
New probes for (membrane) protein structure, function and analysis

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften
(*Dr. rer. nat.*) an der Fakultät für Mathematik, Informatik
und Naturwissenschaften der Universität Hamburg,
Fachbereich Chemie

vorgelegt von

Stephanie Kesgin-Schäfer
geb. Schäfer

Hamburg, Januar 2019

Die vorliegende Arbeit wurde im Zeitraum von Mai 2015 bis Januar 2019 in der Arbeitsgruppe von Prof. Dr. H. Tidow am Institut für Biochemie und Molekularbiologie des Fachbereichs Chemie der Universität Hamburg durchgeführt.

1. Gutachter: Prof. Dr. Henning Tidow

2. Gutachter: Prof. Dr. Michael Kolbe

Datum der Disputation: 01.03.2019

Almost all aspects of life are engineered at the molecular level, and without understanding molecules, we can only have a very sketchy understanding of life itself.

— Francis Crick

TABLE OF CONTENTS

TABLE OF CONTENTS	I
LIST OF FIGURES	III
LIST OF TABLES	V
LIST OF ABBREVIATIONS AND DEFINITIONS	VI
PUBLICATIONS ASSOCIATED WITH THIS THESIS	VIII
ABSTRACT	IX
ZUSAMMENFASSUNG	X
SECTION A	1
INTRODUCTORY PREFACE	1
CHAPTER 1	2
<i>AN OVERVIEW OF PROTEIN ENGINEERING BY UNNATURAL AMINO ACID INCORPORATION</i>	
1.1 The Genetic Code.....	2
1.2 Genetic Code Engineering.....	4
1.3 Introduction to Unnatural Amino Acid Incorporation.....	7
1.3.1 Residue-Specific vs Site-Specific Incorporation.....	9
1.3.2 Residue-Specific Unnatural Amino Acid Incorporation.....	12
1.3.3 Site-Specific Unnatural Amino Acid Incorporation: Amber Codon Suppression.....	12
1.3.4 Frameshift Codons for Site-Specific Incorporation.....	13
1.3.5 O-Ribosome/O-mRNA Pairs for Site-Specific Incorporation.....	14
1.3.6 Engineering Aminoacyl-tRNA Synthetases.....	14
1.4 Designing Novel Proteins by Genetically UAA Incorporation.....	16
AIM OF THIS THESIS	18
SECTION B	20
GENETICALLY ENCODED SITE-SPECIFIC PROTEIN LABELLING	20
CHAPTER 2	21
<i>PHOTOACTIVATABLE GFP</i>	
2.1 Introduction.....	21
2.2 Materials and Methods.....	24
2.3 Results and Discussion.....	27
2.4 Conclusion.....	38
CHAPTER 3	39
<i>OMPX BINDING STUDIES</i>	
3.1 Introduction.....	39

3.2 Materials and Methods.....	42
3.3 Results and Discussion	46
3.4 Conclusion.....	56
SECTION C	58
RESIDUE-SPECIFIC PROTEIN LABELLING	58
CHAPTER 4	59
<i>THIOCYANATE LABELS IN LYSOZYME</i>	
4.1 Introduction.....	59
4.2 Materials and Methods	62
4.3 Results and Discussion.....	66
4.4 Conclusion	74
APPENDIX	76
GENERAL SAMPLE PREPARATION TECHNIQUES	77
INSTRUMENTATION AND CHEMICALS	85
INSTRUMENTATION	85
CHEMICALS USED (GHS CLASSIFICATION)	87
PROTEIN CRYSTALLIZATION SCREENS	92
GHS AND RISK SYMBOLS	93
GHS HAZARD STATEMENTS	93
GHS PRECAUTIONARY STATEMENTS	94
DISPOSAL	95
ACKNOWLEDGEMENT	96
CURRICULUM VITAE	98
EIDESSTÄTTLICHE ERKLÄRUNG	99
REFERENCES	100

LIST OF FIGURES

Figure 1.1: Schematic overview of the prokaryotic translation cycle	4
Figure 1.2: The genetic code.....	5
Figure 1.3: Overview of incorporation of amino acids analogues and unnatural amino acids into target proteins.....	6
Figure 1.4: Overview of selected chemical structures of genetically encoded unnatural amino acids.....	9
Figure 1.5: Selection schemes for engineering orthogonal aminoacyl-tRNA synthetases in <i>E. coli</i>	16
Figure 2.1: Green fluorescent protein.....	22
Figure 2.2: Decaging of the non-fluorescent GFP66ONBY by photolysis results in a reestablished fluorescent protein	23
Figure 2.3: SDS-PAGE showing the expression of sfGFP66ONBY and fractions from its purification.....	28
Figure 2.4: SDS-PAGE and size-exclusion chromatography (SEC) analysis of sfGFP variants	28
Figure 2.5: Excited-state proton transfer (ESPT) in GFP.....	29
Figure 2.6: Spectroscopic characterization of sfGFP66ONBY	30
Figure 2.7: Exemplary crystals of GFP66ONBY	31
Figure 2.8: Crystal structure of the domain-swapped sfGFP66ONBY.....	33
Figure 2.9: Structural comparison of sfGFP66ONBY domain-swapped structure with wild-type sfGFP dimer structure.....	34
Figure 2.10: Postulated folding mechanism of wt-GFP and its implications for the formation of domain-swapped sfGFP66ONBY (pdb: 6H01).....	36
Figure 2.11: Native mass spectrometry analysis of sfGFP variants.....	37
Figure 3.1: Topology and Structure of OmpX.....	40
Figure 3.2: L-Tyrosine and its analogue para-benzoyl-L-phenylalanine (BPA).....	42
Figure 3.3: Schematic overview of experimental setup for adhesion assays.....	47
Figure 3.4: Confocal fluorescence micrographs of liquid adhesion assay.....	49
Figure 3.5: Confocal fluorescence micrographs of fixed-cell adhesion assay.....	50
Figure 3.6: Statistical analysis of adhesion by bacterial cells expressing different OmpX variants.....	52
Figure 3.7: Construction of extracellular loop 2 (EL2) and 3 (EL3) mutants.....	53
Figure 3.8: Confocal fluorescence micrographs of fixed-cell adhesion assay of OmpX extracellular loop variants	54
Figure 3.9: Statistical analysis of adhesion by bacterial cells expressing different OmpX extracellular loop variants	55

Figure 3.10: SDS-PAGE and western blot analysis of OmpX variant membrane preparations.....	56
Figure 4.1: Overview of selected site-specific infrared probes	60
Figure 4.2: T4 Lysozyme as model system to investigate the influence of infrared label in local structural changes	61
Figure 4.3: Cysteine cyanylation	66
Figure 4.4: Full spectra of T4 lysozyme C54CN-C97A at pH 8	68
Figure 4.5: Infrared spectra of the vibrational active thiocyanate label from single S-cyano-L-cysteine labelled T4 lysozyme variant proteins at pH 8 and pH 6	69
Figure 4.6: Spectral behaviour of 4-cyano-L-phenylalanine in different local environments	70
Figure 4.7: Hydrogen bond analysis of the two different cyanylated cysteine systems	71
Figure 4.8: Influence of the cyanate-labelling of cysteines on the protein backbone structure/dynamic	73
Figure 4.9: Inelastic neutron scattering data for the labelled and unlabelled T4 protein variant pairs	74

LIST OF TABLES

Table 1: Comparison of selective pressure incorporation and codon suppression	11
Table 2: Data collection and refinement statistics for dark-state sfGFP66ONBY (pdb:6H01)	32
Table 3: Comparison of the infrared signal of the thiocyanate label of single S-cyano-L-cysteine labelled T4 lysozyme variant proteins at pH 8	70
Table 4: Growth media	77
Table 5: Protein properties used for concentration determination protein molecular weights and extinction coefficients used for the protein concentration determination	78
Table 6: List of buffers and solutions used for SDS-PAGE	79
Table 7: Bacterial strains	80
Table 8: Table of used plasmids	81
Table 9: Table of used primers	83
Table 10: Instrumentation (listed alphabetically)	85
Table 11: Chemicals (listed alphabetically)	87
Table 12: Crystallisation screens with HP-statements	92

LIST OF ABBREVIATIONS AND DEFINITIONS

α	Alpha
Å	Angstrom
aaRS	Aminoacyl-tRNA synthetase
Amp	Ampicillin
β	Beta
C	Celsius
Chlor	Chloramphenicol
Da	Dalton
DNA	Deoxyribonucleic acid
fs	Femtosecond
γ	Gamma
g	Gram
g	Gravitational force
GFP	Green fluorescent protein
h	Hour
IPTG	Isopropyl - β -D-thiogalactopyranosid
K	Kelvin
Kan	Kanamycin
l	Liter
LB	Lennox-Broth
min	Minute
Ni-NTA	Nickel-Nitrilotriacetic acid
mm	Millimeter
mol	mole
ms	Millisecond
M	Molar
MWCO	Molecular weight cutoff
μ	Micro
nm	Nanometer
ns	Nanosecond
O ₂	Oxygen
OD	Optical density
OmpX	Outer membrane protein X
ps	Picosecond
PDB	Protein Data Bank
(t)RNA	(Transfer) ribonucleic acid
rpm	Rounds per minute
s	Second
SEC	Size exclusion chromatography
TB	Terrific broth
UV	Ultraviolet
V	Volt
v/v	Volume per volume
W	Watt
w/v	Weight per volume

Canonical: Limitation of 20 standard amino acids and three stop signals encoded by the consensus genetic code.

Amino acid analog: Structural resemblance with the respective amino acid.

Amino acid surrogate: Strong structural and/or electronic resemblance to the respective amino acid and accepted (with lower efficiency) by the amino acid's aminoacyl-tRNA synthetase.

Noncanonical: Not part of canonical processes (e.g. natural or synthetic amino acids not normally involved in translation).

Orthogonal: Not interfering with and not interfered by natural structures and processes.

Canonical amino acid abbreviations: Alanine (Ala/A); cysteine (Cys/C); aspartic acid (Asp/D); glutamic acid (Glu/E); phenylalanine (Phe/F); glycine (Gly/G); histidine (His/H); isoleucine (Ile/I); phosphoserine (Sep/J); lysine (Lys/K); leucine (Leu/L); methionine (Met/M); asparagine (Asn/N); pyrrolysine (Pyl/O); proline (Pro/P); glutamine (Gln/Q); arginine (Arg/R); serine (Ser/S); threonine (Thr/T); selenocysteine (Sec/U); valine (Val/V); tryptophan (Trp/W); tyrosine (Tyr/Y).

Noncanonical amino acids abbreviations: BPA: *para*-benzoyl-L-phenylalanine; ONBY: *ortho*-nitro-benzoyl-phenylalanine

Standard DNA/RNA nucleotide bases: Adenine (A); cytosine (C); guanine (G); thymine (T); uracil (U)

PUBLICATIONS ASSOCIATED WITH THIS THESIS

Stephanie Kesgin-Schaefer, Johannes Heidemann, Anke Puchert, Knut Koelbel, Briony Yorke, Arwen R. Pearson, Nils Huse, Charlotte Uetrecht, and Henning Tidow. *Crystal structure of a domain-swapped photoactivatable sfGFP variant provides evidence for GFP folding pathway.* Submitted to the The Federation of European Biochemical Societies Letters.

Stephanie Kesgin-Schaefer, Nils Huse, Arwen Pearson, Henning Tidow, and Stephan Niebling. *Investigating local structural heterogeneity in T4 lysozyme by a thiocyanate IR label.* In preparation for submission.

ABSTRACT

The development of new orthogonal tRNA and aminoacyl-tRNA synthetase pairs enabling unnatural amino acid incorporation by amber codon suppression led to the expansion of the genetic code by over hundred unnatural amino acids. These unnatural amino acids represent a variety of structures and functions not found in the canonical 20 amino acids. These novel amino acids offer scientists a powerful tool to generate unique proteins with enhanced or novel properties in a tailored way with high molecular precision as well as probes to study protein structure and function in *Escherichia coli*, yeast, and mammalian cells. This technique has been used not only in fundamental research but also in research applications as well therapeutic drug design, including the generation of bi-specific antibodies and antibody drug conjugates. In the framework of this thesis, several methods related to the incorporation of unnatural amino acids into proteins were used for three different approaches to study protein structure and function.

The incorporation of the photocaging *ortho*-nitrobenzyl-tyrosine into a superfolder green fluorescent protein revealed an unexpected domain-swapped dimeric arrangement, which suggests implications for the folding pathway of green fluorescent protein. The unusual domain-swap suggests that green fluorescence protein indeed folds via equilibrium or kinetic intermediates along one of four postulated pathways.

In a second approach, the photo-crosslinker *para*-benzoyl-phenylalanine could be successfully incorporated in the virulence-related outer membrane protein X and *in vivo* crosslinked to antigen-presenting cells of the human immune system. Extracellular loop truncation variants showed that these are indeed required to function as a molecular fishing rod to attach to eukaryotic cells.

Additional emphasis was placed on the investigation of protein labels and their effect on native structure and dynamics. The analysis of cysteine thiocyanate labelling in lysozyme underlined the use of thiocyanates not only as infrared probes for vibrational spectroscopy but also as pH sensitive probes to characterize the microenvironment of the probe. A first analysis of molecular dynamics simulations and neutron spectroscopy furthermore showed that these small infrared label indeed cause changes on local and global native structure.

ZUSAMMENFASSUNG

Die Entwicklung neuer orthogonaler tRNA- und Aminoacyl-tRNA-Synthetase-Paare, die einen Einbau von unnatürlichen Aminosäuren durch Amber-Codon-Suppression ermöglichen, führte zur Erweiterung des genetischen Codes um über hundert unnatürliche Aminosäuren. Diese unnatürlichen Aminosäuren repräsentieren eine Vielzahl von Strukturen und Funktionen, die nicht in den 20 kanonischen Aminosäuren vorkommen und erlauben die Entwicklung von Proteinen, die durch die ausschließliche Verwendung der 20 kanonischen Aminosäuren nicht möglich wären. Sie bieten Wissenschaftlern daher ein wirkungsvolles Werkzeug, um mit hoher molekularer Präzision maßgeschneiderte Proteine mit verbesserten oder neuartigen Eigenschaften sowie Sonden zur Untersuchung der Proteinstruktur und -funktion in *Escherichia coli*, Hefe und Säugetierzellen zu entwickeln. Diese Technik wird vielfach nicht nur in der Grundlagenforschung, sondern auch in zahlreichen Forschungsanwendungen sowie für die Entwicklung therapeutischer Wirkstoffe einschließlich der Produktion von bi-spezifischen Antikörpern und Antikörper-Wirkstoff-Konjugaten verwendet. Im Rahmen dieser Arbeit wurde für drei verschiedene Ansätze die Änderung der Proteinstruktur und -funktion untersucht, die sich durch den gezielten Einbau unnatürlicher Aminosäuren ergibt. Hierbei kamen unterschiedliche Methoden der Proteinmodifizierung zum Einsatz.

Der Einbau des photo-aktivierbaren *ortho*-Nitrobenzyl-Tyrosins in ein grünes fluoreszierendes Protein zeigte eine unerwartete dimerische Anordnung mit ausgetauschten Domänen, welche auf einen Faltungsweg des grün fluoreszierenden Proteins hindeuten. Der seltene und unerwartete Domain-Swap lässt vermuten, dass sich grün fluoreszierendes Protein über Gleichgewichts- oder kinetische Intermediate entlang eines von vier postulierten Faltungswegen faltet.

In einem zweiten Ansatz konnte der Photocrosslinker *para*-Benzoyl-Phenylalanin erfolgreich in das Außenmembranprotein X, welches Virulenz und Pathogenität beeinflusst, eingebaut und anschließend *in vivo* mit Antigen-repräsentierenden Zellen des menschlichen Immunsystems vernetzt werden. Extrazelluläre Loop-Verkürzungsvarianten zeigten, dass diese als molekulare „Angelrute“ zur Bindung von eukaryotischen Zellen fungieren.

