDKDA AHTNVESLVN QKMTSGKIKK YDVDYEKNAK KTFSHELSDF DGKALERFLQ NFDEIVNNEN KDPNIVIAKM KKYEGGRELS	FESRISDTGS KGIVPLAKVD YDGPRTADGI SIVGFFDDSF EYDDNGEGII FIQENIFGIC GSNYWRNRVM GLESTAGEIP DYFDGNLKRY KDVLIEFYAP DATANDVPSP DFISYLQREA	AGLMLVEFFA CTANTNTCNK VSHLKKQAGP SEAHSEFLKA LFRPSHLTNK PHMTEDNKDL MVAKKFLDAG VVAIRTAKGE LKSEPIPESN WCGHCKNLEP YEVRGFPTIY TNPPVIQEEK	PWCGHCKRLA YGVSGYPTLK ASVPLRTEEE ASNLRDNYRF FEDKTVAYTE IQGKDLLIAY HKLNFAVASR KFVMQEEFSR DGPVKVVVAE KYKELGEKLS FSPANKKLNP PKKKKKAQED L	54260.53	54260.27
HLA-A*03:01 (NCBI: NP_002107.3)	GSHSMRYFFT ASQRMEPRAP LRGYYNQSEA KDYIALNEDL DGTCVEWLRR LRCWALGFYP FQKWAAVVVP TIPIVGIIAG TQAASSDSAQ	SVSRPGRGEP WIEQEGPEYW GSHTIQIMYG RSWTAADMAA YLENGKETLQ AEITLTWQRD SGEEQRYTCH LVLLGAVITG GSDVSLTACK	RFIAVGYVDD DQETRNVKAQ CDVGSDGRFL QITKRKWEAA RTDPPKTHMT GEDQTQDTEL VQHEGLPKPL AVVAAVMWRR V	TQFVRFDSDA SQTDRVDLGT RGYRQDAYDG HEAEQLRAYL HHPISDHEAT VETRPAGDGT TLRWELSSQP KSSDRKGGSY	38312.12	38313.27
tapasin Iso1 (Uniprot: O15533)	GPAVIECWFV DPELYLSVHD ASAKWASGLT PQPEPQQEPV SFAYMPPTSE TPGLNGQMPA VQPFQEGTYL LARAAPGEAP QKAEGQRWLS ACRIHHPSLP FLLLGLFKAL	EDASGKGLAK PAGALQAAFR PAQNCPRALD LITMATVVLT AASSLAPGPP AQEGAVAFAA ATIHLPYLQG PELLCLVSHF ALRHHSDGSV ASGRSAEVTL GWAAVYLSTC	RPGALLLRQG RYPRGAPAPH GAWLMVSISS VLTHTPAPRV PFGLEWRRQH WDDDEPWGPW QVTLELAVYK YPSGGLEVEW SLSGHLQPPP EVAGLSGPSL KDSKKKAE	PGEPPPRPDL CEMSRFVPLP PVLSLSSLLR RLGQDALLDL LGKGHLLLAA TGNGTFWLPR PPKVSLMPAT ELRGGPGGRS VTTEQHGARY EDSVGLFLSA	45711.82	45710.70
TAP2 – Iso1 A665T[1] (NCBI: NP_000535.3)	MRLPDLRPWT EGTLRLGGLW ALVAGASRAP PPGAQEKEQD LAVLGETLIP FGSSLSAGCR FQETKTGELN LYGFMLSISP IQDAVARAGQ QCRQLYWRRD DGELTQGSLL EKVFSYMDRQ	SLLLVDAALL GLLKLRGLLG PARVASAPWS QVNNKVLMWR HYSGRVIDIL GGCFTYTMSR SRLSSDTTLM RLTLLSLLHM VVREAVGGLQ LERALYLLVR SFMIYQESVG PNLPSPGTLA	WLLQGPLGTL FVGTLLLPLC WLLVGYGAAG LLKLSRPDLP GGDFDPHAFA INLRIREQLF SNWLPLNANV PFTIAAEKVY TVRSFGAEEH RVLHLGVQML SYVQTLVYIY PTTLQGVVKF	LPQGLPGLWL LATPLTVSLR LSWSLWAVLS LLVAAFFFLV SAIFFMCLFS SSLLRQDLGF LLRSLVKVVG NTRHQEVLRE EVCRYKEALE MLSCGLQQMQ GDMLSNVGAA QDVSFAYPNR	77724.23	77727.10

9

```
                  PDRPVLKGLT FTLRPGEVTA LVGPNGSGKS TVAALLQNLY
                  QPTGGQVLLD EKPISQYEHC YLHSQVVSVG QEPVLFSGSV
                  RNNIAYGLQS CEDDKVMAAA QAAHADDFIQ EMEHGIYTDV
                  GEKGSQLAAG QKQRLAIARA LVRDPRVLIL DEATSALDVQ
                  CEQALQDWNS RGDRTVLVIA HRLQTVQRAH QILVLQEGKL
                  QKLAQLQEGQ DLYSRLVQQR LMD
```

TAP2 – Iso3 (NCBI: NP_001276972.1)

```
MRLPDLRPWT SLLLVDAALL WLLQGPLGTL LPQGLPGLWL
EGTLRLGGLW GLLKLRGLLG FVGTLLLPLC LATPLTVSLR
ALVAGASRAP PARVASAPWS WLLVGYGAAG LSWSLWAVLS
PPGAQEKEQD QVNNKVLMWR LLKLSRPDLP LLVAAFFFLV
LAVLGETLIP HYSGRVIDIL GGDFDPHAFA SAIFFMCLFS
FGSSLSAGCR GGCFTYTMSR INLRIREQLF SSLLRQDLGF
FQETKTGELN SRLSSDTTLM SNWLPLNANV LLRSLVKVVG
LYGFMLSISP RLTLLSLLHM PFTIAAEKVY NTRHQEVLRE
IQDAVARAGQ VVREAVGGLQ TVRSFGAEEH EVCRYKEALE
QCRQLYWRRD LERALYLLVR RVLHLGVQML MLSCGLQQMQ
DGELTQGSLL SFMIYQESVG SYVQTLVYIY GDMLSNVGAA
EKVFSYMDRQ PNLPSPGTLA PTTLQGVVKF QDVSFAYPNR
PDRPVLKGLT FTLRPGEVTA LVGPNGSGKS TVAALLQNLY
QPTGGQVLLD EKPISQYEHC YLHSQVVSVG QEPVLFSGSV
RNNIAYGLQS CEDDKVMAAA QAAHADDFIQ EMEHGIYTDV
GEKGSQLAAG QKQRLAIARA LVRDPRVLIL DEATSALDVQ
CEQALQDWNS RGDRTVLVIA HRLQTVQRAH QILVLQEGKL
QKLAQL
```
75662.95 | 75665.80

TAP1 - Iso1 (NCBI: NP_000584)

```
ASSRCPAPRG CRCLPGASLA WLGTVLLLLA DWVLLRTALP
RIFSLLVPTA LPLLRVWAVG LSRWAVLWLG ACGVLRATVG
SKSENAGAQG WLAALKPLAA ALGLALPGLA LFRELISWGA
PGSADSTRLL HWGSHPTAFV VSYAAALPAA ALWHKLGSLW
VPGGQGGSGN PVRRLLGCLG SETRRLSLFL VLVVLSSLGE
MAIPFFTGRL TDWILQDGSA DTFTRNLTLM SILTIASAVL
EFVGDGIYNN TMGHVHSHLQ GEVFGAVLRQ ETEFFQQNQT
GNIMSRVTED TSTLSDSLSE NLSLFLWYLV RGLCLLGIML
WGSVSLTMVT LITLPLLFLL PKKVGKWYQL LEVQVRESLA
KSSQVAIEAL SAMPTVRSFA NEEGEAQKFR EKLQEIKTLN
QKEAVAYAVN SWTTSISGML LKVGILYIGG QLVTSGAVSS
GNLVTFVLYQ MQFTQAVEVL LSIYPRVQKA VGSSEKIFEY
LDRTPRCPPS GLLTPLHLEG LVQFQDVSFA YPNRPDVLVL
QGLTFTLRPG EVTALVGPNG SGKSTVAALL QNLYQPTGGQ
LLLDGKPLPQ YEHRYLHRQV AAVGQEPQVF GRSLQENIAY
GLTQKPTMEE ITAAAVKSGA HSFISGLPQG YDTEVDEAGS
QLSGGQRQAV ALARALIRKP CVLILDDATS ALDANSQLQV
EQLLYESPER YSRSVLLITQ HLSLVEQADH ILFLEGGAIR
EGGTHQQLME KKGCYWAMVQ APADAPE
```
80832.50 | 80832.60

ICP47^SBP
^GAM^ICP47^SBP^

```
GAMASWALEM ADTFLDNMRV GPRTYADVRD EINKRGREDR
EAARTAVHDP ERPLLRSPGL LPEIAPNASL GVAHRRTGGT
VTDSPRNPVT RGSGGGSGGG SMDEKTTGWR GGHVVEGLAG
ELEQLRARLE HHPQGQREP
```
14953.39 | 14953.80
14694.07 | 14694.75

MSP2N2

```
GHHHHHHHDY DIPTTENLYF QGSTFSKLRE QLGPVTQEFW
DNLEKETEGL RQEMSKDLEE VKAKVQPYLD DFQKKWQEEM
ELYRQKVEPL RAELQEGARQ KLHELQEKLS PLGEEMRDRA
RAHVDALRTH LAPYSDELRQ RLAARLEALK ENGGARLAEY
HAKATEHLST LSEKAKPALE DLRQGLLPVL ESFKVSFLSA
LEEYTKKLNT QGTPVTQEFW DNLEKETEGL RQEMSKDLEE
VKAKVQPYLD DFQKKWQEEM ELYRQKVEPL RAELQEGARQ
KLHELQEKLS PLGEEMRDRA RAHVDALRTH LAPYSDELRQ
RLAARLEALK ENGGARLAEY HAKATEHLST LSEKAKPALE
DLRQGLLPVL ESFKVSFLSA LEEYTKKLNT Q
```
45540.70 | 45541.24

Supplementary Table 2 Cryo-EM data collection, refinement, and validation statistics

	PLC Editing Module Calreticulin-tapasin-ERp57-MHC I hc-β_2m (EMDB-14119) (PDB ID 7QPD)
Data collection and processing	
Microscope	TITAN-Krios
Camera	Gatan K2 Summit
Magnification	130,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	68
Defocus range (μm)	-1.0 to -2.5
Pixel size (Å)	1.05
Symmetry imposed	C1
Initial particle images (no.)	613,746
Final particle images (no.)	97,952
Map resolution (Å)	3.73
FSC threshold	0.143
Refinement	
Map sharpening B factor (Å^2)	-81.1
Model composition	
Non-hydrogen atoms	11,521
Protein residues	1544
Carbohydrate	10
B factors (Å^2)	
Protein	63.89
Carbohydrate	87.35
R.m.s. deviations	
Bond lengths (Å)	0.0047
Bond angles (°)	0.78
Validation	
MolProbity score	1.85
Clashscore	9.50
Poor rotamers (%)	0.09
Carbohydrates[1]	
Stereochemical problems	0
Unphysical puckering amplitude	0
In unlikely ring conformation	0
Ramachandran plot	
Favored (%)	95.01
Allowed (%)	4.92
Disallowed (%)	0.07

[1]As reported by the Privateer software package

Article

Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity

Ines Katharina Müller [1], Christian Winter [1], Christoph Thomas [1], Robbert M. Spaapen[2,3], Simon Trowitzsch [1] ✉ & Robert Tampé [1] ✉

Adaptive immunity depends on cell surface presentation of antigenic peptides by major histocompatibility complex class I (MHC I) molecules and on stringent ER quality control in the secretory pathway. The chaperone tapasin in conjunction with the oxidoreductase ERp57 is crucial for MHC I assembly and for shaping the epitope repertoire for high immunogenicity. However, how the tapasin–ERp57 complex engages MHC I clients has not yet been determined at atomic detail. Here, we present the 2.7-Å crystal structure of a tapasin–ERp57 heterodimer in complex with peptide-receptive MHC I. Our study unveils molecular details of client recognition by the multichaperone complex and highlights elements indispensable for peptide proofreading. The structure of this transient ER quality control complex provides the mechanistic basis for the selector function of tapasin and showcases how the numerous MHC I allomorphs are chaperoned during peptide loading and editing.

The presentation of antigenic peptides by major histocompatibility complex class I (MHC I) molecules on the cell surface ensures the detection and elimination of infected or cancerous cells by the adaptive immune system. Any misinterpretation can lead to inflammatory and autoimmune diseases. Cytotoxic T lymphocytes, which are licensed to kill, search for non-self peptides presented by MHC I[1,2]. MHC I complexes are composed of a heavy chain (hc) and the light chain β_2-microglobulin (β_2m). Encoded by the most polymorphic gene cluster in humans with more than 24,000 known classical human leukocyte antigen (HLA)-A, -B, and -C alleles, all heavy chains invariantly share their association with β_2m. Prior to presenting peptides on the cell surface, MHC I molecules undergo a stringent quality control during their maturation process in the endoplasmic reticulum (ER) by transiently interacting with specialized proteins, including glycan sensors, disulfide isomerases, and specific MHC I chaperones[3]. These ER chaperone systems play a pivotal role in adaptive immunity not only by stabilizing folding intermediates but also by shaping the peptide repertoire displayed on MHC I molecules.

The two MHC I-specific chaperones, tapasin and TAP-binding protein related (TAPBPR), were shown to share similar binding interfaces on MHC I and were thus suggested to have common catalytic

principles[4–8]. However, the two chaperones differ substantially in their subcellular localization and their molecular liaisons. TAPBPR acts autonomously on a subset of MHC I clients in the peptide-depleted environments of the *cis*-Golgi and ER-Golgi intermediate compartment (ERGIC)[9]. As the major MHC I editor, tapasin fulfils its function predominantly as part of the peptide-loading complex (PLC) in conjunction with the oxidoreductase ERp57 and the lectin-like chaperone calreticulin in the ER[3,8]. The interaction between tapasin and ERp57 relies on a mixed disulfide between Cys[95] of tapasin and Cys[33] of ERp57 (Cys[57] in immature ERp57)[4]. In the oxidizing environment of the ER, Cys[33] of ERp57 toggles between an intermolecular, mixed disulfide with Cys[95] of tapasin and an intramolecular disulfide with Cys[36] (Cys[60] in immature ERp57)[10]. This toggling, known as the escape pathway, has been described for protein-disulfide isomerase-assisted protein folding and prevents the enzyme from becoming trapped in covalent complexes with its substrates[11].

Peptide-deficient MHC I molecules are recruited to the PLC[12,13], undergo peptide loading and editing[14,15], dissociate from the PLC as stable peptide–MHC I (pMHC I) complexes, and traffic to the cell surface for antigen presentation[3]. During peptide loading and editing, tapasin stabilizes peptide-receptive MHC I heterodimers

and catalyzes peptide exchange by accelerating peptide association and dissociation[15–17]. Deletion of tapasin and ERp57 results in an impaired immune response and altered peptide repertoire[18,19]. The various MHC I allomorphs exhibit distinct levels of plasticity, differ in peptide-binding preferences, and depend to different degrees on tapasin for optimal peptide loading[20,21]. Conformational plasticity of MHC I scales with the ability to select high-affinity peptides in the absence of tapasin[22–25]. Insights into the molecular basis of tapasin promiscuity for different MHC I clients as well as structural elements necessary for peptide exchange catalysis at atomic resolution are lacking due to the intrinsically transient nature of the editing complex.

Here, we present the 2.7-Å crystal structure of an MHC I–tapasin–ERp57 editing complex assembled by a photo-triggered approach. The structure of this transient assembly reveals molecular rearrangements within the chaperone complex upon client engagement. Movement of the C-terminal immunoglobulin domain of tapasin extends the interaction interface to MHC I hc and β_2m. We show by functional assays that structure-guided ablation of interface interactions leads to reduced cell surface presentation of pMHC I complexes. The editing loop of tapasin rigidifies upon client recognition and contributes to a widened MHC I F-pocket. Since the loop structure might also stabilize a peptide-bound MHC I pocket, our data suggest a dual function of the editing loop in peptide exchange catalysis. Our structural and functional data reveal the molecular basis of tapasin promiscuity towards MHC I clients and suggest that the level of

plasticity of peptide-receptive MHC I determines chaperone-assisted acquisition of high-affinity peptides.

Results and discussion

Structure of the MHC I–tapasin–ERp57 editing complex

We captured a human tapasin–ERp57 heterodimer with bound MHC I hc and β_2m *in flagrant*e by a photo-triggered approach[7] and determined the crystal structure of the MHC I·chaperone complex to 2.7 Å resolution (Fig. 1). To facilitate crystallization, we utilized the ectodomains of tapasin and MHC I hc lacking the transmembrane and cytosolic regions. After photo-cleavage of loaded peptides, we first screened different MHC I allomorphs for complex formation with the tapasin–ERp57 heterodimer by size exclusion chromatography (SEC), resulting in murine H2-D^b as the most suited MHC I hc candidate (Fig. 1a, b, and Supplementary Fig. 1a–d). Furthermore, we substituted Cys^{36} of ERp57 with Ala (C36A), formerly described as C60A in immature ERp57 (ref. 4), to retain the mixed disulfide between tapasin and ERp57 (ref. 10).

Since a homogeneous population of a kinetically stable chaperone-client complex between disulfide-linked tapasin–ERp57 and refolded MHC I hc/β_2m could not be isolated (Fig. 1a, b), we directly used a photoactivated mix of refolded photo-peptide/H2-D^b/β_2m and tapasin–ERp57 (1.3: 1 molar ratio) for crystallization. Size exclusion chromatography-coupled mass spectrometry (SEC-MS) revealed an almost complete photo-induced cleavage (>96%) of the peptide bound to MHC I (Supplementary Fig. 2). Initial crystals of the
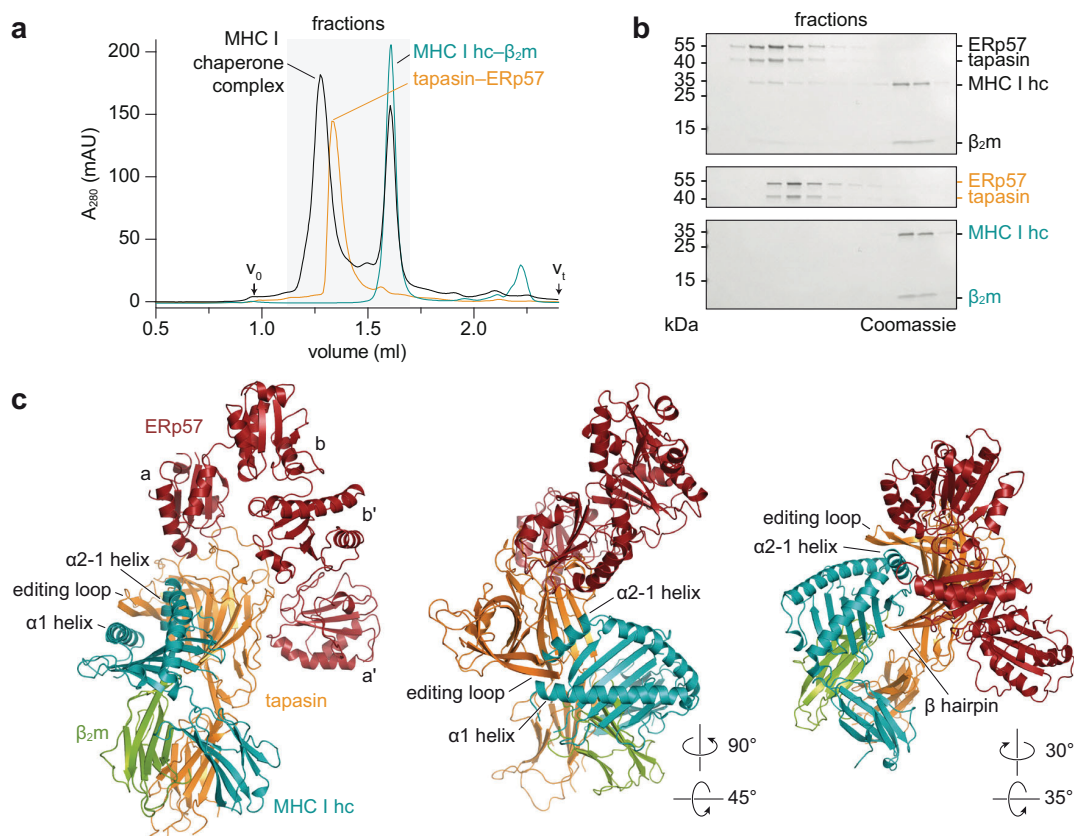
Fig. 1 | Photo-triggered assembly and structural overview of the MHC I–tapasin–ERp57 complex. a Photo-triggered assembly of a transient MHC I–tapasin–ERp57 complex (black) analyzed by size exclusion chromatography (SEC). SEC elution profiles of tapasin–ERp57 and MHC I hc–β_2m are colored in orange and teal, respectively. β_2m β_2-microglobulin; hc heavy chain; A$_{280}$ absorption at 280 nm; V$_0$ void volume; V$_t$ total volume. **b** MHC I–tapasin–ERp57 complex formation monitored by SEC and SDS-PAGE. kDa kilodalton; $n = 1$. Same color code as illustrated in **a** is used. **c** Cartoon representation of peptide-receptive MHC I (MHC I hc, teal; β_2m, green) in complex with the ER chaperone tapasin–ERp57 (tapasin, orange; ERp57, red) in different orientations. a, b, b′, and a′, thioredoxin-like domains of ERp57. Source data for **a** and (**b**) are provided as a Source Data file.
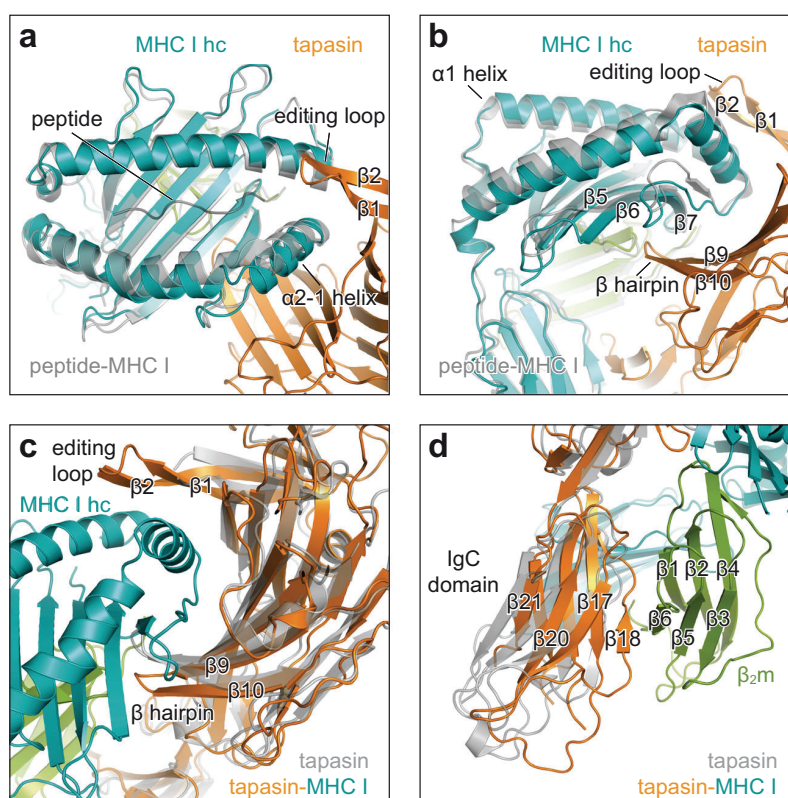
Fig. 2 | Conformational changes upon client-chaperone engagement. a Top view of superpositing peptide-receptive MHC I in the chaperone complex (MHC I hc, teal; tapasin, orange) with peptide-bound H2-D^b (gray, PDB ID 2F74) in cartoon representation. hc heavy chain; β1, β2, β strands of editing loop of tapasin. **b** Side view of the superposition shown in **a** β5, β6, β7, β strands of MHC I hc; β9, β10, β hairpin of tapasin. **c** View onto the concave interface of tapasin in superposition of MHC I-bound tapasin from the chaperone complex (tapasin, orange; MHC I hc, teal; β$_2$m, green) with client-free tapasin (gray, PDB ID 3F8U). **d** Superposition as in **c** but viewed onto the C-terminal domain of tapasin. IgC domain immunoglobulin constant domain, β$_2$m β$_2$-microglobulin.

chaperone-client complex belonging to space group P2$_1$2$_1$2$_1$ diffracted X-rays to 4.0 Å resolution with two heterotetrameric editing complexes in the asymmetric unit. Utilizing microseed matrix screening[26], we obtained crystals in space group P22$_1$2$_1$ with an improved resolution of 2.7 Å and only one complex in the asymmetric unit, enabling us to build an atomic model with good stereochemistry (Fig. 1c and Supplementary Table 1).

As in the client-free structure of tapasin–ERp57 (ref. 4), the four thioredoxin-like domains of ERp57 (a-b-b′-a′) are arranged in their characteristic twisted U form, whereas tapasin forms its typical L-shaped structure[4,8] (Fig. 1c and Supplementary Movie 1). The a and a′ domains of ERp57 interact with the N-terminal region of tapasin composed of a fusion between a seven-stranded β barrel and an immunoglobin (Ig)-like fold[4]. In contrast to the structure of client-free tapasin–ERp57 (ref. 4), we could not model the mixed disulfide between Cys[95] of tapasin and Cys[33] of ERp57 present in the purified tapasin–ERp57 heterodimer (Supplementary Fig. 1). Since disulfide bond reduction did not result from UV illumination performed for photocleavage of the peptide (Supplementary Fig. 1), we speculate that reduction is either facilitated by X-ray irradiation[27] or resulted from the MHC I editing process itself.

Molecular rearrangements within the chaperone complex upon client engagement

The interface between tapasin and MHC I covers a total surface area of 4381 Å^2 (hc: 74%; β$_2$m: 26%) (Supplementary Fig. 3) of which the N-terminal domain of tapasin contributes 2657 Å^2. The concave binding site of the N-terminal domain embeds the α2-1 helix region of MHC I hc like a clamp (Fig. 1c). On one side of the clamp, a loop comprising

residues Glu[11]-Lys[20], referred to as editing loop in the following, sits on top of the F-pocket of the MHC I peptide-binding groove, whereas a β hairpin element contacts the floor of the peptide-binding groove from below (Fig. 2a–c). When superimposed with atomic models of client-free tapasin–ERp57 (ref. 4) and peptide-bound MHC I (ref. 28), our structure reveals that client binding leads to important rearrangements of the α1 and α2-1 helices in MHC I, resulting in a widened F-pocket (Fig. 2a). This widening is stabilized by a hydrogen-bond network between His[70] and Glu[72] of tapasin, which attracts the side chain of Tyr[84] of the MHC I hc (Fig. 3a, b). The side chain of Lys[20] of tapasin thereby facilitates positioning of Tyr[84] by hydrogen bonding with the main-chain oxygen (Fig. 3b). Notably, Glu[72] of tapasin structurally corresponds to Glu[105] of TAPBPR, which was found to contact Tyr[84] of MHC I hc in a similar manner[6,7]. The β-sheet floor (β-strands β5-7) of the peptide-binding groove near the F-pocket is distorted upon binding, presumably by engagement with the β hairpin of tapasin (β strands β9-10) (Fig. 2b). Upon client binding, the β hairpin of tapasin is shifted downwards to accommodate the α2-1 region of MHC I (Fig. 2c) in a similar way as revealed in TAPBPR–MHC I structures[6,7]. Furthermore, the C-terminal Ig-like domain of tapasin swings towards the α3 domain of the MHC I hc and β$_2$m (Fig. 2d), allowing for an additional interaction interface with a buried surface area of 1724 Å^2 (hc: 55%; β$_2$m: 45%; Supplementary Fig. 3a, b).

The editing loop of tapasin contributes to a widened MHC I F-pocket

Upon client binding, the intrinsically disordered editing loop of the client-free state of tapasin[4,29] rigidifies as evidenced by a clear electron density at the 1.5 σ level for most loop residues (Fig. 3a–d). The editing
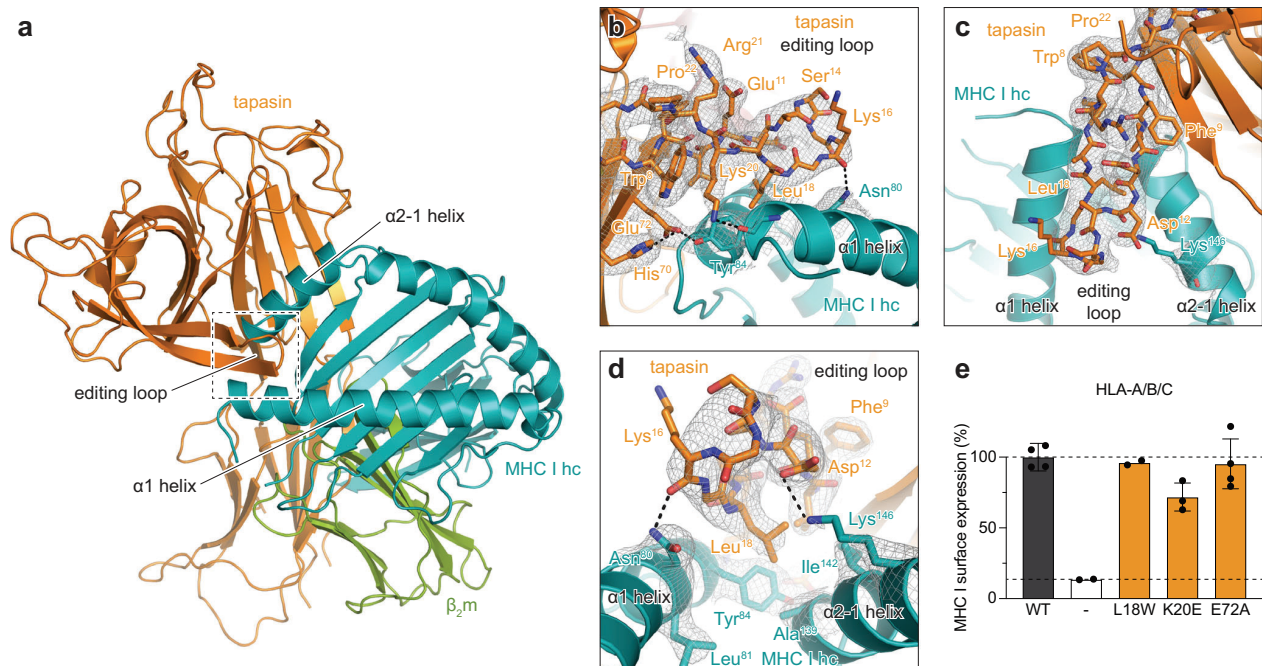
Fig. 3 | The editing loop of tapasin contributes to a widened F-pocket of MHC I.
a Cartoon representation of the chaperone complex as shown in Fig. 1c. Tapasin, MHC I hc and β_2m are colored in orange, teal and green, respectively. Region magnified in **b**−**d** is indicated by a dashed box. ERp57 is not shown for the sake of clarity. **b** Side view onto the editing loop (stick representation) of tapasin with corresponding electron density map (contour level: 1.5 σ). Dashed lines represent hydrogen bonds or salt bridges. The same color code as in **a** is applied. hc heavy chain. **c** Top view onto the editing loop with corresponding electron density and color code as shown in **b**. **d** Close-up view into the F-pocket of MHC I with corresponding electron density and color code as displayed in **b**. **e** MHC I surface expression of tapasin-deficient HAP1 cells (white), expressing wildtype (WT, dark

gray) or different interaction mutants of tapasin (orange). Flow cytometry was performed using an APC-conjugated pan-HLA-A/B/C-specific antibody (W6/32). The lower black dashed line represents the level of MHC I surface expression of cells transfected with a vector devoid of gene of interest. The upper black dashed line represents the level of MHC I surface expression of cells expressing wildtype tapasin. The mean fluorescence intensity of MHC I surface expression was normalized to wildtype tapasin ($\pm$SD; $n = 4$ biologically independent samples; K20E, $n = 3$ biologically independent samples; L18W and mock transfected, $n = 2$ biologically independent samples). The gating strategy is displayed in Supplementary Fig. 8. WT wildtype, − mock transfected. Source data for **e** are provided as a Source Data file.

loop is stabilized by the side chain of Asn[80] of MHC I, which forms a hydrogen bond with the backbone-carbonyl oxygen of Lys[16] of tapasin (Fig. 3d). The side chain of Asp[12] of tapasin is in salt-bridge distance to the ε-amino group of Lys[146] of the MHC I hc that presumably contributes to the widening of the peptide-binding groove and the rigidification of the editing loop of tapasin (Fig. 3d). Furthermore, the side chain of Leu[18] of the editing loop, which was shown to be relevant for MHC I surface expression[30,31], compensates in the peptide-free MHC I F-pocket for the lack of a C-terminal hydrophobic side chain of the antigenic peptide and presumably assists in groove widening. Even though sitting directly above the F-pocket, Leu[18] does not mimic the C terminus of cargo peptides. Thus, Leu[18] is unlikely to act as a peptide surrogate per se[30]. In line with this groove-widening function, substitution of Leu[18] with a bulky hydrophobic side chain (tryptophan) did not negatively impact MHC I surface presentation in our flow cytometry-based cellular assays when compared to wildtype (WT) (Fig. 3e and Supplementary Fig. 3c). Conversely, substitution of Leu[18] with glycine has been shown to nearly abolish peptide exchange activity of tapasin[31].

Surprisingly, abrogation of the hydrogen bond between Tyr[84] of MHC I and Glu[72] of tapasin did not affect MHC I surface presentation (Fig. 3e), as demonstrated by an in vitro peptide-loading assay[4]. Introducing a charge repulsion by mutating tapasin Lys[20], which helps in positioning the α1 helix of MHC I via hydrogen bonding to the backbone-carbonyl oxygen of Tyr[84] (Fig. 3a, b), did not substantially decrease the surface MHC I level ($72 \pm 10\%$ of WT; Fig. 3e), suggesting that the editing loop of tapasin mainly stabilizes the empty F-pocket by a hydrophobic plug. At the same time, MHC I F-pocket residues Asn[80],

Tyr[84], and Lys[146], which coordinate the terminal carboxyl group of bound peptides, are engaged in interactions with tapasin and its editing loop (Fig. 3a, b and Supplementary Fig. 4a). Our data suggest that tapasin acts in a dual manner, opening the peptide-binding groove and stabilizing the empty F-pocket of the MHC I hc.

Domain movements in tapasin extend the interaction interface with MHC I

The interface loop region of tapasin on the lateral side of the MHC I α2-1 helix[4,7] is not involved in specific interactions and therefore only plays a minor role in MHC I surface presentation (Fig. 4a). In contrast, the neighboring Gln[261] of tapasin, which is involved in several hydrogen bonds with the α2-1 helix region, leads to a reduction of MHC I surface presentation to $65 \pm 6\%$ when mutated to alanine (Fig. 4b, f and Supplementary Fig. 3d). An even stronger suppression of surface presentation was observed when interactions between the β hairpin of tapasin and the α2 domain of MHC I hc were disrupted (R187E; $55 \pm 13\%$ compared to WT) (Fig. 4c, f). Mutations in both the interface loop and the β hairpin synergistically reduced MHC I surface presentation to $46 \pm 4\%$ (Fig. 4f). Interestingly, as in free tapasin−ERp57 (ref. 4), we observed an intermolecular contact between side chains of Arg[333] of tapasin and Glu[229] of MHC I hc (Fig. 4d), which, when disturbed by charge repulsion, reduced MHC I surface presentation to $37 \pm 3\%$ (Fig. 4f). The importance of this interaction between the CD8-recognition loop of MHC I and the IgC domain of tapasin was already proposed by the cryogenic electron microscopy structure of the PLC[8]. Disrupting the hydrogen bond between Ser[336] of tapasin and Thr[225] of MHC I hc distant from the β-sheet interaction[32] had almost no
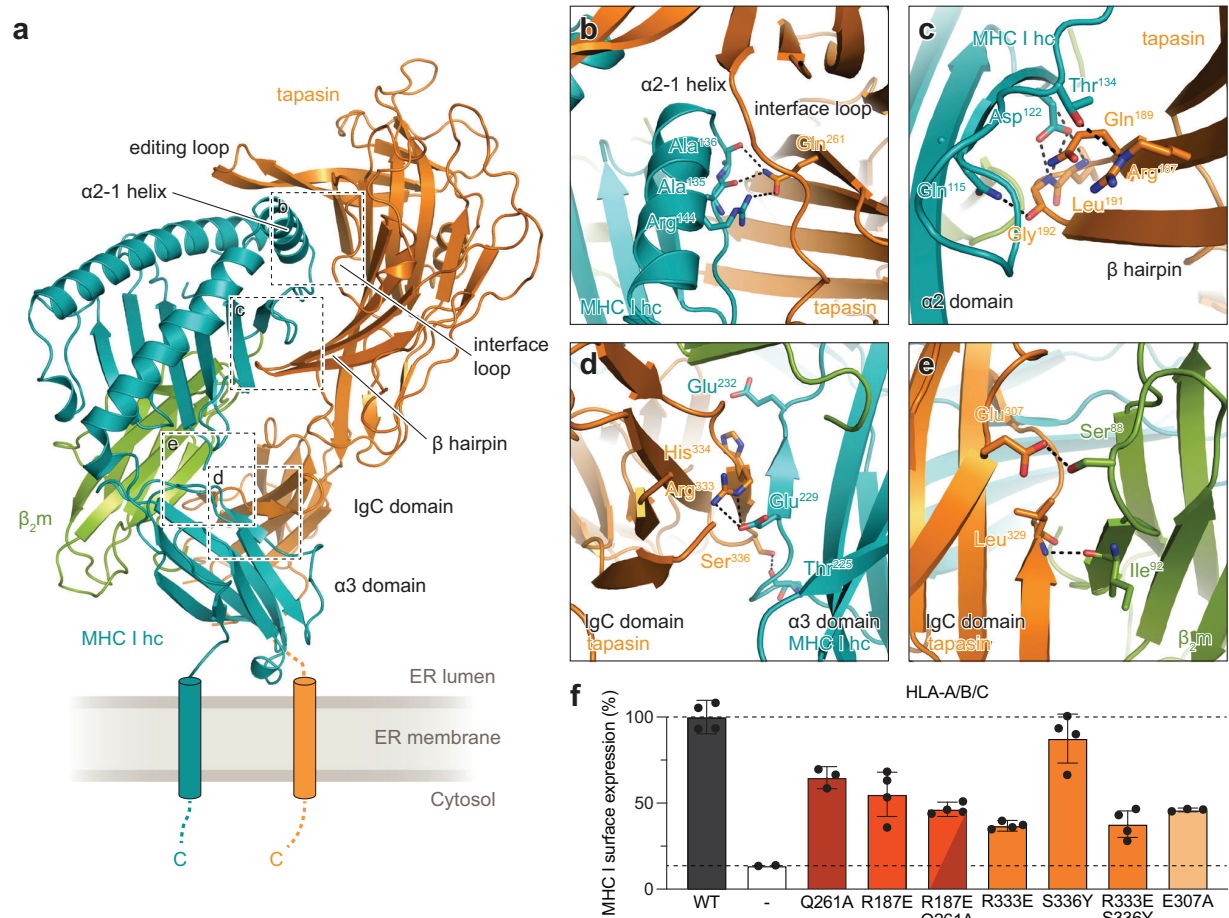
Fig. 4 | Molecular interactions at chaperone-client interfaces affect MHC I surface expression. a Cartoon representation of the chaperone complex as shown in Fig. 1c. The components tapasin, MHC I hc and β_2m are displayed in orange, teal and green, respectively. Regions magnified in **b**–**e** are indicated by dashed boxes. ERp57 is not shown for the sake of clarity. Transmembrane helices and C-terminal regions (C) of MHC I hc and tapasin are indicated schematically. **b** Zoom-in of the α2-1 helix region of MHC I and tapasin. Interface residues forming hydrogen bonds (black dashed lines) are shown as sticks. **c** Magnification of the interface formed between the β hairpin of tapasin and the α1/α2 domains of MHC I. **d** Close-up view of the C-terminal domain of tapasin and the α3 domain of MHC I. **e** Interface formed between β_2m and the IgC domain of tapasin. **f** Restored MHC I surface expression of tapasin-deficient HAP1 cells by interface mutants of tapasin as assessed by flow cytometry. Mean fluorescence intensities of cells stained with APC-conjugated pan-HLA-A/B/C-specific antibody (W6/32) were normalized to the tapasin amount of the wildtype (dark gray, upper dashed line) (±SD, $n = 4$ biologically independent samples; Q261A, R333E/S336Y, and E307A, $n = 3$ biologically independent samples; mock transfected, $n = 2$ biologically independent samples). The lower dashed line represents the level of MHC I surface expression of mock-transfected cells (white). The gating strategy is displayed in Supplementary Fig. 8. α2-1 interface, brown; β-hairpin interface, dark orange; α3 interface, orange; β_2m interface, light orange. WT wildtype, − mock transfected. Source data for **e** are provided as a Source Data file. The same color code as used in **a** is applied in **b**–**e**.

impact on surface presentation (Fig. 4d, f). The displaced IgC domain of tapasin is further stabilized by contacts to β_2m, involving backbone interactions between Ile[92] of β_2m and Leu[329] of tapasin and a side chain hydrogen bond between Ser[88] of β_2m and Glu[307] of tapasin (Fig. 4e). Disruption of the hydrogen bond between Glu[307] of tapasin and Ser[88] of β_2m reduced the MHC I surface level to $46 \pm 1\%$ (Fig. 4e, f); albeit the expression level of mutant tapasin was also lowered compared to WT (Supplementary Fig. 3e).

