Towards the Conformational Dynamics of Multidomain Proteins

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
Christina Sabine Heil
aus Bad Nauheim

Frankfurt 2019
(D30)

vom Fachbereich Biochemie, Chemie und Pharmazie der

Johann Wolfgang Goethe–Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Clemens Glaubitz
Gutachter:	Prof. Dr. Martin Grininger Prof. Dr. Michael Göbel

Datum der Disputation:

Vorwort

Teile der vorliegenden Arbeit wurden vorab veröffentlicht in Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878. sowie Heil, Christina S. *et al.* "Fatty Acid Biosynthesis: Chain-length Regulation and Control" *ChemBioChem* (2019). Mein eigener Anteil am Schreibprozess dieser Publikationen sowie Copyright und Creative Commons Lizenzen sind im Abschnitt Statement of Personal Contributions näher erläutert.
In diese Arbeit sind die Ergebnisse meiner im Jahre 2014 dem Fachbereich 14 - Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe–Universität Frankfurt am Main vorgelegten Masterarbeit "Incorporation of Non-Natural Amino Acids in the Mammalian Fatty Acid Synthase" eingegangen, sowie Ergebnisse weiterer Master- und Bachelorarbeiten des Fachbereichs 14 unter meiner Betreuung. An geeigneten Stellen im Text wird auf diese hingewiesen. Eine Auflistung der inkludierten Arbeiten ist im Abschnitt Statement of Personal Contributions zu finden.

Zusammenfassung

Multidomänen-Enzyme, wie z. B. Fettsäuresynthasen (FASs) oder Polyketidsynthasen (PKSs), spielen eine bedeutende Rolle in der Biosynthese wichtiger Naturstoffe. Sie sind von großer Bedeutung für die Entwicklung neuer Pharmazeutika und viele Forschungsvorhaben spezialisieren sich auf deren Protein-Engineering. Während PKSs vor allem für die Entwicklung neuer Antibiotika – aber auch anderer vielfältiger Pharmazeutika wie z. B. Immunsuppressiva oder Krebstherapeutika – relevant sind,[1] ist die humane FAS des Typs I aufgrund ihrer Schlüsselfunktion in der Lipogenese beispielsweise ein interessantes Zielprotein in der Therapie von Adipositas.[2] Auch in der Behandlung von Diabetes ist die FAS ein beliebtes therapeutisches Zielprotein.[3] Des Weiteren wurde beobachtet, dass die Expression der Fettsäuresynthase in zahlreichen Krebszellen hoch reguliert ist, was sie ebenfalls zu einem Zielprotein in einer möglichen Krebstherapie macht.[4]

Die beiden Klassen von Multidomänen-Enzymen, FASs und PKSs, sind sich strukturell wie auch funktional sehr ähnlich. Die FAS des Typs I aus Säugern wird als evolutionärer Vorläufer der Familie der PKSs angesehen.[5] Da sie bereits sehr gut untersucht und charakterisiert wurde, eignet sie sich besonders gut als Modellprotein für die Masse an strukturell wenig aufgeklärten PKSs.

Darüber hinaus stellen Fettsäuren eine strategisch wichtige und biotechnologisch wertvolle Plattformchemikalie dar, deren Verfügbarkeit über nachhaltige mikrobielle Methoden sichergestellt ist. Beispielweise gelangte Protein-Engineering an FASs mit Hinsicht auf Kettenlängenkontrolle in den letzten Jahren vermehrt in den Fokus der Biotreibstoff-Herstellung. Die Fülle an struktureller Information, die für FASs bereit gestellt ist, führt zur Erkenntnis der molekularen Grundlagen in der Fettsäurebiosynthese. Die Aufklärung der Kettenlängenkontrolle, die Entdeckung und Charakterisierung einer Vielzahl an Substrat-spezifischen Thioesterasen und die Entwicklung des Metabolic Engineerings haben zur gezielten Modulation der Kettenlänge von Fettsäuren geführt.[6]

FASs kommen in unterschiedlicher Form in der Natur vor. Man unterscheidet zwischen einzelnen dissozierten, monofunktionalen Enzymen des Typs II – wie er in Prokaryoten, Pflanzen und der mitochondriellen Fettsäuresynthese vorkommt – und den Multidomänen-Enzym-Komplexen des Typs I – wie er in Eukaryoten und einigen Vertretern der Gruppe von Corynebacterineae (*Corynebacteria, Mycobacteria* und *Nocardia*) vorkommt. Die Multidomänen-Enzyme des Typs I unterteilen sich nochmals in zwei unterschiedliche strukturelle Architekturen. Während sich die Polypeptidketten von Pilz- und Corynebacterineae-FASs des Typs I zu großen käfigähnlichen Strukturen einer Größe im Bereich von MDa zusammenlagern, bilden Säuger-FASs des Typs I Homodi-

mere von ungefähr 540 kDa Größe. Trotz der strukturellen Vielfalt bleibt die Fettsäurebiosynthese zwischen unterschiedlichen Domänen und Organismen konserviert und läuft über den gleichen Mechanismus ab.

Fettsäuren sind essentiell für den Organismus und dienen als Bausteine für Membranen (z. B. als Teil der Phospholipide) und als Proteinanker in denselben, als Energiequelle in Form von Fettreserven (als Teil der Triacylglyceride), oder als Signalmoleküle (z. B. als Eicosanoide).[7–9] Ihre Fettsäurebiosynthese verläuft über einen iterativen Prozess mit strikter Kettenlängenkontrolle. In einem initialen Beladungsschritt wird zunächst eine Acetyl-Einheit (als Startsubstrat dient ein Acetyl-Coenzym A (CoA)-Ester) auf das Enzym geladen, welche dann sukzessive in sich wiederholenden Reaktionssequenzen um eine C_2-Einheit verlängert wird, bis die gewünschte Kettenlänge – in der Regel C_{16} – erreicht ist. Die Knüpfung der kovalenten C–C-Bindung zwischen der Acetyl-Einheit und der Verlängerungseinheit (hierzu dient als Substrat ein Malonyl-CoA-Ester) erfolgt über eine decarboxylierende Claisen Kondensation, katalysiert von der β-Ketoacylsynthase (KS). Das daraus resultierende β-Ketoacyl-Intermediat wird anschließend in drei aufeinanderfolgenden Reaktionen zu einer voll gesättigten Acyl-Kette modifiziert. Hierbei erfolgt zunächst eine Reduktion der β-Keto-Gruppe zu einer β-Hydroxy-Gruppe, katalysiert von der β-Ketoreduktase (KR) in Gegenwart von NADPH. Im Anschluss wird Wasser abgespalten, katalysiert von der Dehydratase (DH). Daraufhin folgt eine weitere Reduktion der resultierenden Alken-Gruppe zur gesättigten Acyl-Kette, katalysiert von der β-Enoylreduktase (ER) in Gegenwart von NADPH.

Die Substrate Acetyl- und Malonyl-CoA werden mittels der Acyltransferase (AT) auf das Enzym geladen. Das Acyl Carrier Protein (ACP) dient als Träger und Shuttle der Substrate und Acyl-Ketten des Fettsäurezyklus. Alle Substrate des ACPs werden kovalent als Thioester an seinen Phosphopantethein-Arm gebunden – eine Armprothese, die über eine post-translationale Modifikation als Phosphatester an die Seitenkette des aktiven Serins angebracht wird, katalysiert von der Phosphopantethein-Transferase (PPT). Das ACP ist eine flexible Einheit der FASs, die eine essentiell wichtige Rolle für die Fettsäurebiosynthese spielt, da sie ihre Ladung von einem Enzym/einer enzymatischen Domäne der FAS zum/zur nächsten transportieren muss.

Ist die gewünschte Kettenlänge erreicht, wird die gebildete Fettsäurekette aus dem sich wiederholenden Fettsäurezyklus ausgeschleust. Das Produkt wird entweder über einen AT-vermittelten Transfer vom Enzym auf ein CoA übertragen (wie in Pilz- und Corynebacerineae-FASs des Typs I), als freie Fettsäure in einer Hydrolyse-Reaktion – katalysiert von der Thioesterase (TE) – vom Enzym abgespalten (wie in Säuger-FASs des Typs I oder Pflanzen-FASs des Typs II), oder das ACP transportiert die finale Acyl-Kette direkt weiter zu anderen Biosynthesewegen – zur Herstellung von Lipiden oder Liponsäuren – oder der Elektronentransportkette (wie in bakteriellen oder mitochondriellen FASs des Typs II).

In dieser These liegt der primäre Fokus auf der Säuger-FAS des Typs I. Vorausgegangene Studien an diesem Enzym bestätigten eine außerordentlich hohe konfor-

mationelle Vielfalt.[10,11] Die Vermutung liegt nahe, dass diese konformationellen Dynamiken einen entscheidenden Beitrag zum reibungslosen Ablauf des katalytischen Zyklus der Fettsäurebiosynthese leisten und das Zustandekommen von effektiven Protein-Protein-Interaktionen des ACPs mit anderen katalytischen Domänen unterstützen.

Der Transport von Substraten und Intermediaten der Fettsäurekette, der über das ACP sichergestellt ist, spielt eine zentrale Rolle im Katalysezyklus der Fettsäurebiosynthese und wirft eine Reihe ungeklärter Fragen auf:
(a) Handelt es sich hierbei um einen zufälligen Bewegungsablauf oder ist die Bewegung des ACPs gerichtet/vorprogrammiert? (b) Wie unterscheidet das ACP produktive von unproduktiven Protein-Protein-Interaktionen? (c) Verändert sich die konformationelle Dynamik des Enzyms über die verschiedenen Zyklen der Fettsäurebiosynthese hinweg, aufgrund der Kettenlänge des an das ACP gebundenen Intermediats? (d) Kann die konformationelle Dynamik des Enzyms durch gezielte Mutationen im Bereich der aktiven Zentren oder der Protein-Protein-Interaktionsflächen von katalytischen Domänen moduliert werden? (e) Welchen Einfluss haben physikalische (z. B. Temperaturänderungen, Änderungen der Ionenstärke des Puffers) oder chemische Effekte (z. B. Substratkonzentrationen, Inhibitoren) auf die konformationelle Dynamik des Enzyms? (f) Wie groß ist die konformationelle Freiheit innerhalb einzelner struktureller Domänen und der Domänen untereinander? (g) Welchen Einfluss hat die konformationelle Dynamik des Enzyms auf die Fettsäurebiosynthese? (h) Kann der genaue Bewegungsablauf des ACPs während des katalytischen Zyklus aufgezeichnet werden? (i) Was sind die Zeitspannen von Protein-Protein-Interaktionen während der Katalyse? (j) Korrelieren die Affinität der Protein-Protein-Interaktionen und katalytische Effektivität? (k) Wird das ACP von der geladenen Kettenlänge dirigiert?

Um diesen Fragenkatalog abzuarbeiten, bedarf es geeigneter Methoden, die in der Lage sind Protein-Dynamiken zu untersuchen. Allen voran stehen spektroskopische Methoden wie Förster Resonanz Energie Transfer (FRET)- oder Elektronen Spin Resonanz (EPR)-Spektroskopie, mit denen Ensembles konformationeller Zustände selbst großer Proteine in nativem Zustand untersucht werden können. Vor allem Einzelmolekülspektroskopie ist eine wertvolle Methode, um der konformationellen Dynamik und dem Katalysezyklus eines einzelnen Enzyms in Echtzeit zu folgen und so die bisher statischen Bilder der konformationellen Dynamik der Säuger-FAS des Typs I mit zeitaufgelösten Daten zu untermauern.

Spektroskopische Untersuchungen wie EPR und FRET erfordern biophysikalische Proben, die mit Fluorophoren oder Spins markiert sind. Um die Säuger-FAS des Typs I mit solchen Markierungen zu versehen, muss auf eine positionsspezifische, bioorthogonale Strategie zurückgegriffen werden. Eine geeignete Markierungsstrategie für solch große Multidomänen-Enzyme stellt die Methode der Amber Codon-Suppression dar. Hierbei wird das Amber Codon benutzt, um eine nicht-natürliche Aminosäure (ncAA) positionsspezifisch in die Proteinsequenz einzubauen. Die natürliche Funktion des Amber Codons – das Signal für den Translations-Stopp – wird hierbei von einer Suppressor-tRNA unterdrückt, die an das Amber Codon bindet und mit der ncAA beladen ist. Für die Beladung der Suppressor-tRNA mit der ncAA ist eine orthogonale

Aminoacyl-tRNA-Synthetase (aaRS) zuständig, die spezifisch auf die ncAA zugeschnitten ist. Amber Codon-Suppression folgt also analog dem Prinzip der Translation und macht sich die natürliche Proteinbiosynthese zu Nutze. Eine Vielzahl an ncAAs mit unterschiedlichen funktionellen Gruppen können über diese Methode in Proteine eingebaut werden, darunter auch funktionelle Gruppen, die in chemoselektiven Reaktionen verwendet werden können. Über solche Reaktionen können Fluorophor- oder Spin-Markierungen positionsspezifisch in Proteine eingeführt werden.

Die vorliegende These beschäftigt sich grundlegend mit der Anwendung der Amber Codon-Suppressions-Technologie an FASs des Typs I aus Säugern, mit der Absicht diese Multidomänen-Enzyme für spektroskopische Methoden bereitzustellen, um komplexe biologische Fragestellungen in Bezug auf die konformationelle Dynamik dieser Enzyme – besonders die Prinzipien des ACP-vermittelten Transports der Substrate und Acyl-Ketten der Fettsäurebiosynthese – aufzuklären.

Einen großen Teil dieser Arbeit stellt der Aufbau eines vielfältigen Werkzeugkastens, basierend auf der Amber Codon-Suppressions-Technologie und dessen Etablierung an Säuger-FASs des Typs I dar. Dies beinhaltete umfassende Klonierungen von Amber Codon-Suppressions-Plasmiden, welche zur heterologen Expression der Suppressor-tRNAs und orthogonalen aaRSs vonnöten sind. Ausgehend von kommerziell erhältlichen Vektoren wurde zunächst der pAC^{U}-Vektor konstruiert, welcher nach seinem Vorbild pUltra[12] eine Kopie der orthogonalen aaRS sowie eine Kopie der Suppressor-tRNA enthält. Das aaRS-Gen des Ausgangsvektors pAC^{U} wurde über mehrere gezielte Punktmutationen variiert und somit für unterschiedliche ncAAs evolviert. Die Gene der Suppressor-tRNAs sowie der evolvierten aaRSs wurden anschließend in ein weiteres Vektorsystem übertragen, dem pAC^{E}-Vektor. Der pAC^{E}-Vektor beruht auf dem pEVOL-Vektor,[13] welcher eine Kopie der Suppressor-tRNAs und zwei Kopien der orthogonalen aaRSs beinhaltet. Beide Vektoren, pAC^{U} und pAC^{E}, wurden ausgiebig in einem eigenes dafür entwickelten Reporter Assay Screen auf ihre Funktionalität getestet und miteinander verglichen.

Der Reporter Assay ist ein weiterer wichtiger Bestandteil dieser Arbeit und eine Schlüsselmethode, die für die Etablierung der Amber Codon Suppressions-Technologie in dieser Arbeit entwickelt wurde. Mit Hilfe des Reporter Assays wurden nicht nur die beiden Vektorsysteme, pAC^{U} und pAC^{E}, miteinander verglichen, sondern auch die Effizienz und Zuverlässigkeit des Einbaus unterschiedlicher ncAAs mittels der evolvierten aaRSs bestimmt, um eine Auswahl effektiver ncAAs und pAC-Plasmide für unsere Zwecke zu treffen.

Der Reporter Assay beruht auf der Expression eines ACP-GFP Fusionsproteins, das eine Amber Codon-Mutation innerhalb der ACP-Sequenz trägt. Erfolgreicher Einbau der ncAA an der Stelle des Amber Codons führt zu einem Volllängen-Protein und GFP-Fluoreszenz. Wird das Amber Codon als Translations-Stopp gelesen und die ncAA nicht eingebaut, wird ein verkürztes Protein erhalten, das keine GFP-Fluoreszenz aufweist. Die GFP-Fluoreszenz dient also als Auslesemethode für erfolgreiche Amber Codon-Suppression und die Intensität des Fluoreszenzsignals kann zur Quantifizierung des

erfolgreichen ncAA-Einbaus herangezogen werden. Im Reporter Assay wird hierfür immer ein ACP-GFP-Wildtypprotein als Referenz exprimiert, das keine Amber-Mutation trägt. Das Fluoreszenzsignal der Referenz wird maximaler Volllängen-Expression gleichgestellt und alle anderen Fluoreszenzsignale der Amber Codon-Suppressions-Expressionen in Relation zur Wildtypprotein-Expression gesetzt. So können Effizienzen unterschiedlicher Amber Codon-Suppressions-Expressionen (Verwendung von verschiedenen ncAAs oder pAC-Plasmiden bzw. evolvierten aaRSs) miteinander verglichen werden.

Neben der Wildtypprotein-Referenz werden im Reporter Assay auch Negativkontrollen durchgeführt. In diesen Kontrollexpressionen werden zwar Suppressor-tRNAs und orthogonale aaRSs bereitgestellt, aber keine ncAA zur Expression hinzugefügt, was einen Translations-Stopp erzwingt und kein GFP-Fluoreszenzsignal aufweisen sollte. Für einige Amber Codon-Suppressions-Expressionen konnte jedoch ein Fluoreszenzsignal und somit Volllängen-Protein detektiert werden. Der Grund dafür ist eine schlechte Orthogonalität der evolvierten aaRS, die nicht spezifisch genug für die ncAA ist. Zu einem gewissen Maß akzeptiert sie endogene Aminosäuren und baut diese an Stelle des Amber Codons ein. Mit den Negativkontrollen wird daher die Zuverlässigkeit der evolvierten aaRSs – in welchem Maß sie ausschließlich die ncAA einbaut – überprüft.

Der Reporter Assay wurde als einfache und schnelle Screening-Methode der Amber Codon-Suppression eingeführt, die mit minimalen Ressourcen auskommt und sowohl kostengünstig als auch zeiteffizient ist. Die Protein-Expression findet in kleinem 2 mL-Maßstab in 96-Deep-Well-Platten statt, sodass mehrere Expressionen in relativ hohem Durchsatz miteinander verglichen werden können. Zur Bestimmung der Fluoreszenzintensität können die Zellen einfach in 96-Well-Mikroplatten überführt und im Platereader ausgelesen werden. Aufwendige Arbeitsschritte wie Zellaufschluss und Protein-Reinigung entfallen, sodass der Reporter Assay in einem schnellen und einfachen Arbeitsprotokoll bereitgestellt wird. Die Vergleichbarkeit der Expressionen wird durch Normierung auf Zellzahlen gewährleistet, die über Messung der Optischen Dichte (OD_{600}) der Zellkulturen bei einer Wellenlänge von 600 nm bestimmt wird. OD_{600} und Fluoreszenzintensität können beide in ein und derselben Probe der 96-Well Mikroplatte am Platereader gemessen werden, sodass kein großer Mehraufwand entsteht.

In dieser These wurden elf verschiedene ncAAs und 27 unterschiedliche pAC-Plasmide im Reporter Assay gescreent. Außerdem wurde mit Hilfe des Reporter Assays die positionsabhängige Effizienz des Einbaus von ncAAs getestet. Hierzu wurden zwei verschiedene ncAAs, *para*-Azido-L-Phenylalanin und Norbornen-L-Lysin, an sechs verschiedenen Positionen im ACP eingebaut. Die Ergebnisse des Reporter Assays wurden des Weiteren auf ihre Übertragbarkeit auf größere Expressionskulturen und im Kontext der Volllängen-FAS des Typs I aus Säugern überprüft.
Es konnte nicht nur bewiesen werden, dass der Reporter Assay zuverlässige Aussagen über die positionsabhängige Einbaueffizienz von ncAAs geben kann, sondern es wurden auch erfolgreich Säuger-FASs des Typs I mit eingebauten ncAAs erzeugt. Der Einbau von ncAAs in Proteine solcher Größe wurde nach meinem besten Wissen zuvor noch nirgends bewerkstelligt und stellt somit eine große Errungenschaft dar.

Die generierten Proteinkonstrukte der Säuger-FAS des Typs I wurden daraufhin in chemoselektiven Reaktionen mit Fluorophor- und Spin-Markierungen versehen, um biophysikalische Proben für spektroskopische Methoden bereitzustellen. Als chemoselektive Reaktionen wurden Kupfer-katalysierte Alkin-Azid-Cycloadditionen (CuAAC) sowie kupferfreie Reaktionen – Strain-Promoted Alkin-Azid-Cycloadditionen (SPAAC), Inverse-Electron-Demand Diels-Alder-Cycloadditionen (IEDDAC) und Oxim-Kupplungsreaktionen – herangezogen, wobei sich die SPAAC und IEDDAC für die hier untersuchten Proteine am geeignetsten herausstellten. Über die gewählte Markierungsstrategien konnten somit Fluoreszenz- und Spin-markierte FASs für FRET- und EPR-Spektroskopie erzeugt werden, die in ersten spektroskopischen Experimenten Anwendung fanden.

Um den zugrunde liegenden Amber Codon-Suppressions-Werkzeugkasten voll auszuschöpfen und eine gute Grundlage für die Beantwortung des oben genannten Fragenkatalogs zu schaffen, wurden gleich mehrere Anwendungen der markierten FASs erprobt. Neben dem Versuch einen Fluoreszenzpolarisations-Assay zu etablieren, wurden die Proben zur Etablierung von Einzelmolekül-FRET-Spektroskopie genutzt. Hierfür wurde ein Protokoll zur Immobilisierung der Säuger-FAS des Typs I auf Objektträgern für die Interne Totalreflexionsfluoreszenz (TIRF)-Spektroskopie entwickelt und erste Einzelmolekül-FRET-Messungen durchgeführt. Des Weiteren konnten Spin-markierte Proteinkonstrukte in Continuous Wave-EPR-Spektroskopie detektiert werden.

Die eingebaute ncAA *para*-Azido-L-Phenylalanin lässt sich nicht nur in chemoselektiven Reaktionen verwenden, sondern ist aufgrund ihrer Photoreaktivität außerdem für Crosslinking-Experimente interessant. Mit solchen Experimenten lassen sich kurzlebige Protein-Protein-Interaktionen einfrieren und untersuchen. Diese Methodik wurde in einem Photocrosslinking-Assay angewandt, der in dieser These entwickelt wurde. Als Zielprotein diente eine Untereinheit der Säuger-FAS des Typs I – das Di-Domänen-Protein bestehend aus KS-MAT mit fusioniertem ACP – welche die Produktion von Triessigsäurelacton (TAL) katalysiert. Im ACP dieses Proteinkonstrukts wurde die photoreaktive ncAA eingeführt. UV-Belichtung des Proteinkonstruktes führte zum Crosslink zwischen ACP und der KS-MAT-Di-Domäne, was mittels SDS-Gelelektrophorese und Tandem-Massenspektrometrieanalyse bestätigt wurde.

In einem nächsten Schritt wurde ein Alaninscan durchgeführt, um Oberflächenmutationen in der KS-Domäne einzuführen und so die Protein-Protein-Interaktion zwischen ACP und KS zu verändern (zu stärken oder zu schwächen). Der Effekt der Oberflächenmutation wurde anhand des Photocrosslinking-Assays über Densitometrie der Crosslinkbanden im SDS-Gel untersucht. Außerdem wurde die Integrität der Proteinmutanten – Aktivität im TAL-Assay und Dimerisierungsgrad ermittelt über analytische HPLC-Größenausschlusschromatographie – untersucht.

Die durchgeführten Experimente und daraus erhaltenen Resultate stellen ein solides Fundament für die weiterführende Untersuchung von Säuger-FASs des Typs I zur Aufklärung prinzipieller Fragen zum ACP-vermittelten Substrat-Shuttle und der konformationellen Dynamik dieser Enzyme dar. In dieser These wurden die Technologie

der Amber Codon-Suppression als Markierungsmethode von Multidomänen-Enzymen etabliert und ein vielfältiger Werkzeugkasten hierfür bereitgestellt. Weiterhin wurde der Reporter Assay als zuverlässiger und gut handhabbarer Screen von Amber Codon-Suppressions-Effizienzen entwickelt.

Fluoreszenz- und Spin-markierte FASs des Typs I aus Säugern sowie Untereinheiten dieses Enzyms wurden für spektroskopische Methoden bereitgestellt und in ersten Messungen getestet. Für die Immobilisierung der Proteine auf Mikroskopie-Objektträgern wurde ein geeignetes Protokoll erstellt um Einzelmolekül-FRET Messungen zu ermöglichen. Um Protein-Protein-Interaktionen zu untersuchen, wurde ein Photocrosslinking-Assay entwickelt.

Ein tiefergehendes Verständnis des katalytischen Zyklus der Fettsäurebiosynthese – vor allem des ACP-vermittelten Substrat-Shuttles und des Einflusses der konformationellen Dynamik des Enzyms – ermöglichen gezieltes Protein-Engineering mit therapeutischer und biotechnologischer Relevanz – besonders im Hinblick auf Kettenlängenkontrolle. Die hohe strukturelle und funktionelle Ähnlichkeit sowie der evolutionäre Zusammenhang zwischen FASs und PKSs lassen darauf schließen, dass sich die hier entwickelten Methoden leicht auf PKSs und andere Multidomänen-Enzyme übertragen lassen. Diese Annahme wurde anhand des Einbaus von ncAAs in Module zweier unterschiedlicher PKSs und anschließende Fluoreszenzmarkierung dieser Proteine in dieser These sogar bewiesen.

Contents

List of Figures

List of Tables

List of Abbreviations

aaRS	aminoacyl-tRNA synthetase
ACP	acyl carrier protein
AT	acyl transferase
ATP	adenosin triphosphate
BCN	bicyclononyne
BSA	bovine serum albumin
CMN-bacteria	*Corynebacteria, Mycobateria* and *Nocardia*
CoA	coenzyme A
CuAAC	copper(I)-catalyzed alkyne-azide cycloaddition
CV	column volume
CW-EPR	continuous-wave EPR
DBCO	dibenzoylcyclooctene
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DNA	deoxyribonucleic acid
DOL	degree of labeling
dSTORM	direct stochastic optical reconstruction microscopy
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EF-Tu	elongation factor Tu
EM	electron microscopy
EPR	electron paramagnetic resonance
ER	β-enoyl reductase
FAS	fatty acid synthase
FP	fluorescence polarization
FRET	Förster resonance energy transfer
GCE	genetic code expansion
GFP	green fluorescent protein
GRO	genomically recoded organism
HPLC	high performance liquid chromatography
IEDDAC	inverse electron-demand Diels Alder cycloaddition
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
IVPS	*in vitro* protein synthesis
KR	β-ketoacyl reductase
KS	β-ketoacyl synthase
LD	linker domain
MAT	malonyl/acetyl transferase

mFAS	murine FAS
MPT	malonyl/palmitoyl transferase
MS	mass spectrometry
MS/MS	tandem MS
MT	methyl transferase
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
ncAA	non-canonical amino acid
NHS	*N*-hydroxysuccinimid
NMR	nuclear magnetic resonance
NTA	nitrilotriacetic acid
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PELDOR	pulsed electron-electron double resonance
PKS	polyketide synthase
PPT	phosphopantetheine transferase
PylRS	pyrrolysyl-tRNA synthetase
RNA	ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
Sfp	4'-phosphopantetheinyl transferase
smFRET	single-molecule FRET
SPAAC	strain-promoted alkyne-azide cycloaddition
TAL	triacetic acid lactone
TCO	transcyclooctene
TE	thioesterase
THPTA	tris(3-hydroxypropyltriazolylmethyl)amine
TIRF	total internal reflection fluorescence
TLC	thin layer chromatography
TPP	thiamine pyrophosphate
tRNA	transporter RNA
$tRNA^{Pyl}$	pyrrolysyl-tRNA
$tRNA^{Tyr}$	tyrosyl-tRNA
TSA	thermal shift assay
TyrRS	tyrosyl-tRNA synthetase

1 Abstract

Multidomain enzymes, such as fatty acid synthases (FASs) or polyketide synthases (PKSs), play a crucial role in the biosynthesis of important natural products. They have a high significance in the development of new pharmaceuticals and various research approaches focus on the engineering of these proteins. For example, human type I FAS is an interesting therapeutic target. Owing to its importance in lipogenesis, upregulation of human type I FAS expression has been observed in numerous cancers.[4] Type I FAS is also regarded as important target in antiobesity treatment.[2] Both multidomain enzyme classes – FASs and PKSs – show high structural and functional similarities. Particularly animal type I FAS is most relevant as evolutionary precursor of the PKS family.[5] Therefore, the well characterized FASs are suitable model proteins for the poorly characterized PKSs, to gain deeper understanding in these megasynthases.

Furthermore, fatty acids are considered to be strategically important platform chemicals accessible through sustainable microbial approaches. The recently acquired structural information on FASs provides an excellent understanding of the molecular basis of fatty acid synthesis. The specific understanding of chain-length control, the characterization of a multitude of substrate-specific thioesterases, and the emerging tools and means for metabolic engineering have fostered targeted approaches for modulating chain length. There is large interest in short-chain fatty acids, since these compounds are biotechnologically valuable platform chemicals and biofuel precursors, and attempts on the synthesis of short-chain fatty acids have been reported during the last years.[6]

Primary focus of this thesis lies on the animal type I FASs, which exhibit large conformational variety, as seen in electron microscopy and high-speed atomic force microscopy.[10,11] Conformational dynamics facilitate productive protein-protein interactions between catalytic domains within the enzyme and aid acyl carrier protein (ACP)-mediated substrate shuttling during the catalytic cycle of fatty acid biosynthesis. To gain deeper insight into the fundamental processes of ACP-mediated substrate shuttling and the underlying conformational dynamics, spectroscopic methods like Förster resonance energy transfer and electron paramagnetic resonance spectroscopy shall be

employed. These spectroscopic methods demand site-specific labeling of proteins with fluorophore or spin labels, which can be accomplished with the amber codon suppression technology. Through amber codon suppression, a non-canonical amino acid (ncAA) with an orthogonal functional group is incorporated site-specifically into the protein sequence, which can be used in chemoselective reactions for protein labeling.

This thesis is at the forefront of employing the technology of amber codon suppression for addressing complex biological questions on megasynthases. The successful production of ncAA-modified FASs is challenging. With the aim of incorporating ncAAs into the multidomain 540 kDa large murine FAS, we by far exceed boundaries of documented application of amber codon suppression. Most of the proteins that are reported by Liu & Schultz in applications of amber codon suppression are in the range of 30 kDa – for example the TE domain of human FAS.[14] In the same review, the largest protein amber codon suppression was applied to is a potassium channel with roughly 80 kDa. Thus, to the best of my knowledge no protein exceeding 100 kDa has been used in amber codon suppression so far.

In this thesis a low-complex, well-plate based reporter assay is presented, based on an ACP-GFP fusion protein for fast and efficient screening of ncAA incorporation. Reliability and applicability of the reporter assay is demonstrated by successful upscaling to larger protein constructs and increased expression scale.

As outlined in this thesis, we have carefully set up methods for the modification of murine FAS and made several achievements:
(i) We have created our own toolbox with a multitude of suppressor plasmids and various orthogonal pairs. pAC^{U} and pAC^{E} plasmids are compatible for fast exchange of cassettes, and cloning procedures are optimized for modification of synthetases by site-directed mutagenesis. (ii) We have organic synthesis of several ncAAs stably running in the lab and synthesis of other ncAAs can be established when required. Therefore, extensive screening at moderate costs is possible. (iii) We have established a reporter assay for screening our own library of vectors for amber codon suppression and for optimizing incorporation of ncAAs. (iv) We successfully incorporated ncAAs into subconstructs and full-length murine FAS, and collected initial promising results for the application of these proteins in spectroscopic methods. Thus, laying the foundation for future studies to address fundamental questions of the ACP-mediated substrate shuttling and other conformational dynamics of these enzymes.

2 Introduction

2.1 The Genetic Code

"Long ago it became evident that the key to every biological problem must finally be sought in the cell; for every living organism is, or at some time has been, a cell" (quote by E. B. Wilson, a pioneering cell biologist, 1925). By the 21st century, a lot of the cell's mysteries have been uncovered and we have gained a quite good understanding of the fundamental units of life. These biochemical factories are made up of basic molecular building blocks that can, broadly speaking,[1] be divided in four major families of small organic molecules: sugars, nucleotides, amino acids, and fatty acids. These small molecules serve either as energy source in intracellular metabolic pathways or as monomer subunits for the biological macromolecules that are: the nucleic acids, namely DNA and RNA, and the proteins, which make up most of the cell's dry mass.[2] While the nucleic acids store all the genetic information of the cell (making up the genome), proteins perform all the functions of the cell (making up the proteome) and carry out key functions in essentially all biological processes, i. e. catalysis, signal transmission, and structural support. Proteins are – from a chemical point of view – structurally the most complex and functionally the most sophisticated molecules. Despite their vast diversity, proteins are made up out of a standard repertoire of only 20 amino acids, which display a limited set of functionalities. Often, chemistries beyond this repertoire are required, realized by cofactors or post-translational modifications. Some organisms also utilize an additional amino acid, selenocystein (the 21st amino acid) or pyrrolysine (the 22nd amino acid), which are catalytic residues specialized for redox reactions and methanogenesis.

In order for the cell to function, the genetic information, which is stored in the DNA, has to be converted into proteins. The flow of genetic information from DNA to RNA and from RNA to proteins is a fundamental principle of molecular biology, known as central dogma. In the early 1960s, researchers got to the bottom of a set of rules how

[1] Actually, not all compounds found in cells do fit into these four categories.
[2] Also sugars can build up polymeric macromolecules, the polysaccharides.

nature translates the genetic information from the four-letter alphabet of DNA/RNA into the 20-letter alphabet of proteins, and thus deciphered the genetic code, which is highly similar among all living organisms. The genetic information is read out in groups of three nucleotides at a time, which code for an amino acid. These nucleotide triplets are called codons. In a four-letter alphabet, there are $4^3 = 64$ possibilities of codons, but a standard or canonical genetic code only has 20 amino acids. Some codons are redundant, which means that several codon synonyms can specify a single amino acid. How the process of translation or protein biosynthesis works, is described in the legend of Figure 2.1.

Synthetic biology aims at creating artificial biological systems and conceives the concept of an expanded or variant genetic code to facilitate the design of proteins with novel functions. Approaches will be discussed in the following section.

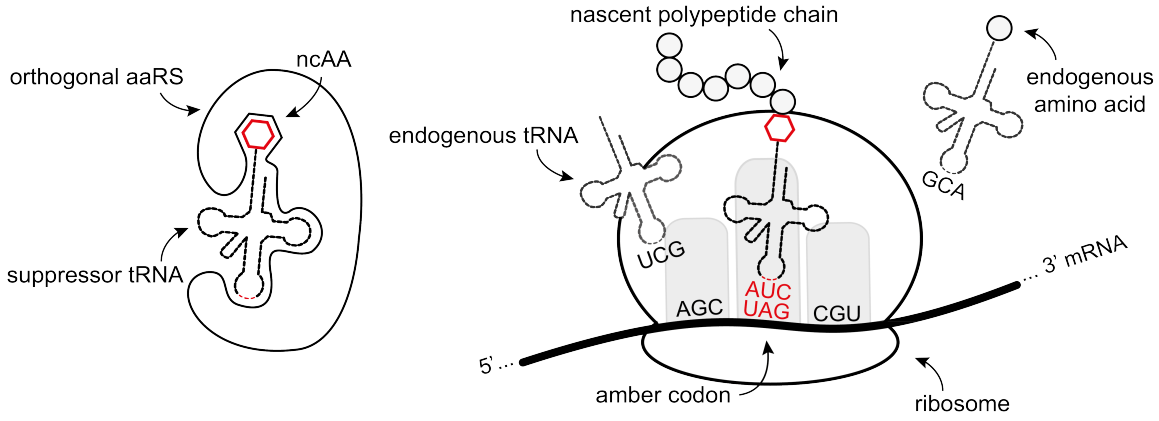


Figure 2.1: Principle of amber codon suppression. Proteinbiosynthesis occurs in the ribosomes, guided by an mRNA template. The genetic information is stored in triplet codons and translated into an amino acid sequence. The decoders of the genetic code are the tRNAs, which bind to the mRNA through anticodons, and carry the encoded amino acids. aaRSs charge the tRNAs with the correct amino acids in respect to the anticodon. The amino acids are linked to a polypeptide, catalyzed by the ribosomes. These biomolecular machineries read along the mRNA strand, which includes start and stop signals for the translation. Amber codon suppression works analog to the natural occuring biosynthesis of proteins. An orthogonal suppressor pair from another kingdom of life, consisting of a tRNA and aaRS, is heterologously expressed and interferes in the endogenous translation process. The amber codon normally serves as translation stop signal, but is overwritten with a new function to encode an ncAA. This is accomplished by providing the suppressor tRNA, which is aminoacylated with the ncAA through the orthogonal aaRS, and incorporates the ncAA in the peptide sequence through the amber anticodon.

2.2 Genetic Code Expansion

Synthetic biologists have long tried to control protein structure and function to design proteins with enhanced or novel activites. By modification, proteins with new physicochemical, biological, or pharmacological properties can be obtained.[14]

Most commonly, solid-phase peptide synthesis is utilized to create designer peptides and to site-specifically incorporate synthetic amino acids[3]. Originally accomplished by R. B. Merrifield, he has been awarded the Nobel Prize in Chemistry 1984 for this achievement. In an automated one-pot reaction, chemists are able to synthesize any peptide. This *in vitro* approach is however limited in size of the synthetic peptides (typically, peptides of over 50-100 amino acids in length are difficult to synthesize in sufficient yields), although in combination with native chemical ligation semisynthetic polypeptides of larger size (up to 200-300 amino acids in length) can be generated. Needing protecting group chemistry, restricted ligation sites, constrained protein folding, and the inherent extracellular nature are general limitations of this synthetic approach.[14]

Other methods directly address amino acids, bypassing the central dogma, like direct manipulation of amino acid side chains.[15] Yet, synthetic biologists were looking for a way to manipulate the genetic code by invasion of non-canonical amino acids (ncAAs) and have been successful exploiting the biomolecular machinery of protein biosynthesis by simply adding new components to it.

In order to add a new amino acid to the genetic code, a blank codon is needed to code for the new amino acid.[14] One simple approach to generate a variant genetic code is the substitution of a canonical amino acid by a non-canonical analog and its incorporation in response to a sense codon. In this approach an auxotrophic strain is used, deficient in one canonical amino acid, and the ncAA substitute is incorporated globally into a protein sequence exploiting endogenous components of the biomolecular proteinbiosynthesis machinery. This leads to a global replacement of the canonical amino acid at all sites in the protein, so this method is rather residue-specific than site-specific. Since wild-type aminoacyl-tRNA synthetases (aaRSs) are used, this method is restricted to very close amino acid analogs, e. g. substitution of methionine with selenomethionine. Another limitation is the constrained exponential cell growth. Although non-dividing

[3]Several notations are found in literature, i. e. unnatural, non-natural, non-proteinogenic, non-standard or non-canonical amino acids. Throughout this thesis we will stick to the term non-canonical amino acids (ncAAs).

cells are still viable and able to express protein, low protein yields are usually the result.

However, to expand the genetic code, simple substitution of amino acids with similar analogs is no adequate approach, so the search for suitable blank codons in order to incorporate diverse ncAAs continues. Of the 64 codons, there are only three nonsense codons (the stop codons: amber, ochre & opal), which do not code for any amino acid. Nonsense codons may be favoured because they are only used once at the end of genes; thus, their reassignment is expected to cause minimal damage to the proteome. Since the amber codon (UAG) is being used the least in nature, it is the most obvious choice to be overwritten with the new function to encode an ncAA (Fig. 2.1). Actually, the original function of the stop codon remains in place and is only suppressed by a suppressor tRNA, which carries the corresponding anticodon to bind to the nonsense codon and is aminoacylated with the ncAA. For this reason, this method is generally termed codon suppression. As a consequence, codon suppression is a competitive reaction between incorporation of ncAAs and translation stop, which usually compromises the expression yield of full-length protein. But all in all, suppression of stop codons is well tolerated by the biomolecular machinery.

The easiest way to manipulate the genetic code is using an extracellular biosynthetic approach like *in vitro* protein synthesis (IVPS). Made of cell extracts, this cell-free environment contains all necessary components for protein production (reconstitutes the fully functional translation system for protein synthesis), from the biomolecular machinery (the ribosome and its cofactors, aaRSs), over a supply of building blocks (nucleotides, tRNAs, amino acids), up to energy source. Such IVPS kits are nowadays commercially available, so the user only needs to insert his gene of interest in form of a DNA plasmid or linear PCR template. Due to the inherent open nature of these cell-free translation systems, the user has direct access and control of the translation environment, which allows for direct manipulation of the genetic code.
For example an expanded genetic code is easily achieved, supplying chemoenzymatically aminoacylated tRNAs, that can be used to incorporate ncAAs in response to nonsense codons. Furthermore a minimal genetic code can be generated, removing all tRNAs recognizing redundant codons. As a consequence thereof, several blank codons are generated, which can in turn be used to generate a maximal genetic code.
Hiroaki Suga brings this system to perfection, having revolutionized this field by the development of highly versatile flexizymes, which can charge a wide variety of ncAA derivatives onto any tRNA.[16] Until the beginning of the new millennium, the chemoenzymatic method used to aminoacylate tRNAs was technically very demanding and

required HPLC purification and handling of the hydrolytically unstable aminoacyl-tRNAs during deprotection. No useful catalyst was available to make use of ncAAs. Nowadays, having flexizymes, a variety of non-canonical aminoacyl-tRNAs can be easily prepared by simply mixing the three components flexizyme, its matched acyl donor, and tRNA in the presence of Mg^{2+} and incubation for 2 h at 4 °C. Furthermore, Suga developed a flexible *in vitro* translation system based on the PURE IVPS, which consists of purified *Escherichia coli* translation components integrated with the flexizyme technology, for reprogramming the genetic code by sense suppressions and reassigning ncAAs to certain codons (artificial division of codon boxes).[17] Limitations of these systems are the general low protein yields (up to 500 µg/mL in batch reaction, or 5 mg/mL in dialysis mode, but for large and complex proteins notably lower).

Chemoenzymatically aminoacylated tRNAs have also been microinjected in living cells (*Xenopus* oocytes) to incorporate ncAAs in response to a nonsense codon. This approach has allowed for detailed structure-function studies on ion channels.[14] Limitations of this transfection technique are the limited accessibility and stability of aminoacyl-tRNAs and their stoichiometric use, since they cannot be delivered continuously. Furthermore the disruptive nature of microinjection is a huge drawback.

The naturally occurring pyrrolysyl-tRNA ($tRNA^{Pyl}$) from *Methanosarcina mazei* and *barkeri* incorporates the aforementioned 22nd amino acid pyrrolysine in response to the amber codon. Since this pair of $tRNA^{Pyl}$ and pyrrolysyl-tRNA synthetase (PylRS) is from another kingdom of life, the archaea, it is actually orthogonal in other organisms like *E. coli* or eukaryotic cells. This benchmark discovery and development of orthogonal aaRS/tRNA pairs led to the very first application of genetic code expansion (GCE) in *E. coli* by the groundbreaking work of the GCE pioneer Peter G. Schultz and coworkers.[18]

In order to exploit the biosynthetic translation machinery in an efficient fashion, certain prerequisites for this methodology have to be met, with orthogonality leading the way: (i) metabolically stable ncAAs with good cellular bioavailability must be delivered, which furthermore must not be a substrate for any endogenous aaRSs, but must be tolerated by the elongation factor Tu (EF-Tu) and the ribosome; (ii) the blank codon encoding the ncAA must only be recognized by the new tRNA but not by any endogenous tRNAs; (iii) an ncAA-specific aaRS/tRNA pair must be provided, which is both functional in the host organism and orthogonal in the context of all the

endogenous aaRS/tRNA pairs.[14]

Most criteria are quite easily met since natural aaRSs have evolved high specificity for their cognate amino acids and the aminoacyl-binding site of EF-Tu and the ribosome are highly promiscuous. Simple base pairing rules determine codon recognition and designing a tRNA with anticodon corresponding to the chosen blank codon is not challenging at all. The real challenge lies in the orthogonality and aminoacylation specificity of the aaRS/tRNA pair. Luckily, suitable aaRSs are found in nature, which show a good degree of plasticity and using a heterologous aaRS/tRNA pair from another domain of life usually fulfills orthogonality. Systematic directed evolution (a field for which the Chemistry Nobel Prize 2018 has been awarded to Frances H. Arnold) of aaRS/tRNA pairs has led to the generation of a vast number of ncAAs to be incorporated.
In 2010, Liu & Schultz reported 71 different genetically encoded ncAAs,[14] only five years later, in 2015, Dumas *et al.* reported 167 ncAAs, which had been genetically encoded into proteins.[19] Most likely, this number has further drastically increased to date, over the course of another five years. A two-step selection process guarantees for orthogonality of the evolved aaRS/tRNA pairs.[14,20]

By now, four main orthogonal pairs have been developed for GCE, which differ in their orthogonality in different hosts: *M. jannaschii* tyrosyl-tRNA synthetase/$tRNA^{Tyr}$ (*Mj*TyrRS/$tRNA^{Tyr}$), *E. coli* tyrosyl-tRNA synthetase/$tRNA^{Tyr}$ (*Ec*TyrRS/$tRNA^{Tyr}$), *E. coli* leucyl-tRNA synthetase/$tRNA^{Leu}$ (*Ec*LeuRS/$tRNA^{Leu}$), and PylRS/$tRNA^{Pyl}$. While the *Mj*TyrRS/$tRNA^{Tyr}$ pair is orthogonal in *E. coli* but not in eukaryotic cells, orthogonality of the *Ec*aaRS/tRNA pairs behaves the other way around. Only PylRS/$tRNA^{Pyl}$ pairs can be used in both bacteria and eukaryotic cells.

By now a variety of possible applications of this method has been utilized, which are summarized in current reviews.[14,21] Still, there are challenges in GCE, like multiple introduction of the same or different ncAAs and the competitive nature of nonsense suppression. Every endeavor has been made to increase codon suppression efficiency, ranging from improving expression levels of the orthogonal aaRS/tRNA pair, to engineering aaRS activity. Many exciting advances have been made in this field. Genome wide codon reassignment led to genomically recoded organisms (GROs), which offer blank codons and eliminate the competitive release function for these codons.[22–24] Other nonsense or quadruplet, even quintuplet codons can be used to generate additional blank codons. Recently, Abhishek Chatterjee reported the incorporation of three

distinct ncAAs in a single protein.[25]
To overcome the drawbacks of quadruplet and quintuplet codon usage, Jason Chin developed orthogonal ribosomes, which are used aside of the natural ribosomes.[26,27] But still, the wobble rules for systems with quadruplet and quintuplet codons are poorly understood. Even new base pairs have been tried to implement in the genome to increase the number of blank codons. Although several non-natural base pairs are implemented into IVPS, major challenges must still be overcome for the full implementation of non-natural base pairs *in vivo*. Replication, transcription and translation of the non-natural base pairs has to be achieved for successful application.[28] Malyshev *et al.* made a crucial step toward *in vivo* implementation, reporting the first semisynthetic organism with two synthetic bases X and Y in 2016.[29] The technology of codon suppression has also spread to multicellular organisms. With the first application in *Caenorhabditis elegans* by Greiss *et al.* in 2011 leading the way,[30] it took only ten years from its first application in such a simple organism like *E. coli*. Nowadays even engineered mice with extended genetic code have been reported.[31]

Despite the rapid development, its application seems limited to smaller proteins and remains a huge challenge on multifunctional proteins. The model protein of this thesis, we planned to use GCE on, is the large and complex multidomain enzyme fatty acid synthase and will be introduced in the next section.

2.3 The Multidomain Enzyme Fatty Acid Synthase

Parts of this chapter have previously been published in: Heil, C. S. *et al.* “Fatty acid biosynthesis: Chain-length regulation and control” *ChemBioChem* (2019). For individual contribution to the writing process as well as copyright and Creative Commons license, please consult the Statement of Personal Contributions.

Fatty Acids:

One of the basic molecular building blocks of the cell are fatty acids. As named in the introduction section “The Genetic Code” 2.1, they make up a major family of small organic molecules. They show typical amphiphilic properties originating from the hydrophobic tail and the hydrophilic head group. Such properties allow fatty acids to act as building blocks of membranes (e. g. as part of the phospholipids), to serve as covalently tethered membrane anchors of proteins, to be packed tightly into storage

fat (as part of the triacylglycerides), or to be processed into signaling molecules (e. g. eicosanoids).[7–9] Thus, fatty acids play an essential role in the cell, as energy source in intracellular metabolic pathways or as structural feature in lipid membranes.
Although termed as small molecules, fatty acids themselves have a polymeric character, which is not so obvious as in the biological macromolecules nucleic acids (polynucleotides) or proteins (polypeptides). While nucleic acids are synthesized by templated polymerization (DNA replication or transcription into RNA) out of four monomeric building blocks (nucleotides A, T or U, C & G), and proteins are also synthesized by templated polymerization (translation) out of 20 monomeric building blocks (amino acids), fatty acids are produced in repetitive reaction of only two monomeric building blocks (acetyl- & malonyl-CoA). Protein and nucleic acid synthesis are guided by mRNA and DNA templates both in the choice of the monomeric unit and in the determination of the length of the polymer. Fatty acid biosynthesis is simpler with regard to the selection of the monomeric unit because it does not vary the polymerizing unit. However, the determination of fatty acid chain length is comparably more challenging because it is defined by the intrinsic properties of the protein – the fatty acid synthase (FAS) – itself.

Fatty Acid Synthases:

FASs can be divided into two systems, on the basis of their protein architectures. Bacteria and plants perform fatty acid biosynthesis with separate, monofunctional enzymes in a dissociated fatty acid biosynthesis system,[32,33] referred to as type II FAS. In eukaryotes and particular bacteria of the subgroup of Corynebacterineae [*Corynebacteria, Mycobacteria* and *Nocardia* (CMN-bacteria)], fatty acid biosynthesis occurs in large multifunctional FASs, termed type I FASs, in which all the required enzymatic activities are integrated into multidomain complexes.[5,8,34] *De novo* fatty acid biosynthesis is also found in mitochondria, resembling that catalyzed by bacterial type II FASs.[35,36] Although FASs occur in diverse kingdoms of life and various structural frames, the chemistry of fatty acid biosynthesis is highly conserved.

Fatty Acid Biosynthesis:

De novo biosynthesis of fatty acids undergoes an iterative process with strict regulation of the length of the produced fatty acids. A fatty acid is always synthesized through a repetitive reaction sequence in which the initial priming step is performed with an

acetyl moiety, which is then successively elongated with C_2 units, originating from malonyl moieties, until an acyl chain of the desired chain length is obtained (Fig. 2.2). The C–C bond formation takes place through decarboxylative Claisen condensation catalyzed by the condensing β-ketoacyl synthase (KS) to yield a β-ketoacyl intermediate bound to the acyl carrier protein (ACP). After condensation, the β-ketoacyl intermediate is modified through sequential processing by a NADPH-dependent β-ketoacyl reductase (KR), a dehydratase (DH) and another NADPH-dependent β-enoyl reductase (ER), eventually to yield the fully saturated acyl intermediate. The acetyl and malonyl moieties are provided as CoA esters and loaded through the action of specific acyl transferases (ATs).

The greatest variation in fatty acid biosynthesis is found in the termination of the repeating elongation. Type I systems of fungal and CMN-bacterial FASs terminate synthesis through the AT-mediated transfer of the acyl moiety from ACP to CoA. Animal type I FASs (including human FAS) terminate synthesis through the action of a thioesterase (TE), releasing the free acid as product. Type II systems in bacteria and mitochondria terminate synthesis through forwarding of acyl-ACP to lipid biosynthetic pathways[37], as well as lipoic acid synthesis and electron transport chain complex assembly.[38] Type II synthesis in plants is terminated by TE-mediated hydrolysis.[39] Most commonly, FASs release palmitic acid (C_{16}) and stearic acid (C_{18}), or the coenzyme A (CoA) derivatives thereof.

ACP Substrate Shuttling:

All substrates and intermediates accepted by the FAS or produced during synthesis are shuttled by the ACP, an inherent protein of type II FAS or a domain of type I systems.[40,41] ACPs are small, monomeric, helical bundle proteins and post-translationally modified with a phosphopantetheine moiety at the active serine residue. ACPs accept a broad range of acyl moieties and even non-natural cargos, all of which become covalently attached to the phosphopantetheine prosthetic group upon thioester formation. After having been loaded through the action of the ATs, the ACPs successively carry their cargo from one enzyme/enzymatic domain to another during the iterative biosynthetic cycles, and thus play a central role in fatty acid biosynthesis. It is therefore clear that the interactions between ACPs and the constituent FAS enzymes/enzymatic domains are critical for the FAS function. Each of these interactions must show appropriate specificity to allow for both loading and unloading of the substrates, but at the same time enable rapid turnover.[42]

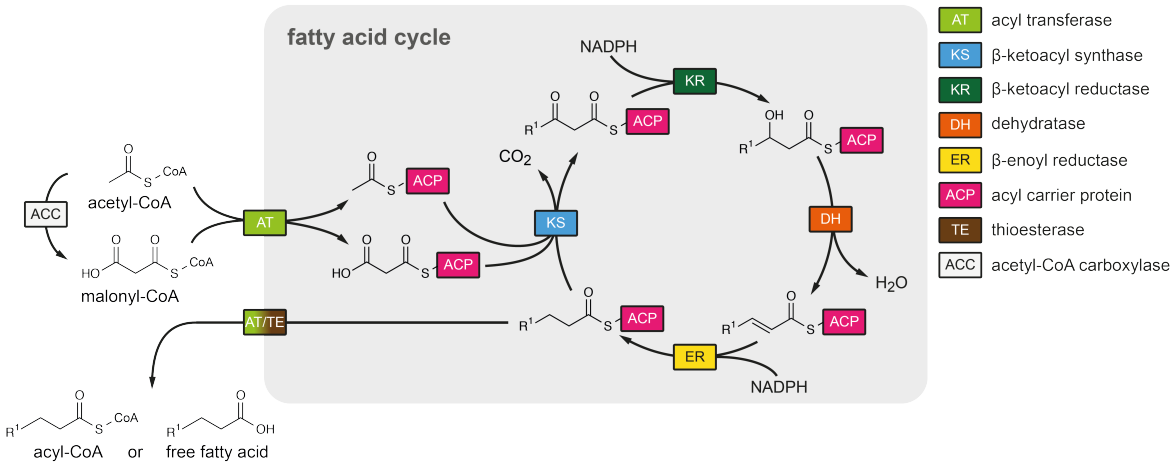


Figure 2.2: General cycle of fatty acid biosynthesis. Multiple catalytic reactions – which are conserved between species – are performed during chain elongation. The growing chain is first extended by a two-carbon (C_2) unit and then processed in a cascade of three reactions (reduction–dehydration–reduction) that blur the traces of the actual C–C bond forming condensation reaction, resulting in a saturated aliphatic chain. The extended acyl chain can be further elongated by running through the same set of reactions until the desired length is achieved. R^1 refers to the growing alkyl chain; in the first cycle it represents CH_3 and in the other cycles $(CH_2\text{-}CH_2)_nCH_3$.