Ein weiterer Schwerpunkt lag auf der Untersuchung von Proteinlabeln und deren Auswirkungen auf

die native Struktur und Dynamik von Proteinen. Die Analyse der Thiocyanat-Label in Lysozym unterstrich die Eignung von Thiocyanaten nicht nur als Infrarotsonden für die Schwingungsspektroskopie, sondern auch als pH-empfindliche Sonden zur Charakterisierung ihrer Mikroumgebung. Eine erste Analyse der Simulation von Molekulardynamiken und von Neutronenspektroskopie-Messungen zeigte außerdem, dass diese kleinen Infrarotmarkierungen Änderungen der lokalen und globalen nativen Proteinstruktur verursachen.

SECTION A

INTRODUCTORY PREFACE

CHAPTER 1

1. AN OVERVIEW OF PROTEIN ENGINEERING BY UNNATURAL AMINO ACID INCORPORATION

1.1. The genetic code

Over 50 years ago, the historic and groundbreaking “The Crick, Brenner et al. experiment” proved that codons consist of three DNA bases and code for one amino acid and thereby established the triplet nature of the genetic code [1]. The same year, Marshall W. Nirenberg and Heinrich J. Matthaei cracked the genetic code by discovering that the amino acid phenylalanine was assigned to the codon UUU [2]. These groundbreaking conclusions revealed the rules by which the genetically encoded information on a nucleotide sequence is translated into the amino acid sequence of a protein's polypeptide chain and confirmed the central dogma of molecular biology, hypothesized already in 1958 by Francis Crick [3]. According to the central dogma of biology, the protein biosynthesis starts with the transcription of the genetic information written in the DNA into an intermediate molecule, the mRNA, which is later translated into a protein by the ribosome, a complex molecular machinery. By 1966, all 64 codons could be mapped to 20 proteinogenic amino acids and three stop codons (TAA “ochre”, TGA “opal” and TAG “amber”) encoding for termination of protein biosynthesis. Since then, scientists have been studying the processes that synthesize biomolecules at the molecular level leading to an enhanced and detailed understanding of the molecular processes involved in protein biosynthesis. Milestones since the discovery of the triplet nature of the genetic code include the isolation of the first DNA polymerase in 1955 by Arthur Kornberg [4, 5] enlightening how DNA sequences are transcribed into mRNA molecules as well as the identification of the first sequence of a tRNA and the identification of the anti-codon in it in 1965 by Robert Holley illuminating how ribosomes translate mRNA into amino acid sequences [6]. These scientific milestones finally made it possible to understand the processes of protein biosynthesis at the molecular level.

DNA transcription, the first step in protein biosynthesis, occurs in the nucleus. The protein-coding gene is transcribed via a three-step process into mRNA [7]. Each stage is highly regulated by a large number of proteins such as transcription factors and coactivators to ensure correct gene expression.

Transcription begins with the binding of RNA polymerase, together with one or more general transcription factors, to the promoter DNA and separates the DNA double strands by hydrogen bond breakings. The RNA polymerase then adds complementary RNA nucleotides to the template DNA strand and forms an RNA sugar-phosphate backbone, thereby synthesizing an mRNA single-strand. After transcription, the hydrogen bonds of the RNA–DNA helix break, releasing the newly synthesized nascent mRNA strand, which is further processed by polyadenylation, capping, and splicing. The mRNA molecules exit to the cytoplasm through the nuclear pore complex.

During translation, mRNA that has been transcribed from the DNA is used as a template to link amino acids into a polypeptide chain determined by codons consisting of three base pairs [8] (Figure 1.1). This is achieved by the use of aminoacylated tRNAs containing anticodons that are able to recognize the codon by Watson-Crick base-pairing [9]. Aminoacyl-tRNA synthetases (aaRS) catalyze the ligation of amino acids to tRNAs enabling the crucial linkage of a codon to its cognate amino acid. Cells contain 20 aaRSs, one for each of the 20 standard proteinogenic amino acids [10]. Each aaRS specifically recognizes and binds only its cognate aminoacyl-tRNA containing the corresponding anticodon to base pair with the codon on the mRNA. Translation itself occurs in the cytoplasm, executed by the ribosomes and proceeds in four phases: activation, initiation, elongation, and termination [11, 12]. First, the amino acid is activated using adenosine-5'-triphosphate (ATP) to form the acid anhydride between the carboxylate and the phosphate group of adenosine-5'- monophosphate (AMP) by the aaRS. The activated amino acid is then esterified with the 2'-OH or 3'-OH of the 3'- terminal ribose of the tRNA, and the AMP is released. In the initiation phase, the ribosome assembles with the help of initiation factors around the mRNA, and the first tRNA, usually a Met-tRNA^{Met}, is attached at the start codon. The elongation of the polypeptide chain is facilitated by peptide bond formation of the amino acids by elongation factors and the energy of GTP hydrolysis [13, 14]. Once the ribosome encounters one of three stop codons, release factors bind and catalyze the hydrolytic cleavage of the nascent polypeptide chain resulting in translation termination and polypeptide chain release from the ribosome [15]. Hence, the ribosome is a molecular machine that is capable of producing unidirectional movement of mRNA and tRNA driven by the hydrolysis of GTP. The ribosome uses Brownian motion for directed movements with elongation factors and the hydrolysis of GTP as driving force for forward movement and fixation of the post state [16].

DNA translation is a highly orchestrated process and due to the complexity of codon recognition and the assembly of amino acids into proteins the most elaborated biosynthetic process. Due to the intricacy of proteins and their distinct functions, it is necessary to guarantee extremely accurate

translation of the genetic information [17]. This high level of accuracy is achieved by the aaRSs, which have each evolved highly precise mechanisms for codon recognition and proofreading [18-20].

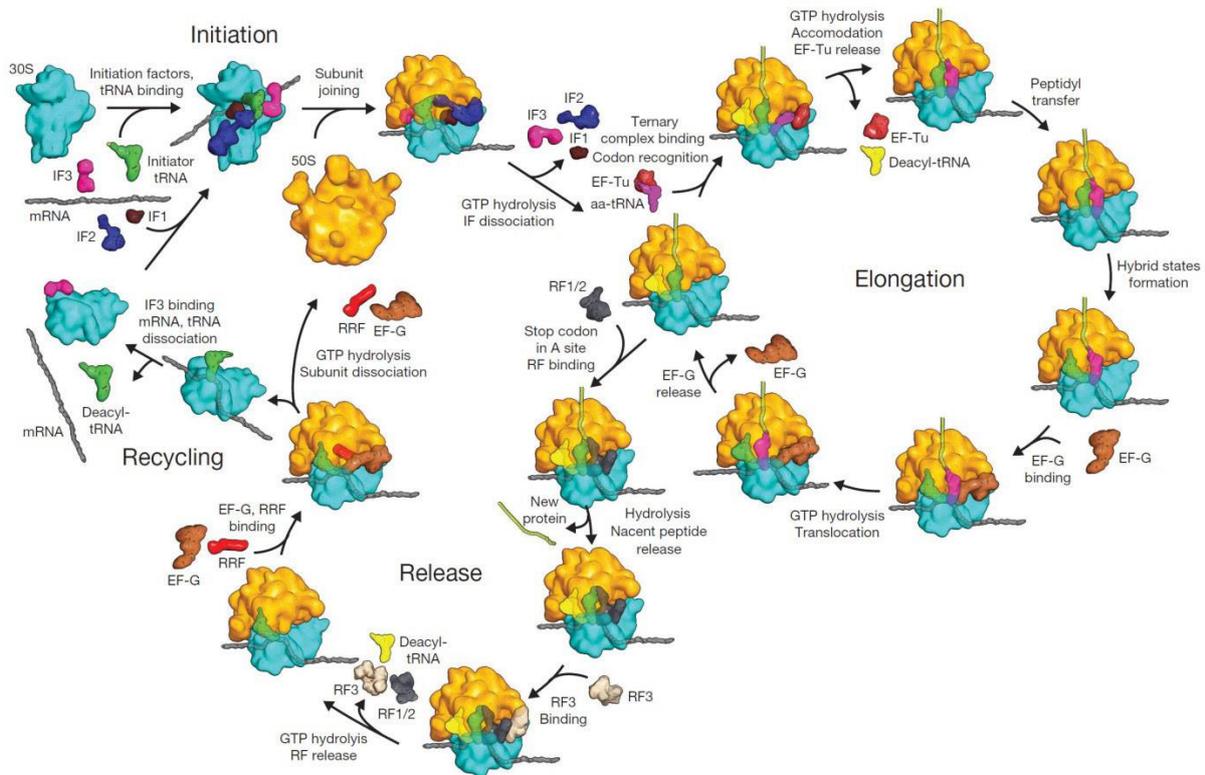


Figure 1.1: Schematic overview of the prokaryotic translation cycle. Translation can be subdivided into: initiation, elongation and termination (or release), which are followed by ribosome recycling. Aminoacyl-tRNA synthetases aminoacylate cognate tRNAs with the corresponding amino acids. Aminoacylated tRNAs bind to elongation factor EF-Tu for trafficking to the ribosome. Aminoacylated tRNAs accommodate the A-site, recognizing the presented codon on the mRNA by base pairing. In the next step, the nascent protein chain is transferred from a tRNA located in the ribosome's P-site, elongating the chain by one amino acid. Finally, the ribosomal subunits briefly dissociate and migrate on the mRNA, ejecting the tRNA from the P-site and placing the tRNA formerly in the A-site, now carrying the nascent protein chain, in the P-site, freeing the A-site. A stop codon in the A-site triggers binding of a release factor instead of a tRNA, leading to dissociation of the ribosome and release of the protein chain (termination). IF, initiation factor; EF, elongation factor; RF, release factor (Adopted from Schmeing & Ramakrishnan, 2009 [8]).

1.2. Genetic code engineering

The genetic code is read by an evolutionary conserved translational machinery in all living organisms and all proteins are synthesized from the limited number of 20 canonical amino acids (Figure 1.2). The set of these 20 building blocks prescribed by the universal genetic code often does not contain all chemistries necessary for the diverse cellular functions of proteins in multicellular organisms.

Therefore, not all proteins have a final covalent structure after DNA transcription and mRNA translation and often contain additional chemical modifications. Thus, the release of a completed polypeptide chain from a ribosome is often not the last step in protein formation. A large part of proteins is post-translationally processed by enzymatic cleavages or selective binding of different chemical functional groups at specific sites of the protein or thereby converted into the active form [21]. Crucial posttranslational modifications of amino acids include hydroxylation (Pro), phosphorylation (Ser/Thr, Tyr, Asp, His), methylation/acetylation (Lys), modification with small molecules, cofactors (non-protein chemical compounds or metallic ions) or proteins (glycosylation, ubiquitination and prenylation of nucleophilic amino acids), and complex modification of ribosomal peptides (macrocyclisation and reduction/oxidation) [22-24]. Post-translational modifications are strictly separated from decoding and proof reading as well as selectively and timely coordinated by specific enzymes and enzymatic complexes.

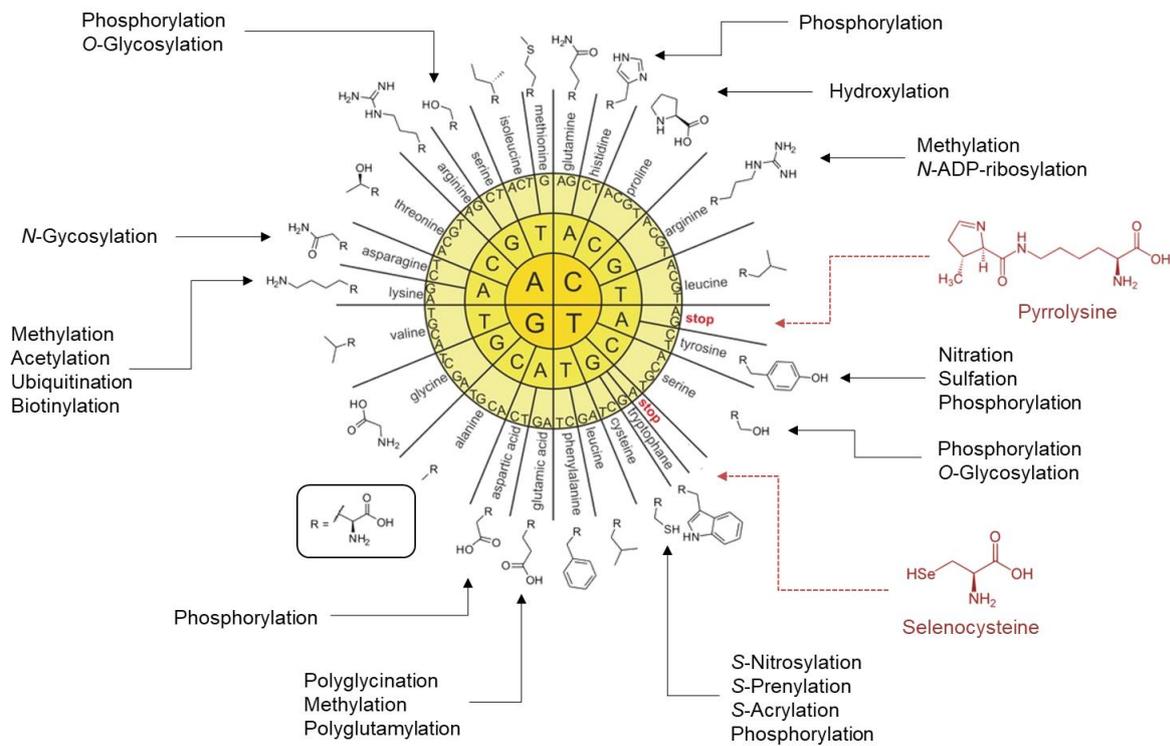


Figure 1.2: The genetic code. The consensus genetic code (DNA notation) used by the majority of organisms with semicanonical amino acids encoded by reassigned stop codons (shown in red). For each amino acid possible chemical functional groups via posttranslational modification are indicated.

Even though most aminoacyl tRNA synthetases are notable specific for their cognate amino acid, infrequently misacylation can occur with structurally similar amino acids [25]. This inability of aminoacyl-tRNA synthetases to distinguish close structural analogues from their cognate amino acid

can result in nascent proteins with residue-specific surrogate noncanonical amino acid analogues or misincorporated canonical amino acids [24, 26] (Figure 1.3). For instance, leucine and isoleucine swapping during aminoacylation can be frequently observed. However, some noncanonical amino acids are structural surrogates to canonical ones and share high structural similarities, these often result in toxic effects for the organism [27, 28].