Dual function of the editing loop in peptide exchange catalysis
Mechanistic studies suggested that the MHC I-specific chaperones tapasin and TAPBPR hold distinct preferences for MHC I allomorphs[33–35]. However, our structure of the MHC I–tapasin–ERp57 editing complex demonstrates that the binding interfaces used for client recognition are similar between the two editors with minor differences for specific residues (Supplementary Figs. 5 and 6). Comparing the impact of the two editors on widening the peptide-binding cleft reveals a slightly stronger displacement of the α2-1 helix of MHC I

when associated with TAPBPR (Supplementary Fig. 4b). In our structure, the editing loop lies on top of the MHC I F-pocket (Fig. 3). In chaperone-free pMHC I, the side chains of residues Asn[80] and Lys[146] point into the F-pocket and coordinate the C terminus of bound peptides (Supplementary Fig. 4a). In an intermediate peptide-/chaperone-bound state the carbonyl group of Gly[15] of the editing loop may present a hydrogen-bond acceptor, which could shield the peptide-bound MHC I F-pocket, consistent with a peptide trapping function[29]. We propose that the editing loop fulfills two putatively contrasting tasks. On the one hand, the editing loop contributes to the widening of the F-pocket in the peptide-receptive state of MHC I, whereas, on the other hand, it shields the peptide C terminus during the epitope-loading process. Such a scenario may also be applicable to TAPBPR as NMR studies showed that the longer editing loop of TAPBPR (residues Gly[24]-Arg[36]) can be observed as a lid on top of the F-pocket[29], whereas in one of the MHC I/TAPBPR structures, the loop dives into the F-pocket and stabilizes peptide-receptive MHC I by acting as a peptide surrogate (scoop loop)[7,36,37] (Supplementary Fig. 4b).
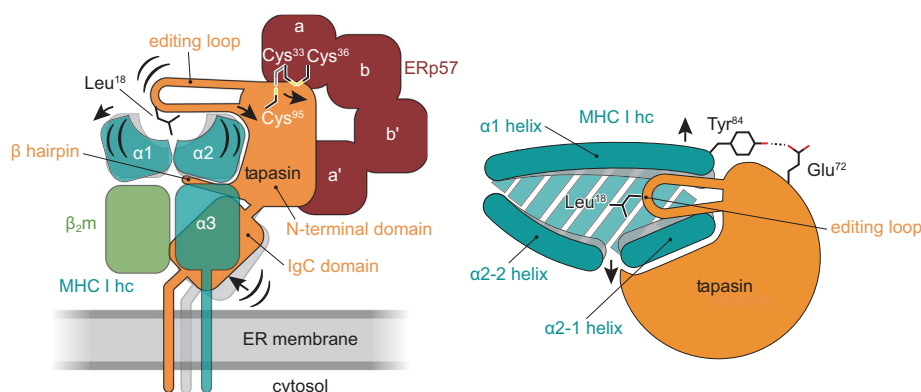
Fig. 5 | Model for chaperone function of tapasin towards peptide-receptive MHC I. Schematic of molecular motions leading to stabilized peptide-receptive clients (MHC I hc, teal; β_2m, green) upon association with tapasin (orange; left, side view; right, top view). Conformational changes are indicated by black arrows. The dotted line indicates a polar interaction between tapasin (Glu^{72}) and chaperoned MHC I (Tyr^{84} of the heavy chain, hc). For reasons of clarity, ERp57 (red) is not shown in the right panel.

Tapasin promiscuity for MHC I clients

In the tapasin-chaperoned state, the position of β_2m in relation to the MHC I hc is similar (RMSD 1.3 Å) to the peptide-bound state[28], whereas binding of TAPBPR leads to a repositioning of β_2m[6,7]. Strikingly, sequence alignments of classical MHC I allomorphs show that the interaction sites for tapasin are conserved among tapasin-dependent and -independent clients and are shared between tapasin and TAPBPR, except for residue Trp^{73} of H2-D^b (Supplementary Fig. 6). Analyzing the primary structure of non-classical MHC I molecules HLA-E, -F, and -G demonstrated that not all potential interacting residues are conserved, but most of the alternative residues are still capable to form hydrogen bonds (Supplementary Fig. 7). In essence, the crystal structure of the MHC I–tapasin–ERp57 editing complex reveals that both tapasin-dependent and -independent allomorphs have the structural prerequisites to interact with the chaperone, suggesting that their degree of plasticity and hence their ability to accommodate high-affinity peptide cargo determines the association, which is stabilized by key structural elements of tapasin (Fig. 5). It needs further structural and functional characterization to understand how *e.g.*, C-terminally extended peptides that protrude from the MHC I F-pocket are edited by tapasin or TAPBPR. As tapasin acts primarily as part of the PLC multichaperone network, we speculate that association of ERp57 with tapasin is not only necessary for coordinating the glycan-chaperone calreticulin, but that the molecular dynamics of this interaction may help unlock tapasin for client engagement and peptide proofreading catalysis.

Methods

Plasmid constructs

The constructs of human β_2-microglobulin (β_2m) (UniProtKB P61769; amino acids 21-119) and of the ectodomain of H2-D^b (UniProtKB P01899; amino acids 21-300) were previously described[7]. DNA fragments encoding the ER-lumenal domain of human tapasin (UniProtKB O15533; amino acids 21-402; tapasin$^{\Delta TMD}$) and human ERp57 (UniProtKB P30101; amino acids 25-505) were amplified by polymerase chain reaction (PCR) and cloned into the modified MultiBac vectors pAMI and pMIDK, respectively[38]. The C36A escape pathway mutation (ERp57^{C36A}) was introduced in the ERp57-encoding construct by QuikChange mutagenesis[11]. A C-terminal His$_6$-tag and a Tobacco Etch Virus (TEV) protease cleavage site were introduced into the pAMI vector encoding tapasin$^{\Delta TMD}$ by sequence- and ligation-independent cloning (SLIC)[39]. The pMIDK_ERp57^{C36A} donor vector was fused by Cre-mediated recombination to the pAMI_tapasin$^{\Delta TMD}$ acceptor vector, resulting in one transfer plasmid for integration into the EMBacY

baculoviral genome via Tn7 transposition[40]. Tapasin-encoding plasmids for flow cytometry analyses were generated by assembling the coding sequence of wildtype tapasin (UniProtKB O15533; amino acids 21-448), an internal ribosomal entry site 2 (IRES2) followed by the enhanced green fluorescent protein (eGFP)-encoding sequence into the pAMI vector. Interaction mutations of tapasin were introduced by SLIC[39]. All vectors were verified by Sanger sequencing.

Protein production and purification

H2-D^b and β_2m were expressed as inclusion bodies in *Escherichia coli* Rosetta(DE3)pLysS as described[41]. In brief, Rosetta(DE3)pLysS cells were transformed with plasmids encoding β_2m or H2-D^b. Overnight cultures of 20-30 ml in TYP media (16 g tryptone, 16 g yeast extract, 5 g NaCl, 1.25 g K$_2$HPO$_4$ per 1 l) supplemented with ampicillin and chloramphenicol were used to inoculate 2 l of TYP main culture. The main culture was incubated at 37 °C and 160 rpm until an optical density at 600 nm of 0.8 was reached. Temperature was reduced to 20 °C and protein production was induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 20 h, cells were harvested by centrifugation at 4 °C and 4500 × g for 15 min. Cells were flash-frozen in liquid nitrogen and stored at −80 °C.

The tapasin–ERp57 heterodimer was overexpressed in *Spodoptera frugiperda* (*Sf*) 21 cells essentially as previously described[40,42]. In brief, *Sf*21 cells were grown in Sf900 II SFM medium (Thermo Fisher Scientific) at 28 °C and transfected with modified EMBacY bacterial artificial chromosomes (BACs) using X-tremeGENE DNA transfection reagent (Roche). After incubation for 72 h at 28 °C, recombinant baculovirus V$_0$ was harvested and utilized for production of amplified baculovirus V$_1$. The tapasin–ERp57 heterodimer was expressed in 800 ml suspension culture at a cell density of 10^6 cells ml^{-1} by infection with 0.5–1.0% (v/v) baculovirus V$_1$. Cells were harvested 72 h post cell proliferation arrest by centrifugation at 4 °C and 1300 × g for 10 min. Cell pellets were flash-frozen in liquid nitrogen and stored at −80 °C.

Peptide synthesis

The photocleavable peptide photo-P18-I10 (RGPGRAF(J*)TI; J*, (S)−3 amino-3-(2-nitrophenyl)-propanoic acid) was synthesized using Fmoc solid-phase chemistry on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation). A Wang resin preloaded with iso-leucine (Iris Biotech GmbH) was utilized as support. After synthesis, peptides were cleaved from the Wang resin using 95% trifluoroacetic acid (TFA), 2.5% H$_2$O and 2.5% triisopropylsilane (TIPS). Peptides were washed four times with diethyl ether, resuspended in a mixture of tert-butanol and water (4:1), and lyophilized. Peptide identity was verified

on a liquid chromatography-mass spectrometer (BioAccord LC-MS, Waters). M_{calc}: 1166.599 Da; M_{obs}: 1166.601 Da.

Refolding and assembly of H2-D^b/β$_2$m complexes

Inclusion bodies of β$_2$m and H2-D^b were isolated and refolded by dialysis according to established protocols[41]. First, cell pellets were resuspended in TE buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM DTT) containing 1% (v/v) Triton X-100, followed by centrifugation at 10,000 × g for 20 min at 4 °C. Inclusion bodies were washed five times followed by two washing steps with TE buffer without Triton X-100. Pure inclusion bodies were resuspended in TE buffer, flash-frozen in liquid nitrogen, and stored at −80 °C in aliquots.

Purified inclusion bodies from a 2 l expression culture of β$_2$m were dissolved in 50 ml denaturation buffer (8 M urea, 100 mM Tris-HCl, pH 8.0) and incubated at RT for 1 h. After centrifugation at 16,000 × g for 5 min at 4 °C, supernatant was dialyzed twice against 2 l of 10 mM Tris-HCl, pH 8.0, for at least 9 h. Refolded β$_2$m was polished by size exclusion chromatography (SEC) using a Superdex75 16/60 column (GE Healthcare) equilibrated with 50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl. Elution fractions corresponding to β$_2$m were concentrated by ultrafiltration (Amicon Ultra 3 kDa, MWCO, Merck), flash-frozen in liquid nitrogen, and stored at −80 °C.

Inclusion bodies of H2-D^b were isolated and refolded by rapid dialysis according to established protocols[7,41]. H2-D^b inclusion bodies were dissolved in denaturation buffer (8 M urea, 100 mM Tris-HCl, pH 8.0) and added dropwise under constant stirring at 4 °C (final concentration of 1 mM) to refolding buffer (100 mM Tris-HCl, pH 8.0, 400 mM L-arginine, 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 2 mM ethylenediaminetetraacetate (EDTA), 1x Protease-Inhibitor Mix HP (Serva, Electrophoresis GmbH)), in the presence of 40 mM photocleavable peptide photo-P18-I10 and 2 mM of purified β$_2$m. The refolding reaction was incubated under stirring for 3 days at 4 °C. Refolded peptide/H2-D^b/β$_2$m complexes were concentrated by ultrafiltration (Amicon Ultra 10 kDa, MWCO, Merck) and polished by SEC using a Superdex200 10/300 Increase column (GE Healthcare) equilibrated in SEC buffer (50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl). Peak fractions of refolded photo-P18-I10/H2-D^b/β$_2$m complexes were pooled, concentrated by ultrafiltration, flash-frozen in liquid nitrogen, and stored at −80 °C.

Purification of tapasin–ERp57 heterodimers

Infected *Sf*21 cell pellets were resuspended in 100 ml lysis buffer (50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl, 25 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM benzamidine) per 1 l expression culture and lysed by sonication. After centrifugation at 20,000 × g, 4 °C for 30 min, the supernatant was incubated with pre-equilibrated Ni^{2+}-nitrilotriacetic acid (NTA) agarose resin (Thermo Fisher Scientific) for 1 h at 4 °C. NTA agarose resin was washed with lysis buffer without protease inhibitors and bound proteins were eluted with elution buffer (50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl, 300 mM imidazole). Protein-containing elution fractions were pooled, and buffer exchanged to 50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl on a PD-10 desalting column (GE Healthcare). The His$_6$-tag was cleaved by addition of TEV protease and incubated overnight at 4 °C. Cleaved tapasin–ERp57 heterodimers were concentrated by ultrafiltration (Amicon Ultra 50 kDa, MWCO, Merck) and polished by SEC using a Superdex200 10/300 Increase column (GE Healthcare) equilibrated with 50 mM HEPES-NaOH, pH 7.4, 150 mM NaCl. Peak fractions were pooled, flash-frozen in liquid nitrogen, and stored at −80 °C.

Crystallization

Purified tapasin–ERp57 and photo-P18-I10/H2-Db/β$_2$m complexes were mixed at a molar ratio of 1.0 to 1.3 at a total concentration of 10 mg ml^{-1}. Aliquots of 20 μl were exposed to UV light (365 nm, 185 mW/cm^2, 120 s) and applied to high-throughput crystallization

trials. Initial crystals belonging to the orthorhombic space group (SG) P2$_1$2$_1$2$_1$ with two tapasin–ERp57/H2-D^b/β$_2$m complexes per asymmetric unit were obtained at 18 °C by sitting drop vapor diffusion using a reservoir solution comprising 900 mM sodium cacodylate, pH 6.2, 270 mM sodium acetate, 6% polyethylene glycol (PEG) 2000, and 6% PEG500 monomethyl ether (MME). Despite extensive manual refinements and additive screening, these crystals diffracted X-rays only to a resolution of 4.0 Å. To obtain higher resolution, crystals were utilized for microseed matrix screening (MMS)[26]. To this end, crystals were crushed, transferred to a tube containing seed beads, and vortexed for 5 min at room temperature (RT). Seeds were used for another round of high-throughput crystallization trials. Crystals belonging to SG P22$_1$2$_1$ with one tapasin–ERp57/H2-D^b/β$_2$m complex per asymmetric unit were obtained by sitting drop vapor diffusion with a reservoir solution of 100 mM Gly-Gly, 100 mM 2-amino-2-methyl-1,3-propanediol (AMPD), pH 8.5, 300 mM lithium sulfate, 300 mM sodium sulfate, 300 mM potassium sulfate, 20% (v/v) PEG8000, 40% (v/v) 1,5-pentanediol, and a seed/protein/reservoir ratio of 0.3/1.0/1.0. Crystals were directly flash-frozen in liquid nitrogen without addition of further cryoprotectants.

Data collection, structure determination, and refinement

The synchrotron data set was collected at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany)[43]. Data were indexed, integrated, and scaled using the XDS package[44]. The structure of the editor complex was solved by molecular replacement using the coordinates of the tapasin–ERp57/MHC I editing module of the peptide-loading complex (PDB ID 6ENY)[8] as search model with the program Phaser within the Phenix software package[45]. The atomic model of the tapasin–ERp57/H2-D^b/β$_2$m complex was manually built in COOT[46] and refined in Phenix applying translation-liberation-screw (TLS) parameters. The atomic model was further improved using the PDB-REDO server[47] followed by manual adjustments in COOT[46] and refinement in Phenix[45]. The residual electron density in the MHC I binding pocket did not allow modeling of the photo-P18-I10 peptide used for refolding of MHC I or photo-induced fragments thereof with high occupancy. This result is consistent with SEC-MS analyses of H2-D^b/β$_2$m/photo-P18-I10 complexes after photo-triggered peptide cleavage that showed residual amounts of uncleaved peptide (4%) or of the 6-mer peptide fragment (2%) still bound to MHC I (Supplementary Fig. 2). Numberings of modeled amino acids correspond to the mature polypeptides lacking the respective signal sequences. Molecular graphics images were prepared using PyMOL (Schrödinger).

Alignments

Multiple sequences were aligned using Clustal Omega[48] and plotted with ESPript 3.0 (ref. [49]).

Cell culture

Tapasin-deficient HAP1 cell line[50] was cultured at 37 °C and 5% CO$_2$ in Iscove's modified Dulbecco's medium (IMDM, Gibco) supplemented with 10% fetal calf serum (FCS, Gibco). Cells were split at about 80% confluency with 0.05% Trypsin-EDTA (Gibco).

Transfection of HAP1 cells

One day prior to transfection, 40 × 10^3 cells per well of a 6-well plate were seeded. On the day of transfection, cells were washed using 1× Dulbecco's phosphate buffered saline, pH 7.4 (DPBS, Gibco), and 1 ml of fresh IMDM medium per well was added. For the transfection mixture, 2 μg DNA and 6 μl X-tremeGENE HP (1:3 ratio) (Roche) were separately dissolved in 50 μl Opti-MEM (Gibco) and incubated for 5 min at RT. The solutions were mixed, incubated for 15 min at RT and added dropwise to the cells. 1 ml of medium was added after 4–6 h, and cells were harvested 48 h after transfection.

MHC I surface expression

MHC I surface expression was analyzed by flow cytometry. All steps were carried out on ice. The cells were washed in 0.2 ml FACS buffer (1x DPBS, 2% BSA, 2 mM EDTA, 0.02% (w/v) sodium azide) and centrifuged for 5 min at $300 \times g$, 4 °C. The supernatant was discarded, and cells were blocked by 10% FcR Blocking Reagent, human (Miltenyi Biotec) in 50 µl FACS buffer for 15 min. Afterwards, cells were washed with 0.2 ml FACS buffer, centrifuged for 5 min at $300 \times g$ and 4 °C, and stained with 2 µl of APC anti-human HLA-A/B/C antibody (64 µg/ml, clone W6/32, BioLegend) in 0.1 ml FACS buffer for 30 min. Subsequently, cells were washed twice with 0.2 ml FACS buffer and resuspended in 0.2 ml FACS buffer for analysis by flow cytometry. Data was recorded on a FACSMelody Cell Sorter (BD Bioscience), processed using FlowJo V10 software and analyzed using Excel (Version 16.7) and GraphPad Prism 8.21 MacOS. The gating strategy is displayed in Supplementary Fig. 8.

Immunoblotting

0.2×10^6 eGFP-positive cells were sorted on FACSMelody Cell Sorter (BD Bioscience), centrifuged at $300 \times g$, 4 °C for 5 min, lysed in 25 µl Pierce RIPA buffer (Thermo Fisher Scientific) supplemented with 1% Benzonase (Novagen, EMD Chemicals) and 1× Protease-Inhibitor Mix HP (Serva, Electrophoresis GmbH), and incubated for 15 min at 500 rpm, 25 °C. 5 µl of 4× sodium dodecyl sulfate (SDS)-loading buffer (250 mM Tris-HCl, pH 7.5, 30% glycerol, 10% SDS, 10% β-mercaptoethanol, 0.02% Bromophenol Blue) were added, and samples were incubated at 95 °C for 10 min. 10 µl of samples were loaded on a Mini-PROTEAN TGX Gel (4-20%, BioRad), and gel electrophoresis was performed according to manufacturer's protocol (BioRad). The samples were blotted onto a PVDF membrane by semi-dry blotting for 30 min at 25 V. The blotted membranes were blocked in 5% milk in TBS-T for 1 h and incubated with Direct-Blot HRP anti-GAPDH (clone FF26A/F9, BioLegend, 1:2000) at 4 °C overnight. Blots were washed in TBS-T at RT for 20 min and then incubated with anti-tapasin (hybridoma, clone 7F6, 1:3000) for 1 h[51]. Anti-rat IgG Peroxidase conjugate (Sigma-Aldrich, 1:20,000) was incubated for 45 min at RT as secondary antibody. Membranes were incubated with Clarity Western ECL reagent (BioRad) or LumiGLO Peroxidase Chemiluminescent Substrate Kit (Seracare) and chemiluminescence was measured with a Fusion FX (Vilber). Due to weak intensities corresponding to GAPDH, membranes were incubated with unconjugated anti-GAPDH (PA1-987, Thermo Fisher Scientific, 1:2,000) at 4 °C overnight before incubation for 45 min with anti-rabbit IgG Peroxidase conjugate (Merck Millipore, 1:20,000).

EndoH digest

For LC-MS analysis, tapasin–ERp57 was deglycosylated using EndoH. Glycoprotein was mixed with Glycobuffer 3 (New England Biolabs), and 625 U of EndoH (New England Biolabs) per 1 µg of protein were added. The mixture was incubated overnight at RT. Deglycosylated tapasin–ERp57 was purified by SEC.

LC-MS analysis

All LC-MS measurements were performed using a BioAccord System (Waters). Peptides were analyzed on an ACQUITY UPLC Peptide BEH C_{18} Column, 130 Å, 1.7 µm, 2.1 mm × 100 mm (Waters), with a linear water/acetonitrile gradient complemented with 0.1% (v/v) formic acid at 60 °C, 30 V cone voltage, 0.8 kV capillary voltage, and a desolvation temperature of 550 °C in positive polarity at 50–2000 m/z. Intact protein LC-MS data was acquired using a cone voltage of 60 V, 1.5 kV capillary voltage, and a desolvation temperature of 500 °C on an ACQUITY UPLC Protein BEH C_4 Column, 300 Å, 1.7 µm, 2.1 mm × 50 mm (Waters) at 80 °C running a linear water/acetonitrile gradient, supplemented with 0.1% (v/v) formic acid. Mass spectra were recorded in positive polarity at 2 Hz in full scan mode at 400–7000 m/z. SEC-MS measurements were performed using an ACQUITY UPLC Protein BEH SEC Column, 200 Å, 1.7 µm, 2.1 mm × 150 mm (Waters) in 20 mM ammonium acetate. Mass spectra were recorded in positive mode at 1 Hz in full scan mode at 400–7000 m/z with a cone voltage of 42 V, capillary voltage of 1.5 kV, and a desolvation temperature of 450 °C. Masses of peptides and proteins were calculated and confirmed in Unify. Intact mass spectra were deconvoluted in Unify using MaxEnt1 algorithm iterating to convergence. Spectra with high background noise were subjected to automatic baseline correction before deconvolution. Deconvoluted spectra were centroidized based on peak height and used for mass calculations. UV spectra were recorded at 280 nm with a sampling rate of 10 Hz.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The structure of the MHC I–tapasin–ERp57 editing complex was deposited to the Protein Data Bank (http://www.rcsb.org) under accession number PDB ID 7QNG. Previously published structural data used in this study are accessible at the PDB (human PLC editing module tapasin-ERp57/calreticulin/MHC I, PDB ID 6ENY; peptide-receptive MHC I, PDB ID 2F74; client-free tapasin–ERp57 heterodimer, PDB ID 3F8U; MHC I–TAPBPR complex, PDB ID 5OPI). Mass spectrometry data are available via Zenodo [https://doi.org/10.5281/zenodo.5939241]. Source data are provided with this paper.

References

1. Neefjes, J., Jongsma, M. L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
2. Blum, J. S., Wearsch, P. A. & Cresswell, P. Pathways of antigen processing. *Annu Rev. Immunol.* **31**, 443–473 (2013).
3. Trowitzsch, S. & Tampé, R. Multifunctional chaperone and quality control complexes in adaptive immunity. *Annu Rev. Biophys.* **49**, 135–161 (2020).
4. Dong, G., Wearsch, P. A., Peaper, D. R., Cresswell, P. & Reinisch, K. M. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* **30**, 21–32 (2009).
5. Hermann, C., Strittmatter, L. M., Deane, J. E. & Boyle, L. H. The binding of TAPBPR and Tapasin to MHC class I is mutually exclusive. *J. Immunol.* **191**, 5743–5750 (2013).
6. Jiang, J. et al. Crystal structure of a TAPBPR-MHC I complex reveals the mechanism of peptide editing in antigen presentation. *Science* **358**, 1064–1068 (2017).
7. Thomas, C. & Tampé, R. Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing. *Science* **358**, 1060–1064 (2017).
8. Blees, A. et al. Structure of the human MHC-I peptide-loading complex. *Nature* **551**, 525–528 (2017).
9. Boyle, L. H. et al. Tapasin-related protein TAPBPR is an additional component of the MHC class I presentation pathway. *Proc. Natl Acad. Sci. USA* **110**, 3465–3470 (2013).
10. Dick, T. P., Bangia, N., Peaper, D. R. & Cresswell, P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. *Immunity* **16**, 87–98 (2002).
11. Walker, K. W. & Gilbert, H. F. Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* **272**, 8845–8848 (1997).
12. Ortmann, B., Androlewicz, M. J. & Cresswell, P. MHC class I/beta 2-microglobulin complexes associate with TAP transporters before peptide binding. *Nature* **368**, 864–867 (1994).

13. Sadasivan, B., Lehner, P. J., Ortmann, B., Spies, T. & Cresswell, P. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* **5**, 103–114 (1996).

14. Wearsch, P. A. & Cresswell, P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. *Nat. Immunol.* **8**, 873–881 (2007).

15. Chen, M. & Bouvier, M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. *EMBO J.* **26**, 1681–1690 (2007).

16. Ortmann, B. et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. *Science* **277**, 1306–1309 (1997).

17. Fleischmann, G. et al. Mechanistic basis for epitope proofreading in the peptide-loading complex. *J. Immunol.* **195**, 4503–4513 (2015).

18. Garbi, N., Tanaka, S., Momburg, F. & Hammerling, G. J. Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57. *Nat. Immunol.* **7**, 93–102 (2006).

19. Grandea, A. G. 3rd et al. Impaired assembly yet normal trafficking of MHC class I molecules in Tapasin mutant mice. *Immunity* **13**, 213–222 (2000).

20. Rizvi, S. M. et al. Distinct assembly profiles of HLA-B molecules. *J. Immunol.* **192**, 4967–4976 (2014).

21. Bashirova, A. A. et al. HLA tapasin independence: broader peptide repertoire and HIV control. *Proc. Natl Acad. Sci. U.S.A.* **117**, 28232–28238 (2020).

22. Sieker, F., Straatsma, T. P., Springer, S. & Zacharias, M. Differential tapasin dependence of MHC class I molecules correlates with conformational changes upon peptide dissociation: a molecular dynamics simulation study. *Mol. Immunol.* **45**, 3714–3722 (2008).

23. van Hateren, A., Bailey, A., Werner, J. M. & Elliott, T. Plasticity of empty major histocompatibility complex class I molecules determines peptide-selector function. *Mol. Immunol.* **68**, 98–101 (2015).

24. Bailey, A. et al. Selector function of MHC I molecules is determined by protein plasticity. *Sci. Rep.* **5**, 14928 (2015).

25. Abualrous, E. T. et al. The carboxy terminus of the ligand peptide determines the stability of the MHC class I molecule H-2Kb: A combined molecular dynamics and experimental study. *PLoS ONE* **10**, e0135421 (2015).

26. D'Arcy, A., Villard, F. & Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 550–554 (2007).

27. Bhattacharyya, R., Dhar, J., Ghosh Dastidar, S., Chakrabarti, P. & Weiss, M. S. The susceptibility of disulfide bonds towards radiation damage may be explained by S⋯O interactions interactions. *IUCrJ* **7**, 825–834 (2020).

28. Achour, A. et al. Structural basis of the differential stability and receptor specificity of H-2Db in complex with murine versus human beta2-microglobulin. *J. Mol. Biol.* **356**, 382–396 (2006).

29. McShan, A. C. et al. TAPBPR promotes antigen loading on MHC-I molecules using a peptide trap. *Nat. Commun.* **12**, 3174 (2021).

30. Hafstrand, I. et al. Successive crystal structure snapshots suggest the basis for MHC class I peptide loading and editing by tapasin. *Proc. Natl Acad. Sci. USA* **116**, 5055–5060 (2019).

31. Lan, H. et al. Exchange catalysis by tapasin exploits conserved and allele-specific features of MHC-I molecules. *Nat. Commun.* **12**, 4236 (2021).

32. Wang, R., Natarajan, K. & Margulies, D. H. Structural basis of the CD8 alpha beta/MHC class I interaction: focused recognition orients CD8 beta to a T cell proximal position. *J. Immunol.* **183**, 2554–2564 (2009).

33. Hermann, C. et al. TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst. *eLife* **4**, e09617 (2015).

34. Morozov, G. I. et al. Interaction of TAPBPR, a tapasin homolog, with MHC-I molecules promotes peptide editing. *Proc. Natl Acad. Sci. USA* **113**, E1006–E1015 (2016).

35. Ilca, F. T., Drexhage, L. Z., Brewin, G., Peacock, S. & Boyle, L. H. Distinct polymorphisms in HLA class I molecules govern their susceptibility to peptide editing by TAPBPR. *Cell Rep.* **29**, 1621–1632 e1623 (2019).

36. Sagert L, Hennig F, Thomas C, Tampé R. A loop structure allows TAPBPR to exert its dual function as MHC I chaperone and peptide editor. *eLife* **9**, e55326 (2020).

37. Thomas, C. & Tampé, R. MHC I assembly and peptide editing—chaperones, clients, and molecular plasticity in immunity. *Curr. Opin. Immunol.* **70**, 48–56 (2021).

38. Sekulovski, S. et al. Assembly defects of human tRNA splicing endonuclease contribute to impaired pre-tRNA processing in pontocerebellar hypoplasia. *Nat. Commun.* **12**, 5610 (2021).

39. Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).

40. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* **172**, 45–54 (2010).

41. Rodenko, B. et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).

42. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004).

43. Cianci, M. et al. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **24**, 323–332 (2017).

44. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).

45. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).

46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).

47. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).

48. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

49. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* **42**, W320–W324 (2014).

50. Jongsma, M. L. M. et al. The SPPL3-defined glycosphingolipid repertoire orchestrates HLA class I-mediated immune pesponses. *Immunity* **54**, 132–150 e139 (2021).

51. Hulpke, S. et al. Direct evidence that the N-terminal extensions of the TAP complex act as autonomous interaction scaffolds for the assembly of the MHC I peptide-loading complex. *Cell Mol. Life Sci.* **69**, 3317–3327 (2012).

Acknowledgements

Author contributions

Conceptualization: S.T., R.T.; Methodology: I.K.M., C.W., C.T., S.T., R.T.; Investigation: I.K.M., C.W., S.T.; Knockout cells: R.M.S.; Visualization: I.K.M., S.T., R.T.; Funding acquisition: R.T.; Supervision: S.T., R.T.; Writing —original draft: I.K.M., S.T., R.T.; Writing—review and editing: I.K.M., C.W., C.T., R.MS., S.T., R.T.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32841-9.

Correspondence and requests for materials should be addressed to Simon Trowitzsch or Robert Tampé.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity

Ines Katharina Müller[1], Christian Winter[1], Christoph Thomas[1], Robbert M. Spaapen[2,3], Simon Trowitzsch[1]*, Robert Tampé[1]*

[1]Institute of Biochemistry, Biocenter, Goethe University Frankfurt; Max-von-Laue Strasse 9, 60438 Frankfurt/Main, Germany.

[2]Department of Immunopathology, Sanquin Research, Amsterdam, The Netherlands.

[3]Landsteiner Laboratory, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands.

*Corresponding authors. Email: trowitzsch@biochem.uni-frankfurt.de (ST); Email: tampe@em.uni-frankfurt.de (RT)

Supplementary Figure 1-8
Supplementary Table 1

SUPPLEMENTARY FIGURES

a

b

96369.1

ERp57–tapasin^{GlcNAc}
M_{calc}= 96368.7 Da

c

42139.8

tapasin^{GlcNAc}
M_{calc} = 42139.4 Da

d

54230.2

ERp57
M_{calc} = 54230.4 Da

e

f

g
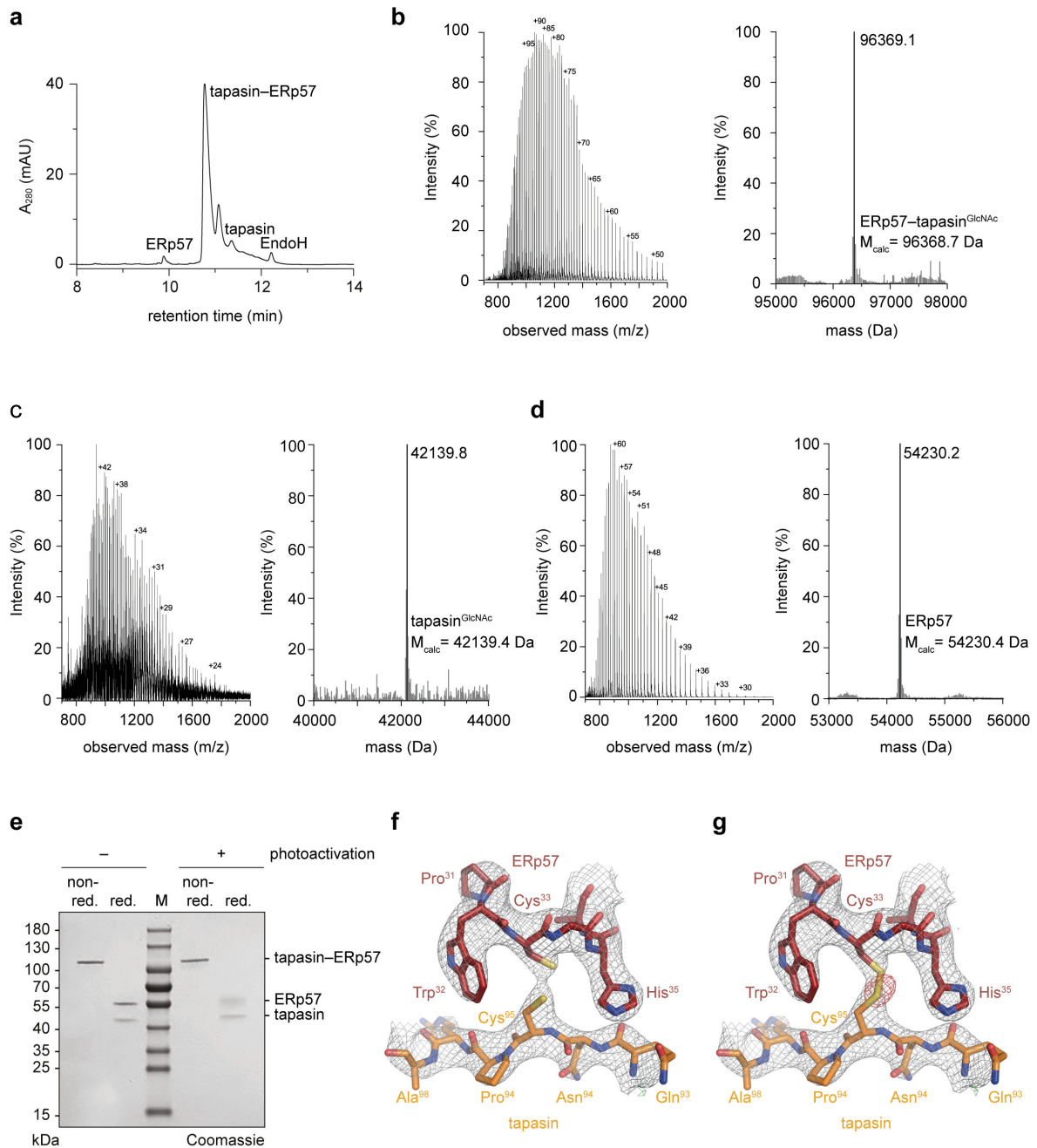
Supplementary Fig. 1 An intact intermolecular disulfide bond is formed in tapasin–ERp57 but is absent in the crystal structure. a The endoH-deglycosylated tapasin–ERp57 complex was analyzed by reversed-phase C$_4$ chromatography combined with mass spectrometry. A$_{280}$, absorption at 280 nm. **b** ESI-MS (left) and MaxEnt1-deconvoluted (right) spectrum of corresponding tapasin–ERp57 complex. M$_{calc}$, calculated mass. **c** ESI-MS (left) and MaxEnt1-

deconvoluted (right) spectrum of tapasin. **d** ESI-MS (left) and MaxEnt1 deconvoluted (right) spectrum of ERp57. **e** SDS-PAGE analysis of the purified tapasin–ERp57 complex under non-reducing (non red.) and reducing (red.) conditions. -/+, samples before and after exposure to UV light; kDa, kilodalton; M, marker; representative SDS-PAGE is shown. **f** Stick representation of residues around Cys^{95} of tapasin (orange) and Cys^{33} of ERp57 (red) with the 2F$_o$-F$_c$ map (grey mesh) displayed at the 2 σ level and the F$_o$-F$_c$ difference map (red mesh) displayed at the -4 σ level. **g** Same representation as in (**f**) but calculated with a modeled disulfide bond between Cys^{95} of tapasin and Cys^{33} of ERp57. Source data for (**a**) and (**e**) are provided as a Source Data file.
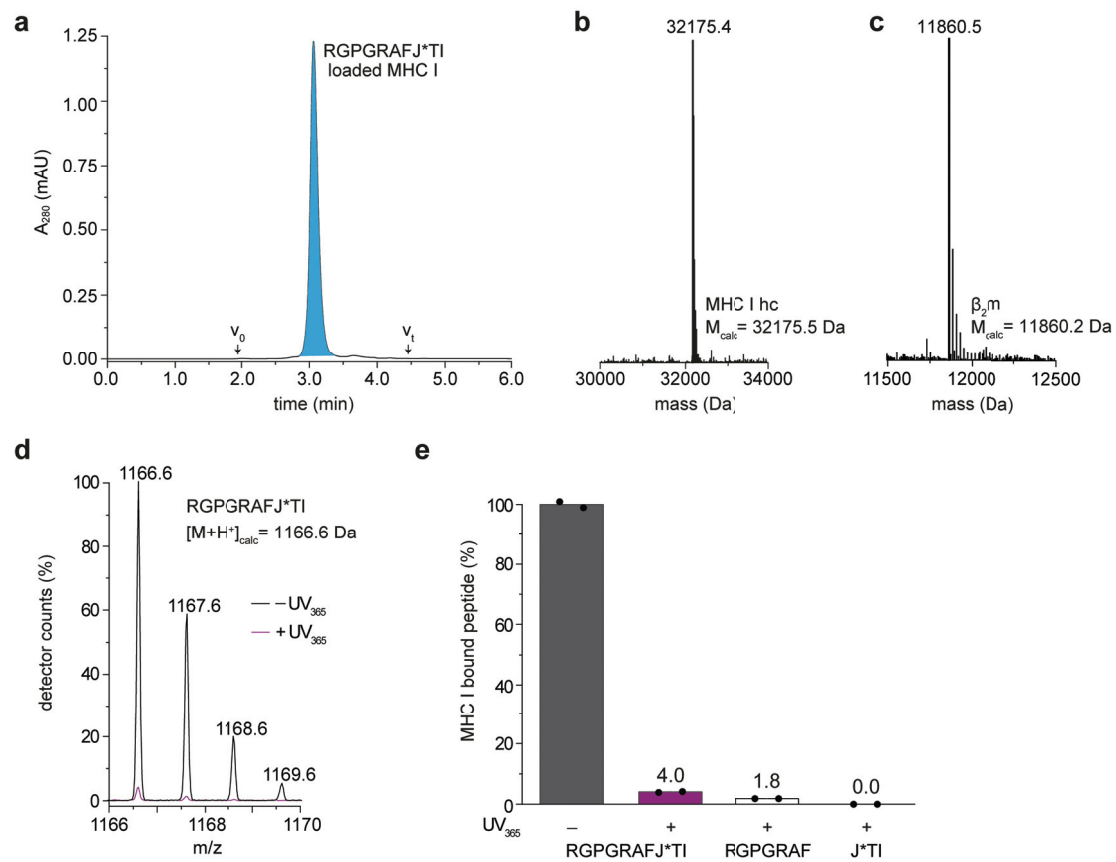
Supplementary Fig. 2 Photo-triggered cleavage of MHC I-associated peptide analyzed by SEC-MS. a Size exclusion chromatogram of RGPGRAFJ*TI loaded MHC I. Combined ESI-MS spectra of the main peak (light blue) were used for MS analysis shown in (**b-e**). The intact pMHC I complex is decomposed into individual components during ionization. A_{280}, absorption at 280 nm; V_0, void volume; V_t, total volume. **b-d** Intact protein masses for MHC I hc (**b**), β_2m (**c**) and MHC I-associated RGPGRAFJ*TI peptide (**d**, black line) were detected by ESI-MS. No covalently linked pMHC I complexes were detected after UV illumination at 365 nm. Amounts of MHC I-associated RGPGRAFJ*TI peptides after photo-triggered peptide removal are shown (**d**, purple line). M_{calc}, calculated mass. **e** Quantification of MHC I-associated photocleavable peptides and peptide fragments before (dark grey) and after illumination at 365 nm (intact peptide, purple; RGPGRAF fragment, white) followed by SEC-MS. Quantification is based on MS detector response and normalized to non-photocleaved pMHC I. The short J*TI fragment was not detected. n=2 biological replicates. Source data for (**a**) and (**e**) are provided as a Source Data file.
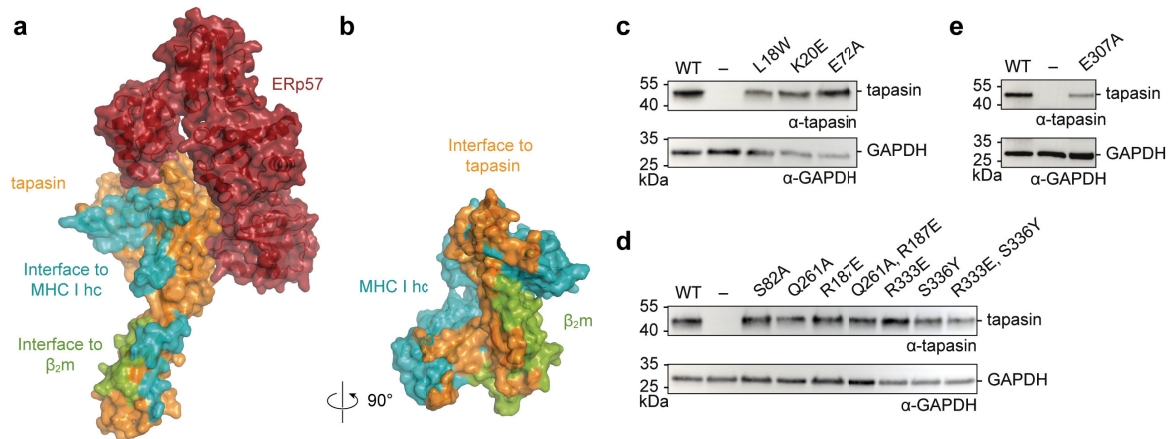
Supplementary Fig. 3 Interaction interfaces in the MHC I–tapasin–ERp57 editing complex. **a** Contact regions (surface representation) of MHC I hc and β_2m on the tapasin–ERp57 heterodimer, colored in orange and green, respectively. hc, heavy chain; β_2m, β_2-microglobulin. **b** Contact regions of tapasin on the MHC I hc and β_2m, colored in orange. **c** Whole cell extracts of eGFP-positive cells, expressing wildtype (WT) or F-pocket interface mutants of tapasin were analyzed by SDS-PAGE and immunoblotting (α-tapasin, α-GAPDH). –, mock transfection; kDa, kilodalton. **d** Immunoblots of whole cell extracts of wildtype and β hairpin, α3-domain interface mutants of tapasin (α-tapasin, α-GAPDH). –, mock transfection. **e** Whole cell extracts of wildtype and β_2m interface mutant of tapasin analyzed by SDS-PAGE and immunoblotting (α-tapasin, α-GAPDH). n=3 biologically independent samples for (**c**), (**d**) and (**e**) –, mock transfection. Source data for (**c**), (**d**), and (**e**) are provided as a Source Data file.
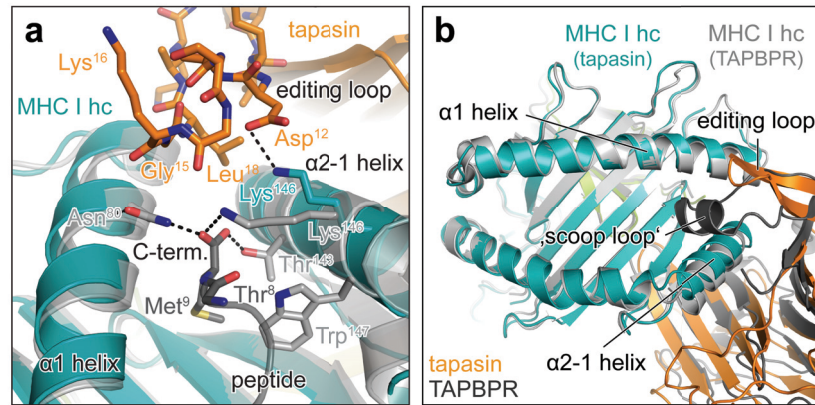
Supplementary Fig. 4 Comparison of peptide-bound with chaperoned MHC I. a View onto the F-pocket region of peptide-bound H2-D^b (grey, PDB ID 2F74) with superposition of tapasin (orange)-complexed H2-D^b (teal). Hydrogen bonds and the salt bridge of Lys^{146} are indicated by black dashed lines. Only backbone atoms are shown for Thr^8 of the MHC I-bound peptide. C-term., C terminus. **b** Comparison between the TAPBPR (dark grey)–MHC I (grey) complex (PDB ID 5OPI) and the tapasin (orange)–MHC I (teal) complex in cartoon representation.