In type I FASs, the ACPs are tethered to the multienzyme through flexible linkers, whereas in type II FASs the ACPs are discrete proteins and diffuse through the cytosol to transport the growing fatty acid chain to the reaction partners. To protect their cargos from thioester hydrolysis, side-reactions, or premature product release, ACPs of some FAS systems sequester the reactive intermediates within the hydrophobic core of the fold.[43–48] This substrate sequestration primarily occurs in type II systems and has not been observed for type I systems.[49–51]

Hammes and colleagues performed fluorescence studies (FRET and fluorescence anisotropy measurements) of chicken liver type I FAS and measured distances between catalytic sites within the multienzyme. The phosphopantetheine arm of the ACP domain is about 20 Å long, not long enough for the growing acyl chain to reach all catalytic sites. In conclusion, relatively large conformational changes of the multienzyme must be part of successful interactions between the ACP domain and the associated enzymatic domains.[52–54]

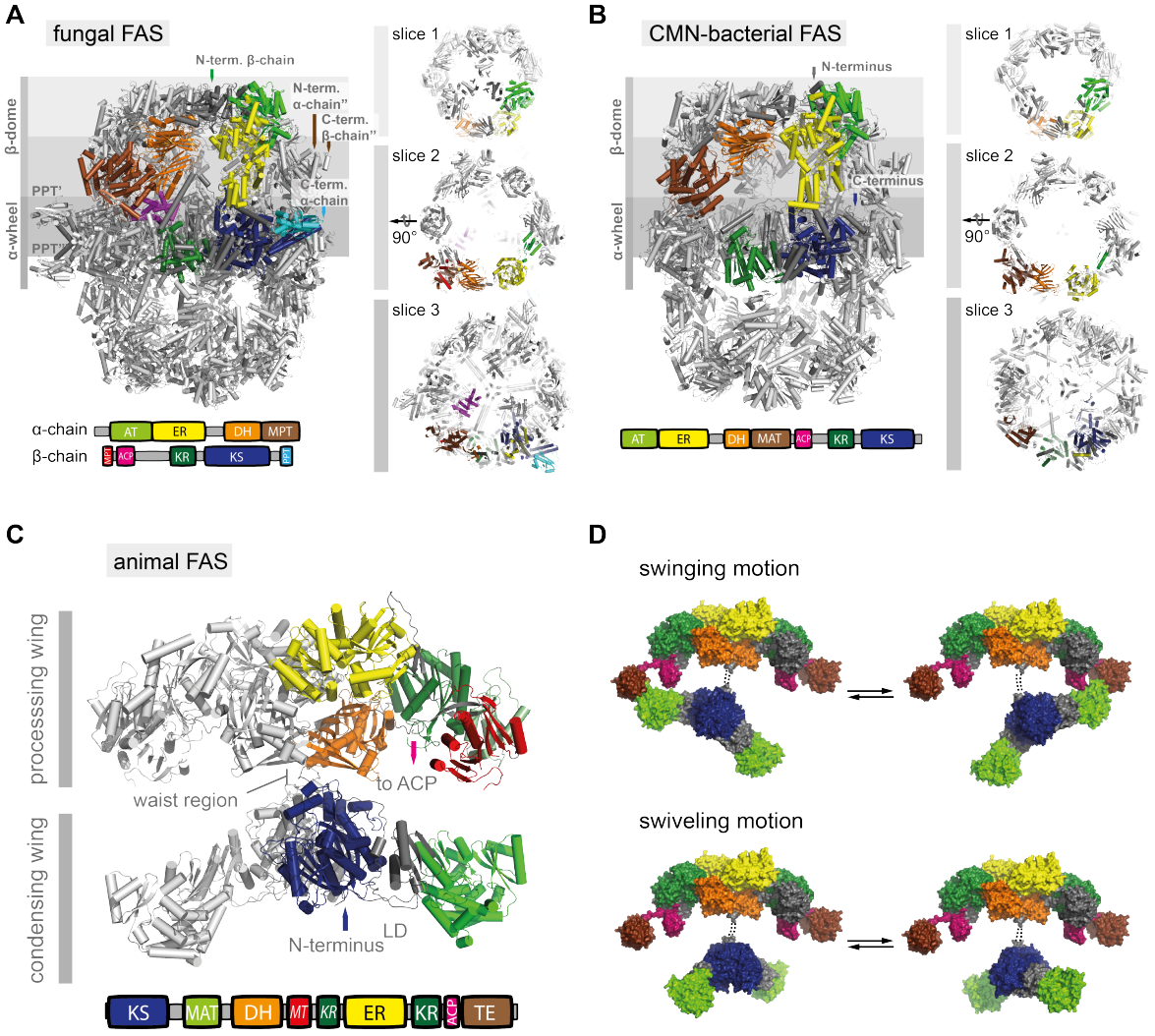


Figure 2.3: Structure and domain organization of the type I FASs. A) Fungal type I FAS. X-ray crystallographic model of baker's yeast type I FAS (PDB ID: 3HMJ).[55] Baker's yeast FAS comprises six α- and six β-chains. One α- and one β-chain are colored as indicated in the domain overview. The phosphopantetheinyl transferase (PPT) domain, which is an integral part of the fungal FAS, is attached at the perimeter of the α-wheel structure. Different chains in the oligomeric assembly are indicated by primes. Slices of the barrel (in orientation as indicated) show the inner volume of the barrel-shaped fold, which harbors the reaction chambers for fatty acid synthesis. The ACP is stalled at the KS under the crystallization conditions. B) Bacterial type I FAS. X-ray crystallographic model of *M. smegmatis* type I FAS (PDB ID: 4V8L).[56] A bacterial type I FAS is evolutionary related to a fungal type I FAS, and can be seen as a minimal fungal type I FAS fold lacking the PPT domains, subdomains, and insertion elements. The ACP domain could not be traced in electron density and is omitted from the model. C) Animal type I FAS. X-ray crystallographic model of porcine type I FAS (PDB ID: 2VZ8).[57] Each chain is colored as indicated in the domain overview. The TE and ACP domains are attached flexibly and could not be traced in electron density. Non-functional domains that fulfill structural tasks are indicated in italic letters in the domain overview. D) Surface representation of a full-length animal type I FAS model with domains TE and ACP from PDB IDs 1XKT[58] and 2PNG,[50] respectively. Condensing and processing wings have been moved apart by about 30 Å for better indication of inherent conformational variability (top: swinging motion; bottom: swiveling motion).

Evolution of Type I FASs:

Type I FASs occur essentially in two structural frames, having originated from separate evolutionary developments.[34,59,60] Whereas fungal and CMN-bacterial FASs form large barrel-shaped complexes (about 2.6 MDa) with enclosed reaction chambers (Fig. 2.3A&B), the animal FASs are dimers of multidomain polypeptides (about 540 kDa) in an overall open structural arrangement (Fig. 2.3C). Herbst *et al.* recently published an exhaustive review on the three-dimensional structures of type I FASs.[5]

There is clear scientific consensus that the multifunctional type I proteins evolved from the separate proteins of the type II systems.[9,60,61] During the course of evolution, two different multifunctional FAS architectures emerged: the homodimeric animal FASs and the oligomeric fungal/CMN-bacterial FASs. Multifunctional architectures provide advanced regulation. The close proximity of the domains improves enzymatic efficiency, minimizes unspecific interactions, and shelters putatively reactive intermediates in reaction chambers or clefts.[62]

It has been suggested that duplication of primordial genes and successive gene fusion at an early evolutionary state, prior to the divergence of prokaryotes and eukaryotes, gave rise to the multifunctional type I FASs.[9,61] Continuous integration of domains led to a single gene encoding a complete FAS as an evolutionary stable system. This gene fusion hypothesis is supported by evidence of FAS-like gene clusters, operons, or partially fused genes in present-day organisms; examples include the *E. coli fab* gene cluster, encoding a type II FAS,[63] the *Rhizobium meliloti fix-23* gene cluster, encoding a FAS-like polypeptide,[64] and the *Mycobacterium tuberculosis* mycocerosic acid synthase, which is encoded by a single gene and shares significant sequence and structure similarity to an animal type I FAS.[65,66] All three examples show sets of clustered or fused genes with a domain order similar to that found in an animal FAS.

A remarkable sequence and functional similarity can also be found for animal FASs and polyketide synthases (PKSs).[5,67] A fully methylating iterative PKS was identified as a common evolutionary ancestor, underlined by the non-functional evolutionary relic of a methyltransferase (MT) domain in the animal FAS scaffold that only fulfills structural tasks.[57,60]

Fungal and CMN-bacterial FASs show a domain arrangement different from that of animal FASs. Phylogenetic and structural characterization identified bacterial mono-

functional proteins from fatty acid and polyketide metabolism as ancestors. Interestingly, these proteins already possessed some of the implemented scaffolding elements that are now involved in stabilizing the intricate barrel-shaped complex. Domains were integrated until a single-gene multidomain precursor resembling a CMN-bacterial-like FAS emerged. Further scaffolding mainly occurred in the fungal protein, likely driven by advantages in catalytic efficiency originating from an increased rigidity of the barrel-shaped structure. The last fungal common FAS ancestor was encoded on one single gene. At a later evolutionary stage, splitting into two genes at various sites occurred. This led to diversified fungal FAS systems with gene topological variants.[5,59]

Alternative scenarios hypothesize that multifunctional FASs first emerged in eukaryotes as products of intronic recombination and at later stage spread into porkaryotes through horizontal gene transfer. Although this alternative evolutionary route cannot be entirely discounted, the gene fusion hypothesis is today the dominant doctrine.[5,9,59,61,68]

Animal FASs:

The structural characterization of FASs was particularly interesting, as an identical chemistry for the synthesis of fatty acids is embedded in extraordinarily different structural frames. Fungal and CMN-bacterial FASs form highly symmetric barrel-shaped complexes, while the mammalian FASs I build large X-shaped homodimers.

The animal FAS consists of two multifunctional polypeptide chains, each of about 270 kDa.[9] Owing to the open fold and conformational variability, structural characterization took considerably longer than for the fungal and CMN-bacterial FASs, and was initially driven by mutational analysis.[69,70] A cryo-electron microscopy (EM) study on human FAS was an initial milestone in providing a low-resolution map.[71,72] The breakthrough came in 2006, when the structure of porcine FAS was determined by X-ray crystallography to 4.5 Å and subsequently to 3.2 Å resolution.[57,73] In neither of these structures could the loosely tethered ACP and TE domains be traced in electron density. However, these domains had previously been structurally characterized as isolated proteins.[50,58]

The X-shaped animal FAS structure can be separated into a condensing and a processing part, connected by an intertwined short linker. The condensing part is named after the condensing KS domain, which forms the central dimeric unit. The KS do-

main is connected to the malonyl/acetyl transferase (MAT) through the linker domain (LD). The processing wing consists of the KR, DH, and ER domains for the processing of the β-carbon. The two wings do not interact permanently through interfaces, but undergo large relative positional changes, as initially observed in a negative stain EM and recently shown by high-speed atomic force microscopy.[10,11] X-ray structures have also been determined for the condensing wings of human FAS and recently for murine FAS (mFAS) with bound substrates.[74,75] Computational molecular docking studies on a human FAS model analyzed the animal type I FAS substrate shuttling and shed light on the overall conformational dynamics of this protein.[76]

Conformational Dynamics:

As aforementioned, all type I FASs carry a covalently attached ACP domain as common feature that is responsible for shuttling substrates and intermediates to the active sites. Productive fatty acid synthesis requires the physical interaction of ACP and the enzymatic domains, which means that ACP and/or enzymatic domains have to undergo large positional variations during a functional cycle. This specific feature of carrier-mediated substrate-shuttling leads to a pronounced conformational concept in type I FASs that is beyond normal local conformational changes at active sites.

ACP-mediated substrate shuttling is best understood in fungal FASs, where the ACP domain is located in the inner compartment of the barrel and doubly tethered by two flexible linkers.[77] While in fungal FASs the barrel-shaped structure is rigid showing almost no positional variations of enzymatic domains, the ACP domain is highly mobile and interacts with enzymatic domains for catalytic turn-over.[78,79] In contrast, the animal FASs have entirely different conformational properties.

Early insight into conformational flexibility of animal FASs was achieved by a fantastic study of Gordon G. Hammes and co-workers in 1986.[54] Authors have modified a FAS with fluorophores at active sites or via fluorescent NADPH to determine geometrical and dynamic features. More detailed insight on the conformational dynamics of animal FASs was finally given by a recent negative stain EM study on rat FAS.[11] Here, an ensemble of structures was identified that, when considering these structures as accessible by structural variability, characterize animal FASs as highly conformationally dynamic proteins. This ensemble of structures was just recently verified by high-resolution atomic force microscopy.[10]

ACP-mediated substrate shuttling in animal FASs does not just involve ACP positional variability, but also enzymatic domains of the catalytic body (Fig. 2.3D). In the current view, the purpose of ACP-mediated substrate shuttling is a concentration effect in fatty acid synthesis. The covalent attachment of substrates and intermediates to ACP allows the compartmentalization of fatty acid synthesis, achieving virtual high concentration in the millimolar range.[77,79] A programming of ACP shuttling in a way that ACP movement is directed towards the enzymatic domain of the consecutive synthetic step appears impossible. Rather the ACP mobility is stochastic but constrained by linkers and the conformational dynamics of the catalytic body.[68]

The basic catalytic cycle of fatty acid synthesis (elongation cycle) includes five enzymatic activities (catalyzed by domains MAT, KS, KR, DH and ER). Each catalytic step requires a dedicated conformational state in which the ACP is in physical contact with the respective enzymatic domain. Several conformational states of animal FASs – recently revealed by single-particle EM and high-speed atomic force microscopy – were also correlated with catalytic states of a fatty acid cycle.[10,11] Analysis on the structural variability of the animal FASs was further performed with mutants loaded with various intermediates.[57] The following insight on animal FAS's conformational variability was received:
(a) The swinging motion mutually induces open and closed states at the reaction clefts at both sides of the protein with potentially different catalytic activities. The closed conformation is required for ACP loading and KS-mediated chain elongation, while the open conformation allows β-keto-modification of the condensation product. The TE domain is excluded from the reaction space in the closed conformation so that TE-mediated hydrolysis of the acyl-chain for termination is just possible in an open conformation. (b) During β-keto-modification, local asymmetry appears in the processing wing of animal FAS. Asymmetry was particularly pronounced in a mutant that accumulated intermediates due to an impaired catalysis of a mutated DH domain. Structures revealed an opening in the processing part that mutually appears in the reaction clefts as the protein swings. The opening might facilitate ACP access to DH and ER active sites (local asymmetry in processing wing is not shown in figure). (c) Entry channels to active sites are located at opposite faces of the proteins, which would require the ACP to shuttle substrates and intermediates back and forth through the cleft. In the proposed functional model of animal FASs, the ACP domain does not pass the cleft. Rather the swiveling motion of the catalytic body of FAS drags the ACP to both faces of the FAS structure. Such large conformational dynamics (swiveling of up to 180° relative postioning) would also explain mutational and crosslinking studies

suggesting that ACP of one protomer can interact with the MAT and/or KS of the other protomer.[69,70]

Chain-Length Control:

As mentioned above, fatty acid biosynthesis underlies strict regulation of chain length. In a nutshell, the process of chain-length regulation can be understood as a competition between the output of a chain elongating C–C bond forming reaction and a terminating fatty acid release function. At the end of each cycle in the iterative process, the synthesizing enzymes need to "decide" whether the growing chain is to be elongated through another cycle or released as the "mature" fatty acid. Recent research has shed light on the factors determining fatty acid chain length and has also achieved control over chain length for the production of the technologically interesting short-chain (C_4-C_8) and medium-chain(C_{10}-C_{14}) fatty acids (see ChemBioChem Review).[6]

Type I FASs are ideal proteins in biotechnological applications because the compartmentalized reaction sequence provides high catalytic efficiency and allows linear synthetic pathways with little branching.[80,81] Estimated from the dimensions of the barrel-shaped structure and with the assumption of full loading of ACPs, fungal *de novo* fatty acid synthesis runs at effective concentrations of active sites and covalently tethered substrates/intermediates of 1.8 mM, when assuming full occupancy (see Fig. 2.3A). An analogous calculation on animal FASs, abstracted as cylindrical reaction spaces, gives an effective concentration of 1.2 mM (see Fig. 2.3C).[82] For the *E. coli* type II FAS system of separate enzymes, copy numbers range from 10,000 for malonyl transferase FabD to 23,000 for dehydratase FabA.[83] If an average volume of 2.5 μm^3,[84] is assumed for the *E. coli* cell, synthesis of fatty acids in *E. coli* proceeds at significantly lower molar enzyme concentrations of about 0.007 to 0.016 mM. Because the ACP domain in a type I system holds substrates and intermediates within the reaction compartment, biosynthetic pathways are highly channeled by proximity and do not suffer from side reactions.[85]

Attempts to engineer the product spectra of animal type I FASs have mainly harnessed the pool of substrate-specific TEs. TEs were either applied as separate *trans*-domains or swapped with the native TE domain. In an animal type I FAS, only 9 % of the total sequence is invested in linkers, and this makes this fold highly susceptible to domain swaps (see Fig. 2.3C).[75] Just recently the modular nature of the animal FAS fold was demonstrated in a protein engineering study.[86] According to this concept,

the animal type I FAS was shown to produce short-chain fatty acids when the cognate TE domains were exchanged for the short-chain, acyl-ACP-specific type II TEs.[87,88] Specifically, the TE of rat type I FAS was swapped with human TE type II, thus shifting the product spectrum towards short-chain fatty acids,[87] and the TE of human type I FAS was swapped with human TE type II or the plant TE FatB1 from *Cuphea palustris* to induce C_6/C_8 and C_8/C_{10} production, respectively, in baker's yeast.[88] In the latter work, type II TE or FatB1 were also co-expressed with human FAS truncated by the TE domain. Although the interaction *in trans* proved less efficient, data reveal that the ACP of the animal type I FAS fold is sufficiently exposed for such *in trans* engineering strategies.
Intriguingly, simply the variation of the TE domain linker length in an animal type I FAS can influence the product spectrum. Whereas addition of residues or deletion of nearly half of the ACP-TE interdomain linker showed no effect, further shortening resulted in reduced FAS activity and shifted the product spectrum towards longer chain lengths of C_{18} and C_{20}.[87] Changes in linker length likely affect the domain mobility and the interplay of ACP and the terminating TE, so the acyl-ACP is steered towards the elongating KS-mediated condensation. It should be noted that any domain swap approach is invasive because it introduces non-cognate interactions between ACP and the enzymatic domain.

Microbial type I FASs have also been engineered for the production of short-chain acyl-CoA. In the microbial type I FAS fold, 50 % of the total sequence is invested in linkers and non-catalytic insertion elements, leading to an overall intricate protein complex of the size of the ribosome (diameter of about 25 nm) (see Fig. 2.3A & B). Linkers and insertion elements are responsible for the assembly and the stability of the barrel-shaped structure.[89] As the enzymatic domains are tightly embedded in the scaffolding matrix, engineering of the protein through domain swapping is difficult and may be envisioned just for highly homologous domains. The fungal type I FASs from baker's yeast and from *Yarrowia lipolytica* have been modified by insertion of the TE domains into the multienzyme complex[90] or by exchanging the malonyl/palmitoyl transferase (MPT) for type II TEs specific for short-chain acyl-ACPs, respectively.[91] It should be noted that the exchange of the bifunctional MPT domain for a TE domain is problematic because such a swap does not specifically target the termination function of MPT as intended but also impairs the malonyl loading function.

The synthesis of short-chain-length acyl-CoAs and fatty acids has recently also been achieved through mutation of selected amino acids in active sites and binding channels

of the domains KS, AT and MPT of baker's yeast type I FAS[92] and corynebacterial type I FAS.[93] A similar approach has also succeeded in producing C_{14}-CoA in *Y. lipolytica* type I FAS. Here, an isoleucine-to-tryptophan mutation at the bottom of the KS-binding pocket spatially constrained the incorporation of long acyl chains.[94] Engineering of the transferase domains can support KS design.[92,93] Reduced affinity for the malonyl moiety increased the export rate of acyl chain transfer and promoted pre-mature termination of synthesis. The direct engineering of the MPT domain to improve the release reaction – by increasing the hydrophobicity of the binding channel in order to favor short acyl chain export, for example – failed in this study.[93] The difficulty in engineering the MPT domain lies in the generally high conformational variability of the AT binding site, and also in lack of structural information on MPT domains in complexation with acyl chains.

Finally, the acetyl-loading AT domains of FASs from baker's yeast and *Corynebacterium ammoniagenes* were mutated to provide an alternative exit tunnel for medium-length fatty acids.[95] Just recently, it was shown for fungal and CMN-bacterial type I FASs that single mutations in the domain-domain interface of ACP and the elongating KS domain can severely shift chain-length spectra.[96] The product's acyl chain length correlated with the strength of the interactions between the ACP and other enzymatic domains of the mutated FAS determined by computational docking. This study demonstrates that targeted mutations distant from the active site can be used to modulate the product spectrum. Direct engineering of fungal and CMN-bacterial type I FASs is beneficial, because these FASs deliver short-chain-length products as activated acyl-CoA esters that can be directly processed to alkyl derivatives. TE-based chain-length control requires reactivation of the free fatty acids to the acyl-CoA esters in an ATP-consuming process.[97]

Engineered FASs have usually been embedded in microbial hosts, preferentially *E. coli, Saccharomyces cerevisiae* and *Y. lipolytica*, in which the metabolisms were optimized towards high product titers. Metabolic engineering aims either to increase the supply of acyl-CoA substrates and the reducing co-factor NADPH, or to block competing pathways such as β-oxidation. Because fatty acids can also be toxic to the host cell, efflux pumps have also been inserted. Metabolic engineering approaches to increase product titers have been reviewed elsewhere and are not considered here.[98–103]

2.4 Underlying Methods for the Proposed Study

Time Scale of Conformational Dynamics in Proteins:

Dynamic fluctuations in proteins cover a wide range of time scales from bond vibrations in the femtosecond range to positional variations of amino acids, secondary structure elements as well as domains in the microsecond to millisecond time scale. Larger conformational changes occur in proteins that are composed of several domains. For animal FASs, we can receive a rough estimate of dwell times of main conformational states when considering the following:
The production of one molecule of palmitic acid takes approximately 1 s as the turnover number of the overall reaction is calculated to k_{cat}<1.3 s^{-1}.[54] One elongation cycle consists of six enzymatic steps (loading by MAT, ping-pong mechanism of KS, three β-carbonyl modifying steps), and elongates the growing fatty acid chain by a C_2 unit. Thus, for the production of palmitic acid (C_{16} fatty acid), seven elongation cycles are passed (giving 42 enzymatic steps).
When assuming that (i) an enzymatic step (as defined here) equals a conformational state, (ii) the transition between states is fast (as implied from cryo-EM data on fungal FAS), and (iii) each conformatinoal state occurs at the same rate, we receive an average dwell time of roughly 24 ms per conformational state. Non-productive interactions might be relevant, but are not considered. I am aware that this assumption is just valid for giving a rough estimate on time scales of conformational states.

Methods to Study Dynamics:

Conformational dynamics of proteins have been analyzed with structural biological, spectroscopic and theoretical methods (quantum mechanical to molecular dynamic calculations).[104]
The primary result of an X-ray crystallographic experiment is a static structure of a macromolecule, and is *per se* not designed for analyzing conformational dynamics. Especially proteins of dynamic nature are challenging to crystallize and diffract poorly; these problems scale rapidly with size of the biomolecule. Although, some interpretations on conformational dynamics can be derived with the help of crystallographic temperature factors (B-factors), and from snapshots of the different static crystal structures, these cannot provide an entire trajectory of the conformational dynamics adopted by proteins, especially in real-time.
Nuclear magnetic resonance (NMR) spectroscopy is much better suited for analyzing

protein dynamics, but the molecular weight limitation of the methods restricts NMR spectroscopy to rather small proteins.[105]
Recent technological progress has developed EM to a powerful method in structure determination of large proteins.[106] As a single-particle method, it can give insight into conformational dynamics when particles are classified during manual and automated sorting,[11] or initial electron maps are analyzed for principle components.[107] Although cryo-EM analysis allows structure analysis at even 3-4 Å resolution,[106] detailed insight into structural changes, as possible with X-ray crystallography and NMR spectroscopy, remains mainly beyond the resolution limit of this method.

While X-ray crystallography and EM deliver static snapshots, spectroscopic methods – e. g., fluorescence spectroscopic methods relying on fluorescence polarization (FP) and Förster resonance energy transfer (FRET), as well as electron paramagnetic resonance (EPR) spectroscopy – can deliver a continuous spectrum of protein conformational dynamics in real-time. As conformational motions in proteins are stochastic and synchronization of populations are hardly possible, spectroscopic methods are particularly powerful when applied at single-molecule level. Very valuable information on the conformational dynamics of proteins and protein complexes has been collected by single-molecule FRET (smFRET).
As an early highlight of smFRET, Diez *et al.* succeeded in observing and analyzing the rotary movement of the γ-subunit during proton-powered ATP synthesis by single F_0F_1-ATP synthase.[108] The proteins were embedded in liposomes and analyzed while traversing the confocal detection volume. Fundamental questions concerning the working mode of F_0F_1-ATP synthase were experimentally addressed, e. g. whether the γ-subunit rotates continuously or stepwise, or whether ATP synthesis and hydrolysis involve opposite rotational movement of the γ-subunit. In a similar approach (although using a prism-based total internal reflection fluorescence (TIRF) microscope), Wang *et al.* reported on the conformational dynamics of a potassium-gated voltage channel.[109]

An integral aspect of spectroscopic methods is the modification of proteins with labels. Except in cases where the intrinsic fluorescence of the natural amino acid tryptophan is used, fluorescence spectroscopy requires the modification of the proteins with dyes. Conventional labeling techniques employing naturally occuring or mutationally introduced cysteines at user-defined sites and labeling them via maleimide chemistry are increasingly replaced by genetic encoding of ncAAs through the codon suppression technology since they are not applicable to proteins that rely on cysteines for their natural function.

ncAAs carry an orthogonal functional group, which can be used to post-translationally modify proteins site-specifically and to add spectroscopic labels. Furthermore photoreactive ncAAs can be used in crosslinking studies to map protein-protein interactions or to attach large, naturally occuring post-translational modifications.

At the start of this thesis, the labeling of proteins at the size and complexity of animal FAS has not been reported. We focused on mFAS which is highly homologous to human FAS. While we clearly aim for analyzing full-length FAS by smFRET, we will be able to also design lower complex mFAS subconstructs that provide insight into some of the fundamental questions addressed in the aim of this thesis and will also be used in other methods.

2.5 Aim of the Thesis

Parts of this chapter have previously been published in: Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878. For individual contribution to the writing process as well as copyright and Creative Commons license, please consult the Statement of Personal Contributions.

During the last decades landmark scientific work has revealed the structure of type I FAS systems and achieved deep understanding about the molecular foundations of *de novo* fatty acid synthesis. We seek for directly monitoring substrate-shuttling and conformational dynamics of animal FAS, ultimately receiving a quantitative description of its kinetic and structural dynamic properties during fatty acid synthesis.
This approach will allow answering fundamental questions to the working mode of FAS and multidomain proteins in general:
(a) Is carrier domain-mediated substrate shuttling based on stochastic or defined/programmed conformational changes? (b) How do productive domain-domain interactions differ from non-productive ones? (c) Do conformational dynamics of the protein change over the several cycles of fatty acid synthesis (with substrate load of carrier domain)? (d) How do conformational dynamics change when mutating the protein in amino acids proximal and distal to the active site, or when mutating domain-domain interaction sites? (e) How do conformational dynamics change when changing physical (e. g. temperature, ionic strength of medium) or chemical conditions and properties (e. g. substrate concentration, inhibitor binding)? (f) How much conformational flexibility is

exhibited within and between different modules? What is the degree of conformational flexibility within and between the modules? (g) What is the impact of conformational freedom on biosynthesis? (h) Can we map the movement of an ACP? (i) What are the time frames of domain-domain interactions? (j) Do affinity and catalytic efficiency correlate? (k) Is there a substrate driven directionality of the ACP domain?

The access to information on the dynamic behavior of large proteins is usually hindered, as spectroscopic methods require the site-specific attachment of biophysical probes. A powerful emerging tool to tackle this issue is amber codon suppression. So far, its application on large and complex multidomain proteins of MDa size has not been reported. Herein, we aim to systematically investigate the feasibility to introduce different ncAAs into a 540 kDa homodimeric type I FAS by genetic code expansion with subsequent site-directed labeling.

Successful introduction of biophysical probes into a molecular machine such as FAS may pave the way to understand the fundamental questions of ACP-mediated substrate shuttling. For example, the conformational variability, which is a primary intrinsic property required for efficient interplay of all catalytic functionalities, can be investigated. The information that will be obtained by such studies will help to engineer these megasynthases towards altered product spectra.

3 Results

3.1 Method Development

Parts of this chapter have previously been published in: Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878. For individual contribution to the writing process as well as copyright and Creative Commons license, please consult the Statement of Personal Contributions.

3.1.1 Preliminary Work

In the course of my master's thesis, a set of methods and protocols that build a solid basis for this thesis have already been established. These include the molecular cloning of the pAC^{U} amber codon suppression vector, incorporation of *para*-azido-L-phenylalanine (AzPhe) and propargyloxy-L-phenylalanine (PrPhe) in the stand-alone ACP, and a prototype of the reporter assay. Furthermore, the supervisor of my master's thesis and later on colleague Dr. Alexander Rittner conducted important preliminary work in his doctoral thesis. Although mFAS is a complex eukaryotic protein, he achieved the recombinant expression of mFAS in *E. coli* at excellent yields and high protein quality. Several subconstructs (truncated variants) of mFAS have been successfully isolated at high yields as well.

3.1.2 The Amber Codon Suppression Toolbox

By the time we started this project in 2014, several amber codon suppression vectors had been reported,[12,13] but only a very limited number was publicly available. The lab of GCE pioneer Peter G. Schultz initially provided two of their pEVOL plasmids for incorporation of AzPhe and *para*-benzoyl-L-phenylalanine (BzPhe) on the nonprofit plas-

mid repository Addgene.[4] These plasmids encode the orthogonal *Mj*TyrRS/tRNA^{Tyr} pair for amber codon suppression in *E. coli*.

At first, we were driven by the purpose of inserting azide or alkyne functional groups in our protein of interest to use them in copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) for the chemoselective bioconjugation with fluorophores.[110] While the available pEVOL-pAzF plasmid[111] was suitable for incorporation of AzPhe, no amber codon suppression vector was available for incorporation of alkyne functional groups. Right from the beginning, we intended to set up a broad range amber codon suppression toolbox in our lab for the versatility of this powerful method. The incorporation of a vast variety of ncAAs with help of evolved aaRSs had been reported. The aaRSs that had been evolved over a two-step selection process carried multiple point mutations, which made them specific for certain ncAAs. The pEVOL vector encodes two copies of the aaRS in order to supply a sufficient amount of orthogonal aaRSs in the expression host.[13] This posed a major problem to our conventional cloning technique, introducing multiple point mutations site-specifically in a protein gene.

The Schultz lab has also published a successor version of the pEVOL vector, named pUltra.[12] The pUltra vector encodes only one copy of aaRS under a stronger promoter, and represents a very puristic vector for efficient amber codon suppression. To keep up with the times, and moreover because this puristic amber codon suppression vector allowed for our conventional and reliable cloning technique, we decided to use the pUltra vector. Unfortunately, at that point, no pUltra vector was publicly available, so we constructed our own pUltra vector, pAC^{U}. The benefits of the pUltra vector as compared with pEVOL not only lie in easier cloning but also in its easy handling. Most genes of interest stand under an IPTG inducible promoter, as does the aaRS gene in pUltra. Expression of the protein of interest and the orthogonal aaRS is induced simultaneously by addition of IPTG in the expression medium. Some might argue that independent induction allows for more fine-tuned expression and efficient amber codon

[4]By now, the GCE technology has gained more and more recognition for its power and versatility and is utilized in an increasing number of scientific studies.[14,21] GCE material is becoming more generally available. The nonprofit plasmid repository Addgene is encouraging scientists to provide the scientific community with amber codon suppression vectors. In the past few years, their collection for GCE has reached 46 entries of plasmids (number certainly further increasing), comprising codon suppression with miscellaneous orthogonal aaRS/tRNA pairs, in different organisms, and aaRSs evolved for the incorporation of various ncAAs. Nowadays, there is no longer necessity to clone an amber codon suppression vector from scratch. Furthermore, several bacteria strains are available, which have been modified to enhance amber codon suppression.

suppression, but in our hands both vectors performed similar.

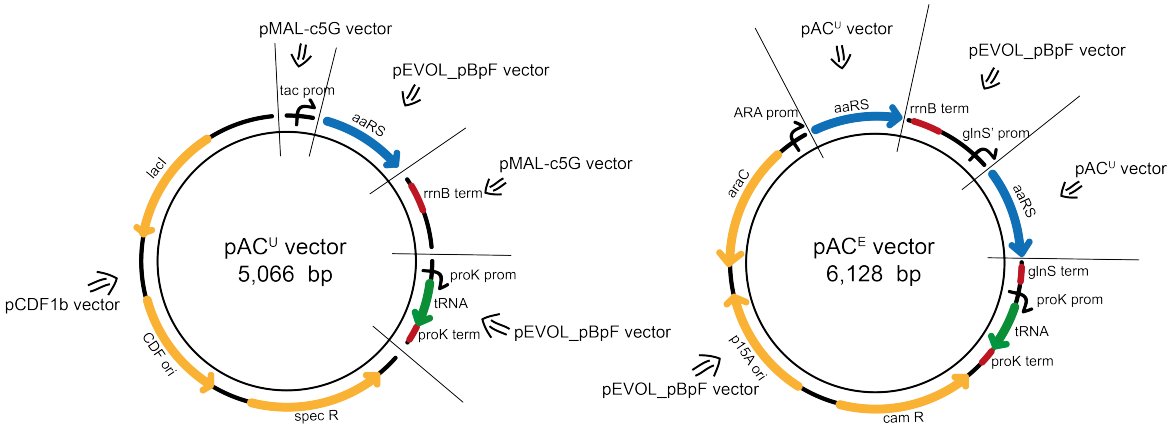


Figure 3.1: Cloning scheme of pAC^{U} and pAC^{E} vectors. The pAC^{U} vectors were constructed using the commercially available vectors pCDF-1b, pMAL-c5G and pEVOL_pBpF [112] (for *Mj*TyrRS/tRNA^{Tyr} pair), pJZ (for *Mm*PylRS/tRNA^{Pyl} pair) or pAcBac1.tR4-MbPyl [113] and pDULE-ABK [114] (for *Mb*PylRS/tRNA^{Pyl}) as templates. PCR fragments were assembled with help of complementary primer overhangs in MegaPrimer PCRs and subsequently cloned into the backbone through ligation independent cloning. Multiple point mutations were introduced to create a set of genes encoding different evolved aaRSs. The amber codon suppression library of pAC^{U} vectors was used as template to create pAC^{E} plasmids by replacing the genes encoding both copies of the aaRS in pEVOL_pBpF [112] with the genes encoding respective mutants of aaRSs, using the same cloning methods.

Molecular cloning of the pAC amber codon suppression vectors is described in detail in the Experimental Procedures 5.1.3 and just briefly outlined here. Based on pUltra and pEVOL, as had been published by the lab of Schultz and co-workers, [12,13] the suppressor plasmids pAC were constructed by assembling cassettes amplified from the commercially available plasmids pCDF-1b, pMAL-c5G and pEVOL_pBpF (which was a gift from Peter Schultz; Addgene plasmid # 31190) [112] (Fig. 3.1). PCR fragments were assembled with help of complementary primer overhang in a MegaPrimer PCR and subsequently cloned into the backbone using In-Fusion® Cloning. Multiple point mutations were introduced into the aaRS gene of pAC^{U} to create a set of evolved aaRSs, specific for certain ncAAs. The genes of evolved aaRSs were subsequently amplified from the pAC^{U} vector and cloned into the pAC^{E} vector. Three different sets of orthogonal suppressor pairs (aaRS/tRNA), derived from *M. jannaschii, M. mazei* and *M. barkeri*, were utilized in this thesis. Genes of the orthogonal pair *Mm*PylRS/tRNA^{Pyl} were obtained from plasmid pJZ, which was a gift from Nediljko Budisa, and *Mb*PylRS was obtained from pAcBac1.tR4-MbPyl (Addgene plasmid # 50832) [113] or pDULE-

ABK (Addgene plasmid # 49086)[114], which were both a gift from Peter Schultz.

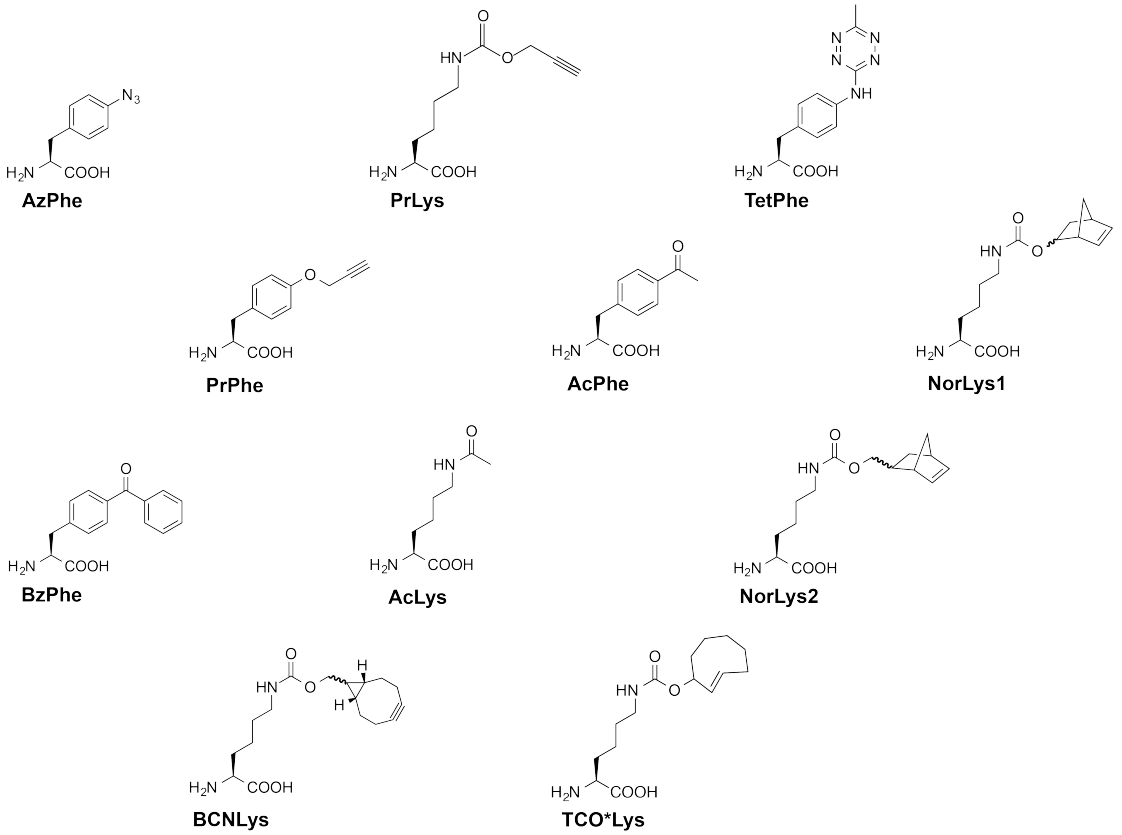


Figure 3.2: Structures of ncAAs compiled in this thesis. ncAAs feature alkyne- (PrLys, PrPhe, BCNLys) and azido-functional groups (AzPhe) for use in CuAAC or SPAAC, norbornene- (NorLys1 & NorLys2), tetrazine- (TetPhe) and strained alkene-functional groups (TCO*Lys) for use in IEDDAC, acetyl-functional groups (AcPhe & AcLys) for use in oxime formation, or photoreactive groups (AzPhe & BzPhe). Except for AcPhe, BzPhe, BCNLys and TCO*Lys, which were purchased from commercial sources, all ncAAs were synthesized in our lab.

In the course of this thesis we did not only construct the two amber codon suppression vectors pAC^{U} and pAC^{E} to incorporate substituted phenylalanines with alkyne and azide functional groups, we further broadened our repertoire of ncAAs by the addition of several new ncAAs.(Fig. 3.2) The field of GCE rapidly advanced especially by an increasing number of ncAAs, which carry fancy functional groups, e. g. for faster and more efficient chemoselective reactions. Besides CuAAC,[110] several copper-free click chemistries have been employed by the GCE technology since copper is in many cases critical for the stability of proteins (see section 3.1.5 in the Results). Examples of these are the strain-promoted alkyne-azide cycloaddition (SPAAC),[115] with azido-

and bicyclononyne (BCN)- or dibenzoylcyclooctyne (DBCO)-functional groups, and the inverse electron-demand Diels-Alder cycloaddition (IEDDAC),[116] with tetrazine- and norbonene-, BCN- or *trans*-cyclooctene (TCO)-functional groups (Fig. 3.3).

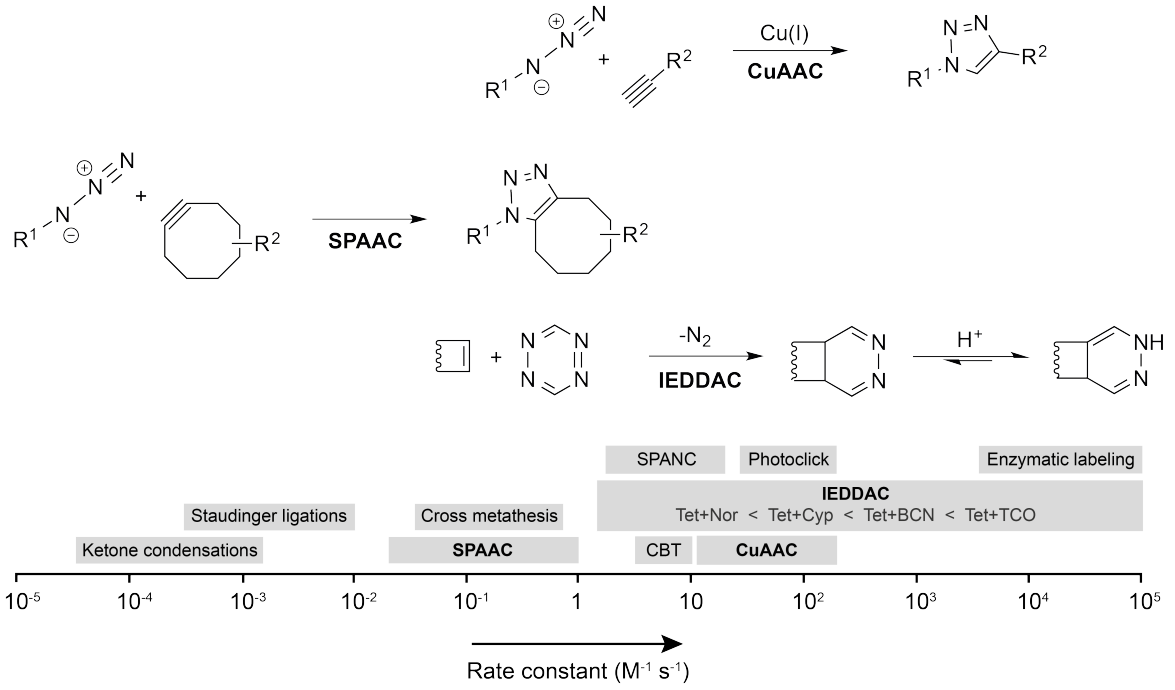


Figure 3.3: Chemoselective click reactions utilized in this thesis to modify proteins. Copper(I)-catalyzed (CuAAC) as well as copper-free chemoselective reactions (SPAAC & IEDDAC) were utilized in this thesis. ncAAs bearing the functional groups depicted in this scheme were incorporated into proteins through amber codon suppression and reacted with functionalized dyes or spin labels as reaction partners. Rate constants of chemoselective reactions are compared in the bottom scheme. The use of cooper is in many cases critical for protein stability. Therefore, the copper-free chemoselective reactions (SPAAC & IEDDAC) were preferably chosen. The IEDDACs have highest potential to yield a sufficient DOL, as their rates approach those of enzymatic labeling. Figure adapted from reference.[20,118]

Most labeling reactions are relatively slow with rate constants ranging from 10^{-4} to $1\,M^{-1}\,s^{-1}$ (Fig. 3.3). In comparison, enzyme-mediated protein labeling – peptide- and protein-tagging approaches such as SNAP and CLIP tagging – are four to five orders of magnitude faster, with rate constants ranging from 10^4 to $10^5\,M^{-1}\,s^{-1}$.[117] Photoclick reactions approach rate constants up to $60\,M^{-1}\,s^{-1}$. IEDDAC is probably one of the fastest and most exciting chemoselective reactions discovered to date, with rate constants up to $10\,M^{-1}\,s^{-1}$ for the reaction between a norbornene and tetrazines, ranging from 80 to $2{,}700\,M^{-1}\,s^{-1}$ for the reaction between a BCN and tetrazines, and ranging from 5,000 to $17{,}000\,M^{-1}\,s^{-1}$ for the reaction between a TCO and tetrazines. These re-

action rates approach those reported for enzymatic labeling.[20,118]

Our toolbox comprises the ncAAs listed in Fig. 3.2, which can mainly be purchased from commercial sources. As this project consumed high amounts of ncAAs, we further established organic chemical synthesis of several ncAAs (for syntheses of respective ncAAs, see Chemical Synthesis 5.2 and Heil *et al.*[119]). Table S1 lists all amber codon suppression plasmids and the orthogonal pairs of evolved aaRSs/tRNAs encoded on them. Out of a growing repertoire of ncAAs that can be introduced by amber codon suppression, we limited our screening to eleven different ncAAs with functional groups that can mainly be used for bioconjugation or crosslinking.[118] Additionally, we also tested incorporation of the ncAA AcLys, as the acyl-functional group can be bioconjugated in oxime formation,[120] and acetylation of lysines naturally occurs as post-translational modification in animal FAS.[121] Further, we compared the two common suppressor vectors pUltra and pEVOL, published by Schultz and co-workers,[12,13] and several evolved aaRS/tRNA pairs from *M. jannaschii, M. mazei and M. barkeri* for their performance.

In order to achieve efficient amber codon suppression, different *E. coli* strains were tested for expression. Very low expression levels of the *Mm*PylRS were observed as also has been reported.[122,123] Our efforts to increase the solubility and expression of *Mm*PylRS in *E. coli* failed. However, some amber codon suppression efficiency was observed. An optimized *Mj*tRNA^{Tyr}, which has a repaired mismatch in the acceptor stem, slightly increased efficiency (cytidine to adenosine nucleotide mutation at 5'-end inhibited interaction with *Mj*TyrRS). In the course of this thesis, we chose *E. coli* strain BL21-Gold(DE3) as expression host. In the following section, the reporter assay is addressed, which was established as screening method for our amber codon suppression toolbox.

3.1.3 The Reporter Assay

Screening of the Amber Codon Suppression Toolbox

To identify the optimal pair of suppression system and ncAA we established a reporter assay. The screening was performed on a fusion construct of ACP from mFAS with GFP (ACP-GFP) placing the amber mutation in a disordered and non-conserved loop region at the Leu54 site (equivalent to Leu2166 in mFAS, UniProtKB accession code P19096), using the homologous rat ACP structure (PDB ID: 2PNG)[50] as template. A Met72Leu mutation was introduced in the ACP-GFP reporter protein to prevent an

alternative translation start and the thereby arising background fluorescence of GFP expression. Incorporation efficiency was read out at 2 mL scale by recording the fluorescence of *E. coli* cell cultures expressing different ACP-GFP mutants. Cultures lacking the ncAA in the medium were taken as negative controls to determine the background signal.

Negative samples showed a fluorescence level of up to 4 % of the wild-type reference, with few exceptions: suppressor plasmid pAC^{U}_CNF showing a high fluorescence level of 16 % ± 8 %, and suppressor plasmids encoding aaRSs evolved for AcPhe and BzPhe incorporation showing fluorescence levels of 8-21 % in negative samples (Fig. 3.4). High incorporation efficiencies were observed for AzPhe (37 % ± 9 %) and NorLys2 (23 % ± 9 %), as well as for AcPhe (59 % ± 7 %) and BzPhe (54 % ± 2 %), although the latter ones are afflicted by general high fluorescence levels of their negative controls. The ncAA BCNLys was incorporated with 12 % ± 2 % efficiency, all other ncAAs (PrPhe, TetPhe, NorLys1 and PrLys) showed efficiencies below 10 %, and AcLys was hardly incorporated at all.
Comparing the suppressor plasmids pAC^{E} and pAC^{U}, the plasmid pAC^{E} seemed to be slightly more efficient in our set-up than its pAC^{U} counterpart (Table 3.1 upper part), except for the incorporation of AcPhe, which was more efficient with the pAC^{U} vector. The D286R mutation in the *Mj*TyrRS,[124] which was reported to have a beneficial effect, did not improve incorporation efficiencies in our hands (Table 3.1 lower part). Comparing the two orthogonal systems, the *Mj*TyrRS seemed to be more efficient in our set-up than *Mm*PylRS or *Mb*PylRS (e. g. 37 % ± 9 % pAC^{U}_AzPhe (*Mj*) vs. 18 % ± 5 % pAC^{U}_NorLys (*Mm*)). Especially the *Mb*PylRS did not perform well, which might be due to a wrong gene sequence. We also tested two less specific suppressor vectors, which incorporate multiple ncAAs. While the suppressor plasmid pAC^{U}_Phe-derivatives of *M. mazei* failed to incorporate any phenylalanine derivatives, the suppressor plasmid pAC^{U}_CNF from *M. jannaschii* showed high incorporation efficiencies of 39 % ± 10 % but suffered from relatively high fluorescence of the negative control (16 % ± 8 %), indicating unspecific incorporation of endogenous amino acids. A compilation of all results from the reporter assay screening can be found in Fig. S1.
Additionally to the results shown here, incorporation of *trans*-cyclooct-2-en-L-lysine (TCO*Lys) was tested in a laboratory course hosted by the Unnatural Protein Facility at the Oregon State University following the protocols of the Ryan Mehl lab using autoinduction media. Unfortunately, no successful incorporation of TCO*Lys was observed in our reporter protein ACP-GFP.

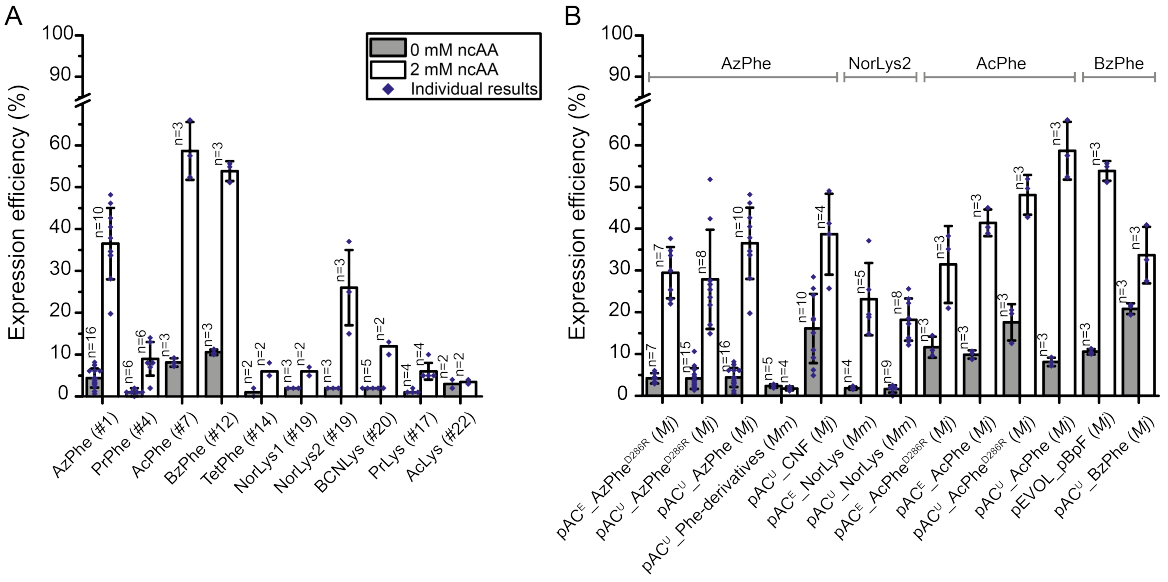


Figure 3.4: Amber codon suppression at site Leu54 in the ACP-GFP fusion construct screened in the reporter assay. A) Best expression efficiency of different ncAAs. Respective plasmids used for incorporation of ncAAs are indicated by plasmid number (#; listed in Table S1). B) Comparison of some representatives of the screening. A compilation of all results from the reporter assay can be found in Fig. S1. Expression efficiency is read out by GFP fluorescence of 2 mL *E. coli* cell cultures and compared to wild-type reference (taken as 100 %). For incorporation, 2 mM ncAAs were supplemented to the medium. Cultures lacking ncAAs were taken as negative control to determine background signal. Dots refer to values of the biological samples. The averages of biological samples are plotted together with standard deviations. Technical errors were below 10 %.

Table 3.1: Comparison of selected representatives in the reporter assay screening.

Suppressor Plasmid	**Efficiency**	**Suppressor Plasmid**	**Efficiency**
Comparison of pAC^{E} & pAC^{U}			
pAC^{E}_AzPheD286R	30 % ± 6 %	pAC^{U}_AzPheD286R	28 % ± 12 %
pAC^{E}_NorLys	23 % ± 9 %	pAC^{U}_NorLys	18 % ± 5 %
pAC^{E}_BzPhe	54 % ± 2 %	pAC^{U}_BzPhe	34 % ± 7 %
pAC^{E}_AcPhe	41 % ± 3 %	pAC^{U}_AcPhe	59 % ± 7 %
pAC^{E}_AcPheD286R	31 % ± 9 %	pAC^{U}_AcPheD286R	48 % ± 3 %
Influence of the D286R mutation in *Mj*TyrRS			
pAC^{U}_AzPheD286R	28 % ± 12 %	pAC^{U}_AzPhe	37 % ± 9 %
pAC^{E}_AcPheD286R	31 % ± 9 %	pAC^{E}_AcPhe	41 % ± 3 %
pAC^{U}_AcPheD286R	48 % ± 5 %	pAC^{U}_AcPhe	59 % ± 7 %

Position Screening of ncAA Incorporation

The specific position of an amber mutation has major effects on incorporation efficiencies as it has been reported.[125] We used the most promising systems from the initial screening to compare incorporation efficiencies at different sites (pAC^{U}_AzPhe with AzPhe as the optimal result and the respective vector pAC^{U}_NorLys for NorLys2). Hence, we selected six positions in the ACP fold to test the acceptance of ncAA incorporation (Ala in the linker region between the N-terminal Strep-tag and ACP-GFP, Gly01 at the N-terminus of the mouse ACP sequence (equivalent to Gly2113 in mFAS), Leu54 in a disordered loop region (equivalent to Leu2166 in mFAS), Gln70 and Glu71 in helix 5 (equivalent to Gln2182 and Glu2183 in mFAS) and Ala79 in the linker between ACP and GFP (equivalent to Ala2191 in mFAS)) (Fig. 3.5A&B)[5].

AzPhe was incorporated with good efficiencies (in average $38\,\% \pm 1\,\%$) throughout all amber mutation sites, whereas the incorporation efficiencies for NorLys2 were strongly dependent on the respective position (Fig. 3.5C). Best incorporation efficiency for NorLys2 was gained for the amber mutation site Gly01 at the N-terminus with $39\,\% \pm 13\,\%$. The amber mutation site Leu54 in the disordered loop region of ACP showed $16\,\% \pm 6\,\%$ incorporation efficiency for NorLys2, whereas amber mutation site Gln70 in the last helix of ACP showed no incorporation at all. All other amber mutation sites showed incorporation efficiencies below 10 %. Higher concentrations (4 mM and 8 mM) of ncAAs in the medium seemed to slightly increase incorporation efficiency of NorLys2 and slightly decrease efficiency of AzPhe (see Fig. S2). However, no significant trend was observed. Therefore, we proceeded with a concentration of 2 mM ncAA.