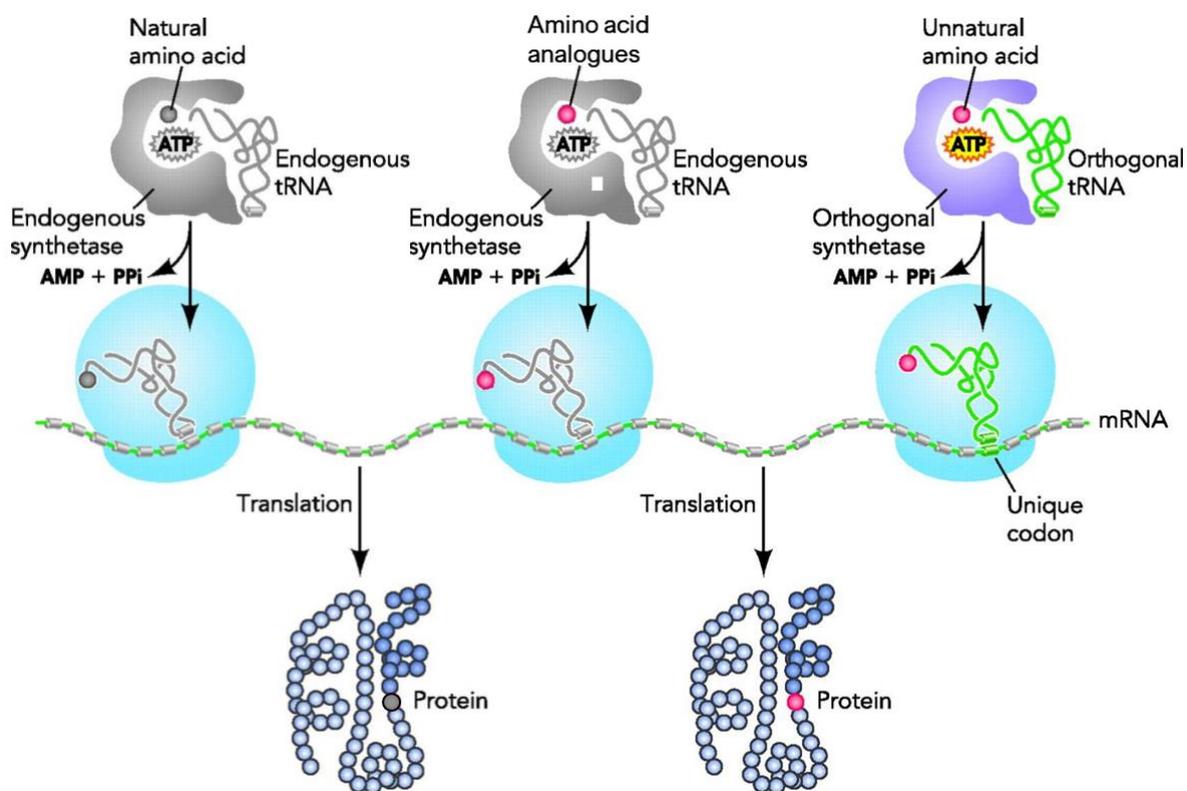


Figure 1.3: Overview of incorporation of amino acids analogues and unnatural amino acids into target proteins. Aminoacyl-tRNA synthetases link amino acids to their complementary tRNAs by aminoacylation. The three major steps in the aminoacylation process are the activation of the amino acid using ATP, the tRNA binding and aminoacylation and the release of aminoacylated tRNA. The aminoacylated tRNA is then delivered to the ribosome for protein synthesis. Whereas amino acid analogues can be incorporated residue-specific by endogenous tRNA and tRNA-synthetase pairs, unnatural amino acids can only be site-specific incorporated by specific orthogonal tRNA and tRNA-synthetase pairs in response to unique codons, for instance “amber” codons (Modified from Wang, 2017).

A limited number of archaea and eubacteria naturally incorporate the special noncanonical amino acids pyrrolysine [29] and selenocysteine [30] co-translationally in a small fraction of their proteins by the reassignment of the termination codons [31, 32] (Figure 1.2 +1.3). Pyrrolysine can be found in methanogenic archaea and bacteria: the archaea *Methanosarcinaceae* genetically encodes pyrrolysine in the methylamine methyltransferase genes via the UAG stop codon (“amber” codon) as well as the

gram-positive *Desulfitobacterium hafniense* in the trimethylamine methyltransferase homolog [33, 34]. The incorporation of pyrrolysine does not only depend on the presence of the “amber” codon but also on the specific structural element pyrrolysine insertion sequence located in the mRNA. The presence of the insertion sequence enables the “amber” codon suppression and allows subsequent pyrrolysine incorporation. Analogically, selenocysteine is encoded by the UGA stop codon (opal codon) and depends on the presence of the selenocysteine insertion sequence located adjacent to the opal codon in the mRNA in prokaryotes or the 3'-untranslated region in eukaryotes [35].

Hence, evolutionary two different strategies evolved to increase the variety of the 20 amino acid side chains contributing to protein structure and function diversity: a majority of proteins that is posttranslational modified with functional groups in a strictly separated process and a small fraction of proteins containing cotranslationally incorporated special proteinogenic amino acids such as pyrrolysine and selenocysteine via stop codon suppression. Interestingly, the second strategy demonstrates that organisms have an efficiently cellular uptake and metabolic stability for useful noncanonical amino acids and a translational machinery tolerating the addition of novel amino acids to the standard collection of amino acids.

1.3. Introduction to Unnatural Amino Acid Incorporation

However, even with the extension of the standard genetic code by the noncanonical amino acids pyrrolysine and selenocysteine, the variety of amino acids side chains is still very limited to generate all natural proteins in living organisms. The number of possible gene products is theoretically infinite due to flexibility in protein length and amino acid position in the polypeptide chain [36, 37]. Though the theoretical variability of proteins is immense, it is practically limited due to functional capacity of proteins including the necessity of functional groups like hydroxyl, methyl and aromatic side chains. Thus, the addition of novel side chains would increase the variety of chemical and physical properties of proteins with the addition of e.g. fluorine, azide, alkene, and alkyne functionalities [36]. The creation of organisms with expanded genetic codes that include additional building blocks might also allow the design of proteins with enhanced or new activities facilitating the analysis, control and imaging of protein structure and function *in vitro* and in living cells [38]. Finally, an expanded genetic code may provide advantage in the study, manipulation and evolution of proteins with new molecular or organismal function [39]. These novel functionalities can be realized through the incorporation of unnatural amino acids into proteins and the rewiring of translation.

The evolutionary posttranslational modification apparatus is exceptionally complex and practically impossible to mimic. Thus, research focuses on alternative possible ways to exploit and divert cellular systems to gain additional chemical variety, including efficient cellular uptake of unnatural amino acids, metabolic stability, translational incorporation, (re)assignment of codons as well as chemical synthesis of useful unnatural amino acids [36]. These methods to engineer proteins containing unnatural amino acids advanced and improved significantly in the last decade and constitute a powerful, easy to use and implement tool valuable for various applications. While research on unnatural amino acid incorporation has seen major interest and growth since the development of recombinant DNA technology [40], this area of research dates back to the 1960's. Dean B. Cowie and Georges N. Cohen were the first to completely replace methionine by selenomethionine in *E. coli* in 1957 [41]. This simple substitution method was for instance widely used since 1990 to improve structure determination by x-ray crystallography using single- or multi-wavelength anomalous diffraction ([42, 43]. Due to heavy atom incorporation, the phase problem in crystallography could be solved. While simple substitution methods marked the beginning of unnatural amino acid incorporation, recent research focuses mainly on the reprogramming or the orthogonalization of the protein's natural translation machinery. This includes the advanced reassignment of the genetic code (codons) to unnatural amino acids, the engineering of orthogonal tRNA and aminoacyl-tRNA synthetase pairs and the increase of mainly synthetic unnatural amino acids in order to expand the scope of protein biosynthesis and the molecular toolbox for protein engineering [36, 38, 39, 44-47].

As mentioned, unnatural amino acids represent an extensive range of chemical and physical properties and can carry out a variety of new structures and functions not found in the common natural amino acids, such as infrared probing [48, 49], photo-induced switching [50], redox sensitive proteins [51], hyperstable proteins [52], heavy atom-proteins for x-ray crystallography [53] and protease-resistant proteins [54]. These methods allowed scientists and engineers to incorporate over hundred unnatural amino acids (Figure 1.4) into proteins for various applications in a residue-specific or site-specific fashion [55] [38, 56-59].

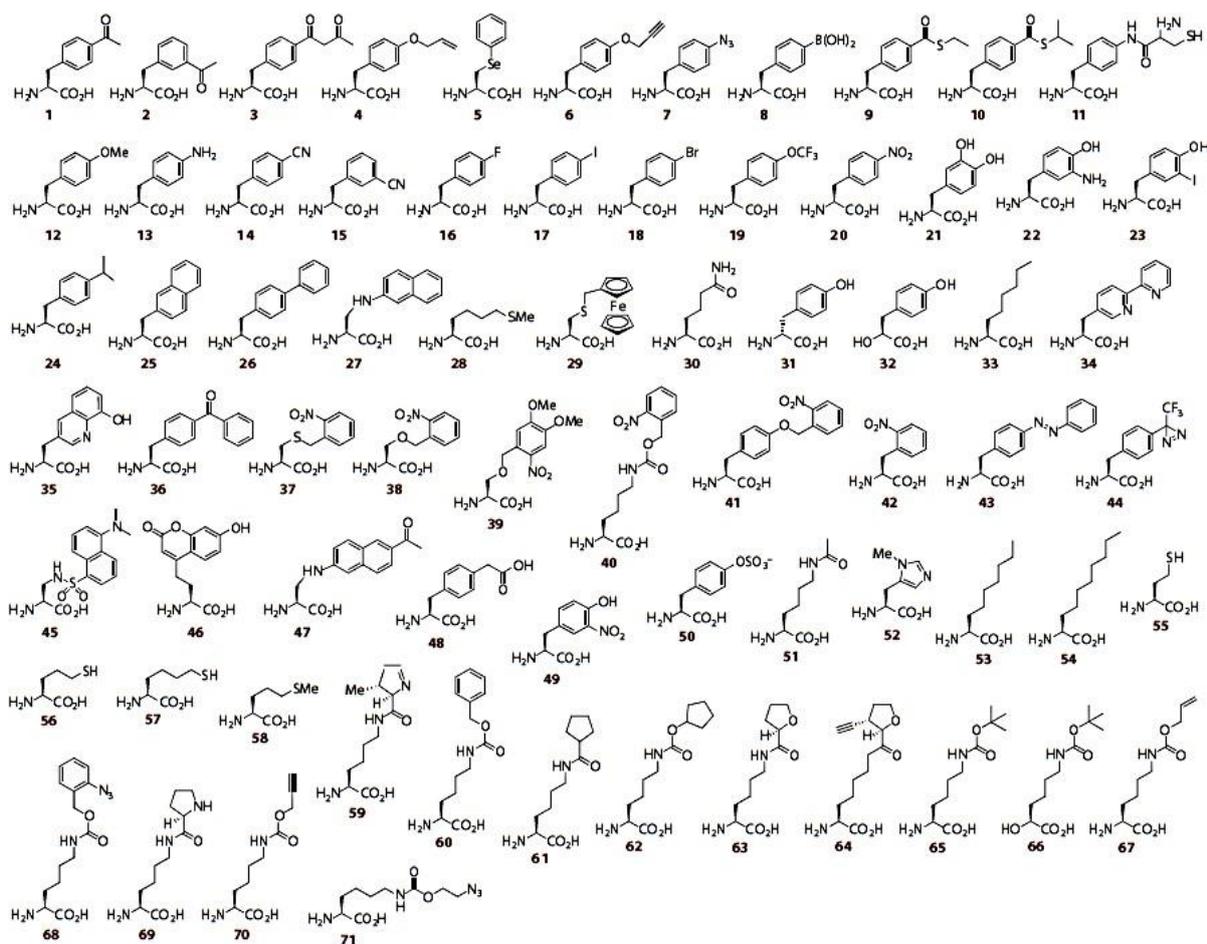


Figure 1.4: Overview of selected chemical structures of genetically encoded unnatural amino acids (Adopted from Liu & Schultz, 2010).

1.3.1. Residue-specific vs. Site-specific Incorporation

As mentioned, for different technological purposes novel unnatural amino acids can be genetically incorporated residue-specific or site-specific by two principal methods *in vivo*: selective pressure incorporation or codon suppression, respectively. While both techniques achieve the incorporation of noncanonical amino acids in a different way, they are not opposing, but rather complementary methods (Table 1).

Residue-specific unnatural amino acid incorporation is easily applicable to an array of diverse surrogate building blocks with readily adaptation for multiple different amino acids and often high efficiencies, which allows global replacement of natural amino acids by their analogues utilizing the loose substrate recognition and inability of aminoacyl-tRNA synthetases to distinguish between analogues and their cognate substrate [36, 60, 61]. Since the analogues are integrated in all positions, where the

corresponding cognate substrate is located, the resulting engineered protein has a greater probability for considerably different chemical and physical properties relative to the wild-type protein and multiple incorporations of surrogates into the protein may not be tolerated. Due to efficient discrimination of non-cognate analogues against their corresponding natural building block, the residue-specific approach requires several necessities including culturing cells in defined growth medium, a media shift procedure to reduce the cognate canonical amino acid, the use of auxotrophic strains for the natural amino acid to be replaced, and an isostructural amino acid analogue to one of the proteinogenic amino acids [61, 62]. The unneeded manipulation of the genetic information constitutes a major advantage of this method.

Site-specific unnatural amino acid incorporation allows the insertion of novel building blocks with minimal perturbation to the overall structure of a protein via point mutations in the gene. This feature makes site-specific incorporation a great versatile tool and especially suitable for applications in protein structure determination [63, 64]. Today, the method is applicable to a large number of organisms including bacteria, yeast, nematodes and mammalian cells [38, 39, 46, 65, 66]. On the downside, the use of site-specific incorporation requires an unassigned codon (sense codon suppression), manipulation of the target sequences to introduce unique stop codons (stop codon suppression) or frameshifts (frameshift/quadruplet codon suppression) as well as the establishment of an orthogonal tRNA and aminoacyl-tRNA synthetase pair in the host system without any cross-reactivity with host's natural translation system as unnatural amino acids are in many cases not accepted by natural orthogonal systems [67-70]. Due to these circumstances, extensive redirected evolution procedures are necessary to generate orthogonal systems activating and incorporating the desired unnatural amino acid often leading to lowered protein yields of approximately 50% compared to normal expression levels for stop codon suppression [36, 38, 65, 66, 68] and suppression of sense and frameshift codons with even lower protein yields [71, 72]. Moreover, incorporation is often limited to only one or a few novel unnatural amino acids, as unique codons, such as the "amber" codon, are the only available triplet codons and the nonsense suppression rate in living cells is extremely low [73]. As an alternative to overcome the limitation problems, new developments, such as quadruplet-decoding ribosomes might be a solution [74]. A final limitation of site-specific incorporation and codon suppression techniques might be the dependency on the codon context. While for the incorporation of pyrrolysine the pyrrolysine insertion sequence located downstream on the mRNA is necessary, it has been shown that these structural elements do not play a significant role for stop codon suppression using the pyrrolysine-tRNA/pyrrolysine-tRNA synthetase pair [25, 75]. In contrast the efficiency is strongly affected by the particular position of the suppressed codon. Generally, stop codons near the

N-terminus of the protein are more efficient, but the codon context can alter suppression events of both stop and sense codons [66, 76, 77]. The underlying mechanism of this phenomenon is still poorly understood, but efforts have been made to decipher the codon context for stop codon suppression efficiency [77]. Consequently, a screening of stop codon positions is advisable for each desired target protein. Nevertheless, the vast possibilities of site-directed unnatural amino acid incorporation makes codon suppression a substantial and versatile tool for protein engineering.

Table 1: Comparison of selective pressure incorporation and codon suppression.

	Selective pressure incorporation	Codon suppression
Mode of incorporation	residue-specific	site-specific
Amino acid composition	unnatural amino acid replaces canonical amino acids	Additional unnatural amino acid to the 20 standard amino acids
Amino acid limitations	only isostructural analogues	undetermined, limited by evolvability of orthogonal tRNA/aaRS
Target gene sequence	unchanged	mutated (insertion of unique suppression codons)
Amino acid activation	natural tRNA/aaRS of host	additional engineered orthogonal tRNA/aaRS
Competition with host's translational machinery	isostructural analogues compete with standard substrate for aminoacylation	unnatural amino acids have own unique tRNA/aaRS
Host requirements	auxotrophic strains (for amino acid to be replaced by surrogate)	absence of strong natural suppressors, additional strain engineering to reduce competing processes advantageous
Multiple incorporations	possible	conditionally possible, might require strain engineering
Multiple incorporations of different unnatural amino acids	possible, requires multiple auxotrophies	difficult, requires multiple orthogonal tRNA/aaRS pairs and two or more suppressed codons

1.3.2. Residue-specific Unnatural Amino Acid Incorporation

Residue-specific approaches exploit the organism's native translational apparatus to replace a natural amino acid with an isostructural analogue. The resulting protein contains 19 standard canonical amino acids and one additional unnatural amino acid analogue. As the isostructural analogue will not only be incorporated in the protein of interest, but rather in the whole proteome, global alterations in protein structure and function can occur [78]. Due to the necessity of reassigning sense codons to the isostructural analogues by exploiting the substrate tolerance of endogenous aminoacyl-tRNA synthetases, this can be achieved by three technical possibilities: enhancing the host's natural aminoacyl-tRNA synthetase activity by overexpression of the synthetase [79, 80], altering the amino acid binding pocket of the aminoacyl-tRNA synthetase [81] [82-84] or manipulating the editing domain of the aminoacyl-tRNA synthetase to incorporate different analogues [85].