editing loop

tapasin_human	GPAVIECWFVEDASGKGLA.....KRPGALLLRQGPGEPPP....RPDLDPELYLSVHDP
tapasin_mouse	GPEAIECWFVEDAGGGGLS.....KKPATLLLRHGPRGPPP....RPDLDPKLYFKVDDP
tapasin_rat	GPQTIECWFVEDAGGGGLS.....KKPATLLLRHGPRGPPP....RPDLDPKLYFKVDDP
tapasin_pig	GPAVIECWFVEDAGRGQLS.....KKPAALLLRQGPGSPPP....RPDLEPELYLKVHDP
tapasin_bovine	RPTVIECWFVEDAGGGKLA.....KKPAALLLRQGAESPPL....RPDLAPERYLKIHDP
TAPBPR_bovine	VDVVLDCFLMEEGGHHGGFASSENMVKAVLVLRQVPVPDDGSLEGLTDFQVDTLTK.DNS
TAPBPR_pig	VDVVLDCFLMEEGGHRGAFARSGNKMKALLVLKQVPVPDDGSLEGLTDFQVGTLAK.DDP
TAPBPR_rat	VDIVLDCFLVTEDKHRGAFGSSEDREKALLVLKQVPVLDDGSLEGFTDFQGSAEAR.QDS
TAPBPR_mouse	TDIILDCFLVTEDRHRGAFAESGDRERALLVLKQVPVLDDGSLECITDFQGSTETK.QDS
TAPBPR_human	VDVVLDCFLVKDGAHRGALASSEDRARASLVLKQVPVLDDGSLEDFTDFQGGTLAQ.DDP

interface loop

tapasin_human	AGAL..........QAAFRRYPRGAPAPHCEMSRFVPLPASAKWASGLTPAQNCPRALD
tapasin_mouse	AGML..........LAAFRRYPAGASAPHCEMSRFIPFPASAKWARSLSPEQNCPRALD
tapasin_rat	AGML..........LAAFRRYPAGAHAPHCEMSSYIPFPASAKWARGLSPQQNCSRALD
tapasin_pig	AGAL..........LAAFRQYPRDAPAPRCEMSHYIPLPASPDWVSGLTPEQRCPRALD
tapasin_bovine	AGAL..........QAAFRQYPRDAPAPRCEMSLYVPLPASAKWLSGLTPEQSCPRALD
TAPBPR_bovine	LITFEASVNLVQIPQAEALLHADCSGKEVTCEISRYFLQARPE............ATVEM
TAPBPR_pig	PITFEASVNLVQLPQAEALLHADCGGKEVTCEISRYFLQARQE............ATVET
TAPBPR_rat	PIIFEASVDLVQIPQAEALLHADCSGKAVTCEIAQYFLQARQE............ATFEK
TAPBPR_mouse	PVIFEASVDLVQIPQAEALLHADCSGKAVTCEISKYFLQARQE............ATFEK
TAPBPR_human	PIIFEASVDLVQIPQAEALLHADCSGKEVTCEISRYFLQMT.E............TTVKT

tapasin_human	GAWLMVSIS..SPVLSLSSLLRPQPEP............QQEPVLITMATVVLTVLTHTP
tapasin_mouse	GDWLLVSVS..STLFSLSSLLRPQPEP............LREPVVITMATVVLTVLTHNP
tapasin_rat	GDWLMVSVS..GTVFSLSSLLRPLPEP............QREPVFITMATVVLTVLTRHP
tapasin_pig	GKWLMVSMS..SPVLSLSSFLQPQSEP............QLEPALITMATAVLTVITHTP
tapasin_bovine	GTWLMVSMS..SPFLSLSSLLRPQSEP............QPEPALITMATVVLTVFTHTP
TAPBPR_bovine	AAWFITNVQVSGGGPGVSMVMKTLEDAENETVLHPTLKLPLSPQGTVRTAVEFQVTTQTP
TAPBPR_pig	AAWFITNVQVSGGGPSISMVMKTLGDAEKGAALHPTLNVPLSPQGTVRTAVEFQVTTQTP
TAPBPR_rat	ADWFISNMQVSRGGPSVSIVMKTLRDDEAGAVRHPSLNIPLSPQGTVKTKVEFQVTAETQ
TAPBPR_mouse	AHWFISNMQVSRGGPSVSMVMKTLRDAEVGAVRHPTLNLPLSAQGTVKTQVEFQVTSETQ
TAPBPR_human	AARFMANVQVSGGGPSISLVMKTPRVAKNEVLWHPTLNLPLSPQGTVRTAVEFQVMTQTQ

β hairpin

tapasin_human	APRVRLGQDALLDLSFATMPPTSEAASSLAPGPPPFGLEWRRQHLGKGHLLAATPGLNG
tapasin_mouse	APRVQLGKDAVLDLRFATAPSALEGSPSLDAGPPPFGLEWRRQHRGKGHLLAATPGLAG
tapasin_rat	APRIHLGKDAVLDLRFATAPPALEDAPSVASGPPPFGLEWRRQHRGKGHLLAAIPGLLE
tapasin_pig	APRIRLGQDALLDLSFATVPPTSKAAISLAPGPPPFGLEWRRQHLGKGHLLAATPGLSG
tapasin_bovine	TPRIPMGQDALLDLSFATVPLTPK.APSLAPGPPPFGLEWRRQHLCKGHLLAVTPGLNG
TAPBPR_bovine	SLNFLLGSTASLHCGY..........SVAPGLDVTSVEWRLQHKGSGRLVYHWT.M..G
TAPBPR_pig	SLNLLLGSTASLHCGF..........SMAPGLDLTSVEWRQQHGGSGHLVYSWTTM..G
TAPBPR_rat	TLNQLLGSSVSLPCRF..........SMAPGLDLTGVEWRLQHKGSGQLYYSWK.T..G
TAPBPR_mouse	TLNHLLGSSVSLHCSF..........SMAPGLDLTGVEWRLQHKGSGQLYYSWK.T..G
TAPBPR_human	SLSFLLGSSASLDCGF..........SMAPGLDLISVEWRLQHKGRGQLYYSWT.A..G

tapasin_human	QMPAAQEGAVAFAAWDDD.EPWGPWTGNGTFWLPRVQPFQEGTYLATIHLPYLQGQVTLE
tapasin_mouse	RMPPAQEKATAFAAWDDD.EPWGPWTGNGTFWLPAVKPSQEGVYLATVHLPYLQGQVSLE
tapasin_rat	KMPPAREKAVAFAAWDDDNDPWGPWTGNGTWLWLPAVSPSQEGVYLATVHLPYLQGQVSLE
tapasin_pig	OVPAAREGAVAFAAWDDD.EPWGPWTGNGSFWLPAVQPFQEGTYLATVHLPYLQGQVAME
tapasin_bovine	QVPAAREGAVAFAAWDDD.EPWGPWTGNGTFWLPAVRPSQEGTYLATVHLPYLQGQITLE
TAPBPR_bovine	QGQAKREGATLEP......EQLLTAGDASLTLPSLTLQDEGAYICQITTSLFRAQQVIQ
TAPBPR_pig	QGQAQREGATLDP......EQLLMAGDATLTLPSLTVKDEGAYICQITTSLYQTQQIVQ
TAPBPR_rat	QGQAKRKGATLKP......EKLLRAGDASLTLPNLTLKDEGNYICQISTSLYQAQQIVP
TAPBPR_mouse	QGQAKRKGATLEP......EELLRAGNASLTLPNLTLKDEGNYICQISTSLYQAQQIMP
TAPBPR_human	QGQAVRKGATLEP......AQLGMARDASLTLPGLTIQDEGTYICQITTSLYRAQQIIQ

tapasin_human	LAVYKPPKVSLMPATLALAAPGEAPPELLCLVSHFYPSGGLEVEWELRGGPGGRSQKAEG
tapasin_mouse	LTVHKAPRVSLTPAPVVIAAPGEAPPELLCLASHFPAEGLEVEWELRGGPEGRSRKVEG
tapasin_rat	LTVHKPPRVSLTPAPIVVAAPGEAPPELLCLVSHFYPEEGLEVEWELRGGPEGRSRKAEG
tapasin_pig	LAVQKSPKVSLTPAPLVVAAPGEAPPELLCLVSHFYPSEGLEVEWELRGGPGQGRFQKAEG
tapasin_bovine	LAVLKPPKVALTPAPLVVAAPGEAPPELLCLVSHFYPSHGLEVEWELRGGPEGRLQKAEG
TAPBPR_bovine	LYIQASPKVRLSLL.....NAAPPATLICNVAGYYPLD.VSVTWI.REEQGGSPAPVSG
TAPBPR_pig	LHVQAAPKVRLSLV.....NEAHSHTLICSVTGYYPLD.VTVTWI.REELGGALVPVSD
TAPBPR_rat	LNILASPKVQLILA.....NKDPLPSLICSIAGYYPLD.VVVTWI.REELGGIPAQVSG
TAPBPR_mouse	LNILAPPKIQLHLA.....NKDPLPSLVCSIAGYYPLD.VGVTWI.REELGGIPAQVSG
TAPBPR_human	LNIQASPKVRLSLA.....NEALLPTLICDIAGYYPLD.VVVTWT.REELGGSPAQVSG

tapasin_human	QRWLSALRHHSDGSVSLSGHLQPPVTTEQHGARYACRTHHPSLPASGRSAEVTLE
tapasin_mouse	KTWLSTIRHHSDGSVSQSGHLQLPPVTAKQHGVHVVCRVYHSLPASGRSADVTLE
tapasin_rat	KTWVSTVRHHSDGSVSQSGHLQLPPVTDKQHGVRMACRVHRHPSLPASGRSAEVTLE
tapasin_pig	QSWHSALSHHADGSVSLSAHLQLPPVTTEQHGARYACRVHHPSLPALGRRSEITLQ
tapasin_bovine	QRWLSALRHHSDGSLSLSAHLQLPPVTSEQHGARYACRVHHPSLSALGRSAEVTLQ
TAPBPR_bovine	A.SLSSLRQSPAGTYSISSSLTVEP...GSAGATYTCQVTHVSLEEPLGVIAWVAP
TAPBPR_pig	A.SFSSLRQSAAGTYSISSSLTADP...GSAGATYTCHITHVSLEEPLDVANTWVAP
TAPBPR_rat	A.SFSSLRQNMMGTYSISGTVIAEP...GPTGATYTCQVAHVSLEEPLTVSMRVLP
TAPBPR_mouse	A.SFSSLRQSTMGTYSISSTVMADP...GPTGATYTCQVAHVSLEEPLTTSMRVLP
TAPBPR_human	A.SFSSLRQSVAGTYSISSSLTAEP...GSAGATYTCQVTHISLEEPLGASTQVVP

Legend:

- tapasin/MHC I hc – H-bonds/salt bridges
- tapasin/MHC I hc – interface residues
- tapasin/β$_2$m – H-bonds/salt bridges
- tapasin/β$_2$m – interface residues
- TAPBPR/MHC I hc – H-bonds/salt bridges
- TAPBPR/MHC I hc – interface residues
- TAPBPR/β$_2$m – H-bonds/salt bridges
- TAPBPR/β$_2$m – interface residues

Supplementary Fig. 5 Structure-based sequence alignment of tapasin and TAPBPR. Conserved residues are highlighted in black, and physiochemically related residues are represented in bold letters. Residues of tapasin that are involved in interactions with MHC I and β_2m are marked with teal and green dots and bars, respectively. Residues of TAPBPR engaged in interaction with MHC I and β_2m are marked with dark and light symbols, respectively. The dots highlight specific interactions, such as hydrogen bonds and salt bridges. hc, heavy chain; β_2m, β_2-microglobulin.
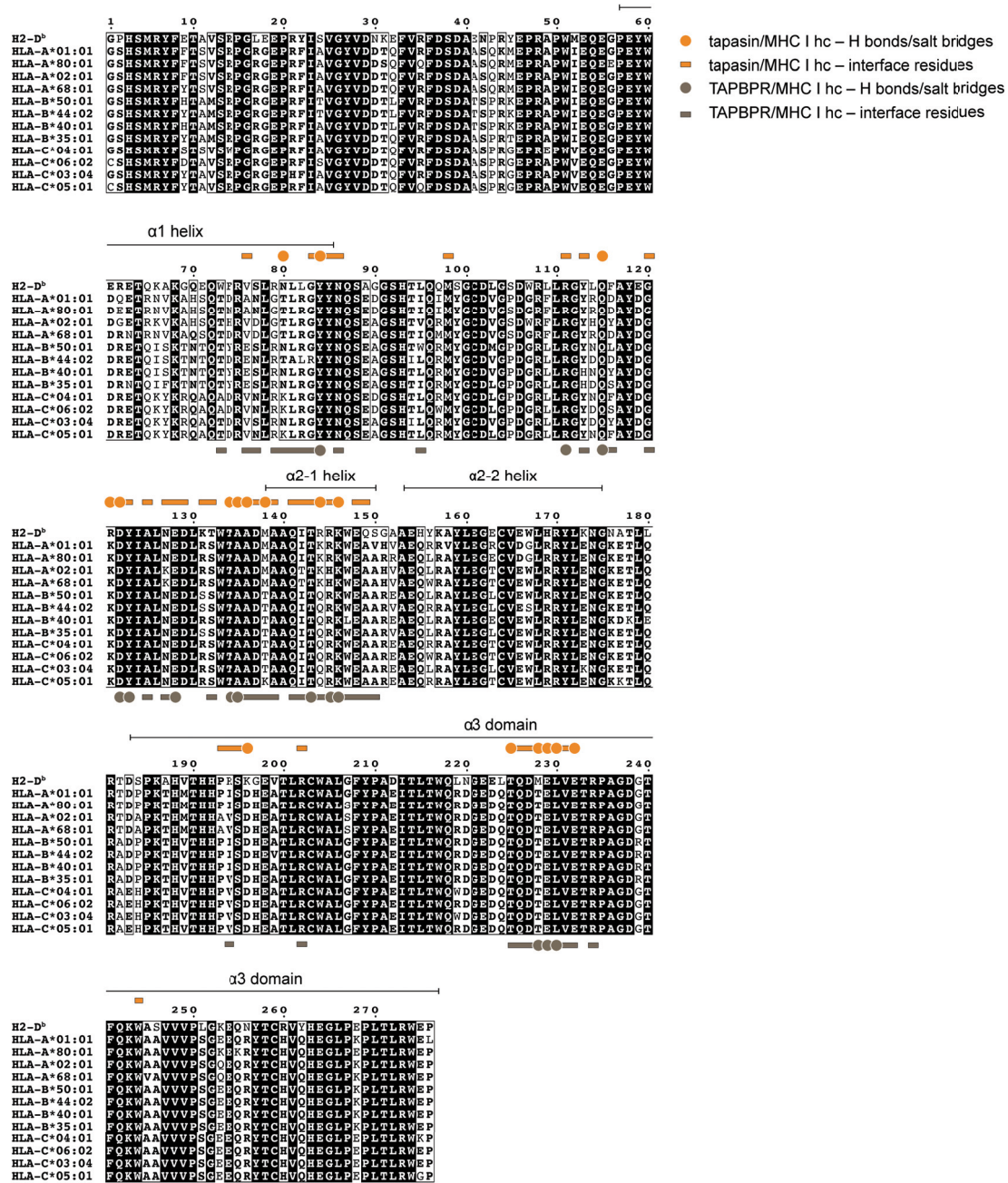
Supplementary Fig. 6 Multiple sequence alignment of classical MHC I heavy chains. Conserved residues are highlighted in black, and physiochemically related residues are represented in bold letters. Orange dots and bars mark residues of MHC I hc that are involved in interactions with tapasin. Residues of MHC I hc engaged in interaction with TAPBPR are marked as for tapasin but with brown symbols. The dots highlight specific interactions, such as hydrogen bonds and salt bridges. hc, heavy chain.
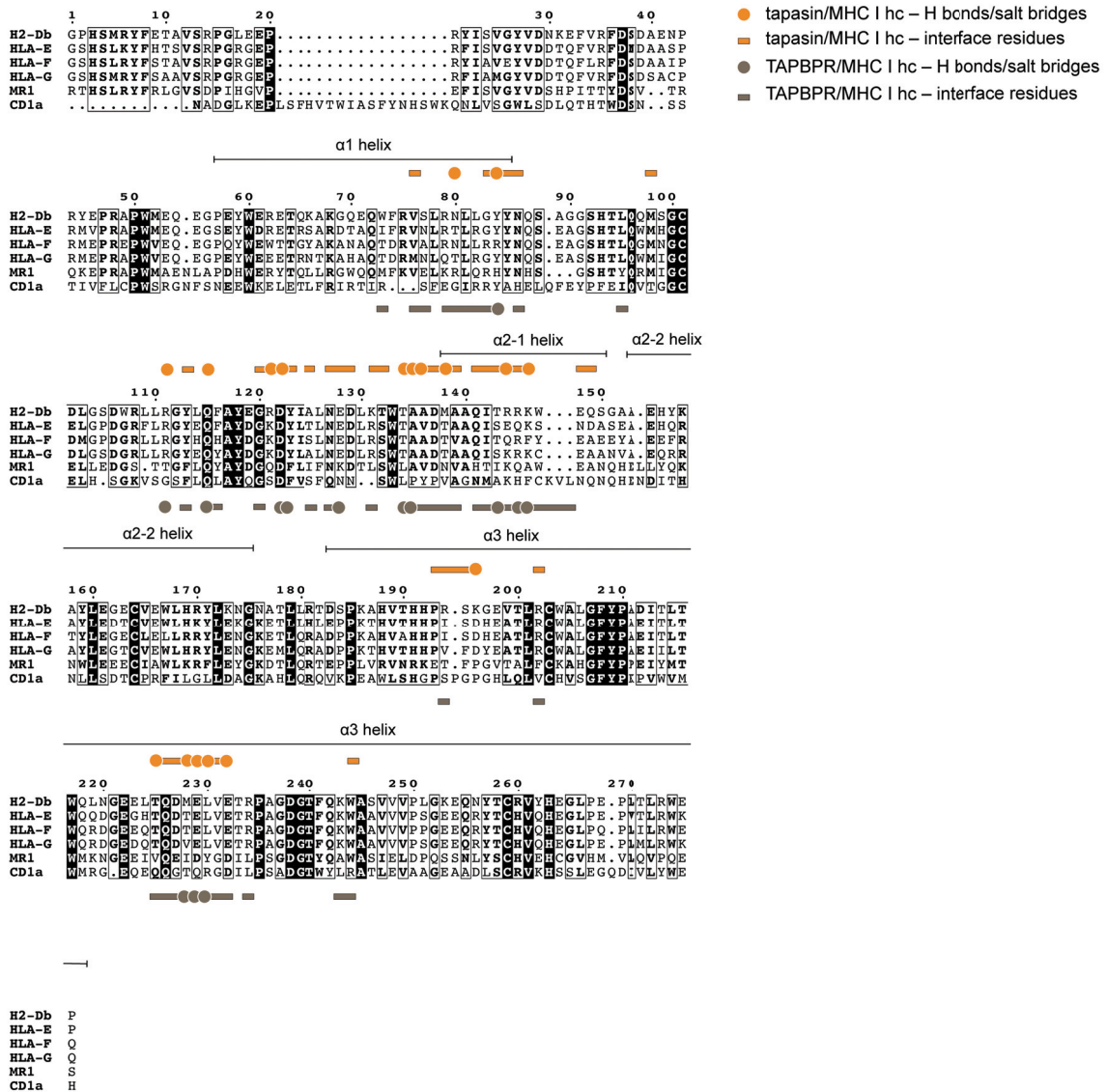
Supplementary Fig. 7 Multiple sequence alignment of non-classical MHC I and related molecules. Conserved residues are highlighted in black and physiochemically related residues are represented in bold letters. Orange dots and bars mark residues of MHC I hc that are involved in interactions with tapasin. Residues of MHC I hc engaged in interaction with TAPBPR are as for tapasin but marked with brown symbols. The dots highlight specific interactions, such as hydrogen bonds and salt bridges. hc, heavy chain.
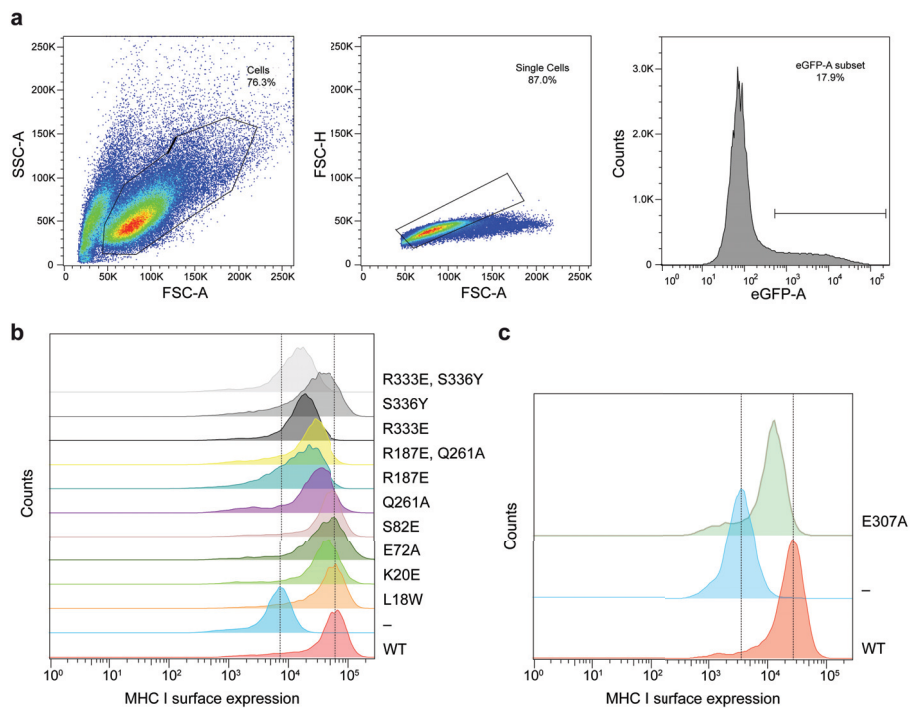
Supplementary Fig. 8 Gating strategy and histograms of flow cytometry analyses. **a** Transfected tapasin-deficient HAP1 cells were first gated based on their size and granularity (SSC-A/FSC-A), duplets were discriminated (FSC-H/FSC-A), and final gate was set based on eGFP fluorescence (histogram). **b** Histogram representation of MHC I surface expression of wildtype (red, WT; mock, blue) and F-pocket (orange, L18W; green, K20E; dark green, E72A), interface loop (purple, Q261A), β hairpin (blue-green, R187E; yellow, R187E, Q261A), α3-domain (dark grey, R333E; grey, S336Y; light grey, R333E, S336Y) interface mutants of tapasin, shown in Fig. 3e and Fig. 4f. **c**, Histogram representation of surface MHC I of wildtype (red, WT; blue, mock) and β_2m interface mutant of tapasin (green, E307A) shown in Fig. 4f. WT, wildtype; −, mock transfection.

SUPPLEMENTARY TABLES

Supplementary Table 1 Data collection and refinement statistics. The structure of MHC I–tapasin–ERp57 complex was determined from one protein crystal. Statistics for the highest-resolution shell are given in parentheses.

	MHC I–tapasin–ERp57 complex
Wavelength	1.033
Resolution range (Å)	48.0 - 2.7 (2.8 - 2.7)
Space group	$P22_12_1$
Unit cell dimensions (Å)	a = 74.76, b = 168.53, c = 187.05
Total reflections	388317 (38711)
Unique reflections	65719 (6491)
Multiplicity	5.9 (6.0)
Completeness (%)	99.76 (99.86)
Mean I/sigma(I)	14.90 (1.02)
Wilson B-factor	87.61
R-merge	0.07144 (1.507)
R-meas	0.0784 (1.65)
R-pim	0.03178 (0.6626)
CC1/2	0.998 (0.443)
CC*	1 (0.783)
Reflections used in refinement	65719 (6488)
Reflections used for R-free	3284 (324)
R-work	0.1983 (0.3229)
R-free	0.2289 (0.3445)
Number of non-hydrogen atoms	9726
macromolecules	9635
ligands	39
solvent	52
Protein residues	1214
RMS(bonds)	0.009
RMS(angles)	1.36
Ramachandran favored (%)	95.00
Ramachandran allowed (%)	5.00
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.10
Clashscore	5.46
MolProbity score	1.64
Average B-factor	106.17
macromolecules	106.23
ligands	128.14
solvent	79.25
Number of TLS groups	21

Supplementary Table 2 Primers used in this study.

Primer	Sequence (5'-3')
BamHI_Tapasin_ΔTMD_for	CTGTTTGGATCCATGAAGTCCCTGTCTCTGCTC
Tapasin_ΔTMD_His6_HindIII_rev	CTCGACAAGCTTTCATGATCCATGATGATGATGATGGTGTGATCCCTCCAGGGTGACCTCAG
BamHI_ERp57_C36A_for	TGTTTGGATCCATGCGTCTCCGTCGCCTAGCGC
ERp57_C36A_XbaI_rev	CAGGCTCTAGATTAGAGATCCTCCTGTGCCTTCTTC
Tapasin_ ΔTMD_TEV_His6_SLIC_for	GAAAACCTGTATTTTCAGGGCCACCATCATCATCATGG
Tapasin_ ΔTMD_TEV_His6_SLIC_rev	CAGACTTACCGAGAGtgCCTGCGCAC
Tapasin_NotI_SLIC_for	AAGCTTGTCGCGGCCGCGAGAAGTACTAG
Tapasin_NotI_SLIC_rev	GCGGCCGCGACAAGCTTTCACTC
NotI_IRES2_eGFP_for	CGCGCGGCCGCCCCCTCTCCCTCCCCCCCCCCTAAC
IRES2_eGFP_ScaI_rev	CGCAGTACTTTACTTGTACAGCTCGTCCATGCCG
Tapasin_L18W_SLIC_for	GAAAGGGCTGGGCCAAGAGACCGGTG
Tapasin_L18W_SLIC_rev	CTTGGCCCAGCCCTTTCCGCTCGC
Tapasin_K20E_SLIC_for	CCTGGCCGAGAGACCCGGTGCAC
Tapasin_K20E_SLIC_rev	GGGTCTCTCGGCCAGGCCCTTTCC
Tapasin_E72A_SLIC_for	CCACACTGCGCGATGAGCCGCTTCG
Tapasin_E72A_SLIC_rev	CTCATCGCGCAGTGTGGTGCGGG
Tapasin_Q261A_SLIC_for	GCAAGGAGCGGTCACCCTGGAGCTTG
Tapasin_Q261A_SLIC_rev	GGGTGACCGCTCCTTGCAGGTATGGC
Tapasin_R187E_SLIC_for	CTAGAGTGGGAACGCCAGCACCTGGG
Tapasin_R187E_SLIC_rev	GTGCTGGCGTTCCCACTCTAGCCCAAAG
Tapasin_R333E_SLIC_for	CGGCCCTGGAACACCATTCCGATGGCTC
Tapasin_R333E_SLIC_rev	GAATGGTGTTCCAGGGCCGAGAGCCACC

Tapasin_S336Y_SLIC_for	GCGCCACCATTACGATGGCTCTGTCAGC
Tapasin_S336Y_SLIC_rev	CATCGTAATGGTGGCGCAGGGCCG
Tapasin_E307A_SLIC_for	CAGGGGCGGCACCCCCGG
Tapasin_E307A_SLIC_rev	GGTGCCGCCCCTGGGGCG

Showcasing research from Professor Robert Tampé's laboratory, Institute of Biochemistry, Goethe University Frankfurt, Germany, in collaboration with Professor Alexander Deiters, Department of Chemistry, University of Pittsburgh, PA, USA.

Light-guided intrabodies for on-demand *in situ* target recognition in human cells

Nanobodies are ideal probes for visualizing intracellular proteins. Inside living cells, unrestrained binding can cause interference with target function or localization. Here, we report a strategy to circumvent interference through the development of photo-conditional intrabodies. To regulate the interaction, we combine optochemical biology and genetic code expansion in stable cell lines. By equipping the paratope with photocaged amino acids, we control target binding with high spatiotemporal precision inside living cells. Due to the highly stable binding, light-guided intrabodies offer a versatile platform for comprehensive imaging and modulation of target proteins.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

Check for updates

Light-guided intrabodies for on-demand *in situ* target recognition in human cells†

Eike F. Joest, [a] Christian Winter, [a] Joshua S. Wesalo, [b] Alexander Deiters [b] and Robert Tampé *[a]

Due to their high stability and specificity in living cells, fluorescently labeled nanobodies are perfect probes for visualizing intracellular targets at an endogenous level. However, intrabodies bind unrestrainedly and hence may interfere with the target protein function. Here, we report a strategy to prevent premature binding through the development of photo-conditional intrabodies. Using genetic code expansion, we introduce photocaged amino acids within the nanobody-binding interface, which, after photo-activation, show instantaneous binding of target proteins with high spatiotemporal precision inside living cells. Due to the highly stable binding, light-guided intrabodies offer a versatile platform for downstream imaging and regulation of target proteins.

Introduction

Nanobodies or V_HHs are single-domain binders derived from heavy chain-only antibodies of camelid or nurse shark species.[1–3] These minimal antigen-binding fragments of 12 to 15 kDa are associated with nanomolar to subnanomolar affinity combined with high chemical and thermal stability.[4,5] They quickly evolved into versatile tools for a broad spectrum of technologies, especially for imaging inside living cells.[6–12] For fluorescent labeling, the nanobodies are typically coupled to organic dyes or fused to auto-fluorescent proteins, named chromobodies.[14] For intracellular live-cell imaging, the nanobodies can be delivered by emerging technologies, such as cell squeezing, or are recombinantly expressed by transient transfection, termed intrabodies.[6,14,15] Although the latter requires minimal efforts, the unbalanced level of unbound nanobodies leads to blurred signal-to-background ratios. Furthermore, it remains unclear whether the uncontrolled binding of nanobodies interferes with the function, assembly, and subcellular dynamics of intracellular target proteins before the actual analysis.[16] To surpass possible artifacts derived from premature binding, a major advancement would be to control intrabody binding by light in living mammalian cells.

Nanobodies are also powerful tools for the modulation of target proteins. Their binding can directly influence target protein function or guide other proteins that offer control. Hence, a photo-conditional intrabody is a versatile platform for

in vivo regulation of target proteins. Here, we established stable human cell lines allowing synthesis of photo-conditional intrabodies. To this end, we expanded the genetic code of mammalian cells to incorporate amino acids equipped with a bulky photocage in the epitope-binding site.[17,18] Genes coding for amber-suppressed intrabodies C-terminally fused to mCherry were site-specifically inserted into the genome of human cells using a recombinase system.[19] Optimized amber codon suppression evoked the site-specific incorporation of photocaged amino acids and hence enabled intrabody expression. We traced the fluorescent intrabodies by flow cytometry or confocal laser-scanning microscopy (CLSM). On demand, we controlled the intrabody binding to its target in single cells. Immediately after a short pulse of light, we observed a complete subcellular reorganization of the intrabody towards the cognate target protein. Furthermore, fine-tuning of the exposed area and the exposure time enabled pulse-chase labeling. Our spatio-temporally controlled intrabody labeling prevents potential artifacts caused by constitutively active binding and hence offers advanced real-time studies of physiologically unrestricted target proteins.

Results and discussion

Light-guided intrabodies by genetic code expansion

As a proof of principle, we selected the well-described GFP-binding nanobody and recombinantly fused it to the red-fluorescent protein mCherry ($Nb^{mCherry}$).[4,13] Based on the dimer crystal structure,[13] we identified several amino acids within the epitope-binding region that could be exchanged towards a photocaged analog to block binding. Based on its crucial location, we selected tyrosine at position 37 (Tyr37) for amber suppression (Fig. 1A). For photo-caging of Tyr37, we used

[a]Institute of Biochemistry, Biocenter, Goethe University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany. E-mail: tampe@em.uni-frankfurt.de

[b]Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, Pennsylvania 15260, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc01331a
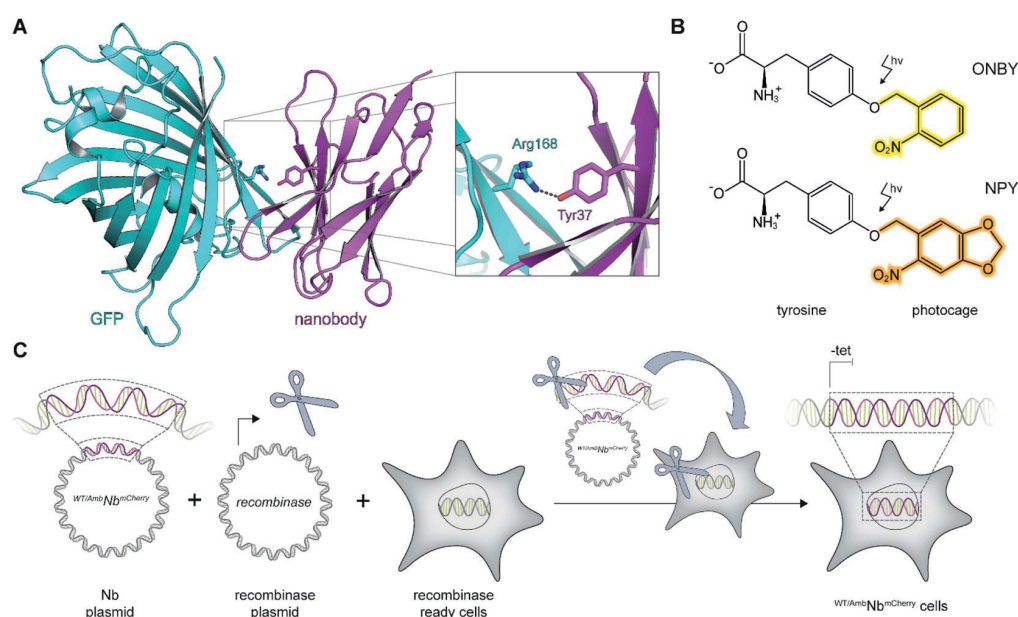
Fig. 1 On-demand tracing of intracellular targets by photo-activatable intrabodies. (A) The X-ray structure of the nanobody-GFP complex (pdb: 3K1K)[13] reveals Tyr37 as a key residue in the binding region to GFP (turquoise). Nanobody, magenta. (B) Tyr37 is replaced by *ortho*-nitrobenzyl- or nitropiperonyl-caged tyrosines (ONBY or NPY, respectively). (C) Schematic illustration of stable cell line generation for expression of photo-activated intrabodies by genetic code expansion. Mammalian cells are co-transfected with plasmids encoding the recombinase and either wild-type (WTNbmCherry) or amber-containing nanobodies (AmbNbmCherry).

either *ortho*-nitrobenzyl-caged tyrosine (ONBY) or the novel nitropiperonyl-caged tyrosine (NPY) with improved light-activation properties (Fig. 1B).[20,21] We used an optimized pyrrolysyl-tRNA synthetase (NPYRS)/tRNA pair[21] for site-specific incorporation.

To prevent multiple transfections, we reasoned that it would be beneficial to stably integrate the nanobody constructs into the cellular genome. Thus, we used the Flp-In™ T-REx™ recombinase system and integrated the wild-type (WTNbmCherry) and amber-containing (AmbNbmCherry) constructs in HeLa cells.[19] The integrated tetracycline inducible promotor allowed tight regulation of the intrabody expression (Fig. 1C).

Expression of intrabodies by stable cell lines

First, we analyzed the expression of WTNbmCherry or AmbNbmCherry in the stable cell lines by using the C-terminally fused mCherry as a reporter. After tetracycline induction, WTNbmCherry was expressed at high levels as monitored by flow cytometry (Fig. 2A and B; ESI Fig. 1 and 2†). Live-cell CLSM imaging visualized fluorescent nanobody localization in the cytosol (Fig. 2A and C; ESI Fig. 3†). For the amber-suppressed AmbNbmCherry, we established monoclonal cell lines based on the strongest mCherry signal referenced to cells cultured in the absence of the caged tyrosine. We combined tetracycline induction (0.1 μg ml^{-1}) with transfection of the NPYRS/tRNA plasmid and added the photocaged amino acid (0.25 mM) 4–6 h later to the media. The expression of AmbNbmCherry in the established monoclonal cell line was further examined by flow cytometry. In contrast to untreated cells or the non-transduced

cell line, a strong mCherry signal of AmbNbmCherry-positive cells was only observed under amber suppression conditions with either ONBY or NPY (Fig. 2D–F; ESI Fig. 1 and 4†). Strong fluorescence with equal cytosolic distribution of AmbNbmCherry was visualized by live-cell CLSM imaging in the presence of ONBY or NPY (Fig. 2G; ESI Fig. 5†), confirming the incorporation of caged tyrosine at the amber codon. In the absence of the photocaged tyrosines or NPYRS/tRNA, the premature amber stop codon aborted the translation of full-length AmbNbmCherry.