Upscale of the Reporter Assay

In a next step, it was tested whether the selected conditions from the reporter assay are reproducible in larger expression cultures of 200 mL. Each culture was analyzed in fluorescence as it was implemented in the reporter assay and further evaluated by the yield of purified protein. Fluorescence data was collected similarly to the reporter assay taking a 2 mL sample of the cell culture.
The ncAA AzPhe was incorporated with overall good efficiency (in average $32\,\% \pm 4\,\%$), whereas large variations were observed for incorporation of NorLys2 at different amber mutation sites (Fig. 3.6A). Best incorporation efficiency for NorLys2 was gained for the amber mutation site Leu54 with $29\,\% \pm 3\,\%$ and no incorporation was achieved at

[5]These positions were first conceived in the master's thesis of Franziska Stegemann (see section 3.2.1)

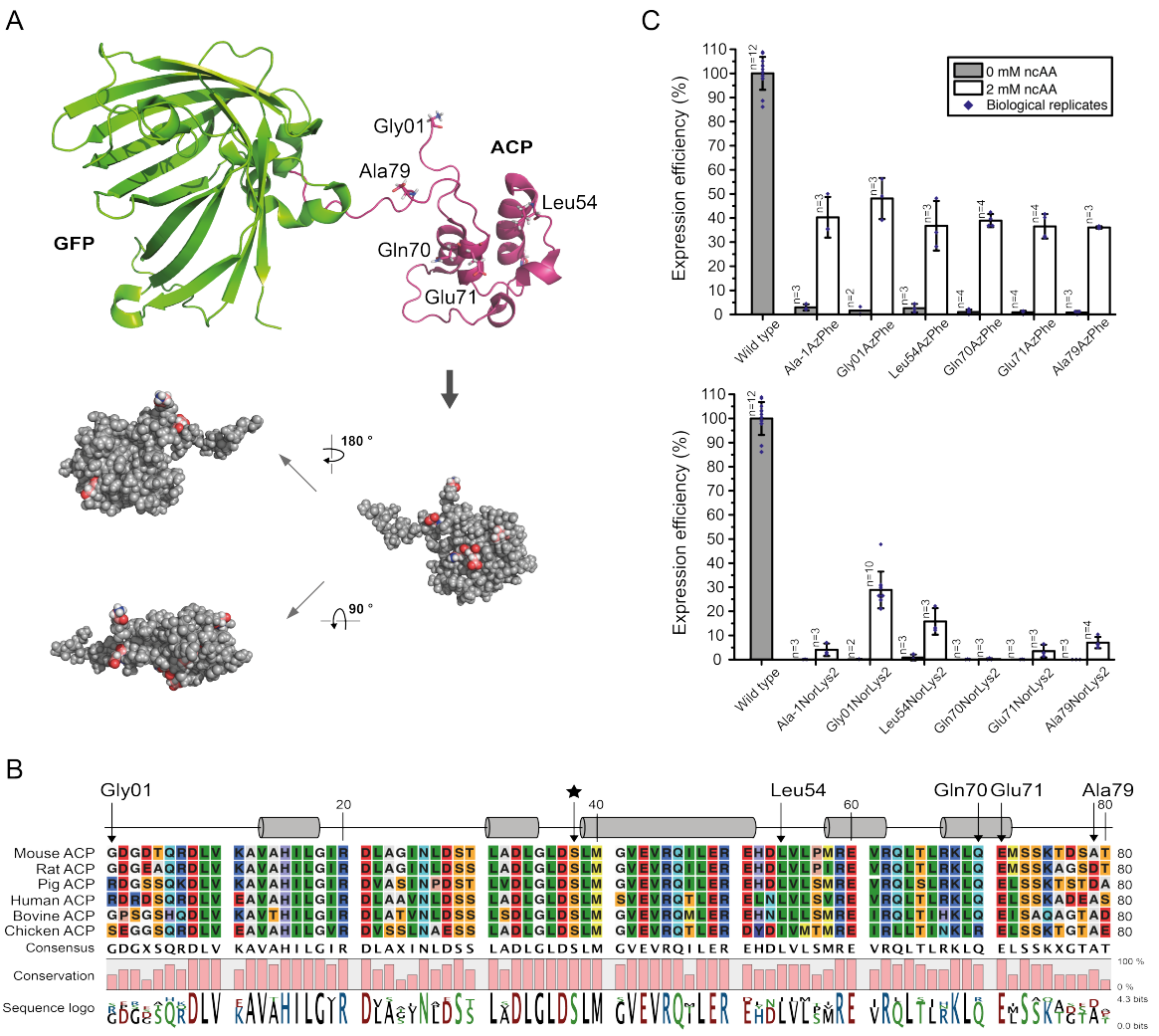


Figure 3.5: Screening of amber codon mutation sites. A) Cartoon representation of the ACP-GFP fusion construct (upper panel; pink: rat ACP (PDB ID: 2PNG) [50] and green: eGFP (PDB ID: 2Y0G) [126]) used in the reporter assay. The five amber mutation sites are labeled and depicted in stick representation (Gly01, Leu54, Gln70, Glu71 and Ala79). Different orientations of the ACP domain (shown in a sphere-filling model) demonstrate the positioning of all amber mutation sites on the surface of the domain (lower panel). Amber mutation sites are colored in red. B) Sequence alignment of six different ACP domains of animal FASs. UniProtKB accession codes: mouse FAS: P19096; rat FAS: P12785; pig FAS: A5YV76; human FAS: P49327; bovine FAS: Q71SP7 and chicken FAS: P12276. The five amber mutation sites are highlighted by arrows, and a star highlights the active-site serine residue. Secondary structure elements received from the rat ACP model (PDB ID: 2PNG) [50] are depicted (α-helices shown as cylinders). C) Expression efficiencies of six different AzPhe mutants (upper panel) and six different NorLys2 mutants (lower panel) in comparison to the wild-type reference, read out by the GFP fluorescence of 2 mL cultures of *E. coli* cells. For incorporation, 2 mM ncAAs were supplemented to the medium. Cultures lacking ncAAs were taken as negative control to determine background signal. The averages of biological replicates are plotted together with standard deviations and the distribution of individual values is indicated as dots. Technical errors were below 10 %.

amber mutation site Gln70. This data was in well agreement with the results from the reporter assay except for the incorporation of NorLys2 in Gly01 failing at larger volume while leading to best incorporation efficiencies in the reporter assay. We observed systematic higher values for NorLys2 and slightly lower values for AzPhe by GFP-fluorescence in the larger expression culture compared to the reporter assay.

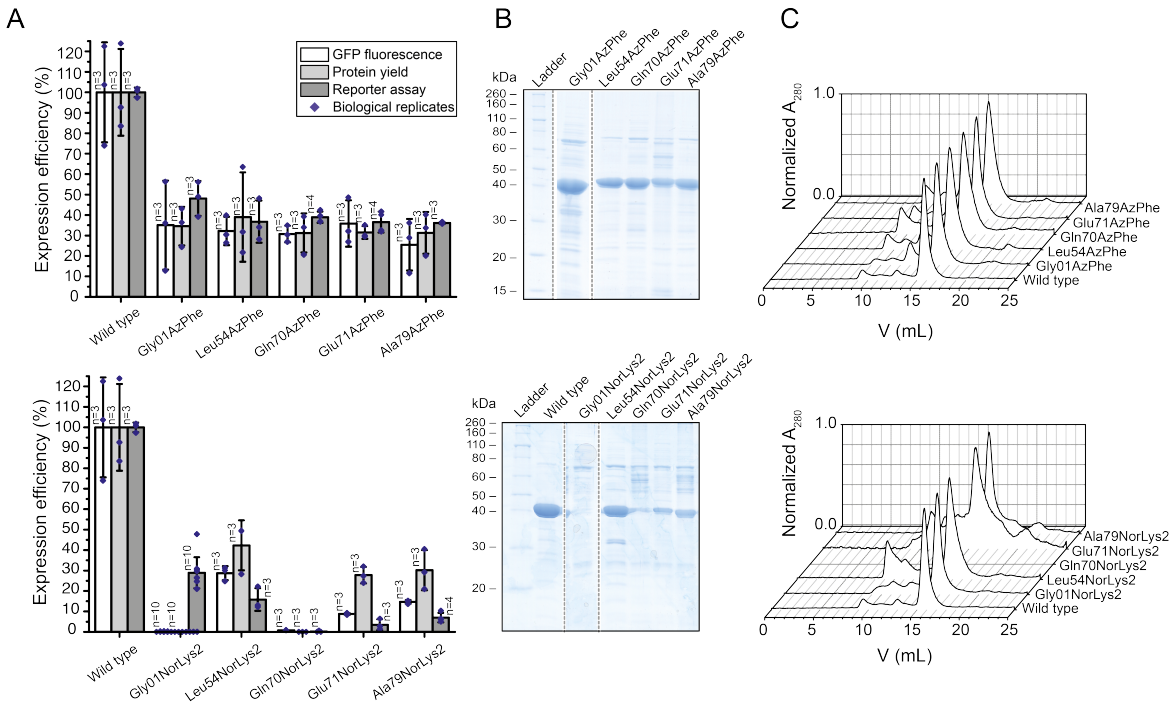


Figure 3.6: Large scale expression and purification of ACP-GFP mutants (upper panels AzPhe mutants, lower panels NorLys2 mutants). A) Comparison of the results from large scale expression cultures (protein yield was read out by GFP fluorescence of a 2 mL sample and by the yield of purified protein after Ni-IMAC) with previous results from the reporter assay. Data compare expression efficiency of wild type and five different AzPhe mutants (upper panel), and of five different NorLys2 mutants (lower panel). All expression efficiencies are related to the wild-type reference. For incorporation, 2 mM ncAA were supplemented to the medium. The averages of biological replicates are plotted together with standard deviations and the distribution of individual values is indicated as dots. Technical errors were below 10 %. B) SDS-PAGE gel (NuPAGE Bis-Tris 4-12 %) of ACP-GFP mutants purified by Ni-chelating chromatography. Lanes have been assembled for clarity (indicated by dashed lines), but scans of the original gels can be found in Fig. S13. SDS-PAGE shows one set of purified proteins (one biological replicate). C) Preparative SEC of ACP-GFP mutants performed with a Superdex 200 Increase 10/300 GL column (the set of proteins shown in B). Peaks at an elution volume of 16 mL correspond to the ACP-GFP variants. The void volume of the column is at approximately 9 mL.

For comparing protein yields, cells received from 200 mL expression cultures were lysed and proteins were isolated by Ni-chelating chromatography (Fig. 3.6B). Compared to ACP-GFP at 53 ± 15 mg, the positive mutants were isolated in average with 14 ± 3 mg yield, which corresponds to about 25 % of the wild-type protein yield. The incorporation efficiency quantified by protein yield correlated well with the trends of the fluorescence data. AzPhe was again incorporated with overall good efficiency (in average $33\,\% \pm 4\,\%$), whereas NorLys2 performed differently throughout the amber mutation sites. The optimal site for NorLys2, i. e. Leu54 in the disordered loop region, showed $42\,\% \pm 12\,\%$ incorporation efficiency, and even the amber mutation sites Glu71 and Ala79 gave up to 30 % incorporation efficiency. Again, no incorporation of NorLys2 at the amber mutation site Gln70 was monitored. NorLys2 mutants led to higher protein yields than expected from fluorescence intensities of cell cultures, which cannot be explained with the collected data.
It has to be noted that protein yield determined after Ni-chelating chromatography is only an estimate as some impurities can still be seen in the gels and size exclusion chromatography (SEC) (Fig. 3.6C) also loss of protein during purification may alter the amount of purified protein. The quality of proteins was analyzed by SEC and mass spectrometry (MS) (see Supplementary Data). The elution profiles of the different ACP-GFP mutants matched very well with the wild-type SEC spectrum. In MS analysis, a mass shift for the Gln70AzPhe mutant was observed (see Supplementary Data), although we were able to specifically label this mutant in further experiments (see section 3.1.4), which confirms presence of the azido group at this stage. Instability of the azido group during protein analysis may account for the reduced molecular weight but cannot explain why this mass shift is only observed for this mutant. Owing to this inconsistency, the Gln70 position was considered unreliable and not chosen in full-length protein constructs.

3.1.4 Site-Specific Labeling

In first experiments, click kinetics for our target protein ACP-GFP were screened (see Fig. S3) and a suited condition of 2 h of incubation at room temperature with 80 equiv. of fluorophore in 10 µL reaction volume for both the SPAAC (AzPhe mutant BCN-POE_3-NH-DY649P1 conjugate) and the IEDDAC (NorLys2 mutant 6-methyltetrazine-ATTO-647N conjugate) were received. For determining the degree of labeling (DOL) by in-gel fluorescence, ACP-GFP enzymatically modified with a fluorescent CoA-label by a 4'-phosphopantetheinyl transferase (Sfp)[127] was used as reference. For better comparison, the different fluorescence intensities were corrected by the re-

spective quantum efficiencies. Three different fluorophores were used in this experiment: DY647P1 at the CoA-label, DY649P1 at the BCN-label and ATTO 647N at the tetrazine-label (spectral properties of the fluorophores are summarized in Table 5.2). The sample of ACP-GFP enzymatically modified by a fluorescent CoA-label showed highest fluorescence and was assumed to be quantitatively labeled based on previous reports[128] and thus set to 100 % as the wild-type reference. As this aspect has a great impact on our data, additional experiments to verify that no unmodified protein is present in the reference sample were performed (see Fig. 3.7).

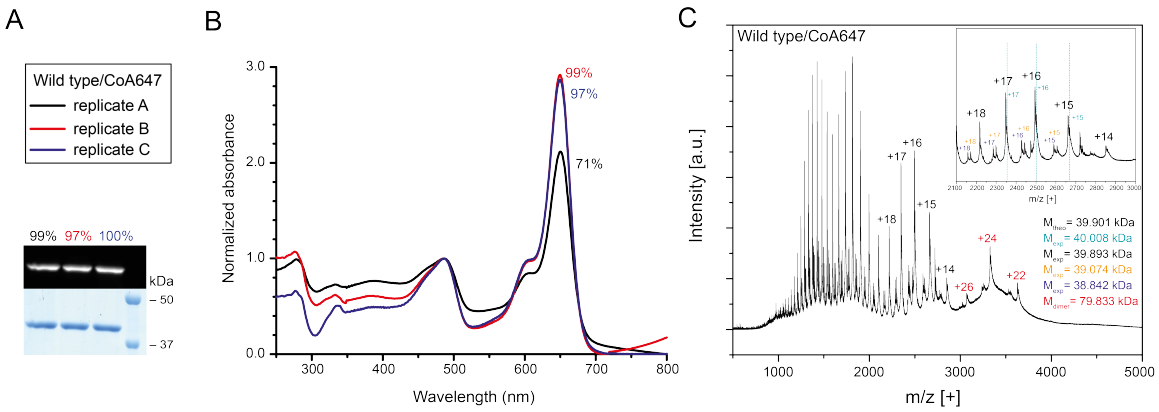


Figure 3.7: Quantification of *in vitro* phosphopantetheinylation of wild-type ACP-GFP with CoA 647-label by Sfp. A) In-gel fluorescence intensities of wild-type ACP-GFP CoA 647 conjugates (SDS-PAGE with 12 % polyacrylamide). Wild-type ACP-GFP was enzymatically modified with a fluorescent CoA 647-label by Sfp in three parallel reactions. In-gel fluorescence intensities were detected at wavelength 650 nm, corrected by the quantum efficiency of the DY647P1 fluorophore and correlated to the protein bands of the Coomassie-stained gel. B) DOL of wild-type ACP-GFP CoA 647 conjugates determined by UV-Vis spectroscopy (equation 5.1). Free fluorophore was removed by purification over Ni-NTA magnetic beads. UV-Vis absorbance spectra were normalized to GFP absorbance at 485 nm wavelength. DOL is read out by comparing absorbance of the fluorophore at 650 nm to absorbance of GFP at 485 nm. C) MS analysis of wild-type ACP-GFP CoA 647 conjugates. Free fluorophore was removed by purification over Amicon® Ultra Centrifugal Filters (Merck Millipore). No unmodified wild-type ACP-GFP mass was detected within the detection range of the instrument. The main peaks match the theoretical mass of modified wild-type ACP-GFP CoA 647 conjugate. Impurities in MS spectra sum up to approximately 10 %, which cannot be assigned to a specific compound.

All fluorescence intensities were further correlated to the intensity of the protein bands of the Coomassie-stained gel (Fig. 3.8A). In average, the AzPhe mutants clicked more efficiently than the NorLys2 mutants (74 % over 23 %). As in-gel fluorescence is always determined relatively to the wild-type reference, accurate comparison of intensities between different gels is difficult, and thus we observed variations in the DOL. Biological replicates of the labeling reaction for both Leu54 mutants were performed

and compared (see inserted box in Fig. 3.8A). The DOL of the Leu54AzPhe mutant was in average 65 % ± 10 % and of the Leu54NorLys2 mutant 24 % ± 10 %. With respect to the given error margin of approximately 20 %, no statement to a position dependency of the labeling reaction can be made.

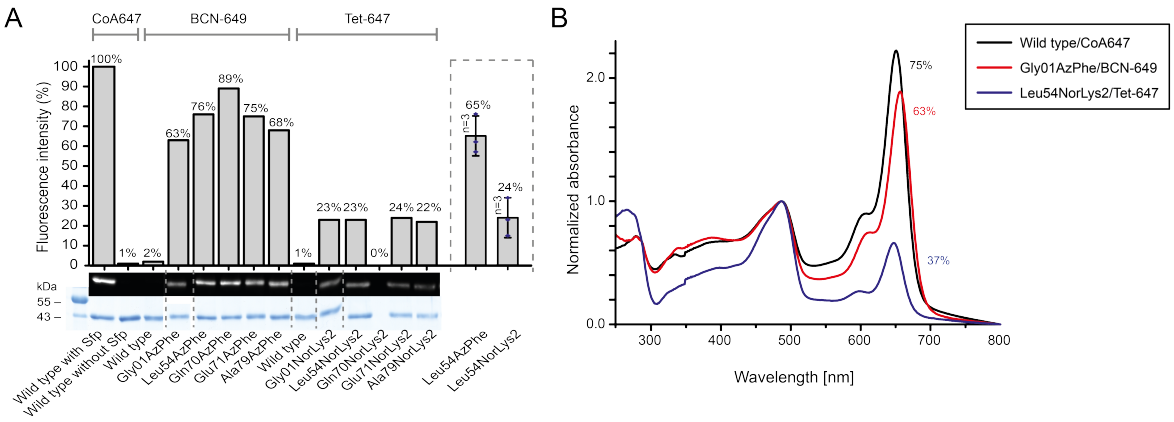


Figure 3.8: Fluorescent labeling of ACP-GFP mutants. A) DOL of ACP-GFP mutants in respect to the amber mutation site determined by relative in-gel fluorescence intensities at wavelength 650 nm (SDS-PAGE on NuPAGE Bis-Tris 4-12 %). The ACP-GFP construct was enzymatically modified by a fluorescent CoA 647-label with Sfp and served as the wild-type reference. Hence, it was put to 100 % fluorescence intensity. AzPhe mutants were labeled with 80 equiv. of BCN-POE$_3$-NH-DY649P1 (BCN-649), NorLys2 mutants were labeled with 80 equiv. of 6-methyl-tetrazine-ATTO-647N (Tet-647) in 10 µL reaction volume for 2 h. All fluorescence intensities were corrected by the quantum efficiency of the respective fluorophore and correlated to the protein bands of the Coomassie-stained gel (lanes have been assembled for clarity, as indicated by dashed lines). Scans of the original gels are presented in Fig. S13. Biological replicates were performed for the Leu54 mutants, comparing DOL determined for three parallel labeling reactions analyzed on different fluorescent gels (inserted box). Individual results were gathered from gels in this Figure and Fig. S3. B) DOL determined by UV-Vis spectroscopy (equation 5.1). 25 equiv. of fluorophore were used in 2 h labeling reactions of ACP-GFP mutants in 50 µL reaction volume. Free fluorophore was removed by purification over Ni-NTA magnetic beads. UV-Vis absorbance spectra were normalized to GFP absorbance at wavelength 485 nm. DOL is read out by comparing absorbance of the fluorophore at 650 nm to absorbance of GFP at 485 nm.

The DOL was alternatively determined by spectroscopy with samples Gly01AzPhe and Leu54NorLys2 after removal of excess free fluorophore by purification via Ni-NTA magnetic beads (see Fig. S4). Several methods were tested for the removal of excess free fluorophore by master's students Maria Dell and Bjarne Goebel but Ni-NTA magnetic beads showed best results. In a single experiment, these proteins were clicked with 25 equiv. of fluorophore in 50 µL reaction volume and the labeling efficiency was monitored with UV-Vis absorbance spectra. For the wild-type reference, a DOL of

75 % was determined, whereas the Gly01AzPhe mutant showed 63 % DOL and the Leu54NorLys2 mutant showed only 37 % DOL (Fig. 3.8B).

In earlier experiments of this thesis, the ncAA NorLys1 was incorporated in proteins, which not only shows poorer incorporation efficiencies than NorLys2, but also poorer labeling efficiency with the methyl-tetrazine dye (see Fig. S5). Site-specific labeling of AzPhe mutants with DBCO functional groups and AcPhe in oxime formation for application in EPR spectroscopy is discussed in section 3.2.2.

3.1.5 Modification of mFAS Subconstructs

The subconstruct KS-MAT-ACP is an interesting truncated variant of mFAS. It comprises the complete condensing wing of animal type I FAS as well as resembles non-reducing PKSs that naturally occur in polyketide synthesis. KS-MAT-ACP is active in triacetic lactone (TAL) production. As outlined in Fig. 3.9A, TAL synthesis requires an alternating sequence of KS:ACP and MAT:ACP interactions. In both interactions, MAT:ACP and KS:ACP, ACP can either be loaded with acetyl or malonyl, giving MAT:ACP$^{\text{acetyl}}$, MAT:ACP$^{\text{malonyl}}$, KS:ACP$^{\text{acetyl}}$ and KS:ACP$^{\text{malonyl}}$. The intermediate acetoacetyl is also expected to unbind and rebind to the KS binding site (KS:ACP$^{\text{acetoacetyl}}$). It has to be noted that for the interactions KS:ACP and MAT:ACP two possible conformation states can appear (intermolecular *vs.* intramolecular interaction). Moreover, spectroscopic labels can either monitor a productive or a non-productive reaction cleft (assuming a negatively cooperative working mode of clefts). In this context, a set of six to eight main conformational states is assumed (Fig. 3.9B). Depending on the specific positioning of spectroscopic labels, not all conformational states are detectable with a specific biophysical probe. For reducing possible conformational states, also a KS-MAT-ACP/*KS-MAT* heterodimer can be constructed (see Fig. 3.9B).

For full-length mFAS, but also truncated subconstructs as described here, conformational states can be stalled by specific crosslinking.[129–131] Although the system is at low complexity as compared to full-length mFAS it gives access to fundamental processes in carrier-mediated synthesis and is therefore used as model protein to establish later applications (see section 3.2). Finally, we aim for kinetically and structurally model KS-MAT-ACP during TAL synthesis. Parameters describing KS-MAT-ACP conformational dynamics can be used for feeding more complex data collected on more complex

models.

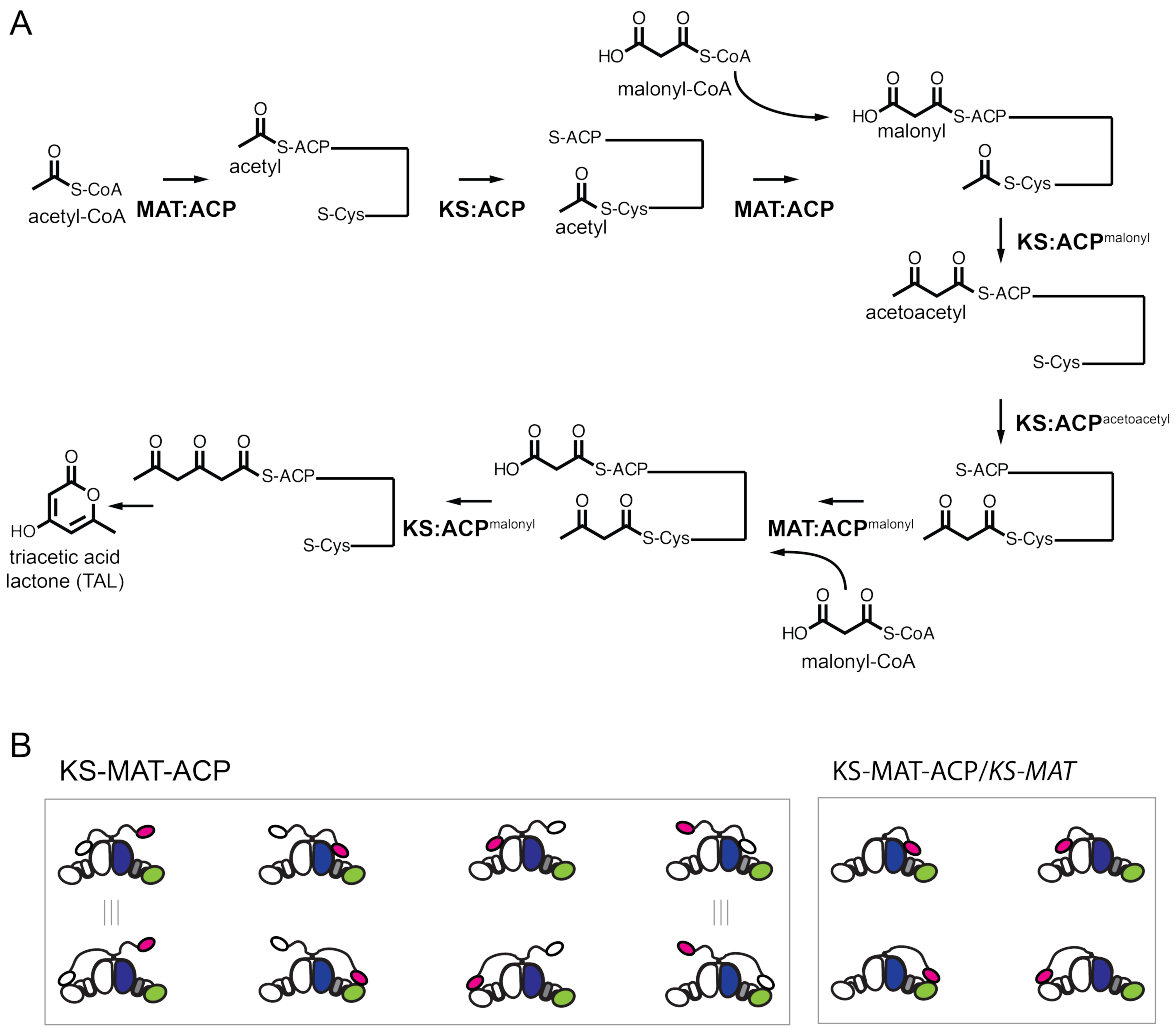


Figure 3.9: Analysis of conformational dynamics of KS-MAT-ACP. A) Catalytic cycle of TAL production with the domain-domain interactions of the specific steps indicated. B) Possible main conformation states of constructs. Putative similar conformations are indicated.

Stability of Proteins in Chemoselective Reactions

To ensure functionality of the protein in later applications, its stability was tested in different buffers (His, Strep, PKS, MOPS and HEPES buffer) and under the conditions of different chemoselective reactions (CuAAC, oxime formation, and copper-free chemistries). Wild-type KS-MAT-ACP was examined in a thermal shift assay (TSA) in technical triplicate measurements. The TSA assesses protein stability on the basis of melting temperatures. The higher the melting temperature the more stable the protein

under the assayed buffer condition.

Table 3.2: Stability of KS-MAT constructs in different buffers and in presence of copper.

Buffer	**0 μM Cu^I**	**100 μM Cu^I**	**50 μM Cu^I**	**250 μM Cu^I**
His	50.5 °C	50.6 °C	50.3 °C	50.1 °C
Strep	50.1 °C	37.8 °C	33.3 °C	34.0 °C
PKS	55.0 °C	39.3 °C	35.2 °C	33.5 °C
MOPS	50.3 °C	n.d.	16.9 °C	17.4 °C
HEPES	51.6 °C	n.d.	11.2 °C	12.3 °C

Table 3.3: Stability of KS-MAT constructs in His buffer and in presence of different components of the CuAAC.

Component	**0 μM Cu^I**	**100 μM Cu^I**	**50 μM Cu^I**	**250 μM Cu^I**
only His buffer	50.5 °C	50.6 °C	50.3 °C	50.1 °C
THPTA ligand	50.5 °C	50.5 °C	50.2 °C	49.6 °C
ascorbic acid	50.7 °C	50.6 °C	50.4 °C	50.5 °C
aminoguanidine	50.3 °C	50.3 °C	50.2 °C	49.8 °C

Our protein of interest is stable in all tested buffer conditions (Table 3.2) giving melting temperatures of 50 °C for His, Strep and MOPS buffer and even higher melting temperatures for HEPES (52 °C) and PKS buffer (55 °C). Commonly, the FAS proteins are stored in Strep buffer as it has been reported as stable protein buffer by Rangan *et al.*.[132] In the presence of Cu^{II}, protein stability drastically decreases. Melting temperatures of 33 to 39 °C were determined for Strep and PKS buffer and even worse for MOPS and HEPES buffer lying in the range of 11 to 17 °C. Only in His buffer the protein remains stable in presence of 50 to 250 μM Cu^{II} giving an unaltered melting temperature of 50 °C. Adding more components of the CuAAC reaction to the protein in His buffer did not affect the melting temperature and hence protein stability did not decrease (Table 3.3). Oxime formation takes place under acidic reaction conditions or at physiological pH catalyzed by *para*-anisidine. From the TSA, it can be concluded that at acidic pH of 4.0 or at presence of *para*-anisidine protein stability is slightly decreased but with a drop in melting temperature of 2 to 9 °C acceptable protein stability is still given.

Table 3.4: Stability of KS-MAT constructs in different buffers at pH 7, at pH 4, and at pH 7 in presence of *para*-anisidine.

Buffer	pH7	pH 7, *p*-anisidine	pH 4
His	50.5 °C	47.9 °C	50.9 °C
Strep	50.1 °C	49.1 °C	41.1 °C
PKS	55.0 °C	52.9 °C	53.5 °C

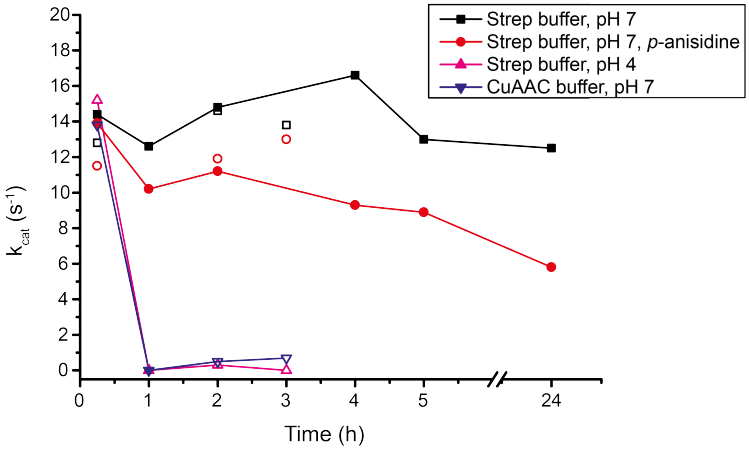


Figure 3.10: Activity of KS-MAT constructs monitored over time. The KS-MAT proteins were incubated at physiological pH in Strep buffer (black curve, squares), in Strep buffer at pH 7 and in presence of *para*-anisidine (red curve, circles), in Strep buffer at acidic pH 4 (pink curve, triangles tip up), or in presence of copper and other components of the CuAAC (blue curve, triangles tip down) for up to 24 h. AT activity assays were performed at different time points. The k_{cat} of the transferase reaction was determined from initial velocities and plotted against time. Filled symbols represent corresponding measurements of samples incubated at 37 °C, unfilled symbols represent corresponding measurements of samples incubated at room temperature.

Nonetheless, the TSA can neither give any information about protein stability over time nor about protein activity. Therefore, the protein was incubated in different buffer conditions and at different temperatures for up to 24 h. AT activity was measured at different time points in technical triplicates (Fig. 3.10). Protein in Strep buffer remained stable and active over time as indicated by a positive slope giving k_{cat} values of about 14 s^{-1}. Even incubation at 37 °C was tolerated for up to 24 h. The k_{cat} value just slightly decreased to 12 s^{-1}. Also in presence of the *para*-anisidine catalyzator the protein remained stable and active at room temperature for 3 h showing only slightly lower k_{cat} values of 12 to 13 s^{-1}. With incubation at 37 °C in presence of *para*-anisidine, the k_{cat} values constantly decreased over time from 14 down to 6 s^{-1} in the course of 24 h. For

acidic pH and the buffer conditions in CuAAC, no linear plot of initial velocities was obtained, hence no k_{cat} value was determined. The protein was not stable over time and inactive after only 1 h incubation at room temperature.

Incorporation of ncAAs in Subconstructs of mFAS and Modules of PKSs

As proof of principle, some examples of successful ncAA incorporation in mFAS subconstructs and even PKS modules are shown in the fluorescent gel in Fig. 3.11. Heterodimers of mFAS subconstructs were generated *in vivo*, designing a pETDuet expression vector for bicistronic expression of two protein genes that contain alternate C-terminal affinity tags. One wild-type KS-MAT-ACP gene contained a C-terminal His-tag, the other KS-MAT gene contained an amber mutation site and a C-terminal FLAG-tag. The KS-MAT-ACP/*KS-MAT* heterodimers were isolated through two subsequent affinity chromatography purification steps using the C-terminal His- and FLAG-tags. Initial FRET probes were generated by labeling these protein constructs with two different fluorophores.

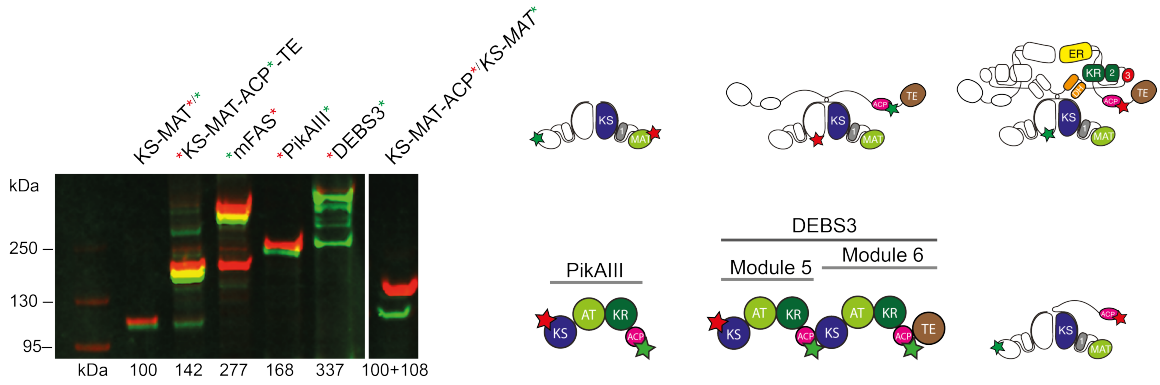


Figure 3.11: In-gel fluorescence of fluorescently labeled mFAS and PKS constructs. Besides mFAS constructs, module PikAIII from the pikromycine type I PKS and module three from DEBS were modified and analyzed (schematic views of the protein constructs are shown in the right pannel). The ACP domains were enzymatically labeled with a fluorescent CoA through Sfp. The ncAA NorLys2 was incorporated into the protein constructs and labeled in a chemoselective IEDDAC with 6-methyl-tetrazine-sulfo-cyanine dyes (green: Cy3 channel; red: Cy5 channel). The KS-MAT homodimer in the first lane was treated with an equivalent mixture of 6-methyl-tetrazine-sulfo-Cy3 and Cy5 fluorophore. The last lane shows a mixture of homodimeric KS-MAT labeled with 6-methyl-tetrazine-sulfo-Cy5 and a KS-MAT/KS-MAT-ACP heterodimer. The full-length mFAS and both PKS constructs show degradation bands. SDS-PAGE was performed on NuPAGE Bis-Tris 4-12 % gels.

3.1.6 Modification of Full-Length mFAS

Incorporation of ncAAs in Full-Length mFAS

The positions identified to have successful incorporation of ncAAs within the excised ACP domain were tested on the full-length protein. For this, three amber mutation sites were chosen and AzPhe was introduced at position Gly2113 (Gly01 in ACP) and NorLys2 at position Leu2166 (Leu54 in ACP) as promising candidates, and NorLys2 at position Gln2182 (Gln70 in ACP) as a negative control in mFAS. In agreement with our previous data the variants Gly2113AzPhe and Leu2166NorLys2 expressed and purified in good yields (66 % and 31 % of wild-type mFAS expression, respectively, for a representative gel of the purification, see Fig. 3.12A), whereas the mutant Gln2182NorLys2 showed poor expression (1 % of wild-type mFAS expression) (Fig. 3.12B). For mFAS and its subconstructs, successful ncAA incorporation and existence of full-length protein cannot be determined by GFP fluorescence anymore. Therefore, these constructs were usually analyzed by western blot, using antibodies against the N-terminal Strep-tag and the C-terminal His-tag. From these western blot analyses it became apparent, that amber codon suppression was insufficient, leading to the formation of heterodimers with truncated protein due to the dimeric nature of mFAS, which could not be separated with the current purification methods (Fig. 3.12C).[75] These truncated fragments were quantified from band intensities in the gel to be about 30 % in the Gly2113AzPhe sample and 40 % in the Gln2182NorLys2 sample. To avoid these truncation bands, expression in the GRO C321.ΔA from the Church lab was tested with limited success.[22,23] This GRO was actually designed to create a real blank codon by replacing all amber codons in the organism and deleting release factor 1 (RF1) function but unfortunately this *E. coli* strain is restricted in cell growth.

Site-Specific Fluorescent Labeling of Full-Length mFAS

The successful incorporation of the respective ncAAs was qualitatively confirmed by clicking complementary fluorophores (Fig. 3.12D), although some amount of unspecific binding was observed as well. The SPAAC reaction was again more efficient than the IEDDAC and yielded interestingly higher in-gel fluorescence intensities than the enzymatic labeling with CoA 647. This may be explained by lower accessibility of the ACP domain in the full-length protein by the complex of Sfp with CoA 647 in an *in vitro* reaction (see Fig. S6). We can exclude a partial activation of the enzyme *in vivo* as mFAS is inactive unless co-expressed with Sfp and hence *E. coli* is not capable to

phosphopantetheinylate the ACP domain.[75]

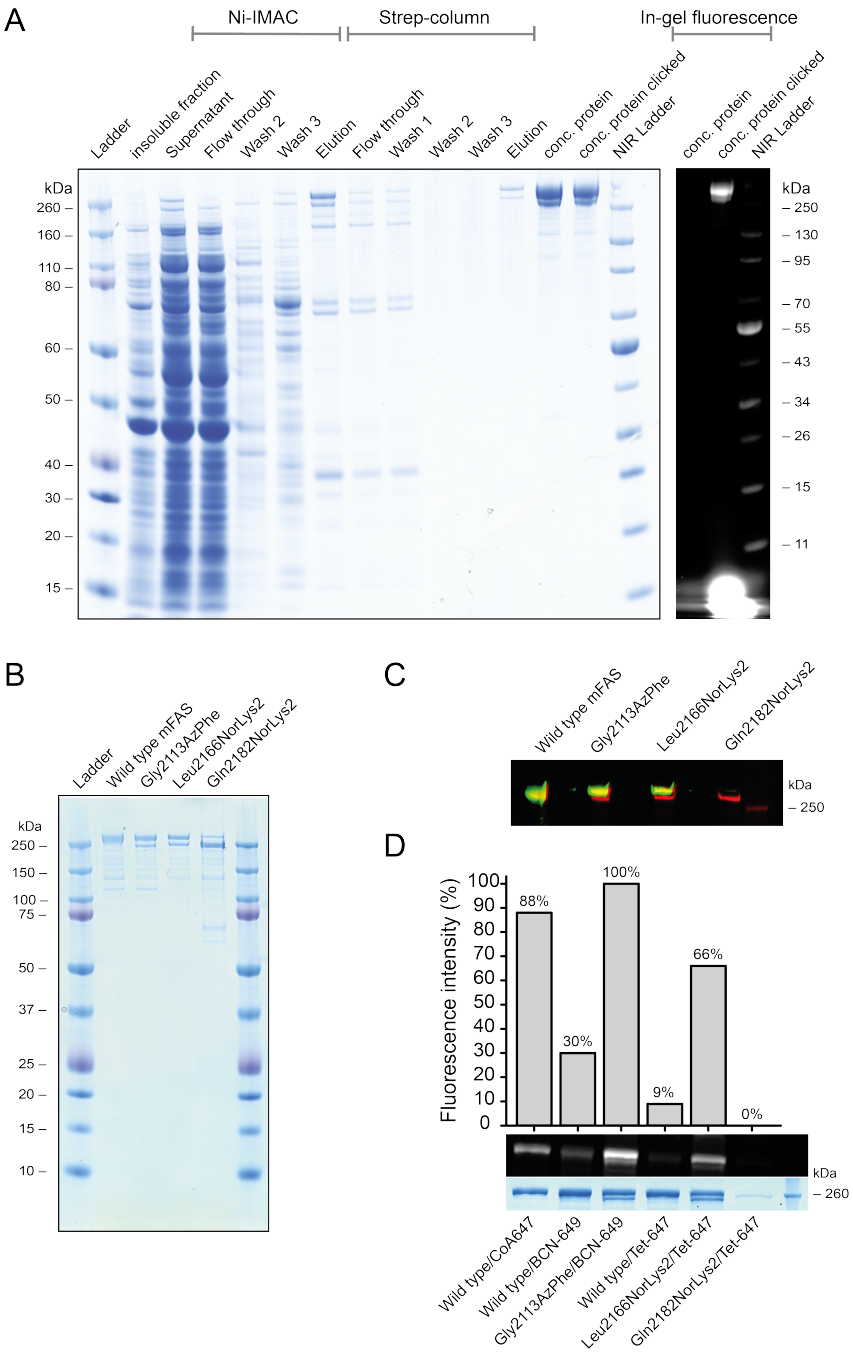


Figure 3.12: Generation of ncAA-modified mFAS mutants. A) SDS-PAGE (NuPAGE Bis-Tris 4-12 %) of a representative purification of the ncAA-modified mFAS mutant (Gly2113AzPhe with additional ACP knock-out mutation Ser2150Ala). The portion of truncated polypeptide chains after the tandem purification strategy was quantified to be roughly 30 %. Truncated proteins reflect insufficient amber suppression and are co-purified

by heterodimer formation. In-gel fluorescence demonstrates successful labeling with BCN-POE_3-NH-DY649P1 (BCN-649). B) SDS-PAGE (NuPAGE Bis-Tris 4-12 %) gel of mFAS mutants. Proteins were purified by a tandem affinity chromatography via the N-terminal Strep-tag and the C-terminal His-tag. C) Qualitative western blot of mFAS mutants. The red channel refers to antibodies conjugated with DyLight 755 against the N-terminal Strep-tag and the green channel to antibodies conjugated with DyLight 633 against the C-terminal His-tag, respectively. The missing green band in lane four indicates that expression of full-length Gln2182NorLys2 mFAS failed and that only a truncated construct without the C-terminal part was obtained. D) Fluorescent labeling of mFAS mutants. AzPhe mutants were labeled with 80 equiv. BCN-POE_3-NH-DY649P1 (BCN-649), NorLys2 mutants were labeled with 80 equiv. 6-methyl-tetrazine-ATTO-647N (Tet-647) for 2 h, and the wild-type mFAS was enzymatically modified at the ACP domain by a fluorescent CoA 647-label by Sfp. DOL is determined by the relative in-gel fluorescence intensities at wavelength 650 nm related to the wild-type reference (SDS-PAGE on NuPAGE Bis-Tris 4-12 %). All fluorescence intensities were corrected by quantum efficiency of the respective fluorophore and correlated to the protein bands of the Coomassie-stained gel.

Physicochemical Analysis of Modified mFAS

We then proceeded with investigating the mutant Gly2113AzPhe in preparative scales. In the first step, a protocol for purification with HisPur Ni-NTA magnetic beads was established to discard excess free fluorophore resulting from the click reaction. About 25-30 % of the protein sample was recovered after clicking (Fig. 3.13A). Three samples were phosphopantetheinylated with Sfp *in vitro* and purified via magnetic beads: mutant Gly2113AzPhe bearing an ACP knock-out (Ser2150Ala) and mutant Gly2113AzPhe both as protein BCN-POE_3-NH-DY649P1 conjugate and wild-type mFAS as control. The integrity of all mFAS samples was confirmed by HPLC-SEC (Fig. 3.13B) showing specific labeling of the AzPhe mutants and little unspecific binding to wild-type mFAS. These three samples were subject to NADPH activity assays showing no activity before the procedure of activation and clicking and giving the results shown in Fig. 3.13C after removal of excess compounds. The wild-type mFAS showed high activity (853 ± 44 nmol/min/mg protein) revealing that it was not compromised by the clicking conditions and that the mild SPAAC reaction performs appropriate to mFAS modification. As expected, the ACP knock-out sample showed loss of activity, whereas the labeled Gly2113AzPhe variant was active although the efficiency was reduced (284 ± 57 nmol/min/mg protein).

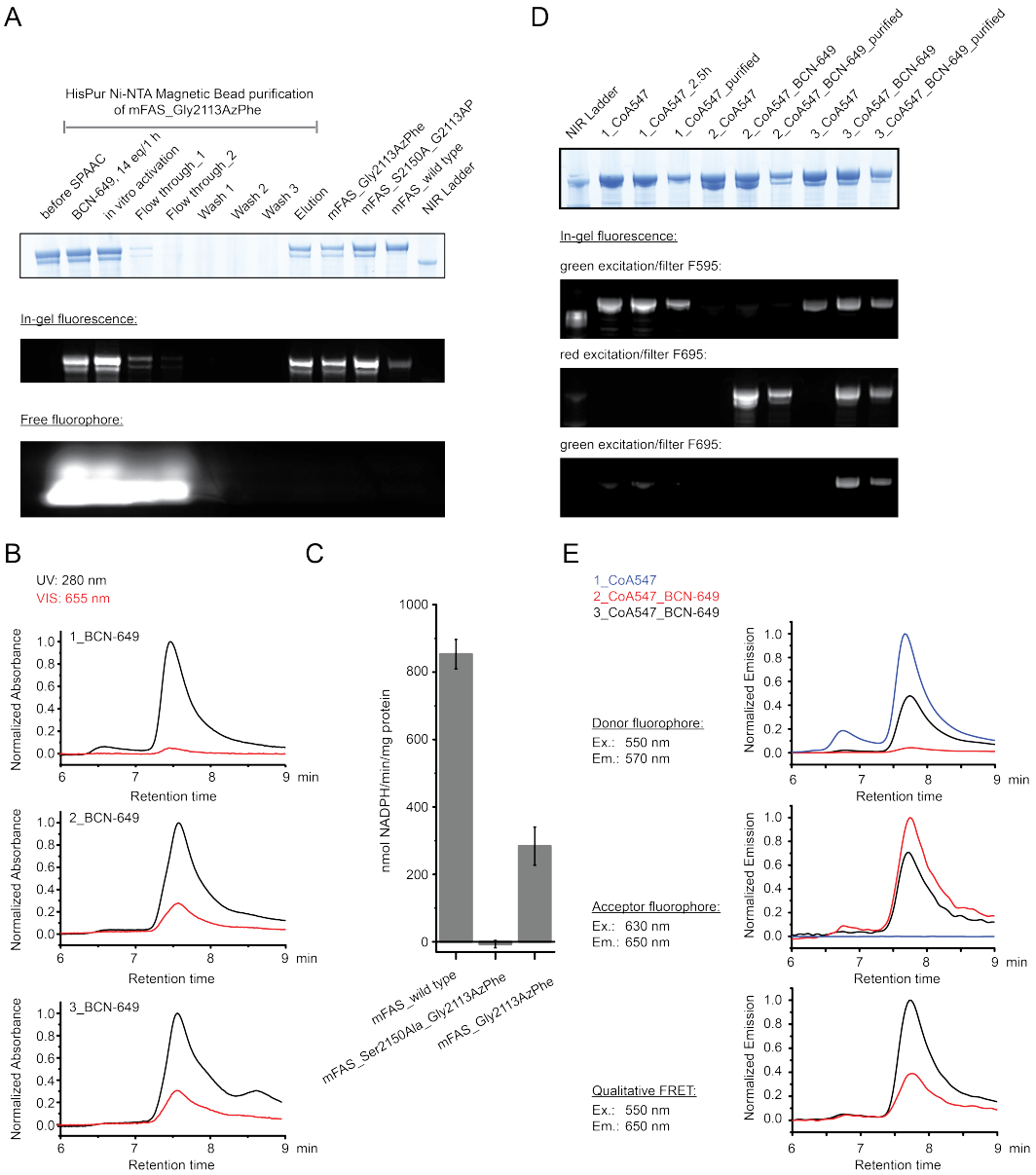


Figure 3.13: Physicochemical analysis of fluorescently labeled mFAS. A) Purification of three full-length mFAS variants (Gly2113AzPhe, Gly2113AzPhe with Ser2150Ala and wild type) in preparative 100 µg scale via Ni-NTA magnetic beads. Samples were clicked with 16 eq. BCN-649 for 1 h followed by *in vitro* phosphopantetheinylation with CoA and Sfp. In-gel fluorescence of an SDS-PAGE (NuPAGE Bis-Tris 4-12 %) was detected before Coomassie-staining. The lowest panel indicates that the majority of free fluorophore is removed after the first washing step. B) SDS-PAGE (NuPAGE Bis-Tris 4-12 %) of the same three variants: 1: wild-type mFAS, 2: mFAS (Gly2113AzPhe and Ser2150Ala) and 3: mFAS (Gly2113AzPhe) after enzymatic labeling with CoA 547 and clicking of the AzPhe containing variants with BCN-649. In-gel fluorescence was detected with three different settings: excitation with green light and filter F595 (for 595 nm); excitation with red light and filter F695 (695 nm) and excitation with green light and filter F695. The gel after Coomassie-staining is attached. All samples show specific fluorescent bands in the respective channels with little unspecific binding due to denaturing conditions. The dual-labeled sample Gly2113AzPhe shows FRET

signal. C) Analysis of the integrity of the three labeled samples after clicking and purification by HPLC-SEC. The main peak at 7.5-7.6 min corresponds to the native dimeric state. Absorbance was monitored at 280 nm and 655 nm and normalized to the highest peak in the UV signal. Both samples containing the ncAA AzPhe were labeled with fluorophore (28 % and 31 % DOL), whereas the wild-type mFAS showed only minor unspecific fluorophore binding (5 %). D) Activity of mFAS variants monitored by a NADPH consumption assay after phosphopantetheinylation and clicking. The variant Gly2113AzPhe showed one third of the wild-type activity, whereas the ACP knock-out (Ser2150Ala) could not produce fatty acids. Error bars indicate the technical reproducability determined from three repeated experiments per construct. E) Fluorescence analysis of the three labeled samples of B) via HPLC-SEC. Emission spectra are shown for the settings: Ex. 550 nm/Em 570 nm; Ex. 630 nm/Em 650 nm and Ex. 550 nm/Em 650 nm. All signals were normalized as described in detail in the methods section. Again, the dual-labeled sample Gly2113AzPhe shows (highest) FRET signal.

As a follow up, we attached two different fluorescent labels to variant Gly2113AzPhe, necessary for FRET experiments. The sample was modified with the BCN-POE_3-NH-DY649P1 conjugate and enzymatically labeled with CoA 547 by Sfp *in vitro*. Data obtained on this sample indeed showed a FRET signal in denaturing SDS-PAGE (Fig. 3.13D) and native HPLC-SEC (Fig. 3.13E) compared to the appropriate control experiments but nevertheless the DOL was low. Our proposed methods require further optimization to achieve high quality samples for single molecule FRET.

3.2 Applications of Site-specific ncAA Incorporation to Address Biological Questions

3.2.1 Analysis of Domain Dynamics Using Fluorescence Spectroscopic Methods

As outlined in section 3.1.4 & 3.1.6 we were able to attach fluorophores site-specifically to our proteins using amber codon suppression. Fluorescent probes are used in living cells to localize proteins or in spectroscopic methods mostly *in vitro* to study protein dynamics. Yuan *et al.* published fluorescence studies on mammalian FAS using dynamic anisotropy measurements to determine segmental flexibility of the ACP and TE domain, and FRET for distance measurements within the protein.[52]

Buchholz *et al.* have reported an FP assay to determine binding affinities of docking domains in the modular type I PKS system of pikromycin.[133] Analog to this study we thought

to determine binding affinities between ACP and catalytic domains of mFAS. Our group has developed an assay to determine kinetic constants of the transferase reaction and is characterizing ATs from different FAS and PKS systems.[75] Feeding in dissociation constants K_D from the FP assay would benefit this ongoing project.

Fluorescence Polarization Assay

Franziska Stegemann took first steps towards the establishment of the FP assay in her master's thesis, which had primarily been supervised by myself (with support from Dr. Alexander Rittner concerning the transferase assay). In the course of her master's thesis we isolated different ACPs from FAS (murine type I FAS and *E. coli* type II FAS) and type I PKS (DEBS3 module 5, PikAIII module 5 and RAPS3 module 14) as well as respective KS-AT di-domains of these systems. Initially, we planned the positions for amber codon placement in the ACP sequence (for all ACPs: Ala-1 in the N-terminal Strep-tag and Leu+1 in the C-terminal His-tag; for murine ACP additionally: Leu54 in the linker region between helix 3 and helix 4, as well as Gln70 and Glu71 in helix 4 (equivalent to Leu2166, Gln2182 and Glu2183 in mFAS, UniProtKB accession code P19096)). These incorporation sites were thoroughly screened in the reporter assay for ACP from mFAS (see section 3.1.3).

The ncAA NorLys2 was incorporated in response to the amber codon and used to site-specifically label the ACPs with 6-methyl-tetrazine-ATTO647N fluorophore. Fluorescently labeled ACP from mFAS with a DOL of approximately 40 % was used to establish the FP measurements. First, the detection limit for the ATTO647N fluorophore was assessed, pipetting a series of concentrations of free fluorophore ranging from 0.08-10 µM. Since FP is independent from fluorophore concentration, a constant FP signal is expected as is the case for free fluorophore concentrations in the range of 0.75-10 µM (Fig. 3.14). The stability of the FP signal was subsequently tested with a dilution series of fluorescently labeled ACP ranging from 0.07-6 µM. A constant FP signal was observed in the concentration range of 0.7-6 µM (Fig. 3.14). Accordingly, the following FP measurements were conducted with a constant concentration of 1 µM fluorescent ACP. Determination of an optimal ACP concentration is important because too low concentrations result in higher variation in the FP signal due to instrument sensitivity, whereas too high concentrations require high levels of binding molecules in the titration experiment. Titration of increasing concentrations of KS-MAT to the fluorescently labeled ACP should lead to a change in FP signal resulting from ACP binding. Smaller proteins exhibit fast rotational dynamics and give a low FP signal, whereas larger proteins exhibit slower rotational dynamics and give a high FP signal. Unbound free ACP should therefore show the lowest FP signal, whereas ACP bound to the KS-AT di-domain should give the maximal FP signal in the FP assay. With increasing concentrations of KS-AT more ACP gets bound and the FP signal should increase showing a sigmoidal curve. From

the point of inflection a K_D can be determined. In order to measure the binding affinity to solely the AT domain KS knock-out mutants of the KS-AT di-domains had to be generated. The FP signal slightly increased with increasing KS-MAT concentrations but the plot showed a lot of scattering (Fig. 3.14). No K_D could be determined.

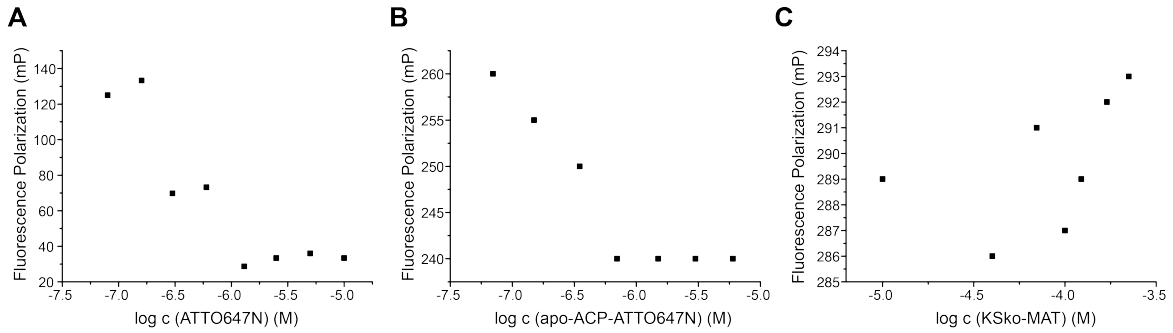


Figure 3.14: Establishment of the FP assay. FP (mP) values are plotted against logarithmic concentrations (M). A) FP measured for a dilution series of free 6-methyl-tetrazine-ATTO647N dye. B) FP measured for a dilution series of apo murine ACP mutant containing NorLys2 and labeled with 6-methyl-tetrazine-ATTO647N. C) FP of the binding experiment for murine ACP and KS-MAT. To a constant concentration of 1 µM apo-ACP-ATTO647N, increasing concentrations of KS-MAT, containing a KS active site knock-out mutation, were titrated.