Since the majority of amino acid analogues are not capable of supporting cell growth, the auxotrophic host organism must first be grown in media containing the standard 20 natural amino acids. To avoid expression of recombinant target protein without the desired isostructural analogue, tight repression of protein expression is required. Upon accumulation of an adequate cell mass, a media-shift is conducted to substitute the natural amino acids with its unnatural analogue. Recombinant DNA technology allows overexpression of the target protein via induction. This method was firstly used to globally replacement methionine with its analogue selenomethionine [41]. Since then this method has been used to incorporate a variety of analogues including analogues of methionine [79, 86, 87], isoleucine [88, 89], leucine [52], phenylalanine [90], proline [60], tryptophan [60], tyrosine [90], and valine [88] in bacterial [89] and mammalian expression systems [91].

1.3.3. Site-specific Unnatural Amino Acid Incorporation: Codon suppression

Unlike residue-specific unnatural amino acid incorporation, the site-specific approach does not utilize the organism's native translational machinery to integrate novel amino acid analogues but reassigns a codon to a new building block ignoring its actual canonical assignment in the genetic code. As mentioned, natural stop codon suppression can be found in several archaea and eubacteria, which naturally incorporate the noncanonical amino acids pyrrolysine and selenocysteine co-translationally in response to unique termination codons [29-31]. Hence, the site-specific incorporation of unnatural amino acids via nonsense suppression is a useful strategy and depends on two crucial components: the presence of a unique termination codon and an orthogonal engineered/evolved suppressor tRNA

and aminoacyl-tRNA synthetase pair. The stop codon, integrated via mutagenesis at a specific site in the gene of interest, is recognized and decoded by the corresponding orthogonal suppressor tRNA resulting in the insertion of the novel unnatural amino acid (Figure 3). Hardly worth mentioning, the used termination codon for the integration of the novel building block must not coincide with the actual termination codon of the gene of interest. Since the “amber” stop codon is the least used stop codon in *E. coli*, it is the most favourable codon for this technique. Nevertheless, the utilization results in a substantial loss of fitness and affects at least 83 peptides by the stop codon read-through in *E. coli* [92]. To minimize the cross-reactivity in the translational apparatus in the host, suitable orthogonal tRNA and aminoacyl-tRNA synthetase pairs normally derive from other organisms [93]. Tremendous work on the engineering of this orthogonal tRNA and aminoacyl-tRNA synthetase have been done by Schultz and coworkers, who developed Tyr-tRNA and tyrosyl-tRNA synthetase from the archaeobacterium *Methanococcus jannaschii* [68]. Since then, several orthogonal tRNA and aminoacyl-tRNA synthetase have been successfully generated not only derived from *M. jannaschii* [47, 48, 68, 89, 94-97], but also from *S. cerevisiae* [98-101], *M. maize* [102], *M. barkeri* [103], *E. coli* [104, 105], *B. stearothermophilus* [106], and *B. subtilis* [96]. These engineered orthogonal pairs can be applied to a variety of host cells including *E. coli* [97, 98, 107, 108], *S. cerevisiae* [102, 104, 109], *P. pastoris* [110], mammalian cell lines [96, 102, 106, 111-113] and even the first multicellular organism, *C. elegans*, [114]. The incorporated unnatural amino acids harbour a great variety of functional groups including alkyne [115], azido [94, 104, 115], bromo [101], chloro [101], fluoro [95, 98, 107], iodo [70] and nitro [116] side chains. The variety of hosts, engineered orthogonal tRNA and aminoacyl-tRNA synthetase pairs and the collection of hundreds of unnatural amino acids demonstrates the modularity, versatility and capacity of possible applications and exciting new opportunities in research.

1.3.4. Frameshift Codons for Site-specific Incorporation

In contrast to stop codon suppression for site-specific incorporation of novel amino acids, frameshift suppression replaces stop codons and uses a quadruplet code utilizing tRNAs with quadruplet anticodons to integrate unnatural amino acids in response to a four-base sequences [117]. This alternative codon system offers the great advantage that the integration of unnatural amino acids is independent from the read-through of the stop codon and no competition with release factors at designated stop codon sites occurs. Quadruplet frameshift assignment has been successfully used in *E. coli* cell free-protein expression systems [118, 119] as well as insect [120] and mammalian cell-free expression systems [119]. Frameshift codon suppression is not limited to a quadruplet code as Sisido and coworkers could even develop a five-base codon [117]. In a joint approach, even frameshift and

stop codon suppression could be simultaneously applied to integrate the two unnatural amino acids L-homoglutamine, via a four-base codon, and *O*-methyltyrosine, in response to the “amber” stop codon, into myoglobin in *E. coli* [121].

1.3.5. O-ribosome/O-mRNA Pairs for Site-specific Incorporation

Similarly to orthogonal tRNAs and aminoacyl tRNA synthetases, since 2005 several orthogonal ribosome and mRNA pairs (O-ribosome/O-mRNA, Ribo-X, Ribo-Q) have been developed to further minimize competition of release factors of the host’s natural translational machinery with stop codon suppression [74, 122-124]. This was achieved by creating an O-ribosome library on the A-site of the *E. coli* ribosome to screen and isolate suitable mutants. O-ribosome/O-mRNA pairs work in parallel to the host’s natural ribosomes [122]. Ribo-X is optimized both for “amber” codon suppression and prohibition of release factor binding [123]. The latest Ribo-Q has redirected quadruplet anti-codons to recognize quadruplet codons, instead of the natural triplet codons [74].

1.3.6. Engineering Aminoacyl-tRNA Synthetases: Rational and Evolved Variants

The same translational mechanisms that guarantee the high selectivity and fidelity in protein biosynthesis also limit the variety of protein structure and function. While the incorporation of an isostructural amino acid analogue may be enhanced by simple alterations such as overexpression of the corresponding wild-type aminoacyl-tRNA synthetase [80], many other unnatural amino acids require further engineering of this evolved enzyme to expand the set of novel amino acids that can be successfully activated and incorporated. Despite the extraordinary high selectivity of aminoacyl-tRNA synthetases, which challenge the engineering of the amino acid specificity, the first engineered orthogonal pair for “amber” codon suppression was created in 2001, based on an engineered *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase and its cognate tRNA^{Tyr} with a changed CUA anticodon, evolved for the incorporation of *O*-methyltyrosine [68]. Nevertheless, the frequent use of incorporation of unnatural amino acids via stop codon suppression propagated only after discovery of the orthogonal pair for pyrrolysine in 2002 [29]. Today, most engineered aminoacyl-tRNA synthetases for incorporation of unnatural amino acids are based on *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase or pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* or *barkeri* [38, 39, 65, 125, 126]. The high evolvability and general applicability of pyrrolysyl-tRNA synthetase makes it the predominantly used orthogonal suppression system [67]. The wild-type pyrrolysyl-tRNA synthetase shows already low substrate specificity and a great acceptance to various noncanonical amino acids

and has been later engineered to incorporate several lysine analogues [125]. Due to the additional direct translatability between different host organisms of the system, this method allowed a range of protein modification based on lysine derivatives. A second-generation of pyrrolysyl-tRNA synthetase variants uses the activation of phenylalanine to encode for various phenylalanine derivatives [66, 127]. Novel pyrrolysyl-tRNA synthetase variants use histidine analogues [128] and cysteine derivatives [129].

As it is generally impossible to rationally predict which residues to change within the amino acid binding pocket to incorporate a certain noncanonical amino acid, the engineering process requires a library of mutant tRNA and aminoacyl-tRNA synthetase pairs, containing randomized residues in the amino acid-binding site, from a foreign organism, for instance archaea or eukaryotes [38, 39] (Figure 1.5). The most common method to create such a library is the site saturation mutagenesis using PCR-based methods to replace codons encoding amino acids to be randomized by the degenerate NNK or NNN codon [130]. The anticodon loops of these tRNAs have to be mutated to generate special tRNA_{SB} suppressing a blank codon that does not encode for a natural amino acid in protein biosynthesis. The engineering of directed evolution of orthogonal tRNA_{SB} and unnatural aminoacyl-tRNA synthetase pairs requires a two-step selection relying on a positive and a negative selection step for each member of the library. For positive selection, the aminoacyl-tRNA synthetase library is cotransformed with an “amber”-disrupted (codon_{BL}) positive selector, for instance the gene for the green fluorescent protein as well as an “amber”-disrupted resistance gene such as the β -lactamase gene. Cells encoding aminoacyl-tRNA synthetase variants capable of charging either the UAA or an endogenous canonical amino acid onto tRNA_{SB} survive and additionally produce GFP as selection marker. The isolated enriched library is then cotransformed with an “amber”-disrupted negative selector, e.g. the toxic gene for barnase. In the presence of the desired unnatural amino acid, only aminoacyl-tRNA synthetase variants that accept the unnatural amino acids allow “amber”-suppression, inactivating the negative selector, whereas aminoacyl-tRNA synthetase variants accepting endogenous natural amino acids will activate the negative selector resulting in cell death. The result is a highly orthogonal tRNA_{SB} and cognate aminoacyl-tRNA synthetase pair in *E. coli*. The two selection steps are iteratively cycled to enrich the desired aminoacyl-tRNA synthetase variants with aminoacylation activity for the unnatural amino acids of interest.

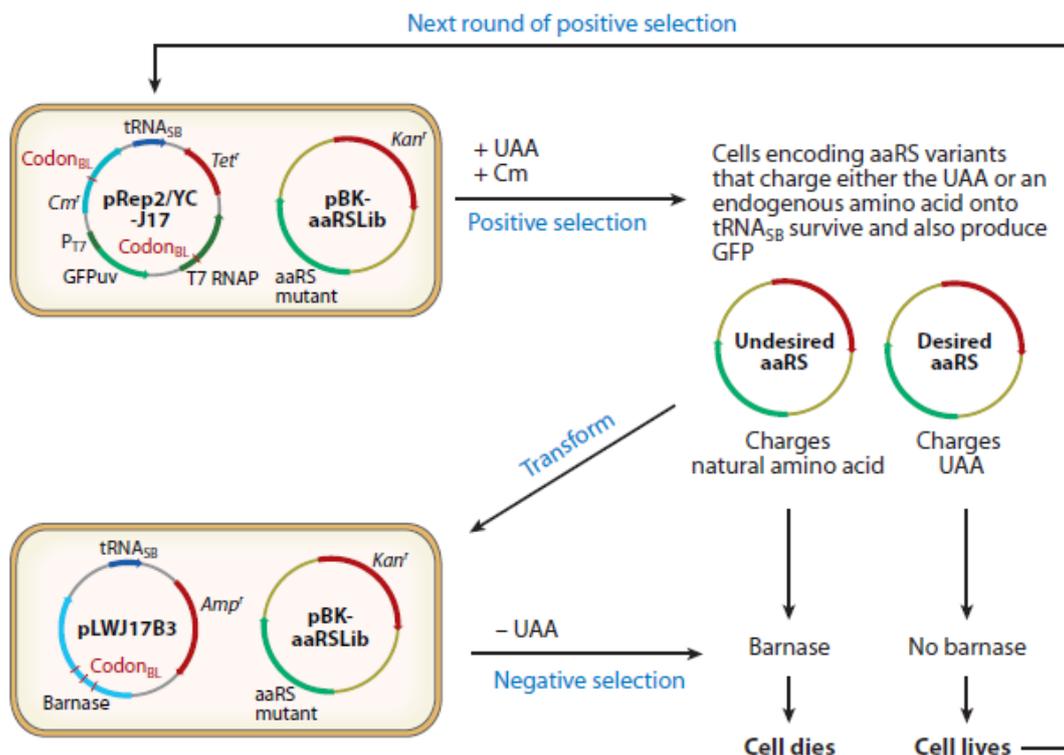


Figure 1.5: Selection schemes for engineering orthogonal aminoacyl-tRNA synthetases in *E. coli*. Engineering of directed evolution of orthogonal tRNA_{S_B} and unnatural aminoacyl-tRNA synthetase pairs require a two-step selection relying on a positive and a negative selection step. For positive selection, the aminoacyl-tRNA synthetase library is cotransformed with an “amber”--disrupted positive selector, e.g. GFP. Cells encoding aminoacyl-tRNA synthetase variants capable of charging either the UAA or an endogenous canonical amino acid onto tRNA_{S_B} survive and additionally produce GFP as selection marker. The isolated enriched library is then cotransformed with an “amber”--disrupted negative selector, e.g. the toxic gene for barnase. In the absence of the desired unnatural amino acid, only aminoacyl-tRNA synthetase variants that accept the unnatural amino acids allow “amber”-codon suppression, inactivating the negative selector, whereas aminoacyl-tRNA synthetase variants accepting endogenous natural amino acids will activate the negative selector resulting in cell death. The two selection steps are iteratively cycled to enrich the desired aminoacyl-tRNA synthetase variants with aminoacylation activity for the unnatural amino acids of interest. Abbreviations: aaRS, aminoacyl-tRNA synthetase; codon_{BL}, blank codon; GFP_{UV}, green fluorescent protein; P_{T7}, bacteriophage T7 promoter; T7 RNAP, bacteriophage T7 RNA polymerase; tRNA_{S_B}, suppressor tRNA; UAA, unnatural amino acid (Modified from Liu & Schultz, 2010).

1.4. Designing Novel Proteins by genetically UAA Incorporation

Although 20 common amino acids are sufficient for all known forms of life, altering proteins using genetically encoded unnatural amino acids to seamlessly integrate new chemistries into biology can provide useful insights into the chemical and physical properties of proteins. Many groups around the world have therefore used unnatural amino acid incorporation to engineer the photophysical properties of recombinant proteins for a number of applications. This section discusses the utility of

unnatural amino acid incorporation for applications in protein structure and function research, highlights various ways in the widespread use of this technique leading to novel therapeutic proteins as well as protein evolution with an expanded genetic code.

A great variety of novel probes have been introduced into proteins through genetic code expansion including redox probes to study electron transfer processes, spin labels for electron paramagnetic resonance, isotopically labelled amino acids for nuclear magnetic resonance spectroscopy, spectroscopic handles for infrared spectroscopy, heavy atom-containing amino acids for X-ray crystallography, photo-crosslinkers for mapping transient protein-interactions, small fluorescent amino acids as direct probes of local structure imaging, novel side chains (including keto, azido, alkenyl and alkynyl groups) participating in bio-orthogonal reactions as well as optical probes for spectroscopy and *in vivo* super-resolution imaging [38, 39, 131, 132]. Combining genetically encoded incorporation of unnatural amino acids with other labelling methods even further enhances potential applications including the recently enabled creation of an *in vitro* metabolic Förster resonance energy transfer sensor with improved dynamic range [133].

The expanded genetic code and its unnatural amino acid mutagenesis is beginning to find many applications not only in protein research, but also in protein therapeutics with the ability to site-specifically modify cytokines, growth factors, antibodies as well as antibody domains with unnatural amino acids bearing bio-orthogonal reactivity opens new field for generating therapeutic proteins that are precisely derivatized with diverse molecules [134]. This includes polyethylene glycosylated proteins, antibody–drug conjugates, antibody–antisense conjugates and bi-specific antibodies for cancer and inflammatory, autoimmune and metabolic diseases [134-136].

From an evolutionary viewpoint, the unnatural mutagenesis methodology can easily be adapted to proteins with novel or enhanced function as the template-based synthesis of polypeptides directly links the information (genome) to the phenotype (proteome). Therefore, phage-based evolution experiments for the use with 21 amino acid organisms were developed [137, 138]. These systems use phage-displayed protein libraries generated in *E. coli* to encode 21 amino acids creating *X-E. coli* strains. The resulting library members contain an “amber” termination codon in a gene of interest as well as the corresponding unnatural amino acid in the specific phage-displayed protein. The phages containing the altered proteins are then subjected to selection, and the surviving phage clones are used to reinfect *X-E. coli* strains. The repeated cycle of mutation, amplification, and selection might result in the discovery of new proteins with novel properties through unnatural amino acid incorporation.