Next, we examined the intrabody properties by monitoring specific binding to different GFP-tagged target proteins. We transfected plasmids encoding the nuclear envelope protein LaminA, which was N-terminally tagged with mEGFP (mEGFPLaminA), or the histone H2B, which was C-terminally tagged with EGFP (H2BEGFP) (Fig. 2A and D).[22,23] After tetracycline induction, WTNbmCherry-positive cells showed the expected colocalization of GFP and mCherry fluorescence. It is worth mentioning that the amber suppression components did not affect the binding of the wildtype intrabody and hence the colocalization of GFP and mCherry (Fig. 2C; ESI Fig. 6 and 7†).

We subsequently analyzed binding in AmbNbmCherry-positive HeLa cells. After transient transfection of the target genes, we did not record a colocalization of the AmbNbmCherry with the two GFP-tagged targets under amber suppression conditions, demonstrating that both photocages, ONBY and NPY, block constitutive intrabody binding (Fig. 2G; ESI Fig. 8–10†). However, after exposing cells to a short 405 nm light pulse to induce photo-cleavage of the caging group, we observed an instantaneous subcellular reorganization of AmbNbmCherry. Corroborated by colocalization in living cells, the intrabody
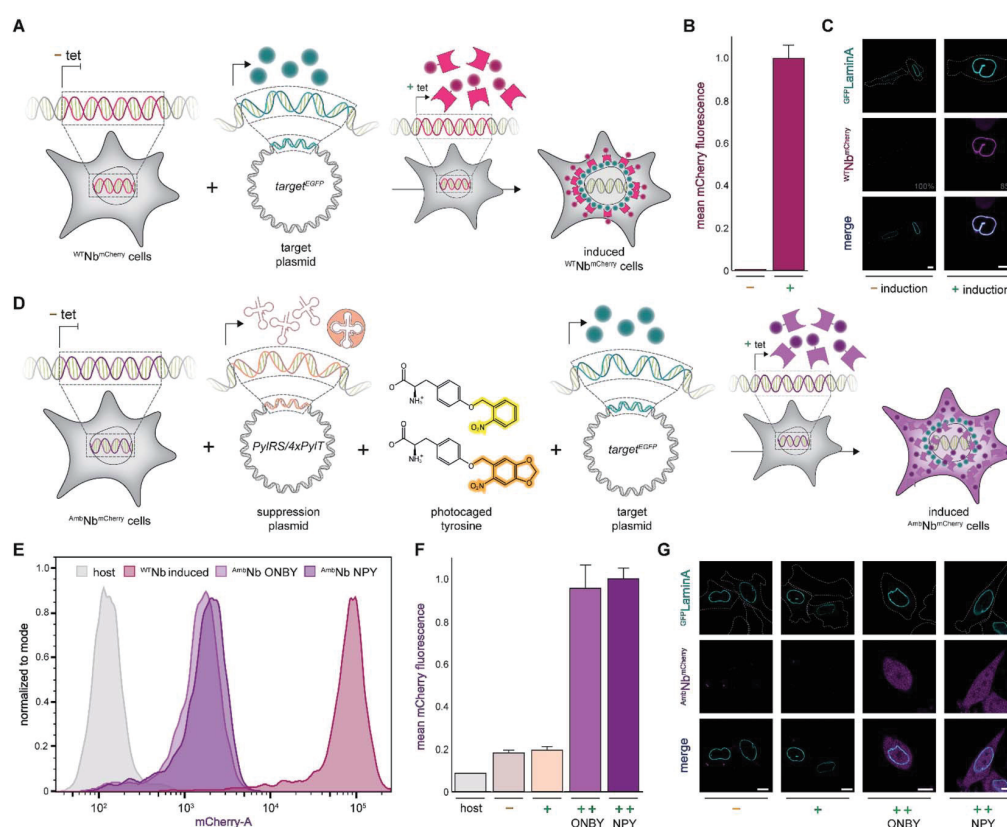
Fig. 2 Expression of photo-activatable intrabodies by stable cell lines. (A) Wild-type intrabody $^{WT}Nb^{mCherry}$ is expressed after tetracycline induction. For intracellular binding analysis, a plasmid encoding a GFP-tagged target is transiently transfected; tet, tetracycline. (B) $^{WT}Nb^{mCherry}$ expression analyzed by flow cytometry without (−) or with (+) tetracycline induction. (C) Live-cell imaging showing expression of $^{WT}Nb^{mCherry}$ and binding of $^{mEGFP}LaminA$ (relative PMT gain settings from the image acquisition in grey). (D) Intrabody $^{Amb}Nb^{mCherry}$ expression by amber suppression with photocaged tyrosines. $^{Amb}Nb^{mCherry}$-positive cells are transiently transfected with plasmids encoding the optimized NPYRS/ tRNA pair and a target protein, respectively. After transfection, $^{Amb}Nb^{mCherry}$ expression was induced when cells were supplied with photocaged tyrosines. (E) Expression of $^{WT}Nb^{mCherry}$ and $^{Amb}Nb^{mCherry}$ monitored by flow cytometry. $^{WT}Nb^{mCherry}$ expression was tetracycline-induced and $^{Amb}Nb^{mCherry}$ was expressed using amber suppression conditions. Cell counts were normalized to mode ($n = 3$). (F) Amber suppression conditions enabling $^{Amb}Nb^{mCherry}$ expression with photocaged amino acids. Mean fluorescence of monoclonal $^{Amb}Nb^{mCherry}$-positive cells analyzed by flow cytometry ($n = 3$). Normalized to expression with amber suppression conditions in the presence of NPY. (G) Live-cell imaging of $^{Amb}Nb^{mCherry}$ revealing blocking of binding to $^{GFP}LaminA$ through incorporation of photocaged ONBY or NPY at Tyr37. In (F) and (G): (−) no supplements, (+) transient transfection of NPYRS/tRNA, (++) transient transfection of NPYRS/tRNA and tetracycline induction. In (C) and (G): scale bar = 10 μm.

bound either to $^{mEGFP}LaminA$ at the nuclear envelope or to histone H2B^{EGFP} in the nucleus, using either ONBY or NPY (Fig. 3A and B; ESI Fig. 9 and 10†). Furthermore, we repeated the experiment with $^{WT}Nb^{mCherry}$-expressing cells to exclude phototoxic artifacts.[24] The location and constitutive binding of $^{WT}Nb^{mCherry}$ remained unaffected, especially in case of cells with saturating nanobody amounts (ESI Fig. 11 and 12†). Finally, we analyzed photo-activated nanobody-EGFP binding in cell lysates. By co-immunoprecipitation, we captured $^{WT/Amb}Nb^{mCherry}$ and associated EGFP from the respective lysates. Subsequently, specific binding in cell lysates was quantified using EGFP fluorescence. After photo-activation, $^{Amb}Nb^{mCherry}$ resembled the EGFP enrichment obtained by the same amount of $^{WT}Nb^{mCherry}$ (ESI Fig. 13†). The results demonstrated the efficiency and specificity of light-triggered restoration of epitope recognition.

Photo-activation with high spatiotemporal precision

We next focused on the intracellular activation of $^{Amb}Nb^{mCherry}$ by light, visualizing the decoration of $^{mEGFP}LaminA$ as a target. For statistical analysis of the induced binding, we quantified the increase in colocalization after illumination. Therefore, we observed strong predominance of cells containing high amounts of target protein (ESI Fig. 14†). To monitor target binding in 3D, we recorded high-resolution z-stacks before and after photo-activation. Before photo-activation, we observed an equally distributed cytosolic fluorescence of $^{Amb}Nb^{mCherry}$ and, immediately after illumination, a background-free colocalization by intrabody binding (Fig. 3C). Full photo-activation was achieved within the first few minutes of illumination. Apart from the temporal resolution, we also investigated the spatial precision of photo-activation by step-wise exposing cells in close
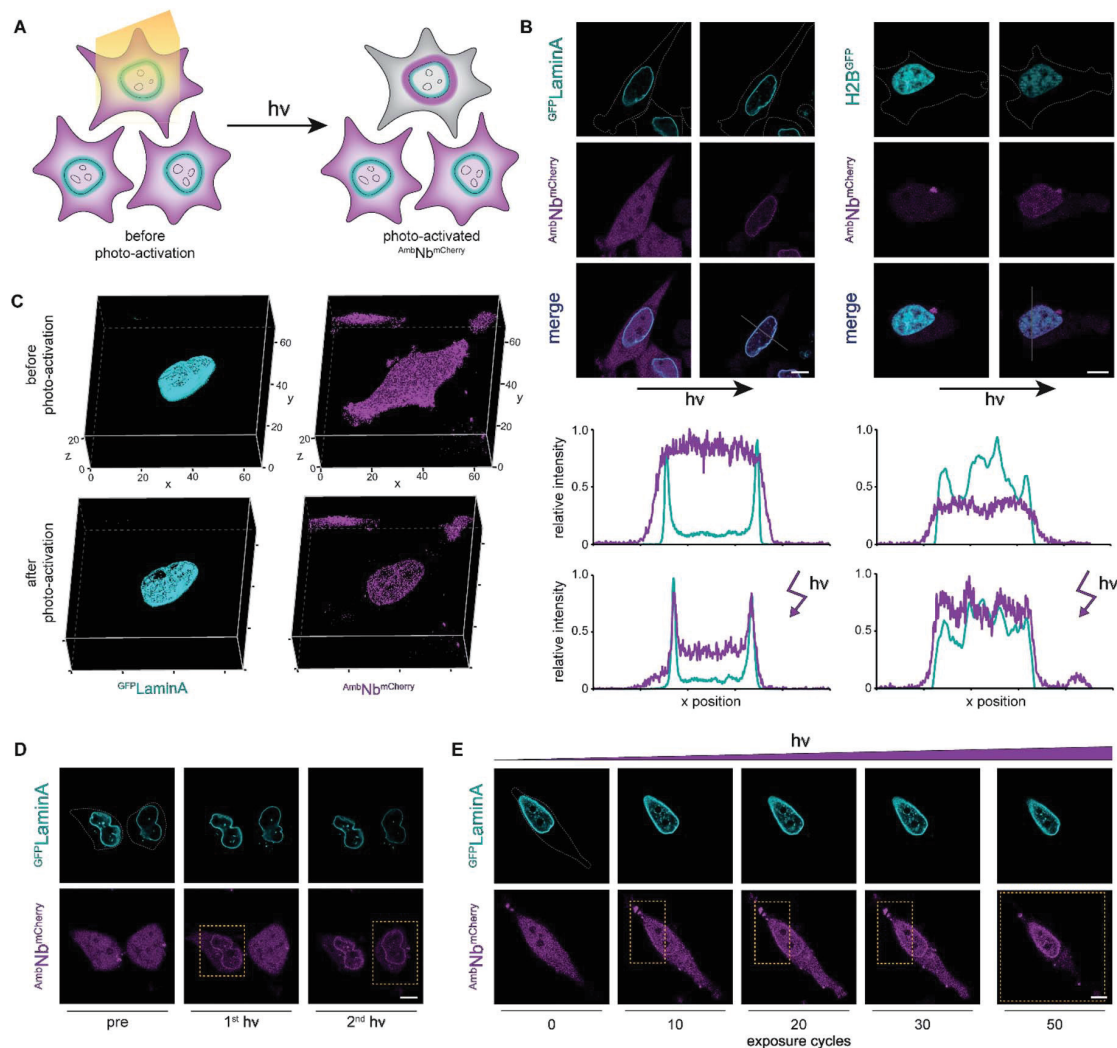
Fig. 3 Photo-activation of intrabodies with high spatiotemporal precision. (A) Light exposure unleashes the intrabody and hence allows instant binding to the target protein. (B) Live-cell imaging and photo-activation of $^{Amb}Nb^{mCherry}$. After photo-activation, $^{Amb}Nb^{mCherry}$ and target protein colocalization was observed. Corresponding line scans highlight intrabody binding. (C) 3D imaging revealing almost background-free target binding of $^{Amb}Nb^{mCherry}$ after illumination. (D) Single-cell intrabody photo-activation with high local precision. (E) Fine-tuned $^{Amb}Nb^{mCherry}$ photo-activation by local energy dosage. In (D) and (E), yellow boxes indicate the light-exposed area. Maximum light exposure corresponds to a bleaching function with 250 iterations and 50 cycles using a 405 nm diode laser (4.5 mW μm^{-2}). The whole region of interest (ROI) or the indicated rectangular selection was illuminated, respectively. Cells were supplemented with NPY.

proximity. We analyzed $^{Amb}Nb^{mCherry}$-positive cells, treated with NPY and expressing $^{mEGFP}LaminA$, that were within less than 5 μm distance. We observed specific activation of individual cells and no target protein engagement in adjacent cells (Fig. 3D). For improved control of binding within an individual $^{Amb}Nb^{mCherry}$-positive cell, we carefully increased the light exposure and recorded the reorganization immediately afterwards. Nanobody binding, monitored by colocalization, strictly correlated with light exposure, offering complete or graded levels of activation (ESI Fig. 15 and 16†). The improved photo-activation properties of NPY allowed saturating activation at 20–40% of light exposure as compared to ONBY (ESI Fig. 17†). Finally, for fine-tuned activation, we only exposed a defined small part of NPY

$^{Amb}Nb^{mCherry}$-positive cells with distinct exposure times. We achieved stringent control of activation and hence intrabody binding, as evidenced by a progressive increase of colocalization (Fig. 3E, ESI Fig. 15†). In summary, these results demonstrate fast photo-controlled nanobody binding and high spatiotemporal precision by using an optimized cellular amber suppression system.

Conclusions

In this study, we generated stable mammalian cell lines for expression of photo-conditional intrabodies by genetic code expansion. We site-specifically incorporated photocages within

the epitope-binding site of the α-GFP nanobody at position Tyr37.[13,20,21] While our studies were ongoing, this design approach was further validated through the incorporation of caged tyrosines into a nanobody expressed in *E. coli*.[25] In our stable mammalian cell line, we monitored amber suppression by flow cytometry and CLSM using a fluorescent reporter. We revealed that incorporation of the photocaged tyrosines ONBY and NPY prevented constitutive binding in the amber-suppressed [Amb]Nb[mCherry]. Cleavage of the photocages by light restored the epitope-binding site and allowed instantaneous target binding. Within a few minutes of illumination, a complete reorganization of the intrabody by colocalization to the target protein was observed, demonstrating efficient photo-activation. Fine-tuning the exposed area and exposure time finally allowed strict control of intrabody binding.

Previous studies have reported on nanobody–photoreceptor hybrids. The combination of optogenetics and nanobodies with fast dissociation rates facilitated reversible intracellular binding.[26–28] Here, we demonstrate a fundamentally different approach using (opto)chemical biology to control a nanobody with kinetically stable, long-term binding.[4,13] Both approaches offer intrinsic advantages. Optogenetic control allows reversible activation after comprehensive bioengineering, while opto-chemical activation by photo-deprotection can be rationally designed based on structural information and restores the native protein. Thus, these two methods form a complementary toolbox to control intrabodies with light. The focus of our study is on the on-demand intrabody binding by genetic code expansion. The controlled interaction circumvents potential interference with the target protein caused by constitutive binding.

We established a stable cell line for expression of the intra-body [WT]Nb[mCherry]. In future studies, a combination of both cell lines can be used to identify target interference caused by the constitutive binding of [WT]Nb[mCherry]. In addition, we used an intrabody with kinetically stable target binding, which facili-tated comprehensive imaging experiments. Stable binding forms the basis for potential target protein modulation.[29] To visualize intrabody expression and binding, we used a C-terminally fused mCherry. Alternatively, the fluorescent reporter of the intrabody could be replaced by a proteasomal degradational signal.[29–31] This will allow on-demand post-translational target knock-downs in defined regions. Addition-ally, several other photocaged amino acids with optimized PylRS/tRNA pairs and nanobodies are available. Combined with our general approach, photo-activatable nanobodies for various functions can be generated. Particularly, the implementation of advanced photocages enables deep tissue activation and intra-cellular labeling lithography.[32–36] Moreover, new CRISP-Cas12a-guided methods allow the versatile tagging of endogenous proteins with small peptide tags, which can be detected by photo-activatable intracellular nanobodies.[37]

Genetic code expansion in stable cell lines harboring a respective PylRS/tRNA pair and an amber-codon construct is challenging and not always beneficial.[38] We chose a hybrid approach by stably integrating the amber-modified target gene using a recombinase and transiently transfecting a plasmid

encoding the NPYRS/tRNA pair. The combination of transient transfection and tetracycline-induction of a stable cell line allows parallel expression of several different proteins. There-fore, we achieved intrabody synthesis with non-natural tyro-sines by amber suppression using the optimized PylRS/tRNA pair, and simultaneously expressed an additional target protein. The parallel expression of different proteins and the incorpo-ration of photo-conditional tyrosines offer further powerful applications in basic and applied research. Post-translational modification of tyrosine residues plays a major role in cellular biochemistry.[39] Due to the easy transferability, our approach can serve as a blueprint to control signaling proteins in complex networks.[40] In particular, in connection with receptor tyrosine kinases, our stable cell line concept allows modulation of *trans*-phosphorylation and the resulting signaling.[41–44] Thus, the parallel expression of other genes will pave the way for comprehensive analysis of downstream processes.

In summary, genetic code expansion facilitates the precise control of intrabody binding by light. The conditional interac-tion circumvents potential artifacts of constitutively binding nanobodies and can guide nanobody-mediated target protein modulation. In general, our stable cell line approach is a versatile proof-of-concept for analyzing proteins in complex pathways *via* photo-caged tyrosine residues. We are convinced that, together with other extracellular photo-activatable binders, these new experimental possibilities will enlarge and improve the broad spectrum of nanobody applications.[25,45,46]

Experimental

Molecular biology

Enzymes for molecular biology were used according to the manufacturers' protocols. DNA amplification was performed with the Phusion™ High-Fidelity DNA Polymerase. For ligation, T4 DNA ligase was applied. The α-GFP[mCherry] (ref. 13) construct was cloned into a pcDNA3.1(+) plasmid by PCR.[6] The amber mutation Tyr37TAG was introduced by site-directed mutagen-esis using the following primers: fwd 5′-AGC ATG CGT TGG **TAG** CGT CAG GCA CCG-3′, rev 5′-CGG TGC CTG ACG **CTA** CCA ACG CAT GCT-3′ (mutation in bold). The target plasmid containing the [mEGFP]LaminA construct was generated by site-directed mutagenesis and PCR as previously described.[22,47] H2B[EGFP] in pEGFP-N1 was a gift from Geoffrey Wahl (Addgene #11680).[23] The plasmid encoding the optimized NPYRS/4xPylT pair was previously generated.[21] The pyrrolysine tRNA synthetase (PylRS) of *Methanosarcina barkeri* was modified with the mutations L270F, L274M, N311G, and C313G.[21] For the generation of the stable cell line by the Flp-In™ T-REx™ system, the amber-free and amber-containing α-GFP[mCherry] constructs were PCR amplified and cloned into the pcDNA5/FRT/TO plasmid (Thermo Fisher Scientific). Therefore, a primer pair introducing an upstream HindIII restriction site and a downstream NotI restriction site was utilized: HindIII fwd 5′-GCG CGC <u>AAG CTT</u> ACC ATG CAG GTT CAG CTG GTT GAA AGC GGT GGT G-3′, NotI rev 5′-GCG CGC <u>GCG GCC GCC</u> TAC TTG TAC AGC TCG TCC ATG CCG CCG-3′ (restriction sites are underlined).

Cell culture and generation of a stable cell line

The generation of the stable HeLa cell line was performed by co-transfection of the respective pcDNA5/FRT/TO constructs, the pOG44 Flp-recombinase expression vector (Thermo Fisher Scientific), and blasticidin (Gibco) and hygromycin (Invitrogen) selection, following the manufacturer's protocol. Transient transfection by Lipofectamine 2000 (Life Technologies) was performed according to the manufacturer's guidelines. Unmodified T-REx™-HeLa cells (Thermo Fisher Scientific) were maintained in DMEM medium containing 4.5 g L^{-1} glucose (Gibco), supplemented with 10% (v/v) FCS tetracycline-negative (BIO&SELL) in T75 cell culture flasks (Greiner). After recombinase reaction, the medium was supplemented with 2 μg ml^{-1} blasticidin (Gibco) and 100 μg ml^{-1} hygromycin (Invitrogen) according to the manufacturer's instructions. Cell passage was performed by using PBS (Sigma-Aldrich) and 0.05% trypsin–EDTA (Gibco) every 2–3 days. Cells were cultivated in a tissue culture incubator at 37 °C and humidified with 5% CO_2. Following established guidelines, mycoplasma contamination tests were regularly carried out.[48] Stable monoclonal cell lines were established by flow cytometry cell sorting based on highest fluorescence (see below). For live-cell imaging, 2.5×10^4 cells were seeded per glass-bottom 8-well imaging slide (Sarstedt). For flow cytometry, 5×10^5 cells were seeded per 6-well (Greiner), harvested using PBS (Sigma-Aldrich) and 0.05% trypsin–EDTA (Gibco), washed once with normal selection medium and once with PBS (Sigma-Aldrich).

Intrabody gene expression and genetic code expansion

The expression cassette of the Flp-In™ T-REx™ expression cell lines contained a tetracycline inducible promotor. For induction of gene expression, we followed the manufacturer's instruction and used 0.1 μg ml^{-1} tetracycline. For genetic code expansion by amber suppression in $^{Amb}Nb^{mCherry}$-positive cells, transient transfection of the NPYRS/4xPylT encoding plasmid was followed 4–6 h later by exchanging the medium with amber suppression medium (DMEM, 4.5 g L^{-1} glucose (Gibco), 10% (v/v) FCS tetracycline-negative (BIO&SELL), 0.1 μg ml^{-1} tetracycline, and 0.25 mM ONBY (Santa Cruz Biotechnology) or NPY). NPY was synthesized as previously described.[21] Transient transfection or exposure to unnatural amino acids is stressful for cells. For optimal culture conditions, we decided to use an established approach for mammalian cells with 4–6 h solely for transfection.[49,50] Unnatural amino acids were dissolved, sterile-filtered and stored in 100 mM NaOH. The compounds were used at the limit of solubility. We recommend preparing a stock solution of hardly soluble NPY by end-over-end rotation for 2 h at room temperature. After supplementing DMEM with unnatural amino acids in NaOH, the pH was neutralized by the same volume of 100 mM sterile filtered HCl. Cells were cultivated in the corresponding medium for 24–48 h before experimental analysis.

CLSM live-cell imaging and photo-activation

For live-cell imaging of $^{WT}Nb^{mCherry}$, cells were transiently transfected with ^{mEGFP}LaminA or H2B^{EGFP}-encoding plasmids.

For $^{Amb}Nb^{mCherry}$-positive cells, the NPYRS/4xPylT plasmid was transiently co-transfected (1 : 1 plasmid ratio) with the corresponding target plasmid, prior to addition of amber suppression medium. For transfection, Lipofectamine 2000 (Life Technologies) was used according to the manufacturer's instructions. Images were recorded 24–48 h after medium exchange or induction, respectively. For live-cell imaging, cells were covered by Ringer solution (145 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, 10 mM glucose, pH 7.4), incubated at 37 °C and humidified with 5% CO_2. Imaging was performed by using a confocal laser-scanning microscope, Zeiss LSM 880 (Carl Zeiss Jena GmbH, Germany), combined with a Plan Appochromat 63x/1.4 oil DIC objective. In order to avoid crosstalk, sequential imaging was used. EGFP excitation was achieved using a 488 nm argon laser and mCherry excitation using a 543 nm helium–neon laser. For EGFP imaging, the gain was adjusted to the level of transient transfection. mCherry was recorded using similar settings for each shown experiment, and relative alterations were indicated. Photo-activation was performed by using a bleaching function with 250 iterations, 50 cycles and a 405 nm diode laser (4.5 mW μm^{-2}) for the whole ROI or with a rectangular selection, if indicated. For 3D rendering of z stacks, the ZenBlack (Carl Zeiss Jena GmbH, Germany) surface function was applied. Images were analyzed using ZenBlue (Carl Zeiss Jena GmbH, Germany), Excel (Microsoft) and Fiji.[51]

Flow cytometry

For monoclonal sorting, $^{Amb}Nb^{mCherry}$-positive cells were transiently transfected with the NPYRS/4xPylT-encoding plasmid, incubated in media supplemented with tetracycline and NPY, and analyzed 48 h after transfection. Cells with highest fluorescence were individually collected in 96-well plates (Greiner) and expanded to culture size. For expression analysis 48 h after transient transfection of the NPYRS/4xPylT-encoding plasmid, the fluorescence of $^{Amb}Nb^{mCherry}$- or $^{WT}Nb^{mCherry}$-positive mono- and polyclonal cells was recorded in biological triplicate by flow cytometry (FACSMelody, BD Biosciences). mCherry fluorescence was obtained using the 561 nm laser line with a 613/18/LP605/10 filter set. Standard gates for doublet discrimination were applied to all samples. FACS data were evaluated with FlowJo 10.6.2 (BD), with the cell count normalized to the mode using the mCherry fluorescence area, or plotted in OriginPro 2020 (OriginLab), reporting the mean mCherry fluorescence area.

EGFP-binding assay

To obtain high amounts of $^{WT/Amb}Nb^{mCherry}$ containing lysates, 5×10^5 HeLa cells were seeded per 6-well (Greiner). Cells were cultured in DMEM medium containing 4.5 g L^{-1} glucose (Gibco), 10% (v/v) FCS tetracycline-negative (BIO&SELL), 2 μg ml^{-1} blasticidin (Gibco), and 100 μg ml^{-1} hygromycin (Invitrogen). On the following day, the medium was renewed and $^{Amb}Nb^{mCherry}$ cells transfected with the NPYRS/4xPylT encoding plasmid using Lipofectamine 2000 (Life Technologies). For increased comparability, $^{WT}Nb^{mCherry}$ cells were also transfected

with the NPYRS/4xPylT plasmid. 4–6 h after transfection, the medium was renewed and supplemented with 0.1 µg ml^{-1} tetracycline for induction. To evoke amber suppression, 0.25 mM ONBY (final) was added to the medium of ^{Amb}Nb^{mCherry}-expressing cells. After 24 h, cells were harvested using PBS (Sigma-Aldrich) and 0.05% trypsin–EDTA (Gibco), washed once with normal selection medium and once with PBS (Sigma-Aldrich). Cells were lysed in Pierce RIPA buffer (Thermo Fisher Scientific) containing 1% (v/v) Benzonase (Merck Millipore) and 1% (v/v) Protease-Inhibitor Mix HP (Serva). For lysis, cells of three individual 6-well plates were pooled and incubated in 300 µl lysis buffer for 1 h at RT with end-over-end rotation. Afterwards, samples were centrifuged at 21 000g for 30 min at 4 °C and the supernatant was collected. Protein concentration of the obtained lysates was determined by using the Pierce Detergent Compatible Bradford Assay Kit (Thermo Fisher Scientific). For the *in vitro* GFP-binding assay, lysates of three individual preparations were pooled. For ^{Amb}Nb^{mCherry} photo-activation, the lysate was exposed to UV light (three times for 3 min with 365 nm at 100 mW, ThorLabs DC 2200 M365L2 λ = 365 nm UV lamp with attached collimator SM2F32-A). Tyrosine deprotection was validated by mass spectrometry (ESI Fig. 18†). To capture $^{WT/Amb}$Nb^{mCherry} from the lysates, RFP-Trap Agarose (Chromotek) was used according to the manufacturer's protocol. In total, 0.5 mg ^{WT}Nb^{mCherry}- or 1.0 mg ^{Amb}Nb^{mCherry}-expressing cell lysate was mixed with 25 µl of agarose bead slurry. Samples were filled to 500 µl with dilution buffer (10 mM Tris/Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA) and EGFP (120 nM final) was added. After 1 h end-over-end rotation at 4 °C, beads were washed three times using washing buffer (10 mM Tris/Cl pH 7.5, 150 mM NaCl, 0.05% Tween 20, 0.5 mM EDTA). Beads were resuspended in 1 ml washing buffer, and EGFP fluorescence was recorded at $\lambda_{ex/em}$ 488/509 nm. As controls, lysates of untransfected monoclonal ^{Amb}Nb^{mCherry} cells cultured in normal selection medium were used. To prevent bead sedimentation during analysis, samples were gently resuspended and measured immediately in triplicate. To normalize the recorded fluorescence intensities, the ratios of captured ^{WT}Nb^{mCherry} and ^{Amb}N^{bmCherry} were determined by immuno-blotting against mCherry.

Immunoblotting

After quantification of the EGFP binding by fluorescence, beads were boiled at 95 °C for 10 min in 15 µl 5× SDS loading dye (0.02% (w/v) bromophenol blue, 30% (v/v) glycerol, 10% (w/v) SDS, 250 mM Tris–HCl, 250 mM DTT, pH 6.8). Samples were separated by 12% Tris–glycine SDS-PAGE. After electrophoresis, the gel was blotted semi-dry on nitrocellulose membrane. As transfer buffer, 25 mM Tris, 100 mM glycine, 0.1% (w/v) SDS and 20% (v/v) methanol was used. The membrane was blocked for 1 h in 5% (w/v) nonfat milk powder containing Tris-buffered saline with TWEEN 20 (TBS-T, pH 7.4). Blocking was followed by three consecutive washing steps with TBS-T. Afterwards, the membrane was incubated with monoclonal primary anti-mCherry antibody (Abcam, EPR20579) derived from rabbit 1 : 1000 in blocking buffer at 8 °C overnight. Unbound primary

antibody was removed in three washing steps with TBS-T. As the secondary antibody, anti-Rabbit IgG antibody (H + L) HRP conjugate produced in goat (Sigma-Aldrich) was used 1 : 10 000 in TBS-T incubated for 1 h. Three TBS-T washing steps followed the incubation with the secondary antibody. For chemilumi-nescent detection *via* the HRP, an ECL solution (Clarity Western ECL Substrate, Bio-Rad) was applied, and visualized by a Fusion FX imaging system (Vilber).

Mass spectrometry

ONBY and NPY, respectively, were analyzed using a Waters BioAccord system running UNIFY 1.9.4. Samples were separated on an Acquity BEH C18 column (1.7 µm, 2.1 mm × 50 mm). Identity was verified for NPY (M_{calc}: 360.0958 Da, M_{obs}: 360.0946 Da [−3.2 ppm]) and ONBY (M_{calc}: 316.1059 Da, M_{obs}: 316.1050 Da [−2.8 ppm]) using a cone voltage of 30 V and capillary voltage of 0.8 kV in positive polarity. Photo-cleavage was performed in neutral pH PBS at 100 µM sample concen-tration. Samples were irradiated with a ThorLabs DC 2200 (M365L2) λ = 365 nm UV lamp with an attached collimator (SM2F32-A) at 100 mW for different timeframes. For NPY and ONBY, the peak height at 214 nm was normalized to the uncleaved sample and fitted with an exponential decay function using OriginPro 2020. For uncaged tyrosine, the ESI-MS detector response was used instead.

Author contributions

E. J. generated the stable cell lines and performed the CLSM experiments. FACS experiments were carried out by C. W. and E. J. Samples for the *in vitro* GFP-binding assay were prepared by E. J. and analyzed by E. J. and C. W. C. W. performed the mass spectrometric analyses. E. J., C. W. and R. T. carried out the data analysis. Photocaged amino acids were synthesized by J. S. W., supervised by A. D. E. J. and R. T. wrote the manuscript with contributions from all authors. R. T. conceived and supervised the project.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

References

1 C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hammers, E. B. Songa, N. Bendahman and R. Hammers, *Nature*, 1993, **363**, 446–448.

2 A. S. Greenberg, D. Avila, M. Hughes, A. Hughes, E. C. McKinney and M. F. Flajnik, *Nature*, 1995, **374**, 168–173.

3 L. Riechmann and S. Muyldermans, *J. Immunol. Methods*, 1999, **231**, 25–38.

4 M. H. Kubala, O. Kovtun, K. Alexandrov and B. M. Collins, *Protein Sci.*, 2010, **19**, 2389–2401.

5 P. Kunz, K. Zinner, N. Mücke, T. Bartoschik, S. Muyldermans and J. D. Hoheisel, *Sci. Rep.*, 2018, **8**, 7934.

6 A. Klein, S. Hank, A. Raulf, E. F. Joest, F. Tissen, M. Heilemann, R. Wieneke and R. Tampé, *Chem. Sci.*, 2018, **9**, 7835–7842.

7 H. Götzke, M. Kilisch, M. Martínez-Carranza, S. Sograte-Idrissi, A. Rajavel, T. Schlichthaerle, N. Engels, R. Jungmann, P. Stenmark, F. Opazo and S. Frey, *Nat. Commun.*, 2019, **10**, 4403.

8 J. Helma, M. C. Cardoso, S. Muyldermans and H. Leonhardt, *J. Cell Biol.*, 2015, **209**, 633–644.

9 J. Ries, C. Kaplan, E. Platonova, H. Eghlidi and H. Ewers, *Nat. Methods*, 2012, **9**, 582–584.

10 D. Virant, B. Traenkle, J. Maier, P. D. Kaiser, M. Bodenhöfer, C. Schmees, I. Vojnovic, B. Pisak-Lukáts, U. Endesfelder and U. Rothbauer, *Nat. Commun.*, 2018, **9**, 930.

11 J. Huo, A. Le Bas, R. R. Ruza, H. M. E. Duyvesteyn, H. Mikolajek, T. Malinauskas, T. K. Tan, P. Rijal, M. Dumoux, P. N. Ward, J. Ren, D. Zhou, P. J. Harrison, M. Weckener, D. K. Clare, V. K. Vogirala, J. Radecke, L. Moynié, Y. Zhao, J. Gilbert-Jaramillo, M. L. Knight, J. A. Tree, K. R. Buttigieg, N. Coombes, M. J. Elmore, M. W. Carroll, L. Carrique, P. N. M. Shah, W. James, A. R. Townsend, D. I. Stuart, R. J. Owens and J. H. Naismith, *Nat. Struct. Mol. Biol.*, 2020, **27**, 846–854.

12 E. A. Della Pia and K. L. Martinez, *PLoS One*, 2015, **10**, e0124303.

13 A. Kirchhofer, J. Helma, K. Schmidthals, C. Frauer, S. Cui, A. Karcher, M. Pellis, S. Muyldermans, C. Casas-Delucchi, M. C. Cardoso, H. Leonhardt, K. Hopfner and U. Rothbauer, *Nat. Struct. Mol. Biol.*, 2010, **17**, 133–138.

14 U. Rothbauer, K. Zolghadr, S. Tillib, D. Nowak, L. Schermelleh, A. Gahl, N. Backmann, K. Conrath, S. Muyldermans, M. C. Cardoso and H. Leonhardt, *Nat. Methods*, 2006, **3**, 887–889.

15 A. L. J. Marschall, S. Dübel and T. Böldicke, *mAbs*, 2015, **7**, 1010–1035.

16 F. Schneider, T. Sych, C. Eggeling and E. Sezgin, *iScience*, 2021, **24**, 101891.

17 K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764–4806.

18 N. Wu, A. Deiters, T. A. Cropp, D. King and P. G. Schultz, *J. Am. Chem. Soc.*, 2004, **126**, 14306–14307.

19 D. Gründemann, S. Harlfinger, S. Golz, A. Geerts, A. Lazar, R. Berkels, N. Jung, A. Rubbert and E. Schömig, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 5256–5261.

20 A. Deiters, D. Groff, Y. Ryu, J. Xie and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2006, **45**, 2728–2731.

21 J. Luo, J. Torres-Kolbus, J. Liu and A. Deiters, *ChemBioChem*, 2017, **18**, 1442–1447.

22 K. Gatterdam, E. F. Joest, M. S. Dietz, M. Heilemann and R. Tampé, *Angew. Chem., Int. Ed.*, 2018, **57**, 5620–5625.

23 T. Kanda, K. F. Sullivan and G. M. Wahl, *Curr. Biol.*, 1998, **8**, 377–385.

24 J. Icha, M. Weber, J. C. Waters and C. Norden, *BioEssays*, 2017, **39**, 1700003.

25 B. Jedlitzke, Z. Yilmaz, W. Dörner and H. D. Mootz, *Angew. Chem., Int. Ed.*, 2020, **59**, 1506–1510.

26 T. A. Redchuk, M. M. Karasev, P. V. Verkhusha, S. K. Donnelly, M. Hülsemann, J. Virtanen, H. M. Moore, M. K. Vartiainen, L. Hodgson and V. V. Verkhusha, *Nat. Commun.*, 2020, **11**, 605.

27 D. Yu, H. Lee, J. Hong, H. Jung, Y. Jo, B. H. Oh, B. O. Park and W. D. Heo, *Nat. Methods*, 2019, **16**, 1095–1100.

28 A. A. Gil, C. Carrasco-López, L. Zhu, E. M. Zhao, P. T. Ravindran, M. Z. Wilson, A. G. Goglia, J. L. Avalos and J. E. Toettcher, *Nat. Commun.*, 2020, **11**, 4044.

29 E. Caussinus, O. Kanca and M. Affolter, *Nat. Struct. Mol. Biol.*, 2012, **19**, 117–121.

30 A. Hermann, J. F. Liewald and A. Gottschalk, *Curr. Biol.*, 2015, **25**, R749–R750.

31 C. Renicke, D. Schuster, S. Usherenko, L. Essen and C. Taxis, *Chem. Biol.*, 2013, **20**, 619–626.

32 J. Luo, R. Uprety, Y. Naro, C. Chou, D. P. Nguyen, J. W. Chin and A. Deiters, *J. Am. Chem. Soc.*, 2014, **136**, 15551–15558.

33 C. A. Hammer, K. Falahati, A. Jakob, R. Klimek, I. Burghardt, A. Heckel and J. Wachtveitl, *J. Phys. Chem. Lett.*, 2018, **9**, 1448–1453.

34 R. R. Nani, A. P. Gorka, T. Nagaya, T. Yamamoto, J. Ivanic, H. Kobayashi and M. J. Schnermann, *ACS Cent. Sci.*, 2017, **3**, 329–337.

35 C. Aonbangkhen, H. Zhang, D. Z. Wu, M. A. Lampson and D. M. Chenoweth, *J. Am. Chem. Soc.*, 2018, **140**, 11926–11930.

36 T. A. Shell, J. R. Shell, Z. L. Rodgers and D. S. Lawrence, *Angew. Chem., Int. Ed.*, 2014, **53**, 875–878.

37 J. Fueller, K. Herbst, M. Meurer, K. Gubicza, B. Kurtulmus, J. D. Knopf, D. Kirrmaier, B. C. Buchmuller, G. Pereira, M. K. Lemberg and M. Knop, *J. Cell Biol.*, 2020, **219**, e201910210.

38 S. J. Elsässer, R. J. Ernst, O. S. Walker and J. W. Chin, *Nat. Methods*, 2016, **13**, 158–164.

39 M. D. Paul and K. Hristova, *Cytokine Growth Factor Rev.*, 2019, **49**, 23–31.

40 E. Arbely, J. Torres-Kolbus, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2012, **134**, 11912–11915.

41 C. L. Neben, M. Lo, N. Jura and O. D. Klein, *Dev. Biol.*, 2019, **447**, 71–89.

42 J. Schlessinger, *Cold Spring Harbor Perspect. Biol.*, 2014, **6**, a008912.

43 M. A. Lemmon and J. Schlessinger, *Cell*, 2010, **141**, 1117–1134.

44 S. R. Hubbard and J. H. Till, *Annu. Rev. Biochem.*, 2000, **69**, 373–398.

45 H. Farrants, V. A. Gutzeit, A. Acosta-Ruiz, D. Trauner, K. Johnsson, J. Levitz and J. Broichhagen, *ACS Chem. Biol.*, 2018, **13**, 2682–2688.

46 T. Bridge, S. A. Shaikh, P. Thomas, J. Botta, P. J. McCormick and A. Sachdeva, *Angew. Chem., Int. Ed.*, 2019, **58**, 17986–17993.

47 A. Kollmannsperger, A. Sharei, A. Raulf, M. Heilemann, R. Langer, K. F. Jensen, R. Wieneke and R. Tampé, *Nat. Commun.*, 2016, **7**, 10372.

48 C. C. Uphoff and H. G. Drexler, *Curr. Protoc. Mol. Biol.*, 2014, **106**, 28.24.1–28.24.14.

49 C. Uttamapinant, J. D. Howe, K. Lang, V. Beránek, L. Davis, M. Mahesh, N. P. Barry and J. W. Chin, *J. Am. Chem. Soc.*, 2015, **137**, 4602–4605.