Establishment of smFRET Measurements

Another spectroscopic method that employs fluorophores is FRET. The ideas envisioned with the FP assay can in principle also be addressed by FRET measurements. Additionally, FRET does also give distance information. While for FP only one fluorophore is needed, FRET requires bioconjugation with two different fluorophores whose spectral shapes must overlap. As outlined in the introduction section 2.2, several approaches exist for multiple incorporation of ncAAs. In principle, a second ncAA can be incorporated in response to another blank codon. Besides a second ncAA and blank codon, this requires an additional orthogonal aaRS/tRNA pair and in some cases additional components like a new synthetic nucleotide or an orthogonal ribosome. We tried to circumvent the technical challenges that come with the use of a second ncAA and chose different approaches instead: (i) the ACP can be modified with a fluorescent CoA; (ii) Rangan *et al.* reported a protocol to generate heterodimeric animal FASs;[132] (iii) stochastic labeling with two fluorophores simultaneously.

All three approaches were tested for their applicability. (i) Fluorescent CoAs have been frequently used in this thesis (see section 3.1.4 & 3.1.6) either as fluorescent label for spectroscopic analysis or as quality control to test whether mutations in the ACP domain influence

its phosphopantetheinylation. The *in vitro* reaction with fluorescent CoA was shown to take place quantitatively (see Fig. 3.7), therefore it was also used as internal fluorescence standard to assess the DOL of other labeling reactions (see Fig. 3.8). Analog to the naturally occuring phosphopantetheinylation the fluorescent prosthetic group is attached at the active-site serine of the ACP. Since this non-natural cargo is not accepted by other catalytic domains of the enzyme FAS activity is inhibited by this modification, which constitutes a major drawback for functional studies we want to perform. (ii) The generation of FAS heterodimers relies on the disassembly and subsequent reassembly of dimeric animal FAS. Under my supervision, Daniel Beyer tested this protocol in his master's thesis. Although incubation of our protein samples for up to 1 h at 37 °C showed a positive effect on the dimerization rate (see Fig. S11&S12), we quickly realized that our proteins were not stable enough in low ionic strength buffer as used in the disassembly protocol. Already overnight the majority of protein precipitated and could not be restored in its active form. (iii) The workaround with stochastic fluorophore labeling as utilized in the master's thesis of Maria Dell showed moderate but better results.

FRET is especially powerful at the single-molecule level. While ensemble FRET measurements in solution give an assembly of conformational states, analysis of single FAS molecules would allow monitoring the conformational dynamics of the catalytic cycle of fatty acid biosynthesis in real time. For this purpose the proteins need to be immobilized on microscopic slides without comprising their functional and structural properties and are examined in total internal reflection fluorescence (TIRF) microscopy. We chose an immobilization protocol described by Pal *et al.* who use a sandwich-like coating of the microscopic slide with different layers (Fig. 3.15A).[134] Maria Dell extensively dealt with the surface immobilization of our protein constructs and optimization of fluorescence detection in TIRF microscopy. In collaboration with the group of Prof. Heilemann from the Goethe University Frankfurt, her joint master's thesis was an important step towards the establishment of the smFRET project and illustrated obstacles that still have to be overcome. As test system to establish and evaluate smFRET measurements, the KS-MAT di-domain of mFAS was chosen and amber mutations were placed at different sites in this protein creating a set of distance matrices in the range of 3-12 nm (Fig. 3.16).

The ncAA NorLys2 was incorporated in response to the amber codon and reacted in an IEDDAC with the fluorophore FRET pairs Cy3/Cy5 or ATTO532/ATTO647N supplied as 6-methyl-tetrazine-dyes. In later experiments the ATTO dyes were preferred over the cyanine dyes since they showed better photochemical behavior like brightness and photostability; the cyanine dyes are prone to rapidly bleach upon excitation. To achieve dual labeling, both components of the FRET pair were reacted at equimolar amounts simultaneously in a one-pot reaction (Fig. 3.15B). Fluorescent labeling was extensively examined and further optimized by adjusting stoichiometry of the fluorophores, temperature and reaction time (see section 3.1.4).

No amount of unspecific labeling of the 6-methyl-tetrazine-dyes to the KS-MAT di-domain was observed (Fig. 3.15C). Efforts were also made to isolate the fluorescently labeled protein from the reaction mixture and to remove free excess fluorophore. Purification over Strep-tactin columns, Ni-chelating affinity chromatography or SEC are preferable to buffer exchange in Amicon® Ultra Centrifugal Filters since the latter one results in loss of protein probably due to the hydrophobic or "sticky" nature of the fluorophore conjugates. In later experiments, purification over Ni magnetic beads was used, which allowed working with small sample volumes and adequate protein concentrations.

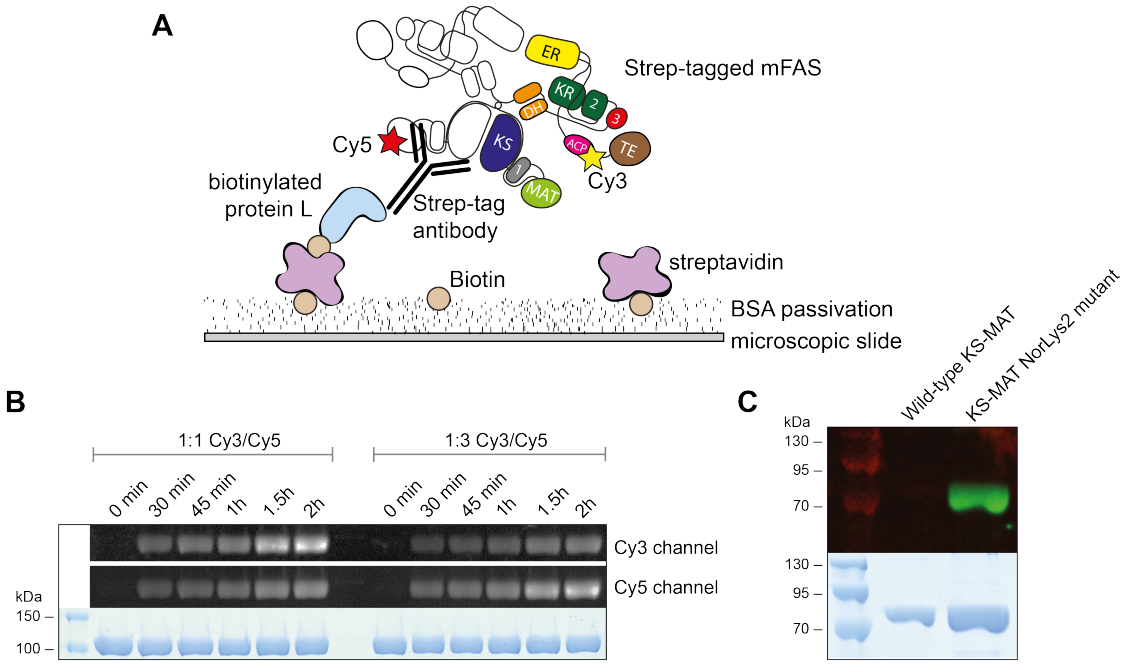


Figure 3.15: Immobilization strategy and fluorescent labeling of KS-MAT constructs. A) A sandwich-like complex consisting of biotinylated BSA, streptavidin, biotinylated protein L and Strep-tag antibody is used to immobilize dual-labeled Strep-tagged mFAS heterodimers onto microscopic slides for smFRET microscopy. B) Generating dual-labeled proteins through stochastic fluorophore labeling. The ncAA NorLys2 was incorporated in response to the amber codon in the KS-MAT di-domain of mFAS and subsequently labeled with an equimolar amount of 6-methyl-tetrazine-sulfo-Cy3 and Cy5 fluorophore. Fluorescence was monitored over time, taking samples of the labeling reaction at different time points and analyzing them on SDS-PAGE (NuPAGE Bis-Tris 4-12 %). Coomassie-stained gel at the bottom, upper panels show in-gel fluorescence from fluorophore conjugation (top: Cy3 channel; middle: Cy5 channel). Two different ratios (1:1 and 1:3) of 6-methyl-tetrazine-sulfo-Cy3 and Cy5 fluorophores were tested. C) Fluorescently labeled KS-MAT di-domain. The ncAA NorLys2 was incorporated in response to the amber codon in the KS-MAT-ACP construct of mFAS and subsequently labeled with 6-methyl-tetrazine-sulfo-Cy3 fluorophore in a IEDDAC. Wild-type KS-MAT-ACP treated with 6-methyl-tetrazine-sulfo-Cy3 fluorophore shows no fluorescence. Left: In-gel fluorescence; Right: Coomassie-stained gel (SDS-PAGE with 7 % polyacrylamide).

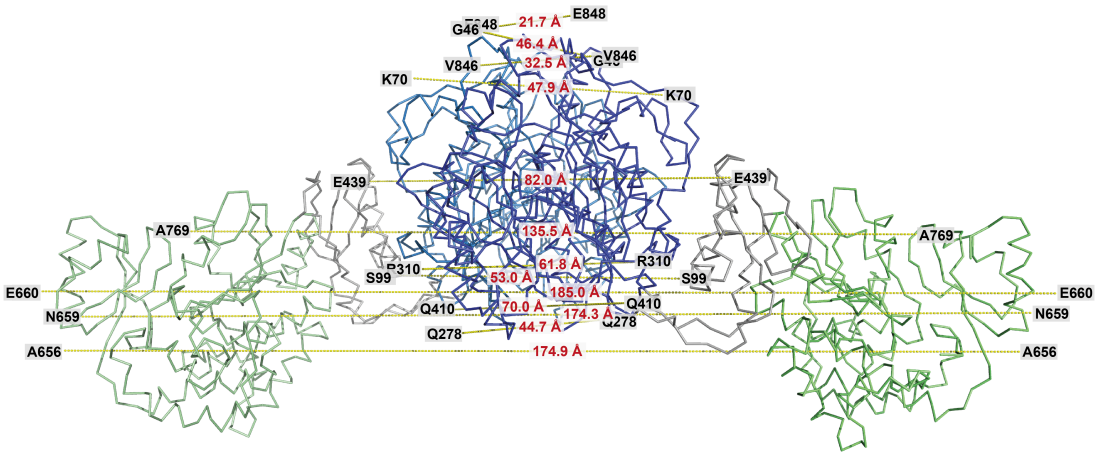


Figure 3.16: Distance matrices in the KS-MAT di-domain of mFAS. A set of distance matrices was planned for the establishment and evaluation of smFRET and EPR measurements. The KS-MAT di-domain of mFAS (PDB ID: 5MY2), here depicted as ribbon, served as test system for this purpose. The KS domains are depicted in blue, and the MAT domains are depicted in green. Amber mutations were placed at different sites in the protein gene. The distance between ncAAs incorporated in response to the amber mutations was estimated proceeding from a lysine mutation at the respective residues and measuring the distance in Å between the Cϵ atoms thereof.

The DOL of the proteins highly varied between protein construct and fluorophores used in this master's thesis; also the click reactions were not very reproducible. The ATTO532 dye gave in general lower DOL (11-29 %) than the ATTO647N dye (40-87 %), and the cyanine dyes only gave a DOL of 3-6 % (for results see Maria's master's thesis). Running behavior of the different dyes on the Illustra NAP-5 columns, which were used for SEC in Maria's project, varies strikingly. Especially free cyanine dye was still present in the protein fractions as was observable in UV-Vis spectra. This might cause failures in the correct determination of DOL.

Despite this, we were able to optimize our immobilization strategy and develop a reliable protocol for stable and reproducible binding of our proteins on microscopic slides. Individual layers of the sandwich-like coating were analyzed by direct stochastic optical reconstruction microscopy (dSTORM) measurements for their optimal surface density, which was determined to be 0.22 ± 0.04 molecules/μm^2 from measurements of fluorescently labeled DNA probes. For dSTORM measurements, particular components of the sandwich-complex (protein L & Strep-antibody) were labeled with Cy5-*N*-hydroxysuccinimid (NHS) esters and a KS-MAT di-domain with fused ACP was modified with CoA 647. Subsequently, the amount of fluorescent molecules immobilized on the microscopic slides was determined for each layer. We were able to regulate the number of CoA 647-labeled KS-MAT-ACP molecules per μm^2 on the surface by adjusting the concentrations of protein L, Strep-antibody and BSA-biotin

used to build up the layers. Furthermore, dilution of fluorescently labeled protein samples with non-fluorescent wild-type protein has proven as adequate method to achieve an optimal surface density so that the individual fluorescent spots could be distinguished. No background fluorescence was observed for each single layer nor the buffer. Unbound fluorophore bleached faster than immobilized fluorophore-protein conjugates on the surface.

The smFRET set-up was validated with high-FRET DNA. The criterion for an unambigious FRET pair is the clear anticorrelation between donor- and acceptor- fluorescence intensity time-traces. If the fluorescence intensity of the acceptor decreases due to bleaching the fluorescence intensity of the donor will have to increase since no FRET can occur. If the fluorescence intensity of the donor decreases due to bleaching the fluorescence intensity of the acceptor upon donor excitation will also have to decrease since no FRET can occur. Only with occurence of blinking or bleaching events these coherences become clearly identifiable. Stable timelines of single molecules on the surface were recorded performing multiple measurements of the same slide. We were not able to clearly identify FRET pairs. Big spots indicated protein aggregates or unspecific fluorophore binding to the protein. This preliminary spectroscopic work was not continued. Nonetheless, the biochemical work of this project was successfully continued, as outlined in section 3.1.6.

3.2.2 Analysis of Domain Dynamics Using EPR Spectroscopy

This work has been conducted in collaboration with the groups of Prof. Michael Göbel and Prof. Thomas Prisner from the Goethe University Frankfurt. Frank Kaiser, doctoral student at the group of Prof. Michael Göbel, synthesized the DBCO spin label used in this thesis and Dr. Alberto Collauto, postdoctoral fellow at the group of Prof. Thomas Prisner, performed the EPR measurements in this thesis.

Next to FP and FRET, EPR is well suited to give information about protein dynamics. EPR and FRET distance measurements are complementary to each other. Instead of fluorophores, paramagnetic centers in form of spin labels are introduced into the protein. The use of two identical spin labels is possible, which bypasses the need for multiple incorporation of different ncAAs. However, EPR spectroscopy cannot be performed at the single-molecule level and is usually carried out at cryogenic temperature giving an ensemble of conformational states.

Our proteins are diamagnetic and thus EPR silent. Therefore we have to site-specifically label them with spin labels. A variety of spin labels exist that can either be conjugated to ncAAs or directly incorporated (genetically encoded) as a paramagnetic ncAA.[136] An *Mm*PylRS mutant has been evolved to genetically encode the nitroxide-bearing ncAA giving side chain SLK-1.[135] Nitroxides are the most popular class of spin labels and are widely employed.

We incorporated *para*-acetyl-L-phenylalanine (AcPhe) or AzPhe into our proteins of interest. The ketone-bearing ncAA AcPhe was reacted with the aminooxy-functionalized nitroxide HO-4120 via oxime formation resulting in side chain K1 (Fig. 3.17A).[137] Kálai *et al.* demonstrated a proof of principle for the conjugation reaction of the cyclooctyne-bearing nitroxide (CO-spin label) with the aryl azide ncAA AzPhe resulting in side chain T1 (Fig. 3.17B).[138] In collaboration with the group of Prof. Göbel, the DBCO-bearing nitroxide (DBCO-spin label)[138] was synthesized by Frank Kaiser, and reacted with AzPhe resulting in the side chain shown in Fig. 3.17C. The spin-labeled proteins were subjected to continuous-wave EPR (CW-EPR) measurements for spin counting (Fig. 3.19), which were performed in the group of Prof. Prisner by Dr. Alberto Collauto.

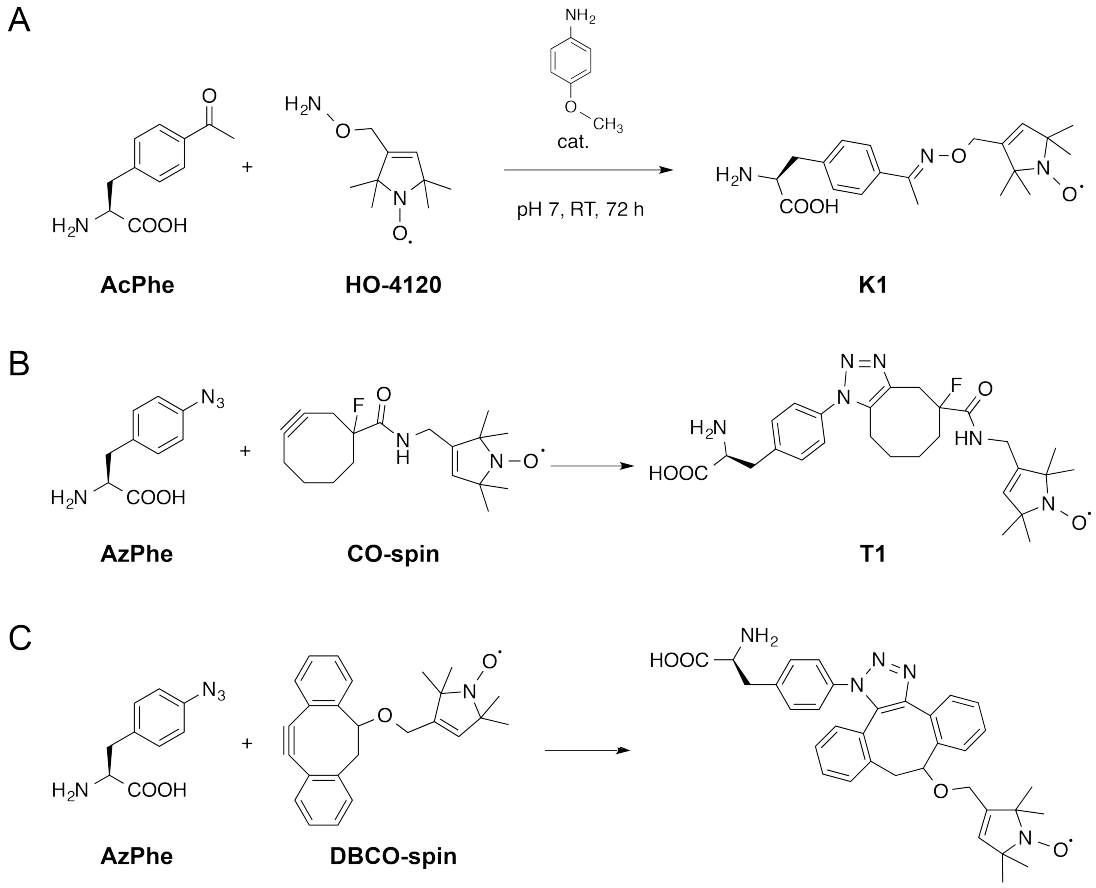


Figure 3.17: Strategies for site-directed spin labeling of proteins. A) Oxime formation of AcPhe and the HO-4120 nitroxide spin label with *para*-anisidine as catalysator at physiological pH, resulting in side chain K1. B) SPAAC of AzPhe and the cyclooctyne-bearing nitroxide (CO-spin label), reported in Schmidt *et al.*,[135] resulting in side chain T1. C) SPAAC of AzPhe and the DBCO-spin label, synthesized by Frank Kaiser.

In previous experiments, the utility of and suitable reaction conditions for the SPAAC between AzPhe and DBCO were tested, labeling AzPhe mutant proteins with a DBCO-Cy5 fluorophore (Fig. 3.18 and Fig. S7). The spin-labeling reactions were performed at room temperature overnight and at three different time points (3 h, 6 h and 24 h) samples for spin counting were taken to assess a sufficient reaction time yielding a high DOL.

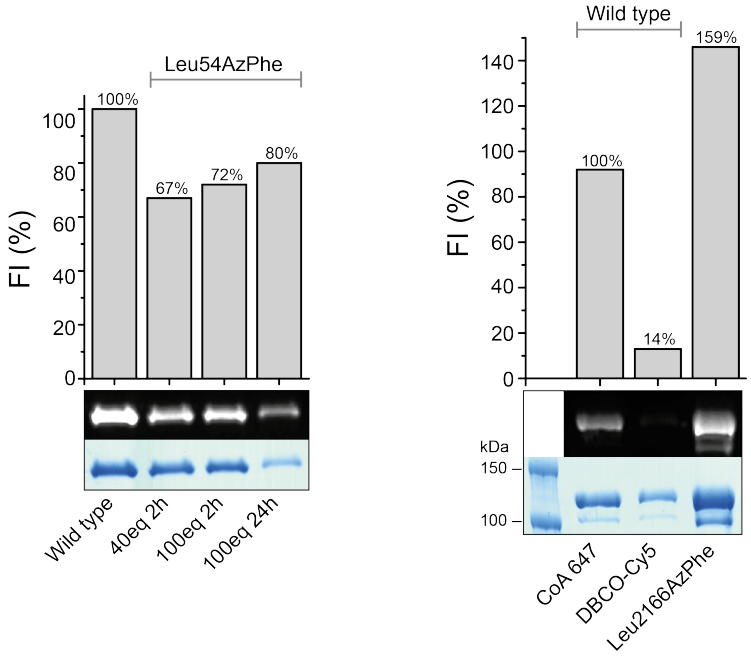


Figure 3.18: Fluorescent labeling of KS-MAT-ACP and ACP-GFP AzPhe mutants in SPAAC with DBCO-functional dyes. AzPhe mutants of KS-MAT-ACP were labeled for 2 or 24 h (left panel) and ACP-GFP for 2 h (right panel) with DBCO-sulfo-Cy5 fluorophore. DOL was determined by relative in-gel fluorescence intensities at wavelength 650 nm compared to wild-type references. All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Coomassie-stained gel. DOL was monitored in dependency of different fluorophore equiv. after labeling reactions for 2 h and 24 h. SDS-PAGE was performed on NuPAGE Bis-Tris 4-12 % gels.

As further site-directed spin labeling reaction, oxime formation was tested. The oxime formation with the AcPhe mutant was not successful and showed no DOL (Table 3.5). The EPR spectra show three sharp lines and give an abnormally high spin concentration (see Fig. 3.19A), which is a clear indication of the presence of free spin label. This in turn prevents the detection of signals from bound species – provided their presence.

Table 3.5: Results from the first spin counting experiment. The DOL is estimated from spin concentrations measured in CW-EPR.

Reaction time	Protein conc.	Spin conc.	DOL
AzPhe mutant in SPAAC			
3 h	29 µM	n.d.	n.d.
6 h	19 µM	1.2 µM	5 %
24 h	32 µM	0.8 µM	5 %
AcPhe mutant in oxime formation			
3 h	9 µM	127 µM	n.d
6 h	8 µM	75 µM	n.d.
24 h	10 µM	60 µM	n.d.

The first spin-labeling experiment with the AzPhe mutant in SPAAC with the DBCO-spin label gave a DOL of about 5 % (Table 3.5). The shape of the EPR spectra is consistent with the molecular weight of the protein (110 kDa) (see Fig. 3.19A). For 100 kDa proteins, a rotational correlation time of 70 ns is expected (t_C=66 ns calculated), which gives rise to the so-called "rigid" spectrum whose outer wings are split by approximately 66 G (lower grey line in the yellow spectrum). In addition, there is a small amount of unbound spin label, which gives rise to the three sharp lines with a separation between edges of approximately 30 G (upper grey line in the yellow spectrum). The labeling efficiency of the SPAAC between the AzPhe mutant and the DBCO-spin label was improved in further experiments (see Fig. 3.19B). A clear signal is coming from the protein indicating a DOL of approximately 49 % after 24 h reaction (and 27 % after 3 h and 6 h) (Table 3.6). The largest part of the spin label is protein-bound. There is a very small amount of free spin label (in the range of 1 µM), which gives practically no contribution to the overall signal. Overall, given the low sample concentrations and the consequent bad signal-to-noise ratio, the accuracy of spin concentration determined by CW-EPR is only at about 10-20 %.

Table 3.6: Results from the second spin counting experiment. The DOL is estimated from spin concentrations measured in CW-EPR.

Reaction time	Protein conc.	Spin conc.	DOL
AzPhe mutant in SPAAC			
3 h	62 µM	17 µM	27 %
6 h	48 µM	13 µM	27 %
24 h	43 µM	21 µM	49 %

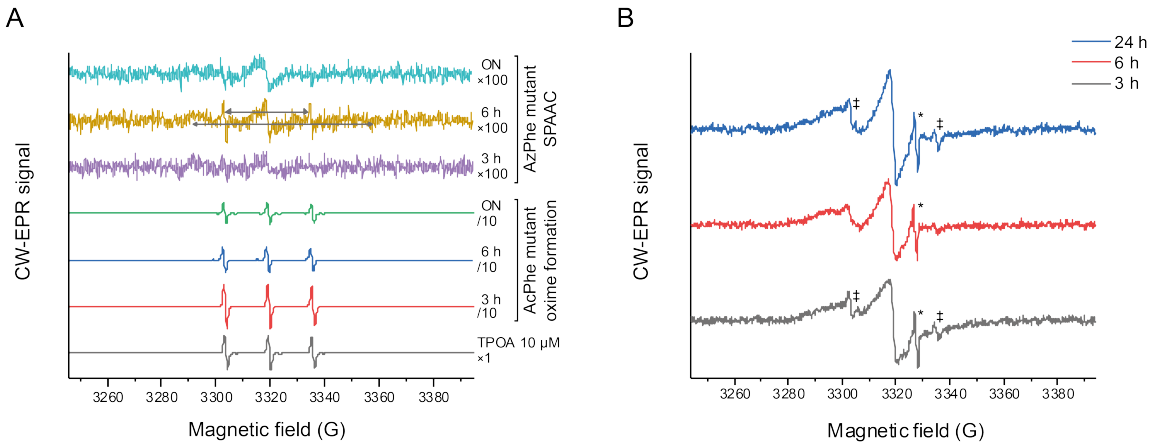


Figure 3.19: Spin counting of spin labeled proteins in CW-EPR measurements. A) First spin counting experiments for KS-MAT-ACP Gln2182 mutants. The AzPhe mutant (purple, yellow and cyan spectrum) was labeled with 10 equiv. of the DBCO-spin synthesized by Frank Kaiser and the AcPhe mutant (red, blue and green spectrum) was labeled with 10 equiv. the HO-4120 spin label in a 150 µL reaction for up to 24 h at room temperature. After 3, 6 and 24 h, samples were taken for the spin counting experiment and measured with CW-EPR. 10 µM nitroxide spin label 2,2,5,5-tetramethyl-pyrrolin-1-oxyl-3-acetylene (TPOA) was used as reference (grey spectrum). Two grey lines are drawn in the yellow spectrum of 6 h SPAAC spin labeling reaction of the AzPhe mutant. The upper grey line shows a separation of approximately 30 G between the edges of the three sharp lines, which arise from unbound spin label. The lower grey line shows a 66 G separation of the outer wings of the rigid spectrum, which arises from spin-labeled protein. B) Second spin counting experiment for the KS-MAT-ACP Leu2166AzPhe mutant. The AzPhe mutant was labeled with 100 equiv. of the DBCO-spin synthesized by Frank Kaiser in a 100 µL reaction for up to 24 h at room temperature. After 3, 6 and 24 h (grey, red and blue, respectively), samples were taken for the spin counting experiment and measured with CW-EPR. The star indicates an impurity or artifact signal from the capillary. The double cross indicates a small signal arising from free spin label.

Our collaborators from the group of Prof. Prisner are interested in the application of pulsed electron-electron double resonance (PELDOR) measurements on biomacromolecules. For the evaluation of their EPR measurements a set of distance matrices would be of advantage. For this purpose, we intend to supply KS-MAT di-domains of mFAS, with spin labels attached at different distances within the protein. Such a set of distance matrices in the KS-MAT di-domain has already been cloned and expressed in the master's thesis of Maria Dell, who used these proteins for the establishment of smFRET measurements (see section 3.2.1, Fig. 3.16). We are especially interested in the following questions addressed to the subconstruct KS-MAT-ACP:

(i) Are the conformational dynamics of this minimal system stochastic or programmed? (ii) Are the conformational dynamics of the protein an intrinsic property or depending on substrate supply? (iii) How does the substrate load of ACP change conformational dynamics? (iv) How do mutation or changes of physical parameters interfere in conformational dyna-

mics?
Therefore, site-directed spin labeling of the ACP domain and comparison of the spectral shapes of KS-MAT-ACP loaded with different cargos will be performed in CW-EPR spectroscopy.

3.2.3 Analysis of Domain-Domain Interactions Using Photocrosslinking

The non-covalent domain-domain interactions in multifunctional type I FASs are relatively weak and transient. Moreover, the mobility of the ACP domain makes the structural elucidation difficult. A strategy to investigate such domain-domain interactions is crosslinking.[129–131,139] Ye *et al.* proposed a photocrosslinking assay for reporting protein interactions in PKSs and FASs.[140,141] They used genetically encoded photoreactive BzPhe incorporated at the position of the active-site serine of ACP in response to an amber codon to crosslink ACP from *E. coli* type II FAS with its cognate KS partner proteins FabB, FabH and FabF. With a panel of ACPs from different sources (6-deoxyerythronolide B type I PKS, actinorhodin type II PKS and *Streptomyces coelicolor* type II FAS) they tested the putative orthogonality of the ACP:KS interactions. Non-cognate protein interactions gave no crosslink. Introducing BzPhe at different sites allowed for probing the influence of phosphopantetheinylation and substrate loading with malonyl and octanoyl moieties on the ACP:KS interaction. Furthermore, the ACP:FabF interaction interface was mapped using mutagenesis studies combined with the photocrosslinking assay.

Table 3.7: Amber mutation sites for photoreactive ncAAs used in the photocrosslinking assay. ACP numbering as in PDB ID: 2PNG[50] and mFAS numbering as in UniProtKB: P19096.

ACP numbering	mFAS numbering
Gly01	Gly2113
Ile16	Ile2128
Leu17	Ile2129
Gly18	Gly2130
Ser38	Ser2150
Leu39	Leu2151
Met40	Met2152
Leu54	Leu2166
Met58	Met2170
Gln70	Gln2182
Glu71	Glu2183
Ala79	Ala2191

Photocrosslinking Assay

Based on this strategy, we established our photocrosslinking assay. This project was in large part performed by Felix Lehmann, who completed his master's thesis under my supervision. We chose to incorporate both BzPhe and AzPhe at different sites in the ACP from mFAS, among them also the active-site serine position (Table 3.7 & Fig. 3.21). The protein mutants expressed well with an average yield of about 3.3 mg per cell pellet for ACP-GFP mutants and 0.45 mg per cell pellet for KS-MAT-ACP mutants (also valid for Ala mutants discussed in the following paragraphs). In average 10 g cell pellet were obtained from 1 L expression culture, varying between protein expressions from 5 to 15 g. No conclusion about an influence of the amber mutation position on expression efficiency can be made because biological replicates are missing. Intermolecular crosslinking between ACP-GFP and the KS-MAT di-domain as well as an intramolecular crosslinking within the KS-MAT-ACP fusion construct were examined. In negative control experiments the wild-type proteins of KS-MAT and KS-MAT-ACP showed a defined double band upon UV irradiation running between 150 and 250 kDa in the SDS-PAGE (Fig. 3.20B, D & E). SDS-PAGE analysis of the photocrosslinking experiments showed specific crosslinked protein bands upon UV irradiation.

The Intermolecular Photocrosslinking Reaction

In the intermolecular photocrosslinking reaction between ACP-GFP and KS-MAT a putative crosslinked protein band appeared at a size of about 140 kDa, which corresponds to the sum of both protein masses ($MW_{ACP\text{-}GFP}$=39 kDa, $MW_{KS\text{-}MAT}$=100 kDa) (Fig. 3.20A, B, D, E & F). The intensity of this band is proportional to the excess of ACP-GFP used in the reaction. With increasing ACP-GFP concentrations also the background increases, caused by unspecific reactions and impurities of the sample. The photocrosslinking probes of ACP-GFP showed some amount of non-specific self-crosslinking. Best signal-to-noise ratio was achieved using an excess of 80 equiv. ACP-GFP.

The Intramolecular Photocrosslinking Reaction

The banding pattern of the intramolecular photocrosslinking of KS-MAT-ACP did not differ from its wild-type negative control upon UV irradiation. A truncation band is observed at around 100 kDa for all KS-MAT-ACP proteins as already discussed in sections 3.1.5 & 3.1.6. Time-course measurements of the photocrosslinking reaction revealed that a putative crosslinked protein band just under 250 kDa rapidly forms after only 1 min UV irradiation. While the double band observed for the wild type, the negative control, starts to appear much later at 2 min 30 s and is readily visible after 5-10 min (Fig. 3.20B, D & E). With a size of approximately 220 kDa the crosslinked protein band corresponds to dimeric KS-MAT-ACP

protein. Intensity of the crosslinked protein band increases over time and blurs into the double band observed also for the wild type. Additionally to the defined protein bands a light shade was observed in the photocrosslinking reaction right above the KS-MAT-ACP protein band at 110 kDa, which was not present in the wild-type negative control. At longer UV irradiation times (>15 min) the protein bands started to fade and background shade started to blend in.

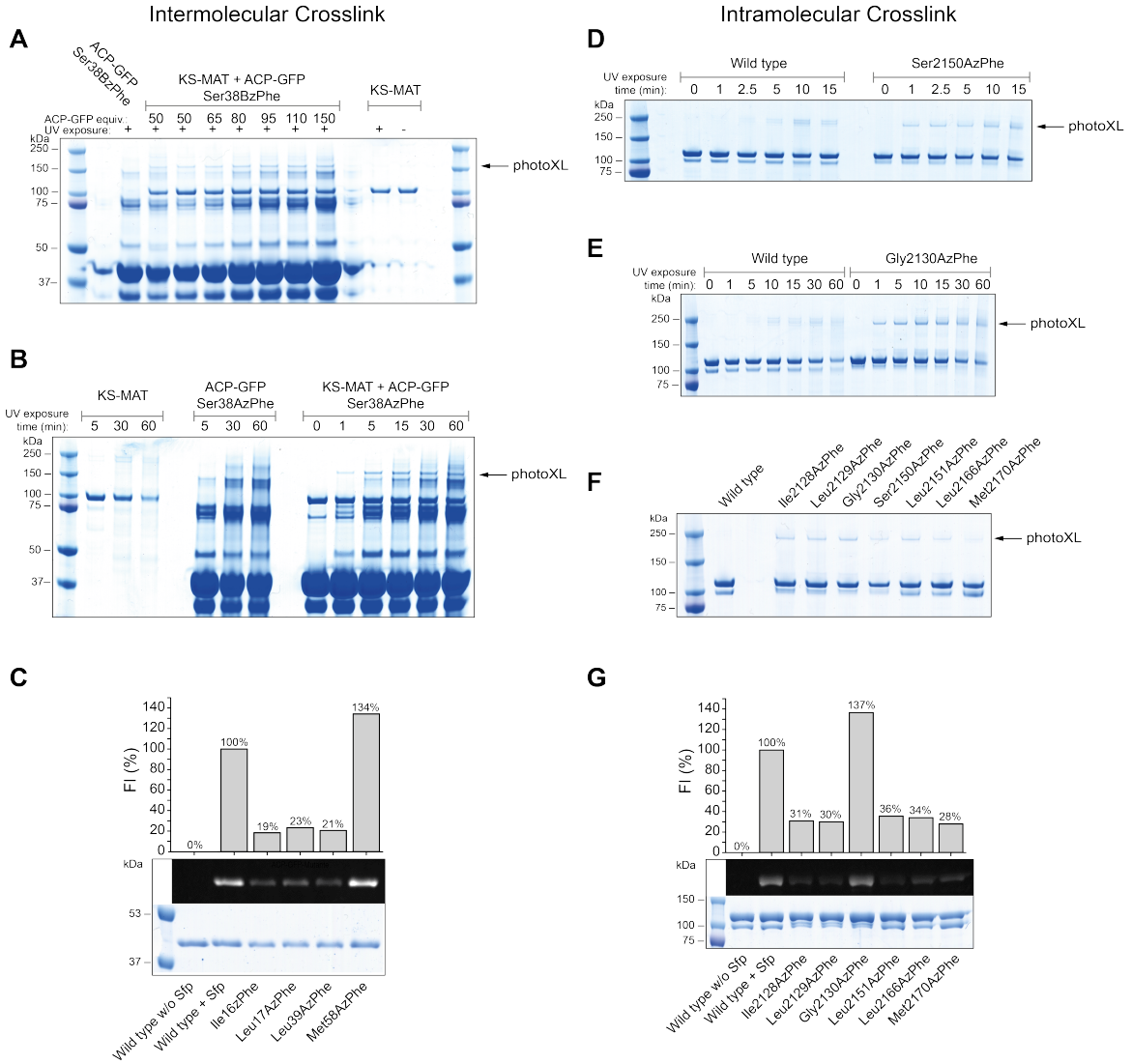


Figure 3.20: Establishment of the intermolecular (left panel) and intramolecular (right panel) photocrosslinking assay. Crosslink bands were detected using SDS-PAGE (NuPAGE Bis-Tris 4-12 %). A) Intermolecular photocrosslinking reaction between KS-MAT and ACP-GFP Ser38BzPhe mutant. Multiple photocrosslinking reactions were performed keeping the KS-MAT concentration constant at 0.5 mg/mL and increasing the equiv. of the ACP-GFP mutant. B) Intermolecular photocrosslinking reaction between KS-MAT and ACP-GFP Ser38AzPhe mutant. Photocrosslinking was monitored for UV exposure times up to 1 h and

samples were taken at different time points. Control experiments were performed, exposing individual KS-MAT and the ACP-GFP mutant proteins to UV light. to UV light. Specific crosslink bands are observed at approximately 150 kDa, indicated by an arrow (A & B). C) CoA 488 assay of ACP-GFP AzPhe mutants. Wild-type ACP-GFP and AzPhe mutants were enzymatically labeled with fluorescent CoA 488 by Sfp. DOL determined by relative in-gel fluorescence intensities at wavelength 488 nm compared to wild-type ACP-GFP reference. All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Coomassie-stained gel. D) Intramolecular photocrosslinking reaction of KS-MAT-ACP Ser2150AzPhe mutant. Photocrosslinking was monitored for UV exposure times up to 15 min and samples were taken at different time points. E) Intramolecular photocrosslinking reaction of KS-MAT-ACP Gly2130AzPhe mutant. Photocrosslinking was monitored for UV exposure times up to 1 h and samples were taken at different time points. F) Intramolecular photocrosslinking reaction of KS-MAT-ACP AzPhe mutants exposed to UV light for 1 min. Wild-type KS-MAT-ACP was exposed to UV light in control experiments. Specific crosslink bands are observed between 150 and 250 kDa, indicated by an arrow (D, E & F). G) CoA 488 assay of KS-MAT-ACP AzPhe mutants. Wild-type KS-MAT-ACP and AzPhe mutants were enzymatically labeled with fluorescent CoA 488 by Sfp. DOL determined by relative in-gel fluorescence intensities at wavelength 488 nm compared to wild-type KS-MAT-ACP reference. All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Coomassie-stained gel.

Position Dependency of the Photocrosslinking Reaction

Crosslinking was investigated dependent on the incorporation site of the photoreactive ncAA AzPhe. Both assay systems, the intermolecular crosslinking between ACP-GFP and KS-MAT as well as the intramolecular crosslinking within KS-MAT-ACP gave comparable results (Fig. 3.20F & S8). Crosslinks were observed for AzPhe mutation sites Gly2113, Ile2128, Leu2129, Gly2130, Ser2150 and Leu2151. Position Leu2166 showed no crosslink in the intermolecular assay system but did show a crosslinking band for KS-MAT-ACP. Met2170 behaved the other way around. No crosslinks were observed for Met2152, Gln2182, Glu2183 and Ala2191. Not all positions were tested in both assay systems. All proteins were isolated as apo proteins and subsequently tested in an *in vitro* CoA 488 assay for their phosphopantetheinylation ability. Wild-type protein was used in the same reaction as reference and quantitative reaction was assumed. Phosphopantetheinylation rate was calculated from fluorescence intensities of the fluorescent gel. Most proteins showed a phosphopantetheinylation rate of 20-30 % (Fig. 3.20C & G). Only mutation Met2170AzPhe in ACP-GFP and Gly2130AzPhe in KS-MAT-ACP showed high fluorescence intensities in the CoA 488 assay,

which even exceeded the reference. Surprisingly, the Met2170AzPhe mutation showed different behaviour in the ACP-GFP and the KS-MAT-ACP construct giving a high (quantitative) and a low (30 %) phosphopantetheinylation rate, respectively. Enzymatic digestion and subsequent mass spectrometric analysis of the protein bands confirmed crosslinked species. Tandem mass spectrometry (MS/MS) was performed by our collaborator Andreas Linden from the group of Prof. Urlaub at the MPI in Göttingen. Although reasonable results were obtained, the data were not meaningful enough with overall very few protein fragments detected in MS due to low sample quality.

Alanine Scanning Mutagenesis on the KS Surface Analyzed by the Photocrosslinking Assay

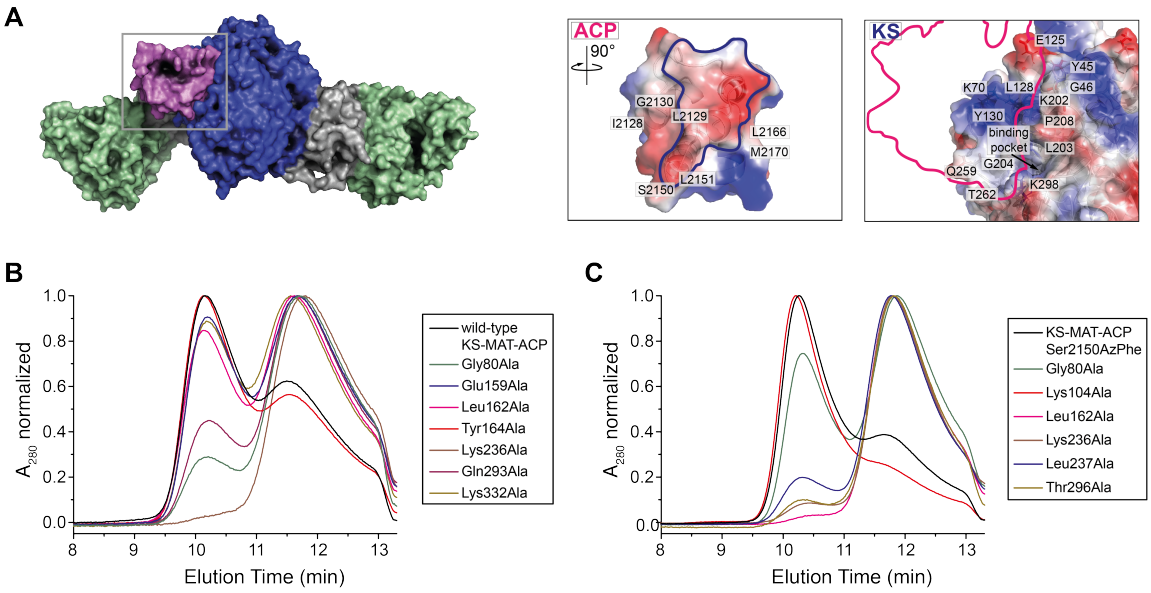


Figure 3.21: KS surface mutation analyzed in the photocrosslinking assay. A) Surface model of the KS-MAT di-domain with ACP bound to the KS (PDB IDs: 5MY2 for KS-MAT and 2PNG[50] for ACP; blue: KS, green: MAT, grey: interdomain linker, pink: ACP). The grey box indicates the section, which is zoomed in on the right. The surface of the ACP and the KS interface is colored according to its relative electrostatics (blue: positive and red: negative). Amber mutation sites in the ACP and alanine mutation sites in the KS are labeled, as well as the binding pocket of the KS. The silhouette of the KS interface is depicted in blue and in pink for the ACP. B & C) HPLC-SEC spectra of the KS-MAT-ACP mutants from the alanine scan without (B) and with Ser2150AzPhe mutation (C). Dimeric proteins elute at approximately 10 min and monomeric proteins at approximately 12 min. Individual spectra for all mutants can be found in Fig. S11&S11.

To investigate the ACP:KS interface more precisely and to identify so called hotspots of this protein-protein interaction, alanine scanning mutagenesis was performed on the KS surface. Viegas *et al.* modeled the interaction of ACP with the catalytic domains KS, MAT, KR, DH and ER in human FAS in molecular dynamics simulations.[76] Proceeding from these structures, surface mutations were planned (Tyr45Ala, Gly46Ala, Lys70Ala, Glu125Ala, Leu128Ala, Tyr130Ala, Lys202Ala, Leu203Ala, Gly204Ala, Pro208Ala, Gln259Ala, Thr262Ala and Lys298Ala; mFAS numbering, UniProtKB accession code P19096) (Fig. 3.21). The KS-MAT mutants were analyzed by HPLC-SEC and TAL activity assay. The effect of the surface mutations on photocrosslinking was assayed within the KS-MAT-ACP construct with AzPhe incorporated at the active-site serine (Ser2150). All surface mutants showed significantly different pronounced crosslinked protein bands in the SDS-PAGE, which were quantified by densitometric analysis (see Fig. S9). HPLC-SEC of the proteins showed different degrees of dimerization (Fig. 3.21B & C), which can have huge impact on the photocrosslinking assay. Therefore, we tried to count back photocrosslinking efficiency taking into account the monomer/dimer ratio of the proteins. Likewise does the TAL activity assay largely depend on the monomer/dimer ratio of the proteins since only the dimeric KS-MAT-ACP is the active form. For monomeric proteins like the Lys236Ala mutant we were not able to measure any activity in the TAL assay. All other protein mutants that showed some degree of dimeric state, showed activity. After correction for the monomer/dimer ratio, all activities lay in the range between 60-120 nmol/min/mg protein (wild-type reference from literature (152 ± 15) nmol/min/mg protein)[86].

3.2.4 Influence of Post-Translational Modifications on FAS Activity

Although the mFAS heterologously expressed in *E. coli* seems to retain integrity – regarding stability, oligomeric state and phosphopantetheinylation – the protein seems to be less active than enzymes isolated from tissue. Catalytic activity of animal FAS might be increased by post-translational modifications, which are absent when expressed in *E. coli*. Several acetylation and phosphorylation sites are present in animal FASs. Although our amber codon suppression toolbox includes acetylated amino acids, we have not really exploited their use so far mostly owing to their low incorporation rate. Dr. Alexander Rittner tested the incorporation of acetyllysine in the KS-MAT di-domain once and obtained some amount of full-length protein. The protein sequence of the *Mb*PylRS, which was derived from the pAcBac1.tR4-MbPyl plasmid available from Addgene and which was used in our studies,[113] contains several point mutations and does not resemble the wild-type *Mb*PylRS gene. To further improve this application, we started to clone *Mb*PylRS from another source into our vector systems pAC^{U} and pAC^{E}. The sequence of these plasmids has not yet been fully verified.

First tests in the reporter assay did not show incorporation of PrLys by the new orthogonal *Mb*PylRS/tRNA^{Pyl} pair. Amber codon suppression of phosphoserine has been reported,[142] but has not been added to our amber codon suppression toolbox, yet. Nonetheless, further effort will be made to establish this application and test the effect of those post-translational modifications in our enzyme kinetic assays.

4 Discussion & Outlook

Parts of this chapter have previously been published in: Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878. For individual contribution to the writing process as well as copyright and Creative Commons license, please consult the Statement of Personal Contributions.

Conformational variability is a fundamental property for the catalysis of many enzymes and especially megaenzymes such as the animal type I FAS.[11,53,57,76] In the last decade, amber codon suppression has emerged as a powerful tool to investigate structure and dynamics of proteins. The opportunity to label proteins site-specifically with bioorthogonal functional groups has led to novel biophysical probes *in vitro* and *in vivo*.[14,20,21,118,143] So far, this technique has commonly been used to study small proteins, but has not been applied for large and complex multidomain systems. In this thesis, an efficient way to incorporate ncAAs site-specifically into animal type I FAS with subsequent labeling was established. Such probes make a range of biological applications accessible such as photocrosslinking, fluorescence and EPR spectroscopy.[21,135,144] In case of animal type I FAS, labeled proteins offer for example the prospect of addressing some of the fundamental questions to carrier domain-mediated substrate shuttling, i. e. towards the time scale of domain-domain interactions or the influence of loaded intermediates on the mobility of the ACP domain.

Reporter Assay Design

Incorporation of ncAAs by amber codon suppression into such a complex enzyme like animal type I FAS remains a challenge, owing to the sensitivity of the protein to buffer and temperature changes as well as salt concentrations.[132] To circumvent high consumption of resources we decided to approach this task by reducing complexity. Specifically, focus was set on the small ACP domain and expression volumes were scaled down to 2 mL cultures, which can be conveniently handled and analyzed in 96-well format. In order to achieve fast screening of amber codon suppression conditions, the fluorescent protein GFP was C-terminally fused to the ACP domain. This set-up allows to easily monitor the incorporation of the ncAA by the fluorescence emerging from the full-length fusion construct only. While a universal test system like GFP is typically used to assess general efficiency of amber codon suppression systems, a C-terminal fusion of GFP to a desired domain from a multidomain protein has the

advantage of reflecting position dependencies of ncAA incorporation. The initial screening at low complexity was followed by upscaling protocols to eventually make the multidomain protein available as labeled probe.

Screening of the Amber Codon Suppression Toolbox

Employing this set-up, two reported plasmid systems[12,13] were used and the incorporation efficiencies across eleven different ncAAs were compared, which predominantly allow for copper-free click chemistry (see Fig. 3.2). In our set-up, the two suppressor plasmids pAC^{U} and pAC^{E} derived from pUltra and pEVOL vectors, respectively, gave similar performance. The pAC^{U} vector was used for its ease in protein production, as not requiring additional induction of the suppression system along with the protein of interest. In general, the TyrRS from *M. jannaschii* performed much better than the PylRS from *M. mazei* or *M. bakeri* as the latter suffered from a high tendency to aggregate in *E. coli*.[122,123]
As the two ncAAs AzPhe and NorLys2 showed optimal incorporation efficiencies of 26-35 % and allow for copper-free click chemistry – e. g. in fluorescent bioconjugation or site-directed spin labeling for fluorescence or EPR studies – we chose to proceed further with these ncAAs incorporated with the pAC^{U} vector. Another important and practical aspect of these two selected ncAAs is that they can be resourced by well established and easy synthetic routes, which provide the compounds in multigram scale with good yields but yet at relatively low cost. Furthermore, the ncAAs BzPhe and AcPhe, which showed high incorporation efficiencies of 54-59 % but suffered in incorporation fidelity with 8-21 % fluorescence of the negative samples found application in our photocrosslinking and EPR studies.

Positionscreening of ncAA Incorporation

In addition to the choice of a suppression system, the reporter assay revealed that the site of ncAA incorporation in the ACP fold was critical for suppression efficiency. Although all amber mutation sites were located on the surface of the ACP domain, to minimize the disturbance of the overall protein fold or any protein-protein interactions in mFAS, both AzPhe and NorLys1 were differently sensitive to incorporation sites. Whereas AzPhe was tolerated at all tested sites, NorLys2 was only introduced sufficiently into a disordered loop region. The higher tolerance to incorporate AzPhe at different positions may be explained by its smaller size preserving the integrity of the ACP fold. We cannot draw any conclusions to the position dependency of incorporation efficiency in the sequence context of the incorporation sites.[125]

Upscale of the Reporter Assay

Overall, the results of the upscale experiment and the reporter assay agree with one another. Incorporation efficiencies of the two different ncAAs at five different mutation sites were in line with data received from the reporter assay with the only exception of NorLys2 incorporation at site Gly01. From SEC profiles of purified proteins, we were further able to conclude that modification with ncAAs did not disturb the protein fold. With a drop in expression yield to 30–40 % of the non-mutated reference construct also the access to the protein remained satisfyingly high.

Site-Specific Labeling

As a quality control, the constructs were finally also employed in testing bioconjugation with fluorophores. Although there is a discrepancy between the DOL determined using relative in-gel fluorescence intensities and UV-Vis absorbance both methods agreed that for our set-up the azido-BCN reaction is more efficient than the norbornene-tetrazine reaction. The ncAA NorLys2 was not only preferred over NorLys1 because it showed better incorporation efficiencies, but also because it achieved higher DOL in the reaction with the 6-methyl-tetrazine dyes (see Fig. S5). AzPhe gave gratifying high DOL both in the SPAAC with BCN- and DBCO-functional groups. The amount of unspecific binding of the BCN and DBCO dyes to larger protein constructs like KS-MAT-ACP was somewhat higher than for the 6-methyl-tetrazine dyes, but mostly still in a low range of about 9-13 %.
The difference in DOL determined by in-gel fluorescence intensities and UV-Vis absorbance spectra can be explained as originating from different reaction conditions and sample preparations performed for analysis in SDS-PAGE and spectroscopy. The quantum efficiency is determined for the free fluorophore and may be differently affected by the microenvironment within the native and the denatured protein. Furthermore, the formula to calculate the DOL (equations 5.2–5.4) is moreover weak in determining correct values for samples of low protein concentrations (<0.02 mg/mL). This has to be kept in mind analyzing the results of Maria's master's thesis since they were obtained with low protein concentrations.
Data may also indicate photobleaching of the CoA 647 conjugate in the reference sample (see Fig. 3.7).

Modification of Full-Length mFAS

Application of the identified conditions on the excised ACP domain to the full-length mFAS has led to fluorescently labeled samples that fulfill the criteria of high protein quality including a maintained oligomeric state and enzymatic activity. Nevertheless, initial experiments

have revealed two major issues, which have to be tackled in future studies.

First of all, the dimeric nature of mFAS, which is an analogous property of structurally related PKSs, has resulted in pairing of truncated polypeptide chains from insufficient suppression of the amber codon with full-length chains. These unwanted heterodimers were not separated with the applied purification strategy and might interfere in downstream experiments where dual-labeling or native-like full-length mFAS are required. The reduced activity of the ncAA-modified mFAS may possibly be attributed to truncated proteins within the sample, but this cannot fully explain the complete effect and it further seems that the modification decreases the catalytic efficiency of the enzyme.
Although Rangan *et al.* provided protocols to disassemble and reassemble mFAS *in vitro*,[132] this method is unfortunately not suitable to generate full-length homodimers owing to protein instability. The use of a GRO, which releases the amber codon from its translation stop function and creates an actual blank amber codon with default RF1 competition would be able to solve this problem. In our experiments the GRO C321.ΔA we received from the Church lab[22,23] did not perform well. The strain C321.ΔA was engineered from *E. coli* MG1655 with complete amber codon replacement and RF1 deletion and suffers from comprised reproductive strength of the host cells. Another GRO engineered from the BL21 (DE3) *E. coli* strain is available from the RIKEN laboratories Yokoyama and Sakamoto,[24] which has only 95 amber codon replacements and is told to perform much better without comprising the reproductive strength of the expression host. We reached out for this GRO B-95.ΔA to test its performance in our set-up; the material transfer agreement is still under review. In the meantime, we are currently working on forming heterodimers *in vivo*. A second gene encoding mFAS with an alternative C-terminal FLAG-tag was included on the mFAS expression vector creating a pETDuet vector. Both mFAS genes are expressed bicistronically and mFAS dimers form co-translationally *in vivo*. Full-length dimers were isolated over the two C-terminal tags using two subsequent affinity chromatography purification steps. This might further provide the opportunity to place fluorescent labels at different positions in both polypeptide chains of the heterodimer, which would be beneficial for our FRET applications. In parallel we are expanding our GCE toolbox for multiple incorporation of ncAAs in response to a second nonsense codon.