AIM OF THE THESIS

The aim of this study was to investigate and utilize the incorporation of novel unnatural amino acids into (membrane proteins) with the objective to use them as photo-caged probes to investigate protein dynamics and conformational changes in serial X-ray crystallography, as infra-red probes for the analysis of local structural heterogeneity in proteins, and as light-activated crosslink probes to study protein-protein interactions. In this work, three different model proteins were used: the green fluorescent protein, the integral outer membrane protein X and recombinant T4-lysozyme.

GFP, originally discovered in the jellyfish *Aequorea Victoria* [139], and its engineered derivatives are often used as reporter and model proteins due to their unique spectroscopic features [140, 141]. Their spectral characteristics can be controlled and modified by irradiating them with light of a specific wavelength, intensity, and duration, leading to a fluorescence state [142]. The protein forms a cylindrical barrel structure of eleven β -sheets interconnected by loop regions, with an α -helix harbouring the chromophore spanning straight through the barrel centre [162]. Due to its ability to auto-catalytically form its highly visible, efficiently emitting internal chromophore [143-145], GFP has become a well-established and powerful tool for many applications in cell and molecular biology [146]. As time-resolved structural biology is the key to understand the mechanisms and principles of biological processes [147] we wanted to develop and apply a GFP-based system that can be used for simultaneous time-resolved structural and spectroscopic studies utilizing photocaged unnatural amino acids as triggers.

The integral outer membrane protein X (OmpX) of *Escherichia coli* belongs to the family of outer membrane proteins [148] and consists of an 8-stranded antiparallel β -barrel which shows two girdles of aromatic amino acid residues and a ribbon of nonpolar residues that attach to the membrane interior [149, 150]. The core of the barrel consists of an extended hydrogen-bonding network building up an inverse micelle. OmpX seems to promote bacterial adhesion via elongated β -strands elongated into the extracellular space and bacterial cell internalization to mammalian cells and confer resistance against the human complement system [149]. Understanding the function and biochemical interactions of OmpX structure, as well as factors determining protein interaction with its surrounding

lipid environment, is critical to decipher its role in virulence and pathogenicity. Therefore the photo-crosslinker para-benzoyl-L-phenylalanine was introduced into OmpX at probable binding sites at the extracellular β -strands to identify protein interaction partners *in vivo*.

Lysozyme, accidentally found in 1922 by Alexander Flemming during a deliberate search for medical antibiotics [151], is a small and stable enzyme, making it an ideal system for research into protein structure and function. Lysozyme is a cornerstone of innate immunity and plays a crucial role in cell wall hydrolysis of pathogens as well as the modulation of the host immune response to infection [152]. Today, it is a well-characterized model protein system in research and was due to its characteristics chosen for the analysis of the effects of infrared labels on native protein structure and dynamics. These studies aim to elucidate how vibrational labels can be used to obtain information about local heterogeneity and assess changes in protein structure and dynamics due to incorporation and labelling with unnatural amino acids by simple one-dimensional infrared spectroscopy.

SECTION B

GENETICALLY ENCODED SITE-SPECIFIC PROTEIN LABELLING

CHAPTER 2

2. PHOTOACTIVATABLE GFP

Parts of this work were accomplished in collaboration with Dr. Johannes Heidemann & Dr. Knut Kölbel (Utrecht Lab, HPI, Germany), Dr. Briony Yorke (Pearson Lab, University Hamburg, Germany) & Anke Puchert (Huse Lab, University Hamburg, Germany).

2.1 Introduction

Photoactivatable fluorescent proteins (PA-FPs) are molecular probes exhibiting naturally fluorescence. Their spectral characteristics can be controlled and modified by irradiating them with light of a specific wavelength, intensity, and duration, leading to a fluorescence state [142]. Inactivated PA-FPs are non-fluorescent, whereas activated PA-FPs yield bright signals in contrast to a dark background. These properties allow the spatial and temporal labelling of specific subcellular structures and thus enable the visualization of protein, organelle, and cell dynamics making PA-FPs a powerful non-invasive tool for high-resolution live-cell imaging [153] [154]. Various PA-FPs have been engineered, especially proteins of the green fluorescent protein (GFP) family, with altered properties such as shifted emission wavelengths, enhanced fluorescence and higher quantum yields [153] [155] [156] [157] [158] [159]. GFP, originally discovered in the jellyfish *Aequorea victoria*, exhibits bright green fluorescence when exposed to light in the blue to ultraviolet range [160]. The eleven β -strands form a cylindrical β -barrel with the α -helix harbouring the chromophore spanning straight through the barrel centre (Figure 2.1) [161] [162]. Due to its ability to auto-catalytically form its highly visible, efficiently emitting internal chromophore [143] [144] [145], GFP has not only become a well-established and powerful tool for many applications in cell and molecular biology [146] such as reporter genes for the *in vivo* detection of gene expression [163], fusion tags to monitor protein localization [144], pH biosensors [164], and Ca^{2+} sensors [165] [166], but also a well-characterized model protein.

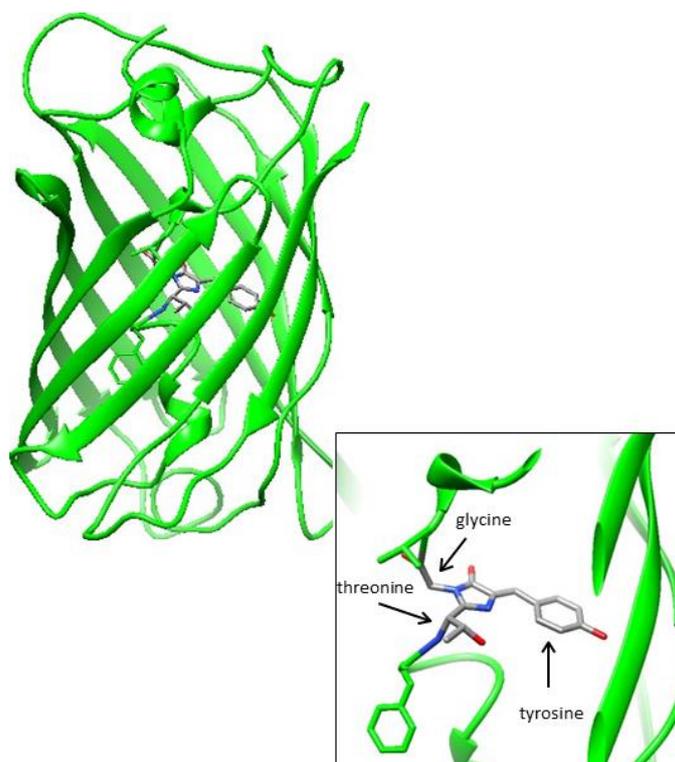


Figure 2.1: Green fluorescent protein. The protein backbone forms a cylindrical β -barrel with the α -helix harbouring the chromophore spanning straight through the barrel centre. The tight packing of the protein allows the shielding of the chromophore from the surrounding environment. The internal chromophore (shown in the close-up) is auto-catalytically formed by the three central amino acids glycine, tyrosine and threonine creating an unusual five-membered ring (pdb: 1ema).

The most commonly used radiation sources for high-intense X-ray experiments are currently synchrotrons. First generation synchrotrons with an electron storage ring have been built in the 1970s. Nowadays synchrotrons of the third generation are in use providing a multitude of energy and potential applications in many research fields. In parallel, X-ray sources of the fourth generation are currently constructed at several locations worldwide. These new third and fourth generation radiation sources allow new applications in the field of time-resolved structural biology such time-resolved X-ray diffraction or time-resolved small-angle X-ray scattering as well as serial femtosecond X-ray crystallography [167-169]. Time-resolved structural biology is a central key to understand the mechanisms of biological processes [147]. In order to perform time-resolved experiments an ensemble of molecules first needs to be populated/trapped in a defined state. Conformational changes can then be triggered by means of rapid mixing or light irradiation [147] [170] [171] [172]. The use of light in principle allows observations in the ps-regime if the decaying reaction is fast enough. Currently, time-resolved X-ray crystallography is the technique providing the greatest spatial and temporal resolution. However, the requirement for well diffracting crystals still limits its application. Although great

advances have been made in this field, crystallization remains an art rather than an exact science [147]. The specific intrinsic fluorescence makes GFP a perfect tool to study dynamics with time-resolved techniques. In order to develop and apply a GFP-based system that can be used for simultaneous time-resolved structural and spectroscopic studies, a photoactivatable GFP variant, originally designed by Schultz and coworkers [173], that contained the photocaged tyrosine analogue *ortho*-nitrobenzyl-tyrosine (ONBY) was further investigated [116] [174]. Photocaged unnatural amino acids are synthetically modified amino acids which can be controlled by light, usually by photolytic conversion from an inactive to an active form [175]. The most common photocaging groups are *ortho*-nitrobenzyl-groups and their derivatives, which can be readily cleaved upon irradiation with 365 nm light [176]. The incorporation of photocaged ONBY in *E. coli* is achieved by genetically encoding it using an “amber” stop codon in combination with an orthogonal suppressor tRNA / aminoacyl-tRNA synthetase pair as mentioned before [116]. Incorporation of ONBY at residue 66, a key residue of the chromophore, leads to a non-fluorescent (“dark-state”) GFP molecule. Upon irradiation with light, the photocaging *ortho*-nitrobenzyl-group dissociates and the fluorescent state is re-established (Figure 2.2). However, the structure of this photocaged GFP variant, the decaging mechanism, as well as its suitability for time-resolved structural studies are still not understood.

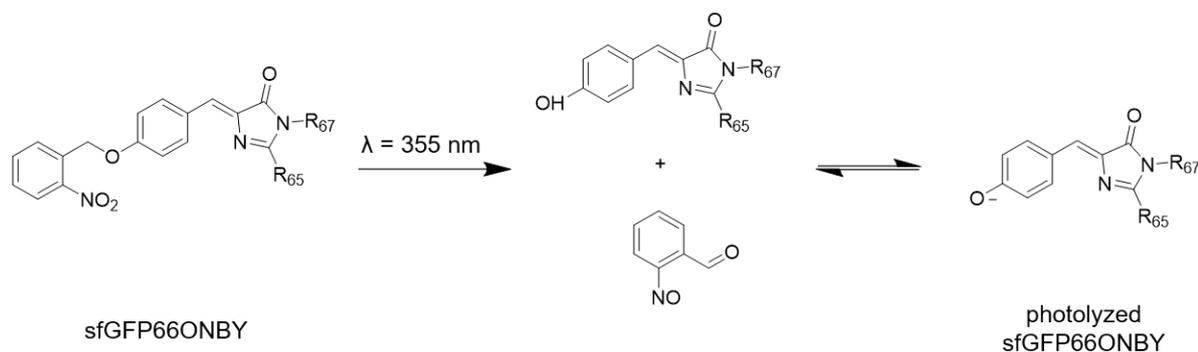


Figure 2.2: Decaging of the non-fluorescent GFP66ONBY by photolysis results in a reestablished fluorescent protein.

2.2 Materials and Methods

2.2.1 *Ortho*-nitrobenzyl-tyrosine (ONBY) synthesis

The protocol for the synthesis of ONBY has been initially established by Henry A. Lester and coworkers [177] and has been adapted. L-Tyrosine (2.0 g, 11.0 mmol) was dissolved in 2 M NaOH aq. (10 ml), and CuSO₄·5H₂O (1.9 g, 7.28 mmol), dissolved in a minimal amount of water, was added slowly at room temperature (RT). The solution was heated to 60 °C and stirred for 20 min and then allowed to cool to RT before adjusting to pH = 7 using 1 M HCl. The light-blue solid was filtered and washed three times with 25 ml water, before it was suspended in 75% aqueous dimethylformamide (60 ml). K₂CO₃ (1.5 g, 11.04 mmol) and *ortho*-nitrobenzyl bromide (1.8 g, 8.49 mmol) were added and the reaction was allowed to proceed for 72 hours at RT while kept in the dark. The solid was filtered, washed with 75% aqueous dimethylformamide (40 ml x 2), water (40 ml x 2), 75% aqueous acetone (40 ml), and ice-cold acetone (10 ml), and then suspended in 1 M HCl (100 ml) to stir for 2 hours at RT. The white solid was filtered and stirred once more with fresh 1 M HCl (100 ml) for another 30 min. The solid was finally filtered, washed with water (40 ml x 2) and ice-cold acetone (10 ml), and dried to give *ortho*-nitrobenzyl-tyrosine as an off-white solid (1.85 g, 68%).

2.2.2 Cloning, protein expression and purification

The gene for wild-type ‘superfolder GFP’ [178] was a gift from Scott Gradia (Addgene plasmid #29663). The gene was sequenced prior usage. To produce photoactivatable *ortho*-nitrobenzyl-photocaged sfGFP (sfGFP66ONBY), the stabilized GFP variant ‘superfolder GFP’ was cloned into a pET28a vector with an N-terminal fusion consisting of a His₆-tag and a TEV protease cleavage site. For incorporation of ONBY at residue position 66, tyrosine 66 was mutated to the “amber” codon, TAG, via site-directed mutagenesis, to generate pET28a-sfGFP66TAG. This plasmid was cotransformed into *E. coli* BL21 (DE3) Gold cells together with the plasmid pEVOL-ONBY, containing the orthogonal aminoacyl tRNA synthetase (aaRS) and an amber suppressor tRNA, both derived from the *Methanococcus jannaschi* tyrosyl tRNA/aaRS pair [56] [116]. To produce the sfGFP66ONBY-F145A variant, phenylalanine 145 was mutated to alanine by site-directed mutagenesis. This was followed by the cotransformation of this plasmid into *E. coli* BL21 (DE3) Gold cells together with the plasmid pEVOL-ONBY, containing the orthogonal aminoacyl tRNA synthetase (aaRS) and an amber suppressor tRNA, both derived from the *Methanococcus jannaschi* tyrosyl tRNA/aaRS pair [56] [116]. A single colony of the cells, grown on a LB-agar plate with kanamycin (25 µg/ml) for the wild-type sfGFP, with kanamycin (25 µg/ml) and

chloramphenicol (34 µg/ml) for the sfGFP66ONBY and sfGFP66ONBY-F145A variants overnight at 37 °C, was used to inoculate a 20 ml preculture with LB-medium containing the same antibiotics. The preculture was grown for 16 hours at 37 °C at 180 rpm and subsequently used to inoculate the expression culture (1L Terrific Broth medium at 37 °C und 180 rpm, supplemented with the same antibiotics selection). Cells were grown at 37 °C to an $OD_{600} = 0.9$ in Terrific Broth medium and shifted to 20 °C. ONBY was added (for sfGFP66ONBY and sfGFP66ONBY-F145) to a final concentration of 1 mM and arabinose to a final concentration of 0.02 % for induction of the tRNA/aaRS pair, and protein production was induced by the addition of 1 mM IPTG with a delay of 60 min. After 20 h, the cells were harvested by centrifugation at 4000 x g for 20 min. The cell pellet was resuspended in lysis buffer (40 mM TrisHCl, pH 8.0, 300 mM NaCl, 5 % glycerol), supplemented with a 0.5 µl lysozyme (100 mg/ml) (*Gallus gallus*, Sigma Aldrich, Germany). The cells were disrupted twice by sonication on ice for three minutes each (30 % duty cycle). The lysate was centrifuged at 4 °C at 20 000 x g for 30 minutes to remove cell debris and the protein was obtained in the supernatant. Metal ion affinity chromatography using a Ni-NTA column with gravity flow was applied as a first purification step. All steps were performed at room temperature with ice-cold buffers. The column was equilibrated with lysis buffer and the cleared lysate was loaded on the matrix. The flow-through was collected and the resin was washed three times with 20 ml lysis buffer supplemented with 20 mM imidazole to remove unspecifically bound protein. The protein was eluted in several steps with 10 ml elution buffer containing 50 mM TrisHCl, pH 8.0, 200 mM imidazole. The elution fractions were analysed by SDS-PAGE and suitable fractions were combined. The combined fraction was dialyzed overnight against imidazole free elution buffer at 4 °C. Subsequently, the sample was concentrated to a suitable volume and a second purification step using a size exclusion chromatography (SEC) was applied and operated by an ÄKTA_{pure} (GE Healthcare, USA). For the SEC a Superdex 75 (GE Healthcare, USA) column was used at 4 °C and equilibrated with 40 mM Tris, pH 8.0. SEC was used to separate the monomer and dimer species from aggregates. Finally, the purified protein was concentrated using an Amicon-15 centrifugal filter concentrator (Merck Millipore, Germany) with a MWCO = 10 kDa. Protein identity and purity were assessed using mass spectrometry and SDS-PAGE as described in Appendix – General Sample Preparation Techniques. For all proteins the final protein concentration was determined using a Nanodrop 2000 (Thermo-Scientific, Germany).