50 M. Cigler, T. A. Nguyen and K. Lang, in Oxidative Folding of Proteins: Basic Principles, *Cellular Regulation and Engineering*, The Royal Society of Chemistry, 2018, pp. 399–420.

51 J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona, *Nat. Methods*, 2012, **9**, 676–682.

Supporting Information

Light-guided intrabodies for on-demand *in-situ* target recognition in human cells

Eike F. Joest[1], Christian Winter[1], Joshua S. Wesalo[2], Alexander Deiters[2], Robert Tampé[1,*]

[1] Institute of Biochemistry, Biocenter, Goethe University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany

[2] Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, Pennsylvania 15260, USA

Table of Contents for Supporting Information

Supplementary Fig. 1 | Flow cytometry gating strategy. For generation of monoclonal cell lines and amber-construct expression analysis. 5000 events were recorded per sample. Shares of parent populations are provided for each gate. (A) representative FACS plots for selection of monoclonal cells. Single cells with high $^{Amb}Nb^{mcherry}$ expression were sorted into distinct wells after doublet discrimination. (B) representative FACS plots for amber-construct expression analysis. mCherry excitation laser intensities were different.
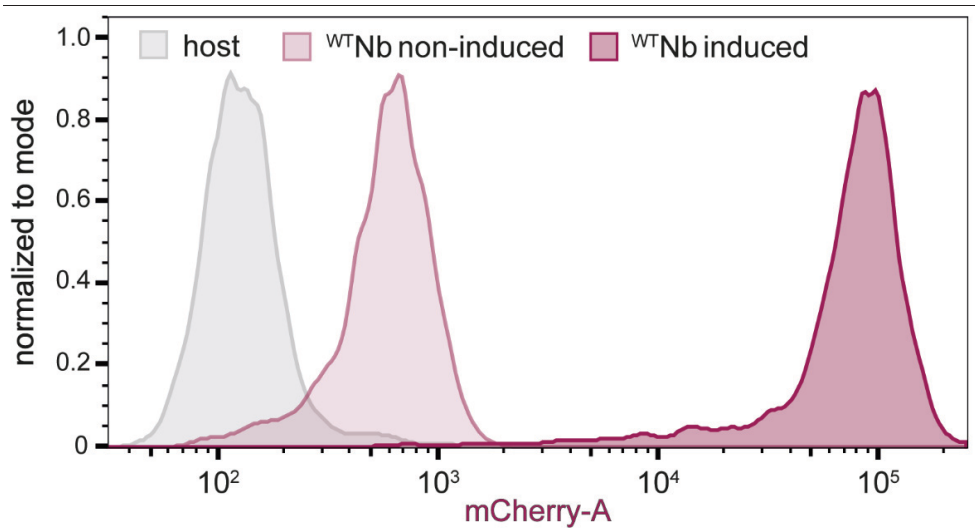
Supplementary Fig. 2 | Expression of wild-type intrabody in stable cells monitored by flow cytometry. Tetracycline-induced expression of $^{WT}Nb^{mCherry}$ evokes a strong mCherry fluorescence. Cell counts were normalized to mode (n = 3).
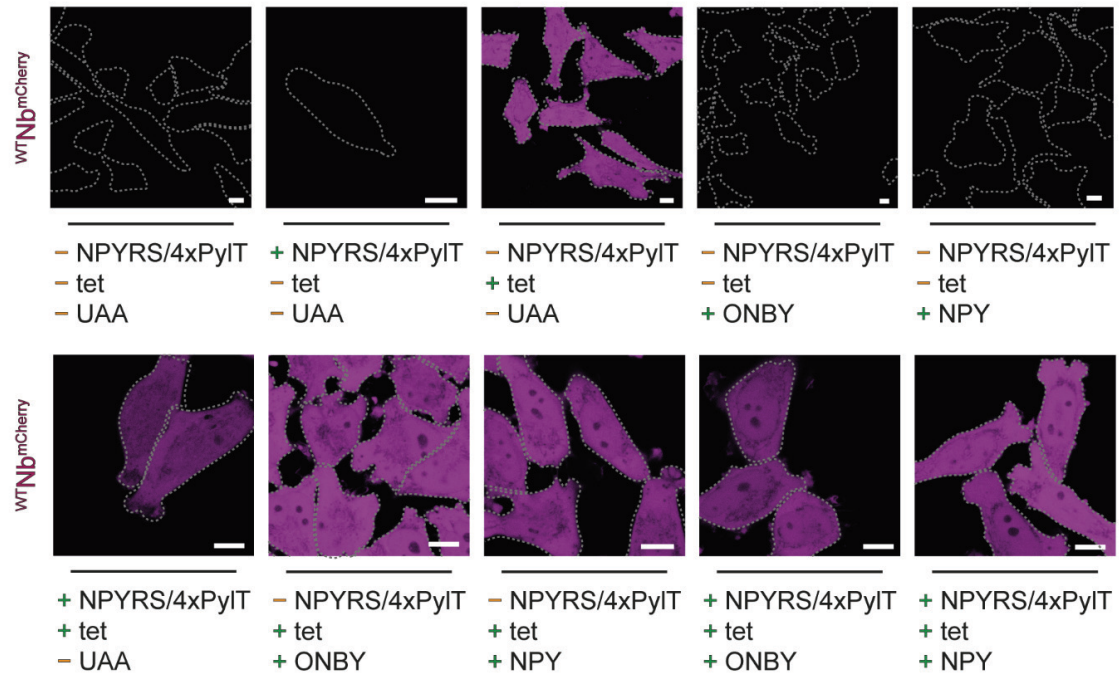
Supplementary Fig. 3 | Expression of wild-type intrabody ^{WT}Nb^{mCherry} in stable cells. Tetracycline-induced expression of intrabodies monitored by mCherry. Expression was not affected by the presence of amber suppression components. Live-cell CLSM imaging. Dashed lines indicate cellular borders. Scale bar, 10 µm.
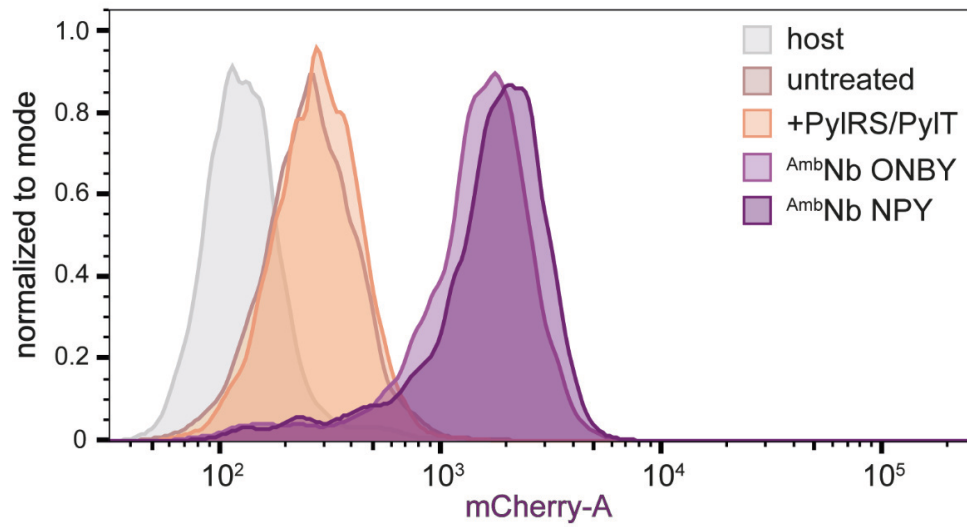
Supplementary Fig. 4 | Expression of amber-construct ^{Amb}**Nb**^{mCherry} **monitored by flow cytometry**. The early amber stop codon mutation Tyr37TAG abolished expression without optimized amber-suppression conditions. Suppression of the amber stop codon resulted in incorporation of photocaged amino acids and hence in intrabody expression. Expression was observed by downstream encoded mCherry. ^{Amb}Nb ONBY/NPY show amber suppression conditions with respective photocaged tyrosine. Counts of stable cells were normalized to mode (n = 3).
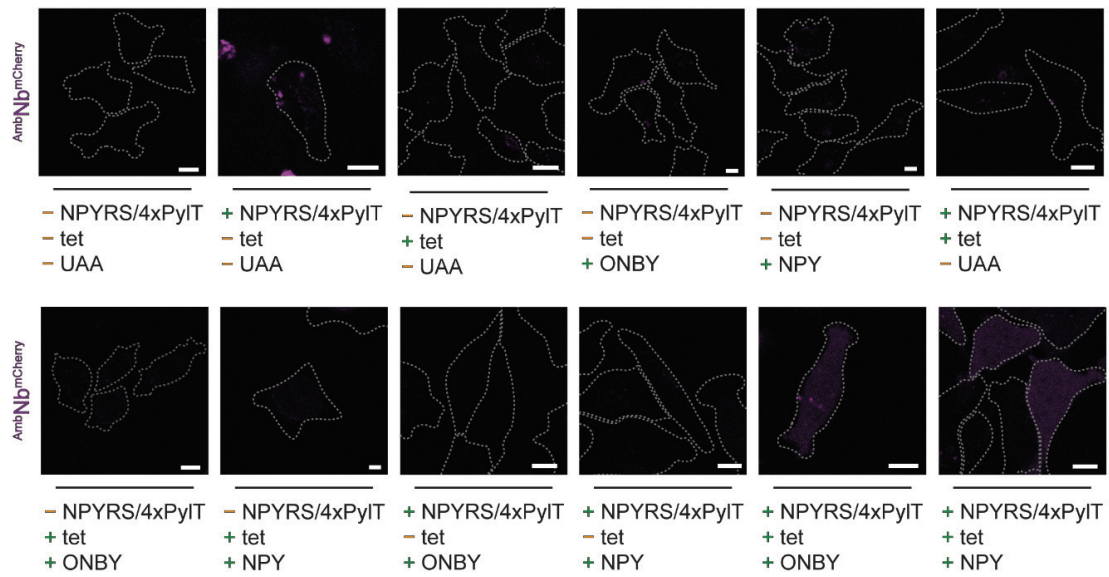
Supplementary Fig. 5 | Expression of amber-construct ^{Amb}Nb^{mCherry} by CLSM. The early amber stop codon (Tyr37TAG) abolished expression in the absence of the tRNA/NYPRS pair and/or the unnatural amino acid (UAA). Suppression of the amber stop codon resulted in incorporation of photocaged amino acids and hence in intrabody expression. Expression was observed by downstream encoded mCherry. Live-cell imaging of stable cell line. Dashed lines indicate cell borders. Scale bar, 10 µm.
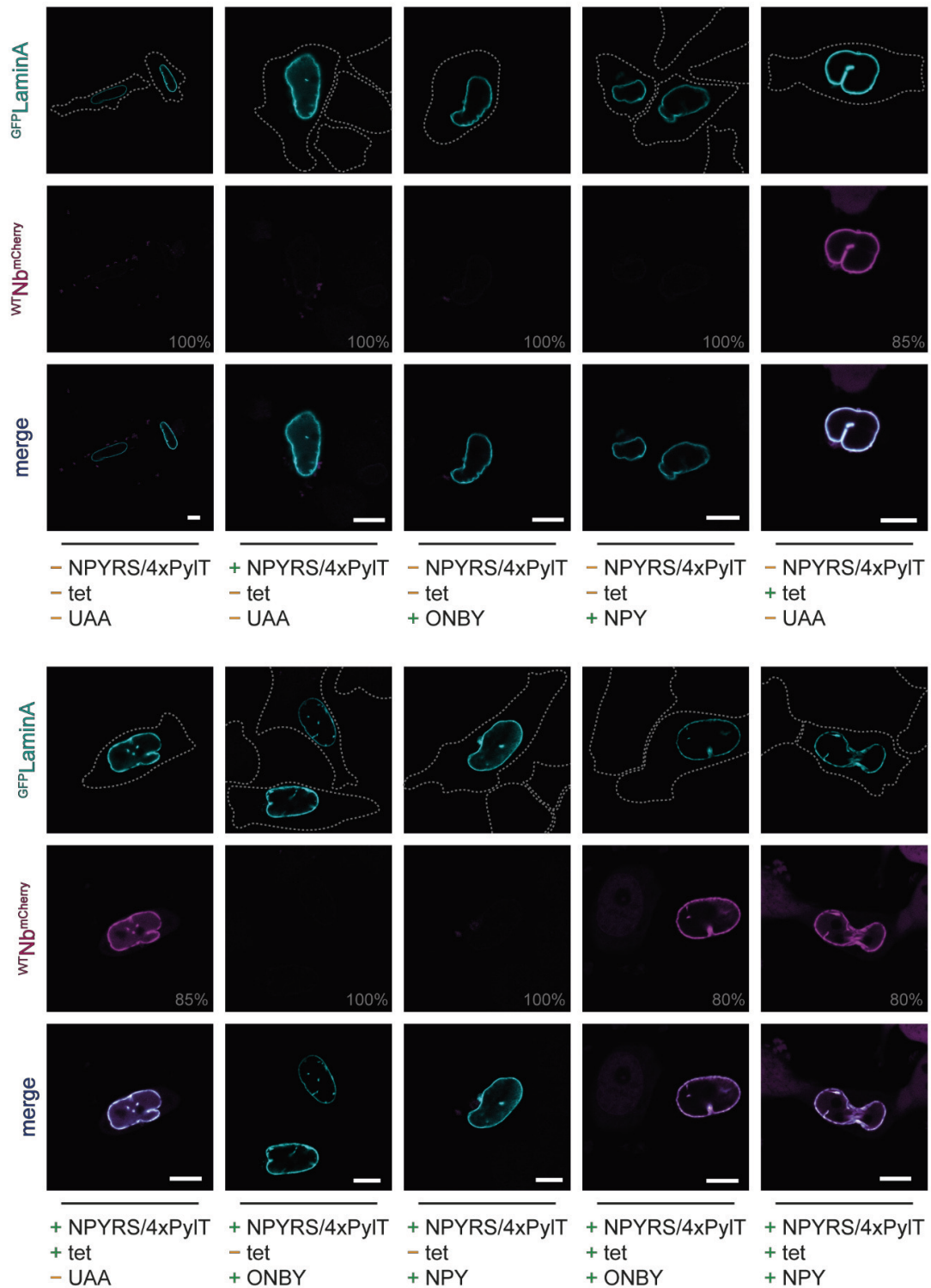
Supplementary Fig. 6 | Constitutively active binding of wild-type intrabody $^{WT}Nb^{mCherry}$ to $^{mEGFP}LaminA$. Colocalization visualized by intrabody $^{WT}Nb^{mCherry}$ binding of target located at the nuclear envelope. Tetracycline-induced expression of a constitutively active intrabody resulted in binding. Binding was unaffected by the presence of amber suppression components. CLSM live cell imaging of stable cell line and relative gain indicated in grey. Dashed lines indicate cellular borders. Scale bar, 10 μm.

7
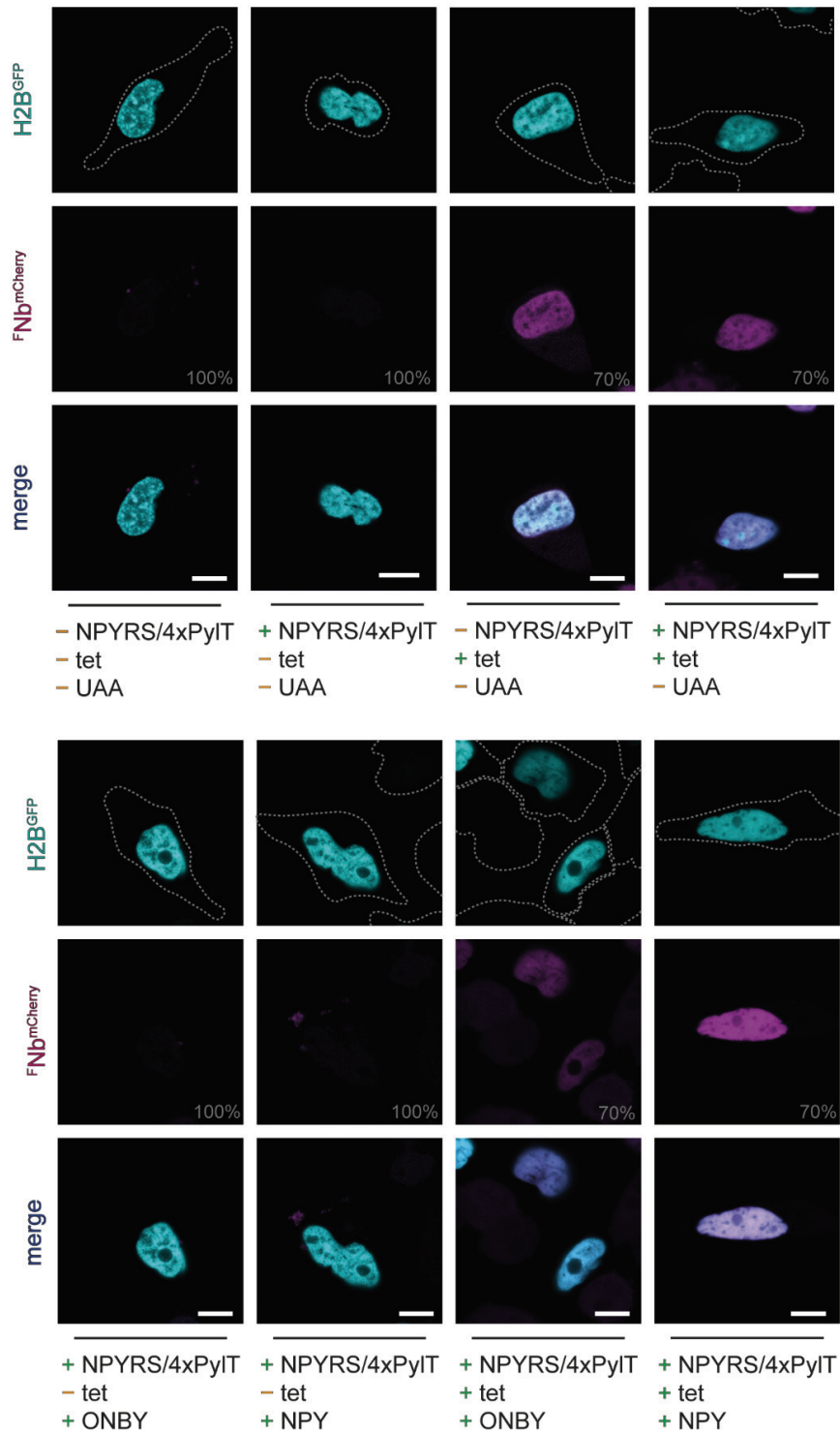
Supplementary Fig. 7 | Constitutively active binding of wild-type ^WTNb^mCherry to histone H2B^EGFP. Colocalization visualized by intrabody binding of target located inside the nucleus. Tetracycline-induced expression of a constitutively active intrabody resulted in binding. Binding was unaffected by the presence of amber suppression components. CLSM live cell imaging of stable cell line and relative gain is indicated in grey. Dashed lines indicate cellular borders. Scale bar, 10 µm.
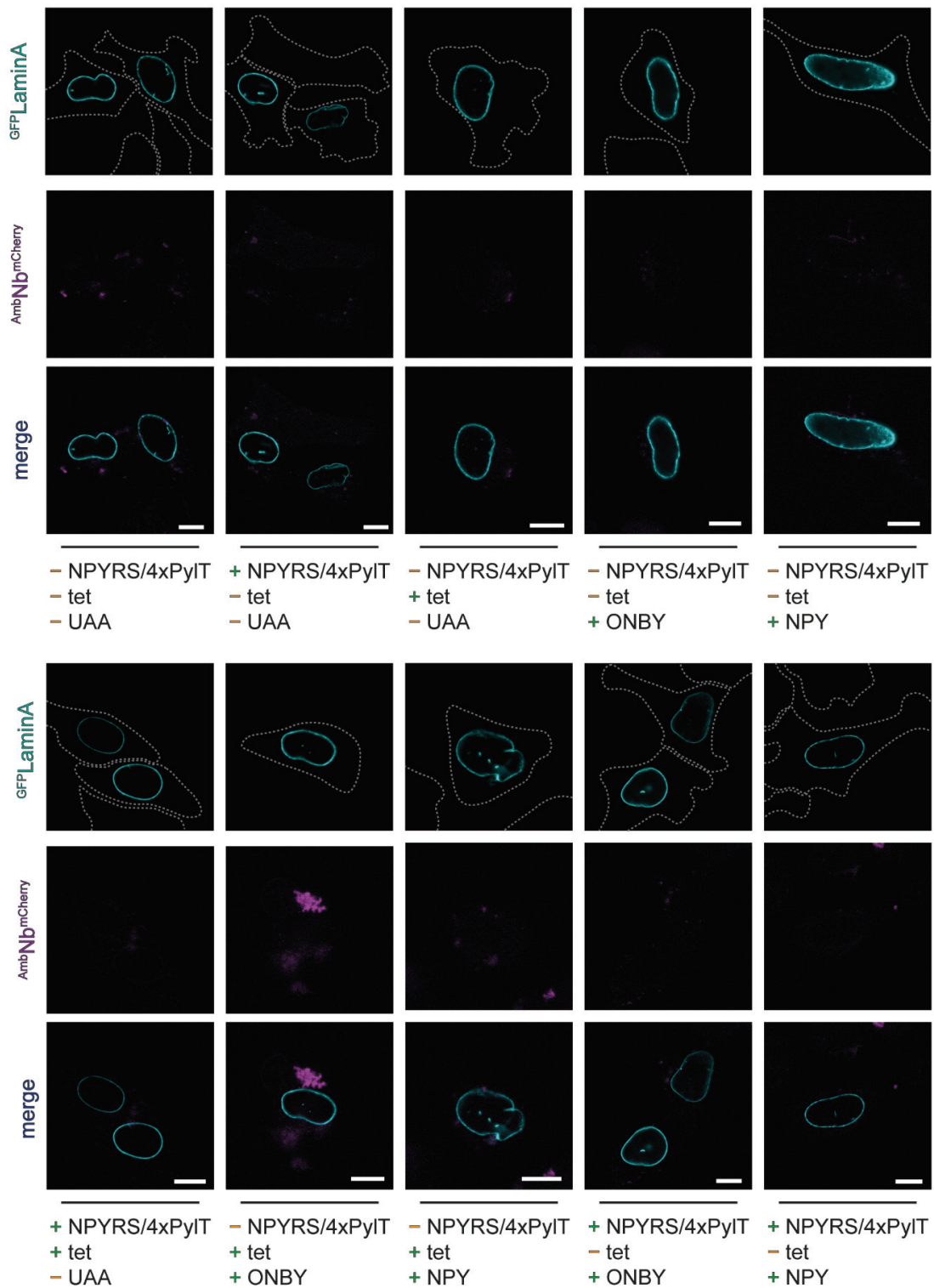
Supplementary Fig. 8 | Abolished translation of Amb**Nb**mCherry **in presence of** mEGFP**LaminA.** The amber stop codon (Tyr37TAG) abolished expression in the absence of conditions for complete amber suppression. CLSM live cell imaging of stable cell line and relative gain are indicated in grey. Dashed lines indicate cellular borders. Scale bar, 10 µm.
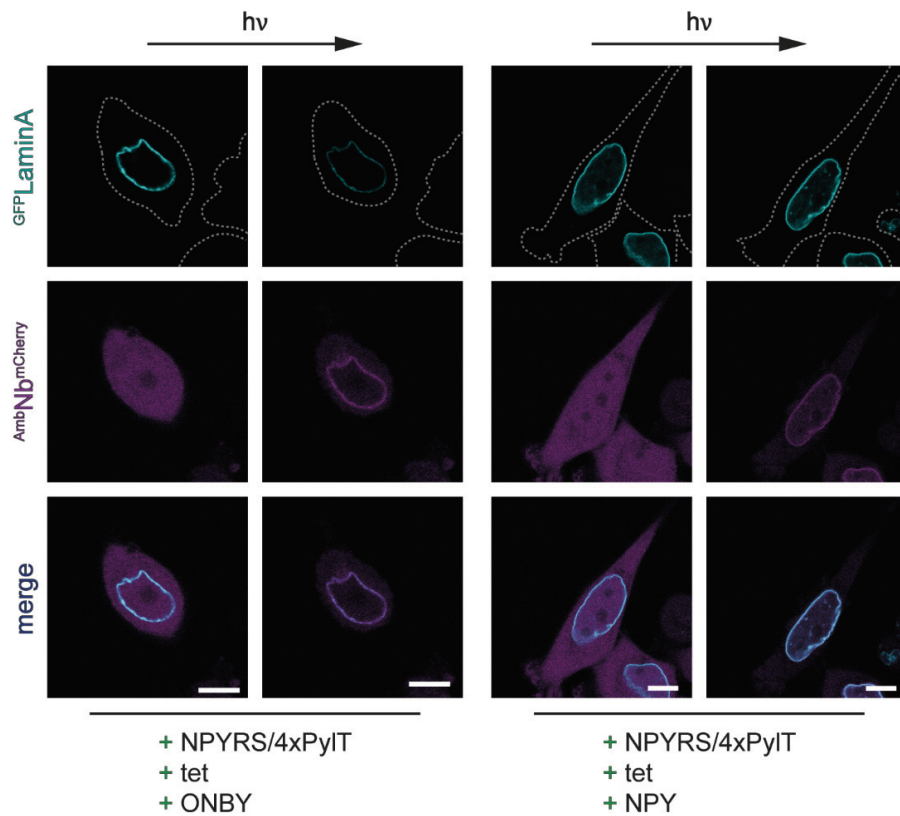
Supplementary Fig. 9 | Photo-activated binding of ^{Amb}Nb^{mCherry} **to** ^{mEGFP}LaminA.
Optimized amber-suppression conditions resulted in incorporation of photocaged amino acids. Expression of the amber-suppressed intrabody was monitored by downstream encoded mCherry. The photocaged tyrosines within the epitope-binding site prevented binding. After photo-activation, ^{Amb}Nb^{mCherry} binding of target was visualized by colocalization. CLSM live cell imaging of stable cell line. Dashed lines indicate cellular borders. Scale bar, 10 μm. In addition to Supplementary Fig. 7.
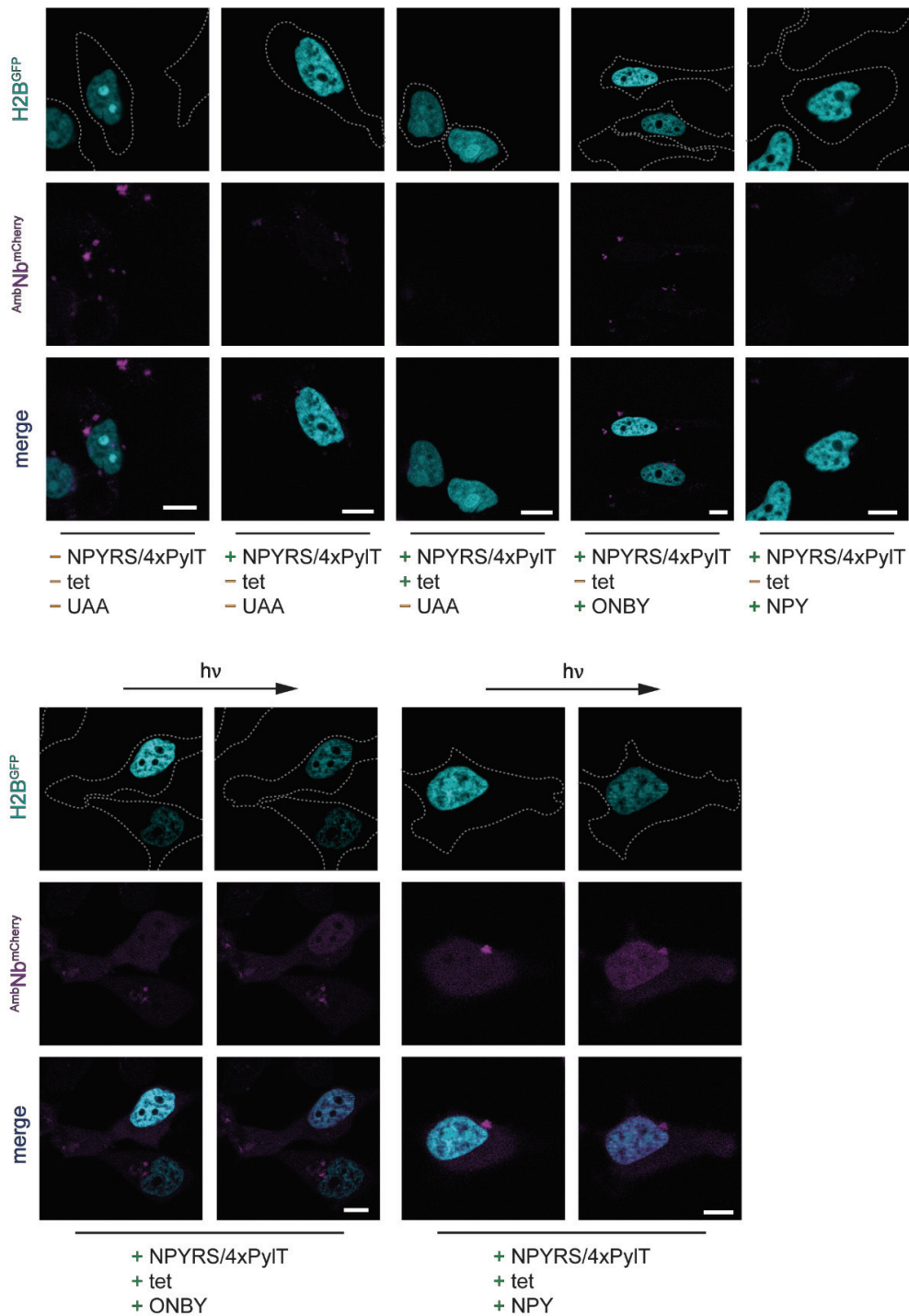
Supplementary Fig. 10 | Photo-induced binding of ᴬᵐᵇNbᵐᶜʰᵉʳʳʸ to histone H2BᴱᴳᶠP. The early amber stop codon (Tyr37*TAG*) abolished expression without complete amber-suppression conditions. Optimized amber-suppression conditions resulted in incorporation of photocaged amino acids. Expression of the amber-suppressed intrabody was monitored by downstream-encoded mCherry. The photocaged tyrosines within the epitope-binding site prevented binding. After photo-activation, instantaneous ᴬᵐᵇNbᵐᶜʰᵉʳʳʸ binding of target was visualized by colocalisation. CLSM live cell imaging of stable cell line. Dashed lines indicate cellular borders. Scale bar, 10 µm.
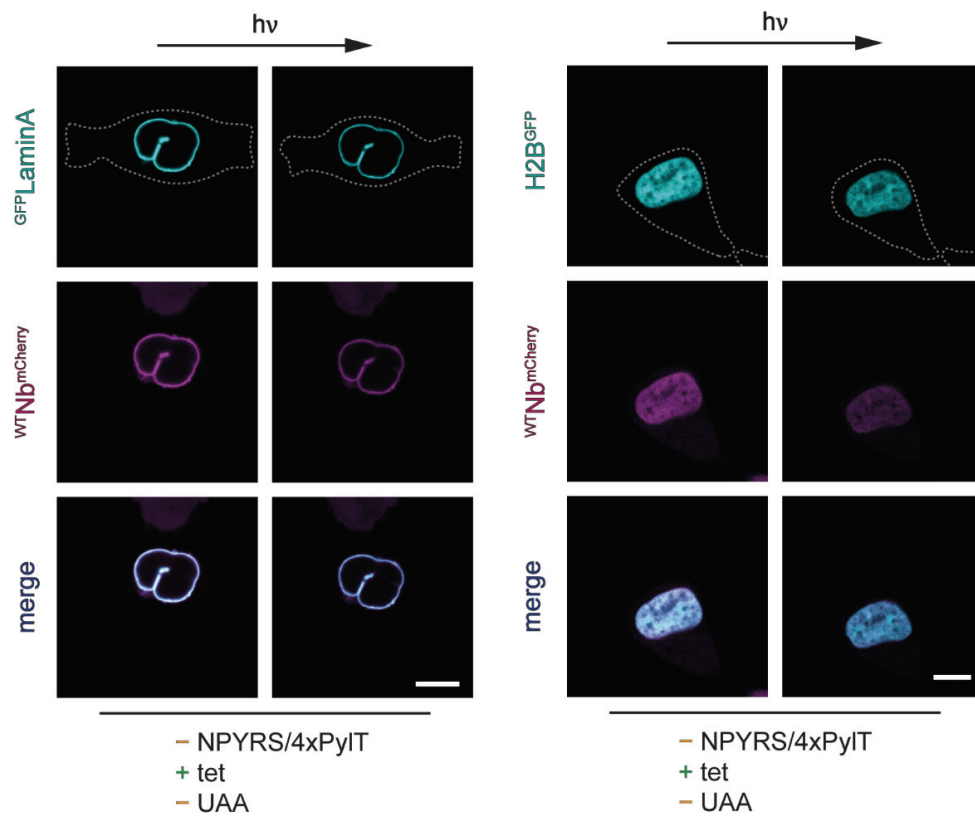
Supplementary Fig. 11 | Constitutively active binding of wild-type intrabody ^{WT}Nb^{mCherry} to target proteins is unaffected by light exposure. After tetracycline-induced expression, the constitutively active intrabody showed target binding. Colocalization remained unaffected after light exposure (compared to Supplementary Fig. 8 and 9). Live-cell CLSM imaging of stable cell line. Dashed lines indicate cell border. Scale bar, 10 µm.
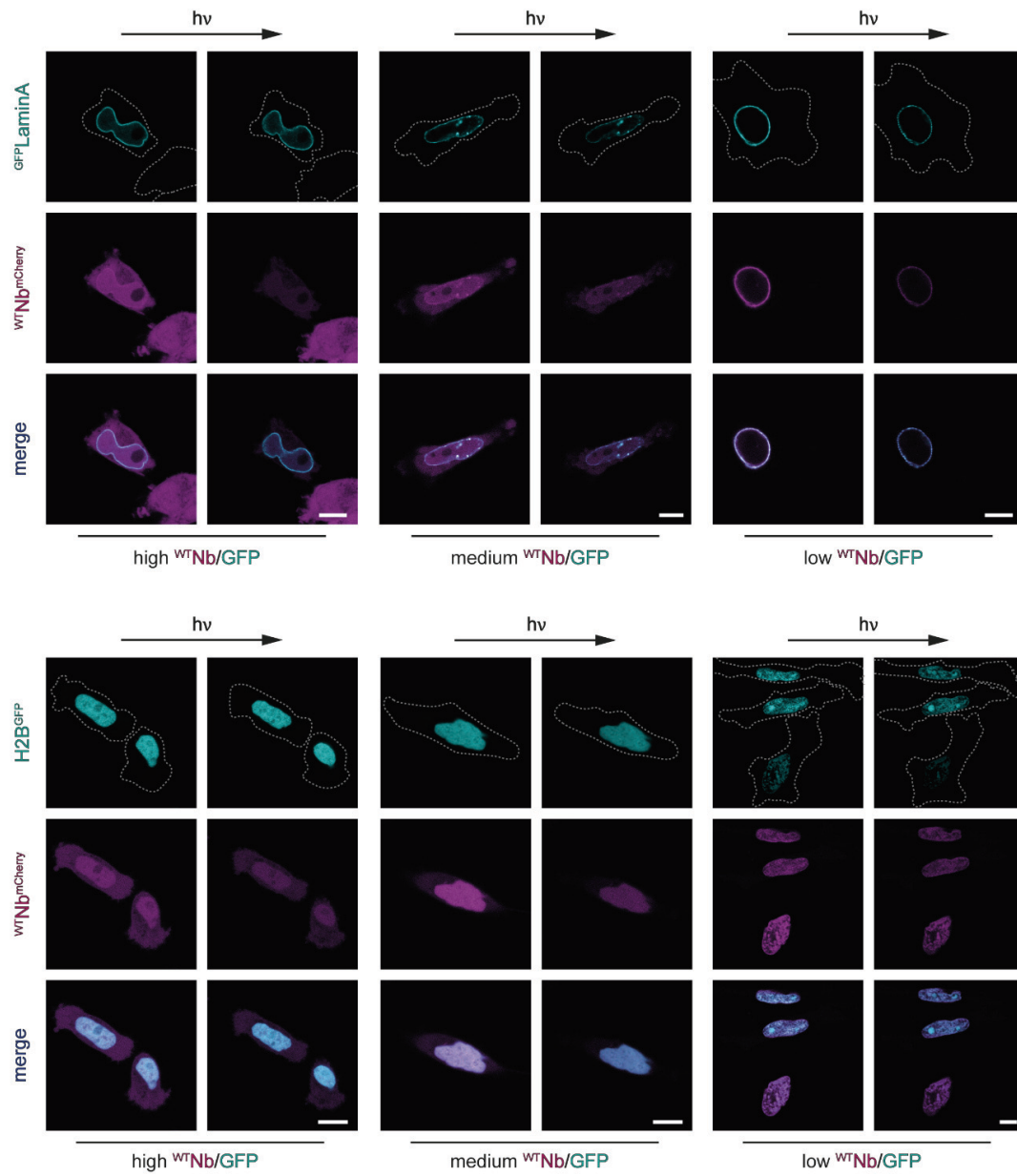
Supplementary Fig. 12 | Unaffected binding of ^WTNb^mCherry after illumination. After tetracycline-induced expression, the intrabody showed target binding. Colocalization remained unaffected after light exposure (compared to Supplementary Fig. 7 and 8). Live-cell CLSM imaging of stable cell line. Dashed lines indicate cell border. Scale bar, 10 µm.
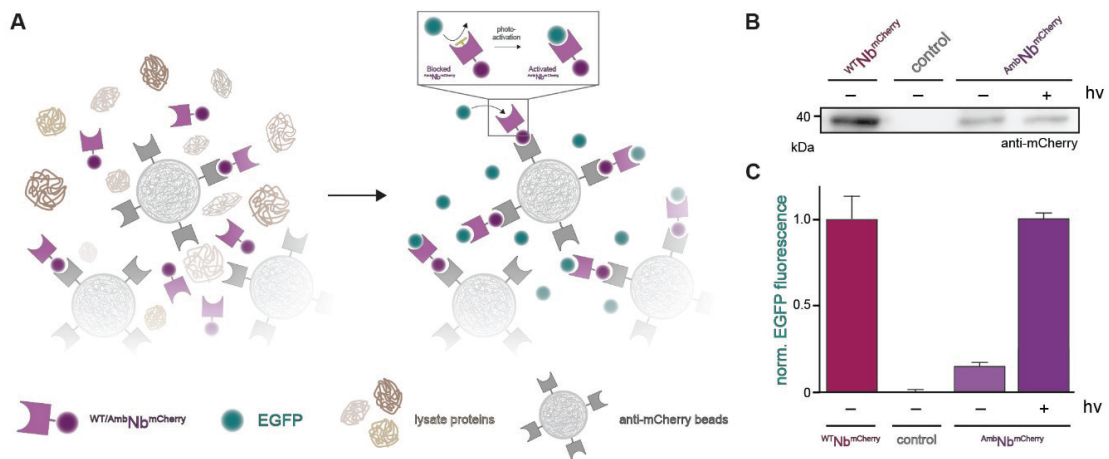
Supplementary Fig. 13 | Photo-activated nanobody binding in human cell lysates. (A) Scheme of the GFP-binding analysis in cell lysates. ^{WT}Nb^{mCherry} or ^{Amb}Nb^{mCherry} were captured in the cell lysate using anti-mCherry affinity beads. Beads loaded with ^{WT}Nb^{mCherry} or ^{Amb}Nb^{mCherry} were examined for EGFP binding. After washing, the amount of EGFP coimmunoprecipitated on the beads was quantified by fluorescence. (B) The amount of catpured ^{WT}Nb^{mCherry} or ^{Amb}Nb^{mCherry} was revealed by immunoblotting. Beads were incubated with 0.5 mg of ^{WT}Nb^{mCherry} containing lysate or 1 mg of ^{Amb}Nb^{mCherry} containing lysate and 120 nM EGFP. Lysates were pooled from three individual cell preparations. For ^{Amb}Nb^{mCherry} amber suppression, ONBY was used. (C) EGFP binding by ^{WT}Nb^{mCherry} or ^{Amb}Nb^{mCherry}. Exposing the cell lysate to UV light restored the epitope binding of ^{Amb}Nb^{mCherry}. In contrast, only neglectable amounts of EGFP were recorded using beads with unexposed ^{Amb}Nb^{mCherry} lysates. EGFP fluorescence was normalized to the amount of captured ^{WT}Nb^{mCherry} as illustrated in (B). Means ± SD (n = 3, technical replicates) are displayed.
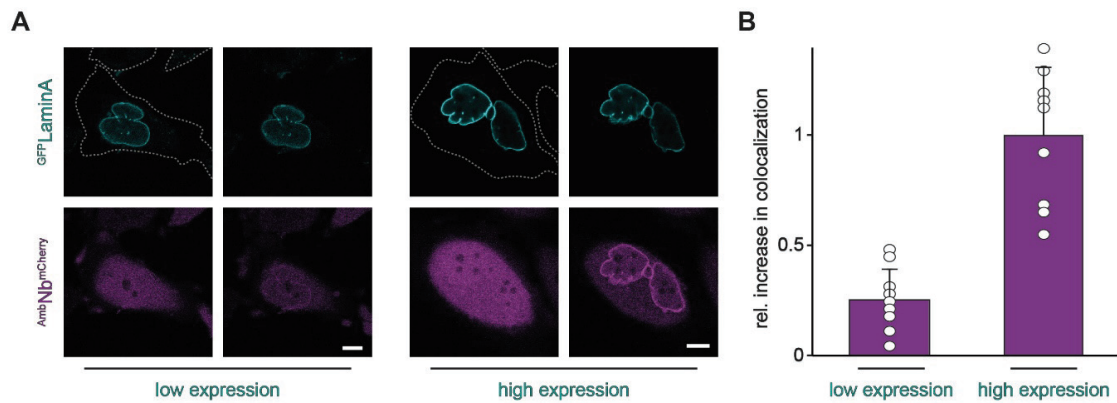
Supplementary Fig. 14 | Statistical analysis of photo-activated intrabody binding. (A) Intrabody binding after illumination inside cells containing low (GFP gain 75%) or high (GFP gain 100%) amounts of target protein. Exemplary cells of two separate groups used in (B). Scale bar, 10 μm. (B) Statistical analysis of photo-activated binding inside two groups of cells with different amount of target protein. Mean increase in colocalization after illumination observed within individual cells (n = 10). For quantification, the absolute increase in colocalization denoted by the Pearson's coefficient was determined and normalized using the high expression group.
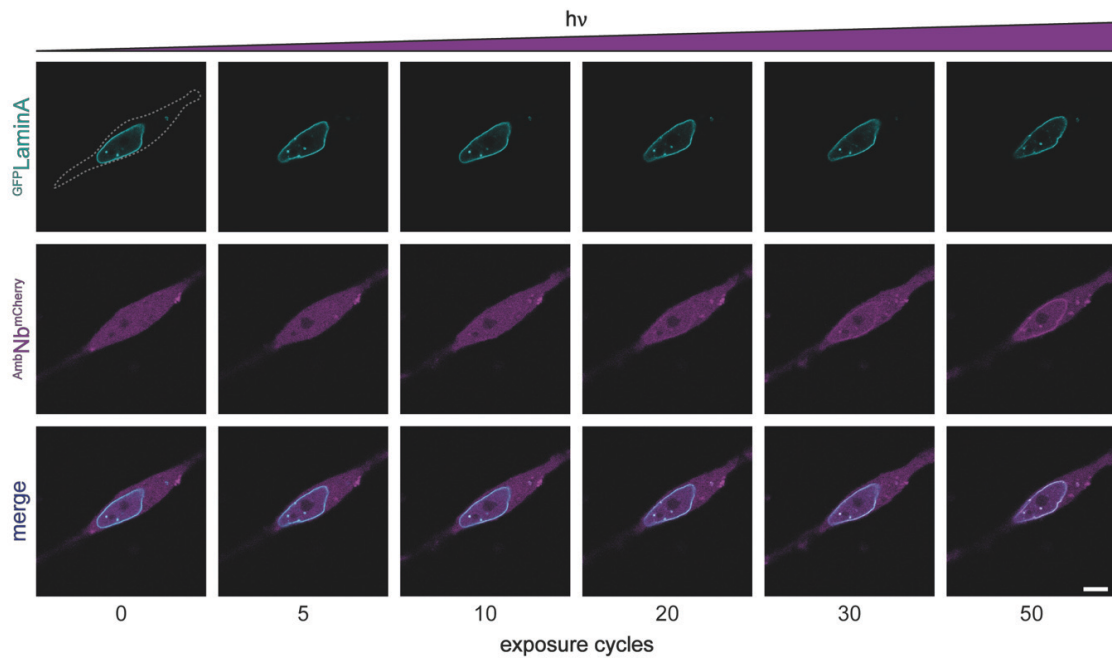
Supplementary Fig. 15 | Dose-dependent ONBY photo-cleavage. After amber suppression using ONBY, no target binding of $^{Amb}Nb^{mCherry}$ was observed. Progressive increase of light exposure allowed tight control of $^{Amb}Nb^{mCherry}$ activation. Intrabody binding was visualized by a correlative increase in target colocalization. Maximum exposure time for photo-activation corresponds to a bleaching function with 250 iterations, 50 cycles, and a 405-nm diode laser (4.5 mW/μm^2) for the whole region of interest. Live-cell CLSM imaging of stable cell line. Dashed lines indicate cell borders. Scale bar, 10 μm.
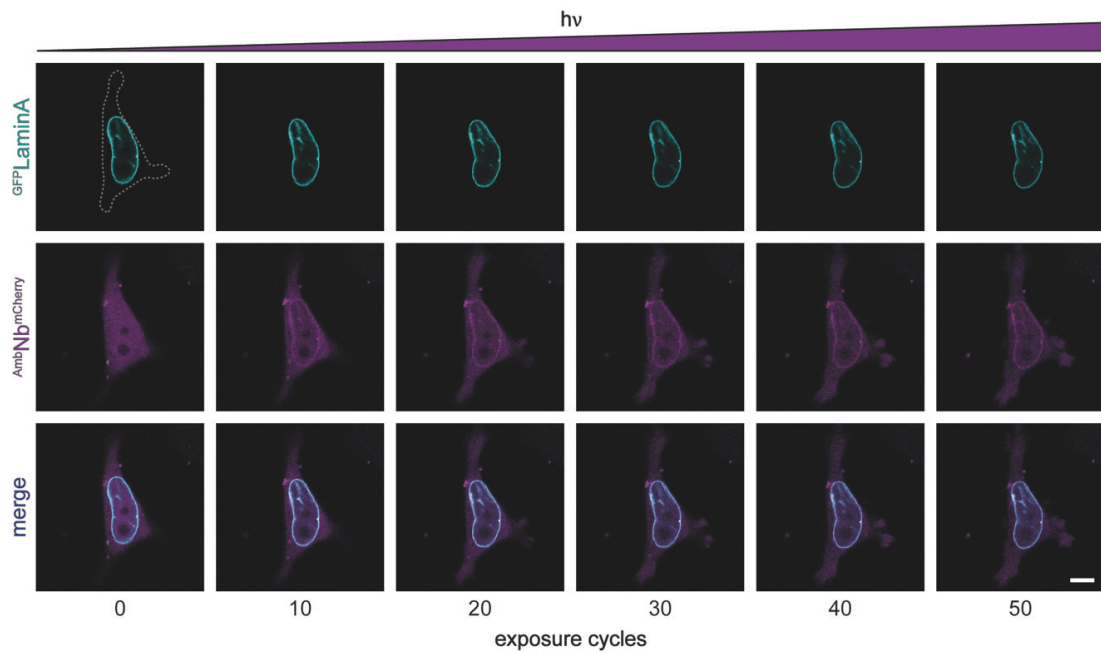
Supplementary Fig. 16 | Dose-dependent NPY photo-cleavage. After amber suppression using NPY, no target preterm binding of $^{Amb}Nb^{mcherry}$ was observed. Progressive increase of light exposure allowed tight control of $^{Amb}Nb^{mCherry}$ activation. Intrabody binding was visualized by a correlative increase in target colocalization. Maximum exposure time for photo-activation corresponds to a bleaching function with 250 iterations, 50 cycles and a 405-nm diode laser (4.5 mW/µm^2) for the whole region of interest. Live-cell CLSM imaging of stable cell line. Dashed lines indicate cell borders. Scale bar, 10 µm.
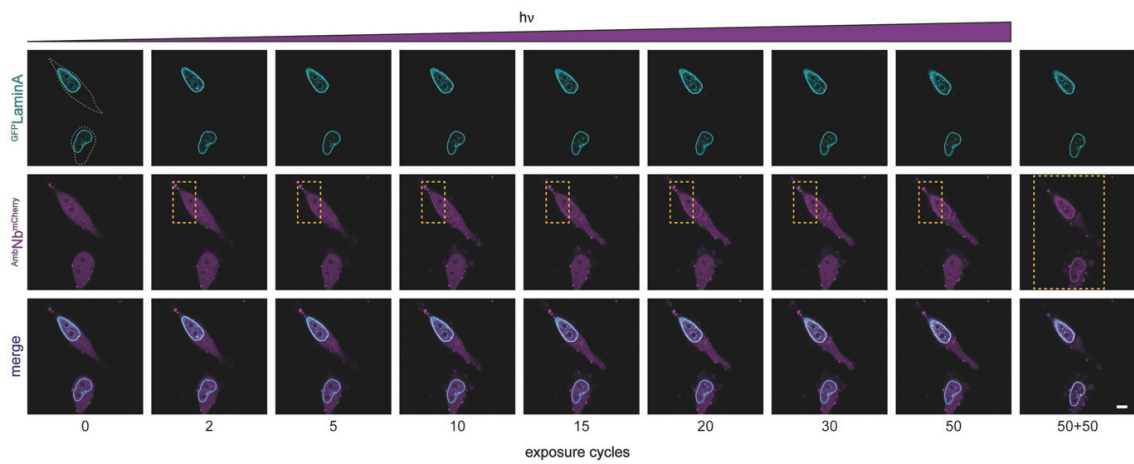
Supplementary Fig. 17 | Fine-tuned intrabody photo-activation by local energy dosage. After amber suppression using NPY, no target-preterm binding of $^{Amb}Nb^{mcherry}$ was observed. A small section of the cell was illuminated with increasing exposure time. Intrabody binding was visualized by a correlative increase in target colocalization. Maximum exposure time for photo-activation corresponds to a bleaching function with 250 iterations, 50 cycles and a 405-nm diode (4.5 mW/µm^2) for the whole region of interest. Live-cell CLSM imaging of stable cell line. Dashed lines indicate cell borders. Scale bar, 10 µm.
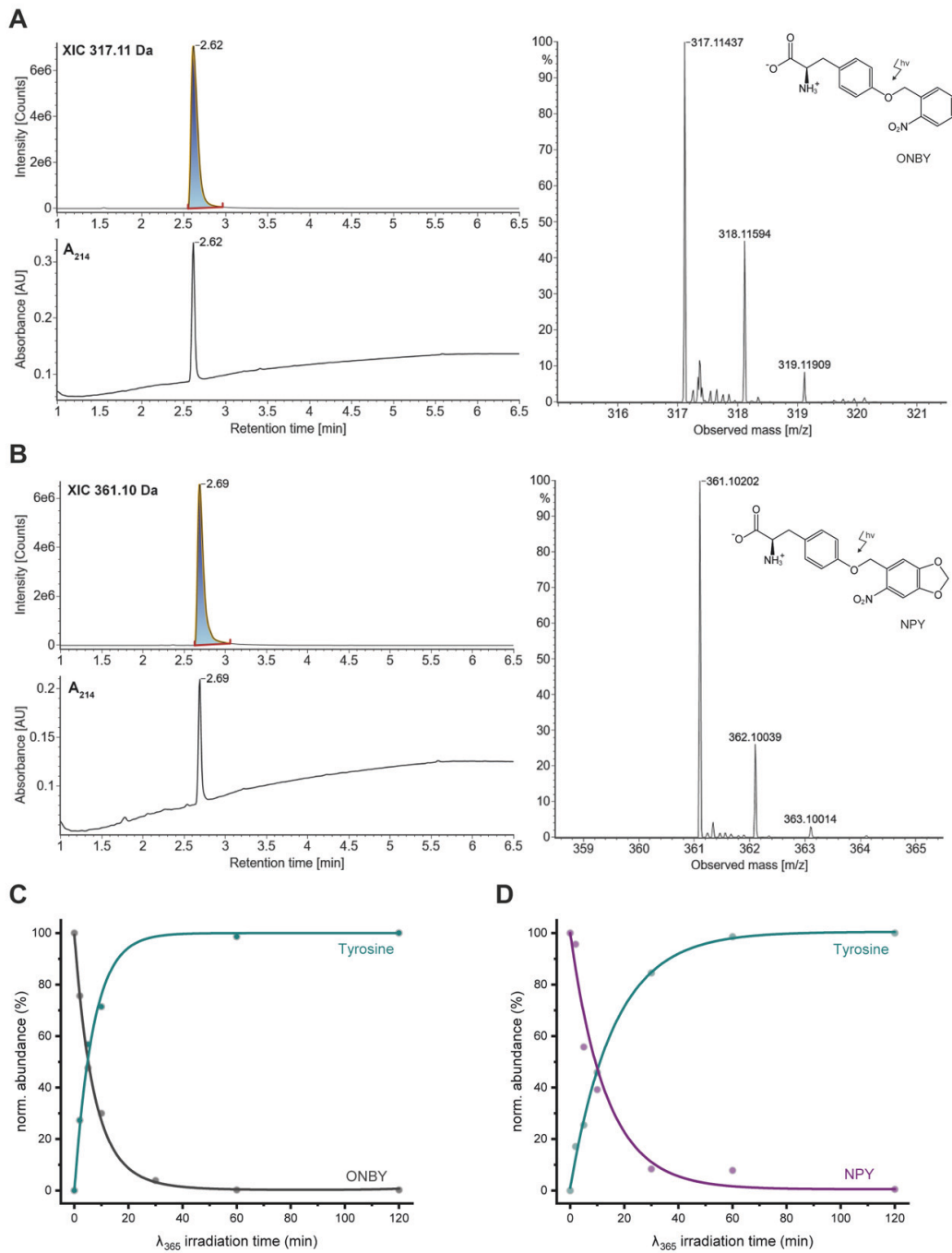
Supplementary Fig. 18 | High-resolution mass analysis of photocaged tyrosines. (A, B) The photocaged tyrosines ONBY (A) or NPY (B) were analyzed by LC-MS. Integrity of (A) ONBY (M_{calc}: 316.1059 Da; M_{obs}: 316.1050 Da [-2.8 ppm]) and (B) NPY (M_{calc}: 360.0958 Da; M_{obs}: 360.0946 Da [-3.2 ppm]) was verified by a cone voltage of 30V, capillary voltage of 0.8 kV in positive polarity. (C, D) Photo cleavage performed in neutral pH PBS using 100 µM of either ONBY (C) or NPY (D). Samples were irradiated with 365 nm (100 mW) for different periods of time. For ONBY and NPY, the peak height at 214 nm was normalized to the uncleaved sample and fitted with a monoexponential decay function. Half-life times were estimated to be 5.0 s for ONBY and 9.1 s for NPY. For uncaged tyrosine, the ESI-MS detector response was used instead.