Secondly, the relative low degree of fluorescent labeling in the model constructs and even more importantly in the full-length mFAS is another important aspect. The discrepancy in DOL suggests that efficient amber mutation sites may not be as accessible in a multidomain protein, though placed on the surface. This is a major drawback in fluorescent studies. The utilized clickable fluorescent probes in this thesis, which were chosen due to their commercially availability suffer from relatively low reaction kinetics.[115,116] A limitation for an enzyme such as mFAS is that it cannot be incubated in click reactions overnight. Especially

the click reactions and ncAAs used in our fluorescent studies, which would greatly benefit from higher DOL were not optimal. A higher DOL could be achieved by improved chemoselective reactions with faster click-kinetics. It is noted that the focus of this field of chemistry is optimizing reactive probes and inventing new interesting compounds.[145] We observed very slow click kinetics for NorLys1 and improved the DOL significantly by the use of NorLys2. Still, the IEDDAC rate can be further improved by choosing other reaction partners. Even slight electronic changes in the tetrazine structure have a major impact on the reaction rate.[116] The DOL can possibly be improved by adding another tetrazine-fluorophore to our NorLys2-bearing proteins since the 6-methyl-tetrazine that was used in our labeling reactions has the lowest reaction rates reported. Tetrazine-fluorophore conjugates would have to be synthesized in a simple one-step reaction ourselves since they are not commercially available. The reaction between BCN and tetrazines is even more efficient than the norbornene-tetrazine reaction and the IEDDAC of tetrazines with TCOs have the fastest reaction rates reported to date. All components (norbornene, tetrazine, BCN, TCO) can be introduced into proteins as ncAAs bearing these functional groups. Incorporation efficiencies vary between those different ncAAs. In our screenings, only the incorporation of NorLys1 and NorLys2 was sufficient enough and made us stick to rather slow IEDDAC. The lab of Ryan Mehl reported the synthesis and incorporation of a significantly improved tetrazine-containing ncAA Tet-v2.0,[146] which was shown to have better incorporation efficiencies and rate constants in the click reaction than its precursor Tet-v1.0 reported by Seitchik *et al.*,[147] which was used in our reporter assay. Utilizing Tet-v2.0 in our studies would enable labeling with BCN- or even better TCO-fluorophores and thus exploiting rapid and quantitative labeling. The fluorophore conjugates would have to be synthesized ourselves and will be somewhat more challenging and expensive than the tetrazine-conjugates. The balance between high DOL, effort and costs has to be found.

Applications to Analyze Domain Dynamics

The projects presented in chapter 3.2 of this thesis are a first step towards the investigation of the conformational dynamics of the multidomain enzyme mFAS. Partly initial experiments were performed to exploit the amber codon suppression toolbox and its range of possible applications, establishing this technology in our lab.

Fluorescence Polarization Assay

Fluorescent probes find a variety of biological and biophysical applications. Spectroscopic analysis of fluorescently labeled mFAS probes in FP or FRET microscopy, especially on the single-molecule level, is a desirable goal. The FP assay, introduced in this thesis, was not

investigated in its full extend. It suffered not only from very low DOL but also from low protein yield. Due to the low amount of protein, replicates are missing and the labeled protein samples were repeatedly measured and diluted. Repeated excitation of the fluorophores caused photobleaching and negatively influenced FP results. The inherent low affinity of protein-protein interactions within FAS requires a high concentration of KS-AT protein to achieve full saturation of ACP binding. The measurement points and concentration of KS-MAT were by far not high enough to reach the saturation level in our first experiments. With the improved protein preparation protocols we will be able to generate a sufficient amount of labeled protein to conscientiously establish the FP assay.
With the envisaged FP assay we will be able to compare binding affinities of apo and holo ACPs. Furthermore, we will be able to investigate chain-length effects on the binding affinity by loading different cargos onto the ACP.

FRET Spectroscopy

Although no FRET pairs were identified unequivocally in the smFRET experiments of this thesis, the FRET project carried out within Maria's master's thesis build a great foundation for further spectroscopic work on mFAS. Additionally to low and non-reproducible DOL, which can be improved by better choice of the ncAA and underlying chemoselective reactions, dual-labeling posed a problem in the FRET project. The disassembly/reassembly protocol by Rangan *et al.*[132] would be a nice method to generate heterodimeric dual-labeled mFAS constructs. In theory, two different mFAS preparations, which are labeled with different fluorophores, are mixed together and disassembled in low ionic strength buffer at 4 °C. Subsequent reassembly of this mixture of mFAS polypeptide chains should lead to the formation of mFAS heterodimers, which can be isolated from the mixture by dual affinity chromatography if two different affinity tags were used in the protein preparations. While our proteins of interest are stable at the reassembly conditions for 1 h at 37 °C, they are too sensitive for the disassembly conditions and tend to aggregate in low ionic strength buffer.
A fluorescent CoA analog was used to phosphopantetheinylate the ACP domain. This method has proven reliable and reproducible giving close to quantitative DOL. In combination with amber codon suppression, FRET pairs can be easily introduced into proteins carrying an ACP domain. Unfortunately, the introduction of the fluorescent label is restricted to the active-site serine of the ACP. Furthermore, modification with the fluorescent prosthetic group generates inactive enzymes. For functional studies, which investigate active enzymes, holo ACP that can be loaded with different substrates is required.
For this reason, another way to site-specifically label proteins with two different fluorophores has to be accomplished. In a one-pot reaction proteins were treated with two different fluorophores simultaneously. Although this stochastic fluorophore labeling led to generation of dual-labeled proteins, this method has some clear limitations: (i) Stochastic distribution

of dual-labeled mFAS. The portion of heterodimeric mFAS, which is labeled with two different fluorophores of the FRET pair, is sparsely populated. The portion of homodimeric mFAS, which is labeled with the same fluorophore twice, makes up already two thirds (double ATTO532 and double ATTO647N labeling), if not taking into account the single-labeled or non-labeled protein portion, which did not react quantitatively. This assumption is of course not valid, since no quantitative labeling was reached, for a fact also owing to truncated polypeptide chains in the protein preparation. (ii) Discrepancy in DOL for donor and acceptor fluorophore. The DOL of the ATTO647N dye, which served as acceptor fluorophore of the FRET pair, was much higher than for the donor fluorophore ATTO532. In smFRET measurements a donor/acceptor stoichiometry of 0.1 to 0.3 was calculated from fluorescence intensities. It is not clarified if this stems from dissimilar reaction rates of the two fluorophore conjugates. This resulted in mainly double ATTO647N labeled homodimeric FAS. Treatment with higher donor/acceptor ratios in the one-pot labeling reaction could solve this problem. Both (i) and (ii) strongly affect the occurence of FRET pairs observed in smFRET measurements. (iii) Unspecific binding of fluorophore to the protein. The high discrepancy in DOL between donor and acceptor fluorophore and poor stoichiometries of 0.1 might result from unspecific binding of the acceptor fluorophore to the protein, indicated by continuously decreasing fluorescence intensities upon direct acceptor excitation. It is not clear why the ATTO647N dye should be more prone to unspecific binding than the ATTO532 dye. As seen from fluorescence gels, unspecific binding of the 6-methyl-tetrazine-ATTO647N dye was hardly detectable for ACP-GFP and KS-MAT constructs, and at a low amount of 9 % for full-length mFAS (see Fig. 3.8, 3.12, 3.15 and S5). Obviously – if it was the case – a high amount of unspecific bound acceptor fluorophore would render the identification of FRET pairs in the smFRET measurement impossible.

The FRET project would benefit from the more technical advanced method of multiple ncAA incorporation in future. At this point we are experienced enough in the GCE technology to tackle this challenge. The introduction of another nonsense codon (the ochre codon UAA) and e. g. the combined incorporation of AzPhe and NorLys2 using the evolved *Mj*TyrRS in combination with *Mm*PylRS, respectively, are planned. Co-transformation of more than two plasmids can be circumvented by cloning a dual-incorporation suppressor vector, encoding both aaRS/tRNA pairs on a single plasmid. The two ncAAs can be placed on the same polypeptide chain, resulting in homodimers containing four ncAAs in total for the introduction of two FRET pairs. Usually, only one FRET pair per protein is preferred for analysis. This can be achieved by forming heterodimers *in vivo* as mentioned above, since the disassembly/reassembly protocol[132] cannot be realized for the aforementioned reasons. Leaving one mFAS gene unaltered, and introducing two nonsense codons in the other gene, will generate heterodimers that can have one FRET pair within only one polypeptide chain of the dimer. Introducing a nonsense codon in each of the two genes will generate heterodimers that can

have one FRET pair allocated on both polypeptidechains of the dimer.

In the course of Maria's master's thesis a stable and reproducible immobilization protocol was developed, based on the sandwich-like coating reported by Pal *et al.*.[134] Each individual layer was thoroughly evaluated by dSTORM measurements to find the optimal surface density for smFRET measurements. Still, improvements could be made: (i) The less layers on the microscopic slide, the easier the adjustment of surface density becomes. Reducing the number of layers, removing the protein L layer from the sandwich complex by making use of biotinylated antibody can be envisioned with little effort. Introducing a biotin-tag into our proteins of interest would be accompanied by huge effort to establish a whole new protein purification protocol. Another option makes use of the IMAC principle. His-tagged proteins can be immobilized on microscopic slides functionalized with Ni^{2+} or Cu^{2+} chelated groups.[148] (ii) Use of an alternative passivation agent. The microscopic glass slides need to be thoroughly cleaned and passivated in order to avoid contamination, that will lead to background fluorescence in smFRET measurements. In this thesis BSA was used as passivation, which may cause unspecific binding of proteins. Another passivation agent commonly used is polyethylenglycol.[149] It has to be tested if this passivation is better qualified for our experiments. (iii) To further adjust surface density of our proteins of interest, labeled protein preparations were diluted with unlabeled wild-type preparations. The ratio of labeled/unlabeled protein lies somewhere between 1:2 and 1:5. Further dilutions should be tested to find the optimal surface density for smFRET measurements. (iv) A K_D of the immobilization complex should be determined, to estimate how strong our proteins of interest are bound to the surface and if it is possible that they can diffuse from the surface. A dilution series of protein concentrations was examined for its binding on the surface. No reasonable data fit over the Hill equation was possible to determine the K_D. For smFRET measurements a K_D in the nanomolar range is required.[150]

Because enzymatically active enzymes should be analyzed with smFRET, the dimerization of mFAS polypeptide chains is a crucial factor as well.[70] Mikkelsen *et al.* reported the immobilization of enzymatically active FAS enzymes using specific antibodies against the TE domain.[151] In our study a similar approach was utilized, but it was not fully investigated if the proteins immobilized on the surface and analyzed in smFRET measurements exhibited enzymatic activity. The surface immobilization protocol or the measurement buffer might disturb the dimeric fold. It has to be tested whether the mFAS proteins are stable and active in the measurement buffer. In our measurements a mix of an enzymatic oxygen-scavenging system (including glucose oxidase, catalase and β-D-glucose)[152] and the triplet-state-quencher Trolox[153] was used. Molecular oxygen on the one hand is an efficient triplet-state-quencher, whereas reactive oxygen species on the other hand ultimately cause photobleaching.[154] That implies that removal of molecular oxygen as source of highly reactive oxygen species reduces on the one hand photobleaching, but on the other hand the residence time in the dye's unfa-

vorable triplet dark state is prolonged.[153] Therefore the combination of oxygen-scavenging system and triplet-state-quencher is inevitable. A downside of the popular glucose oxidase–catalase system are the chromatic co-factors, required to scavenge oxygen, which absorb wavelengths in the visible spectrum. An improved system uses the combination of protocatechuic acid and protocatechuate-3,4-dioxygenase for enzymatic oxygen-scavenging, which follows contrary to the glucose oxidase–catalase system a one-step mechanism.[155] It has to be tested whether these components are compatible with our proteins of interest. Enzymatic activity of the immobilized mFAS proteins has to be evaluated by established methods described in section 5.1.25. If it actually turns out that the surface immobilization protocol leads to instable mFAS dimers, another surface immobilization strategy has to be established. One option are surface-tethered phospholipid vesicles, which encapsulate the protein of interest in native-like surrounding.[156–158] As long as smFRET is not fully established, also ensemble FRET measurements can be performed, which is simpler to implement and will give valuable information.

Site-Directed Spin Labeling for EPR spectroscopy

Analog to fluorophore labeling, the ncAA-bearing proteins generated in this thesis were used for site-directed spin labeling. Whereas the low click kinetics of the oxime formation make site-directed spin labeling of mFAS difficult, the reaction of AzPhe with DBCO-bearing nitroxide seemed to work fine. In our initial experiments, we achieved 30-60 % spin-labeled protein, which gave an EPR signal in the CW-EPR experiment. The DOL of the spin labeling experiment achieved only half the DOL obtained in fluorophore bioconjugation with the DBCO-functional group. This may be explained by low solubility of the spin label in the protein buffer. Removal of excess free spin label works fine using Amicon centrifugal filters, without loss of protein. The spin label is less hydrophobic than the fluorophores, which makes it less sticky and easier to handle. At the moment some problems with spin-label solubility are encountered. Different solvents have to be tested. Maybe a two-phase system can be envisioned for site-directed spin labeling of proteins. We are confident, that the DOL can be further increased.

Photocrosslinking Assay

The photocrosslinking assay shows promising results. We chose AzPhe over BzPhe, because its chemical synthesis is well established in our lab and AzPhe showed high incorporation efficiency with better fidelity in the reporter assay. We screened several incorporation sites for their crosslinking efficiency. While mutations were still placed on the surface of the protein in order to not disturb the protein fold, for trapping protein-protein interactions, the mu-

tations were intentionally placed at interaction sites. The photocrosslinking reaction seems to be very sensitive and specific, because only mutations at the ACP binding interface gave crosslinks (Gly2113, Ile2128, Leu2129, Gly2130, Ser2150 and Leu2151), whereas mutations far from this interface gave no crosslinks (Gln2182, Glu2183 and Ala2191). Position Ser2150 and Gly2130 were the most promising incorporation sites. Identification of crosslink species in MS/MS gave moderate but reasonable results. Despite good overall sequence coverage of about 81%, the sequence coverage at the mutation site in the ACP was very low and only a few crosslinked peptides were detected. The size of crosslinked peptide fragments obtained through tryptic digestion might have been too big, so that ionization was not sufficient enough and these fragments were not transferred into the gas phase or not successfully accelerated in the electric field, and therefore were not detected. Other endoproteinases like GluC and AspN were tested in combination with different mutation sites, to obtain shorter fragments and a good sequence coverage of the peptide region of interest.

One major benefit of the Gly2130 position is the possibility to phosphopantetheinylate the ACP domain and load different cargos. Both the intramolecular (KS-MAT + ACP-GFP) as well as the intermolecular system (KS-MAT-ACP) gave positive crosslinks and have their advantages and disadvantages. In the case of intramolecular crosslinking, the analysis of crosslinking efficiency by densitometric analysis is easier, because it shows less background. Moreover the natural occurance of ACP:catalytic domain stoichiometry is resembled. On the other hand, the intermolecular system is more interchangeable and needs less cloning and protein expression. However, the most critical factor of the photocrosslinking assay is the dimerization rate of the proteins, since only dimeric proteins are functional. In our initial studies not all proteins fulfilled this quality standard as seen in the HPLC-SEC spectra. This led to low activity in the TAL assay and possibly influenced crosslinking. We tried to estimate this bias, but do not claim reliability of these results. Clearly, the experiments need to be performed with better protein quality.

Biomimicry of Post-Translational Modifications

Post-translational modifications like acetylation of lysines or phosphorylation of serines likely influence FAS activity and are interesting to analyze. Several post-translational modifications have been identified in FASs. First steps have been made and will be continued to add such modifications to our amber codon suppression toolbox. Especially the AT assay is a powerful tool to analyze the effect of such modifications and would enrich our expert understanding of these protein systems.

4.1 Conclusion

The results compiled in this thesis demonstrate the successful incorporation of ncAAs into a 540 kDa homodimer by amber codon suppression with subsequent fluorescent labeling. A systematic approach to tackle the challenges in application of amber codon suppression in an excised small domain instead of the whole multidomain protein was devised, which facilitated the rapid identification of mutation sites. The reporter assay with subsequent upscaling of the culture volume and extending the ncAA modification to the full-length protein allows for possible biophysical methods on such a megaenzyme for the first time. This approach led to a minimal resource usage of both chemical and biological samples, and was analogously applied to other expressible subconstructs of mFAS, as e. g. the KS-MAT di-domain or the minimal functional KS-MAT-ACP construct active in TAL production, which were used in different applications. Together, this procedure may be applied to any comparable biological system – as shown in a proof of principle for pikromycin synthase and DEBS PKS modules – and has the potential to become a powerful tool to elucidate structure and conformational properties of multidomain proteins, as e. g. the homologous PKS family.

Furthermore, the suitability of the generated probes was tested in exciting applications: an FP assay, smFRET microscopy, EPR spectroscopy, and a photocrosslinking assay; all of which have the power to answer fundamental questions in ACP-mediated substrate shuttling that were postulated in the aim of this thesis.
With the spectroscopic methods (FP, FRET and EPR), ensembles of conformational dynamics of these megasynthases can be analyzed. The influence of chain-length on the ACP-mediated substrate shuttling, and protein-protein interactions of ACP and other catalytic domains can be investigated, comparing proteins loaded with different substrates and intermediates of the catalytic cycle, and monitoring the dynamic information obtained through these spectroscopic methods. smFRET furthermore allows for time-resolved monitoring of such dynamics and interactions of single enzymes during their catalytic cycle.
The photocrosslinking assay can be used to map the transient protein-protein interactions between ACP and other catalytic domains. As read-out for successful interactions, crosslink bands between the two catalytic domains are analyzed on SDS-PAGE. In combination with alanine scanning mutagenesis on the surface of the KS domain, these interactions can be fine-tuned. This allows for chain-length modulation of FASs, as has been reported for *C. ammoniagenes* type I FAS.[96] Such chain-length modulation enables the fatty acid production of various chain-lengths, which are valuable platform chemicals, e. g. for biofuel production.

Under these aspects, this thesis resembles an important foundation and powerful method for future studies to investigate conformational dynamics and protein-protein interactions in multidomain proteins, and to engineer them towards altered product spectra.

5 Experimental Procedures

Experimental Procedures have in parts been previously published in: Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878 (for individual contribution to the writing process as well as copyright and Creative Commons license, please consult the Statement of Personal Contributions).

5.1 Material and Methods

5.1.1 General Cloning Procedure

The genes for mFAS proteins (including sub-constructs) and ACP-GFP fusion constructs were cloned into a pET-22b vector (Novagen), which contains a pBR322 origin, ampicillin resistance and *lacI* gene. They are encoded under a T7 or a tac promoter and a T7 terminator, and feature an N-terminal Strep-tag and a C-terminal His8-tag or FLAG-tag. Vectors were linearized either through restriction enzymes or PCR. Genes of interest were amplified from plasmid templates by PCR using Phusion® High-Fidelity DNA Polymerase (NEB) with specific primers containing complementary overhangs, and inserted into the linearized vector through ligation independent cloning with the In-Fusion® HD Cloning Kit (Takara Bio). Multiple inserts were assembled in a MegaPrimer PCR, so In-Fusion® Cloning was conducted with three fragments at most. Templates for large PCR products >4,000 bp were digested with DpnI (NEB). Point mutations (including amber mutations) were generated with the use of a universal reverse primer, annealing with its 5'-end next to the position of interest, and an overhang forward primer, including the mutation. All generated cloning products were transformed into Stellar™competent cells (Clontech) and grown on LB-agar (Lennox; Carl Roth) selection plates, supplemented with 100 µg/mL ampicillin and 1 % glucose, at 37 °C overnight or at room temperature over weekend. Selection plates were stored at 4 °C up to several weeks. Single colonies were picked to inoculate 2 mL LB-media (Lennox; Carl Roth), supplemented with 100 µg/mL ampicillin and 1 % glucose, and cultivated at 37 °C overnight. Plasmids were isolated from the overnight-culture using either the PureYield™Plasmid Miniprep System (Promega) or the GeneJET™Plasmid Miniprep Kit (Thermo Fisher Scientific). The sequences of all plasmids were confirmed by Sanger sequencing at GATC Services (Eurofins

Genomics) or Microsynth Seqlab.
The ACP-GFP construct termed "wild type" in this thesis contained a Met72Leu mutation, to prevent an alternative translation start and to reduce background GFP-fluorescence. Amber mutations were introduced site-specifically in the wild-type genes and their positions were varied throughout the protein sequence to generate various mutant constructs, with diverse incorporation sites. In the following, the cloning of the pETDuet and pAC vectors is described in detail.

5.1.2 Cloning of pETDuet Vectors for Generation of *in vivo* Heterodimers

The pETDuet vectors were generated by inserting a second gene for mFAS proteins (including sub-constructs) into the pET-22b vector (Novagen) already encoding an mFAS protein (including sub-constructs) at the restriction site for XbaI (right between the T7 promoter and the gene of interest). The pET-22b vector was linearized either through the action of the restriction enzyme XbaI (NEB) or through PCR using the primers CH148 & CH149. The second gene of interest was amplified from plasmid templates by PCR with the primers AR81 & CH130. The amplified insert of the second gene of interest was inserted into the linearized pET-22b vector through ligation independent cloning using In-Fusion® HD Cloning (Takara Bio).

5.1.3 Cloning of Suppressor Plasmids pAC^{U} & pAC^{E}

For a list of primers used for cloning of pAC vectors and information about all pAC plasmids constructed and utilized in this thesis, see Supplementary Tables in the Appendix.

5.1.3.1 Cloning of pAC^{U} Vectors

The pAC^{U} vector was assembled from cassettes amplified from different commercially available vectors (see Fig. 3.1). The backbone was derived from the pCDF-1b vector (Novagen), and was cloned with primers CH02 & CH03. The tac promoter and rrnB terminator region were derived from the pMAL-c5G vector (NEB, discontinued), and were cloned as inserts with primers CH04 (with 5'-overhang to the pCDF-1b backbone) & CH05 and CH32 & CH19. The pAC^{U} plasmids encode one copy of aaRS and suppressor tRNA under the tac promoter and rrnB terminator, and the proK promoter and proK terminator, respectively. The backbone of the pAC^{U} contains a CDF origin, spectinomycin resistance and *lacI* gene.

Methanococcus jannaschii Derived TyrRS-tRNA^{Tyr} Pair

The genes encoding *Mj*TyrRS and tRNA^{Tyr} were cloned as inserts from pEVOL-pBpF[112] with primers CH06 (with 5'-overhang to the tac promoter region from pMAL-c5G) & CH33

(with 5'-overhang to the rrnB terminator region from pMAL-c5G), and CH34 (with 5'-overhang to the rrnB terminator region from pMAL-c5G) & CH11 (with 5'-overhang to the pCDF-1b backbone insert). The pEVOL-pBpF plasmid was a gift from Peter G. Schultz (Addgene plasmid # 31190) and contains genes encoding an optimized $tRNA^{Tyr}$ and an evolved *Mj*TyrRS for the incorporation of BzPhe.[112] Phusion® High-Fidelity DNA Polymerase (NEB) was used to generate PCR fragments, which were assembled with help of complementary primer overhang in a MegaPrimer PCR and subsequently cloned into the backbone using In-Fusion® HD Cloning (Takara Bio). The resulting pAC^{U}_BzPhe plasmid was used as template for multiple point mutations in the gene of *M. jannaschii* derived TyrRS to generate a variety of pAC^{U} plasmids encoding evolved *Mj*TyrRSs, specific for incorporation of various substituted phenylalanine ncAAs.

Methanosarcina mazei/ barkeri Derived PylRS-$tRNA^{Pyl}$ Pair

In order to incorporate lysine-derivative ncAAs, a *M. mazei* or *M. barkeri* derived PylRS and corresponding $tRNA^{Pyl}$ is used. Therefore, the genes encoding *Mj*TyrRS and $tRNA^{Tyr}$ in the pAC^{U}_BzPhe plasmid were exchanged by genes encoding *Mm*PylRS/ *Mb*PylRS and $tRNA^{Pyl}$. The insert between the genes of PylRS and $tRNA^{Pyl}$ was cloned from pAC^{U} with primers CH32 & AR366 and contained the rrnB terminator and proK promoter sequences. The pAC^{U} backbone was cloned with primers AR369 & CH05. The inserts encoding *Mm*PylRS and $tRNA^{Pyl}$ were derived from the plasmid pJZ, which was a gift from Nediljko Budisa and contains genes encoding a wild-type $tRNA^{Pyl}$ and an N-terminally Strep-tagged *Mm*PylRS evolved for the incorporation of BCNLys.**cite** The *Mm*PylRS insert was cloned without Strep-tag with primers CH53 (with 5'-overhang to the tac promoter region) & AR365 (with 5'-overhang to the rrnB terminator region) and the $tRNA^{Pyl}$ insert with primers AR367 (with 5'-overhang to the rrnB terminator region) & AR368 (with 5'-overhang to the pAC^{U} backbone). Phusion® High-Fidelity DNA Polymerase (NEB) was used to generate the PCR fragments, that were assembled with the help of complementary primer overhangs in MegaPrimer PCRs and subsequently cloned into the backbone using In-Fusion® HD Cloning (Takara Bio). The resulting pAC^{U}_BCN plasmid was used as template for multiple point mutations in the gene of *M. mazei* derived PylRS to generate a variety of pAC^{U} plasmids encoding evolved *Mm*PylRS, specific for incorporation of various lysine-derivative ncAAs. The wild-type $tRNA^{Pyl}$ contains an unfavourable U:G wobble pair in the anticodon stem, which was mutated to a C:G pair to improve suppression efficiency. This mutagenesis was performed with primers CH57 & CH56. The *Mb*PylRS insert was derived from the pAcBac1.tR4-MbPyl plasmid, which was a gift from Peter G. Schultz (Addgene plasmid # 50832) and contains genes encoding an undescribed *Mb*PylRS and two copies of synthetic $tRNA^{Pyl}$ derived from *Desulfitobacterium hafniense*.[113] The *Mb*PylRS insert was cloned with primers CH51 (with 5'-overhang to the tac promoter region) & CH52 (with 5'-overhang to the rrnB terminator region). Another *Mb*PylRS insert was derived from the pDULE-ABK plasmid, which was a gift

from Peter G. Schultz (Addgene plasmid # 49086) and contains a gene encoding an evolved *Mb*PylRS for the incorporation of an aliphatic diazirine ncAA.[114] The evolved *Mb*PylRS insert was cloned with primers CH51 (with 5'-overhang to the tac promoter region) & CH158 (with 5'-overhang to the rrnB terminator region), and subsequently multiple point mutations were introduced in the gene of the evolved *Mb*PylRS to generate the wild-type *Mb*PylRS. Same improved $tRNA^{Pyl}$ was used as for the *M. mazei* derived PylRS-$tRNA^{Pyl}$ pair. Phusion® High-Fidelity DNA Polymerase was used to generate the PCR fragments, that were assembled with the help of complementary primer overhangs in MegaPrimer PCRs and subsequently cloned into the backbone using In-Fusion® Cloning resulting in the pAC^{U}_PyLys plasmid. The *M. barkeri* derived PylRS was only used in its wild-type form.

5.1.3.2 Cloning of pAC^{E} Vectors

Genes encoding different mutants of aaRSs in the pAC^{U} vectors were used as templates to create pAC^{E} plasmids (see Fig. 3.1). The pAC^{E} backbone was derived from the pEVOL-pBpF[112] plasmid and was amplified with primers CH123 & CH124. The insert between the two copies of the gene encoding aaRS (pEVOL insert) was also derived from the pEVOL-pBpF[112] plasmid and was cloned with primers CH100 & CH101. The pAC^{E} plasmids encode two copies of aaRS, one under the arabinose (ARA) promoter and the rrnB terminator, and one under the glnS' promoter and glnS terminator, as well as one copy of suppressor tRNA under the proK promoter and proK terminator. The backbone of the pAC^{E} contains a p15A origin, chloramphenicol resistance and *araC* gene.

Methanococcus jannaschii Derived TyrRS-$tRNA^{Tyr}$ Pair

The two copies of the gene encoding *Mj*TyrRS were derived from the corresponding pAC^{U} plasmids and were amplified with primers CH102 (with 5'-overhang to the ARA promoter region) & CH103 (with 5'-overhang to the rrnB terminator region) and CH104 (with 5'-overhang to the glnS' promoter region) & CH105 (with 5'-overhang to the glnS terminator region), and combined with the pEVOL insert in a MegaPrimer PCR using primers CH116 & CH118 (with 5'-overhangs to the pAC^{E} backbone) and Phusion® High-Fidelity DNA Polymerase (NEB), and subsequently cloned into the backbone using In-Fusion® Cloning (Takara Bio).

Methanosarcina mazei/ barkeri Derived PylRS-$tRNA^{Pyl}$ Pair

For the *M. mazei* and *M. barkeri* pAC^{E} plasmids, the gene encoding $tRNA^{Tyr}$ had to be replaced with the respective gene for $tRNA^{Pyl}$. Therefore, primers AR369 & AR366 were used to amplify the pAC^{E} backbone and primers AR367 & AR368 (with 5'-overhangs to the pAC^{E} backbone) were used for the $tRNA^{Pyl}$ insert, respectively. Again, the two copies

of the gene encoding PylRS were derived from the corresponding pAC^{U} plasmids. For the *Mm*PylRS gene primers CH96 (with 5'-overhang to the ARA promoter region) & CH97 (with 5'-overhang to the rrnB terminator region), and CH98 (with 5'-overhang to the glnS' promoter region) & CH99 (with 5'-overhang to the glnS terminator region) were used to amplify a fragment, which was combined with the pEVOL insert in a MegaPrimer PCR using primers CH115 & CH117 (with 5'-overhangs to the pAC^{E} backbone) and Phusion® High-Fidelity DNA Polymerase (NEB), and subsequently cloned into the backbone using In-Fusion® Cloning (Takara Bio). To amplify the gene encoding the *Mb*PylRS primers CH106 (with 5'-overhang to the ARA promoter region) & CH107 (with 5'-overhang to the rrnB terminator region) and CH108 (with 5'-overhang to the glnS' promoter region) & CH109 (with 5'-overhang to the glnS terminator region) were used, and the respective PCR product was combined with the pEVOL insert in a MegaPrimer PCR using primers CH115 & CH119 (with 5'-overhangs to the pAC^{E} backbone), and subsequently cloned into the backbone using In-Fusion® Cloning (Takara Bio).

5.1.4 General Protein Expression Procedure

All constructs were transformed in *E. coli* BL21-Gold (DE3) competent cells (Agilent Technologies) following the provided heat shock protocol. For incorporation of ncAAs, plasmids encoding ACP-GFP and mFAS constructs (including sub-constructs) with amber mutations (for a list of used plasmids, see Table S1) were co-transformed with the appropriate suppressor plasmid pAC^{U} or pAC^{E}. For *in vivo* phosphopantetheinylation, co-transformation with a pETcoco vector (Novagen) encoding the 4'-phosphopantetheinyl transferase Sfp from *B. subtilis* was performed. LB-agar (Lennox; Carl Roth) selection plates contained 1 % glucose to suppress leaky expression, and were supplemented with either 100 µg/mL ampicillin for transformation of the wild-type protein, 50 µg/mL ampicillin and 25 µg/mL spectinomycin for co-transformation with pAC^{U} plasmids, 50 µg/mL ampicillin and 17 µg/mL chloramphenicol for co-transformation with pAC^{E} plasmids, or with the pETcoco plasmid. Colonies were grown at 37 °C overnight or at room temperature over weekend and stored at 4 °C up to several weeks. A randomly picked single clone was used to inoculate a pre-culture of LB-media (Lennox; Carl Roth), supplemented with 1 % glucose and respective antibiotics, which was grown at 37 °C and 140-180 rpm overnight. A sufficient amount of the densely-grown pre-culture was transferred into TB-medium to yield a 1:50 cell dilution. Expression cultures were incubated at 37 °C and 140 rpm until an OD_{600} of about 0.6-0.8 was reached. After cooling down the expression cultures for 15 min at 4 °C or at the bench at room temperature, expression was induced with 0.25 mM IPTG. In case of ncAA incorporation, 1 to 2 mM ncAA was added at the time of induction. Expression took place at 20 °C and 140 rpm overnight.

5.1.5 Reporter Assay

The protein expressions for the reporter assay were performed in 2 mL scale in 96-well deep well plates (VWR) in technical triplicates. An ACP-GFP wild-type construct without amber mutation was used as reference and negative samples of each construct without addition of ncAA were assayed as well. The cells were harvested by centrifugation (3,220 rcf for 5 min at 4 °C), washed and resuspended in 300 µL PBS.

5.1.6 Test Expression Incorporating TCO*Lys

The test expression incorporating the ncAA TCO*Lys (SiChem) in response to the amber codon in our protein of interest was carried out at the Unnatural Protein Facility of the Oregon State University following the protocols of the Mehl lab. In the laboratory course, the ACP-GFP fusion construct encoded on a pET-22b vector (Novagen) and a GFP protein encoded on a pET-28a vector (Novagen) were used as wild-type proteins or with an amber mutation placed at site Leu54 and Val150 of the proteins, respectively. Wild-type proteins served as positive control. All constructs were transformed in electrocompetent BL21-AI *E. coli* cells (Thermo Fisher Scientific) following the provided electrotransformation protocol (voltage: 1,500 V; discharge time: 5 ms). For incorporation of ncAAs, the pET-22b and pET-28a plasmids, encoding ACP-GFP and GFP with amber mutation, were co-transformed with the pEVOL plasmid containing a *Mm*PylRS mutant specific for incorporation of TCOLys (provided by the Mehl lab). LB-agar (Lennox; Carl Roth) selection plates were supplemented with respective antibiotics (100 µg/mL ampicillin for pET-22b, 50 µg/mL kanamycin for pET-28a, and additional 25 µg/mL chloramphenicol for co-transformation with pEVOL). Colonies were grown at 37 °C overnight. A randomly picked single clone was used to inoculate a pre-culture of non-inducing media (prepared following the protocol by Hammill *et al.*)[159] supplemented with respective antibiotics. The pre-culture was grown at 37 °C and 140-180 rpm for 8 to 10 h. 50 µL of the pre-culture was transferred into 50 mL auto-induction media (prepared following the protocol by Hammill *et al.*)[159] and incubated at 37 °C and 250 rpm. After 30 min, 1 mM TCO*Lys, dissolved in 0.2 M NaOH, was added to the expression culture. The expression culture was incubated for up to 48 h at 37 °C and 250 rpm. OD_{600} and fluorescence intensity of the expression culture were monitored over time. Samples for SDS-PAGE were taken at different time points.

5.1.7 Expression of ACP-GFP Constructs

Large scale expression of ACP-GFP constructs was carried out in 200 mL expression cultures. Prior to harvesting the cells by centrifugation (4,000 rcf for 20 min at 4 °C), 2 mL samples of cell cultures were taken for further quantification using GFP-fluorescence. All cell pellets were flash frozen in liquid nitrogen and stored at -80 °C until use.

5.1.8 Expression of mFAS Constructs

Large scale expression of mFAS constructs (including sub-constructs) was carried out in 1 L expression cultures following the protocol by Rittner *et. al.*[75] The cells were harvested by centrifugation (4,000 rcf for 20 min at 4 °C) and subsequently purified.

5.1.9 Purification of ACP-GFP Constructs

The cell pellets were thawed on ice and resuspended in 10 mL His buffer (50 mM KPi, 200 mM NaCl, 20 mM imidazole, 10 % glycerol, pH 7.4) containing DNase I (Sigma-Aldrich) and 1 mM EDTA. French pressure cell press (Aminco) was used for mechanical disruption at a pressure of 1,000 bar and the cell debris was removed by centrifugation (50,000 rcf for 30 min at 4 °C). After addition of 2 mM $MgCl_2$, the lysate was subjected to 3 mL (bead capacity 50 mg/mL) Ni-NTA Superflow resin (QIAGEN) and incubated for 1 h at 4 °C. Unbound protein was washed off with 5 CVs His buffer and bound His-tagged protein was eluted with 2.5 CVs His buffer containing 300 mM imidazole and additional 2 CVs of His buffer with imidazole increased to 500 mM. The elution fractions were analyzed by SDS-PAGE and SEC over a Superdex 200 Increase 10/300 GL or HiLoad Superdex 200 16/60 pg (both GE Healthcare) column (His buffer filtered and degassed). SEC columns were calibrated with the Gel Filtration LMW and HMW Calibration Kits (GE Healthcare). Protein samples were concentrated using an Amicon® Ultra Centrifugal Filters (Merck Millipore), flash frozen in liquid nitrogen, and stored at -80 °C.

5.1.10 Purification of mFAS Constructs

The cell pellets were resuspended in 30 mL His buffer (50 mM NaPi, 450 mM NaCl, 10 mM imidazole, 20 % glycerol, pH 7.6) containing DNase I (Sigma-Aldrich) and 1 mM EDTA. French pressure cell press (Aminco) was used for mechanical disruption at a pressure of 1,000 bar and the cell debris was removed by centrifugation (50,000 rcf for 30 min at 4 °C). After addition of 2 mM $MgCl_2$, the protein was bound to Ni-NTA Superflow resin (QIAGEN) and eluted at 300 mM imidazole. Additionally to Ni-IMAC the mFAS constructs were purified over a Strep-Tactin® Sepharose® column (Iba), eluted with 2.5 mM D-desthiobiotin (Iba) (Strep buffer: 250 mM KpPi, 1 mM EDTA, 1 mM DTT, 10 % glycerol, pH 7.4). Heterodimers generated *in vivo* were purified over ANTI-FLAG® M2 Affinity Gel (Sigma-Aldrich) instead of Strep-Tactin®. FLAG-tagged protein was eluted with 5 CVs 100 µg/mL FLAG-peptide in Strep buffer. Further purification was performed by SEC over a Superdex 200 Increase 10/300 GL or HiLoad Superdex 200 16/60 pg (both GE Healthcare) column (Strep buffer filtered and degassed). SEC columns were calibrated with the Gel Filtration LMW and HMW Calibration Kits (GE Healthcare). Protein samples were concentrated using an Amicon® Ultra Centrifugal Filters (Merck Millipore), flash frozen in liquid nitrogen, and stored at -80 °C.

5.1.11 Quantification of GFP-Fluorescence

The samples of the reporter assay, test expressions at the Unnatural Protein Facility and the 2 mL samples from large scale ACP-GFP expression were analyzed by their GFP-fluorescence. A 10-fold dilution sample of the resuspended cells in PBS was transferred into black cyclo-olefin-copolymer 96-well UV-Star® μclear® flat-bottom microplates (Greiner Bio-One) and OD_{600} and GFP fluorescence were measured at the CLARIOstar (BMG). Blank-corrected fluorescence values were normalized by OD_{600}. Fluorescence intensity of the wild type was set to 100 % and all other fluorescence signals were related to the wild type.
Fluorescence intensity of 1:20 cell dilutions from expression cultures grown at the Unnatural Protein Facility was measured at a VersaFluor™Fluorometer (Bio-Rad).

5.1.12 Protein Concentration

Protein concentrations were calculated from the absorbance at 280 nm, which was recorded on a NanoDrop 2000c (Thermo Fisher Scientific). Extinction coefficients were calculated from the primary sequence with CLC Main workbench (QIAGEN Bioinformatics). Absorbance 1 g/L at 280 nm (10 mm) is given for all proteins (without *N*-formylmethionine) in Table 5.1.

Table 5.1: Absorbance of 1 g/L proteins at 280 nm (10 mm).

Protein	Species	Abs.
ACP	mouse	0.475
	E. coli	0.131
	DEBS3M5	0.842
	PikAIIIM5	0.857
	RAPS3M14	0.917
ACP-GFP	mouse	0.558
KS-AT	mouse	1.053
	E. coli	1.171
	DEBS3M5	1.021
	PikAIIIM5	1.1126
	RAPS3M14	1.210
KS-MAT-ACP	mouse	0.956
KS-MAT-ACP-TE	mouse	0.846
FAS	mouse	0.893

5.1.13 SDS-PAGE Analysis

Proteins were analyzed by SDS-PAGE, loading a 10 µL sample of about 0.05 mg/mL protein in water, that was treated with Laemmli buffer for 5 min at 95 °C. Electrophoresis was performed at 200 V for 45-60 min in Tris/Glycine/SDS running buffer (Bio-Rad) or NuPAGE™MES SDS running buffer (Invitrogen). Self-cast 7 % (for 50-500 kDa proteins), 10 % (for 20-300 kDa proteins), 12 % (for 10-200 kDa proteins), 15 % (for 3-100 kDa proteins) polyacrylamide gels, NuPAGE™4-12 % Bis-Tris Protein Gels (Invitrogen) or 4-20 % Mini-PROTEAN® TGX™Precast Protein Gels (Bio-Rad) were used. Gels were stained in InstantBlue™Coomassie Protein Stain (Expedeon) or light-protected in a 5,000× dilution of SYPRO® Red (Thermo Fisher Scientific) in 7.5 % acetic acid solution. Protein bands on SYPRO® Red-stained gels and in-gel fluorescence of fluorescently labeled proteins were detected at the Fusion SL Fluorescence Imaging System (VilberLourmat). Illumination/emission filter: blue epi/F-532 Y2; epi green/F-695 Y5; epi near infrared/F-695 Y5; epi infrared/F-820.

5.1.14 Western Blot Analysis

For western blot analysis, samples of mFAS constructs (including sub-constructs) were analyzed with SDS-PAGE. The proteins were transferred from the analytical polyacrylamide gel onto a Immobilon-FL PVDF membrane (Merck Millipore) by an electrophoretic tank-blot method (25 V for 1 h). The membrane was subsequently blocked with 0.2 % I-Block (Tropix) and 0.1 % Tween-20 (Carl Roth) in PBS, treated first with monoclonal mouse anti-Strep antibody (StrepMAB classic, Iba) and monoclonal rabbit anti-His antibody (Bethyl Laboratories) (at 4 °C overnight), and secondly with IgG donkey anti-mouse DyLight 755 and IgG goat anti-rabbit DyLight 633 (both Thermo Fisher Scientific) (light-protected at room temperature for 1 h). Excess antibodies were washed off in several washing steps with 0.2 % I-Block and 0.1 % Tween-20 in PBS. The western blot was developed at the Fusion SL Fluorescence Imaging System (VilberLourmat) with excitation in the near infrared to infrared range and using the emission filters F-695 Y5 and F-820.

5.1.15 Fluorescent Bioconjugation

Bioconjugation of fluorophore (20-150 equiv.) and protein (1 equiv.) was performed in a copper-free environment at room temperature in the dark. The reaction proceeded in 25-100 µL His buffer for up to 4 h. Protein mutants with the AzPhe were treated with the fluorophore BCN-POE_3-NH-DY649P1 (Synaffix, discontinued), and mutants with the NorLys2 were treated with 6-methyl-tetrazine dyes (Jena Biosciences). Wild-type ACP-GFP served as negative control, treated with the same fluorophores under the same conditions. As reference, wild-type ACP was phosphopantetheinylated with fluorescent CoA substrate (NEB) (see section 5.1.22). The phosphopantetheinylation was performed in presence of 10 mM $MgCl_2$ for

30-45 min at 37 °C with mild agitation in the dark. Subsequent to bioconjugation, an analytical SDS-PAGE was performed and fluorescent protein bands were detected at the Fusion SL Fluorescence Imaging System (Vilber Lourmat) with excitation in the near infrared range and using emission filter F-695 Y5. The FusionCapt Advance Solo 4 16.08a software was used to quantify the fluorescence of the protein bands on the polyacrylamide gel.

5.1.16 Site-Directed Spin Labeling

100 µM protein was labeled in a SPAAC with 100 equiv. DBCO-spin label (Frank Kaiser, AK Göbel) in a 100 µL reaction volume for up to 24 h at room temperature and 350 rpm. 50 µM protein was labeled in an oxime formation with 10 equiv. HO-4120 nitroxide spin label (Sigma-Aldrich), catalyzed by 100 mM *para*-anisidine, in a 150 µL reaction volume for up to 24 h at room temperature and 350 rpm. Subsequently, the reaction was applied to an Amicon® Ultra Centrifugal Filter (Merck Millipore) for exchange with Strep-buffer and to wash out unreacted spin label (12,000 rcf, 5 min). Spin-labeled protein was subjected to CW-EPR measurement and spin counting at the group of Prof. Prisner at the Goethe University Frankfurt, performed by Dr. Alberto Collauto.

5.1.17 Purification of Protein Fluorophore Conjugates

Purification Over HisPur Ni-NTA Magnetic Beads

Excess fluorophore from bioconjugation reaction was removed by purification over 1 mg HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific). At each purification step the beads were shortly vortexed, spun down and placed in a magnetic stand, so the liquid phase could be taken up with a pipette. The Ni-NTA beads were first equilibrated with 160 µL and additional 400 µL His buffer (50 mM KPi, 200 mM NaCl, 20 mM imidazole, 10 % glycerol, pH 7.4). The bioconjugation reaction was diluted with one volume of His buffer and incubated with the Ni-NTA beads in the dark, either for 30 min on an end-over-end rotator (for ACP-GFP proteins) or for 15 min on ice (for mFAS and its sub-constructs). Unbound protein was washed off with two times 400 µL His buffer (the flow through can be applied again to Ni-NTA magentic beads to maximize yield). In two elution steps, the bound His-tagged protein was incubated for 30 min, and 15 min, respectively, in the dark on an end-over-end rotator or on ice with 50 µL His buffer containing 300 mM imidazole.

Purification Over Gravity-Flow Columns

Alternatively, proteins were isolated from fluorophore bioconjugation reaction using Ni-IMAC, Strep-affinity chromatography or illustra NAP-5 columns (GE Healthcare), which are packed with Sephadex G-25 DNA Grade for SEC. For Ni-IMAC, the protein was bound to Ni-NTA Superflow resin (QIAGEN) and eluted at 300 mM imidazole (His buffer: 50 mM KPi, 200 mM

NaCl, 20 mM imidazole, 10 % glycerol, pH 7.4). For Strep-affinity chromatography, the protein was bound to Strep-Tactin® resin (Iba) and eluted with 2.5 mM D-desthiobiotin (Iba) (Strep buffer: 250 mM KPi, 1 mM EDTA, 1 mM DTT, 10 % glycerol, pH 7.4). For SEC over illustra NAP-5 columns (GE Healthcare), also Strep buffer was used.

5.1.18 Determining the Degree of Labeling

UV-Vis spectra were recorded on a Carry 100 UV-Vis spectrophotometer (Agilent Technologies) from 800 nm to 220 nm wavelength in quartz glass cuvettes (50 µL sample). Excess fluorophore had been removed by purification over HisPur Ni-NTA magnetic beads (Thermo Fisher Scientific). The reference sample contained His buffer with 300 mM imidazole. Absorption at 650 nm (maximum absorption of the acceptor fluorophore A), at 550 nm (maximum absorption of the donor fluorophore D), at 280 nm (maximum absorption of the protein) and at 485 nm (maximum absorption of GFP) were used to determine the degree of labeling, following equations (5.1)–(5.4). With ϵ_{prot} and ϵ_{GFP} being the molar extinction coefficients of the protein and GFP, respectively, and ϵ_{550} and ϵ_{650} being the molar extinction coefficients of the acceptor (A) and the donor (D) fluorophores, respectively. The correction factors $CF_{280} = A_{485}/A_{\mathrm{max}}$ and $CF_{485} = A_{485}/A_{\mathrm{max}}$, as well as $CF_{550,\mathrm{D}} = A_{550}/A_{\mathrm{max}}$ and $CF_{650,\mathrm{A}} = A_{650}/A_{\mathrm{max}}$ were determined from the absorption spectra of the free fluorophores in water. Table 5.2 summarizes absorption maxima (λ_{max}) and molar extinction coefficients (ϵ_{max}) of the dyes used in this thesis.

$$DOL = (A_{\mathrm{max}}\epsilon_{\mathrm{GFP}})/((A_{\mathrm{GFP}} - A_{\mathrm{max}}CF_{485})\epsilon_{\mathrm{max}}), \tag{5.1}$$

for ACP-GFP proteins.

$$DOL = (A_{\mathrm{max}}\epsilon_{\mathrm{prot}})/((A_{280} - A_{\mathrm{max}}CF_{280})\epsilon_{\mathrm{max}}), \tag{5.2}$$

for proteins without GFP-fusion.

$$DOL_{\mathrm{D}} = ((A_{550} - A_{650}CF_{550,\mathrm{A}})\epsilon_{\mathrm{prot}})/((A_{280} - A_{550}CF_{280,\mathrm{D}} - A_{650}CF_{280,\mathrm{A}})\epsilon_{550}), \tag{5.3}$$

for the acceptor (A) fluorophore of dual-labeled proteins.

$$DOL_{\mathrm{A}} = ((A_{650} - A_{550}CF_{650,\mathrm{D}})\epsilon_{\mathrm{prot}})/((A_{280} - A_{550}CF_{280,\mathrm{D}} - A_{650}CF_{280,\mathrm{A}})\epsilon_{650,\mathrm{A}}), \tag{5.4}$$

for the donor (D) fluorophore of dual-labeled proteins.

Table 5.2: Spectral properties of fluorophores, according to the manufacturers (ATTO-TEC, Dyomics and Jena Bioscience), measured in PBS.

Dye	λ_{max} (nm)	ϵ_{max} ($M^{-1}cm^{-1}$)	Quantum efficiency
Cy3	553	151,000	15 %
Cy5	647	251,000	27 %
ATTO532	532	115,000	90 %
ATTO647N	647	150,000	65 %
DY647P1	647	250,000	30 %
DY649P1	649	250,000	30 %

5.1.19 Mass Spectrometric Protein Analysis

Purified ACP-GFP proteins were analyzed using nanoESI (Synapt G2-S) mass spectrometry. Protein buffer of the sample was changed to 0.1 or 1 M ammonium acetate in Amicon® Ultra Centrifugal Filters (Merck Millipore). Protein concentration of the samples was 1 mg/mL.

5.1.20 HPLC-SEC Analysis of mFAS Constructs

HPLC analysis of mFAS and sub-constructs was performed using an UltiMate 3000 RSLC (Dionex) equipped with an RS fluorescence detector. Chromatographic separation was performed on a SEC column (bioZen 1.8 µm SEC-3) in the buffer (150 mM KPi, pH 6.8). Proteins were detected by monitoring the absorbance at 280 nm. All buffers and samples were filtered through 0.1 µm. A protein sample of 6 µL was injected into the HPLC-SEC and analyzed at 4 °C with a flow rate of 0.3 mL/min and a pressure of 410-420 bar.

HPLC-SEC Analysis of Fluorescently Labeled mFAS

Fluorescently labeled proteins were detected by monitoring the absorbance at 280 nm, 554 nm and 674 nm or the fluorescence at Em./Ex.: 550/580 nm; 630/650 nm and 550/650 nm. Dimeric mFAS was found to elute at 7.5-7.6 min when monitored by fluorescence. Emission spectra of the donor (Ex. 550 nm/Em. 570 nm: 550/570) and acceptor (Ex. 630 nm/Em. 650 nm: 630/650) channels were normalized to the sample with the highest intensity in the respective channel. Additionally, the signal was corrected for differences in the amount of protein by using the absorption at 280 nm. Analyzing the channel (Ex. 550 nm/Em. 650 nm: 550/650) was more difficult, as the Dy547 fluorophore of CoA-547 shows relatively strong emission in the respective channel, falsifying the qualitative FRET signal. Hence, we calculated a correction factor of 0.013 by dividing the fluorescence signal of the CoA 547 labeled mFAS sample at (550/650) by the signal at (550/570). We multiplied both fluorescence signals of AzPhe containing samples at (550/570) with this correction factor to obtain the potential crosstalk

of Dy547 in the channel (550/650). All signals were multiplied with 0.32 (Gly2113AzPhe) and with 0.59 (Gly2113AzPhe and Ser2150Ala) and than normalized to the highest peak of sample Gly2113AzPhe. Crosstalk of the Dy649P1 fluorophore in channel (550/650) was not considered.

5.1.21 Thermal Shift Assay

Protein stability in different buffer conditions was determined with the thermal shift assay at the CFX Connect Real-Time System (Bio-Rad) monitoring the fluorescence intensity of SYPRO® Orange (Thermo Fisher Scientific) at excitation 470 nm and emission 569 nm upon temperature changes increasing from 5 to 94.5 °C in steps of 0.5 °C. All measurements were performed in technical triplicates. 1 µM protein was assayed in white Multiplate™96-well PCR plates (Bio-Rad) in a total sample volume of 25 µL. 2 µL of a SYPRO® Orange 62.5× stock was added to the protein.

5.1.22 Phosphopantetheinylation of apo ACP

Holo ACP was generated either through co-expression with 4'-phosphopantetheinyl transferase Sfp from *B. subtilis* encoded on a pETcoco vector (Novagen) *in vivo* or in an *in vitro* reaction with isolated Sfp. 1 equiv. apo ACP was phosphopantetheinylated with 5 equiv. of CoA substrate (Sigma-Aldrich), catalyzed by 0.5 equiv. Sfp. The phosphopantetheinylation was performed in presence of 10 mM $MgCl_2$ for 30-45 min at 37 °C with mild agitation.

5.1.23 TAL Activity Assay

TAL production was monitored photometrically with the NanoDrop 2000c (Thermo Fisher Scientific) in a sample volume of 120 µL in UV-transparent cuvettes (Sarstedt) at 37 °C by observing the increase in absorbance at 298 nm. All TAL activity assays were performed in technical triplicates. 3× stock solutions of the proteins as well as the substrates acetyl- and malonyl-CoA (both Sigma-Aldrich) were prepared freshly in the assay buffer (50 mM KPi, 10 % glycerol, 0.03 mg/mL BSA, pH 7.4) and stored at 4 °C. Shortly before the measurements, aliquots of the 3× stock solutions were heated to 37 °C. The reaction was started by addition of acetyl-CoA. Final concentrations of the reaction components were 100 µM CoA substrates and 700 nM protein. The absorbance at 298 nm was monitored over 5 min. Absorbance units were converted into concentrations using the molar extinction coefficient ϵ=2540 $M^{-1}cm^{-1}$ for TAL.[160]

5.1.24 AT Activity Assay

AT activity was determined using an enzyme-coupled assay in black polystyrene 96-well flat-bottom microplates (Greiner Bio-One) monitoring NADH fluorescence at the CLARIOstar

(BMG).[75] Settings: excitation: 348±20 nm; emission: 476±20 nm; gain: 1900; focal height: 5.2 mm; number of flashes: 70; orbital averaging: 4 mm. Four different stock solutions were prepared in filtered and degassed assay buffer (50 mM NaPi, 10 % glycerol, 1 mM DTT, 1 mM EDTA, pH 7.6): solution 1 (3× stock): 3 nM protein, 0.1 mg/mL BSA; solution 2 (4× stock): 8 mM α-ketoglutaric acid, 1.6 mM NAD^+, 1.6 mM TPP, 60 mU/100 µL α-ketoglutarate dehydrogenase; solution 3 (4× stock): 79 µM malonyl-CoA; solution 4 (5× stock): 300µM holo ACP. The stock solutions were pipetted and mixed in their respective order (30µL solution 1, 25µL solution 2, 25µL solution 3). The reaction was started by addition of 20µL of solution 4 through the dispenser, installed at the CLARIOstar (BMG labtech). Equidistant kinetic measurements were taken every 20 s for 5 min. For data analysis, the average of technical triplicates was calculated and the background of a sample without protein was subtracted. Relative fluorescence units were converted into µM concentrations using an experimentally derived NADH calibration of 14,948 RFU/µM.

5.1.25 Overall Fatty Acid Synthase Activity

FAS activity was measured by following the oxidation of NADPH at 25 °C in 75 mM KPi (pH 7.0), 0.5 mM EDTA, 1 mM ascorbic acid, 130 µM acetyl-CoA, 130 µM malonyl-CoA and 80 µM NADPH. The absorbance at 334 nm was monitored with the NanoDrop 2000c (Thermo Fisher Scientific) in UV-transparent cuvettes (Sarstedt), using an extinction coefficient for NADPH of 6,220 $M^{-1}cm^{-1}$. The enzyme was prepared in a 4× stock containing 20 % (v/v) PEG 400 for stabilization, resulting in a final protein concentration of 20 nM and 5 % (v/v) PEG 400 in the assay. The reaction was initiated by addition of the enzyme to the reaction mixture. Every measurement was performed in technical triplicates and the corresponding background of a sample without enzyme was subtracted.

5.1.26 Fluorescence Polarization Assay

Fluorescence polarization was measured in black polystyrene 384-well flat-bottom microplates (Greiner Bio-One) at the CLARIOstar (BMG) at 30 °C. Settings: excitation: 590±50 nm; dichroic filter: 639 nm; emission: 675±50 nm; gain: 641 channel A and 657 channel B; settling time: 0.5 s; number of flashes per well: 500; focal height: 12.5 mm. Protein samples were prepared in filtered buffer containing 100 mM NaCl and 5 % glycerol in a final volume of 20 µL. Gain and focus adjustment were performed on a 10 µM sample of ATTO647N dye (NEB). A dilution series of ATTO647N dye in a range of 0.08 to 10 µM, and a dilution series of fluorescently labeled ACP in a range of 0.07 to 6 µM was measured. To 1 µM fluorescently labeled ACP, increasing concentrations of murine KS-MAT from 0 to 225 µM were titrated and measured in fluorescence polarization mode. Addition of BSA to the samples kept the total protein concentration constant. Every sample was measured three times in a row.