2.2.3 Crystallization and data collection

For the crystallization of the purified GFP66ONBY variant the protein solution was concentrated to a range of 8 - 11 mg/ml by using a Amicon-15 centrifugal filter concentrator (Merck Millipore, Germany)

with an MWCO =10 kDa. The protein solution was centrifuged for 15 minutes at 15,000 x g prior usage for crystallization experiments. For the crystallization experiments 48-well SWISSCI MRC2 plates were used (Molecular Dimensions, USA). Initial crystallization conditions were screened using commercially available matrices from Hampton Research and Molecular Dimensions using the vapour-diffusion technique in sitting drops. Sitting drops containing 2 μ l of protein solution and 2 μ l of precipitant solution were equilibrated at 20 °C against 1 ml of the same buffer in the reservoir. The plates were sealed, stored at room temperature and regularly inspected for crystal formation using a microscope. Crystals in nylon loops were mounted and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K at the MASSIF1 beamline at ESRF, Grenoble, France at fixed energy of 12.835 keV (0.966 Å). Full datasets with an oscillation angle of 0.1° per image were collected to a resolution limit of 2.7 Å. All datasets were processed with XDS [179] and merged with POINTLESS [180].

2.2.4 Structure determination, refinement, and analysis

Structures were solved by molecular replacement with PHASER [181] using pdb:1ema [161] as search model. Building of the photocaged chromophore was achieved using *phenix.eLBOW* [182]. Subsequent rounds of manual building using *Coot* [183] and refinement using *phenix.refine* [184] allowed complete model building. Molecular graphics images were produced using *UCSF Chimera* [185]. The structural data have been deposited in the Protein Data Bank with accession code pdb: 6H01.

2.2.5 Spectroscopy

Spectroscopy was performed in collaboration with Dr. Briony Yorke (Pearson Lab, University Hamburg, Germany) and Anke Puchert (Huse Lab, University Hamburg, Germany). UV/Vis absorption spectra of sfGFP66ONBY during pulsed 355-nm irradiation were measured using a QE Pro spectrometer and a DH-2000 light source (both from Ocean Optics). A Q-switched diode-pumped Nd:YAG laser (Standa Q1TH) operating at 1 kHz and producing 35 mW average power of 355 nm light was used to initiate photo-dissociation of the ONBY group. The laser was continuously pulsed with pulse durations of 750 ps and a beam size of 600 microns FWHM. 1000 μ l protein samples (30 μ M protein in 40 mM Tris, pH 8.0) was used in a rectangular quartz cuvette (4 mm x 10 mm lateral dimensions) without stirring with excitation along the short cuvette axis and absorption/fluorescence measurements in orthogonal direction. The integration time of each absorption spectrum was 22 ms. Fluorescence spectra of photolysed sfGFP66ONBY were recorded with an integration time of 100 ms using an Ocean Optics QE Pro spectrometer via a quartz lens and a UV grade optical fibre with a diameter of 400 microns.

Samples were photolysed for 30 min before fluorescence spectra were recorded. The Standa-Q1-TH laser was used in the operating mode described above for both, sfGFP66ONBY photolysis and fluorescence spectroscopy.

2.2.6 Native Mass Spectrometry

Native mass spectrometry was performed by Dr. Johannes Heidemann and Dr. Knut Kölbl (Utrecht Lab, HPI, Germany). Proteins were exchanged into 50 mM ammonium acetate pH 8.0 using centrifugal filter units (Vivaspin 500, 10k MWCO, Sartorius, Germany). Samples were filled into ESI capillaries that were prepared as follows: Borosilicate capillaries (1.2 mm OD, 0.68 mm ID, with filament, World Precision Instruments, Germany) were produced with a micropipette puller (P-1000, Sutter Instruments, USA), equipped with a squared box filament (2.5 x 2.5 mm, Sutter Instruments, USA). Capillaries were gold-coated using a sputter coater (Q150R, Quorum Technologies, United Kingdom, 40 mA, 200 s, tooling factor 2.3, and end bleed vacuum 8×10^{-2} mbar). Samples were analysed in positive ion mode on a QToF2 (Waters, USA and MS Vision, Germany) modified for high mass experiments [186]. Data were acquired with 7 mbar source pressure and 1.5×10^{-2} mbar argon as collision gas, 1.3 kV capillary voltage and 100 V sample cone voltage. MS measurements were performed with 5 μ M protein concentration and spectra were acquired with 30 V collision energy.

2.3 Results and Discussion

2.3.1 Recombinant expression and purification

In order to investigate the suitability of GFP variants for time-resolved structural and spectroscopic studies, a superfolder GFP (sfGFP) variant was generated, which had incorporated the genetically encoded unnatural amino acid ONBY at position 66, replacing the central tyrosine residue of the chromophore. The photocaged sfGFP66ONBY was expressed recombinantly in *E. coli* BL21 (DE3) cells for 16 hours. An expression profile before and after induction of expression is shown in Figure 2.3 A. The sfGFP66ONBY variant protein was purified from the cleared lysate by Ni-NTA affinity chromatography and the resulting elution fraction contained around 90 % pure protein, as can be seen in the SDS-PAGE in Figure 2.3 B. No significant amount of sfGFP66ONBY is observed in the flow through (FT) and the both washing steps (W1 + W2). In the elution fraction several bands are visible, revealing that the sample contained low amounts of impurities. Subsequently, a second purification

step was performed by size exclusion chromatography to remove the remaining impurities. The resulting chromatogram is shown in Figure 2.4 A (blue curve). For comparison, a chromatogram of the wild-type sfGFP is presented (green curve). In the chromatogram the first peak represents the void fraction containing aggregated protein as well as other impurities whereas both other distinct peaks could be assigned to monomeric and dimeric species after calibration of the size exclusion column. Compared to the wt-sfGFP, sfGFP66ONBY shows an increased height of the void peak and a slightly reduced monomer to dimer ratio. For further experiments of sfGFP66ONBY, the central fractions for the monomeric species under the third peak were combined and further concentrated. Without heat treatment, the protein migrates as a dimer on an SDS-PAGE gel (Figure 2.4 B).

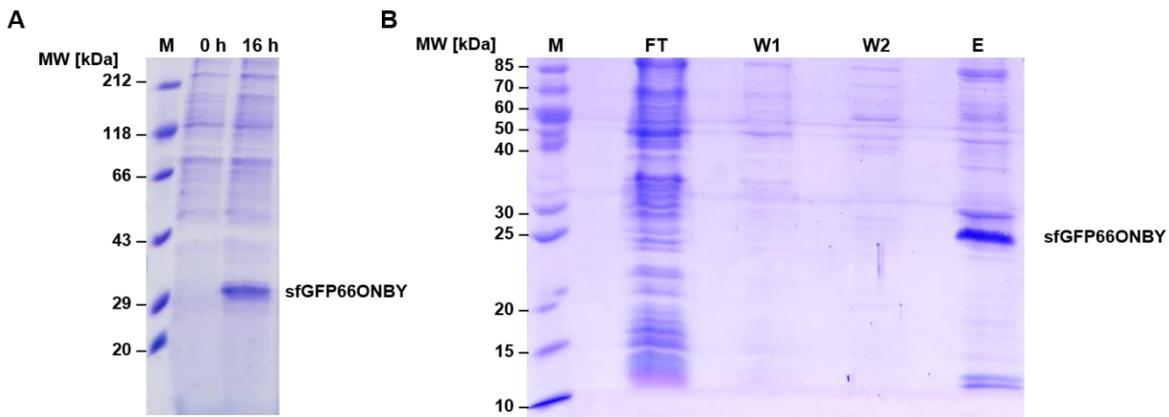


Figure 2.3: SDS-PAGE showing the expression of sfGFP66ONBY and fractions from its purification. A. Expression profile of sfGFP66ONBY in *E. coli* BL21(DE3) Gold cells before and 16 hours after induction of expression by IPTG at 20 °C (M = molecular weight marker). **B.** Elution fractions of the first purification step, the Ni-NTA affinity chromatography (M = molecular weight marker, FT = flow through, W1/2 = washing step 1 + 2, E = elution).

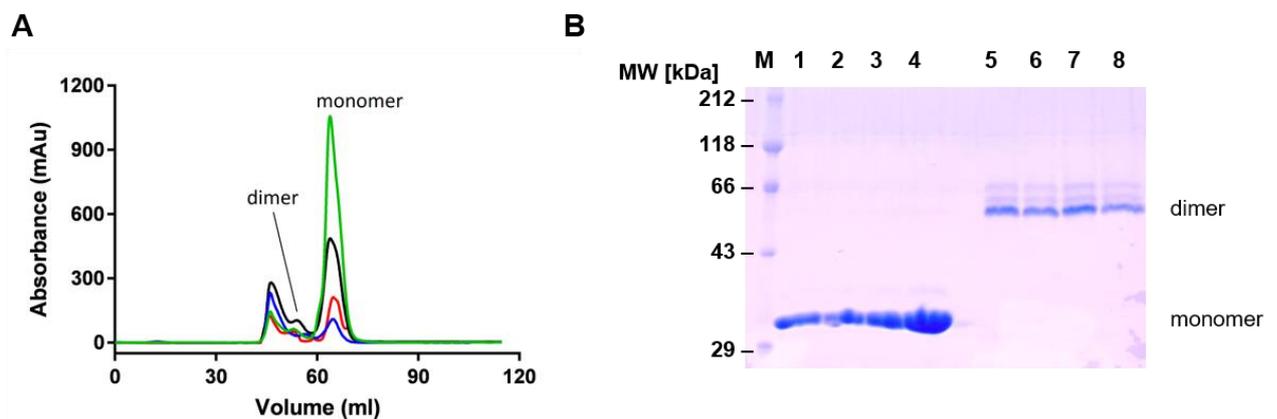


Figure 2.4: SDS-PAGE and SEC analysis of sfGFP variants. A. wt-sfGFP (green), sfGFP66ONBY (blue), sfGFP66ONBY-F145A (red) and decaged sfGFP66ONBY (black) predominantly elute as monomers from SEC **B.** Without heat treatment wt-sfGFP, sfGFP66ONBY, sfGFP66ONBY-F145A and decaged sfGFP66ONBY migrate as dimers (54 kDa) in the SDS-PAGE gel (5 – 8). In contrast, after heat treatment (1 – 4) proteins migrate as monomer (27kDa) (M = molecular weight marker, 1 + 5 = wt-sfGFP, 2 + 6 = sfGFP66ONBY, 3 + 7 = sfGFP66ONBY-F145A, 4 + 8 = decaged sfGFP66ONBY).

2.3.2 Spectroscopic and hydrodynamic properties of sfGFP66ONBY

The wt-GFP chromophore is formed by the autocatalytic cyclization of the Ser/Thr 65-Tyr 66-Gly 67 tripeptide [161] [162], which is associated to an intramolecular hydrogen-bonding network linking the two ends of the chromophore in the protein pocket (Figure 2.5). Upon excitation with light of wavelength 355 nm, an excited-state proton transfer (ESPT) is induced which enables structural transitions and involves multiple proton transfers through the hydrogen bonding network resulting in fluorescence [187] [188]. In contrast to wt-GFP, the sfGFP66ONBY variant protein is non-fluorescent. The bulky *ortho*-nitrobenzyl-group most likely blocks the proton-wire, resulting in a non-functional chromophore [189]. Moreover, it has been reported, that *ortho*-nitrobenzyl-groups quench excited fluorophores through a photo-induced electron transfer from the S1 excited chromophore to the *ortho*-nitrobenzyl-group upon irradiation [190] [191].

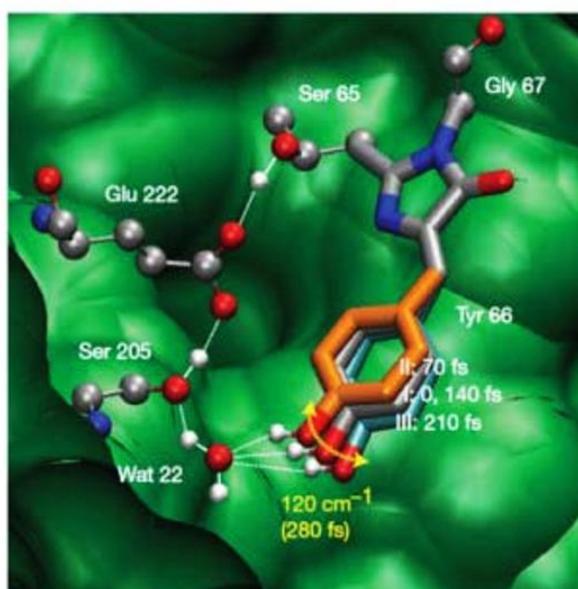


Figure 2.5: Excited-state proton transfer (ESPT) in GFP. The chromophore is associated to an intramolecular hydrogen-bonding network where the hydroxyl-group of Tyr 66 is linked to a water molecule (Wat 22), the side-chain hydroxy-group of Ser/Lys 205, the carboxy-group of Glu 222, and the sidechain hydroxyl-group of Ser 65. Upon excitation with light of wavelength 365 nm, the ESPT is induced and the proton from Tyr 66 is transferred to Glu 222. This enables structural transitions assigned to ESPT and involves multiple proton motions through the proton wire resulting in fluorescence (Adopted from Fang et al., 2009 [189]).

The decaying characteristics of sfGFP66ONBY were investigated by absorption spectroscopy. Upon photolysis with 355 nm pulsed light, the UV/Vis absorption spectrum exhibits a loss of the absorption centred at 390 nm for the caged protein while a new absorption band centred at 488 nm for the photolysed protein appeared (Figure 2.6 A, 2.6 B). Photolysed sfGFP66ONBY shows an absorption spectrum similar to wild-type sfGFP. Additionally, fluorescence spectra of photolysed GFP66ONBY

using 355 nm excitation and wild-type sfGFP are similar with fluorescence maxima at 509 nm and 517 nm respectively, indicating a shift to higher energy by 18 nm (Figure 2.6 C). A pronounced shoulder at 441 nm is also present in both spectra. The similarity of absorption and fluorescence spectra between photolysed sfGFP66ONBY and wild-type GFP confirm previous measurements and dissociation of the *ortho*-nitrobenzyl group from the mutant [20].