Efficient Amber Suppression *via* Ribosomal Skipping for *In Situ* Synthesis of Photoconditional Nanobodies

Eike F. Joest, Christian Winter, Joshua S. Wesalo, Alexander Deiters, and Robert Tampé*

Cite This: *ACS Synth. Biol.* 2022, 11, 1466−1476

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Genetic code expansion is a versatile method for *in situ* synthesis of modified proteins. During mRNA translation, amber stop codons are suppressed to site-specifically incorporate non-canonical amino acids. Thus, nanobodies can be equipped with photocaged amino acids to control target binding on demand. The efficiency of amber suppression and protein synthesis can vary with unpredictable background expression, and the reasons are hardly understood. Here, we identified a substantial limitation that prevented synthesis of nanobodies with N-terminal modifications for light control. After systematic analyses, we hypothesized that nanobody synthesis was severely affected by ribosomal inaccuracy during the early phases of translation. To circumvent a background-causing read-through of a premature stop codon, we designed a new suppression concept based on ribosomal skipping. As an example, we generated intrabodies with photoactivated target binding in mammalian cells. The findings provide valuable insights into the genetic code expansion and describe a versatile synthesis route for the generation of modified nanobodies that opens up new perspectives for efficient site-specific integration of chemical tools. In the area of photopharmacology, our flexible intrabody concept builds an ideal platform to modulate target protein function and interaction.

KEYWORDS: genetic code expansion, intrabodies, optochemical biology, photoactivation, protein modification, photopharmacology
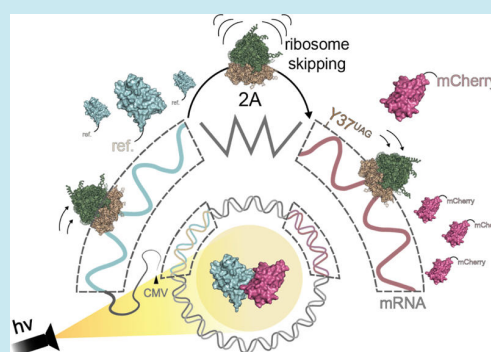
INTRODUCTION

Conditional protein modifications are fundamental in the life sciences. Genetic code expansion connects the versatility of chemical synthesis to protein expression in living systems.[1−10] Protein modifications can be inserted site-specifically *via* natural translation machineries.[11] By reprogramming the genetic code, non-canonical amino acids bearing a modification of choice are used in ribosomal polypeptide synthesis.[12−14] Non-canonical amino acids have many applications in protein engineering, as they can be equipped with isotopes for structural studies,[15−17] photoreactive groups and post-translational modifications for functional studies,[18−29] reactive groups for bio-orthogonal coupling,[13,15,30−34] photocross-linkers,[1,22,35−40] infrared-active probes to follow conformational dynamics,[6,36−41] fluorescent dyes for imaging,[42−46] biotin analogues for high-affinity interactions with streptavidin,[47] and stable phosphotyrosine analogues for analysis of signal transduction.[23,48] Site-specific incorporation is achieved by suppressing a stop codon with an additional aminoacyl-tRNA-synthetase (aaRS)/tRNA pair.[4,14,49,50] In bacterial or mammalian cells, the rarest amber codon (TAG) is most often used to minimize suppression throughout the proteome.[5,6,12,33,51] However, the underlying processes of the genetic code expansion are complex and hardly understood.[6,7] For efficient protein synthesis, an ideal interplay of all chemical and biological components is required.[12,50] To date, no standard approach for the screening of optimal parameters is available. In mammalian cells, the *in vivo* synthesis of modified proteins is further challenged by the relatively high abundance (23%) of the amber codon, in contrast to only 9% in *Escherichia coli*.[52−54] The potential for incorporation in response to endogenous amber codons can be partially prevented by the cellular compartmentalization of the mRNA using liquid-phase segregation or membrane recruitment.[55,56] However, the problem of cells incorporating natural amino acids at TAG codons is well known, particularly for Phe and Tyr analogues. Though background-causing amino acids may be dropped out of the media in *E. coli*,[11,57] methods for addressing this problem in eukaryotic species have not been reported.

Here, we systematically identified a potential interference factor and reported a novel synthesis strategy to circumvent limitations for synthesis of modified nanobodies. After
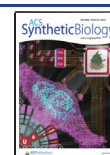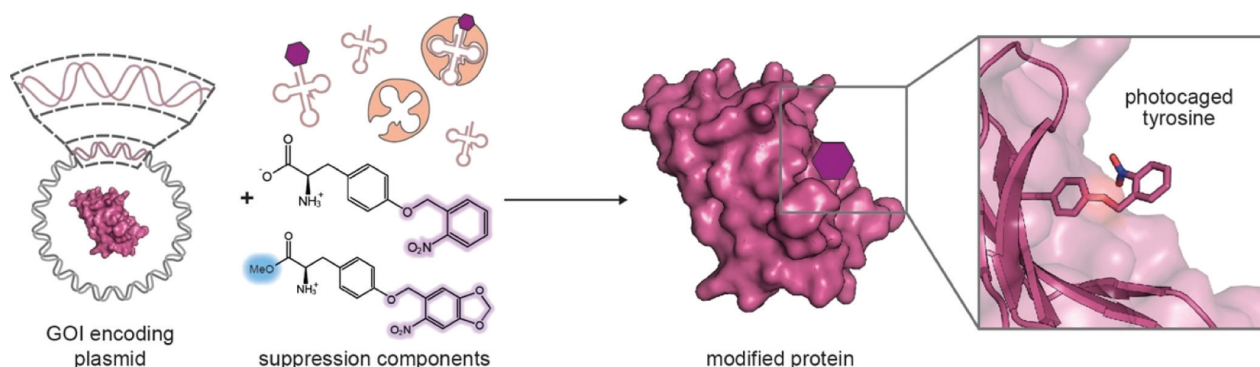
Figure 1. Genetic code expansion enables the synthesis of the modified proteins (pink). To generate the functionalized proteins in living cells, natural translation machineries are reprogrammed. Using an orthogonal aaRS/tRNA pair (beige), amber stop codons are suppressed to site-specifically introduce a non-canonical amino acid (magenta). For photoactivated binding, photocaged tyrosine analogous are introduced at position Tyr37 of the α-GFP nanobody (enhancer).

introducing an amber stop codon at a strategically defined position in the 5′ region of the gene of interest (GOI), we observed an inefficient termination of translation. The resulting background expression impeded the synthesis of the modified protein by genetic code expansion. A systematic analysis revealed frequent read-through across the premature stop codon. Recent studies using ribosome display technologies reported reduced accuracy in early phases of translation in the 5′ region.[58−61] We rationalized that premature amber stop codons in nanobodies are prone to such infidelity. For amber suppression with high efficiency, we present a new approach that improves the fidelity of ribosomal synthesis, particularly for amber stop codons present in the 5′ region of the mRNA.

As an outstanding example, we generated intrabodies with photoactivated binding to intracellular antigens. The strategy opens new avenues because intrabodies are ideal for imaging or modulating intracellular targets. Controlling the protein function with light enables experimental and therapeutic applications with minimal interference and high spatiotemporal precision. In general, genetic code expansion allows the synthesis of photoactivatable (PA) proteins.[21,26] To this end, amino acids equipped with large photolabile protection groups are site-specifically introduced at crucial positions. Active sites and interaction interfaces, for example, constitute excellent targets. In this context, we described the synthesis of a conditional interaction pair *via* a genetic code expansion in a stable human cell line.[62] To generalize the concept to other proteins and expression hosts and to reduce the amount of work required, it would be beneficial to avoid stable genomic integration. Moreover, synthesis based on transient transfection would facilitate flexible tagging to modulate the target protein.

■ RESULTS AND DISCUSSION

To generate a PA version of the α-GFP nanobody (enhancer), an amber mutation was introduced at position Tyr37.[63,64] Incorporation of a tyrosine analogue bearing a bulky photolabile protection group controls target binding (Figure 1). After photocleavage, the native paratope is restored, allowing rapid high-affinity binding of the target. For transient transfection, we cloned the construct in a plasmid for expression in mammalian cells (Figure 2a, construct I). To visualize protein synthesis, we used a C-terminal fusion to the red fluorescent protein mCherry. An engineered variant of the

pyrrolysyl-tRNA (PylT) synthetase (PylRS) pair from *Methanosarcina barkeri* was co-transfected in order to suppress the premature amber stop codon.[26] First, we selected *ortho*-nitrobenzyl-caged tyrosine (ONBY) to replace Tyr37 (Figure 2b).[22] After transient transfection of HeLa cells, we analyzed the synthesis efficiency of the PA nanobody ($^{PA}Nb^{mCherry}$) *via* flow cytometry. The fluorescence intensities of cells expressing the unmodified (UM) construct ($^{UM}Nb^{mCherry}$) were used for normalization (Figure S1). After solely transfecting the $^{PA}Nb^{mCherry}$-encoding plasmid in the absence of ONBY, we observed up to 40% cells with similar fluorescence intensities.[11] High ratios of mCherry-positive cells indicated that the premature amber codon did not terminate mRNA translation in the absence of amber suppression components. In contrast, 30% positive cells were recorded under amber suppression conditions by co-transfecting $^{PA}Nb^{mCherry}$ with the orthogonal aaRS/tRNA pair and supplementing the medium with ONBY (Figure 2c). These results demonstrated insufficient suppression of the premature termination signal.

To improve amber suppression, we equipped the $^{PA}Nb^{mCherry}$ encoding plasmid with four additional copies of the orthogonal tRNA (PylT)[65,66] (Figure 2a, construct II). Here, 15% of mCherry-positive cells were observed under amber suppression conditions, and 20% after solely transfecting the plasmid in the absence of ONBY (Figure 2c). We excluded the background expression of unfused mCherry by replacing the start methionine with an alanine (Met1Ala) to eliminate the possibility that this could be used as a translational initiation site (Figure 2a, construct III). Again, high ratios of mCherry-positive cells were recorded in the absence of amber suppression components. To identify the source of mCherry fluorescence, we analyzed the translation products in the cell lysates by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS-PAGE) and in-gel fluorescence. We observed the full-length fusion construct Nb^{mCherry} in all samples. Moreover, amber construct samples showed similar fluorescence intensities (Figure S2). These surprising results indicated that the amber stop codon Tyr37^{TAG} failed to terminate translation efficiently. Consequently, the read-through caused a background protein expression, competing with efficient synthesis of the modifed protein.

One reason for insufficient termination of translation could be an unbalanced mRNA/tRNA ratio. To harmonize the RNA level, we cloned the constructs into a plasmid with a previously
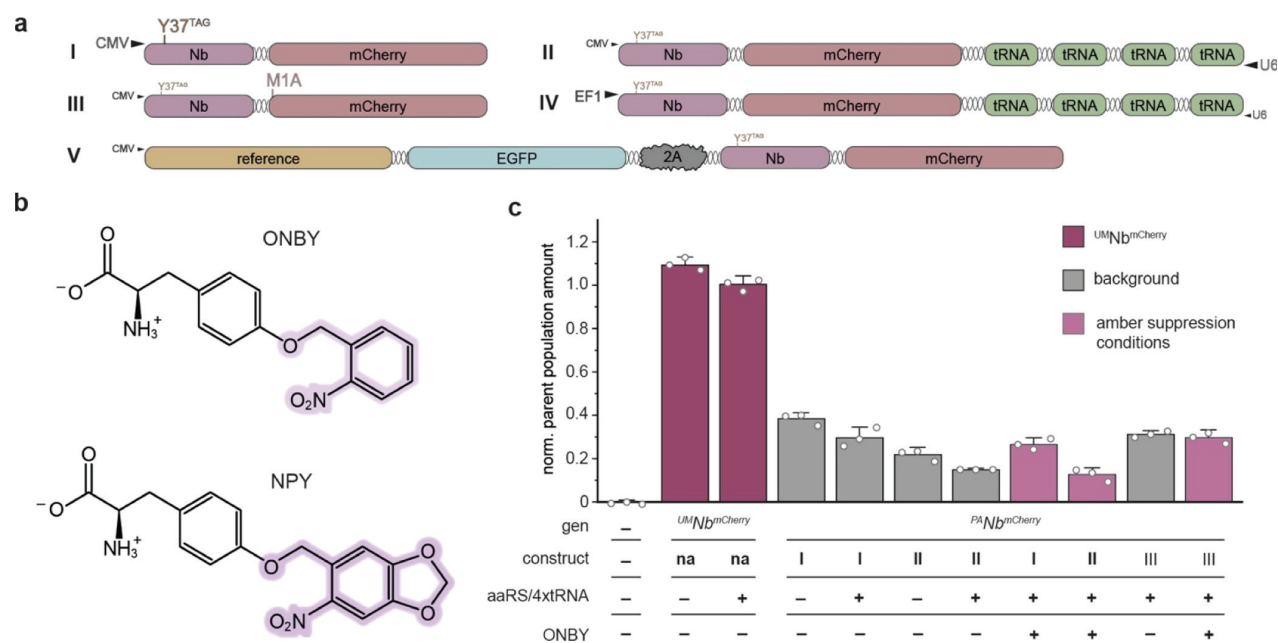
Figure 2. Synthesis of the modified protein in mammalian cells by amber suppression. (a) Expression cassettes for the generation of $^{PA}Nb^{mCherry}$ intrabodies by transient transfection. To site-specifically insert photocaged tyrosines (b), an amber codon was introduced at position Tyr37. For amber suppression, an orthogonal aaRS/4xtRNA pair was co-transfected. (c) Quantity of $Nb^{mcherry}$-positive cells after transient transfection of expression cassettes under the control of the CMV promoter (constructs I–III), monitored by flow cytometry. The premature amber codon at position Tyr37-inadequately terminated translation. Background expression limited sufficient amber suppression for the synthesis of $^{PA}Nb^{mCherry}$ (construct I). Efficient amber suppression was not achieved by encoding four additional copies of the orthogonal tRNA (construct II). The parallel synthesis of unfused mCherry was excluded by the substitution of the potential start codon with an mCherry-M1A mutation (construct III). aaRS/4xtRNA = NPYRS/4xPylT co-transfection, ONBY = medium supplemented with 0.25 mM ONBY. Living HeLa cells 24 h after transfection, mean amount of $n = 3$ biological replicates ± SD.

described design for amber suppression.[12,13,67] In this case, the GOI is under control of the strong EF-1 alpha promoter and amber suppression supported by four additional PylT copies (Figure 2a, construct IV). After transient transfection, we measured background expression by flow cytometry. Relative to the UM construct, up to 90% cells showed similar fluorescence intensities in the absence of suppression components (Figure S3). Again, the amber stop codon Tyr37^{TAG} failed to terminate translation, indicating that altering the mRNA/tRNA ratio did not prevent an undesired read-through.

Recent advances in ribosome profiling and other technologies demonstrated reduced fidelity in early phases of translation.[58−61] We rationalized that the early amber codon at position Tyr37 is prone to such inaccuracy. In a recent study, we observed slightly increased amber suppression efficiency by using a bicistronic translation approach.[29] In order to circumvent the $Nb^{mCherry}$ background expression, we developed a new approach inspired by the translation strategy of the foot-and-mouth disease virus (Figure 1a, construct V). First, we implemented an upstream encoded GOI to our expression cassette to reduce translational slowdown and inaccuracy. To analyze $^{PA}Nb^{mCherry}$ binding, we co-expressed the cognate target protein (EGFP), which can be fused to many reference proteins *via* a multiple cloning site. As an initial reference, we used the ER-resident transporter associated with antigen processing (TAP1).[68,69] Second, we connected both gene constructs *via* the foot-and-mouth disease virus ribosomal skipping site (F2A), which allows bicistronic translation of two proteins (target and nanobody) in a single ribosome passage

(Figure 3a).[29,70] Polycistronic mRNAs are a valuable strategy for the expression of certain genes in mammalian cells that is gaining increasing recognition.[71] Parallel protein synthesis was monitored using flow cytometry to detect the two fluorescent reporters (Figure S4). In the absence of amber suppression components, no mCherry-positive cells were detected (Figure 3b), while the expression of the GFP-tagged reference protein indicated efficient transient transfection (Figure S4). Gratifyingly, the new approach restored the translational stop function of the amber codon at position Tyr37. In contrast to all previous approaches (constructs I−IV), efficient amber suppression was observed after this complete silencing of the background expression. By suppressing the amber codon with ONBY, we observed up to 50% $Nb^{mCherry}$-positive cells (Figure 3b). To investigate the synthesis of $^{PA}Nb^{mCherry}$ with improved photodeprotection properties, we suppressed the amber codon with nitropiperonyl-caged tyrosine (NPY). After transfection, we observed 60% of cells that contained the PA intrabody (Figure 3b). Background expression was minimal in the absence of amber-suppression components. These results demonstrate high fidelity in amber suppression and photocaged tyrosine incorporation. Importantly, the upstream reference protein restored the functionality of the amber codon at position Tyr37 and allowed efficient genetic code expansion for the generation of photoconditional intrabodies. For simplicity, we termed the new plasmid p2A-Amber (Figure 3a).

In contrast to the synthesis in stable cell lines, expression by transient transfection is flexible toward other hosts, with no need for stable transfection and tedious selection. To scale up
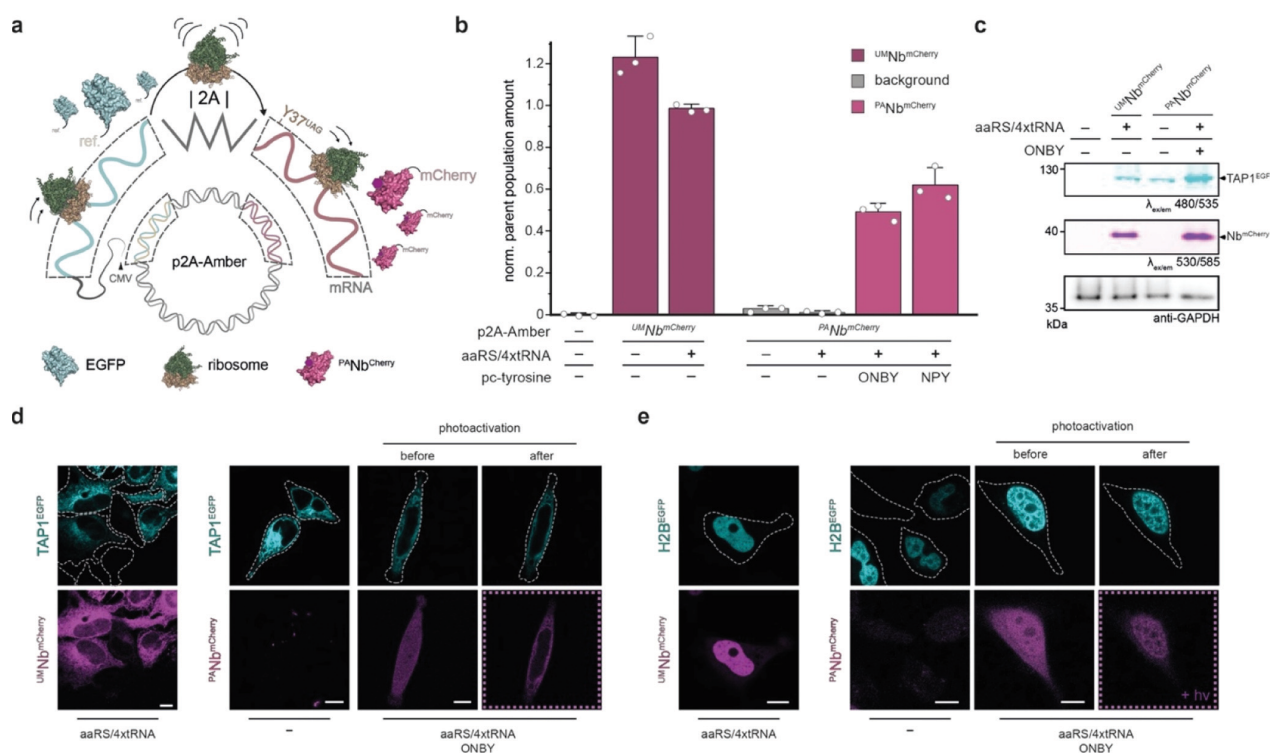
Figure 3. Synthesis of the modified protein in mammalian cells using p2A-Amber. (a) Plasmid design circumvents ribosomal inaccuracy in early phases of translation for efficient amber suppression by adding an upstream protein. As an example, the photoactivatable intrabody $^{PA}Nb^{mCherry}$ is generated. To facilitate intracellular binding analysis, a GFP-tagged reference protein was encoded upstream of the intrabody gene. The proteins are separated during translation *via* a viral ribosomal skipping (F2A) site. (b) Amber suppression in HeLa cells monitored by flow cytometry. The plasmid design of p2A-Amber restored the termination signal at the position Tyr37 of the $^{PA}Nb^{mCherry}$ gene. Thereby, efficient amber suppression generated site-specifically modified intrabodies by transient transfection. Relative to the unmodified (UM) protein $^{UM}Nb^{mCherry}$, high quantity of $^{PA}Nb^{mCherry}$-containing cells were observed *via* the C-terminal mCherry. Living HeLa cells 24 h after transfection, mean quantity of $n = 3$ biological replicates ± SD. (c) Large-scale synthesis of the modified proteins by HEK-F cells after the transient transfection of p2A-Amber. In-gel fluorescence of cell lysates (2.5 μg total protein), displaying efficient generation of $^{PA}Nb^{mCherry}$. Before SDS-PAGE, cells were analyzed by flow cytometry and $^{PA}Nb^{mCherry}$-containing cells were enriched. Importantly, in the absence of ONBY, no $^{PA}Nb^{mCherry}$ were observed. (d) *In situ* photoactivation of $^{PA}Nb^{mCherry}$ inside living cells. Intracellular binding of the ER resident reference protein TAP1^{EGFP} was corroborated by co-localization. (e) Photo-activated $^{PA}Nb^{mCherry}$ binding of the reference histone protein H2B^{EGFP}. Imaging conditions were adjusted to the level of the fluorescent protein expression for cells transfected with H2B^{EGFP} | F2A | $^{UM}Nb^{mCherry}$. (d,e) CLSM live-cell imaging of living HeLa cells. Scale bar = 10 μm. (b−e) Mammalian cells were transiently transfected with p2A-Amber containing the UM construct ($^{UM}Nb^{mCherry}$) or the genetic code expansion construct ($^{PA}Nb^{mCherry}$) with the premature amber codon at position Tyr37. For the suppression of the amber codon, cells were co-transfected with a plasmid encoding an orthogonal aaRS/tRNA pair (aaRS/4xtRNA) and the medium was supplemented with photocaged tyrosine (0.25 mM, ONBY).

the synthesis of the modified protein, we analyzed our p2A-Amber strategy in HEK293-F cells. After transient transfection, we enriched $^{PA}Nb^{mCherry}$-positive cells by flow cytometry and analyzed protein synthesis by SDS-PAGE (Figure S5). In-gel fluorescence showed high levels of $^{PA}Nb^{mCherry}$ and the reference protein (TAP1^{EGFP}) when ONBY was present. In the absence of ONBY components, the amber codon terminated the synthesis of the mCherry-tagged intrabody. The analysis further validated efficient translational processing and cleavage into the two proteins by the ribosomal skipping site (Figure 3c).

Next, we analyzed the functionality of the inserted protein modification. Using confocal laser scanning microscopy (CLSM), we visualized $^{PA}Nb^{mCherry}$ in living HeLa cells that were transiently transfected with our new p2A-Amber plasmids. Under amber suppression conditions with ONBY, we observed a homogeneous distribution of $^{PA}Nb^{mCherry}$ in the cytosol (Figure 3d,e). The results confirmed that the photocage blocked the high-affinity interaction with EGFP-tagged target proteins in comparison to the $^{UM}Nb^{mCherry}$

(Figure S6). Immediately after *in situ* photodeprotection, we detected $^{PA}Nb^{mCherry}$ binding of the antigen transport complex (TAP1^{EGFP})[69] or histone H2B located at the ER membrane or the nucleus, respectively (Figure 3d,e). The same results were obtained by generating $^{PA}Nb^{mCherry}$ with NPY *via* a p2A-Amber-mediated amber suppression (Figure S7). Collectively, these results confirmed the efficient processing and synthesis of PA intrabodies *via* genetic code expansion and ribosomal skipping.

In general, synthesis of non-canonical amino acids is quite laborious, and material consumption for genetic code expansion is high. Especially for the purification of recombinantly expressed proteins to homogeneity, large quantities of non-canonical amino acids are required. To optimize amber suppression, we combined our p2A-Amber approach with a methyl ester NPY concept.[72,73] The generation of methyl esters promises enhanced membrane crossing and protein incorporation.[73] Accordingly, we synthesized the methyl ester of NPY (NPY-Me) and quantified genetic code expansion by flow cytometry (Figure S8a). By

using various concentrations of NPY and NPY-Me, we systematically identified the optimal concentrations for amber suppression. At high concentrations, NPY-Me seemed to be toxic. However, we observed that a 25-fold lower concentration of NPY-Me (only 10 μM) was sufficient to evoke more than 60% of NPY-mediated amber suppression (Figure S8). For large-scale expression studies, these results will help to significantly decrease the amount of the precious materials. These results pave the way for improved future applications of this versatile sterically bulky UAA that has excellent photophysical properties.[26]

CONCLUSIONS

We systematically identified a deadlock of a genetic code expansion with nanobodies/intrabodies and circumvented inefficient protein synthesis by a new translation strategy. In this context, we identified that an early amber stop codon is potentially prone to ribosomal infidelity during early phases of mRNA translation corresponding to the N-terminal region of the nascent chain. The resulting background expression can severely affect the efficiency of amber suppression. We sought to eliminate background by pre-starting translation with an upstream gene. Inspired by the translational strategy of the foot-and-mouth disease virus, we implemented a ribosomal skipping site to separate both proteins during translation. The new approach establishing plasmid p2A-Amber allowed synthesis of modified nanobodies/intrabodies in mammalian cells with high efficiency. Due to its productive translational initiation, the resulting mRNA is likely to be highly stable,[74] which will allow efficient expression of a variety of single-chain binders for genetic code expansion. Correspondingly, we generated photoactivatable intrabodies in mammalian cells, which targeted different proteins in the ER membrane and nucleus. Similar to the viral ribosome skipping site, N-terminal fusions with the STELLA-tag, the GCE-tag, ubiquitin, SUMO, and other ubiquitin-like proteins can be applied, which could allow efficient genetic code expansion in 5′-regions of the GOI.[75−77] However, some proteins are potentially affected by N-terminal fusions. For nanobodies, the N-terminus is located close to the interaction-mediating CDR loops, which limits the use of N-terminal fusions in general. Here, the p2A-Amber approach facilitates simple and efficient co-translation of the reference and target proteins, with no potentially interfering changes to the latter.