5.1.27 Immobilization on Microscopic Slides

Microscopic slides and coverslips (VWR) were thoroughly cleaned before the immobilization protocol by the following procedure: 25 min sonicated in isopropanol, three times rinsed with MilliQ water and dried under a stream of nitrogen, treated with the Zepto plasma cleaner (Diener electronic) for 15 min under nitrogen. Sample chambers of about 50 µL volume were provided by sandwiching 0.5 cm wide stripes of parafilm between a microscopic slide and a coverslip and by sealing it at 95 °C. The individual layers of the sandwich-like immobilization complex were applied to the microscopic slide by processing the respective component in solution through the sample chambers. 50 µL of each solution was filled into the sample chambers, incubated for the respective time span, and afterwards removed by suction applied through a tissue, which pulled out the liquid. Between each layer application, the chambers were washed three times with 100 µL PBS. Layers were applied in the following order with the following concentrations and incubation times: BSA-biotin in 5 % BSA solution (Sigma-Aldrich) (0.15 mg/mL, 1 h incubation), streptavidin (Sigma-Aldrich) (0.05 mg/mL, 20 min incubation), biotinylated protein L (Thermo Fisher Scientific) (17 µg/mL, 20 min incubation), StrepMAB-Immo antibody (Iba) (10 µg/mL, 5 min incubation), Strep-tagged protein (1-17 µg/mL, 5 min incubation).

5.1.28 dSTORM Measurements

Individual layers of surface modifications were characterized by dSTORM measurements in the group of Prof. Heilemann at the Goethe University Frankfurt. Proteins were fluorescently labeled using Cy5-NHS ester (Lumiprobe), which reacts with primary amino groups found at the N-terminus or in lysine side chains of polypeptides. The reaction took place in 0.1 M $NaHCO_3$ buffer pH 8 with 8 equiv. Cy5-NHS ester at 4 °C overnight. dSTORM measurements were recorded at the NSTORM microscope (Nikon) in dSTORM buffer (1 mM Tris, 2.5 mM KCl, 20 % glucose, 50 U/mL glucose oxidase, 100 mM β-mercaptoethylamin, 0.2 mM TCEP, 2.5 % glycerol, pH 7.5). Settings: 1 % (0.012 mW) laser intensity at 405 nm; 60 % (0.64 mW) laser intensity at 647 nm; gain 300; exposure time 25 ms; frames 5,000; filter ET band pass 705/72; dichroic mirror 660 nm. Data evaluation was performed with the rapidSTORM 3.2 [161] and the LAMA program. [162,163]

5.1.29 smFRET Measurements

smFRET measurements were performed on a self-built TIRF microscope in the group of Prof. Heilemann at the Goethe University Frankfurt in GLOX buffer (1 mM Tris, 2.5 mM KCl, 50 U/mL glucose oxidase, 2.5 % glycerol, 0.2 mM TCEP) with alternating laser excitation. Settings: 25-32 mW laser intensity at 637 nm; 65 mW laser intensity at 532 nm; gain: 300; integration time: 100 ms; number of images: 2,000-5,000; frequence circular TIRF: 100 Hz;

amplitude circular TIRF: 0.403. Data evaluation was performed with the iSMS and imageJ software.

5.1.30 EPR Measurements

The measurements were performed at room temperature on a Bruker E500 X-Band spectrometer equipped with a Bruker ER4122SHQE resonator containing a quartz insert. The sample was transferred inside a 50 µL micropipette (BLAUBRAND intraMark; ID=0.86 mm) and loaded into the resonator using a 2 mm ID, 3 mm OD Suprasil EPR tube (Wilmad). The spectra were recorded with 1,024 points over a 150 G range using the following parameters: 2 mW microwave power (20 dB attenuation), 81.92 ms conversion time (1.79 G/s), 20.48 ms time constant, 0.5 G modulation amplitude, 100 kHz modulation frequency. The spin concentration was determined by double integration of the CW-EPR signal; the calibration of the signal intensity was performed by measuring a 100 µM water solution of 3-carbamoyl-2,2,5,5-tetramethyl-3-pyrrolin-1-oxyl (Sigma-Aldrich) under the same conditions.

5.1.31 Photocrosslinking Assay

The photocrosslinking assay was performed in 50 µL reaction volume in µclear® transparent cyclo-olefin-copolymer UV-Star® 96-well half-are flat-bottom microplates (Greiner Bio-One) on ice. Protein samples were prepared in Strep buffer and illuminated with a Nu-8 UV lamp (Herolab) at 256 nm (AzPhe) or 365 nm (BzPhe) and 8 W in a distance of 5 cm. 1 µL reaction samples were analyzed on SDS-PAGE using NuPAGE 4-12 % Bis-Tris Protein Gels (Invitrogen). Efficiency of the crosslinking reaction was estimated by densitometric analysis using the software imageJ.

5.1.32 In-Gel Digestion

Crosslink bands were cut out from the polyacrylamide gel and rinsed with sterile water. Incubation of the gel slices in 50 % acetonitrile for 5 min at 300 rpm at room temperature was followed by incubation in absolut acetonitrile for another 5 min at 300 rpm at room temperature. Subsequently, the gel slices were dried for 20 min in the SpeedVac (SavantTM) at 5 mbar. Enzymatic in-gel digestion of the crosslinked proteins was performed at the Urlaub research group for bioanalytic mass spectrometry at the MPI Göttingen by Andreas Linden. UV-irradiated proteins in dried gel slices were reduced with DTT (10 mM final concentration), alkylated with iodoacetamide (55 mM final concentration) and digested with either trypsin or GluC (depending on the construct). Extracted and dried peptides were resuspended in 2 % acetonitrile/0.05 % trifluoroacetic acid (v/v) and subjected to liquid chromatography mass spectrometry (LC-MS) analysis.

5.1.33 Tandem MS Measurements

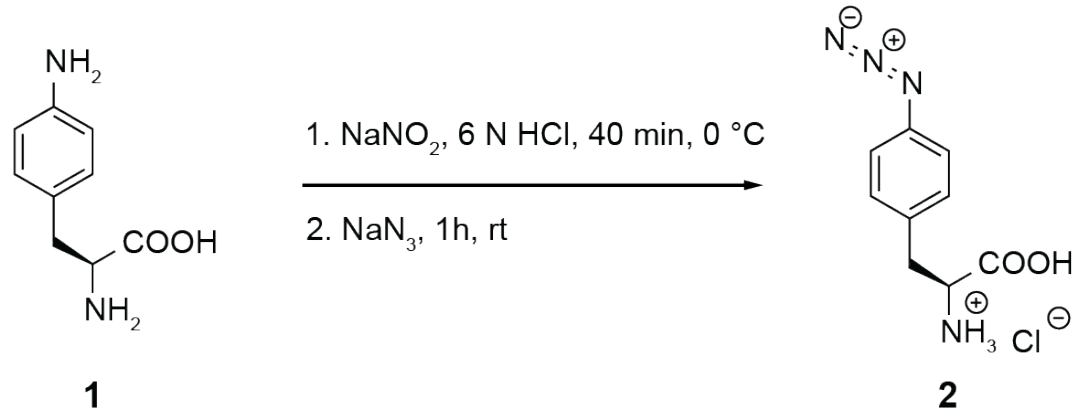
Tandem MS analysis of the crosslinked proteins was performed at the Urlaub research group for bioanalytic mass spectrometry at the MPI Göttingen by Andreas Linden. Peptides were measured in technical duplicates on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer or on a Q Exactive HF-X coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) equipped with an in house-packed C_{18} column (ReproSil-Pur 120 C18-AQ, 1.9 µm pore size, 75 µm inner diameter, 30 cm length, Dr. Maisch GmbH). Samples were separated applying the following gradient: mobile phase A consisted of 0.1 % formic acid (v/v), mobile phase B of 80 % acetonitrile/0.08 % formic acid (v/v). The gradient started at 5 % B, increasing to 10 % B within 3 min, followed by a continuous increase to 46 % B within 45 min, then keeping B constant at 90 % for 8 min. After each gradient the column was again equilibrated to 5 % B for 2 min. The flow rate was set to 300 nL/min. MS1 survey scans were acquired in the orbitrap (OT) with a resolution of 120,000, an injection time (IT) of 60 ms (50 ms on HF-X) and an automatic gain control (AGC) target of 5×105 (1×106). Dynamic exclusion was set to 10 s and charge states between +2 and +8 were considered for fragmentation. MS2 spectra were acquired in the OT of the 20 (30 on HF-X) most abundant precursor ions, resolution 30,000, IT 120 ms (12 ms) and AGC target 5×104 (1×105). Fragmentation was enforced by higher-energy collisional dissociation (HCD) at 30 %.
For data analysis, raw files were analyzed with pLink2.3 [164] to identify cross-linked peptides. Here, default settings were applied with carbamidomethylation of cysteines as fixed and oxidation of methionines as variable modification. FDR was set to 0.01 (or was fully relaxed to 1). Benzophenone A and azidophenylalanine were configured as individual amino acids and corresponding crosslinkers, respectively. Each construct was searched against its modified sequence.

5.2 Chemical Synthesis

All chemicals were used as purchased from Sigma-Aldrich without further processing. Protected amino acids were ordered at Iris Biotech. All syntheses except for AzPhe were performed under dry conditions and argon atmosphere.

5.2.1 Synthesis of Azidophenylalanine (AzPhe) (2)

4-Aminophenylalanine **1** (7.20 g, 34.7 mmol, 1.0 equiv.) was dissolved in 6 N HCl solution (20 mL) and cooled down to 0 °C. Subsequently, a solution of $NaNO_2$ (2.80 g, 40.5 mmol, 1.2 equiv.) in water (5 mL) was prepared and slowly added. The reaction mixture was diluted with 6 N HCl solution (33 mL) and further stirred for 20 min at 0 °C. Then, a solution of NaN_3 (2.50 g, 38.5 mmol, 1.1 equiv.) in water (6 mL) was prepared and added slowly. The reaction mixture was kept at 0 °C for another 15 min and allowed to warm to room temperature for



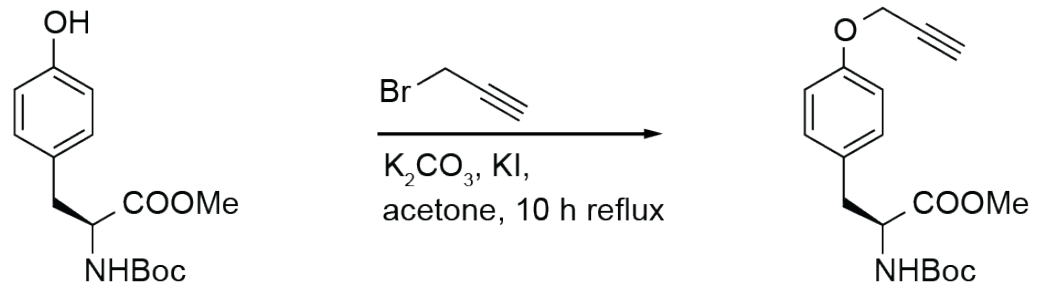
1 **2**

1 h. The hydrochloride was filtered and washed with cold diethyl ether. Recrystallization with absolute ethanol and three volumes of diethyl ether overnight at –20 °C afforded **2** as pale yellow solid (7.30 g, 81.0 %).

^{1}H-NMR (250 MHz, DMSO-d_6) δ (ppm) = 7.34 (d, ^{3}J = 8.4 Hz, 2H), 7.10 (d, ^{3}J = 8.4 Hz, 2H), 4.13 (dd, ^{3}J = 6.4 Hz, 6.3 Hz, 1H), 3.17-3.03 (m, 2H). IR (KBr): Azide band at 2,140 cm^{-1} matches with the literature.

5.2.2 Synthesis of Propargyloxyphenylalanine (PrPhe) (6)

Synthesis of Methyl-2-(*tert*-butoxycarbonyl)amino)-3-(4-(ethynyloxy)phenyl)propanoate (4)



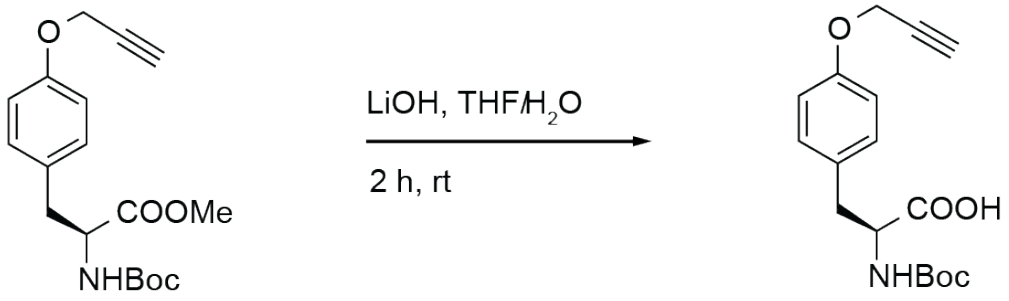
3 **4**

K_2CO_3 (9.10 g, 67.5 mmol, 1.5 equiv.) and KI (0.72 g, 4.50 mmol, 0.1 equiv.) were added to a stirred solution of Boc-L-Tyr-OMe **3** (13.1 g, 45.0 mmol, 1.0 equiv.) in acetone (150 mL) at

room temperature. Propargyl bromide (6.0 mL, 45.0 mmol, 1.0 equiv.) was added dropwise over 30 min and the resulting mixture was heated under reflux for 10 h. The reaction mixture was cooled on ice and filtered two times. The filtrate was concentrated in vacuo and the residue was re-dissolved in ethyl acetate (100 mL), washed twice with water (30 mL), and dried over $MgSO_4$. Evaporation of the solvent in vacuo afforded **4** as yellow oil (14.1 g, 94.0 %).

^{1}H-NMR (300 MHz, $CDCl_3$): δ (ppm) = 7.06 (d, ^{3}J = 8.75 Hz, 2H), 6.91 (d, ^{3}J = 8.75 Hz, 2H), 4.97-4.96 (m, 1H), 4.67 (d, ^{3}J = 2.50 Hz, 2H), 4.54-4.53 (m, 1H), 3.71 (s, 3H), 3.11-3.03 (m, 1H), 2.53-2.52 (m, 1H), 1.42 (s, 9H).

Synthesis of 2-((*tert*-Butoxycarbonyl)amino)-3-(4-(ethynyloxy)phenyl)propanoic Acid (5)

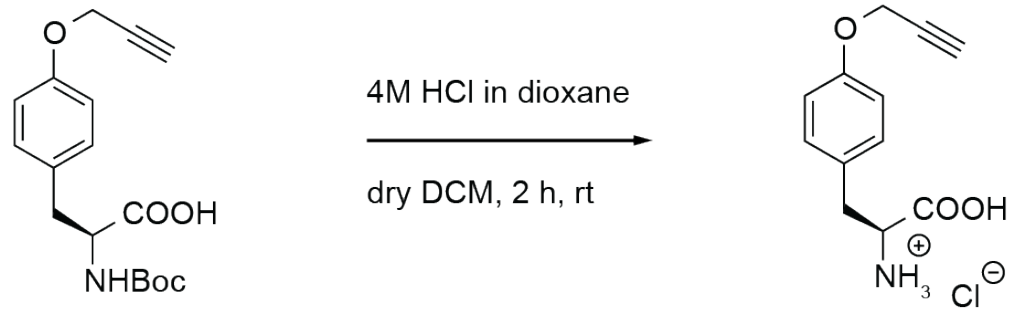


4 **5**

The propargyl derivative Boc-L-PrPhe-OMe **4** (13.3 g, 40.0 mmol, 1.0 equiv.) was dissolved in THF (100 mL) and 1 M LiOH solution (85.0 mL, 85.0 mmol, 2.0 equiv.) was added carefully to the stirred solution at room temperature. After 1 h, completion of the reaction was checked by TLC (1:1 EtOAC/n-Hexan, R_f = 0.8). The reaction mixture was acidified with conc. HCl (10 mL) and the two phases were separated. The aqueous phase was extracted three times with ethyl acetate and the combined organic phases were washed with brine, and dried over $MgSO_4$. Evaporation of the solvent in vacuo afforded **5** as yellow oil (11.7 g, 91.0 %).

^{1}H-NMR (250 MHz, $CDCl_3$): δ (ppm) = 7.13 (d, ^{3}J = 8.51 Hz, 2H), 6.92 (d, ^{3}J = 8.70 Hz, 2H), 4.68 (d, ^{3}J = 2.40 Hz, 2H), 3.19-2.90 (m, 2H), 2.52 (t, ^{3}J = 2.40 Hz, 1H), 2.05 (s, 1H), 1.43 (s, 9H).

Synthesis of 2-Amino-3-(4-(ethynyloxy)phenyl)propanoic Acid Hydrochloride (6)



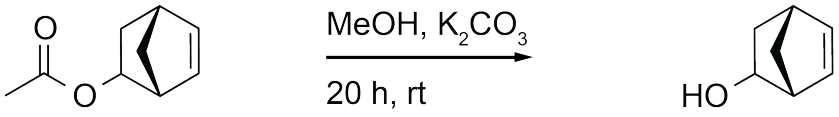
5 **6**

The carboxylic acid Boc-L-PrPhe **5** (11.7 g, 37.0 mmol, 1.0 equiv.) was dissolved in DCM (25 mL) and 4 M HCl in dioxane (15.0 mL, 60.0 mmol, 1.6 equiv.) were added to the stirred solution at room temperature. After 2.5 h of reaction, the solvent was evaporated in vacuo to afford **6** as colourless solid (9.33 g, 99.0 %).

^{1}H-NMR (400 MHz, DMSO-d$_6$): δ (ppm) = 8.40 (br s, 2H), 7.21 (d, ^{3}J = 8.56 Hz, 2H), 6.94 (d, ^{3}J = 8.56 Hz, 2H), 4.77 (d, ^{3}J = 2.32 Hz, 2H), 4.11-4.08 (m, 1H), 3.57-3.56 (m, 1H), 3.08 (d, ^{3}J = 6.11 Hz, 2H).
MALDI-MS (m/z): $[M+Na]^+$ calcd. for $C_{12}H_{13}NO_3$ 242.07876, meas. 242.07914.

5.2.3 Synthesis of Norbornenelysine-1 (NorLys1) (12)

Synthesis of Bicyclo[2.2.1]hept-5-en-2-ol (8)



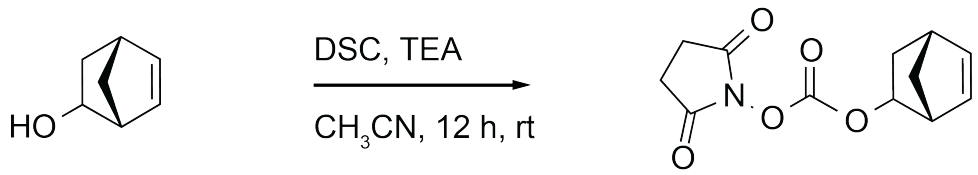
7 **8**

A mixture of 5-norbornene-2-yl acetate **7** (10.0 g, 65.7 mmol, 1.0 equiv.) and K_2CO_3 (9.60 g, 69.7 mmol, 1.1 equiv.) in methanol (100 mL) was stirred for 20 h at room temperature. After evaporation of the solvent in vacuo, the residue was re-dissolved in water (300 mL), and extracted five times with ethyl acetate (100 mL). Evaporation of the solvent in vacuo afforded

8 as off-white crystals (6.20 g, 85.7 %).

^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm) = 6.46-6.45 (m, 1H), 6.07-6.06 (m, 1H), 4.48-4.47 (m, 1H), 3.00 (s, 1H), 2.82 (s, 1H), 2.14-2.10 (m, 1H), 1.33-1.28 (m, 3H).

Synthesis of Bicyclo[2.2.1]hept-5-en-2-yl 2-(2,5-dioxopyrrolidin-1-yl)acetate (9)



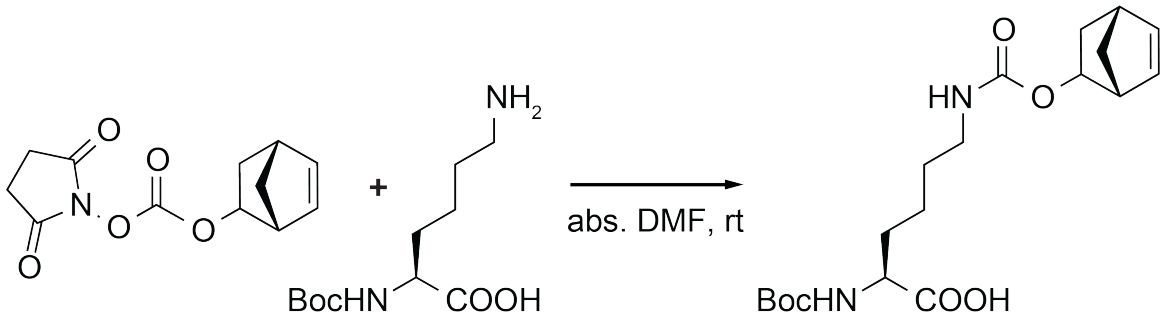
8 **9**

To a solution of the norbornene derivative **8** (3.00 g, 27.0 mmol, 1.0 equiv.) in acetonitrile (100 mL), triethylamine (12.0 mL, 87.0 mmol, 3.2 equiv.) and N,N'-disuccinimidyl carbonate (11.3 g, 44.0 mmol, 1.2 equiv.) were added, and the mixture was stirred for 12 h at room temperature. After evaporation of the solvent in vacuo, the crude product was purified on silica gel (99.5:0.5 DCM/Et_2O) to afford **9** as slightly yellow crystals (6.20 g, 92.0 %).

^{1}H-NMR (250 MHz, $CDCl_3$) δ (ppm) = 6.41-6.30 (m, 1H), 6.04-5.94 (m, 1H), 5.36 and 4.73 (m, 1H), 3.27 and 3.08 (m, 1H), 2.91 (s, 1H), 2.82 (s, 4H), 2.24-2.14 (m, 1H), 1.68-1.10 (m, 3H).

Synthesis of 6-(((Bicyclo[2.2.1]hept-5-en-2-yloxy)carbonyl)amino)-2-((*tert*-butoxycarbonyl)amino)hexanoic Acid (11)

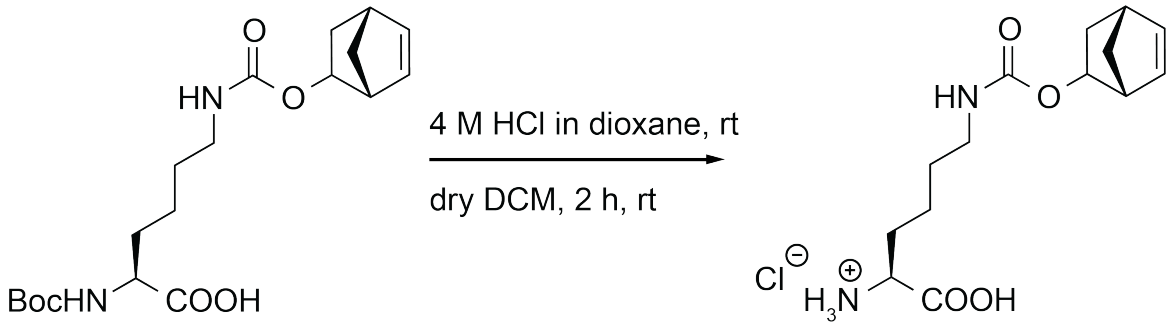
To a solution of the activated norbornene **9** (2.50 g, 10.0 mmol, 1.0 equiv.) in DMF (35 mL), the Boc-protected amino acid Boc-Lys-OH **10** (3.20 g, 13.0 mmol, 1.3 equiv.) was added, and the mixture was stirred overnight at room temperature. After addition of water (300 mL), the mixture was extracted with ethyl acetate (150 mL). The organic phases were combined, washed with water (150 mL) and once with brine (75 mL), and dried over $MgSO_4$. Evaporation of the solvent in vacuo afforded **11** as white foam, which was used directly in the next step without further purification.



9 **10** **11**

^{1}H-NMR (400 MHz, $CDCl_3$) δ (ppm) = 6.29-6.19 (m, 1H), 5.95-5.93 (m, 1H), 5.30-5.22 (m, 2H), 4.93-4.09 (m, 2H), 3.11 (br s, 2H), 2.80-2.79 (m, 1H), 2.09-2.08 (m, 1H), 1.82-1.30 (m, 15H), 0.88-0.87 (m, 1H).

Synthesis of 2-Amino-6-(((bicyclo[2.2.1]hept-5-en-2-yloxy)carbonyl)amino)hexanoic Acid Hydrochloride (12)



11 **12**

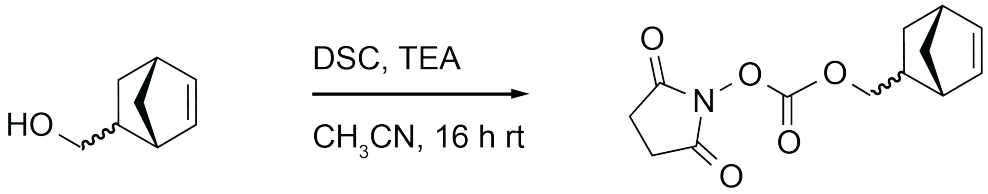
A solution of the Boc-protected norbornene lysine **11** in DCM (25 mL) and 4 M HCl in dioxane (20.0 mL, 80.0 mmol, 8.0 equiv.) was stirred for 1 h at room temperature. After evaporation of the solvent in vacuo, the residue was taken up in diethyl ether and filtrated, to afford **12** as colourless solid (3.10 g, 97.0 %).

^{1}H-NMR (400 MHz, DMSO-d_6) δ (ppm) = 7.08-6.91 (m, 1H), 6.31-6.24 (m, 1H), 6.00-5.92 (m, 1H), 5.11-5.07 and 4.44-4.42 (m, 1H), 3.86-3.78 (m, 1H), 3.45-3.27 (m, 1H), 3.03 (s, 1H),

2.97-2.88 (m, 2H), 2.81-2.77 (m, 1H), 2.06-2.00 (m, 1H), 1.77-1.75 (m, 2H), 1.59-1.55 (m, 1H), 1.42-1.29 (m, 5H), 0.80-0.77 (m, 1H).
MALDI-MS (m/z): $[M+Na]^+$ calcd. for $C_{14}H_{22}N_2O_4$ 305.14718, meas. 305.14741.

5.2.4 Synthesis of Norbornenelysine-2 (NorLys2) (16)

Synthesis of Bicyclo[2.2.1]hept-5-en-2-ylmethyl (2,5-Dioxopyrrolidin-1-yl) Carbonate (14)



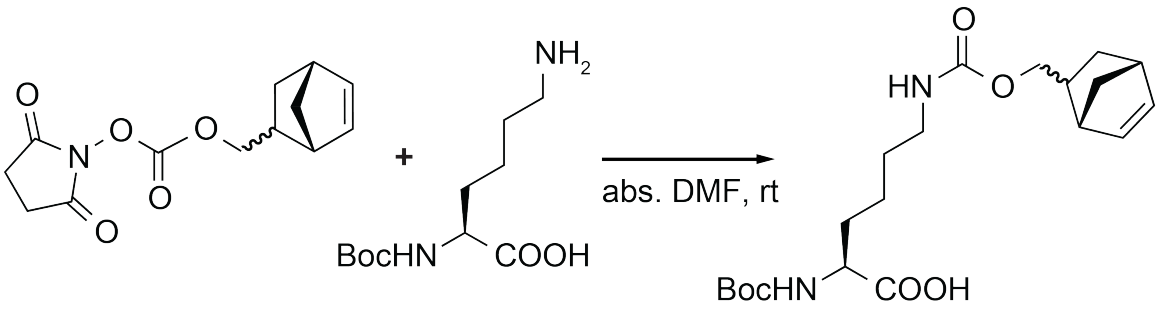
13 **14**

To a solution of N,N'-disuccinimidyl carbonate (13.6 g, 53.6 mmol, 1.1 equiv.) and 5-norbornene-2-methanol **13** (5.80 mL, 48.0 mmol, 1.0 equiv.) in acetonitrile (120 mL), triethylamine (24 mL) were added, and the mixture was stirred at room temperature overnight. After evaporation of the solvent in vacuo, the crude product was purified on silica gel (99.5:0.5. DCM/Et_2O) to afford **14** as colourless solid (10.2 g, 80.0 %).

^{1}H-NMR (500 MHz, $CDCl_3$) δ (ppm) = 6.20 (dd, ^{3}J = 3.1 Hz, 5.7 Hz, 1H), 6.13-6.10 (m, 1H), 5.99 (dd, ^{3}J = 2.9 Hz, 5.7 Hz, 1H), 4.41 (dd, ^{3}J = 6.5 Hz, 6.4 Hz, 0.5H), 4.23 (m, 0.5H), 4.11 (dd, ^{3}J = 6.6 Hz, 10.3 Hz, 1H), 3.94 (dd, ^{3}J = 9.8 Hz, 10.1 Hz, 1H), 2.97 (s, 1H), 2.55-2.49 (m, 1H), 1.92-1.83 (m, 2H), 1.50 (dd, ^{3}J = 2.08 Hz, 8.4 Hz, 1H), 1.42-1.18 (m, 3H), 0.61-0.57 (m, 1H).
ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_{13}H_{15}NO_5$ 288.25, meas. 288.16.

Synthesis of 6-(((Bicyclo[2.2.1]hept-5-en-2-ylmethoxy)carbonyl)amino)-2-((*tert*-butoxycarbonyl)amino)hexanoic Acid (15)

A solution of compound **14** (9.66 g, 36.4 mmol, 1.0 equiv.) and Boc-protected lysine **10** (9.88 g, 40.0 mmol, 1.1 equiv.) in DMF (100 mL) was stirred at room temperature overnight. After addition of water (400 mL), the mixture was extracted with ethyl acetate (500 mL). The organic

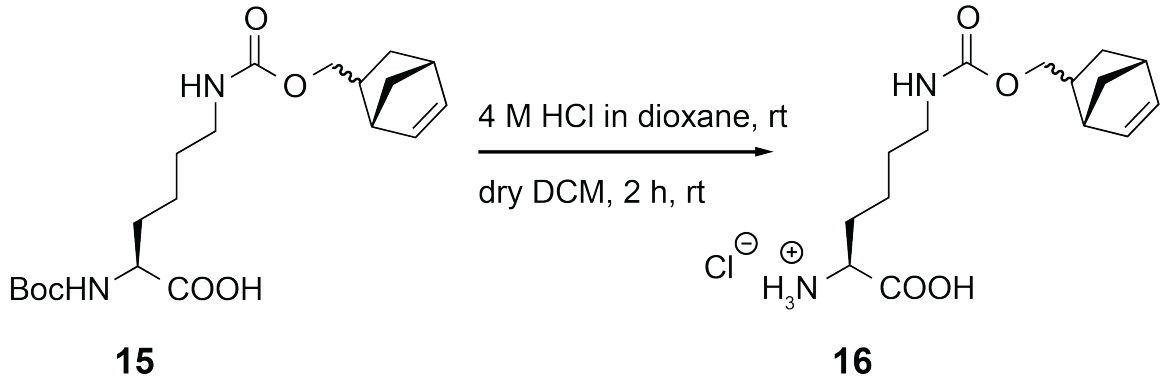


14 **10** **15**

phase was washed with water (150 mL) and subsequently, the aqueous phase was extracted twice with ethyl acetate (400 mL). The organic phases were combined and washed again with water (400 mL) and brine (200 mL), and dried over $MgSO_4$. Evaporation of the solvent in vacuo afforded **15** as slightly yellow foam (14.4 g, 99.0 %).

^{1}H-NMR (500 MHz, $CDCl_3$) δ (ppm) = 6.17-12 (m, 1H), 6.11-6.06 (m, 1H), 5.95-5.94 (m, 1H), 5.29-5.27 (m, 1H), 4.36-4.26 (m, 1H), 3.20-3.17 (m, 3H), 2.98 (s, 1H), 2.89-2.85 (m, 1H), 1.92-1.80 (m, 3H), 1.57-1.51 (m, 3H), 1.48-1.39 (m, 14H).
ESI-MS (m/z): $[M+H]^+$ calcd. for $C_{20}H_{32}N_2O_6$ 397.48, meas. 397.25.

Synthesis of 2-Amino-6-(((bicyclo[2.2.1]hept-5-en-2-ylmethoxy)carbonyl)amino)hexanoic Acid Hydrochloride (16)



15 **16**

To a solution of the Boc-protected intermediate **15** (14.4 g, 36.0 mmol, 1.0 equiv.) in DCM (100 mL), 4 M HCl in dioxane (80.0 mL, 0.32 mol, 8.8 equiv.) were added and the solution

was stirred for 1 h at room temperature. Evaporation of the solvent in vacuo afforded **16** as colourless solid (10.4 g, 87.0 %).

^{1}H-NMR (250 MHz, DSMO-d_6) δ (ppm) = 8.31 (m, 2H), 7.14-7.06 (m, 1H), 6.20-6.17 (m, 1H), 6.11-6.09 (m, 1H), 5.95-5.93 (m, 1H), 4.38-4.34 (m, 0.3H), 4.07-4.01 (m, 1H), 2.90 (s, 0.7H), 2.82-2.80 (m, 2H), 2.25-2.23 (m, 1H), 1.86-1.70 (m, 8H), 1.16-1.10 (m, 1H).
MS-ESI (m/z): $[M+H]^+$ calcd. for $C_{15}H_{24}N_2O_4$ 297.36, meas. 297.20.

References

[1] S. Smith and S.-C. Tsai, "The Type I Fatty Acid and Polyketide Synthases: a Tale of Two Megasynthases," *Nat. Prod. Rep.*, vol. 24, no. 5, pp. 1041–1072, 2007.

[2] G. Ronnett, E. Kim, L. Landree, and Y. Tu, "Fatty Acid Metabolism as a Target for Obesity Treatment," *Physiology & Behavior*, vol. 85, no. 1, pp. 25–35, 2005.

[3] J. A. Menendez, A. Vazquez-Martin, F. J. Ortega, and J. M. Fernandez-Real, "Fatty Acid Synthase: Association with Insulin Resistance, Type 2 Diabetes, and Cancer," *Clinical Chemistry*, vol. 55, no. 3, pp. 425–438, 2009.

[4] F. P. Kuhajda, "Fatty Acid Synthase and Cancer: New Application of an Old Pathway.," *Cancer Res*, vol. 66, no. 12, pp. 5977–5980, 2006.

[5] D. A. Herbst, C. A. Townsend, and T. Maier, "The Architectures of Iterative Type I PKS and FAS," *Nat. Prod. Rep.*, vol. 35, no. 10, pp. 1046–1069, 2018.

[6] M. Grininger, C. S. Heil, S. S. Wehrheim, and K. S. Paithankar, "Fatty Acid Biosynthesis: Chain Length Regulation and Control," *ChemBioChem*, pp. 1–25, 2019.

[7] O. Tehlivets, K. Scheuringer, and S. D. Kohlwein, "Fatty Acid Synthesis and Elongation in Yeast," *Biochim. Biophys. Acta*, vol. 1771, no. 3, pp. 255–270, 2007.

[8] E. Schweizer and J. Hofmann, "Microbial Type I Fatty Acid Synthases (FAS): Major Players in a Network of Cellular FAS Systems," *Microbiol. Mol. Biol. R.*, vol. 68, no. 3, pp. 501–517, 2004.

[9] S. Smith, "The Animal Fatty Acid Synthase: One Gene, One Polypeptide, Seven Enzymes," *FASEB J.*, vol. 8, no. 15, pp. 1248–1259, 1994.

[10] F. M. C. Benning, Y. Sakiyama, A. Mazur, H. S. T. Bukhari, R. Y. H. Lim, and T. Maier, "High-Speed Atomic Force Microscopy Visualization of the Dynamics of the Multienzyme Fatty Acid Synthase," *ACS Nano*, vol. 11, no. 11, pp. 10852–10859, 2017.

[11] E. J. Brignole, S. Smith, and F. J. Asturias, "Conformational Flexibility of Metazoan Fatty Acid Synthase Enables Catalysis," *Nat. Struct. Mol. Biol.*, vol. 16, no. 2, pp. 190–197, 2009.

[12] A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao, and P. G. Schultz, "A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*," *Biochemistry*, vol. 52, no. 10, pp. 1828–1837, 2013.

[13] T. S. Young, I. Ahmad, J. A. Yin, and P. G. Schultz, "An Enhanced System for Unnatural Amino Acid Mutagenesis in *E. coli*," *J. Mol. Biol.*, vol. 395, no. 2, pp. 361–374, 2010.

[14] C. C. Liu and P. G. Schultz, "Adding New Chemistries to the Genetic Code," *Annu. Rev. Biochem.*, vol. 79, no. 1, pp. 413–444, 2010.

[15] P. G. Isenegger and B. G. Davis, "Concepts of Catalysis in Site-Selective Protein Modifications," *J. Am. Chem. Soc.*, vol. 141, no. 20, pp. 8005–8013, 2019.

[16] H. Suga, "Max-Bergmann Award Lecture: A RaPID Way to Discover Bioactive Nonstandard Peptides Assisted by the Flexizyme and FIT Systems," *J. Pept. Sci.*, vol. 24, no. 1, pp. 3055–3057, 2018.

[17] Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga, "Expanding the Amino Acid Repertoire of Ribosomal Polypeptide Synthesis via the Artificial Division of Codon Boxes," *Nat. Chem.*, vol. 8, no. 4, pp. 317–325, 2016.

[18] L. Wang, A. Brock, B. Herberich, and P. G. Schultz, "Expanding the Genetic Code of *Escherichia coli*," *Science*, vol. 292, pp. 498–500, 2001.

[19] A. Dumas, L. Lercher, C. D. Spicer, and B. G. Davis, "Designing Logical Codon Reassignment – Expanding the Chemistry in Biology," *Chem. Sci.*, vol. 6, no. 1, pp. 50–69, 2015.

[20] J. W. Chin, "Expanding and Reprogramming the Genetic Code of Cells and Animals," *Annu. Rev. Biochem.*, vol. 83, no. 1, pp. 379–408, 2014.

[21] S. Tyagi and E. A. Lemke, "Single-Molecule FRET and Crosslinking Studies in Structural Biology Enabled by Noncanonical Amino Acids," *Curr. Opin. Struc. Biol.*, vol. 32, pp. 66–73, 2015.

[22] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H. R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, and F. J. Isaacs, "Genomically Recoded Organisms Expand Biological Functions," *Science*, vol. 342, no. 6156, pp. 357–360, 2013.

[23] F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C.

Jewett, J. M. Jacobson, and G. M. Church, "Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement," *Science*, vol. 333, no. 6040, pp. 348–353, 2011.

[24] T. Mukai, H. Hoshi, K. Ohtake, M. Takahashi, A. Yamaguchi, A. Hayashi, S. Yokoyama, and K. Sakamoto, "Highly Reproductive *Escherichia coli* Cells with No Specific Assignment to the UAG Codon," *Sci. Rep.*, vol. 5, no. 1, pp. 1056–1059, 2015.

[25] J. S. Italia, P. S. Addy, S. B. Erickson, J. C. Peeler, E. Weerapana, and A. Chatterjee, "Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites," *J. Am. Chem. Soc.*, vol. 141, no. 15, pp. 6204–6212, 2019.

[26] K. Wang, H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Evolved Orthogonal Ribosomes Enhance the Efficiency of Synthetic Genetic Code Expansion," *Nat. Biotechnol.*, vol. 25, no. 7, pp. 770–777, 2007.

[27] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, and J. W. Chin, "Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome," *Nature*, vol. 464, no. 7287, pp. 441–444, 2010.

[28] M. J. Lajoie, D. Söll, and G. M. Church, "Overcoming Challenges in Engineering the Genetic Code," *J. Mol. Biol.*, vol. 428, no. Part B, pp. 1004–1021, 2016.

[29] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Corrêa, and F. E. Romesberg, "A Semi-Synthetic Organism with an Expanded Genetic Alphabet," *Nature*, vol. 509, pp. 385–388, 2014.

[30] S. Greiss and J. W. Chin, "Expanding the Genetic Code of an Animal," *J. Am. Chem. Soc.*, vol. 133, no. 36, pp. 14196–14199, 2011.

[31] R. J. Ernst, T. P. Krogager, E. S. Maywood, R. Zanchi, V. Beránek, T. S. Elliott, N. P. Barry, M. H. Hastings, and J. W. Chin, "Genetic Code Expansion in the Mouse Brain," *Nat. Chem. Biol.*, vol. 12, pp. 776–778, 2016.

[32] J. Beld, D. J. Lee, and M. D. Burkart, "Fatty Acid Biosynthesis Revisited: Structure Elucidation and Metabolic Engineering," *Mol. BioSyst.*, vol. 11, no. 1, pp. 38–59, 2015.

[33] S. W. White, J. Zheng, Y.-M. Zhang, and C. O. Rock, "The Structural Biology of Type II Fatty Acid Biosynthesis," *Annu. Rev. Biochem.*, vol. 74, pp. 791–831, 2005.

[34] T. Maier, M. Leibundgut, D. Boehringer, and N. Ban, "Structure and Function of Eukaryotic Fatty Acid Synthases," *Q. Rev. Biophys.*, vol. 43, no. 3, pp. 373–422, 2010.

[35] A. J. Kastaniotis, K. J. Autio, J. M. Kerätär, G. Monteuuis, A. M. Mäkelä, R. R. Nair, L. P. Pietikäinen, A. Shvetsova, Z. Chen, and J. K. Hiltunen, "Mitochondrial Fatty Acid Synthesis, Fatty Acids and Mitochondrial Physiology," *Biochim. Biophys. Acta*, vol. 1862, no. 1, pp. 39–48, 2017.

[36] J. K. Hiltunen, M. S. Schonauer, K. J. Autio, T. M. Mittelmeier, A. J. Kastaniotis, and C. L. Dieckmann, "Mitochondrial Fatty Acid Synthesis Type II: More Than Just Fatty Acids," *J. Biol. Chem.*, vol. 284, no. 14, pp. 9011–9015, 2009.

[37] Y.-J. Lu, Y.-M. Zhang, K. D. Grimes, J. Qi, R. E. Lee, and C. O. Rock, "Acyl-Phosphates Initiate Membrane Phospholipid Synthesis in Gram-Positive Pathogens," *Mol. Cell*, vol. 23, no. 5, pp. 765–772, 2006.

[38] J. G. Van Vranken, S. M. Nowinski, K. J. Clowers, M. Y. Jeong, Y. Ouyang, J. A. Berg, J. P. Gygi, S. P. Gygi, D. R. Winge, and J. Rutter, "ACP Acylation Is an Acetyl-CoA-Dependent Modification Required for Electron Transport Chain Assembly," *Mol. Cell*, vol. 71, no. 4, pp. 567–580, 2018.

[39] M. R. Pollard, L. Anderson, C. Fan, D. J. Hawkins, and M. H. Davies, "A Specific Acyl-ACP Thioesterase Implicated in Medium-Chain Fatty Acid Production in Immature Cotyledons of *Umbellularia californica*," *Arch. Biochem. Biophys.*, vol. 284, no. 2, pp. 306–312, 1991.

[40] J. Crosby and M. P. Crump, "The Structural Role of the Carrier Protein - Active Controller or Passive Carrier," *Nat. Prod. Rep.*, vol. 29, no. 10, pp. 1111–1137, 2012.

[41] R. N. Perham, "Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions.," *Annu. Rev. Biochem.*, vol. 69, pp. 961–1004, 2000.

[42] Y.-M. Zhang, M. S. Rao, R. J. Heath, A. C. Price, A. J. Olson, C. O. Rock, and S. W. White, "Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on β-Ketoacyl-ACP Synthase III," *J. Biol. Chem.*, vol. 276, no. 11, pp. 8231–8238, 2001.

[43] S. E. Evans, C. Williams, C. J. Arthur, S. G. Burston, T. J. Simpson, J. Crosby, and M. P. Crump, "An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein," *ChemBioChem*, vol. 9, no. 15, pp. 2424–2432, 2008.

[44] S. E. Evans, C. Williams, C. J. Arthur, E. Płoskoń, P. Wattana-amorn, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson, and M. P. Crump, "Probing the Interactions of Early Polyketide Intermediates with the Actinorhodin ACP from *S. coelicolor* A3(2)," *J. Mol. Biol.*, vol. 389, no. 3, pp. 511–528, 2009.

[45] R. W. Haushalter, F. V. Filipp, K.-s. Ko, R. Yu, S. J. Opella, and M. D. Burkart, "Binding and pK_a Modulation of a Polycyclic Substrate Analogue in a Type II Polyketide Acyl Carrier Protein," *ACS Chem. Biol.*, vol. 6, no. 5, pp. 413–418, 2011.

[46] M. J. Jaremko, D. J. Lee, S. J. Opella, and M. D. Burkart, "Structure and Substrate Sequestration in the Pyoluteorin Type II Peptidyl Carrier Protein PltL," *J. Am. Chem. Soc.*, vol. 137, no. 36, pp. 11546–11549, 2015.

[47] G. Shakya, H. Rivera Jr., D. J. Lee, M. J. Jaremko, J. J. La Clair, D. T. Fox, R. W. Haushalter, A. J. Schaub, J. Bruegger, J. F. Barajas, A. R. White, P. Kaur, E. R. Gwozdziowski, F. Wong, S.-C. Tsai, and M. D. Burkart, "Modeling Linear and Cyclic PKS Intermediates Through Atom Replacement," *J. Am. Chem. Soc.*, vol. 136, no. 48, pp. 16792–16799, 2014.

[48] G. A. Zornetzer, B. G. Fox, and J. L. Markley, "Solution Structures of Spinach Acyl Carrier Protein with Decanoate and Stearate," *Biochemistry*, vol. 45, no. 16, pp. 5217–5227, 2006.

[49] D. R. Perez, M. Leibundgut, and G. Wider, "Interactions of the Acyl Chain with the *Saccharomyces cerevisiae* Acyl Carrier Protein," *Biochemistry*, vol. 54, no. 13, pp. 2205–2213, 2015.

[50] E. Ploskon, C. J. Arthur, S. E. Evans, C. Williams, J. Crosby, T. J. Simpson, and M. P. Crump, "A Mammalian Type I Fatty Acid Synthase Acyl Carrier Protein Domain Does Not Sequester Acyl Chains," *J. Biol. Chem.*, vol. 283, no. 1, pp. 518–528, 2008.

[51] L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli, and K. J. Weissman, "Insights into Protein-Protein and Enzyme-Substrate Interactions in Modular Polyketide Synthases," *Chem. Biol.*, vol. 17, no. 7, pp. 705–716, 2010.

[52] Z. Yuan and G. G. Hammes, "Fluorescence Studies of Chicken Liver Fatty Acid Synthase," *J. Biol. Chem.*, vol. 261, no. 29, pp. 13643–13651, 1986.

[53] S.-I. Chang and G. G. Hammes, "Amino Acid Sequences of Pyridoxal 5'-Phosphate Binding Sites and Fluorescence Resonance Energy Transfer in Chicken Liver Fatty Acid Synthase?," *Biochemistry*, no. 28, pp. 3781–3788, 1989.

[54] S.-I. Chang and G. G. Hammes, "Structure and Mechanism of Action of a Multifunctional Enzyme: Fatty Acid Synthase," *Acc. Chem. Res.*, vol. 23, no. 11, pp. 363–369, 1990.

[55] P. Johansson, B. Mulinacci, C. Koestler, R. Vollrath, D. Oesterhelt, and M. Grininger, "Multimeric Options for the Auto-Activation of the *Saccharomyces cerevisiae* FAS Type I Megasynthase.," *Structure*, vol. 17, no. 8, pp. 1063–1074, 2009.

[56] D. Boehringer, N. Ban, and M. Leibundgut, "7.5-Å Cryo-EM Structure of the Mycobacterial Fatty Acid Synthase," *J. Mol. Biol.*, vol. 425, no. 5, pp. 841–849, 2013.

[57] T. Maier, M. Leibundgut, and N. Ban, "The Crystal Structure of a Mammalian Fatty Acid Synthase," *Science*, vol. 321, no. 5894, pp. 1315–1322, 2008.

[58] B. Chakravarty, Z. Gu, S. S. Chirala, S. J. Wakil, and F. A. Quiocho, "Human Fatty Acid Synthase: Structure and Substrate Selectivity of the Thioesterase Domain," *Proc. Natl. Acad. Sci. USA*, vol. 101, no. 44, pp. 15567–15572, 2004.

[59] H. S. T. Bukhari, R. P. Jakob, and T. Maier, "Evolutionary Origins of the Multienzyme Architecture of Giant Fungal Fatty Acid Synthase," *Structure*, vol. 22, no. 12, pp. 1775–1785, 2014.

[60] H. Jenke-Kodama, A. Sandmann, R. Müller, and E. Dittmann, "Evolutionary Implications of Bacterial Polyketide Synthases.," *Mol. Biol. Evol.*, vol. 22, no. 10, pp. 2027–2039, 2005.

[61] C. Vogel, M. Bashton, N. D. Kerrison, C. Chothia, and S. A. Teichmann, "Structure, Function and Evolution of Multidomain Proteins," *Curr. Opin. Struc. Biol.*, vol. 14, no. 2, pp. 208–216, 2004.

[62] L. J. Sweetlove and A. R. Fernie, "The Role of Dynamic Enzyme Assemblies and Substrate Channelling in Metabolic Regulation," *Nat. Commun.*, vol. 9, no. 1, pp. 1–12, 2018.

[63] M. Rawlings and J. E. J. Cronan, "The Gene Encoding *Escherichia coli* Acyl Carrier Protein Lies within a Cluster of Fatty Acid Biosynthetic Genes," *J. Biol. Chem.*, vol. 267, no. 9, pp. 5751–5754, 1992.

[64] G. Petrovics, P. Putnoky, B. Reuhs, J. Kim, T. A. Thorp, K. D. Noel, R. W. Carlson, and A. Kondorosi, "The Presence of a Novel Type of Surface Polysaccharide in *Rhizobium meliloti* Requires a New Fatty Acid Synthase-Like Gene Cluster Involved in Symbiotic Nodule Development," *Mol. Microbiol.*, vol. 8, no. 6, pp. 1083–1094, 1993.

[65] D. A. Herbst, R. P. Jakob, F. Zähringer, and T. Maier, "Mycocerosic Acid Synthase Exemplifies the Architecture of Reducing Polyketide Synthases," *Nature*, vol. 531, no. 7595, pp. 533–537, 2016.

[66] M. Mathur and P. E. Kolattukudy, "Molecular Cloning and Sequencing of the Gene for Mycocerosic Acid Synthase, a Novel Fatty Acid Elongating Multifunctional Enzyme, from *Mycobacterium tuberculosis* var. *bovis Bacillus Calmette-Guerin*," *J. Biol. Chem.*, vol. 267, no. 27, pp. 19388–19395, 1992.

[67] K. J. Weissman, "The Structural Biology of Biosynthetic Megaenzymes," *Nat. Chem. Biol.*, vol. 11, no. 9, pp. 660–670, 2015.

[68] M. Grininger, "Perspectives on the Evolution, Assembly and Conformational Dynamics of Fatty Acid Synthase Type I (FAS I) Systems," *Curr. Opin. Struc. Biol.*, vol. 25, pp. 49–56, 2014.

[69] A. K. Joshi, V. S. Rangan, A. Witkowski, and S. Smith, "Engineering of an Active Animal Fatty Acid Synthase Dimer with Only One Competent Subunit," *Chem. Biol.*, vol. 10, no. 2, pp. 169–173, 2003.

[70] S. Smith, A. Witkowski, and A. K. Joshi, "Structural and Functional Organization of the Animal Fatty Acid Synthase," *Progress Lipid Res.*, vol. 42, no. 4, pp. 289–317, 2003.

[71] J. Brink, S. J. Ludtke, C.-Y. Yang, Z.-W. Gu, S. J. Wakil, and W. Chiu, "Quaternary Structure of Human Fatty Acid Synthase by Electron Cryomicroscopy," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 1, pp. 138–143, 2002.

[72] S. S. Chirala, "Coordinated Regulation and Inositol-Mediated and Fatty Acid-Mediated Repression of Fatty Acid Synthase Genes in *Saccharomyces cerevisiae*," *Proc. Natl. Acad. Sci. USA*, vol. 89, no. 21, pp. 10232–10236, 1992.

[73] T. Maier, S. Jenni, and N. Ban, "Architecture of Mammalian Fatty Acid Synthase at 4.5 Å Resolution.," *Science*, vol. 311, no. 5765, pp. 1258–1262, 2006.

[74] G. Pappenberger, J. Benz, B. Gsell, M. Hennig, A. Ruf, M. Stihle, R. Thoma, and M. G. Rudolph, "Structure of the Human Fatty Acid Synthase KS–MAT Didomain as a Framework for Inhibitor Design," *J. Mol. Biol.*, vol. 397, no. 2, pp. 508–519, 2010.

[75] A. Rittner, K. S. Paithankar, K. V. Huu, and M. Grininger, "Characterization of the Polyspecific Transferase of Murine Type I Fatty Acid Synthase (FAS) and Implications for Polyketide Synthase (PKS) Engineering," *ACS Chem. Biol.*, vol. 13, no. 3, pp. 723–732, 2018.

[76] M. F. Viegas, R. P. P. Neves, M. J. Ramos, and P. A. Fernandes, "Modeling of Human Fatty Acid Synthase and in Silico Docking of Acyl Carrier Protein Domain and Its Partner Catalytic Domains," *J. Phys. Chem. B*, vol. 122, no. 1, pp. 77–85, 2017.

[77] M. Leibundgut, S. Jenni, C. Frick, and N. Ban, "Structural Basis for Substrate Delivery by Acyl Carrier Protein in the Yeast Fatty Acid Synthase," *Science*, vol. 316, no. 5822, pp. 288–290, 2007.

[78] P. Gipson, D. J. Mills, R. Wouts, M. Grininger, J. Vonck, and W. Kühlbrandt, "Direct Structural Insight Into the Substrate-Shuttling Mechanism of Yeast Fatty Acid

Synthase by Electron Cryomicroscopy," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 20, pp. 9164–9169, 2010.

[79] C. Anselmi, M. Grininger, P. Gipson, and J. D. Faraldo-Gómez, "Mechanism of Substrate Shuttling by the Acyl-Carrier Protein Within the Fatty Acid Mega-Synthase," *J. Am. Chem. Soc.*, vol. 132, no. 35, pp. 12357–12364, 2010.

[80] M. Castellana, M. Z. Wilson, Y. Xu, P. Joshi, I. M. Cristea, J. D. Rabinowitz, Z. Gitai, and N. S. Wingreen, "Enzyme Clustering Accelerates Processing of Intermediates Through Metabolic Channeling," *Nat. Biotechnol.*, vol. 32, no. 10, pp. 1011–1018, 2014.

[81] I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov, and M. Sigman, "Substrate Channelling as an Approach to Cascade Reactions," *Nat. Chem.*, vol. 8, no. 4, pp. 299–309, 2016.

[82] M. Fischer and M. Grininger, "Strategies in Megasynthase Engineering – Fatty Acid Synthases (FAS) as Model Proteins," *Beilstein J. Org. Chem.*, vol. 13, pp. 1204–1211, 2017.

[83] Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner, and D. Frishman, "Protein Abundance Profiling of the *Escherichia coli* Cytosol," *BMC Genomics*, vol. 9, pp. 102–118, 2008.

[84] B. Volkmer and M. Heinemann, "Condition-Dependent Cell Volume and Concentration ofEscherichia colito Facilitate Data Conversion for Systems Biology Modeling," *PLoS ONE*, vol. 6, no. 7, p. e23126, 2011.

[85] T. Maier, "Fatty Acid Synthases: Re-Engineering Biofactories," *Nat. Chem. Biol.*, vol. 13, no. 4, pp. 344–345, 2017.

[86] A. Rittner, K. S. Paithankar, D. J. Drexler, A. Himmler, and M. Grininger, "Probing the Modularity of Megasynthases by Rational Engineering of a Fatty Acid Synthase Type I," *Protein Sci.*, vol. 28, no. 2, pp. 414–428, 2018.