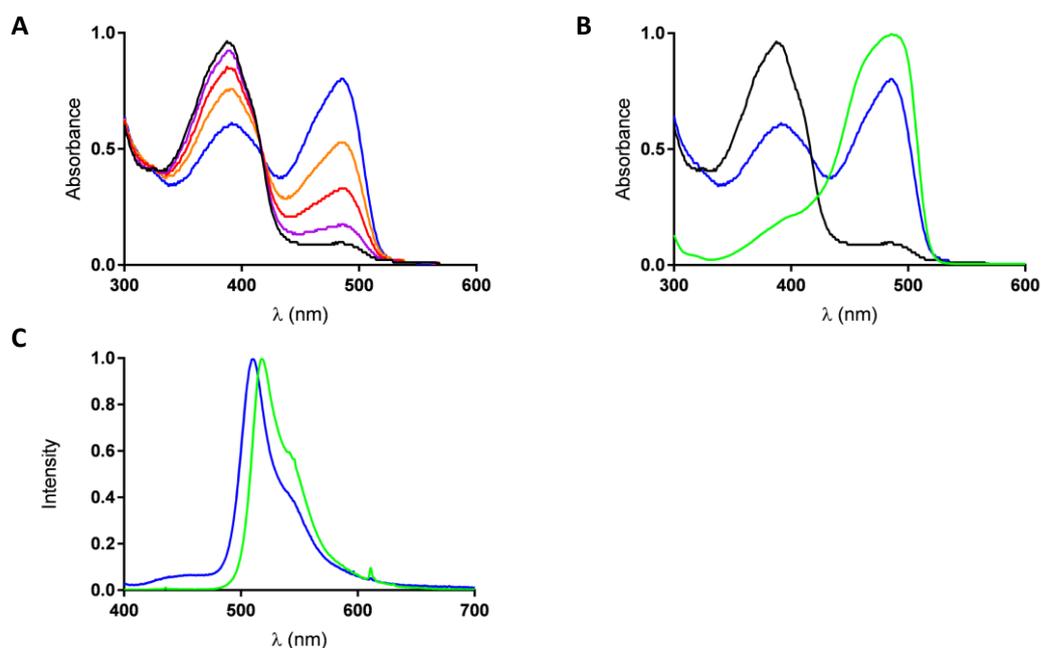


Figure 2.6: Spectroscopic characterization of sfGFP66ONBY. Spectra of sfGFP66ONBY before and after photolysis were multiplied by 0.8 for better comparison with wt-sfGFP **A.** Evolution of UV/vis absorption spectra of sfGFP66ONBY during 355-nm pulsed irradiation. Samples were irradiated for 0 min (black), 1 min (purple), 4 min (red), 9 min (orange), and 30 min (blue). **B.** UV/vis absorption spectra of sfGFP66ONBY before (black) and after photolysis (blue) and wt-sfGFP (green). **C.** Fluorescence emission spectra of photolysed sfGFP66ONBY (blue) and wt-sfGFP (green) after pulsed excitation at 355 nm. For B) and C) samples were irradiated with 355-nm pulses at 1 kHz and 35 mW average power for 30 min.

2.3.3 Crystallization, data collection and refinement of dark-state sfGFP66ONBY

To visualize the chemical environment of the caged chromophore, the dark-state (photocaged) sfGFP66ONBY was crystallized. Crystallization trials were performed using commercially available crystallization screens according to the protocol described in section 2.2.3. Protein crystals formed after 6 – 8 weeks in several crystallization conditions (see Figure 2.7). Successful crystal formation has been observed in crystallization experiments, which have been set up with a protein concentration of 8 - 11 mg/ml. Crystals were colourless and possessed cubic or rhombic shapes. Best diffracting crystals

were obtained in a condition containing 1.1 M tri-Na-citrate, 0.1 M Na-HEPES pH 7.5 using 11 mg/ml protein (Figure 2.7). For X-ray diffraction data collection, crystals were mounted in a nylon loop, flash cooled in liquid nitrogen at 100 K and mounted at the MASSIF1 beamline at the European Synchrotron Radiation Facility. Diffraction was recorded and the structure was determined to a resolution of 2.7 Å. The GFP66ONBY crystal belonged to space group $P4_12_12$ with unit-cell dimensions $a = 155.3$, $b = 155.3$, $c = 162.0$, $\alpha = \beta = \gamma = 90^\circ$. A summary of the data collection statistics is given in Table 2. The final model yielded crystallographic R factors of 0.20/0.24 ($R_{\text{work}}/R_{\text{free}}$) with 97% of residues falling within the Ramachandran favoured region and no outliers in disallowed regions, and a MolProbity [192] clashscore of 4.34. Molecular graphics images were produced using the *UCSF Chimera package* [185]. The structural data have been deposited in the Protein Data Bank with accession code pdb: 6H01.

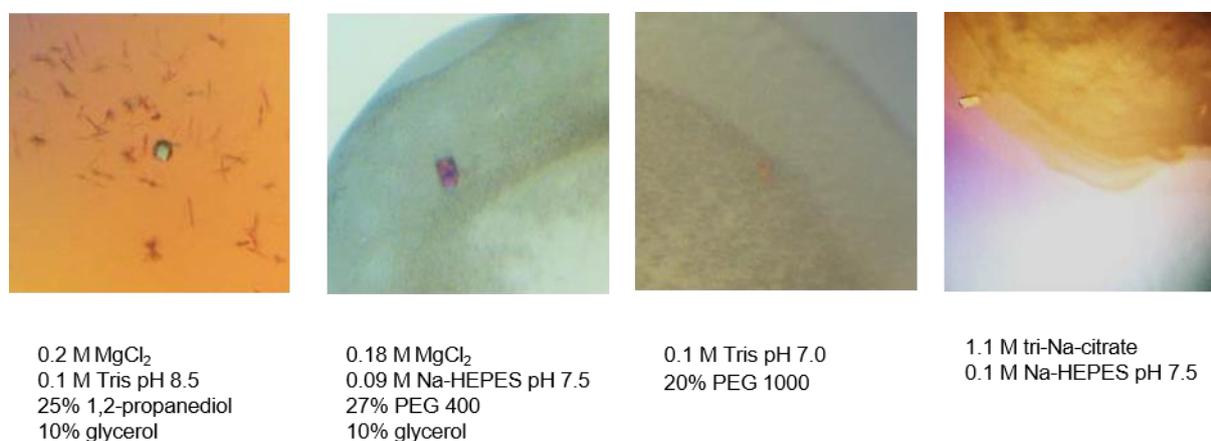


Figure 2.7: Exemplary crystals of GFP66ONBY. Crystallization conditions were screened using the vapour-diffusion technique in sitting drops. Sitting drops containing 2 μ l of protein solution and 2 μ l of precipitant solution were equilibrated at 20 °C against 1 ml of the same buffer in the reservoir.

Unexpectedly, the dark-state sfGFP66ONBY structure revealed a domain-swapped dimeric arrangement with residues 1- 143 (domain 1) of the first molecule associating with the residues 148 – 235 (domain 2) of the second molecule (Figure 2.8 A). Clear electron density identified the connecting loops ranging from residues Tyr143 to His148 (Figure 2.8 C), while no electron density was present in the position corresponding to β -strand β_7 in the search model (as previously observed by Groff *et al.* [173]). The electron density extends away from the search model, leading to the second monomer where the protein backbone is also shifted due to the steric interference of the bulky *ortho*-nitrobenzyl-group with the normal tight packing of the β_7 -strand against the chromophore. The pep-

Table 2: Data collection and refinement statistics for dark-state sfGFP66ONBY (pdb:6H01)

Data collection	
Space group	P 4 ₁ 2 ₁ 2
Cell dimensions	
α, b, c (Å)	155.3, 155.3, 162.0
α, β, γ (°)	90, 90, 90
Resolution (Å)	90.93 - 2.70 (3.34 – 2.70)
R_{merge}	0.055 (0.609)
$I / \sigma I$	9.7
CC(1/2)	0.997 (0.546)
Completeness (%)	99.90 (99.87)
Multiplicity	2.0 (2.0)
Refinement	
Resolution (Å)	49.32 - 2.70 (2.78-2.70)
No. reflections	54873 (5387)
$R_{\text{work}} / R_{\text{free}}$	0.201/ 0.239 (0.35/0.37)
No. atoms	
Protein	7273
Chromophore	128
Water	133
Average B-factors (Å ²)	
Protein	47.14
Chromophore	44.28
Water	48.20
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.37
Ramachandran plot analysis	
Most favoured region [%]	96.7
Allowed regions [%]	3.2
Generously allowed regions [%]	0.1

*Values in parentheses are for the highest-resolution shell.

side chains of the β_7 -strand are reoriented and the bulky side chain of Phe 145 is rotated by 180° to the outside of the barrel. These backbone rearrangements lead to the exchange of half of the structure of one monomer with the other monomer, as the backbone of each monomer folds back on itself between residues 143 and 148 to form intertwined domain-swapped dimers (Figure 2.8 A). The electron density clearly shows the presence of the *ortho*-nitrobenzyl group and the cyclized backbone of the chromophore (Figure 2.8 B) indicating that the *ortho*-nitrobenzyl-group allows proper chromophore formation.

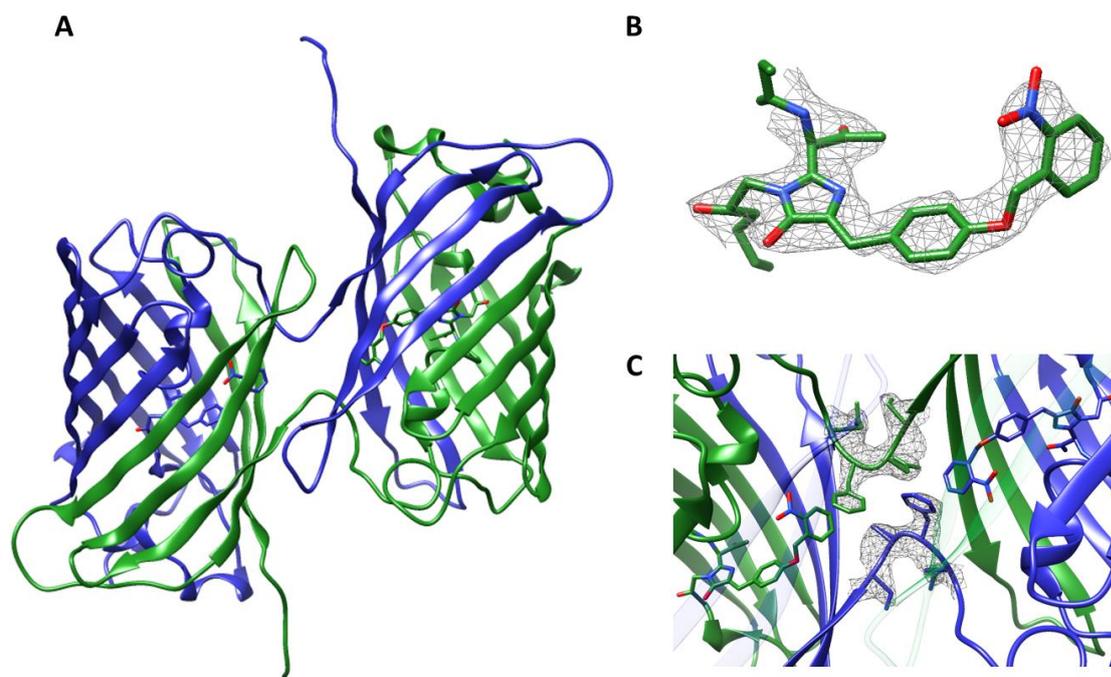


Figure 2.8: Crystal structure of the domain-swapped sfGFP66ONBY. **A.** Structure of domain-swapped sfGFP66ONBY dimer with each chain coloured in blue and green, respectively (pdb: 6H01). **B.** sfGFP66ONBY was purified and crystallized to yield crystals that diffracted to 2.7 Å. The 2Fo-Fc map contoured at 1 r.m.s.d. clearly shows the presence of the *ortho*-nitrobenzyl group and the cyclized backbone forming the chromophore. **C.** Omit map of the open interface and hinge loop region of the domain-swapped dimer contoured at 1 r.m.s.d showing the continuous electron density along the loop connecting the two domains of each chain. This open interface exists only in the domain-swapped dimer, but not in the wild-type monomer or dimer.

The domain-swapped dimer structure of sfGFP66ONBY shows a significantly different dimer interface compared to the wild-type sfGFP structure, which also crystallizes as a dimer (Figure 2.9 A). In the wt-sfGFP structure the dimer interface is formed by β -strands β_7 and β_{10} . In the domain-swapped sfGFP66ONBY structure the β_{10} strands are far apart and the only connection between the two barrels are the loops formed by residues 143-148 (Figure 2.9 A). While the loops (residues 143-148) connecting the swapped-domains are clearly distinct from the wild-type sfGFP structure, the overall structures of the individual barrel moieties are very similar with an all-atom RMSD of 1.27 Å (0.35 Å for C α atoms).

Both structures show an 11-strand β -barrel as well as a mature cofactor derived from cyclisation of the Thr 65 – Tyr 66/ONBY 66 – Gly 67 tripeptide differing only in the presence of the photocaged tyrosine (ONBY) in the sfGFP66ONBY structure, which is clearly visible in the electron density (Figure 2.8 B). The phenomenon of domain-swapped oligomers has been described in several proteins [193] [194], and in *3DSwap*, the curated knowledgebase of proteins involved in 3D domain swapping, nearly 300 depositions can be found [195]. As domain-swapped structures have been mainly observed in crystals, clear evidence for their functional role *in vivo* is lacking. However, three possible *in vivo* functions have been suggested: regulation of protein function, a mechanism for protein misfolding and aggregation, and structural diversification during evolution [193]. The sfGFP66ONBY structure presented here constitutes a *bona fide* domain-swapped structure [196], an arrangement where the dimer adopts a domain swapped conformation and the monomer adopts a closed conformation [195], as previously observed for diphtheria toxin or RNase A [197] [198].

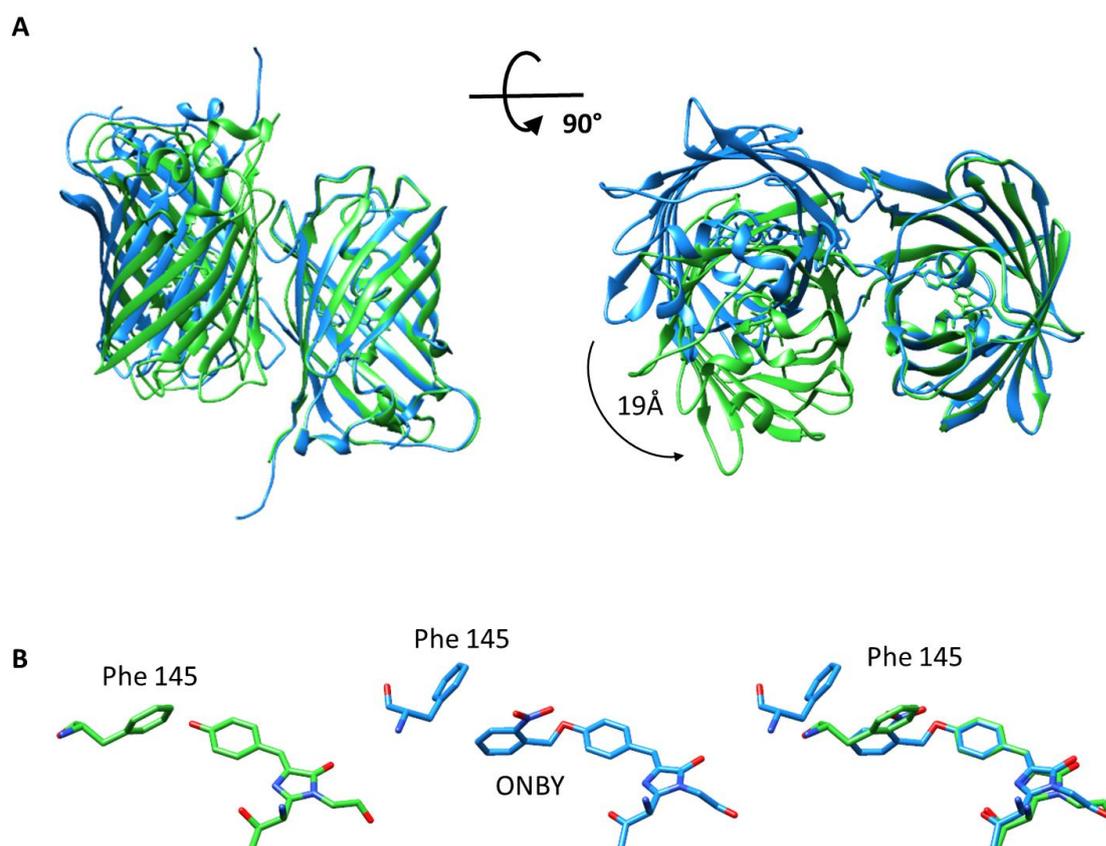


Figure 2.9: Structural comparison of sfGFP66ONBY domain-swapped structure (cyan, (pdb: 6H01)) with wild-type sfGFP dimer structure (green, pdb: 2B3Q). A. One monomer of the dimer is superimposed. The different orientation of the second monomer in the domain-swapped structure and the corresponding lack of dimer interface are clearly visible. The second dimer is shifted by 19 Å. **B.** Chromophores are shown as sticks. Tyr 66 of the chromophore is superimposed. Comparison of the orientation of Phe 145 in the sfGFP66ONBY structure (cyan) with Phe 145 in the wt-sfGFP structure (green) indicates that a clear steric clash between Phe 145 in its wt-sfGFP conformation and the *ortho*-nitrobenzyl-group of ONBY at position 66 would occur, thus causing the structural rearrangement.