Photoactivatable intrabodies are ideal tools to monitor or modulate target proteins in living cells. We controlled binding by introducing a photocaged tyrosine into the interaction interface of the α-GFP enhancer nanobody (^{PA}Nb^{mCherry}). Previously, we reported synthesis of photoactivatable intrabodies in stable mammalian cell lines. The new synthesis route based on transient transfection allows flexible upgrades of ^{PA}Nb^{mCherry} with additional fusion tags for target protein modulation. For example, combining ^{PA}Nb^{mCherry} with a degradation signal will allow photoconditional protein knock-downs to study essential genes and complex pathways. Additionally, the plasmid set facilitates the expression in other cells and higher protein yields.

The direct contrast to our stable cell lines[62] reveals further details about underlying processes. Stable transfection enabled a background-free synthesis of the modified nanobody. In this case, gene expression was carefully controlled via tetracycline induction to generate nearly endogenous protein levels. By transient transfection, a strong overexpression is evoked. In the case of transient transfection using conventional expression plasmids, the high mRNA levels can increase the likelihood of read-through events. Accordingly, we observed the highest quantity of Nb^{mCherry} in the absence of amber suppression components under these conditions (constructs I−IV). In the presence of an orthogonal aaRS/tRNA pair, a competition of read-through and amber suppression events further compli-cates the procedure. By implementing a bicistronic expression of an upstream encoded gene via p2A-Amber, this competition is shifted toward amber suppression events.

Antibody/nanobody drug conjugates offer a broad range of pharmaceutical applications. Our approach offers a direct route to photopharmacology[78,79] using therapeutic antibodies. For conjugate generation, efficient synthesis and precise mod-ification are key requirements.[80,81] By using our p2A-Amber strategy, high quantity of site-specifically modified nanobodies can be expressed in human cells. Because our approach is customizable, it can be a robust route for the synthesis of nanobodies harboring other non-canonical amino acids for various applications such as bioconjugation.

In summary, we acquired new insights about genetic code expansion by amber suppression. We showed that an early amber codon in nanobodies is prone to translational read-through. Consequently, efficient incorporation of unnatural amino acids is impeded. The novel plasmid design p2A-Amber, inspired by the foot-and-mouth-disease virus, circumvents the limitation. Hence, pA2-Amber offers efficient synthesis of modified nanobodies or similar proteins. Our findings will help to expand the toolbox of genetic code expansion and allow new in vivo applications of chemically modified proteins.

MATERIALS AND METHODS

Molecular Biology. Enzymes for molecular biology were used by following the manufacturers' instructions. For DNA amplification by PCR, Phusion high-fidelity DNA polymerase was used. Ligation was performed using T4 DNA ligase. The mCherry-tagged α-GFP enhancer (Nb^{mCherry}) was previously cloned into pcDNA3.1(+) by PCR and restriction enzymes.[82] The amber mutation at position Tyr37 was inserted as previously described by site-directed mutagenesis using the following primer pair: fwd 5′-AGC ATG CGT TGG TA**G** CGT CAG GCA CCG-3′, rev 5′-CGG TGC CTG ACG C**TA** CCA ACG CAT GCT-3′ (mutation in bold).[62] For the generation of a pcDNA3.1(+) construct with additional tRNA copies (plasmid II), the plasmid SE323 with four PylT copies under the control of individual U6 promotors was used. The plasmid SE325 was kindly provided by Kathrin Lang (TU Munich).[13] All four tRNA genes were cloned by PCR and endonuclease restriction into pcDNA3.1(+) encoding Nb^{mCherry} with the amber mutation at position Tyr37. For PCR, a primer pair introducing an upstream MfeI restriction site and a downstream Bpu10I restriction site was applied. *MfeI* fwd: 5′-GTG GCG CAA TTG GGG GAT ACG GGG AAA AGG-3′, *Bpu10I* rev: 5′-GCG GTG CCT AAG CGG CAC CGG AGC GAT CGC AGA T-3′ (restriction sites are underlined). The mCherry mutant Met1Ala (plasmid III) was generated by site-directed mutagenesis using the pcDNA3.1-(+) plasmid encoding Nb^{mCherry} with amber mutation. Base-exchange was performed with the following primer pair: fwd 5′-GTT AGC AGC GGT ACC **GCG** GTG AGC AAG GGC GAG-3′ and rev 5′-CTC GCC CTT GCT CAC **CGC** GGT ACC GCT GCT AAC-3′ (mutation in bold). For cloning of the mammalian expression vector with additional tRNA copies

Scheme 1. Synthesis of NPY (1) and NPY-Me (2) *via* Reaction of a Dityrosine−Cu²⁺ Complex with *o*-NPBr and Removal of the Cu²⁺ by Heating with a Solution Containing Excess Ethylenediaminetetraacetic acid (EDTA)
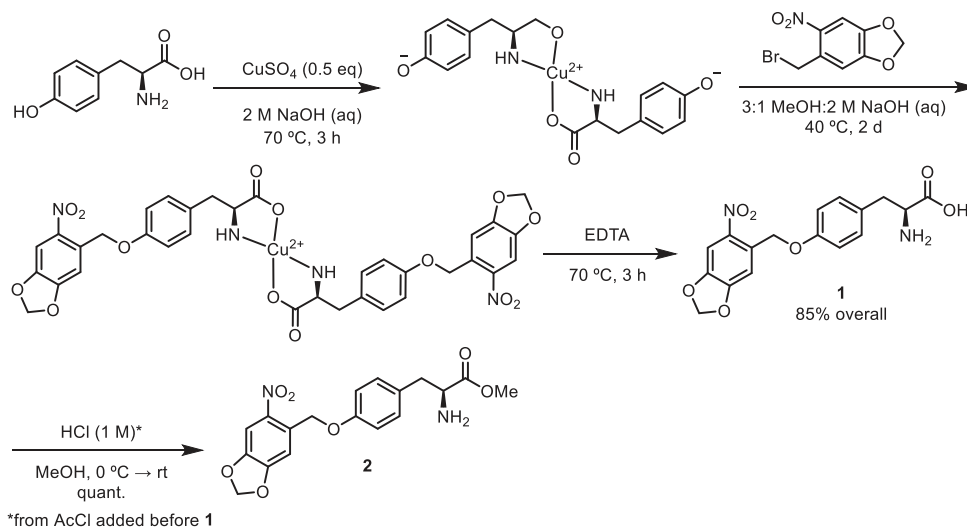
and a moderate EF-1 alpha promotor (plasmid IV), the Nb^mCherry genes were cloned *via* endonuclease restriction into SE323 by exchanging a PylRS gene. Therefore, the UM Nb^mCherry gene and the amber mutant encoded by pcDNA3.1-(+) were amplified by PCR using a primer pair that maintained an upstream NheI restriction site and inserted a downstream BamHI restriction site. *NheI* fwd: 5′-AGA CCC AAG CTG GCT AGC ACC-3′, *BamHI* rev 5′-GTG GCG GGA TCC GGT TTA AAC GGG CCC TCT AGA-3′ (restriction sites are underlined). A pcDNA3.1(+) plasmid encoding a multiple cloning site and downstream mEGFP| 2A |Nb^mCherry with the amber mutation at position Tyr37 of the nanobody was designed for amber suppression after transient transfection (construct V, p2A-Amber). The plasmid was synthesized by GenScript. TAP1 or H2B was cloned into the multiple cloning site *via* endonuclease restriction. The TAP1 gene was amplified by PCR originating from a previously generated plasmid encoding TAP1^myc-SBP| 2A |TAP2^mVenus-His10.[29] A primer pair binding an upstream NheI restriction site and inserting a downstream BamHI restriction site was applied. *NheI* fwd: 5′-CCA AGC TGG CTA GCG TTT AAA CTT AAG CTT AGC-3′, *BamHI* rev: 5′-GCG GGC GCT CAC GGA TCC TTG TGG ACC ATC AGG AGC GTC AGC AGG AGC CTG GAC-3′ (restriction sites are underlined). Furthermore, the H2B gene for the generation of H2B^EGFP| 2A |^PANb^mCherry was obtained by endonuclease restriction, based on a previously described plasmid encoding H2B^EGFP. The plasmid H2B^EGFP in pEGFP-N1 was a gift from Geoffrey Wahl (Addgene #11680). After endonuclease restriction using HindIII upstream and BamHI downstream, the resulting frameshift was removed by site-directed mutagenesis using the following primer pair fwd: 5′-ACC AGC GCT AAG GAT CCA ATG GTG AGC AAG GGC GAG-3′ and rev: 5′-CTC GCC CTT GCT CAC CAT TGG ATC CTT AGC GCT GGT-3′ (insertion bold). For the generation of the UM intrabody (^UMNb^mCherry), the amber mutation Tyr37^TAG of the p2A-Amber constructs was exchanged toward the wild-type tyrosine by site-directed mutagenesis. The following primer pair with fwd 5′-CGT TAT AGC ATG CGT TGG **TAT** CGT CAG GCA CCG-3′ and rev 5′-ACG TTC TTT ACC CGG

TGC CTG ACG **ATA** CCA ACG CAT GCT-3′ (mutation underlined) was used. The plasmid encoding the orthogonal aaRS/tRNA pair (NPYRS/4xPylT) was previously generated.[26,62] The pyrrolysine tRNA synthetase (PylRS) of *M. barkeri* contained the following mutations: L270F, L274M, N311G, and C313G.[26]

Synthesis and Analytics of Non-Canonical Amino Acids. NPY was synthesized using a procedure modified from a published report.[26] ONBY was purchased from Santa Cruz Biotechnology. Compound identity was validated by mass spectrometry (Figures S8−S11).

NPY-Me was synthesized as follows. All reagents and solvents were obtained from commercial suppliers and used without further purification. Reactions were monitored by thin-layer chromatography (TLC) using glass-backed silica gel (60 Å, F_254) plates (EMD Millipore). Flash chromatography was performed with silica gel (Sorbtech, 60 Å, 230−400 mesh). Nuclear magnetic resonance (NMR) spectra were obtained on 500 MHz or 600 MHz Bruker NMR spectrometers, and chemical shifts are in δ units (ppm) relative to the solvent signal. HRMS data were obtained using a Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific) (Scheme 1).

o-Nitropiperonyl Bromide. This compound was synthesized by the Appel reaction according to literature procedures and was obtained as a yellowish tan solid in 90% yield (2.39 g, 9.19 mmol). Characterization data matched reported results.[83]

(S)-2-Amino-3-(4-((6-nitrobenzo[d][1,3]dioxol-5-yl)-methoxy)phenyl)propanoic Acid (NPY, 1). A solution of copper sulfate pentahydrate (444.4 mg, 1.780 mmol, 0.5 equiv) in water (1.5 mL) was added dropwise to a stirred solution of L-tyrosine (645.0 mg, 3.564 mmol) in 2 M NaOH (2.8 mL, 5.6 mmol, 1.6 equiv) at 70 °C. After 1 h, the reaction mixture was cooled to 40 °C, and MeOH (10 mL) and additional 2 M NaOH (1.9 mL, 1.8 mmol, 0.5 equiv) was added. *o*-Nitropiperonyl bromide (984.1 mg, 3.784 mmol, 1.06 equiv) was added in 10 portions over 15 min. The mixture was stirred for 2 days at 40 °C in the exclusion of light (in a vial wrapped in aluminum foil). The mixture was centrifuged (500*g*, 1 min) and washed three times by resuspending in 1:3 MeOH/water

(10 mL) with a vortex mixer, repeating the centrifugation step, and discarding the supernatant. A solution of disodium EDTA dihydrate (796.6 mg, 2.140 mmol, 0.6 equiv) in water (10 mL) was added, and the mixture was stirred at 70 °C overnight. The product was cooled and was filtered in a funnel containing a glass frit and was washed with ice-cold water (20 mL) and ice-cold acetone (2 × 15 mL). The remaining solids were dried *in vacuo*, affording the product as a beige solid in a 85% yield overall (1.08 g, 3.01 mmol). Characterization data matched reported results.[26]

Methyl (S)-2-Amino-3-(4-((6-nitrobenzo[d][1,3]dioxol-5-yl)methoxy)phenyl)propanoate (NPY-Me, 2). NPY (99.0 mg, 0.275 mmol) was added to a methanolic solution of 1 M HCl that had been prepared by adding AcCl (210 μL, 3.00 mmol) dropwise to anhydrous MeOH (3.0 mL) in a flame-dried vial on ice. The reaction mixture was allowed to warm to room temperature and was stirred overnight [note: TLC analysis (84:15:1 EtOAc/MeOH/Et$_3$N) of the reaction mixture shows a spurious spot at the baseline; a 1 M HCl in MeOH blank also produces this spot, suggesting that the solution destroys the TLC fluorescent indicator]. The reaction mixture was concentrated onto silica gel *in vacuo*. A silica gel column was neutralized by equilibration with 0.1% (v/v) Et$_3$N in EtOAc, and the sample was purified by flash chromatography (99:1 → 75:25 EtOAc/MeOH), affording the product as a pale yellow powder in quantitative yield (104.3 mg, 0.2789 mmol). ^{1}H NMR (600 MHz, DMSO-d_6): δ (ppm): 3.06 (m, 1H), 3.14 (m, 1H), 3.66 (s, 3H), 4.17 (t, J = 6.6 Hz, 1H), 5.35 (s, 2H), 6.26 (s, 2H), 6.96 (d, ^3J = 8.4 Hz, 2H), 7.17 (d, ^3J = 8.4 Hz, 2H), 7.26 (s, 1H), 7.74 (s, 1H), 8.74 (br s, 2H); ^{13}C NMR (151.1 MHz, DMSO-d_6): δ (ppm): 35.4, 53.0, 53.8, 67.0, 104.1, 106.0, 108.3, 115.4, 127.8, 130.5, 131.2, 141.8, 147.7, 152.6, 157.5, 169.9; HRMS-ESI (m/z) [M + H]$^+$ calcd for C$_{18}$H$_{19}$O$_7$N$_2$: 375.11868 Da, found 375.11741 Da.

Mass Spectrometry. Photocaged tyrosines, ONBY and NPY solved in phosphate buffered saline (PBS), were analyzed using a Waters BioAccord system running UNIFY 1.9.4. Individual samples were separated within an Acquity BEH C18 column (1.7 μm, 2.1 mm × 100 mm) and subjected to ESI-TOF mass spectrometry using a cone voltage of 30 V and capillary voltage of 0.8 kV in positive polarity.

Cell Culture. Adherent mammalian cells (HeLa, HEK293-F) were maintained in Dulbecco's modified Eagle medium (DMEM) containing 4.5 g/L glucose, supplied with 10% (v/v) fetal calf serum (FCS) in T75 cell culture flasks. For passaging, cells were detached by using PBS (Sigma-Aldrich) and 0.05% trypsin−EDTA (Gibco) every 2−3 days. Cells were cultivated in a tissue culture incubator at 37 °C and humidified with 5% CO$_2$. Following the established guidelines, mycoplasma contamination tests were regularly carried out.[84] For live-cell imaging of HeLa cells, 2.5 × 10^4 cells were seeded per glass-bottom 8-well imaging slide. For flow cytometry analysis of HeLa cells, 5 × 10^5 cells were seeded per 6-well, harvested using PBS and 0.05% trypsin−EDTA, washed once with DMEM and once with PBS. To establish suspension cultures of HEK293-F cells for high-scale synthesis of ^{PA}Nb^{mCherry}, adherent cells were detached using PBS and 0.05% trypsin/0.02% EDTA/PBS. Cells were pelleted, washed in PBS (800 g, 5 min, RT), and resuspended in 100 mL of FreeStyle 293 medium. The suspension culture was incubated at 37 °C, 5% CO$_2$, and 125 rpm in 1 L Erlenmeyer flasks. For flow cytometry and cell sorting, cells were pelleted (800 g, 8 min, 4 °C) and washed twice with PBS before analysis.

Cell Transfection and Amber Suppression. For the expression analysis of the various plasmid I−V, HeLa cells were transiently transfected using Lipofectamine 2000 according to the manufacturer's instructions. Solubilized HEK293-F cells were transiently transfected using PEI for a large-scale expression. Two days before transfection, HEK293-F cultures were adjusted to 1 × 10^6 cells/mL. One day before transfection, cell concentrations were reduced to 0.5 × 10^6 cells/mL. For the transfection of a 100 mL culture, two mixtures were prepared. 100 μg of the plasmid (200 μg for co-transfection, 1:1 ratio) was diluted in 3.9 mL of Opti-MEM. In parallel, 400 μL of PEI (18 mM) was added to 3.25 mL of Opti-MEM. After 5 min, both solutions were mixed and incubated for 30 min at room temperature. For transfection, the solution was carefully added to the cell flasks. In order to ensure an optimal mixture, flasks were gently agitated during the procedure. To enable genetic code expansion by amber suppression, cells were co-transfected with the respective plasmids (constructs I−V, aaRS/4xtRNA), and 4−6 h later the medium was supplemented with non-canonical amino acids. For adherent cells, the medium was exchanged with DMEM, 4.5 g/L glucose, 10% (v/v) FCS, containing 250 μM ONBY or NPY, respectively. Stock solutions of the non-canonical amino acids were prepared in 100 mM NaOH and sterile-filtered. After stock addition to DMEM medium, the pH was neutralized by the same volume of 100 mM sterile-filtered HCl. NPY-Me was dissolved in dry DMSO at 50 mM to generate a stock solution. For suspension HEK293-F cultures, stocks of non-canonical amino acid and HCl were added in parallel to the cells cultured in the FreeStyle 293 medium. Again, flasks were gently agitated during the procedure. For optimal amber suppression, cells were cultivated in the corresponding medium for 24−48 h after transfection.

Flow Cytometry and Analysis. To analyze the synthesis of the modified proteins, the C-terminal mCherry tag was recorded *via* flow cytometry. For quantification, amounts of cells showing fluorescence intensities similar to the UM intrabody were used. Samples were recorded in biological triplicates by fluorescent-activated cell sorting (FACSMelody, BD Biosciences). mCherry fluorescence was excited with a 561 nm laser line and passed through a 613/18/LP605/10 filter set. To all samples, standard gates for doublet discrimination were applied. Flow cytometry data were evaluated using FlowJo 10.6.2 (BD), with the cell count of a mCherry transfection gate evaluated and plotted in OriginPro 2020 (OriginLab). For amber suppression efficiency analysis, the cell portions of the doublet-discriminated cells in the amber-supression gate were background corrected and normalized to the highest applied NPY concentration. Background correction was performed with a UAA-free sample. For the sorting of HEK293-F cells showing mCherry fluorescence in the range of ^{UM}Nb^{mCherry}, a 100 mL culture was used.

Fluorescence Microscopy and Intrabody Photoactivation. For live cell CLSM imaging, cells cultivated in 8-well cell culture chambers were rinsed once and covered with a Ringer's solution (145 mM NaCl, 5 mM KCl, 2 mM CaCl$_2$, 1 mM MgCl$_2$, 10 mM HEPES, 10 mM glucose, pH 7.4). During imaging, chambers were incubated at 37 °C and humidified with 5% CO$_2$. Images were acquired by a confocal laser-scanning microscope (Zeiss LSM 880) combined with a Plan-Apochromat 63x/1.4 Oil DIC objective. Excitation of EGFP was achieved by using a 488 nm argon laser and of mCherry by a 543 nm helium-neon laser. To exclude fluorescence cross-

talk, sequential track imaging was used. For photoactivation of $^{PA}Nb^{mCherry}$, a bleaching function with 250 iterations, 50 cycles using a 405 nm diode (4.5 mW/μm^2) was applied. The whole region of interest was illuminated.

In-Gel Fluorescence and Immunoblotting. Synthesis of $^{UM/PA}Nb^{mCherry}$ was analyzed by in-gel fluorescence. To visualize the expression in adherent cells, HeLa cells were transiently transfected in 6-well plates. 5×10^5 cells were seeded in a 6-well plate. After 24 h, transient transfection using Lipofectamine 2000 was performed to initialize amber suppression. Cells were harvested on the following day by using PBS/0.05% trypsin–EDTA. The samples were pelleted (800 g, 4 °C) and washed once with DMEM to remove trypsin and once in PBS. For suspension HEK293-F cells, two PBS washing steps were applied. The obtained pellets were lysed in RIPA buffer containing 1% (v/v) benzonase and 1% (v/v) protease-inhibitor mix HP. The pellet of one 6-well plate was incubated in 30 μL of lysis buffer for 1 h (end-over-end rotation, RT). For HEK293-F cells, 100 μL lysis buffer was used. Subsequently, lysates were centrifuged (21,000 g, 30 min, 4 °C) and the supernatant collected. The obtained lysate protein concentration was determined by the Bradford assay (Pierce Detergent Compatible Bradford Assay Kit). To analyze the size of the synthesized protein by in-gel fluorescence, lysates were separated by discontinuous Tris/glycine SDS-PAGE (12%). Samples were prepared using a fivefold loading buffer concentrate with 30% (v/v) glycerol, 10% (w/v) SDS, 0.02% (w/v) Bromophenol Blue, and 250 mM Tris/HCl pH 6.8. After electrophoresis, the gel was washed three times with ddH$_2$O. In-gel fluorescence was analyzed using a gel and blot imaging system (Fusion FX imaging system, Vilber). EGFP was excited with 480 nm and detected using a 535 nm narrow band pass filter. For the detection of mCherry, a 530 nm and a 585 nm narrow band pass filter was used. As a loading control for in-gel fluorescence, an anti-GAPDH immunoblotting was performed. After imaging the samples, gels were semi-dry blotted on nitrocellulose membranes. The membranes were blocked in TBS-T supplemented with a 5% (w/v) skim milk powder (1 h, RT). Subsequently, membranes were washed three times in TBS-T. For immediate GAPDH detection, Direct-Blot HRP anti-GAPDH antibody was diluted 1:2000 in blocking buffer. Antibodies and membranes were incubated overnight at 4 °C. Before analysis, three washing steps with TBS-T were applied to remove unbound antibodies. For the chemiluminescent detection of the HRP conjugate, an ECL solution (Clarity Western ECL Substrate, Bio-Rad) was used in a gel and blot imaging system (Fusion FX imaging system, Vilber).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.1c00471.

Experimental methods; experimental data; experimental approaches; gating strategies for flow cytometry quantification of $^{UM/PA}Nb^{mCherry}$ synthesis using the various constructs (I–V); in-gel fluorescence analysis highlighting translational read-through across amber mutation close to N-terminus; flow cytometry analysis of $^{UM/PA}Nb^{mCherry}$ synthesis *via* the EF1 promotor (construct IV); flow cytometry data of large-scale $^{UM/PA}Nb^{mCherry}$ synthesis by the transient transfection

of solubilized mammalian cells using p2A-Amber; live cell CLSM controls showing constitutive $^{UM}Nb^{mCherry}$ target binding before and after light exposure; live-cell CLSM photoactivation of $^{PA}Nb^{mCherry}$ *via* NPY; NPY-Me-mediated $^{PA}Nb^{mCherry}$ synthesis monitored by flow cytometry; live-cell CLSM photoactivation of NPY-Me-synthesized $^{PA}Nb^{mCherry}$; and mass spectrometric analysis of ONBY, NPY, and NPY-Me compounds (PDF)

AUTHOR INFORMATION

Corresponding Author

Robert Tampé − *Institute of Biochemistry, Biocenter, Goethe University Frankfurt, 60438 Frankfurt/M, Germany;* orcid.org/0000-0002-0403-2160; Email: tampe@em.uni-frankfurt.de

Authors

Eike F. Joest − *Institute of Biochemistry, Biocenter, Goethe University Frankfurt, 60438 Frankfurt/M, Germany*

Christian Winter − *Institute of Biochemistry, Biocenter, Goethe University Frankfurt, 60438 Frankfurt/M, Germany*

Joshua S. Wesalo − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*

Alexander Deiters − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acssynbio.1c00471

Author Contributions

E.J. cloned the various constructs, performed amber suppression experiments, and imaged the cells by CLSM. FACS analysis was done by C.W. and E.J. C.W. performed mass spectrometric analyses of the various compounds. E.J., C.W., and R.T. carried out the data analysis. Photocaged amino acids were generated by J.S.W. and the synthesis was supervised by A.D. E.J. and R.T. wrote the manuscript with contributions from all authors. R.T. conceived and supervised the project.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An expanded eukaryotic genetic code. *Science* **2003**, *301*, 964−967.

(2) Xie, J.; Schultz, P. G. An expanding genetic code. *Methods* **2005**, *36*, 227−238.

(3) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **2001**, *292*, 498−500.

(4) Hancock, S. M.; Uprety, R.; Deiters, A.; Chin, J. W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino

acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.* **2010**, *132*, 14819−14824.

(5) Lang, K.; Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **2014**, *114*, 4764−4806.

(6) Zhou, W.; Wesalo, J. S.; Liu, J.; Deiters, A. Genetic code expansion in mammalian cells: a plasmid system comparison. *Bioorg. Med. Chem.* **2020**, *28*, 115772.

(7) Brown, W.; Liu, J.; Deiters, A. Genetic code expansion in animals. *ACS Chem. Biol.* **2018**, *13*, 2375−2386.

(8) Wang, Q.; Parrish, A. R.; Wang, L. Expanding the genetic code for biological studies. *Chem. Biol.* **2009**, *16*, 323−336.

(9) de la Torre, D.; Chin, J. W. Reprogramming the genetic code. *Nat. Rev. Genet.* **2021**, *22*, 169−184.

(10) Shandell, M. A.; Tan, Z.; Cornish, V. W. Genetic code expansion: A brief history and perspective. *Biochemistry* **2021**, *60*, 3455−3469.

(11) Wang, Y.-S.; Fang, X.; Wallace, A. L.; Wu, B.; Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **2012**, *134*, 2950−2953.

(12) Elsässer, S. J.; Ernst, R. J.; Walker, O. S.; Chin, J. W. Genetic code expansion in stable cell lines enables encoded chromatin modification. *Nat. Methods* **2016**, *13*, 158−164.

(13) Uttamapinant, C.; Howe, J. D.; Lang, K.; Beránek, V.; Davis, L.; Mahesh, M.; Barry, N. P.; Chin, J. W. Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. *J. Am. Chem. Soc.* **2015**, *137*, 4602−4605.

(14) Gubbens, J.; Kim, S. J.; Yang, Z.; Johnson, A. E.; Skach, W. R. In vitro incorporation of nonnatural amino acids into protein using tRNACys-derived opal, ochre, and amber suppressor tRNAs. *RNA* **2010**, *16*, 1660−1672.

(15) Liu, W.; Brock, A.; Chen, S.; Chen, S.; Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **2007**, *4*, 239−244.

(16) Xie, J.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **2004**, *22*, 1297−1301.

(17) Hayashi, A.; Hino, N.; Kobayashi, T.; Arai, R.; Shirouzu, M.; Yokoyama, S.; Sakamoto, K. Dissecting cell signaling pathways with genetically encoded 3-iodo-L-tyrosine. *ChemBioChem* **2011**, *12*, 387−389.

(18) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **2004**, *126*, 14306−14307.

(19) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chem.* **2009**, *121*, 4112−4115.

(20) Tsai, Y.-H.; Essig, S.; James, J. R.; Lang, K.; Chin, J. W. Selective, rapid and optically switchable regulation of protein function in live mammalian cells. *Nat. Chem.* **2015**, *7*, 554−561.

(21) Chou, C.; Young, D. D.; Deiters, A. A light-activated DNA polymerase. *Angew. Chem.* **2009**, *121*, 6064−6067.

(22) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chem., Int. Ed.* **2006**, *45*, 2728−2731.

(23) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **2007**, *3*, 769−772.

(24) Peters, F. B.; Brock, A.; Wang, J.; Schultz, P. G. Photocleavage of the polypeptide backbone by 2-nitrophenylalanine. *Chem. Biol.* **2009**, *16*, 148−152.

(25) Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **2012**, *134*, 11912−11915.

(26) Luo, J.; Torres-Kolbus, J.; Liu, J.; Deiters, A. Genetic encoding of photocaged tyrosines with improved light-activation properties for the optical control of protease function. *ChemBioChem* **2017**, *18*, 1442−1447.

(27) Jedlitzke, B.; Yilmaz, Z.; Dörner, W.; Mootz, H. D. Photobodies: light-activatable single-domain antibody fragments. *Angew. Chem., Int. Ed.* **2020**, *59*, 1506−1510.

(28) Jedlitzke, B.; Mootz, H. D. Photocaged nanobodies delivered into cells for light activation of biological processes. *ChemPhotoChem* **2021**, *5*, 22−25.

(29) Brunnberg, J.; Herbring, V.; Castillo, E. G.; Krüger, H.; Wieneke, R.; Tampé, R. Light control of the peptide-loading complex synchronizes antigen translocation and MHC I trafficking. *Commun. Biol.* **2021**, *4*, 430.

(30) Fleissner, M. R.; Brustad, E. M.; Kalai, T.; Altenbach, C.; Cascio, D.; Peters, F. B.; Hideg, K.; Peuker, S.; Schultz, P. G.; Hubbell, W. L. Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21637−21642.

(31) Zeng, H.; Xie, J.; Schultz, P. G. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5356−5359.

(32) Saal, K.-A.; Richter, F.; Rehling, P.; Rizzoli, S. O. Combined use of unnatural amino acids enables dual-color super-resolution imaging of proteins via click chemistry. *ACS Nano* **2018**, *12*, 12247−12254.

(33) Heil, C. S.; Rittner, A.; Goebel, B.; Beyer, D.; Grininger, M. Site-specific labelling of multidomain proteins by amber codon suppression. *Sci. Rep.* **2018**, *8*, 14864.

(34) Aloush, N.; Schvartz, T.; König, A. I.; Cohen, S.; Brozgol, E.; Tam, B.; Nachmias, D.; Ben-David, O.; Garini, Y.; Elia, N.; Arbely, E. Live cell imaging of bioorthogonally labelled proteins generated with a single pyrrolysine tRNA gene. *Sci. Rep.* **2018**, *8*, 14527.

(35) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020−11024.

(36) Grunbeck, A.; Huber, T.; Abrol, R.; Trzaskowski, B.; Goddard, W. A., III; Sakmar, T. P. Genetically encoded photo-cross-linkers map the binding site of an allosteric drug on a G protein-coupled receptor. *ACS Chem. Biol.* **2012**, *7*, 967−972.

(37) Grunbeck, A.; Huber, T.; Sachdev, P.; Sakmar, T. P. Mapping the ligand-binding site on a G protein-coupled receptor (GPCR) using genetically encoded photocrosslinkers. *Biochemistry* **2011**, *50*, 3411−3413.

(38) Ye, S.; Zaitseva, E.; Caltabiano, G.; Schertler, G. F. X.; Sakmar, T. P.; Deupi, X.; Vogel, R. Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **2010**, *464*, 1386−1389.

(39) Naganathan, S.; Ye, S.; Sakmar, T. P.; Huber, T. Site-specific epitope tagging of G protein-coupled receptors by bioorthogonal modification of a genetically encoded unnatural amino acid. *Biochemistry* **2013**, *52*, 1028−1036.

(40) Huber, T.; Naganathan, S.; Tian, H.; Ye, S.; Sakmar, T. P. Unnatural amino acid mutagenesis of GPCRs using amber codon suppression and bioorthogonal labeling. *Methods Enzymol.* **2013**, *520*, 281−305.

(41) Kang, H. J.; Baker, E. N. Intramolecular isopeptide bonds: protein crosslinks built for stress? *Trends Biochem. Sci.* **2011**, *36*, 229−237.

(42) Wang, W.; Li, H.; Wang, J.; Zu, L. Enantioselective organocatalytic tandem Michael−Aldol reactions: One-pot synthesis of chiral thiochromenes. *J. Am. Chem. Soc.* **2006**, *128*, 10354−10355.

(43) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9785−9789.

(44) Charbon, G.; Wang, J.; Brustad, E.; Schultz, P. G.; Horwich, A. L.; Jacobs-Wagner, C.; Chapman, E. Localization of GroEL determined by in vivo incorporation of a fluorescent amino acid. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 6067−6070.

(45) Charbon, G.; Brustad, E.; Scott, K. A.; Wang, J.; Løbner-Olesen, A.; Schultz, P. G.; Jacobs-Wagner, C.; Chapman, E. Subcellular protein localization by using a genetically encoded fluorescent amino acid. *ChemBioChem* 2011, 12, 1818−1821.

(46) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. A genetically encoded fluorescent probe in mammalian cells. *J. Am. Chem. Soc.* 2013, 135, 12540−12543.

(47) Hohl, A.; Mideksa, Y. G.; Karan, R.; Akal, A.; Vogler, M.; Groll, M.; Rueping, M.; Lang, K.; Feige, M. J.; Eppinger, J. Genetically encoded biotin analogues: Incorporation and application in bacterial and mammalian cells. *ChemBioChem* 2019, 20, 1795−1798.

(48) Xie, J.; Supekova, L.; Schultz, P. G. A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* 2007, 2, 474−478.

(49) Drabkin, H. J.; Park, H. J.; RajBhandary, U. L. Amber suppression in mammalian cells dependent upon expression of an Escherichia coli aminoacyl-tRNA synthetase gene. *Mol. Cell. Biol.* 1996, 16, 907−913.

(50) Serfling, R.; Lorenz, C.; Etzel, M.; Schicht, G.; Böttke, T.; Mörl, M.; Coin, I. Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. *Nucleic Acids Res.* 2018, 46, 1−10.

(51) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* 2012, 4, 298−304.

(52) Mitra, N. Incorporating unnatural amino acids into recombinant proteins in living cells. *Mater. Methods* 2013, 3, 1.

(53) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* 2010, 464, 441−444.

(54) Fredens, J.; Wang, K.; de la Torre, D.; Funke, L. F. H.; Robertson, W. E.; Christova, Y.; Chia, T.; Schmied, W. H.; Dunkelmann, D. L.; Beránek, V.; Uttamapinant, C.; Llamazares, A. G.; Elliott, T. S.; Chin, J. W. Total synthesis of *Escherichia coli* with a recoded genome. *Nature* 2019, 569, 514−518.

(55) Reinkemeier, C. D.; Girona, G. E.; Lemke, E. A. Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* 2019, 363, No. eaaw2644.

(56) Reinkemeier, C. D.; Lemke, E. A. Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation. *Cell* 2021, 184, 4886−4903.

(57) Fottner, M.; Brunner, A.-D.; Bittl, V.; Horn-Ghetko, D.; Jussupow, A.; Kaila, V. R. I.; Bremm, A.; Lang, K. Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase. *Nat. Chem. Biol.* 2019, 15, 276−284.

(58) Dana, A.; Tuller, T. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.* 2012, 8, No. e1002755.

(59) Sharma, A. K.; Sormanni, P.; Ahmed, N.; Ciryam, P.; Friedrich, U. A.; Kramer, G.; O'Brien, E. P. A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLoS Comput. Biol.* 2019, 15, No. e1007070.

(60) Riba, A.; Di Nanni, N.; Mittal, N.; Arhné, E.; Schmidt, A.; Zavolan, M. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. U.S.A.* 2019, 116, 15023−15032.

(61) Gobet, C.; Naef, F. Ribosome profiling and dynamic regulation of translation in mammals. *Curr. Opin. Genet. Dev.* 2017, 43, 120−127.

(62) Joest, E. F.; Winter, C.; Wesalo, J. S.; Deiters, A.; Tampé, R. Light-guided intrabodies for on-demand in situ target recognition in human cells. *Chem. Sci.* 2021, 12, 5787−5795.

(63) 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. Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* 2010, 17, 133−138.

(64) Kubala, M. H.; Kovtun, O.; Alexandrov, K.; Collins, B. M. Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. *Protein Sci.* 2010, 19, 2389−2401.

(65) Schmied, W. H.; Elsässer, S. J.; Uttamapinant, C.; Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* 2014, 136, 15577−15583.

(66) Aloush, N.; Schvartz, T.; König, A. I.; Cohen, S.; Brozgol, E.; Tam, B.; Nachmias, D.; Ben-David, O.; Garini, Y.; Elia, N.; Arbely, E. Live cell imaging of bioorthogonally labelled proteins generated with a single pyrrolysine tRNA gene. *Sci. Rep.* 2018, 8, 14527.

(67) Zhou, W.; Wesalo, J. S.; Liu, J.; Deiters, A. Genetic code expansion in mammalian cells: A plasmid system comparison. *Bioorg. Med. Chem.* 2020, 28, 115772.

(68) Blees, A.; Januliene, D.; Hofmann, T.; Koller, N.; Schmidt, C.; Trowitzsch, S.; Moeller, A.; Tampé, R. Structure of the human MHC-I peptide-loading complex. *Nature* 2017, 551, 525−528.

(69) Fischbach, H.; Döring, M.; Nikles, D.; Lehnert, E.; Baldauf, C.; Kalinke, U.; Tampé, R. Ultrasensitive quantification of TAP-dependent antigen compartmentalization in scarce primary immune cell subsets. *Nat. Commun.* 2015, 6, 6199.

(70) Donnelly, M. L. L.; Luke, G.; Mehrotra, A.; Li, X.; Hughes, L. E.; Gani, D.; Ryan, M. D. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip. *J. Gen. Virol.* 2001, 82, 1013−1025.

(71) Karginov, T. A.; Pastor, D. P. H.; Semler, B. L.; Gomez, C. M. Mammalian polycistronic mRNAs and disease. *Trends Genet.* 2017, 33, 129−142.

(72) Takimoto, J. K.; Xiang, Z.; Kang, J.-Y.; Wang, L. Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *ChemBioChem* 2010, 11, 2268−2272.

(73) Zhou, H.; Cheung, J. W.; Carpenter, T.; Jones, S. K., Jr.; Luong, N. H.; Tran, N. C.; Jacobs, S. E.; Galbada Liyanage, S. A.; Cropp, T. A.; Yin, J. Enhancing the incorporation of lysine derivatives into proteins with methylester forms of unnatural amino acids. *Bioorg. Med. Chem. Lett.* 2020, 30, 126876.

(74) Chan, L. Y.; Mugler, C. F.; Heinrich, S.; Vallotton, P.; Weis, K. Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* 2018, 7, No. e32536.

(75) Segal, I.; Nachmias, D.; Konig, A.; Alon, A.; Arbely, E.; Elia, N. A straightforward approach for bioorthogonal labeling of proteins and organelles in live mammalian cells, using a short peptide tag. *BMC Biol.* 2020, 18, 5.

(76) Lafranchi, L.; Schlesinger, D.; Kimler, K. J.; Elsässer, S. J. Universal Single-Residue Terminal Labels for Fluorescent Live Cell Imaging of Microproteins. *J. Am. Chem. Soc.* 2020, 142, 20080−20087.

(77) Wesalo, J. S.; Luo, J.; Morihiro, K.; Liu, J.; Deiters, A. Phosphine-activated lLysine analogues for fast chemical control of protein subcellular localization and protein SUMOylation. *ChemBioChem* 2020, 21, 141−148.

(78) Lerch, M. M.; Hansen, M. J.; van Dam, G. M.; Szymanski, W.; Feringa, B. L. Emerging targets in photopharmacology. *Angew. Chem., Int. Ed.* 2016, 55, 10978−10999.

(79) Hüll, K.; Morstein, J.; Trauner, D. In vivo photopharmacology. *Chem. Rev.* 2018, 118, 10710−10747.

(80) VanBrunt, M. P.; Shanebeck, K.; Caldwell, Z.; Johnson, J.; Thompson, P.; Martin, T.; Dong, H.; Li, G.; Xu, H.; D'Hooge, F.; Masterson, L.; Bariola, P.; Tiberghien, A.; Ezeadi, E.; Williams, D. G.; Hartley, J. A.; Howard, P. W.; Grabstein, K. H.; Bowen, M. A.; Marelli, M. Genetically encoded azide containing amino acid in mammalian cells enables site-specific antibody-drug conjugates using Click cycloaddition chemistry. *Bioconjugate Chem.* 2015, 26, 2249−2260.

(81) Roy, G.; Reier, J.; Garcia, A.; Martin, T.; Rice, M.; Wang, J.; Prophet, M.; Christie, R.; Dall'Acqua, W.; Ahuja, S.; Bowen, M. A.;

Marelli, M. Development of a high yielding expression platform for the introduction of non-natural amino acids in protein sequences. *MAbs* **2020**, *12*, 1684749.

(82) Klein, A.; Hank, S.; Raulf, A.; Joest, E. F.; Tissen, F.; Heilemann, M.; Wieneke, R.; Tampé, R. Live-cell labeling of endogenous proteins with nanometer precision by transduced nanobodies. *Chem. Sci.* **2018**, *9*, 7835−7842.

(83) Tietze, L. F.; Müller, M.; Duefert, S.-C.; Schmuck, K.; Schuberth, I. Photoactivatable prodrugs of highly potent duocarmycin analogues for a selective cancer therapy. *Chemistry* **2013**, *19*, 1726−1731.

(84) Uphoff, C. C.; Drexler, H. G. Detection of mycoplasma contamination in cell cultures. *Curr. Protoc. Mol. Biol.* **2014**, *106*, 28.4.1−28.4.14.