[87] A. K. Joshi, A. Witkowski, H. A. Berman, L. Zhang, and S. Smith, "Effect of Modification of the Length and Flexibility of the Acyl Carrier ProteinThioesterase Interdomain Linker on Functionality of the Animal Fatty Acid Synthase," *Biochemistry*, vol. 44, no. 10, pp. 4100–4107, 2005.

[88] C. Leber and N. A. Da Silva, "Engineering of *Saccharomyces cerevisiae* for the Synthesis of Short Chain Fatty Acids," *Biotechnol. Bioeng.*, vol. 111, no. 2, pp. 347–358, 2014.

[89] M. Fischer, B. Mulinacci, M. Joppe, R. Vollrath, K. Konstantinidis, P. Kötter, L. Ciccarelli, J. Vonck, D. Oesterhelt, and M. Grininger, "Molecular Mechanisms in Fungal Fatty Acid Synthase (FAS) Assembly," *bioRxiv*, pp. 1–21, 2018.

[90] Z. Zhu, Y. J. Zhou, A. Krivoruchko, M. Grininger, Z. K. Zhao, and J. Nielsen, "Expanding the Product Portfolio of Fungal Type I Fatty Acid Synthases," *Nat. Chem. Biol.*, vol. 13, pp. 360–362, 2017.

[91] P. Xu, K. Qiao, W. S. Ahn, and G. Stephanopoulos, "Engineering *Yarrowia lipolytica* as a Platform for Synthesis of Drop-In Transportation Fuels and Oleochemicals," *Proc. Natl. Acad. Sci. USA*, vol. 113, no. 39, pp. 10848–10853, 2016.

[92] J. Gajewski, R. Pavlovic, M. Fischer, E. Boles, and M. Grininger, "Engineering Fungal de Novo Fatty Acid Synthesis for Short Chain Fatty Acid Production," *Nat. Commun.*, vol. 8, pp. 14650–14657, 2017.

[93] J. Gajewski, F. Buelens, S. Serdjukow, M. Janßen, N. S. Cortina, H. Grubmüller, and M. Grininger, "Engineering Fatty Acid Synthases for Directed Polyketide Production," *Nat. Chem. Biol.*, vol. 13, no. 4, pp. 363–365, 2017.

[94] C. Rigouin, M. Gueroult, C. Croux, G. Dubois, V. Borsenberger, S. Barbe, A. Marty, F. Daboussi, I. André, and F. Bordes, "Production of Medium Chain Fatty Acids by Yarrowia lipolytica: Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase," *ACS Synth. Biol.*, vol. 6, no. 10, pp. 1870–1879, 2017.

[95] V. S. Rangan and S. Smith, "Alteration of the Substrate Specificity of the Malonyl-CoA/Acetyl-CoA:Acyl Carrier Protein S-Acyltransferase Domain of the Multifunctional Fatty Acid Synthase by Mutation of a Single Arginine Residue," *J. Biol. Chem.*, vol. 272, no. 18, pp. 11975–11978, 1997.

[96] E. Rossini, J. Gajewski, M. Klaus, G. Hummer, and M. Grininger, "Analysis and Engineering of Substrate Shuttling by the Acyl Carrier Protein (ACP) in Fatty Acid Synthases (FASs)," *Chem. Commun.*, vol. 54, no. 82, pp. 11606–11609, 2018.

[97] Y. J. Choi and S. Y. Lee, "Microbial Production of Short-Chain Alkanes," *Nature*, vol. 502, no. 7472, pp. 571–574, 2013.

[98] M.-K. Kang and J. Nielsen, "Biobased Production of Alkanes and Alkenes Through Metabolic Engineering of Microorganisms," *J. Ind. Microbiol. Biot.*, vol. 44, no. 4-5, pp. 613–622, 2017.

[99] E. R. Marella, C. Holkenbrink, V. Siewers, and I. Borodina, "Engineering Microbial Fatty Acid Metabolism for Biofuels and Biochemicals," *Curr. Opin. Biotech.*, vol. 50, pp. 39–46, 2018.

[100] W. Runguphan and J. D. Keasling, "Metabolic Engineering of *Saccharomyces cerevisiae* for Production of Fatty Acid-Derived Biofuels and Chemicals," *Metab. Eng.*, vol. 21, pp. 103–113, 2014.

[101] S. Sarria, N. S. Kruyer, and P. P. Peralta-Yahya, "Microbial Synthesis of Medium-Chain Chemicals from Renewables," *Nat. Biotechnol.*, vol. 35, no. 12, pp. 1158–1166, 2017.

[102] Y. Zhang, J. Nielsen, and Z. Liu, "Metabolic Engineering of *Saccharomyces cerevisiae* for Production of Fatty Acid-Derived Hydrocarbons," *Biotechnol. Bioeng.*, vol. 115, no. 9, pp. 2139–2147, 2018.

[103] Y. J. Zhou, N. A. Buijs, Z. Zhu, J. Qin, V. Siewers, and J. Nielsen, "Production of Fatty Acid-Derived Oleochemicals and Biofuels by Synthetic Yeast Cell Factories.," *Nat. Commun.*, vol. 7, no. 1, pp. 11709–11717, 2016.

[104] S. Hammes-Schiffer and S. J. Benkovic, "Relating Protein Motion to Catalysis," *Annu. Rev. Biochem.*, vol. 75, pp. 519–541, 2006.

[105] L. E. Kay, "NMR Studies of Protein Structure and Dynamics," *J. Magn. Reson.*, vol. 213, no. 2, pp. 477–491, 2005.

[106] W. Kühlbrandt, "The Resolution Revolution," *Science*, vol. 343, no. 6178, pp. 1443–1444, 2014.

[107] L. Ciccarelli, S. R. Connell, M. Enderle, D. J. Mills, J. Vonck, and M. Grininger, "Structure and Conformational Variability of the *Mycobacterium tuberculosis* Fatty Acid Synthase Multienzyme Complex," *Structure*, vol. 21, no. 7, pp. 1251–1257, 2013.

[108] M. Diez, B. Zimmermann, M. Börsch, M. König, E. Schweinberger, S. Steigmiller, R. Reuter, S. Felekyan, V. Kudryavtsev, C. A. M. Seidel, and P. Gräber, "Proton-Powered Subunit Rotation in Single Membrane-Bound F_0F_1-ATP Synthase," *Nat. Struct. Mol. Biol.*, vol. 11, no. 2, pp. 135–141, 2004.

[109] S. Wang, R. Vafabakhsh, W. F. Borschel, T. Ha, and C. G. Nichols, "Structural Dynamics of Potassium-Channel Gating Revealed by Single-Molecule FRET," *Nat. Struct. Mol. Biol.*, vol. 23, no. 1, pp. 31–36, 2015.

[110] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless, "A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes," *Angew. Chem. Int. Edit.*, vol. 41, no. 14, pp. 2596–2599, 2002.

[111] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, and P. G. Schultz, "Addition of p-Azido-L-Phenylalanine to the Genetic Code of *Escherichia coli*," *J. Am. Chem. Soc.*, vol. 124, no. 31, pp. 9026–9027, 2002.

[112] J. W. Chin, A. B. Martin, D. S. King, L. Wang, and P. G. Schultz, "Addition of a Photocrosslinking Amino Acid to the Genetic Code of *Escherichia coli*," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 17, pp. 11020–11024, 2002.

[113] A. Chatterjee, H. Xiao, M. Bollong, H.-W. Ai, and P. G. Schultz, "Efficient Viral Delivery System for Unnatural Amino Acid Mutagenesis in Mammalian Cells," *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 29, pp. 11803–11808, 2013.

[114] H.-W. Ai, W. Shen, A. Sagi, P. R. Chen, and P. G. Schultz, "Probing Protein-Protein Interactions with a Genetically Encoded Photo-Crosslinking Amino Acid," *ChemBioChem*, vol. 12, no. 12, pp. 1854–1857, 2011.

[115] J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl, and F. L. van Delft, "Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells," *Angew. Chem. Int. Edit.*, vol. 49, no. 49, pp. 9422–9425, 2010.

[116] K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, and J. W. Chin, "Genetically Encoded Norbornene Directs Site-Specific Cellular Protein Labelling via a Rapid Bioorthogonal Reaction," *Nat. Chem.*, pp. 298–304, 2012.

[117] M. J. Hinner and K. Johnsson, "How to Obtain Labeled Proteins and What to Do with Them," *Curr. Opin. Biotech.*, vol. 21, no. 6, pp. 766–776, 2010.

[118] K. Lang and J. W. Chin, "Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins.," *Chem. Rev.*, vol. 114, no. 9, pp. 4764–4806, 2014.

[119] C. S. Heil, A. Rittner, B. Goebel, D. Beyer, and M. Grininger, "Site-Specific Labelling of Multidomain Proteins by Amber Codon Suppression," *Sci. Rep.*, vol. 8, no. 1, pp. 14864–14878, 2018.

[120] A. Hahn, S. Reschke, S. Leimkühler, and T. Risse, "Ketoxime Coupling of p-Acetylphenylalanine at Neutral pH for Site-Directed Spin Labeling of Human Sulfite Oxidase," *J. Phys. Chem. B*, vol. 118, no. 25, pp. 7077–7084, 2014.

[121] C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, and M. Mann, "Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions," *Science*, vol. 325, no. 5942, pp. 834–840, 2009.

[122] W. Wan, J. M. Tharp, and W. R. Liu, "Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme But an Outstanding Genetic Code Expansion Tool," *Biochim. Biophys. Acta*, vol. 1844, no. 6, pp. 1059–1070, 2014.

[123] R. Jiang and J. A. Krzycki, "PylSn and the Homologous N-Terminal Domain of Pyrrolysyl-tRNA Synthetase Bind the tRNA That is Essential for the Genetic Encoding of Pyrrolysine.," *J. Biol. Chem.*, vol. 287, no. 39, pp. 32738–32746, 2012.

[124] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, and S. Yokoyama, "Structural Basis for Orthogonal tRNA Specificities of Tyrosyl-tRNA Synthetases for Genetic Code Expansion," *Nat. Struct. Biol.*, vol. 10, no. 6, pp. 425–432, 2003.

[125] M. Pott, M. J. Schmidt, and D. Summerer, "Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids," *ACS Chem. Biol.*, vol. 9, no. 12, pp. 2815–2822, 2014.

[126] A. Royant and M. Noirclerc-Savoye, "Stabilizing Role of Glutamic Acid 222 in the Structure of Enhanced Green Fluorescent Protein," *Journal of Structural Biology*, vol. 174, no. 2, pp. 385–390, 2011.

[127] L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, and C. T. Walsh, "Characterization of Sfp, a *Bacillus subtilis* Pohospohpantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases," *Nat. Commun.*, vol. 37, pp. 1585–1595, 1998.

[128] K. K. M. Lee, N. A. Da Silva, and J. T. Kealey, "Determination of the Extent of Phosphopantetheinylation of Polyketide Synthases Expressed in *Escherichia coli* and *Saccharomyces cerevisiae*," *Anal. Biochem.*, vol. 394, pp. 75–80, 2009.

[129] A. S. Worthington, H. Rivera Jr., M. D. Alexander, and M. D. Burkart, "Mechanism-Based Protein Cross-Linking Probes to Investigate Carrier Protein-Mediated Biosynthesis," *ACS Chem. Biol.*, vol. 1, no. 11, pp. 687–691, 2006.

[130] C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. L. Markwick, J. Bruegger, G. Caldara-Festin, K. Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S.-C. Tsai, and M. D. Burkart, "Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis," *Nature*, vol. 505, no. 7483, pp. 427–431, 2014.

[131] F. Ishikawa, R. W. Haushalter, D. J. Lee, K. Finzel, and M. D. Burkart, "Sulfonyl 3-Alkynyl Pantetheinamides as Mechanism-Based Cross-Linkers of Acyl Carrier Protein Dehydratase," *J. Am. Chem. Soc.*, vol. 135, no. 24, pp. 8846–8849, 2013.

[132] V. S. Rangan, A. K. Joshi, and S. Smith, "Mapping the Functional Topology of the Animal Fatty Acid Synthase by Mutant Complementation in Vitro," *Biochemistry*, vol. 40, no. 36, pp. 10792–10799, 2001.

[133] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, and D. H. Sherman, "Structural Basis for Binding Specificity Between Subclasses of Modular Polyketide Synthase Docking Domains," *ACS Chem. Biol.*, vol. 4, no. 1, pp. 41–52, 2009.

[134] P. Pal, J. F. Lesoine, M. A. Lieb, L. Novotny, and P. A. Knauf, "A Novel Immobilization Method for Single Protein spFRET Studies," *Biophys. J.*, vol. 89, no. 2, pp. L11–L13, 2005.

[135] M. J. Schmidt, A. Fedoseev, D. Bücker, J. Borbas, C. Peter, M. Drescher, and D. Summerer, "EPR Distance Measurements in Native Proteins with Genetically Encoded Spin Labels," *ACS Chem. Biol.*, vol. 10, no. 12, pp. 2764–2771, 2015.

[136] P. Roser, M. J. Schmidt, M. Drescher, and D. Summerer, "Site-Directed Spin Labeling of Proteins for Distance Measurements in Vitro and in Cells," *Org. Biomol. Chem.*, vol. 14, no. 24, pp. 5468–5476, 2016.

[137] M. R. Fleissner, E. M. Brustad, T. Kálai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz, and W. L. Hubbell, "Site-Directed Spin Labeling of a Genetically Encoded Unnatural Amino Acid," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 12, pp. 5693–5693, 2010.

[138] T. Kálai, M. R. Fleissner, J. Jekő, W. L. Hubbell, and K. Hideg, "Synthesis of New Spin Labels for Cu-Free Click Conjugation," *Tetrahedron Lett.*, vol. 52, no. 21, pp. 2747–2749, 2011.

[139] A. Witkowski, A. K. Joshi, V. S. Rangan, A. M. Falick, H. E. Witkowska, and S. Smith, "Dibromopropanone Cross-linking of the Phosphopantetheine and Active-Site Cysteine Thiols of the Animal Fatty Acid Synthase Can Occur Both Inter- and Intrasubunit," *J. Biol. Chem.*, vol. 274, no. 17, pp. 11557–11563, 1999.

[140] Z. Ye, M. Bair, H. Desai, and G. J. Williams, "A Photocrosslinking Assay for Reporting Protein Interactions in Polyketide and Fatty Acid Synthases," *Mol. BioSyst.*, vol. 7, no. 11, pp. 3152–3155, 2011.

[141] Z. Ye and G. J. Williams, "Mapping a Ketosynthase: Acyl Carrier Protein Binding Interface via Unnatural Amino Acid-Mediated Photo-Cross-Linking," *Biochemistry*, vol. 53, no. 48, pp. 7494–7502, 2014.

[142] H.-S. Park, M. J. Hohn, T. Umehara, L.-T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, and D. Söll, "Expanding the Genetic Code of *Escherichia coli* with Phosphoserine," *Science*, vol. 333, no. 6046, pp. 1151–1154, 2011.

[143] C. H. Kim, J. Y. Axup, and P. G. Schultz, "Protein Conjugation with Genetically Encoded Unnatural Amino Acids," *Curr. Opin. Chem. Biol.*, vol. 17, no. 3, pp. 412–419, 2013.

[144] W. L. Hubbell, D. S. Cafiso, and C. Altenbach, "Identifying Conformational Changes with Site-Directed Spin Labeling," *Nat. Struct. Biol.*, vol. 7, no. 9, pp. 735–739, 2000.

[145] K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, and J. W. Chin, "Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions," *J. Am. Chem. Soc.*, vol. 134, no. 25, pp. 10317–10320, 2012.

[146] R. J. Blizzard, D. R. Backus, W. Brown, C. G. Bazewicz, Y. Li, and R. A. Mehl, "Ideal Bioorthogonal Reactions Using a Site-Specifically Encoded Tetrazine Amino Acid," *J. Am. Chem. Soc.*, vol. 137, no. 32, pp. 10044–10047, 2015.

[147] J. L. Seitchik, J. C. Peeler, M. T. Taylor, M. L. Blackman, T. W. Rhoads, R. B. Cooley, C. Refakis, J. M. Fox, and R. A. Mehl, "Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with trans-Cyclooctenes ," *J. Am. Chem. Soc.*, vol. 134, no. 6, pp. 2898–2901, 2012.

[148] T. Cha, A. Guo, and X.-Y. Zhu, "Enzymatic Activity on a Chip: The Critical Role of Protein Orientation," *Proteomics*, vol. 5, no. 2, pp. 416–419, 2005.

[149] T. Ha, "Single-Molecule Fluorescence Resonance Energy Transfer," *Methods*, vol. 25, pp. 78–86, 2001.

[150] A. Gust, A. Zander, A. Gietl, P. Holzmeister, S. Schulz, B. Lalkens, P. Tinnefeld, and D. Grohmann, "A Starting Point for Fluorescence-Based Single-Molecule Measurements in Biomolecular Research," *Molecules*, vol. 19, pp. 15824–15865, 2014.

[151] J. Mikkelsen, J. Knudsen, and S. Smith, "Novel Procedure for the Preparation and Characterization of Catalytically Active Fatty Acid Synthetase Immobilized on Sepharose Beads," *J. Biol. Chem.*, vol. 262, no. 4, pp. 1566–1569, 1987.

[152] R. E. Benesch and R. Benesch, "Enzymatic Removal of Oxygen for Polarography and Related Methods," *Science*, vol. 118, no. 3068, pp. 447–448, 1953.

[153] I. Rasnik, S. A. McKinney, and T. Ha, "Nonblinking and Long-Lasting Single-Molecule Fluorescence Imaging," *Nat. Methods*, vol. 3, no. 11, pp. 891–893, 2006.

[154] C. G. Hübner, A. Renn, I. Renge, and U. P. Wild, "Direct Observation of the Triplet Lifetime Quenching of Single Dye Molecules by Molecular Oxygen," *J. Chem. Phys.*, vol. 115, no. 21, pp. 9619–9622, 2001.

[155] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, "An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments," *Biophys. J.*, vol. 94, no. 5, pp. 1826–1835, 2008.

[156] B. Okumus, T. J. Wilson, D. M. J. Lilley, and T. Ha, "Vesicle Encapsulation Studies Reveal That Single Molecule Ribozyme Heterogeneities Are Intrinsic," *Biophys. J.*, vol. 87, no. 4, pp. 2798–2806, 2004.

[157] J. J. Benítez, A. M. Keller, P. Ochieng, L. A. Yatsunyk, D. L. Huffman, A. C. Rosenzweig, and P. Chen, "Probing Transient Copper Chaperone Wilson Disease Protein Interactions at the Single-Molecule Level with Nanovesicle Trapping," *J. Am. Chem. Soc.*, vol. 130, no. 8, pp. 2446–2447, 2008.

[158] E. Rhoades, E. Gussakovsky, and G. Haran, "Watching Proteins Fold One Molecule at a Time," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 6, pp. 3197–3202, 2003.

[159] J. T. Hammill, S. Miyake-Stoner, J. L. Hazen, J. C. Jackson, and R. A. Mehl, "Preparation of Site-Specifically Labeled Fluorinated Proteins for^{19}F-NMR Structural Characterization," *Nat. Protoc.*, vol. 2, no. 10, pp. 2601–2607, 2007.

[160] D. Xie, Z. Shao, J. Achkar, W. Zha, J. W. Frost, and H. Zhao, "Microbial Synthesis of Triacetic Acid Lactone," *Biotechnol. Bioeng.*, vol. 93, no. 4, pp. 727–736, 2006.

[161] S. Wolter, A. Löschberger, T. Holm, S. Aufmkolk, M.-C. Dabauvalle, S. van de Linde, and M. Sauer, "rapidSTORM: Accurate, Fast Open-Source Software for Localization Microscopy," *Nat. Methods*, vol. 9, no. 11, pp. 1040–1041, 2012.

[162] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proc. Second Int. Conf. Knowl. Discov. Data Min.*, pp. 226–231, July 1999.

[163] S. Malkusch and M. Heilemann, "Lama: The LocAlization Microscopy Analyzer (Documentation)," *Single Molecule Biophysics*, pp. 1–42, 2016.

[164] B. Yang, Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He, and M.-Q. Dong, "Identification of Cross-Linked Peptides From Complex Samples," *Nat. Methods*, vol. 9, no. 9, pp. 904–906, 2012.

[165] A. Deiters and P. G. Schultz, "In Vivo Incorporation of an Alkyne into Proteins in *Escherichia coli*," *Bioorg. Med. Chem. Lett.*, vol. 15, no. 5, pp. 1521–1524, 2005.

[166] L. Wang, Z. Zhang, A. Brock, and P. G. Schultz, "Addition of the Keto Functional Group to the Genetic Code of *Escherichia coli*," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 1, pp. 57–61, 2002.

[167] D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon, and P. G. Schultz, "An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity," *Biochemistry*, vol. 50, no. 11, pp. 1894–1900, 2011.

[168] K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, and P. G. Schultz, "A Genetically Encoded Infrared Probe," *J. Am. Chem. Soc.*, vol. 128, no. 43, pp. 13984–13985, 2006.

[169] D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, and J. W. Chin, "Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins *via* Pyrrolysyl-tRNA Synthetase/tRNA$_{CUA}$ Pair and Click Chemistry," *J. Am. Chem. Soc.*, vol. 131, no. 25, pp. 8720–8721, 2009.

[170] E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, and T. Carell, "A Genetically Encoded Norbornene Amino Acid for the Mild and Selective Modification of Proteins in a Copper-Free Click Reaction," *Angew. Chem. Int. Edit.*, vol. 51, no. 18, pp. 4466–4469, 2012.

[171] S. Schneider, M. J. Gattner, M. Vrabel, V. Flügel, V. López-Carrillo, S. Prill, and T. Carell, "Structural Insights into Incorporation of Norbornene Amino Acids for Click Modification of Proteins," *ChemBioChem*, vol. 14, no. 16, pp. 2114–2118, 2013.

[172] A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. M. Verkade, M. Wießler, C. Schultz, J. C. M. van Hest, F. L. van Delft, and E. A. Lemke, "Genetic Encoding of a Bicyclo[6.1.0]nonyne-Charged Amino Acid Enables Fast Cellular Protein Imaging by Metal-Free Ligation," *ChemBioChem*, vol. 13, pp. 2094–2099, 2012.

[173] T. Umehara, J. Kim, S. Lee, L.-T. Guo, D. Söll, and H.-S. Park, "N-Acetyl Lysyl-tRNA Synthetases Evolved by a CcdB-Based Selection Possess N-Acetyl Lysine Specificity in Vitro and in Vivo," *FEBS Lett.*, vol. 586, no. 6, pp. 729–733, 2012.

[174] Y.-S. Wang, X. Fang, A. L. Wallace, B. Wu, and W. R. Liu, "A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant with a Broad Substrate Spectrum," *J. Am. Chem. Soc.*, vol. 134, no. 6, pp. 2950–2953, 2012.

Appendix

List of Suppression Plasmids

Table S1: List of suppression plasmids constructed and utilized in this thesis and ncAAs tested with these plasmids. Plasmids^{D286R} contain an additional D286R mutation in the *Mj*TyrRS.[124]

#	Plasmid Name	aaRS Variant	ncAAs
1	pAC^{U}_AzPhe	*Mj*TyrRS[111] Y32T, E107N, D158P, I159L, L162Q	AzPhe
2	pAC^{U}_AzPhe^{D286R}		
3	pAC^{E}_AzPhe^{D286R}		
4	pAC^{U}_PrPhe	*Mj*TyrRS[165] Y32A, E107P, L110F, D158A, L162A	PrPhe
5	pAC^{U}_PrPhe^{D286R}		
6	pAC^{E}_PrPhe^{D286R}		
7	pAC^{U}_AcPhe	*Mj*TyrRS[166] Y32L, D158G, I159C, L162R	AcPhe
8	pAC^{E}_AcPhe		
9	pAC^{U}_AcPhe^{D286R}		
10	pAC^{E}_AcPhe^{D286R}		
11	pAC^{U}_BzPhe	*Mj*TyrRS[112] Y32G, E107S, D158T, I159S	BzPhe
12	pAC^{E}_BzPhe		
13	pAC^{U}_TetPhe^{D286R}	*Mj*TyrRS[147] Y32E, L65A, F108P, Q109S, D158G, L162G	TetPhe
14	pAC^{E}_TetPhe^{D286R}		
15	pAC^{U}_CNF	*Mj*TyrRS[167,168] Y32L, F108W, Q109M, D158G, I159A	AzPhe, PrPhe, TetPhe
16	pAC^{U}_PyLys	*Mm*PylRS[116,169]	NorLys1, NorLys2, PrLys, AcLys
17	pAC^{E}_PyLys		

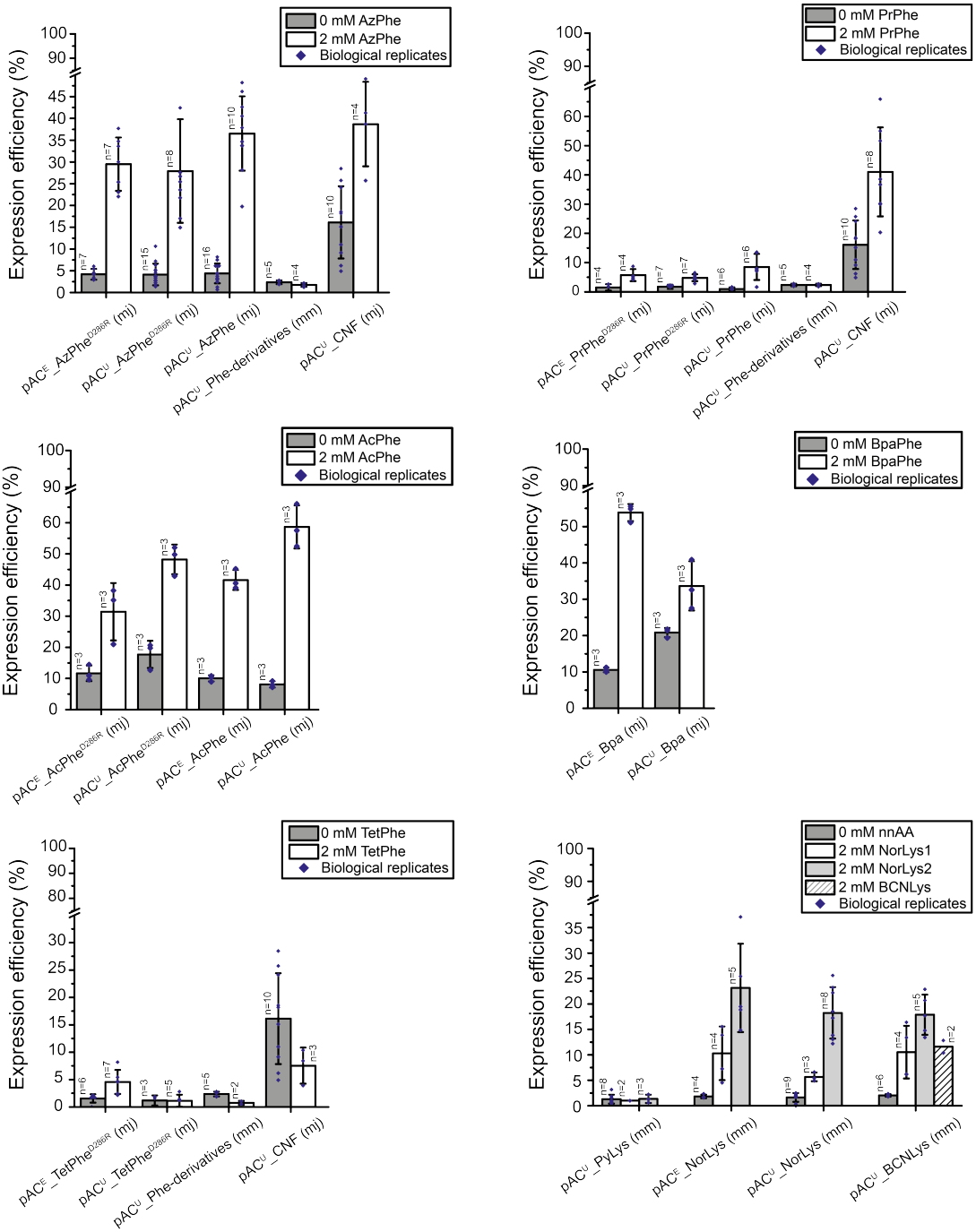
18 19	pAC^{U}_NorLys pAC^{E}_NorLys	*Mm*PylRS[170,171] Y306G, Y384F, I405R	NorLys1, NorLys2
20	pAC^{U}_BCN	*Mm*PylRS[172] Y306A, Y348A	BCNLys, NorLys1, NorLys2
21 22	pAC^{U}_AcLys pAC^{E}_AcLys	*Mm*PylRS[173] L301M, Y306L, L309A, C348F	AcLys
23	pAC^{U}_Phe-derivatives	*Mm*PylRS[174] N346A, C348A	AzPhe, PrPhe, TetPhe
24 25 26 27	pAC^{U}_PyLys pAC^{E}_PyLys pAC^{U}_PyLys pAC^{E}_PyLys	*Mb*PylRS[116,169]	ncAA

List of Primers for the Cloning of pAC Vectors

Primer	Name	Length	Orientation	T_m (°C)	Sequence (5'-3')
CH02	pCDF1b_for	24	forward	55.7	CTGAAACCTCAGGCATTTGAGAAG
CH03	pCDF1b_rev	26	reverse	56.4	ATTTCCTAATGCAGGAGTCGCATAAG
CH04	pCDF_tac_for	44	forward	55.9	CCTGCATTAGGAAATTTGACAATTAATCATCGGCTCGTATAATG
CH05	tac_Nde1-rev	27	reverse	55.7	ATGCTATGGTCCTTGTTGGTCAATTGC
CH06	tac_mjTyrS_for	56	forward	54.8	CAATTGACCAACAAGGACCATAGCATATGGACGAATTTGAAATGATAAAGAGAAAC
CH11	mjtRNA_pCDF1b_rev	35	reverse	53.8	TGCCTGAGGTTTCAGCAAATTCGACCCTGAGCTGC
CH19	mjtRNA_rrnBterm_rev	46	reverse	54	GCGGAATTATTTGATCGGCGAGGTCTCATGAGCGGATACATATTTG
CH32	rrnBterm_for	19	forward	55.4	GGTGCGGCACACTACGTAC
CH33	mjTyrS_rrnBterm_rev	50	reverse	53.8	GTACGTAGTGTGCCGCACCTTATAATCTCTTTCTAATTGGCTCTAAAATC
CH34	rrnBterm_mjtRNA_for	46	forward	53.8	CTCGCCGATCAAATAATTCCGCTAATTCCGCTTCGCAACATGTGAG
CH51	tac_mbPylRS	50	forward	56.2	CAAGGACCATAGCATATGGATAAAAAACCATTAGATGTTTTAATATCTGC
CH52	mbPylRS_rrnBterm_rev	47	reverse	56.7	GTAGTGTGCCGCACCTCATAGATTGGTTGAAATCCCATTATAGTAAG
CH53	tac_notag-mmPylRS_for	47	forward	56.7	CAAGGACCATAGCATATGGATAAAAAACCACTAAACACTCTGATATC
CH56	PyltRNA_wobble_rev	28	reverse	57.0	TTCGATCTACATGATCAGGTTTCCAATG
CH57	PyltRNA_wobble_for	37	forward	56.7	ATCATGTAGATCGAACGGACTCTAAATCCGTTCAGCC
CH96	ara_mmPylRS_forward	49	forward	56.7	GGAGGAATTAGATCTATGGATAAAAAACCACTAACACTCTGATATCTG
CH97	mmPylRS_rrnB_rev	59	reverse	57.0	CTCAATGATGATGATGATGGTCGACTTATTACAGGTTGGTAGAAATCCCGTTATAG
CH98	glnsprom_mmPylRS_for	61	forward	58.4	CGTTGTTTACGCTTTGAGGAATCCCATATGGATAAAAAACCACTAACACTCTGATATCTG
CH99	mmPylRS_glnsterm_reverse	43	reverse	57.0	CGTTTGAAACTGCAGTTACAGGTTGGTAGAAATCCCGTTATAG
CH100	pEVOL_insert_forward	25	forward	54.4	TAAGTCGACCATCATCATCATC
CH101	pEVOL_insert_reverse	26	reverse	56.4	ATGGGATTCCTCAAAGCGTAAACAAC
CH102	ara_mjTyrRS_forward	45	forward	54.8	GGAGGAATTAGATCTATGGACGAATTTGAAATGATAAAGAGAAAC
CH103	mjTyrRS_rrnB_rev	66	reverse	55.1	CTCAATGATGATGATGATGGTCGACttaTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTA
CH104	glnsprom_mjTyrRS_for	57	forward	54.8	CGTTGTTTACGCTTTGAGGAATCCCATATGGACGAATTTGAAATGATAAAGAGAAAC
CH105	mjTyrRS_glnsterm_rev	50	reverse	55.1	CGTTTGAAACTGCAGTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTA
CH106	ara_mbPylRS_for	49	forward	54.8	GGAGGAATTAGATCTATGGATAAAAAACCATTAGATGTTTTAATATCTG
CH107	mbPylRS_rrnBterm_rev	59	reverse	54.1	CTCAATGATGATGATGATGGTCGACTTATCATAGATTGGTTGAAATCCCATTATAG
CH108	glnsprom_mbPylRS_for	61	forward	54.8	CGTTGTTTACGCTTTGAGGAATCCCATATGGATAAAAAACCATTAGATGTTTTAATATCTG
CH109	mbPylRS_glnsterm_rev	43	reverse	54.1	CGTTTGAAACTGCAGTCATAGATTGGTTGAAATCCCATTATAG
CH115	MegaPrimer_mmpEvol_for	28	forward	54.1	GGAGGAATTAGATCTATGGATAAAAAAC
CH116	MegaPrimer_mjpEvol_for	28	forward	57.0	GGAGGAATTAGATCTATGGACGAATTTG
CH117	MegaPrimer_mmpEvol_rev	25	reverse	56.0	CGTTTGAAACTGCAGTTACAGGTTG
CH118	MegaPrimer_mjpEvol_rev	29	reverse	55.9	CGTTTGAAACTGCAGTTATAATCTCTTTC
CH119	MegaPrimer_mbpEvol_rev	26	reverse	56.4	CGTTTGAAACTGCAGTCATAGATTGG
CH123	pEvol_backbone_for	25	forward	56.0	CTGCAGTTTCAAACGCTAATTGCC
CH124	pEvol_backbone_rev	24	reverse	55.7	AGATCTAATTCCTCCTGTTAGCCC
CH130	XbaI_pET22b_rev	36	reverse	56.3	CAAAATTATTTCTAGCCAACTCAGCTTCCTTTCGGG
CH148	XbaI_seq_for	33	forward	54.5	CTA GAA ATA ATT TTG TTT AAC TTT AAG AAG GAG

Primer	Name	Length	Orientation	T_m (°C)	Sequence (5'-3')
CH149	XbaI_seq_rev	22	reverse	54.8	AGGGGAATTGTTATCCGCTCAC
CH158	mbPylRS_rrnBterm_rev	47	reverse	55.4	GTAGTGTGCCGCACCTTATAGATTGGTTGAAATCCCATTATAGTAAG
AR81	pET22b_Xba1_for	31	forward	56.4	GGATAACAATTCCCCTCTAGAAATAATTTTG
AR365	mmPylS_rrnB_rev	47	reverse	57	GTACGTAGTGTGCCGCACCTTACAGGTTGGTAGAAATCCCGTTATAG
AR366	proK_rev	21	reverse	56.3	AATGCGGGGCGCATCTTACTG
AR367	proK_mmtRNA_for	47	forward	56.4	CAGTAAGATGCGCCCCGCATTGGAAACCTGATCATGTAGATCGAATG
AR368	mmtRNA_terK_rev	35	reverse	57.9	GCTTTTCGAATTTGGTGGCGGAAACCCCGGGAATC
AR369	terK_for	24	forward	55.7	CCAAATTCGAAAAGCCTGCTCAAC

Supplementary Figures



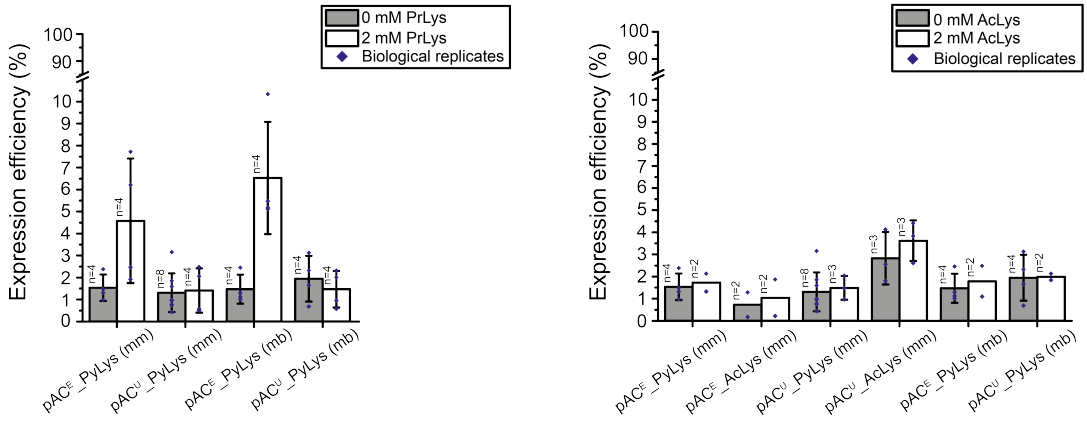


Figure S1: Screening for the most efficient suppression vector system to introduce selected ncAAs with the reporter assay. The ACP-GFP construct with amber codon at site Leu54 was used for screening. Expression efficiency was read out by GFP fluorescence of 2 mL cell cultures and compared to the wild-type reference. For incorporation, 2 mM ncAAs were supplemented to the medium. Cultures lacking ncAAs were taken as negative control to determine background signal. The averages of biological replicates are plotted together with standard deviation and the distribution of individual values is indicated as dots. Technical errors were below 10 %.

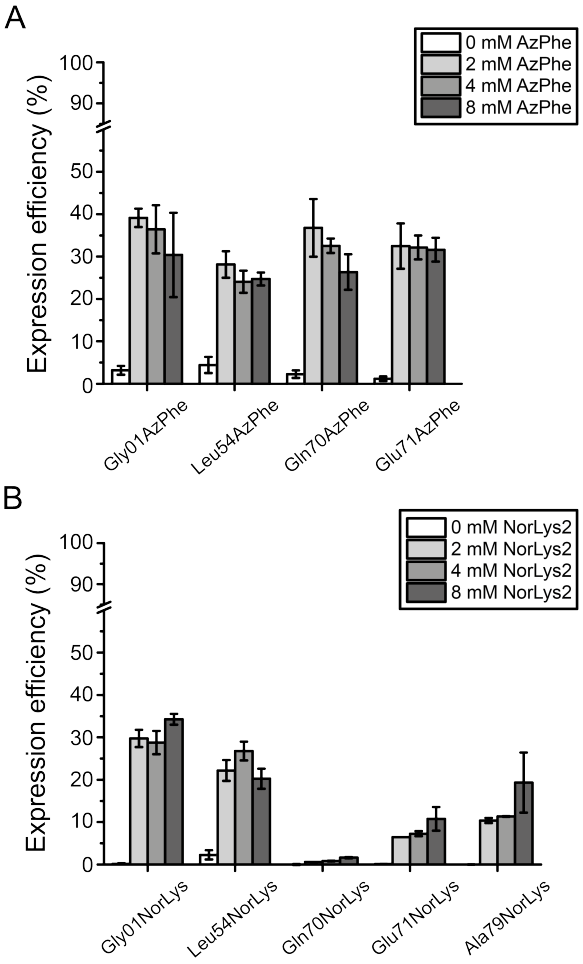


Figure S2: Testing the effect of different ncAA concentrations. Incorporation at different sites in the ACP-GFP construct was investigated with the reporter assay in 2 mL scale. Concentrations of 0, 2, 4 and 8 mM ncAA were supplemented to the medium. Expression efficiency was read out by GFP fluorescence and compared to the wild-type reference. Error bars reflect the standard deviation of technical triplicates. A) Incorporation of AzPhe. B) Incorporation of NorLys2.

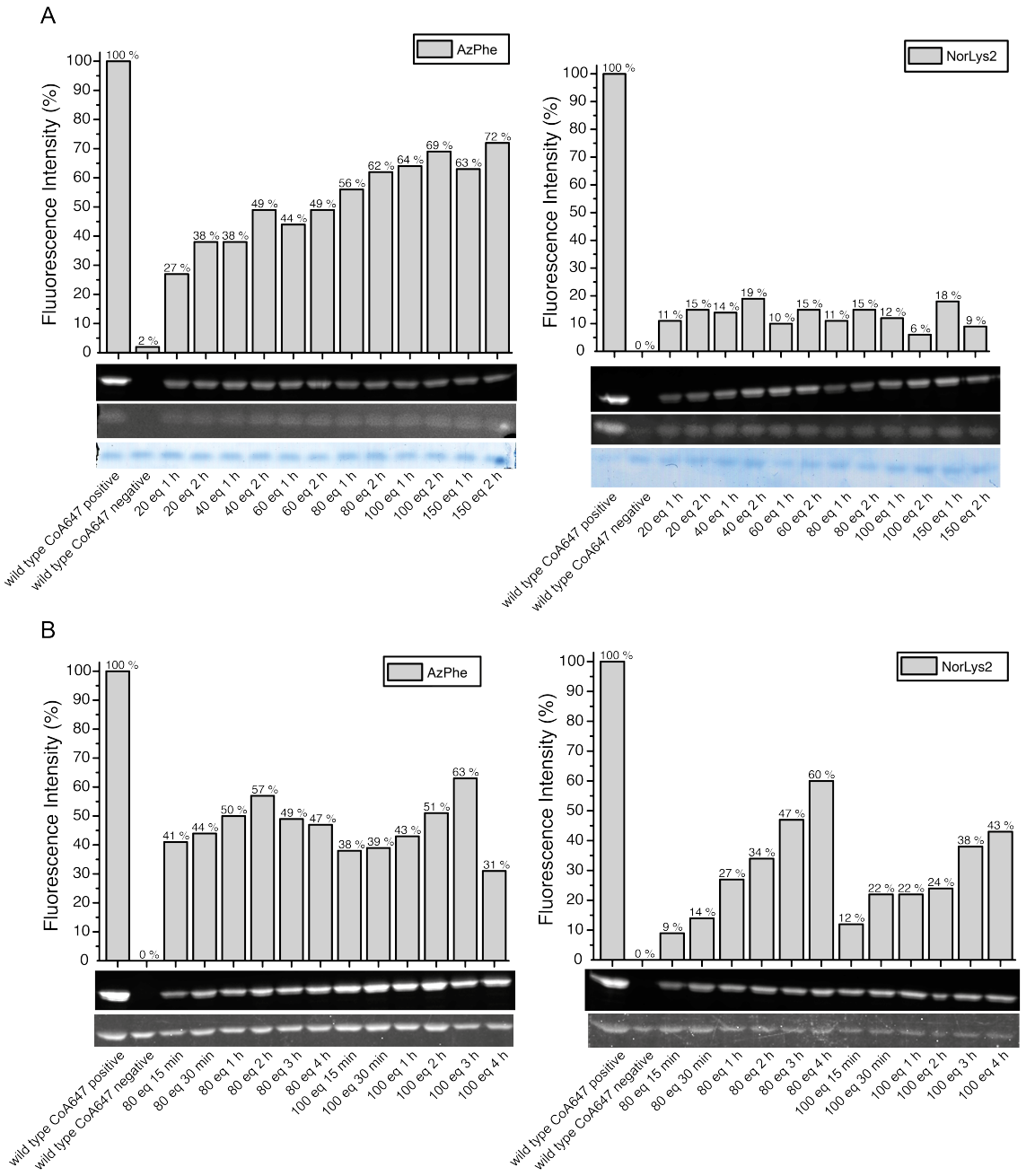


Figure S3: Optimizing the click reaction conditions for fluorescent labeling of ACP-GFP mutants. AzPhe mutants were labeled with BCN-POE3-NH-DY649P1, NorLys2 mutants were labeled with 6-methyl-tetrazine-ATTO-647N and the wild-type ACP-GFP was modified enzymatically with a fluorescent CoA 647-label by Sfp. DOL determined by relative in-gel fluorescence intensities at wavelength 650 nm compared to wild-type ACP-GFP reference. All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Syprored- or Coomassie-stained gel. A) DOL of AzPhe mutants (left panel) and NorLys2 mutants (right panel) monitored in dependence of different fluorophore equiv. after labeling reactions for 1 h and 2 h. B) Time dependency of the DOL of AzPhe mutants (left panel) and NorLys2 mutants (right panel) monitored for labeling reactions with 80 and 100 equiv. of fluorophore. SDS-PAGE performed on NuPAGE Bis-Tris 4-12 %.

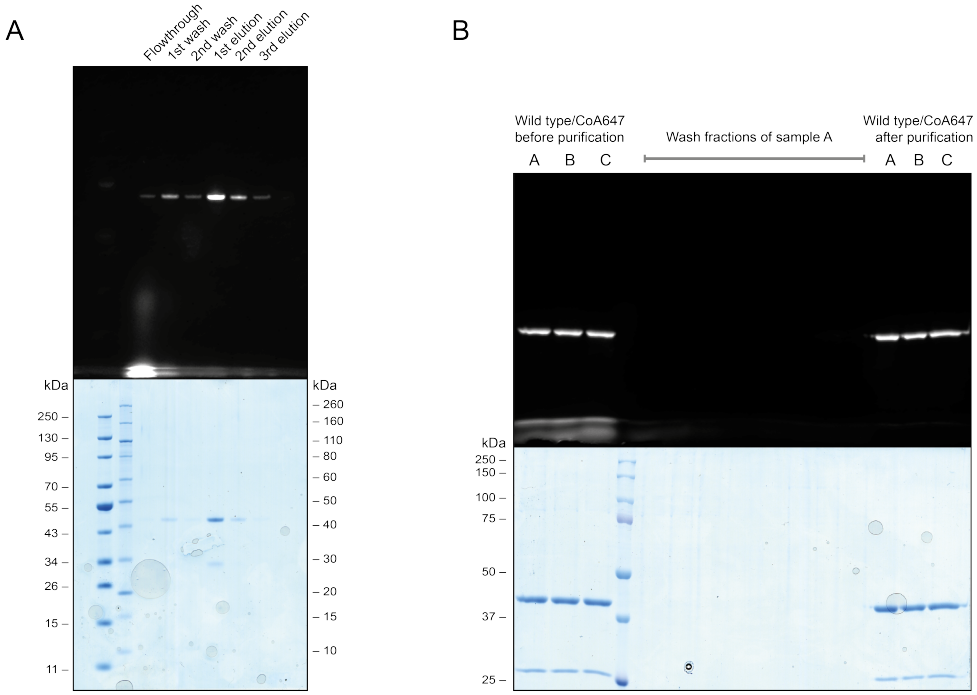


Figure S4: Removal of excess free fluorophore. A) SDS-PAGE gel (NuPAGE Bis-Tris 4-12 %) of wild-type ACP-GFP CoA 647 conjugate purified over Ni-NTA magnetic beads. Non-covalently bound dye molecules are washed off. The elution fractions show only protein dye conjugate and no free fluorophore. B) SDS-PAGE gel (12 % polyacrylamide) of wild-type ACP-GFP CoA 647 conjugate purified over Amicon® Ultra Centrifugal Filters (Merck Millipore). Wild-type ACP-GFP was enzymatically modified with a fluorescent CoA 647-label by Sfp (additional band at 26 kDa) in three parallel reactions. For MS analysis, the protein buffer was exchanged by 0.1 M ammonium acetate buffer in an Amicon® Ultra Centrifugal Filters (Merck Millipore). After eight centrifugation steps, non-covalently bound dye molecules are washed off. In-gel fluorescence intensities were detected at wavelength 650 nm.

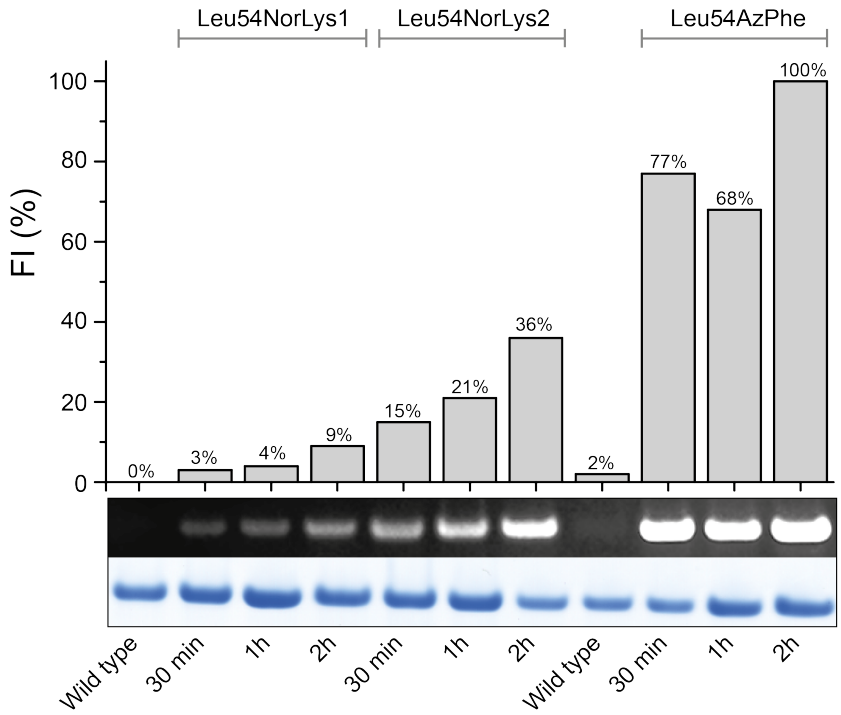


Figure S5: Comparing the labeling efficiency of ACP-GFP mutants with different ncAAs incorporated at position Leu54. AzPhe mutants were labeled with 80 equiv. BCN-POE3-NH-DY649P1, NorLys1 and NorLys2 mutants were both labeled with 80 equiv. 6-methyl-tetrazine-ATTO-647N. Wild-type ACP-GFP was treated with the respective fluorophores in a negative control to estimated the amount of unspecific fluorophore binding to the protein. DOL after 30 min, 1 h and 2 h labeling reaction was determined by relative in-gel fluorescence intensities at wavelength 650 nm compared to the highest DOL of the AzPhe mutant (SDS-PAGE on NuPAGE Bis-Tris 4-12 %). All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Coomassie-stained gel.

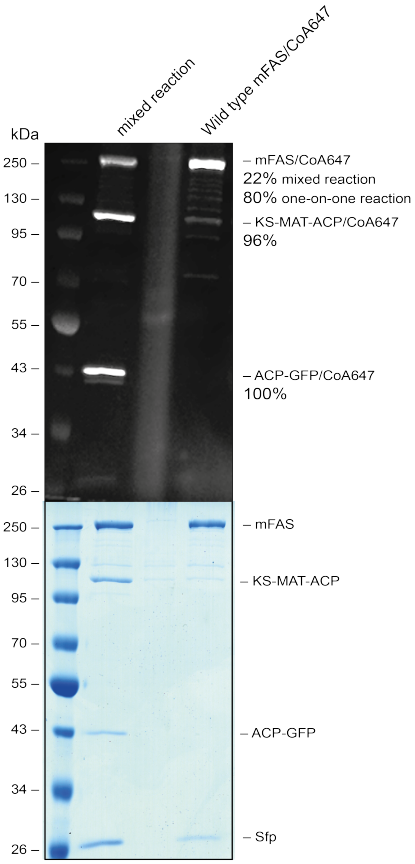


Figure S6: *In vitro* phosphopantetheinylation of wild-type mFAS with CoA 647-label by Sfp. In-gel fluorescence intensities of different *in vitro* phosphopantetheinylation reactions with CoA 647-label were compared (SDS-PAGE on NuPAGE Bis-Tris 4-12 %). In a one-on-one reaction, 1 equiv. wild-type mFAS was treated with 1 equiv. Sfp and 5 equiv. CoA 647 fluorohpore. In a mixed reaction, equimolar amounts (1.85 µM each, 5.55 µM total protein) of wild-type mFAS, wild-type ACP-GFP and of a KS-MAT protein construct with fused ACP domain, were treated with 1 equiv. Sfp and 5 equiv. CoA 647 fluorophore. *In vitro* phosphopantetheinylation of wild-type mFAS seems to be less efficient than the reaction of ACP-GFP or KS-MAT-ACP. In-gel fluorescence intensities were detected at wavelength 650 nm, corrected by the quantum efficiency of the DY647P1 fluorophore and correlated to the protein bands of the Coomassie-stained gel.

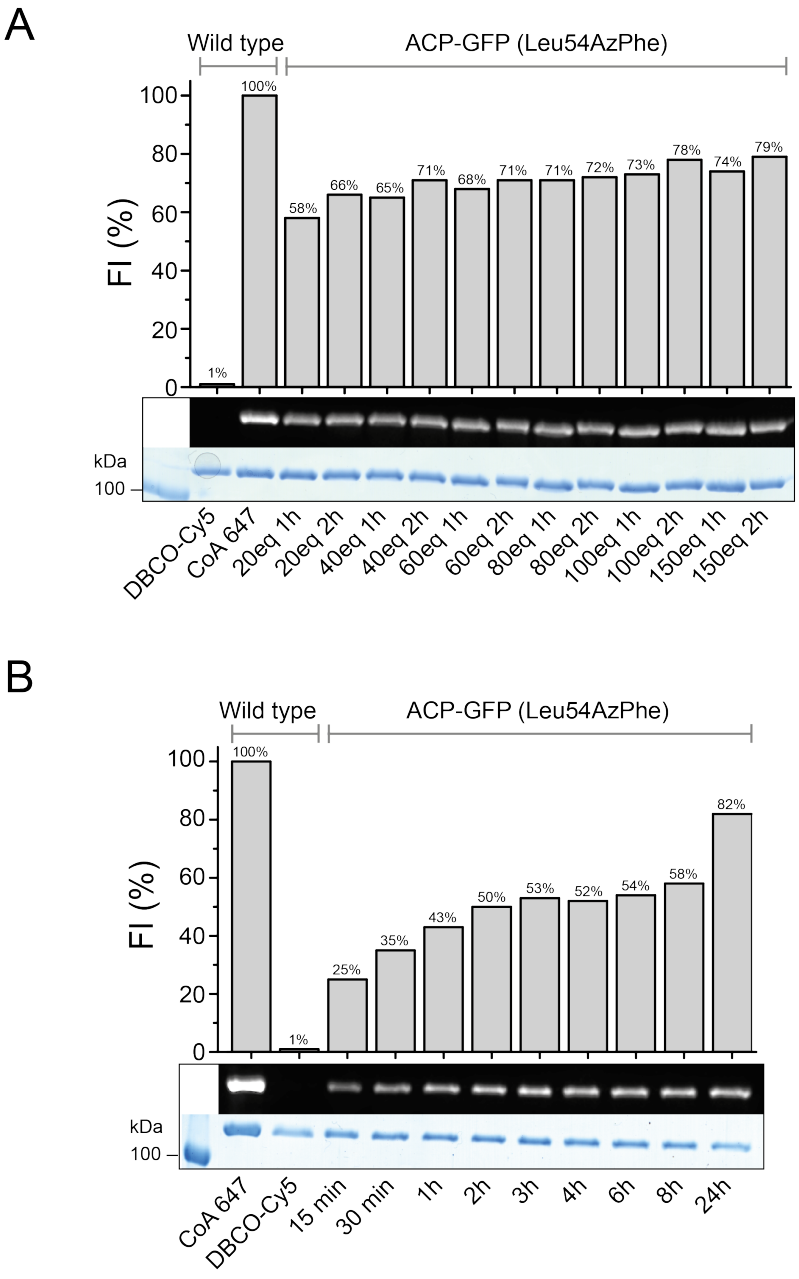


Figure S7: Optimizing the click reaction conditions for fluorescent labeling of ACP-GFP mutants in SPAAC with DBCO-functional dyes. ACP-GFP Leu54AzPhe mutants were labeled with DBCO-sulfo-Cy5 fluorophore. Wild-type ACP-GFP was modified enzymatically with a fluorescent CoA 647-label by Sfp. DOL determined by relative in-gel fluorescence intensities at wavelength 650 nm compared to wild-type ACP-GFP reference. All fluorescence intensities were corrected by the quantum efficiencies of the respective fluorophores and correlated to the protein bands of the Coomassie-stained gel. A) DOL of AzPhe mutants monitored in dependence of different fluorophore equiv. after labeling reactions for 1 h and 2 h. B) Time dependency of the DOL of AzPhe mutants monitored for labeling reactions with 100 equiv. of fluorophore. SDS-PAGE was performed on NuPAGE Bis-Tris 4-12 % gels. SDS-PAGE (12 % polyacrylamide)

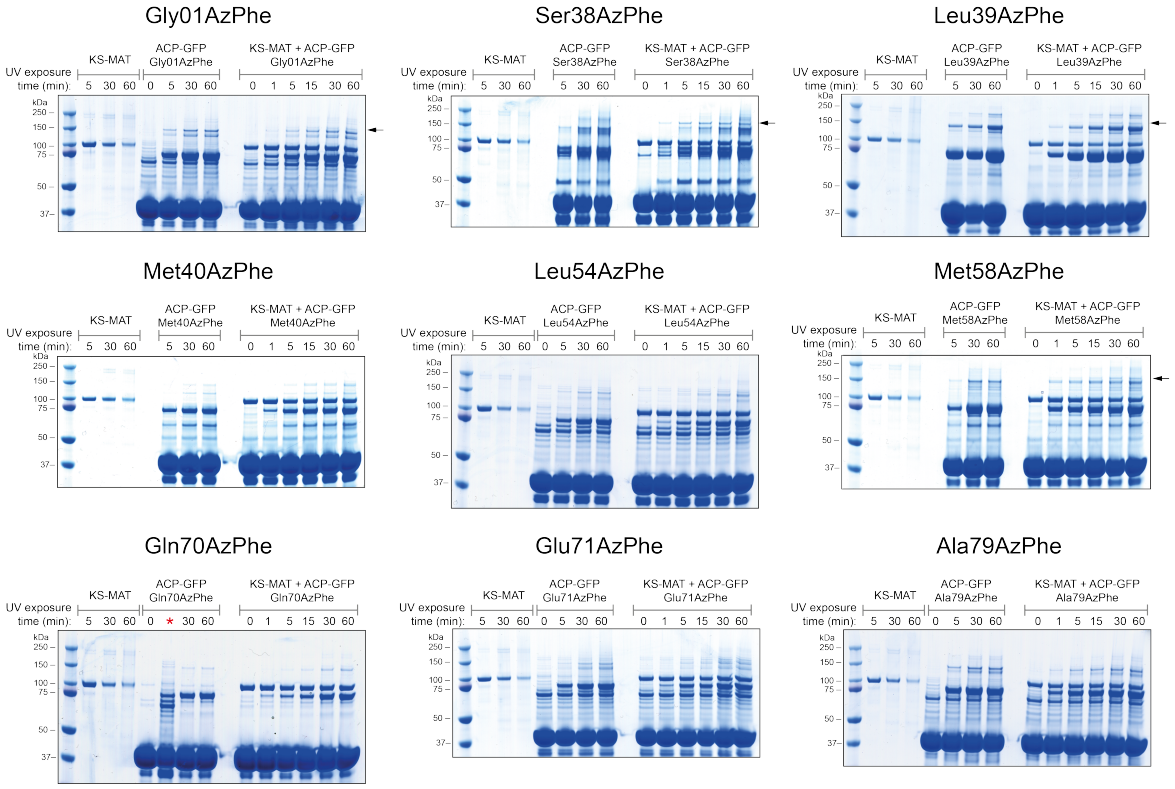


Figure S8: Position dependency of the intermolecular photocrosslinking reaction between KS-MAT and ACP-GFP AzPhe mutants. In every photocrosslinking assay, control experiments were performed, exposing individual KS-MAT and the ACP-GFP mutant proteins to UV light. Photocrosslinking was monitored for UV exposure times up to 1 h and samples were taken at different time points for SDS-PAGE (NuPAGE Bis-Tris 4-12 %). Specific crosslink bands are observed at approximately 150 kDa, indicated by an arrow. The red star in the photocrosslinking assay for Gln70AzPhe indicates a wrong protein sample.

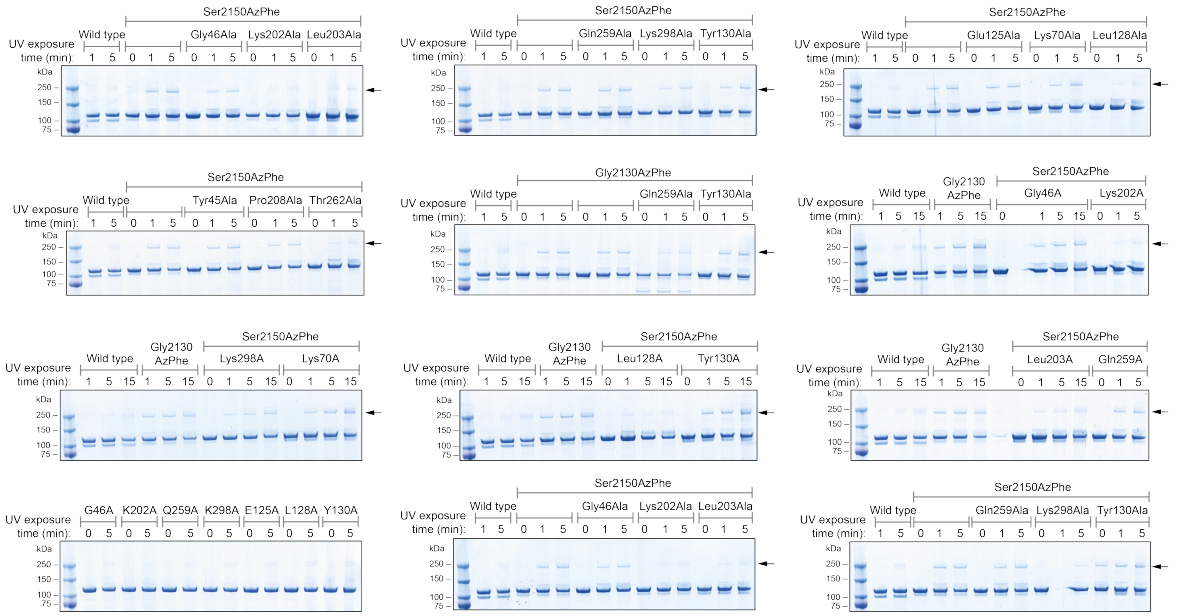


Figure S9: Position dependency of the intramolecular photocrosslinking reaction of KS-MAT-ACP AzPhe mutants. In every photocrosslinking assay a control experiment was performed, exposing wild-type KS-MAT-ACP to UV light. Photocrosslinking was monitored for UV exposure times up to 5 min and samples were taken at different time points for SDS-PAGE (NuPAGE Bis-Tris 4-12 %). Specific crosslink bands are observed between 150 and 250 kDa, indicated by an arrow.

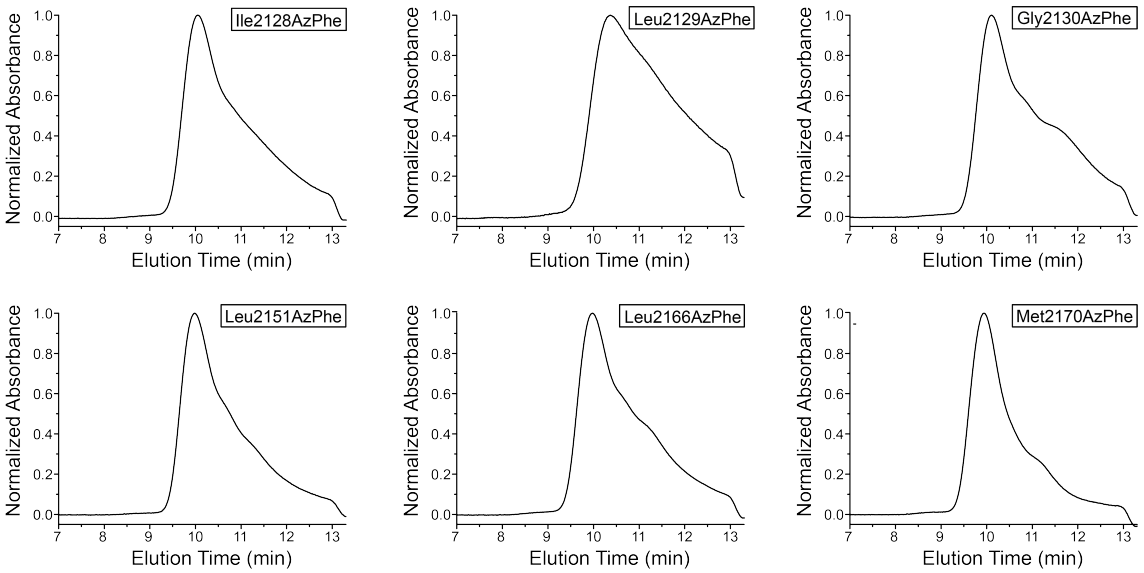


Figure S10: HPLC-SEC spectra of KS-MAT-ACP AzPhe mutants. Dimeric proteins elute at approximately 10 min elution time. Monomeric proteins elute at approximately 12 min elution time.

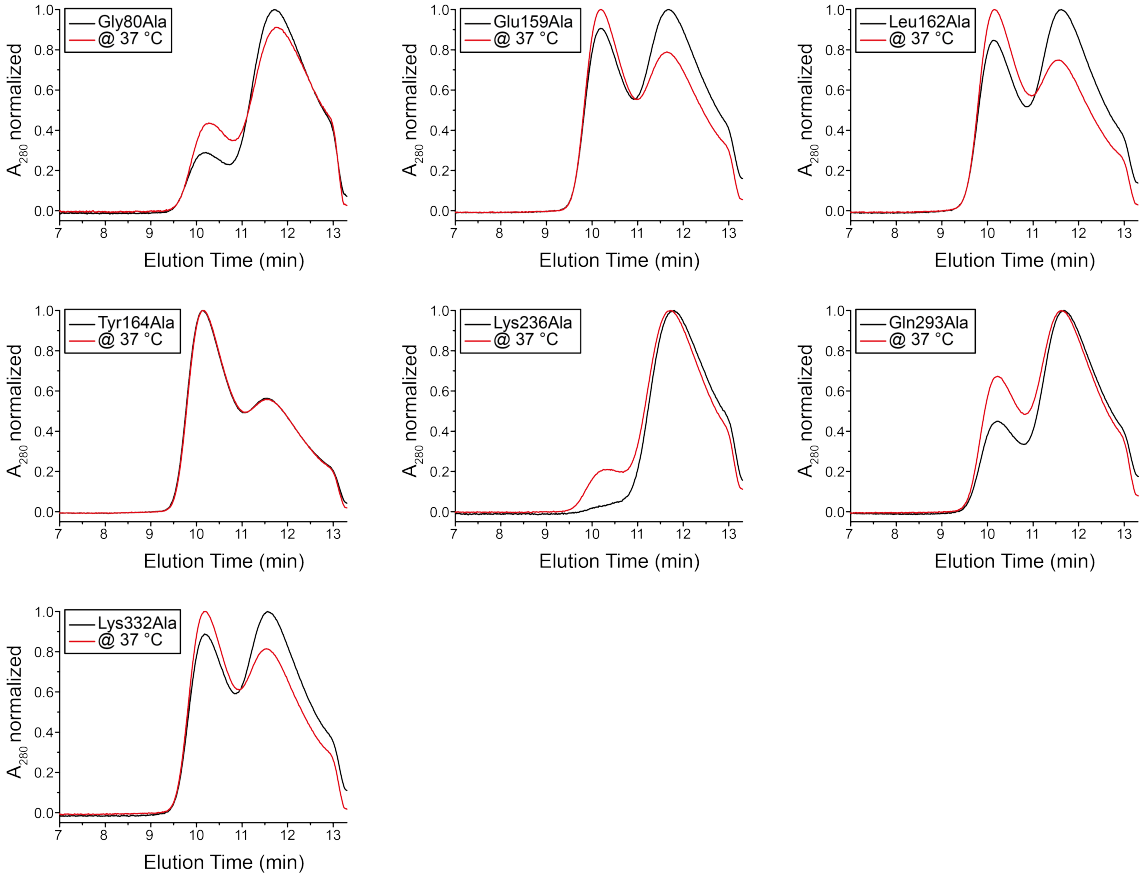


Figure S11: HPLC-SEC spectra of the alanine scanning mutagensis on the KS surface of KS-MAT-ACP. Dimeric proteins elute at approximately 10 min elution time. Monomeric proteins elute at approximately 12 min elution time.

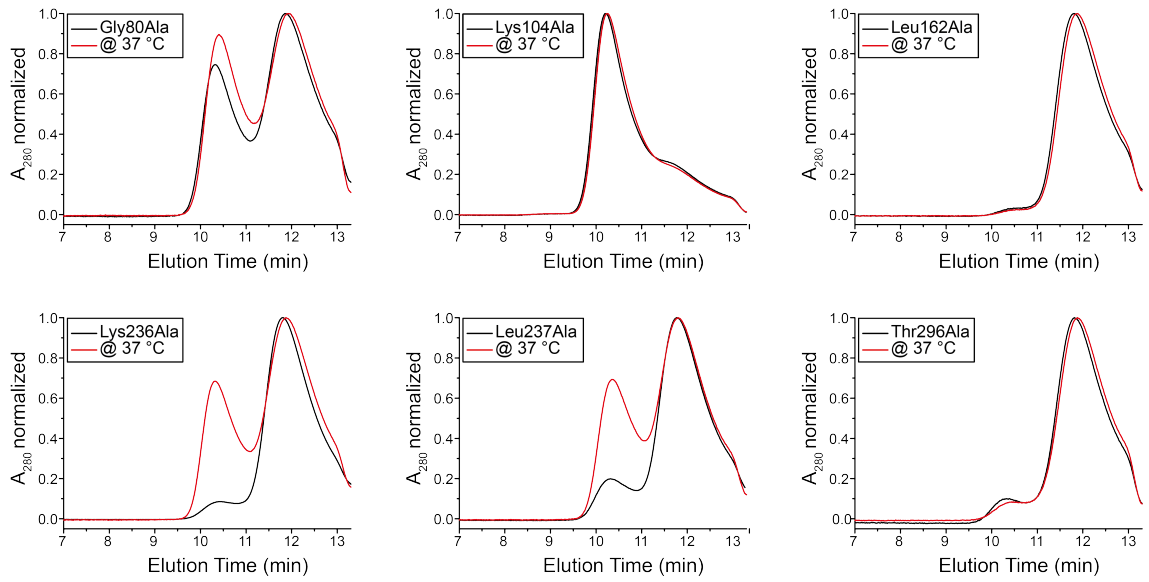


Figure S12: HPLC-SEC spectra of the alanine scanning mutagenesis on the KS surface of in the KS-MAT-ACP Ser2150AzPhe mutant. Dimeric proteins elute at approximately 10 min elution time. Monomeric proteins elute at approximately 12 min elution time.

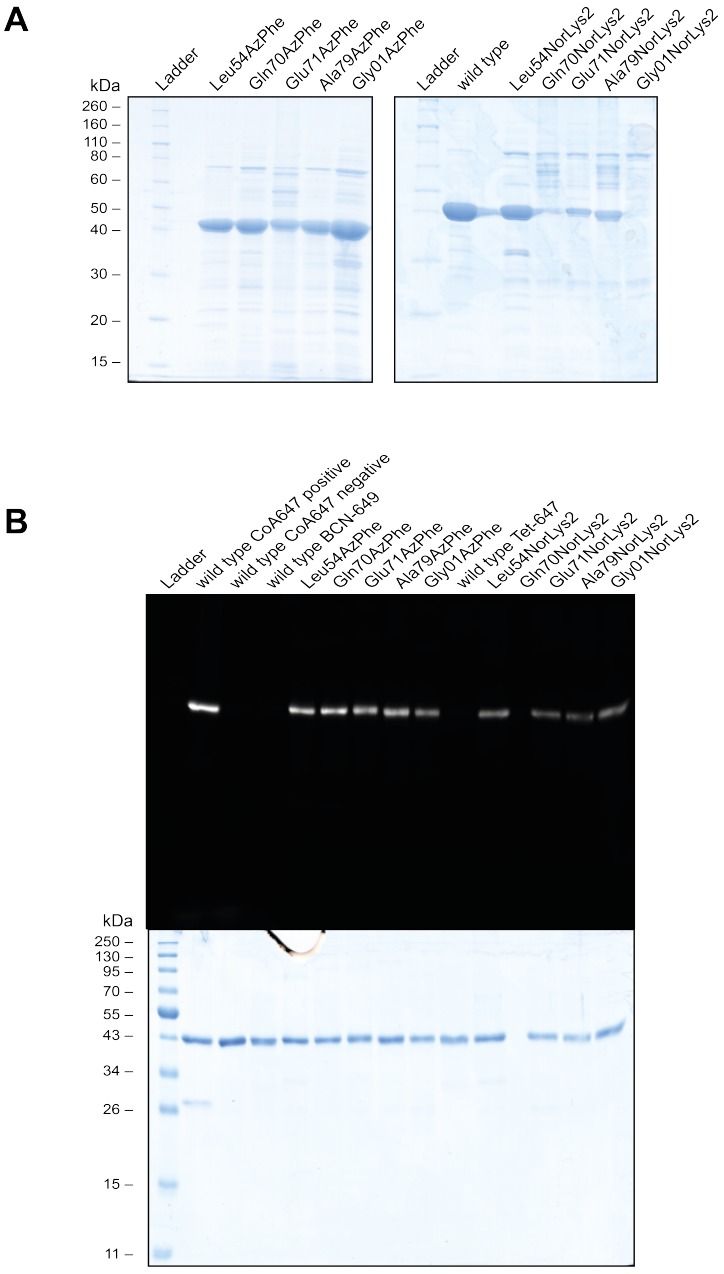
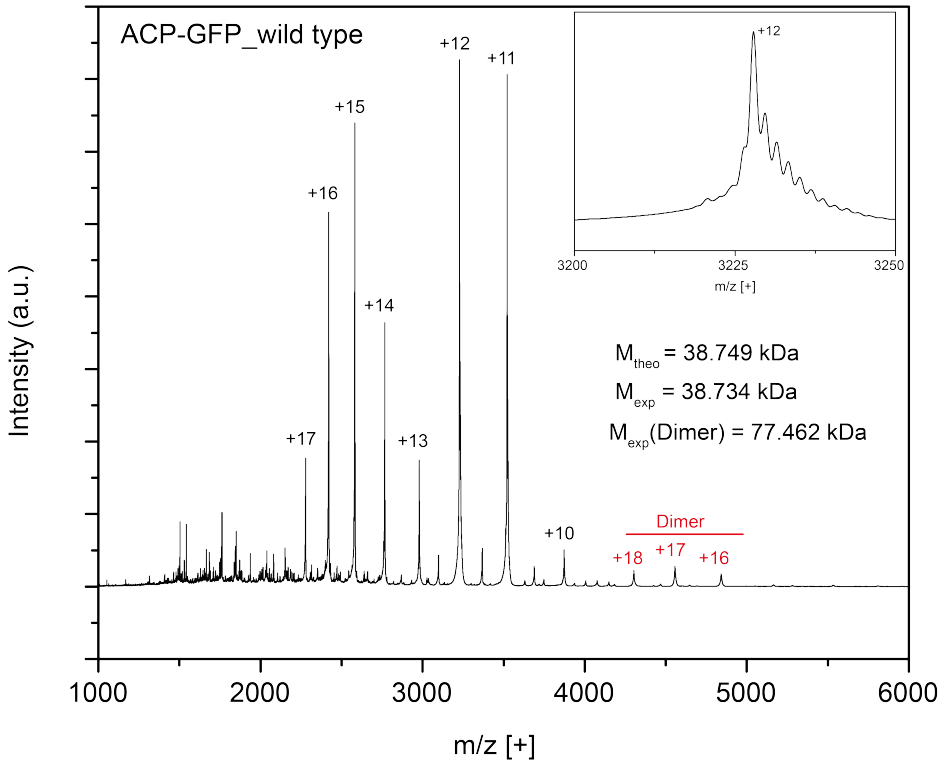
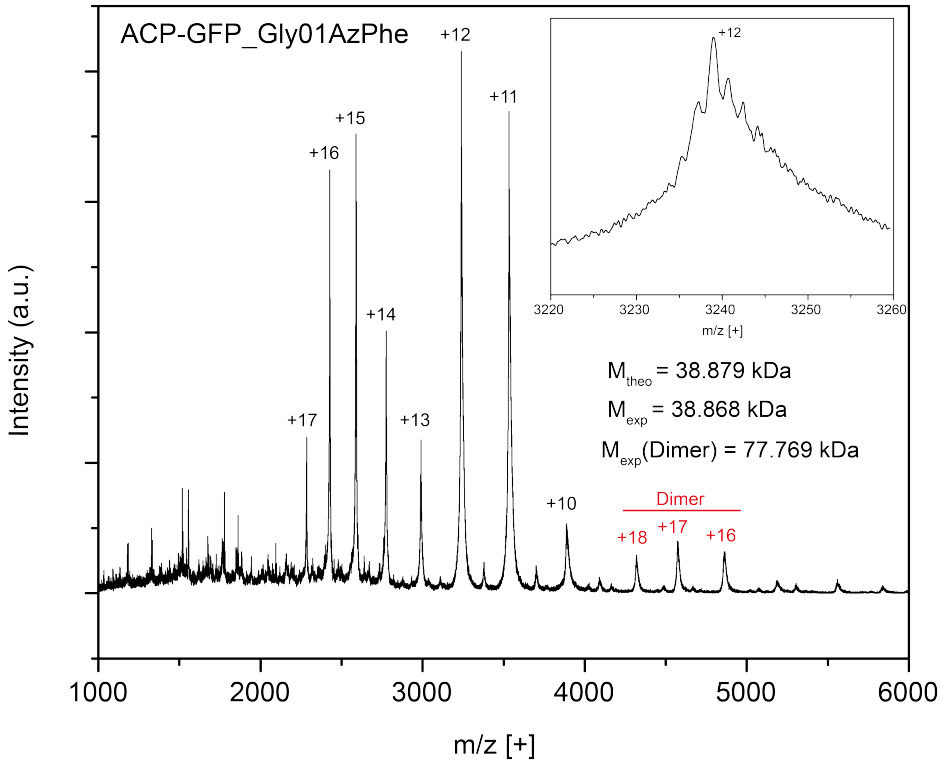


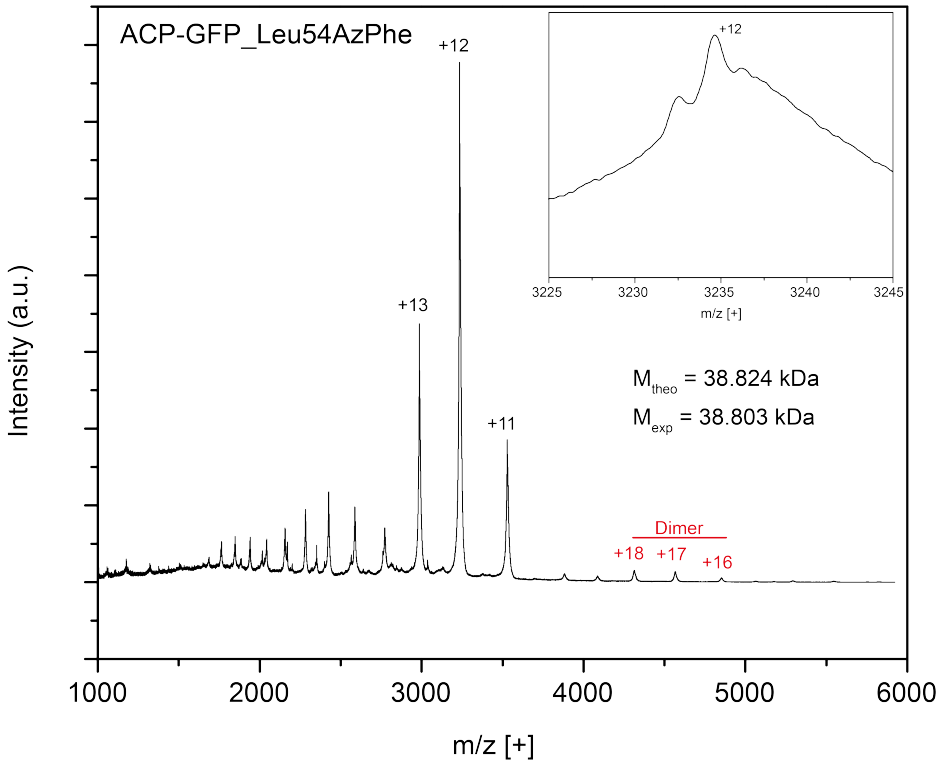
Figure S13: Original gels for assembled gels in the thesis. A) Original SDS-PAGE (NuPAGE Bis-Tris 4-12 %) gel of ACP-GFP mutants purified by Ni-chelating chromatography. Referring to Fig. 3.6 in section 5.1.5. B) Original fluorescent gel and Coomassie-stained gel (NuPAGE Bis-Tris 4-12 %) of fluorescently labelled ACP-GFP mutants. Referring to Fig. 3.8 in section 3.1.4.

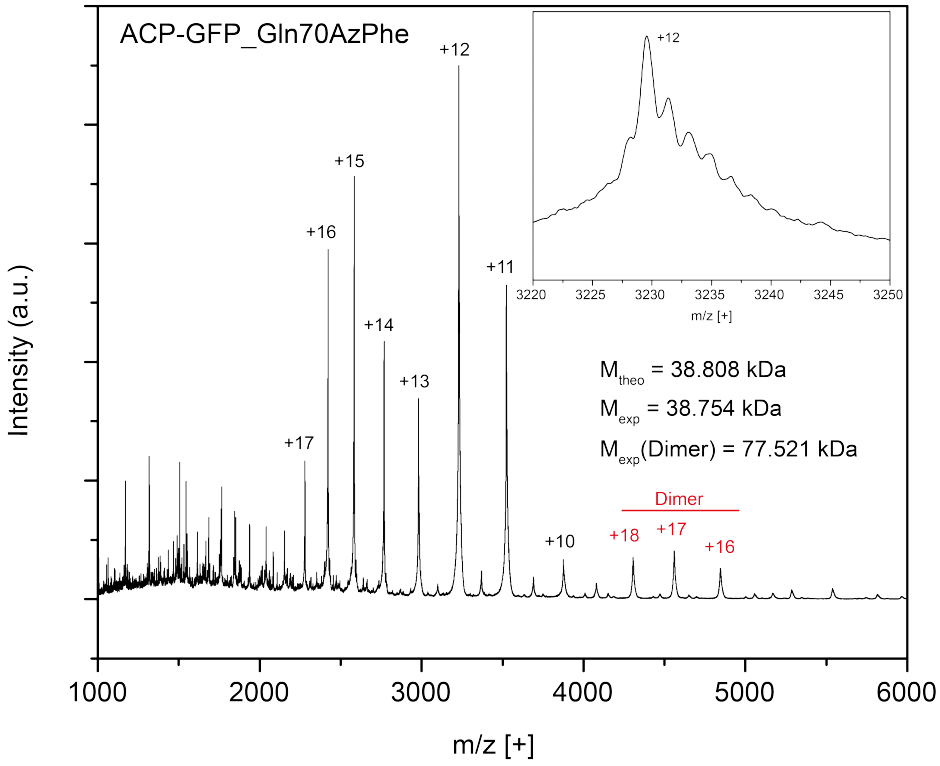
Supplementary Data

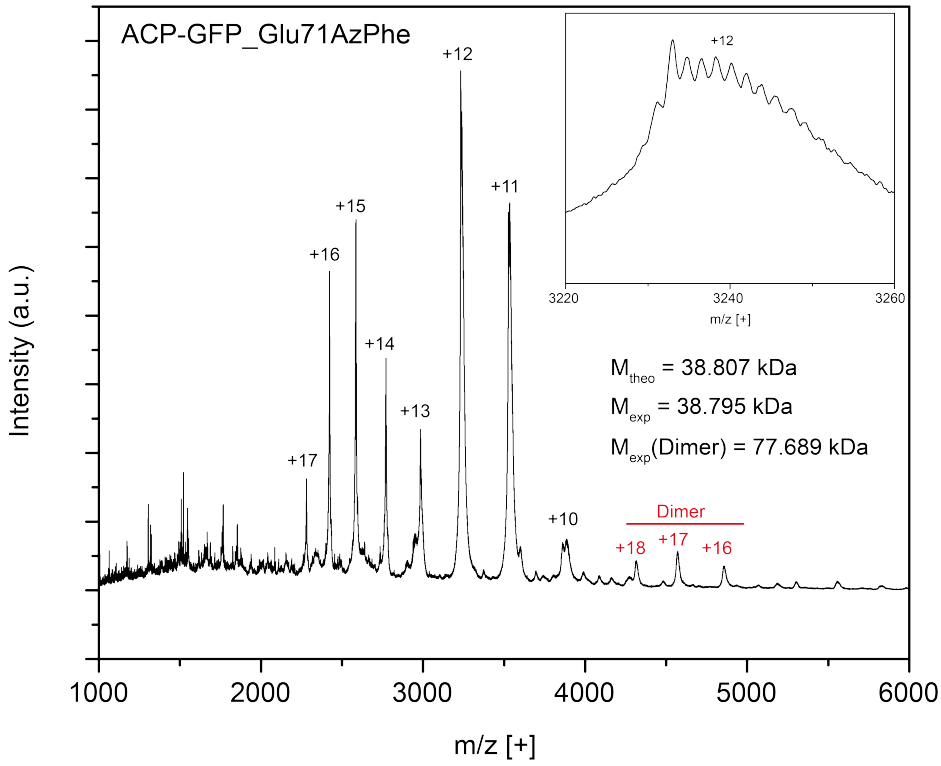
MS analysis of purified ACP-GFP constructs was performed on a nanoESI (Synapt G2-S). For each construct the highest peak of the +12 ionized species was picked to calculate the protein mass. Within accuracy of the mass spectrometry device, found protein masses were typically 8-21 Da smaller than theoretical protein masses (except for ACP-GFP_Gln70AzPhe which showed a mass difference of 54 Da). The experimental mass differences between wild type and ACP-GFP with incorporated ncAAs were determined with a deviation of ±7 Da (except for ACP-GFP_Gln70AzPhe which showed a deviation of -39 Da).

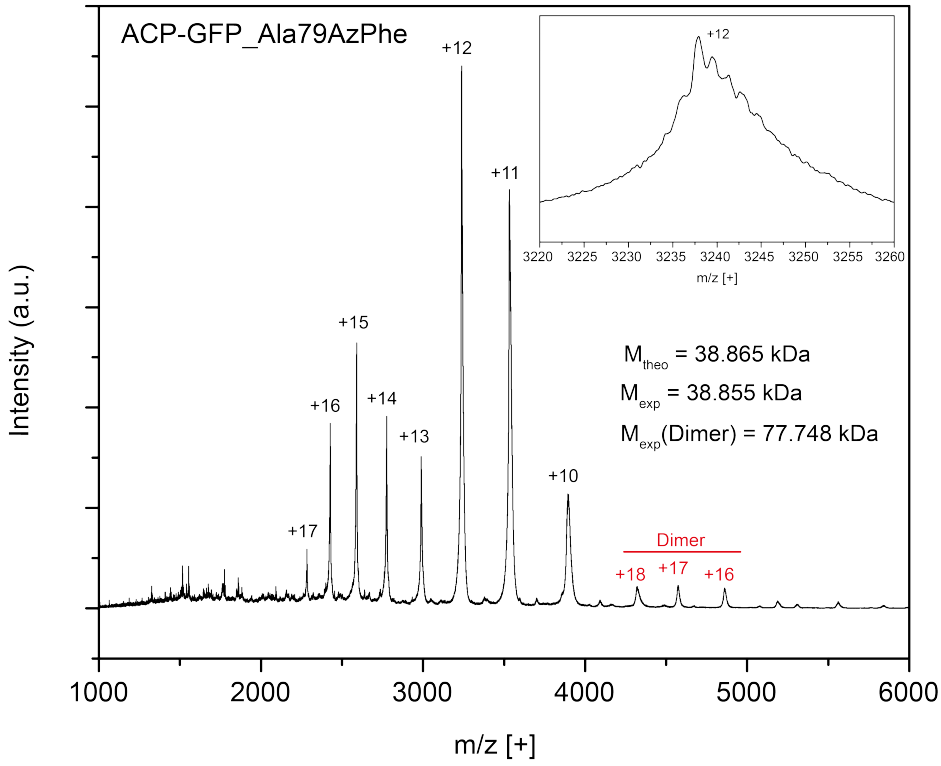

ACP-GFP_wild type
+12
+11
+15
+16
+14
+17
+13
+10
Dimer
+18
+17
+16
+12
3200
3225
3250
m/z [+]
M theo = 38.749 kDa
M exp = 38.734 kDa
M exp(Dimer) = 77.462 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]

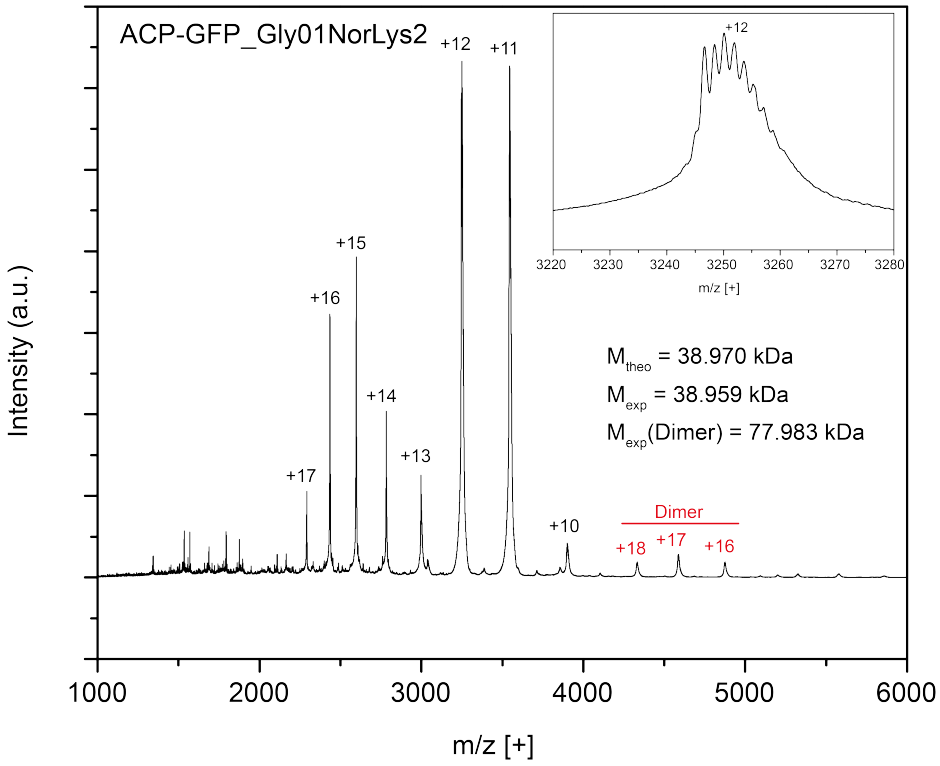

ACP-GFP_Gly01AzPhe
+12
+11
+15
+16
+14
+17
+13
+10
Dimer
+18
+17
+16
+12
3220
3230
3240
3250
3260
m/z [+]
M theo = 38.879 kDa
M exp = 38.868 kDa
M exp(Dimer) = 77.769 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]

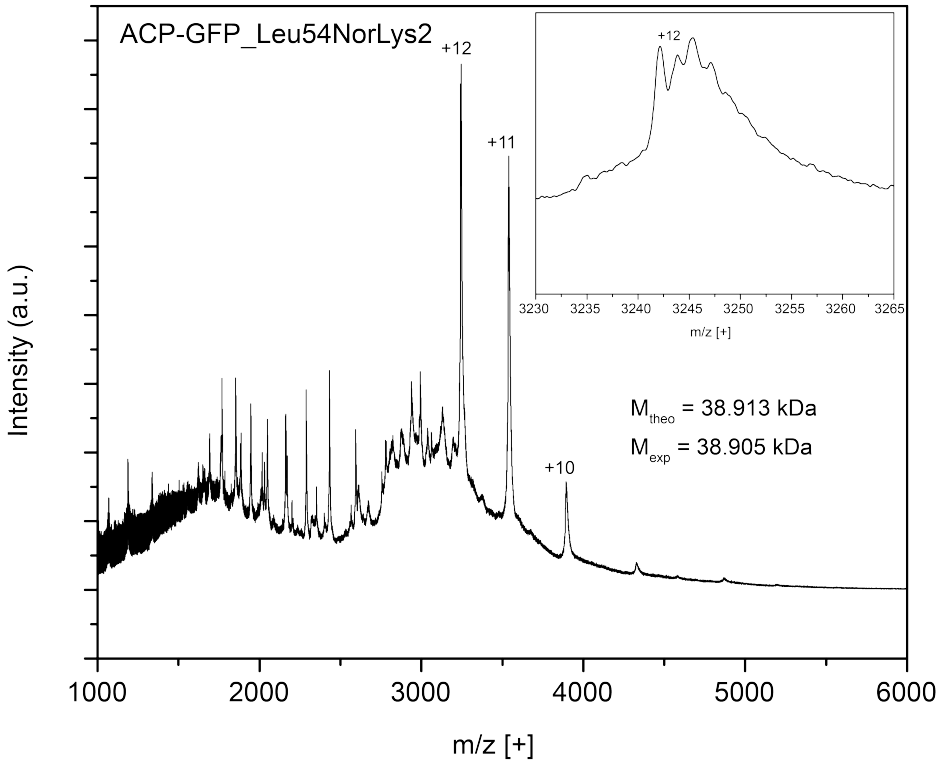

ACP-GFP_Leu54AzPhe
+12
+13
+11
Dimer
+18 +17 +16
M_theo = 38.824 kDa
M_exp = 38.803 kDa
Intensity (a.u.)
m/z [+]
1000 2000 3000 4000 5000 6000
3225 3230 3235 3240 3245

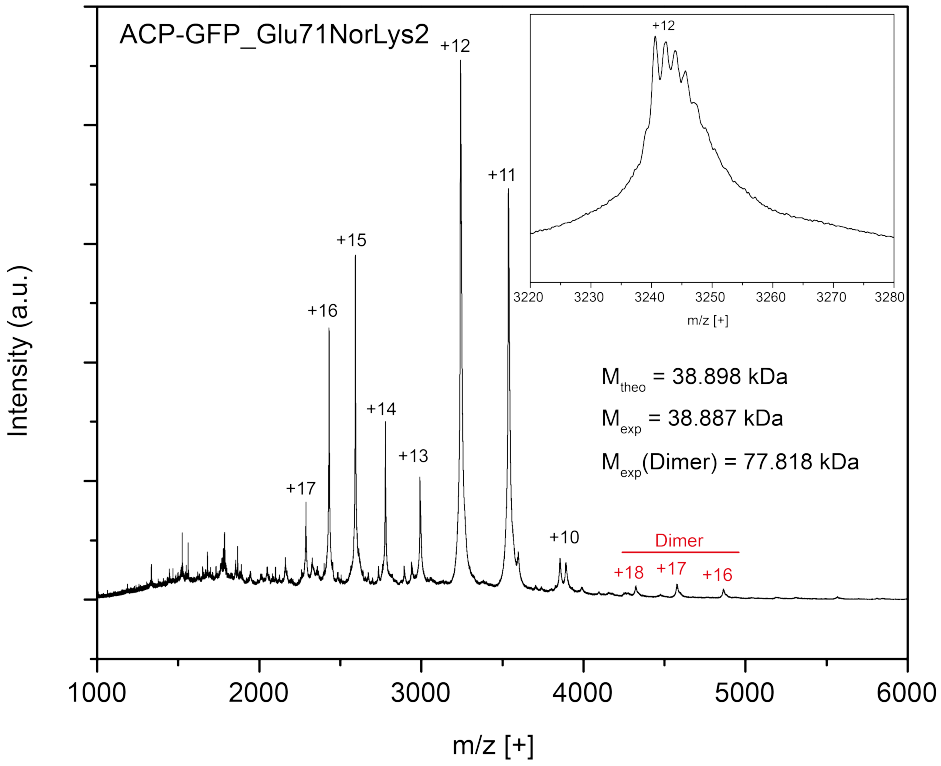

ACP-GFP_Gln70AzPhe
+12
+15
+16
+14
+13
+17
+11
+10
Dimer
+18 +17 +16
M_theo = 38.808 kDa
M_exp = 38.754 kDa
M_exp(Dimer) = 77.521 kDa
Intensity (a.u.)
m/z [+]
1000 2000 3000 4000 5000 6000
3220 3225 3230 3235 3240 3245 3250

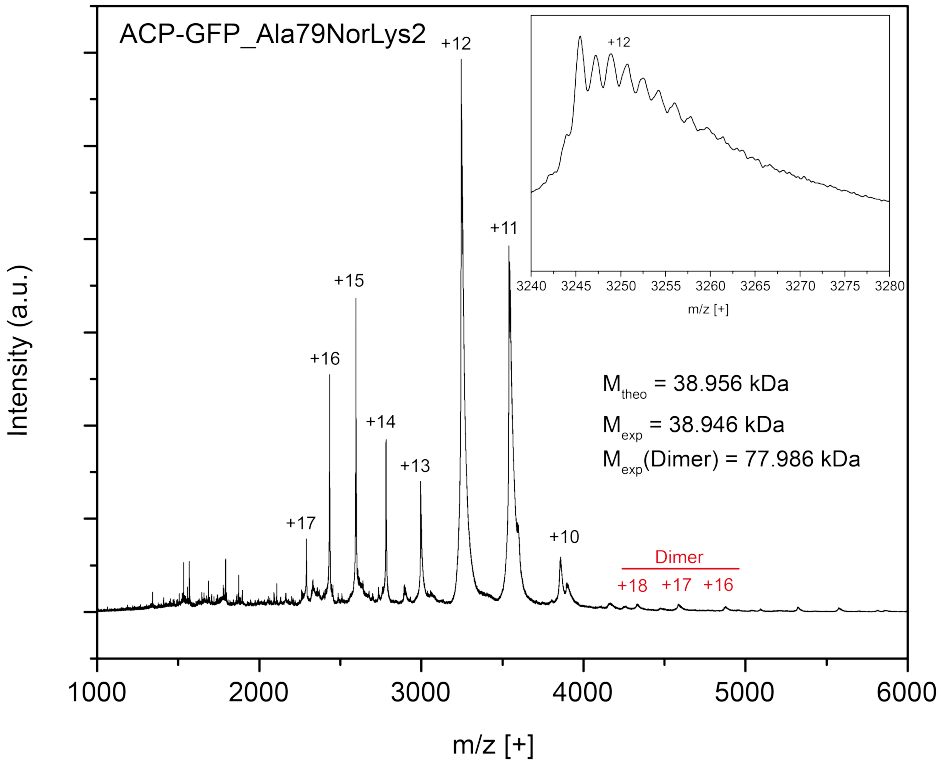

ACP-GFP_Glu71AzPhe
+12
+11
+15
+16
+14
+13
+17
+10
Dimer
+18
+17
+16
3220
3240
3260
m/z [+]
M theo = 38.807 kDa
M exp = 38.795 kDa
M exp (Dimer) = 77.689 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]


ACP-GFP_Ala79AzPhe
+12
+11
+15
+16
+14
+13
+17
+10
Dimer
+18
+17
+16
3220
3225
3230
3235
3240
3245
3250
3255
3260
m/z [+]
M theo = 38.865 kDa
M exp = 38.855 kDa
M exp (Dimer) = 77.748 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]


ACP-GFP_Gly01NorLys2
+12
+11
+15
+16
+14
+13
+17
+10
Dimer
+18
+17
+16
+12
3220
3230
3240
3250
3260
3270
3280
m/z [+]
M theo = 38.970 kDa
M exp = 38.959 kDa
M exp (Dimer) = 77.983 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]


ACP-GFP_Leu54NorLys2
+12
+11
+10
+12
3230
3235
3240
3245
3250
3255
3260
3265
m/z [+]
M theo = 38.913 kDa
M exp = 38.905 kDa
Intensity (a.u.)
1000
2000
3000
4000
5000
6000
m/z [+]


ACP-GFP_Glu71NorLys2
+12
+11
+15
+16
+14
+13
+17
+10
Dimer
+18 +17 +16
M_theo = 38.898 kDa
M_exp = 38.887 kDa
M_exp(Dimer) = 77.818 kDa
Intensity (a.u.)
m/z [+]
1000 2000 3000 4000 5000 6000
3220 3230 3240 3250 3260 3270 3280


ACP-GFP_Ala79NorLys2
+12
+11
+15
+16
+14
+13
+17
+10
Dimer
+18 +17 +16
M_theo = 38.956 kDa
M_exp = 38.946 kDa
M_exp(Dimer) = 77.986 kDa
Intensity (a.u.)
m/z [+]
1000 2000 3000 4000 5000 6000
3240 3245 3250 3255 3260 3265 3270 3275 3280

Statement of Personal Contributions

Except where stated otherwise by reference or acknowledgment, the work presented herein was generated by myself under the supervision of my advisors during my doctoral studies. All contributions from colleagues or students under my supervision are explicitly referenced in the thesis.

The material listed below was obtained in the context of collaborative research:

Fig. 3.19 "Spin counting of spin labeled proteins in CW-EPR measurements": Dr. Alberto Collauto, AK Prisner, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry
Fig. 3.13 "Physicochemical analysis of fluorescently labeled mFAS": Dr. Alexander Rittner, AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology
Fig. 3.11 "In-gel fluorescence of fluorescently labeled mFAS and PKS constructs": Dr. Alexander Rittner, AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology, 50 % his contribution, 50 % my own contribution
Chapter 3.1.6 "**Modification of Full-Length mFAS**": Dr. Alexander Rittner, AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology, 50 % his contribution, 50 % my own contribution
Fig. 3.14 "Establishment of the fluorescence polarization assay": Franziska Stegemann, AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology
Experimental procedure 5.1.30 "**EPR Measurements**": Dr. Alberto Collauto, AK Prisner, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry
Experimental procedure 5.1.32 "**In-Gel Digestion**": Andreas Linden, AK Urlaub, Max-Planck-Institute for Biophysical Chemistry in Göttingen, 50 % his contribution, 50 % my own contribution
Experimental procedure 5.1.33 "**Tandem MS measurements**": Andreas Linden, AK Urlaub, Max-Planck-Institute for Biophysical Chemistry in Göttingen
Supplementary Data "MS analysis of purified ACP-GFP constructs": Khanh Vu Huu, AK Morgner, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry

The following persons contributed to the results of this thesis:

Dr. Alexander Rittner (AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology) contributed through his supervision during my master's thesis and conceived together with Prof. Dr. Martin Grininger (Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology) the initial idea of this project. His preliminary work, providing a protocol for the recombinant expression of mFAS and subconstructs in *E. coli* and several expression vectors utilized herein, is an important fundament of this thesis. Furthermore, he contributed partly in cloning of suppressor vectors pAC and was mainly responsible for the chemical syntheses of ncAAs. As co-first author of the publication Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878., he largely contributed to the physicochemical analysis of fluorescently labeled mFAS, performing HPLC-SEC and activity assays.

Student Bjarne Goebel was largely involved in the establishment of ncAA chemical syntheses under the supervision of Dr. Alexander Rittner (AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology).

Students Bjarne Goebel, Daniel Beyer, Sina Manger and Vanessa Bause performed reporter assays and upscale experiments thereof, under the supervision of myself and partly Dr. Alexander Rittner (AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology).

Khanh Vu Huu (AK Morgner, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry) performed MS analysis of ACP-GFP mutants containing ncAAs.

Student Maria Dell conducted important research on the establishment of smFRET for mFAS. She established the immobilization protocol and performed dSTORM and smFRET measurements at the group of Prof. Dr. Heilemann (Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry), under the joint supervision of Dr. Sebastian Malkusch (AK Heilemann, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry), Dr. Alexander Rittner (AK Grininger, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology) and myself.

Student Franziska Stegemann conducted the FP measurements under my supervision.

Frank Kaiser (AK Göbel, Goethe-University Frankfurt, Institute for Organic Chemistry and Chemical Biology) synthesized the DBCO-spin label utilized in this thesis, with help of his student Janik Weckesser.

Dr. Alberto Collauto (AK Prisner, Goethe-University Frankfurt, Institute for Physical and Theoretical Chemistry) performed the EPR measurements.

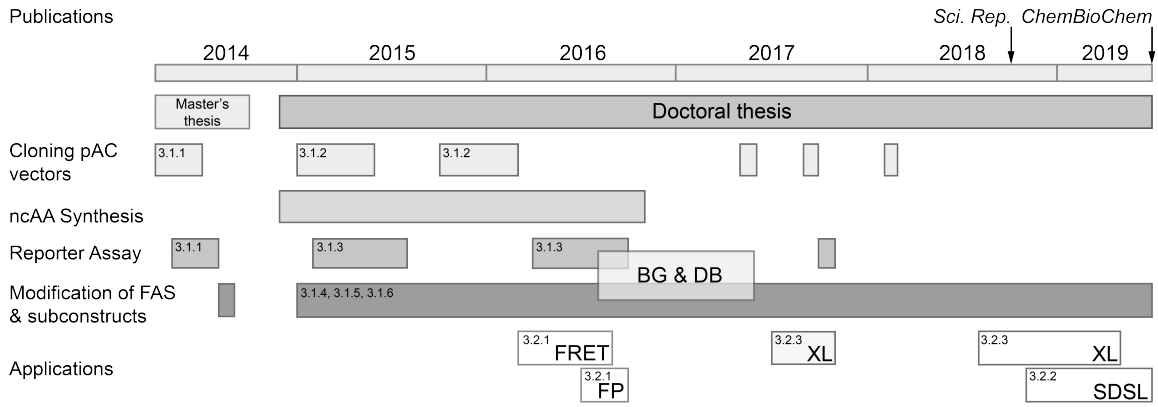
Student Felix Lehmann conducted the photocrosslinking assay under my supervision. He performed photocrosslinking experiments and surface mutations on the KS domain, as well as HPLC-SEC and TAL assays on the mutant constructs.

Andreas Linden (AK Urlaub, Max-Planck-Institute for Biophysical Chemistry in Göttingen) performed tandem MS experiments on the crosslinked protein constructs.

Students Tobias Matzel, Shuxin Jin and Rahimeh Ormoz assisted in cloning pAC suppressor vectors containing a wild-type *Mb*PylRS gene.

Time-Course of this Thesis

This thesis was compiled over the course of four years and several projects were implemented. The Gantt chart gives a better overview for the timeline of these projects. The results of my master's thesis were included as preliminary work. While in the beginning, the focus of my thesis lay on cloning of pAC vectors, ncAA synthesis and the reporter assay, modification of full-length mFAS and subconstructs was performed throughout the whole course of my doctoral thesis, with first implementations of spectroscopic methods in later stages. First applications included fluorescent spectroscopic studies. The FRET project was mainly performed during Maria Dell's master's thesis, the FP project during Franziska Stegemanns master's thesis. Later on, the photocrosslinking project was performed, mainly during Felix Lehmanns master's thesis. Master students Bjarne Goebel (BG) and Daniel Beyer (DB) both contributed with experiments to the modification of full-length mFAS and the reporter assay screening, which was published in *Scientific Reports*.[119] Site-directed spin labeling was performed by myself to the end of this thesis. Chapter numbers of the individual projects discussed in this thesis are given in the Gantt chart. Time points of publications are highlighted as well.



Copyright and Creative Commons Licences

Whenever a figure, table or text is identical to a previous publication, it is hereby stated explicitly that copyright permission and/or co-author agreement has been obtained.

The following parts of the thesis have been previously published:

- Chapters:

2.3 "**The Multidomain Enzyme Fatty Acid Synthase**" [6]
2.5 "**Aim of the Thesis**" [119]
2.4 "**Method Development**" [119]
4 "**Discussion & Outlook**" [119]
5 "**Experimental Procedures**" [119]

- Figures:

Fig. 2.2 "General cycle of fatty acid biosynthesis." [6]
Fig. 2.3 "Structure and domain organization of the type I FASs." [6]
Fig. 3.1 "Cloning scheme of pAC^{U} and pAC^{E} vectors." [119]
Fig. 3.4 "Amber codon suppression screened in the reporter assay." [119]
Fig. 3.5 "Screening amber codon mutation sites." [119]
Fig. 3.6 "Large scale expression and purification of ACP-GFP mutants." [119]
Fig. 3.7 "Quantification of ACP-GFP phosphopantetheinylation." [119]
Fig. 3.8 "Fluorescent labeling of ACP-GFP mutants." [119]
Fig. 3.12 "Generation of ncAA-modified mFAS mutants." [119]
Fig. 3.13 "Physicochemical analysis of fluorescently labeled mFAS." [119]
Fig. S1 "Screening suppression vector systems with reporter assay." [119]
Fig. S2 "Testing the effect of different ncAA concentrations." [119]
Fig. S3 "Optimizing fluorescent labeling of ACP-GFP mutants." [119]

Fig. S4 "Removal of excess free fluorophore." [119]
Fig. S6 "*In vitro* phosphopantetheinylation of wild-type mFAS." [119]

Tables:
Table S1 "List of suppression plasmids" [119]

Supplementary Data "MS analysis of purified ACP-GFP constructs" [119]

The complete draft of the manuscript published as Heil, Christina S. *et al.* "Site-specific Labelling of Multidomain Proteins by Amber Codon Suppression" *Scientific Reports* 8 (2018): 14864-14878. was initially written by myself. Dr. Alexander Rittner and Prof. Dr. Martin Grininger participated significantly with editing and phrasing to the final version in the process of publication.

The Review published as Heil, Christina S. *et al.* "Fatty Acid Biosynthesis: Chain-Length Regulation and Control" *ChemBioChem* (2019). was written by myself with minor contribution by S. Sophia Wehrheim and significant contribution by Prof. Dr. Martin Grininger. Dr. Karthik S. Paithankar participated with editing and phrasing to the final version in the process of publication.

Copyright and Creative Commons declaration of the *Scientific Reports* journal:

"*Scientific Reports* articles are published open access under a CC BY license (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials [...]. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor (read full legal code). [...] Under Creative Commons, authors retain copyright in their articles."
Source: https://www.nature.com/srep/about/open-access (accessed 29th June 2019)

Copyright and Creative Commons declaration of the *ChemBioChem* journal:

"Authors retain copyright of their article and are given a choice of Creative Commons license under which to publish their work."
Source: https://onlinelibrary.wiley.com/page/journal/14397633/homepage/onlineopen

(accessed 29th June 2019)

The following master's and bachelor's theses, which were submitted to the department 14 - Biochemistry, Chemistry and Pharmacy of the Goethe-University Frankfurt, contributed to the results of this thesis:

Christina Heil "Incorporation of Non-Natural Amino Acids in the Mammalian Fatty Acid Synthase" Master's thesis (2014).
Maria Dell "Konformationelle Analyse von Multienzym Proteinkomplexen des Fettsäure-Synthase I Systems mittels Einzelmolekülspektroskopie" Master's thesis (2016).
Franziska Stegemann "Fluorescence Studies on Dynamic Domains of the Mammalian Fatty Acid Synthase and Different Polyketide Synthases" Master's thesis (2016).
Sina Manger "Screening Incorporation Efficiencies of Non-Natural Amino Acids into ACP-GFP" Bachelor's thesis (2016).
Bjarne Goebel "Site Specific Labelling of Multidomain Enzymes via Amber Codon Suppression" Master's thesis (2017).
Daniel Beyer "Generation of Labeled Mammalian Fatty Acid Synthase for Application in Fluorescence Spectroscopy" Master's thesis (2017).
Felix Lehmann "Mapping Binding Interfaces of ACP within Mammalian FAS Using a Photocrosslinking Assay" Master's thesis (2019).

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung im mathematisch-naturwissenschaftlichen Bereich unterzogen habe. Des Weiteren erkläre ich, dass ich die vorgelegte Dissertation “Towards the Conformational Dynamics of Multidomain Proteins” selbstständig angefertigt und mich keiner anderen Hilfsmittel als der in ihr angegebenen bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere die Grundsätze der guten wissenschaftlichen Praxis beachtet und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den