📖 **Recommended by ACS**

Optimized Loopable Translation as a Platform for the Synthesis of Repetitive Proteins

Sea On Lee, Stephen D. Fried, *et al.*

SEPTEMBER 24, 2021
ACS CENTRAL SCIENCE

READ ☑

Generation of Recombinant Mammalian Selenoproteins through Genetic Code Expansion with Photocaged Selenocysteine

Jennifer C. Peeler, Eranthie Weerapana, *et al.*

APRIL 24, 2020
ACS CHEMICAL BIOLOGY

READ ☑

Genetic Code Expansion in the Engineered Organism Vmax X2: High Yield and Exceptional Fidelity

Sebasthian Santiago González, Alanna Schepartz, *et al.*

AUGUST 31, 2021
ACS CENTRAL SCIENCE

READ ☑

Mutation Maker, An Open Source Oligo Design Platform for Protein Engineering

Kaori Hiraga, Danny A. Bitton, *et al.*

JANUARY 12, 2021
ACS SYNTHETIC BIOLOGY

READ ☑

Get More Suggestions >

Supporting information

Efficient Amber Suppression *via* Ribosomal Skipping for *In Situ* Synthesis of Photoconditional Nanobodies

Eike F. Joest[1], Christian Winter[1], Joshua S. Wesalo[2], Alexander Deiters[2], Robert Tampé[1],*

[1]Goethe University Frankfurt, Institute of Biochemistry, Biocenter, Max-von-Laue-Str. 9, 60438 Frankfurt/M, Germany; [2]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA

*To whom correspondence may be addressed: Email: tampe@em.uni-frankfurt.de

Figure S1. Analysis of amber suppression in mammalian cells by flow cytometry. Protein synthesis is monitored via C-terminal mCherry. **a**, Scheme illustrating experimental approach. Cells are transiently transfected with POI-encoding plasmids I–IV. At the position of a premature amber stop codon (Y37[TAG]), non-canonical amino acids can be incorporated by amber suppression. To accomplish this, cells are co-transfected with a plasmid encoding an orthogonal aaRS/4xtRNA pair and the medium is supplemented with non-canonical amino acid (ONBY). **b**, Gating strategy for analysis of plasmids I–IV. After transient transfection, no amber suppression was reported by a significant increase in mCherry-positive cells. Living HeLa cells 24 h after transfection, supplemented with 0.25 mM ONBY, respectively.
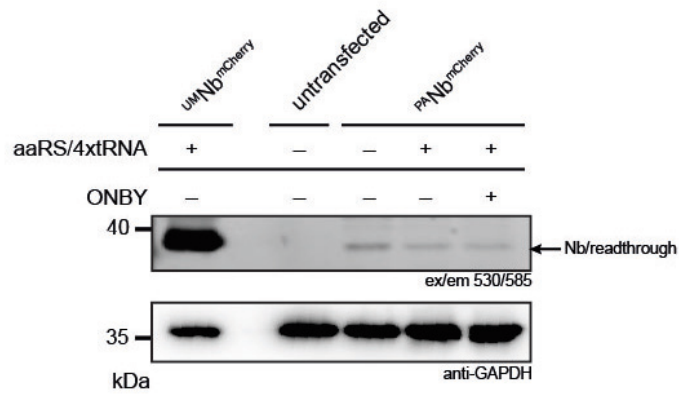
Figure S2. In-gel fluorescence revealing amber codon read-through. Lysates of HeLa cells transiently transfected with ^{UM/PA}Nb^{mCherry} encoding plasmid under control of a CMV promotor (I). ONBY-dependent amber suppression was not observed. 10% SDS-PAGE, 20 µg total protein of cell lysates.
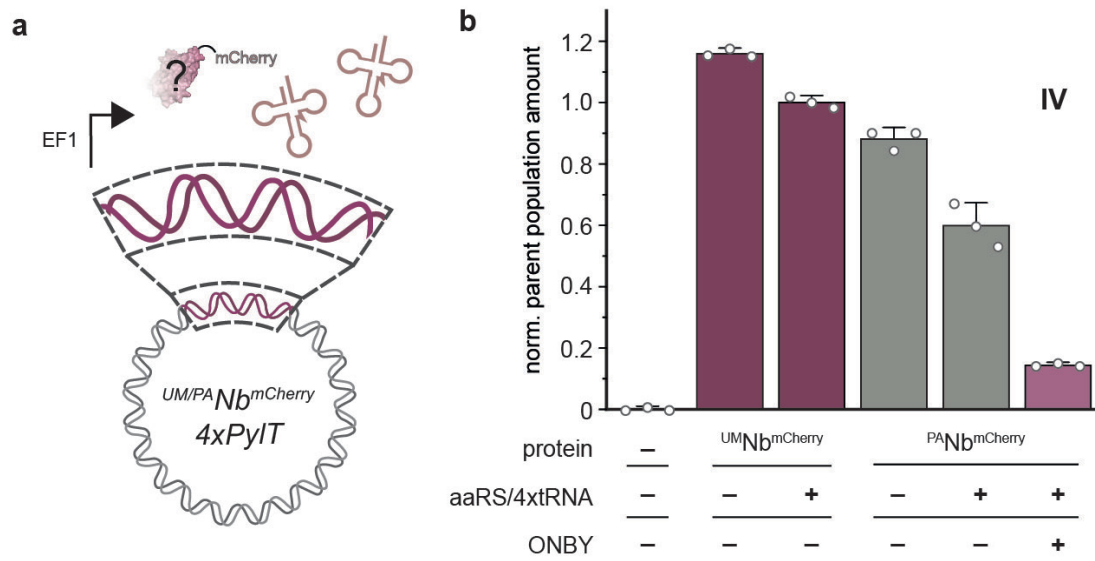
Figure S3. Amounts of Nb^mCherry-positive cells after transient transfection of expression cassette under control of the moderate EF1-alpha promotor (**a**, IV) monitored by flow cytometry (**b**). For optimal amber suppression, the plasmid was equipped with four additional tRNA copies (4xPylT). The premature amber codon at position Y37 did not terminate translation. Background expression limited sufficient amber suppression for synthesis of ^PANb^mCherry. aaRS/4xtRNA = NPYRS/4xPylT co-transfection, ONBY = medium supplemented with 0.25 mM ONBY. Living HeLa cells 24 h after transfection, mean amounts of *n* = 3 biological replicates ± SD.
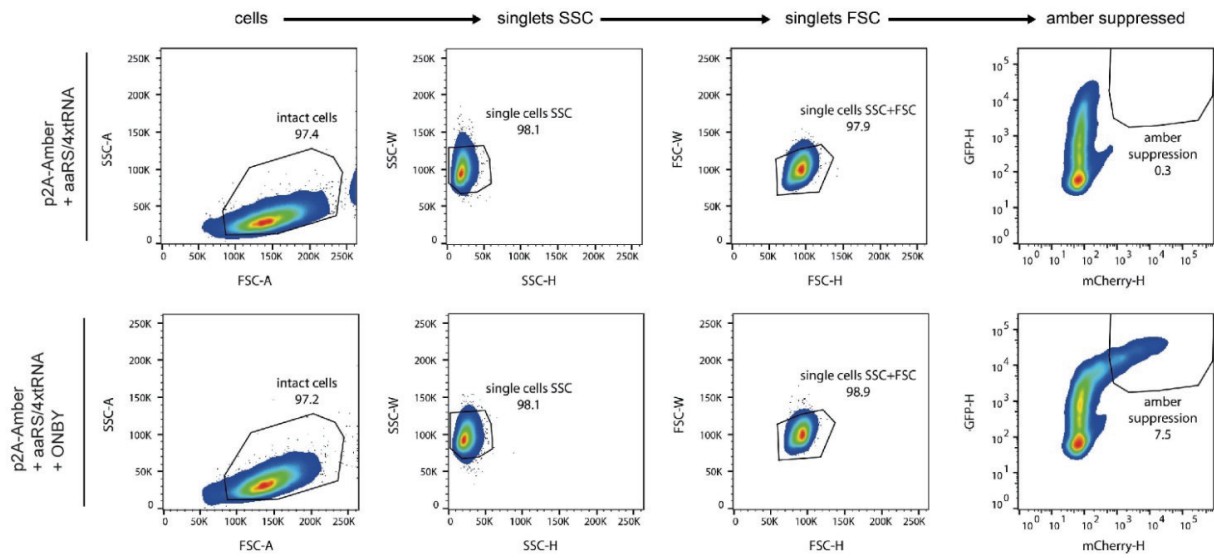
Figure S4. Gating strategy for analysis of cells transiently transfected with p2A-Amber and NPYRS/4xtRNA. After transient transfection, a strong increase in mCherry-positive cells reported efficient amber suppression for generation of $^{PA}Nb^{mCherry}$. As a reference, p2A-Amber encoding TAP1^{EGFP} was used. HeLa cells 24 h after transfection, grown in the absence (top) or presence (bottom) of 0.25 mM ONBY.
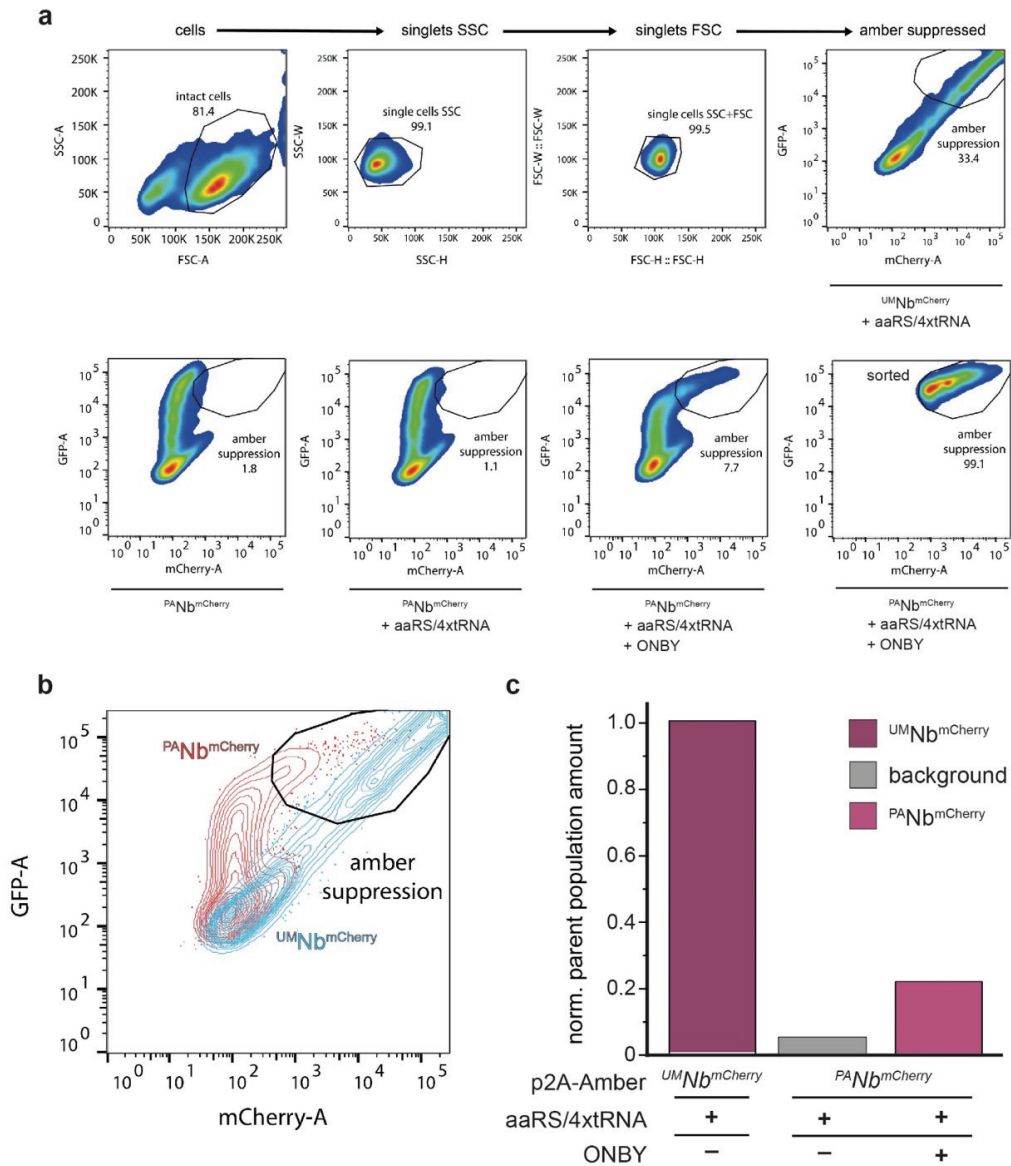
Figure S5. Large-scale synthesis of modified protein by transient transfection of mammalian cells in suspension culture with p2A-Amber. **a**, Gating strategy for flow cytometry analysis and sorting of transiently transfected cells. After transient transfection, a strong increase in mCherry-positive cells reported efficient amber suppression for generation of ^{PA}Nb^{mCherry}. **b**, The ribosome skipping site of p2A-Amber allows the bicistronic (parallel) synthesis of POI ($^{UM/PA}$Nb^{mCherry}) and reference protein (TAP1^{EGFP}). Gene expression was monitored *via* the fluorescent protein tags using flow cytometry. **c**, Number of cells showing amber suppression relative to the construct without premature amber codon (^{UM}Nb^{mCherry}). The p2A-Amber design restored the premature amber codon at position Y37 for efficient amber suppression. In absence of suppression components, the premature amber codon initiated the end of translation. **a–c**, As a reference gene, p2A-Amber encoding TAP1^{EGFP} was used. Living HEK-F cells were analyzed 24 h after transfection, supplemented with 0.25 mM ONBY.
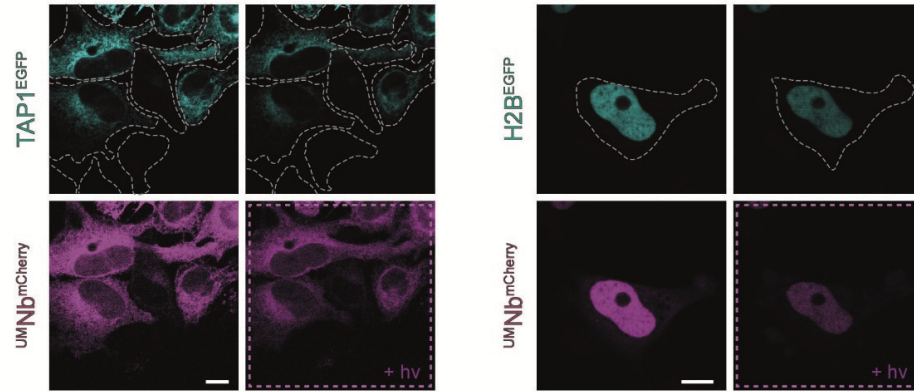
Figure S6. Constitutive binding of wildtype intrabody (^{UM}Nb^{mCherry}). Cells transiently transfected with p2A-Amber encoding the unmodified construct and EGFP-tagged reference protein (TAP1 or H2B). Intrabody binding of the cognate target (EGFP) is corroborated by colocalization in live-cell CLSM. For increased comparability, cells were co-transfected with a plasmid encoding the orthogonal aaRS/tRNA pair (aaRS/4xtRNA). After light exposure, no intracellular reorganization of ^{UM}Nb^{mCherry} was observed. HeLa cells 24 h after transfection, scale bar = 10 μm.
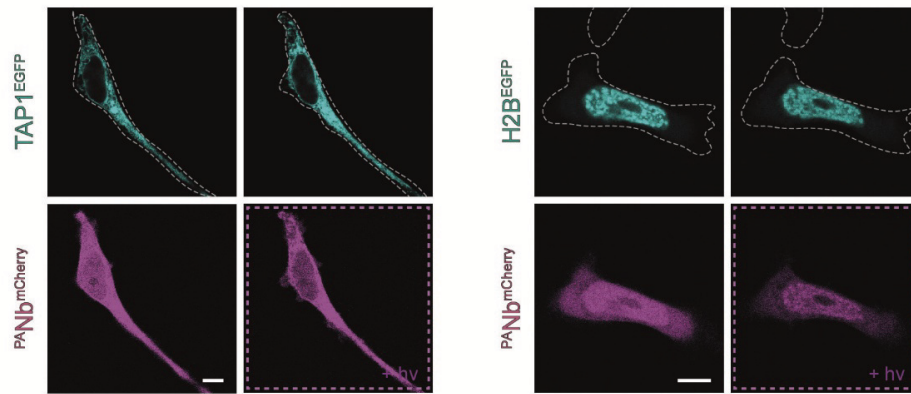
Figure S7. *In situ* photo-activation of ^{PA}Nb^{mCherry} synthesized with NPY. After photo-deprotection, colocalization with ER resident reference protein (TAP1^{EGFP}) or reference protein located inside the nucleus (H2B^{EGFP}) demonstrated intrabody binding. Mammalian cells were transiently transfected with p2A-Amber encoding ^{PA}Nb^{mCherry} and the respective reference protein. For suppression of the amber codon at position Y37, cells were co-transfected with a plasmid encoding an orthogonal aaRS/tRNA pair (NPYRS/4xPylT) and the medium was supplemented with photo-caged tyrosine (0.25 mM NPY). Live-cell CLSM of HeLa cells 24 h after transfection, scale bar = 10 µm.
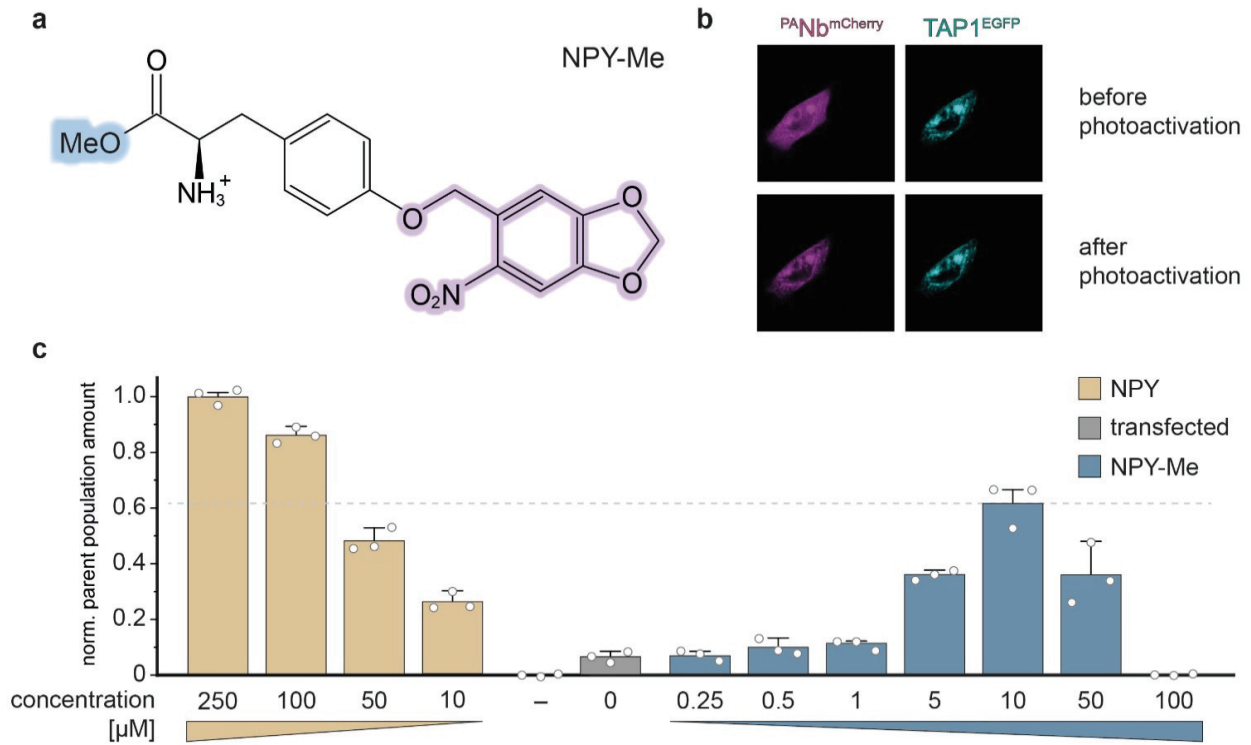
Figure S8. Combination of p2A-Amber and NPY-Me for synthesis of modified protein. **a**, Methyl ester of NPY with increased solubility and cell penetration for improved amber suppression. **b**, Synthesis of PANbmCherry by mammalian cells supplemented with minimal amounts of NPY-Me (10 µM). Immediately after *in situ* photoactivation, intrabody binding of the cognate target protein was corroborated by colocalization. Live-cell CLSM imaging of HeLa cells 32 h after transfection, scale bar = 10 µm. **c**, Amber suppression with various amounts of NPY or NPY-Me. Intrabody synthesis was monitored via C-terminal mCherry in flow cytometry. Relative to cells treated with NPY at the limit of solubility (250 µM), ~60% PANbmCherry-positive cells were obtained using lower amounts of the methyl ester (NPY-Me, 10 µM). *n* = 3 biological replicates of HeLa cells. **b-c**, Cells were transiently co-transfected with p2A-Amber encoding TAP1EGFP reference protein and a plasmid encoding the orthogonal aaRS/tRNA pair (NPYRS/4xPylT).
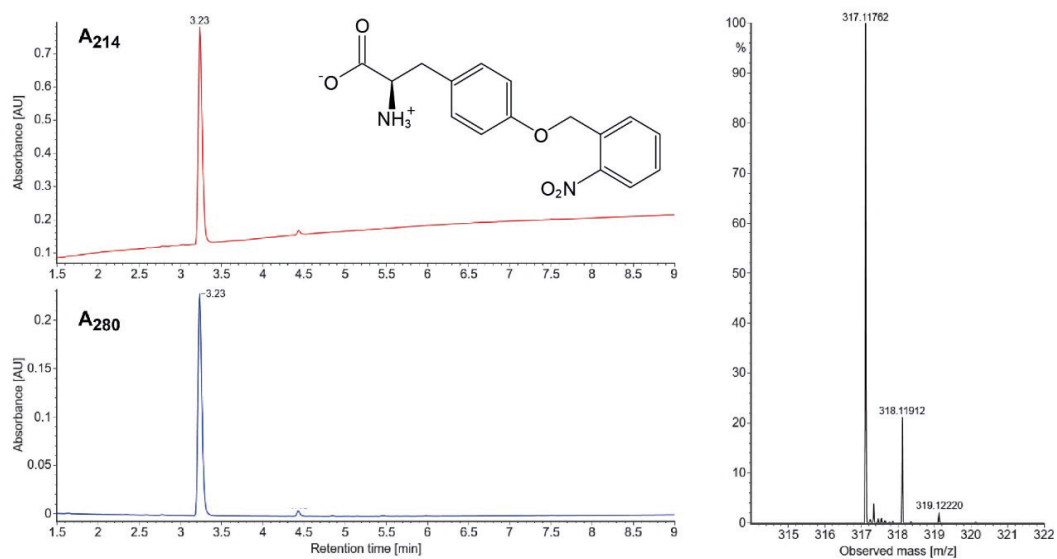
Figure S9. Mass spectrometric analysis of ONBY. The stock of photocaged tyrosine was analyzed by LC-MS. Identity of ONBY ($[M^{+}+H^{+}]_{calc}$: 317.1132 Da, $[M^{+}+H^{+}]_{obs}$: 317.1176 Da) was determined using a cone voltage of 30 V and capillary voltage of 0.8 kV in positive polarity.
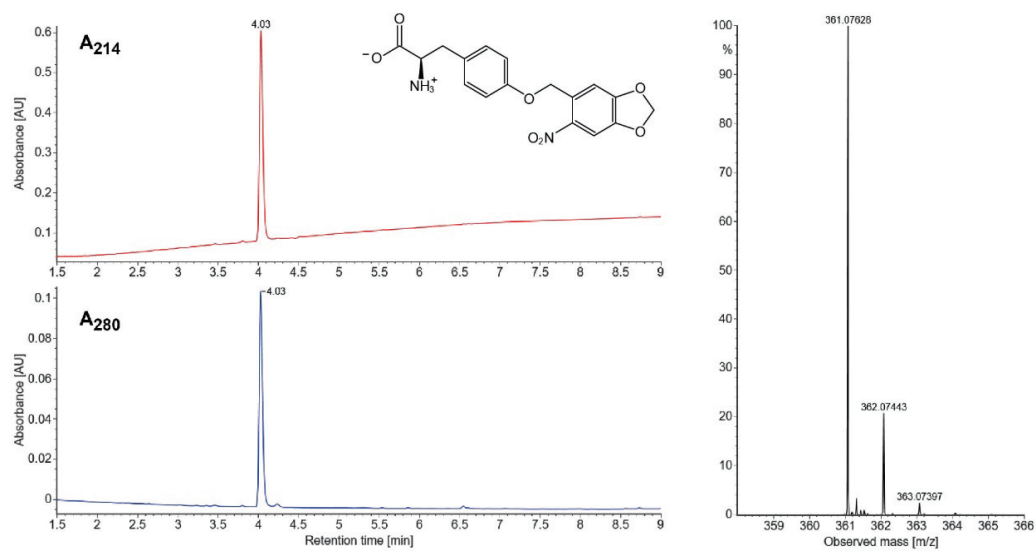
Figure S10. Mass spectrometric analysis of NPY. The stock of photocaged tyrosine was analyzed by LC-MS. Identity of NPY ($[M^++H^+]_{calc}$: 361.1030 Da, $[M^++H^+]_{obs}$: 361.0762 Da) was determined using a cone voltage of 30 V and capillary voltage of 0.8 kV in positive polarity.

Figure S11. Mass analysis of NPY-Me. The stock of photocaged tyrosine ester was analyzed by LC-MS. Identity of NPY-me ([M$^+$+H$^+$]$_{calc}$: 375.1187 Da, [M$^+$+H$^+$]$_{obs}$: 375.1181 Da) was determined using a cone voltage of 30 V and capillary voltage of 0.8 kV in positive polarity.

Plasmid sequence p2A-Amber (TAP1^EGFP | F2A | ^Y37TAG^Nb^mCherry^)

```
GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGA
GCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTAT
TGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG
TACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTG
ATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAAT
TAATACGACTCACTATAGGGAGACCCAAGCTGgctagcgtttaaacttaagcttagcgcagaggcttggggcagccgagcggcagccaggccccggcccgggcctcggttcagaagggagaggagcccgccaaggcgcgcaagag
agcgggctgcctcgcagtccgagccggagaggggagcgcgagccgcgccggccccggacggcctccgaaaccatgtcttccagatcccctgctcctagaggctccagatctctgcctggccgcttccctggcttggctgggcaccgtcctgctgctgctggc
tgactgggtcctgctgagaaccgctctgcctagaatcttctccctgctggtccctaccgccctgcctctgctgagagtctgggctgtcggcctgtccagatgggctgtcctgtggctgggtgcttccggagtcctgagagctaccgtcggctccaagtccgaa
aacgctggcgctcagggctggctggctgctctgaagccctcbggccgctgctctggggctggcctgttcagagaactgatctcctggggcgctcccgggtccggctgactccaccagactgctgcactggggctcccatccaaccgc
tttcgtcgtctctacgctgctgccctgcctgctgccgccctgtggcacaagctgggttccctgtgggtcccaggcggtcagggcggctccggcaacctgtgagaagactgctgggctccctgggttccgaaaccagaagactgtccctgttcctggtcct
ggtcgtcctgtcctctctgggcgaaatggctatccctttcttcaccggcagactgaccgactggattctgcaggatggctccgctgacaccttcacgcgtaacttaaccctgatgtccatcctgaccatcgcttccgctgtcctggaattcgtcggcgacggc
atctacaacaacaccatgggccacgtccactcccacctgcagggcgaagtcttcggtgccgtcctgagacaggaaaccgaattcttccagcagaaccagaccggcaacatcatgtccagagtcaccgaagatacgtccaccctgtctgactccctgtcc
gaaaacctgtccctgttcctgtggtacctggtcaagggcctggccctgctgggcatcatgctgtggggctccgtctccctgactatgggtgacccgtcccctgatcaccctgccctctgctgttcctgctgcctaagaaggtcggcaagtggtaccagctgctggaagtc
caggtcagagaatccctggctaagtcctcccaggtcgctatcgaagccttaagtgctatgctctaccgtcagatccttcgctaacgaagaaggcgaagctcagaagttcagaaaaagtcggcaggaaatcaagaccctgaaccagaaggaagctgtcg
cttacgctgtcaactcctggaccacctccatctccggcatgctgctgaaagtcggtatcctgtacatcggcggccaactggtgacctccggcgctgtgagctctggcaacctggtcaccttcgtcctgtaccagatgcagttcacccaggccgtcgaagtcc
tgctgtccatctaccctagagtccagaaggctgtcggcctcctccgaaaagatcttcgaatacctggacagaaccctagatccccacctagtggcctgctgaccctctgcatctggaaggactggtccagttccaggacgtctccttcgcttaccctaaca
gacctgacgtgctggtcctgcagggtctgaccttcaccctgagacctggcgaagtcaccgcactggtcggccctaacggctccggcaagtcaccgtcgctgctctgctgcagaacctgtaccagcctaccggcggccagctgctgctggacggcaagc
ctctgcctcagtacgaacacagatacctgcacagcaggtcgctgctgtcggccaggaacctcaggtcttcggcaggatctctgcaggaaaacatcgcttacggcctacggcgtgacccagaagctaccattggaagaaataaccgccgccgctgtcaagtccggg
gcccactccttcatctccggcctgcctcagggctacgacaccgaagtagacgaagctggctctcagctgtccggcggccagacaggctgtcgctctggctagagccctgatcagaaagcctcctcgtcgatcctggacgacgctacctccgctctgg
acgctaactcccagctgcaagtcgaacaactgctgtacgaatcccctgaaagatactccagatccgtcctgctgatcacccacgcacctgtccctggtcgaacaggctgaccacatcctgttcctggaaggtggcgctatcagagaaggcggcacccacc
agcagctgatggaaaagaaggagctcctactgggctatggtccaggctcctgctgacgctcctgaaTTAGCGGCCGCGTTTACCGGTCCACAAGGATCCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCC
CATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG
GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGC
GATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGgcattagaagtttttgtttCAAGGTCCAGAATTCaccGAACAAAAACTCATCTCAGAAGa
ggatctgagggggggccagcatggacgagaagaccactggttggccgaggtggacacgttgttgaaggactggctgggggaacttgaacaacttcgtcgcacgactggagcatcacccacaaggtcaacgtgaaccagtgactgaactgctttaccgcatg
aagagggccgaaacatactgcccccggcctcttttggctattcacccaagcgaagctagacacaaacaaaagattgttgcgcctgtgaaacagctttgaactttgacctgctcaagttggcaggagacgtcgagtccaaCCCTGGGCCTggcCA
GGTTCAGCTGGTTGAAAGCGGTGGTGCACTGGTTCAGCCAGGTGGTAGCCTGCGTCTGAGCTGTGCAGCAAGCGGTTTCCGGTTAATCGTTATAGCATGCGTTGGTAGCGTCAGGCACCGG
GTAAAGAACGTGAATGGGTTGCAGGTATGAGCAGTGCCGGTGATCGTAGCAGCTATGAAGATAGCGTTAAAGGTCGTTTTACCATCAGCCGTGATGATGCACGTAATACCGTTTATCTGCAA
ATGAATAGCCTGAAACCGGAAGATACCGCAGTGTATTATTGCAATGTTAACGCTGGGCTTTGAATATTGGGGTCAGGGCACCCAGGTTACCGTTAGCAGCGGTACCgtgagcaagggcgaggaggat
aacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgccc
ttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgaccc
aggactcctccctgcaggacggcgagttcatctacaaggtgaagcgcgtgatgaacttcccctccgacggccccgtaatgcagaagaagaccatgggctggggaggcctccctccgagcggatgtaccccgaggacgccctgaaggtgcgagt
caagcagaggctgaagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgactacaacaagcaacatcaagttggacatcacctcccaacgaggactacacccatcgtggaac
agtacgaacgcgccgagggccgccactccaccggccatggacgagctgtacaagcccgcggccccggatcgcccggaaggcattgaagaacaccatcatcatcatcaccaccatcactaactcgagTCTAGAGGGCCCGTTTAAAC
CCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTG
CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTT
CTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCT
AGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTC
GGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTTATTTATGCAGAGGCCG
AGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGAT
GAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTG
TTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTG
CGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTG
ATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTG
GACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGG
AAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGC
TTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCT
GCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCA
CCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA
TAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG
AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAAT
CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCC
TTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG
CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
AAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGT
AGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
CACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGT
AAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACA
TGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCAT
GCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAG
CAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCT
TTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC
```

V.   Declaration of personal contributions

In accordance with common practice in this scientific field, all work included in this cumulative dissertation was conducted in close cooperation with other scientists. This section lists personal contributions to individual parts of each study. If a method or figure is not listed here, the corresponding research was conducted by one of the co-authors.

Semisynthetic viral inhibitor for light control of the MHC I peptide loading complex
C. Winter‡, A. Domnick‡, D. Cernova, R. Tampé
Angewandte Chemie **2022**, e202211826

Study	involved in		
	conception	execution	revision & proofreading
Initial draft	✓	✓	/
Manuscript writing	✓	✓	✓
Data availability	✓	✓	/

Figures	involved in	
	conception	design & execution
Figure 1 A-B, Figure 2 A-D, Figure 3 A-F, Figure 4 A-B	✓	✓
Figure S1 A-C, Figure S2 A-C, Figure S4 A-B, Figure S6 A-D, Figure S7	✓	✓
Figure S3 A-V	✓	
Figure S5		✓

Methods	involved in		
	conception	execution	analysis & interpretation
Synthesis of 3-cholor-2-hydroxypropanoic acid	✓	✓	✓
Synthesis of 2-hydroxy-3-(triphenylmethyl)thio-propanoic acid	✓	✓	✓
Synthesis of pc-ICP47$_{2\text{-}23}$-Hmp	✓	✓	✓
ULP1 SUMO protease preparation	✓	✓	✓
^{Cys}ICP47^{SBP} preparation	✓	✓	✓

ICP47SBP preparation	✓	✓	✓
ICP47AF647 preparation	✓	✓	✓
Native chemical ligation of pc-ICP47$_{2-23}$-Hmp and ^{Cys}ICP47SBP	✓	✓	✓
Desulfurization of pc-ICP47$^{23Cys-SBP}$	✓	✓	✓
LC-MS analysis	✓	✓	✓
pc-ICP47SBP cleavage by UV illumination	✓	✓	✓
Biochemical analysis of the PLC			✓
PLC pH tolerance test			✓
Analysis of PLC bound pc-ICP47SBP cleavage kinetics	✓	✓	✓
ICP47AF647 rebinding in the native PLC	✓		✓
Peptide synthesis, labeling, and purification	✓	✓	✓
Peptide binding to the PLC analyzed by fluorescence polarization	✓		
PLC ATP hydrolysis rate analyzed by radioactive phosphate release from [γ-^{32}P]ATP	✓	✓	✓
Microsome preparation	✓	✓	✓
Microsome size determination	✓	✓	✓
PLC topology in microsomes analyzed by proteinase K protection	✓	✓	✓
Flow cytometry with microsomes	✓	✓	✓
Peptide translocation assay in microsomes	✓	✓	✓
Immunostaining of microsomes for flow cytometry	✓	✓	✓

Molecular basis of MHC I quality control in the peptide loading complex
A. Domnick[‡], **C. Winter**[‡], L. Sušac, L. Hennecke, M. Hensen, N. Zitzmann,
S. Trowitzsch, C. Thomas, R. Tampé
Nature Communications **2022**, 13, 4701

Study	involved in		
	conception	execution	revision & proofreading
Initial draft	✓	✓	/
Manuscript writing	✓	✓	✓
Data availability	✓	✓	/

Figures

	involved in	
	conception	design & execution
Fig. 1 a-d, Fig. 3 a-c, Fig. 4 a-c	✓	✓
Supplementary Fig. 1, Supplementary Fig. 2, Supplementary Fig. 7 a-b	✓	✓
Fig. 5		✓

Methods

	involved in		
	conception	execution	analysis & interpretation
ICP47^{SBP} preparation	✓	✓	✓
Biochemical analysis of the PLC			✓
Peptide design, labeling, and purification	✓	✓	✓
Peptide binding and editing in the PLC	✓		
LC-MS analysis	✓	✓	✓
Deglycosylation by PNGase F	✓	✓	✓
Deglucosylation by GluII	✓	✓	✓

Structure of an MHC I–tapasin–ERp57 editing complex defines chaperone promiscuity

I. Müller, **C. Winter**, C. Thomas, R. Spaapen, S. Trowitzsch, R. Tampé

Nature Communications **2022**, 13, 5383

Study

	involved in		
	conception	execution	revision & proofreading
Initial draft			✓
Manuscript writing		✓	✓
Data availability		✓	✓

Figures

	involved in	
	conception	design & execution
Supplementary Fig. 1 a-d, Supplementary Fig. 2 a-e,	✓	✓

Methods

	involved in		
	conception	execution	analysis & interpretation
Peptide synthesis	✓		✓
EndoH digest	✓	✓	✓
LC-MS analysis	✓	✓	✓

Light-guided intrabodies for on-demand in situ target recognition in human cells

E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
Chemical Science **2021**, 12, 5787-5795

Study

	involved in		
	conception	execution	revision & proofreading
Initial draft			/
Manuscript writing		✓	✓
Cover design		✓	/

Figures

	involved in	
	conception	design & execution
Fig. 1 B		✓
Fig. 2 B	✓	
Fig. 2 E-F	✓	✓
Supplementary Fig. 1 A-B, Supplementary Fig. 2, Supplementary Fig. 4, Supplementary Fig. 18 A-D	✓	✓

Methods

	involved in		
	conception	execution	analysis & interpretation
Flow cytometry		✓	✓	✓
EGFP-binding assay			✓	
Mass spectrometry		✓	✓	✓

Efficient amber suppression via ribosomal skipping for in situ synthesis of photoconditional nanobodies

E. F. Joest, **C. Winter**, J. S. Wesalo, A. Deiters, R. Tampé
ACS Synthetic Biology **2022**, 11, 1466-1476 b

Study

	involved in		
	conception	execution	revision & proofreading
Initial draft				/
Manuscript writing			✓	✓
Cover design		✓	✓	/

Figures

	involved in	
	conception	design & execution
Figure 2 c, Figure 3 b		✓	✓
Figure S1 b, Figure S3 b, Figure S4, Figure S5 a-c, Figure S8 c, Figure S9, Figure S10		✓	✓

Methods

	involved in		
	conception	execution	analysis & interpretation
Mass Spectrometry		✓	✓	✓
Flow Cytometry and Analysis		✓	✓	✓

TAPBPR is necessary and sufficient for UGGT1-mediated quality control of MHC I

L. Sagert, **C. Winter**, M. Zehetmaier, I. Rupert, C. Thomas, R. Tampé
manuscript in preparation

Study

	involved in		
	conception	execution	revision & proofreading
Initial draft			/
Manuscript writing		✓	/
Data availability	✓	✓	/

Figures

	involved in	
	conception	design & execution
Figure 1 a-e	✓	✓
Figure 3 b-c, Figure 4 a		✓
SI Figure 3	✓	✓
SI Figure 1 b, SI Figure 2		✓

Methods

	involved in		
	conception	execution	analysis & interpretation
LC-MS analysis	✓	✓	✓
UGGT1-mediated Glycan processing assay		✓	✓
Peptide synthesis and purification	✓		✓

VI. Acknowledgements

In the end, I would like to gratefully acknowledge everyone who has contributed to the accomplishment of this work and who supported me throughout the process:

First, I would like to express my sincere gratitude to my PhD supervisor Prof. Robert Tampé for providing me with the opportunity to demonstrate my talents on not just one, but a wide variety of projects. Thank you for the numerous, detailed discussions, your inspiring ideas and visions, your conveyed knowledge, and your attention to detail. Thank you for always having an open door for results, meetings, and concerns.

Furthermore, I would like to express my gratitude to Alexander Domnick for the successful, professional, and harmonious cooperation. Without your contribution, large parts of this dissertation would not have been possible.

I would also like to thank Dr. Eike Frank Jöst for his excellent cooperation in our two publications and his unmatched sense for style and design.

In addition, I would like to thank Ralph, Simon, and Rupert for many helpful, inspiring, and instructive discussions.

In absence of a competent laboratory management, nothing would have worked, hence a big thank you goes to Renate, Katrin and Stefan. Many thanks also to Andrea and Inga for your swift support, at any time and in all matters.

I would like to thank my friends and colleagues at Tampé Lab for the wonderful atmosphere, constant support, and encouragement. Especially I would like to acknowledge Alex, Eike, Heike, Holger, Ines, Jamina, Katharina, Lina, Martina, Max, Samoil, Tamina and Tim.

A big thanks to Sophie, Alex, Lina and for your efforts proofreading my work.

Finally, I would like to thank my family and in particular my parents for their everlasting and loving support in all situations of life. Special thanks to my beloved partner Sophie for her amazing, passionate, and generous support in all my endeavors, as she has been throughout my doctoral studies.

VII.   Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt,

dass ich die vorgelegte Dissertation mit dem Titel

Mechanistic and structural insights into the quality control of the MHC I antigen processing pathway

selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Christian Winter

Frankfurt am Main, den...

VIII.   Curriculum vitae