2.3.4 Structural model interpretation

The surprising observation of a domain-swapped GFP structure prompts to speculate about its assembly mechanism. Due to the huge thermodynamic stability of folded GFP [107] [199] [200] it is very unlikely that the domain-swapped structure forms via unfolding of properly folded monomeric GFP, for example during crystallization. Therefore it was examined whether conclusions regarding possible folding pathways and intermediates from this domain-swapped structure could be inferred. . The folding of GFP and its variants has been excessively investigated experimentally and theoretically [201] [202] [203] [204] [205] [206] [207] [107]. Thirumalai and coworkers used molecular dynamics simulations to map the folding landscape of GFP and, consistently together with previous studies, suggest that GFP folds via equilibrium or kinetic intermediates along one of four possible pathways (EQL, KIN1, KIN2, KIN3) with the so-called EQL pathway characterized by an equilibrium intermediate [208]. The sfGFP protein chain is built by eleven β -strands forming a cylindrical barrel, with one α -helix threading straight through the β -barrel. The β -strands can be divided into four folding units: the N-terminal β -strands (strand 1 – 3), the chromophore-containing helix at the centre of the β -strand barrel, the three β -strands in the centre (strands 4 – 6), which form local contacts, and the five C-terminal β -strands (strands 7 – 11) (Figure 2.10). In the EQL folding pathway, the N-terminal β -sheets $\beta_1 - \beta_6$ and the central helix fold and join to build half of the β -barrel structure. The C-terminal β -sheets $\beta_7 - \beta_{11}$ are unstructured and flexible and are not part of the β -barrel structure. Hence, the C-terminal β -sheets do not interact with the ordered N-terminal strands $\beta_1 - \beta_6$ in this equilibrium intermediate state. Finally, the C-terminal β -sheets $\beta_7 - \beta_{11}$ fold and join with the rest of the structure folding into the β -barrel structure. This EQL folding pathway agrees well with experimental HDX/NMR data that fit a three-state model and indicate higher flexibility for β -strands $\beta_7 - \beta_{10}$ as well as stable intermediate states along the folding pathway of GFP. These folding intermediates retain considerable secondary and tertiary structure [202].

The domain-swapped sfGFP66ONBY structure presented here contains two domains, $\beta_1 - \beta_6$ (res. 1-143) and $\beta_7 - \beta_{11}$ (res. 148-235), which makes it tempting to speculate that the domain swapped structure is formed from an intermediate resembling the EQL intermediate postulated by Reddy *et al.* [208]. In the sfGFP66ONBY protein, steric clashes of the Phe 145 sidechain with the bulky *ortho*-nitrobenzyl group interfere with tight packing around the chromophore, which prevents the folding and joining of $\beta_7 - \beta_{11}$ to the structured $\beta_1 - \beta_6$ and inhibits folding into the correct monomeric form. Hence, it is not surprising that sfGFP66ONBY is able to access an alternate domain-swapped folded structure (Figure 2.8, Figure 2.9, Figure 2.10 C) under certain conditions. It is likely that the domain-

swapped form of sfGFP66ONBY derives from the same folding intermediate as that postulated for the EQL folding pathway.

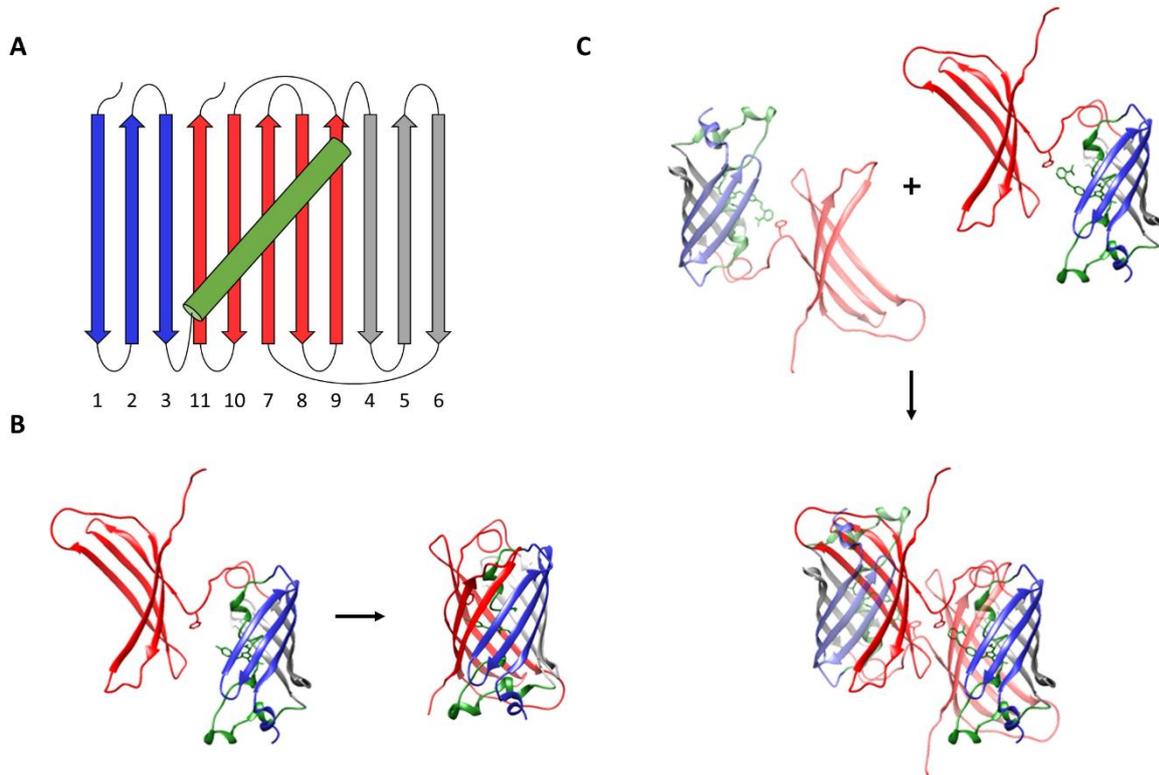


Figure 2.10: Postulated folding mechanism of wt-GFP and its implications for the formation of domain-swapped sfGFP66ONBY (pdb: 6H01). **A.** Topological depiction of wt-GFP. The N-terminal β -strands are represented in blue, the helix containing the chromophore in the centre of the β -strand barrel is in green, the central three β -strands are in grey, and the C-terminal β -strands are in red. **B.** The postulated folding of wt-GFP in the EQL pathway occurs in a two-state manner via equilibrium intermediates. The EQL intermediate comprises two domains: a structured N-terminus ($\beta_1 - \beta_6$) and structured C-terminus ($\beta_7 - \beta_{11}$ in red) connected by a loop. The EQL intermediate is able to fold into the correct monomeric species. **C.** In the sfGFP66ONBY protein, steric clashes of Phe145 with the *ortho*-nitrobenzyl group interferes with folding and joining of C-terminal $\beta_7 - \beta_{11}$ (red) to the structured $\beta_1 - \beta_6$ at the N-terminus and prevent folding of the correct monomer. Hence, the intermediate of the EQL pathway can only fold into the unusual domain-swapped sfGFP66ONBY dimer.

Phe 145 was identified as key residue preventing folding of sfGFP66ONBY into a monomeric structure due to its steric clashes with the nitrobenzyl moiety of the ONBY chromophore. I investigated whether it adopts the conformation observed in wild-type or sfGFP66ONBY (see above). Therefore, a sfGFP66ONBY-F145A point mutant was expressed and purified and compared regarding its oligomeric status to wild-type sfGFP and sfGFP66ONBY. Wild-type GFP and sfGFP are active as a monomers, however, sfGFP migrates as a dimer on an SDS-PAGE gel without heat treatment and as monomer with heat treatment and predominantly elute as a monomer from SEC [178] (Figure 2.4). The ONBY variants sfGFP66ONBY and sfGFP66ONBY-F145A show similar behaviours in SDS-PAGE and SEC as wt-sfGFP. In

order to verify the oligomeric state of the sfGFP66ONBY variant and to gain further insight into the postulated folding intermediate as a building block for the domain-swapped structure, native mass spectrometry was employed (NMS) [209]. For both wt-sfGFP and sfGFP66ONBY, monomeric and dimeric species were present at low μM concentration with the dimeric fraction representing the minor population (Figure 2.11). For wt-sfGFP the monomeric and dimeric species were assigned to the properly folded monomer and dimer structures as observed in the crystal structure [178]. For sfGFP66ONBY, the dimer mass was assigned to the domain swapped dimer, while the monomer mass could correspond to the partly unfolded intermediate (Figure 2.10B). This suggests that the samples used for spectroscopy and crystallography (see above) contained a mixture of monomeric and dimeric domain-swapped sfGFP66ONBY and that only decaging of the monomeric sfGFP66ONBY population is able to restore the wt-sfGFP phenotype and fluorescence, and that the crystallization process may force the monomers into domain-swapped dimers. In contrast to wt-sfGFP and sfGFP66ONBY, the sfGFP66ONBY-F145A variant is predominantly monomeric in NMS analysis and shows less dimeric species (Figure 2.11). These results suggest that the F145A mutation indeed allows proper folding of sfGFP66ONBY.

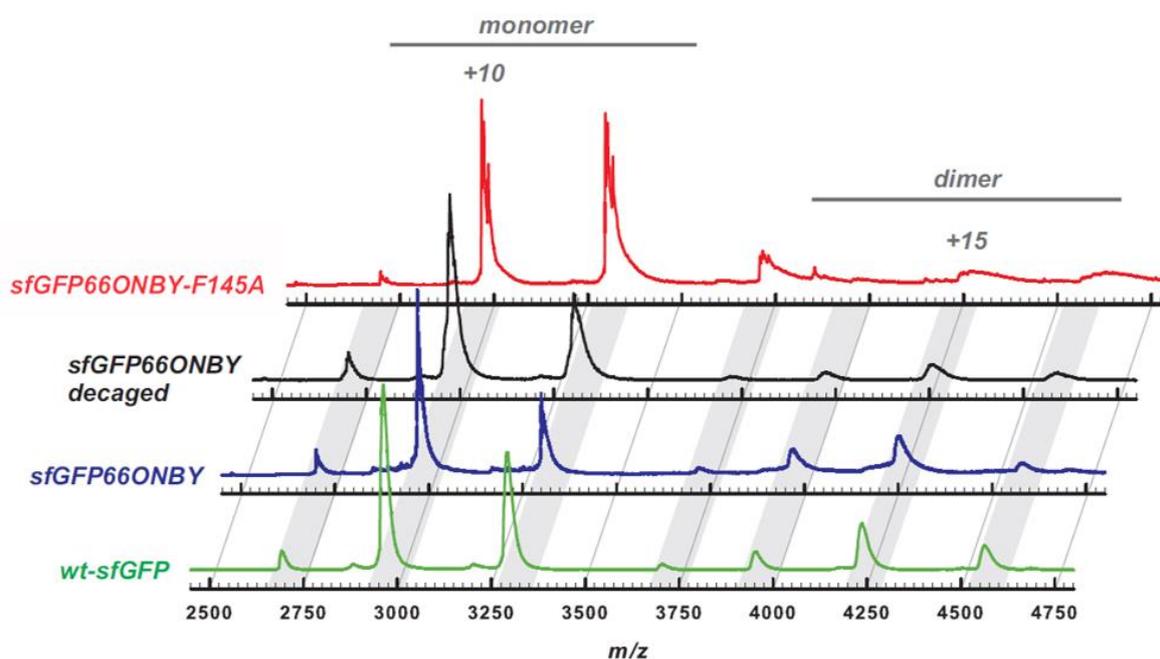


Figure 2.11: Native mass spectrometry analysis of sfGFP variants. Native mass spectra indicate the presence of monomeric and dimeric species for wt-sfGFP (green), sfGFP66ONBY (blue) as well as for decaged sfGFP66ONBY (black) and predominantly monomeric species for sfGFP66ONBY-F145A (red). Main peaks are annotated with their charge states; positions of equal charge states are depicted in grey.

2.4 Conclusion

In summary, the crystal structure of sfGFP66ONBY containing a genetically incorporated unnatural amino acid as part of the chromophore revealed a novel and unexpected domain-swapped arrangement. While the suitability of this construct for time-resolved structural studies remains to be investigated, the domain-swapped structure provides experimental evidence in support of the EQL GFP folding pathway.

CHAPTER 3

3. OMPX BINDING STUDIES

3.1 Introduction

Gram-negative bacteria such as *E. coli* are surrounded by two concentric lipid bilayer membranes. In contrast to the phospholipid containing inner cytoplasmic membrane, the outer membrane is highly asymmetric, with its inner side having the same lipid composition as the inner cytoplasmic membrane and the outer side composed of lipopolysaccharides. This outer membrane is not only protecting bacteria against their harsh environment, but also plays a crucial role in the transport of molecules and information. About 50% of the outer membrane mass consists of protein, either in the form of integral membrane proteins or as lipoproteins that are anchored to the membrane, fulfilling the variety of completely different functions that are crucial to the bacteria such as solute and protein translocation, signal transduction and metabolic processes [210]. So far, more than a dozen different outer membrane proteins have been identified in *E. coli* [211] and classified into six outer membrane protein families [210]. Characteristically, integral outer membrane proteins do not consist of transmembrane α -helices, but instead fold into anti-parallel β -barrels [212]. In recent years, the atomic structures of several outer membrane proteins have been determined by X-ray crystallography including the OmpA membrane domain [213], the OmpX protein [214], phospholipase A [215], general porins such as OmpF [216] and PhoE [217], substrate-specific porins such as LamB [218] and ScrY [219] and the TonB-dependent iron siderophore transporters FhuA and FepA [220-222]. These crystallographic studies did not only yield remarkable advances in the understanding of the function of these outer membrane proteins, but may additionally allow conceptual models for membrane proteins with similar functions or common principles with unknown structures. Moreover, β -barrels have been studied excessively since then and are well-described systems as well as suitable objects for engineering, due to their simple structure and the production in inclusion bodies allowing recovery therefrom in the native conformation via micelles [223].

The integral outer membrane protein X (OmpX) of *E. coli* belongs to a family of highly conserved proteins that appear to be important for virulence by neutralizing the host's defence mechanism and

enabling cell adhesion and entry [224]. The protein consists of an 8-stranded antiparallel β -barrel [149, 150], which shows two girdles of aromatic amino acid residues and a ribbon of nonpolar residues that attach to the membrane interior (Figure 3.1). The core of the barrel consists of an extended hydrogen-bonding network building up an inverse micelle. A peculiarity of OmpX compared to other known outer membrane proteins is the organization of four elongated β -strands extending beyond the region of the outer membrane into the extracellular space that feature much less conserved residues compared to the membrane-spanning part of the barrel [150]. This structural motif, similar to a 'fishing rod', has been proposed to promote cell adhesion and bacterial cell internalization as well as to confer resistance to the complement system by binding at least one essential external protein containing complementary β -strands [150]. Consistently, it has been shown that in other bacteria OmpX is also involved in the invasion of host cells [225-227], neutralizing host defence mechanisms, and bacterial defence against the complement systems of the host [224]. As there is no direct pathway between the periplasmic space and the extracellular end of the barrel, OmpX is very unlikely to function as a pore and involved in secretion or transport processes.

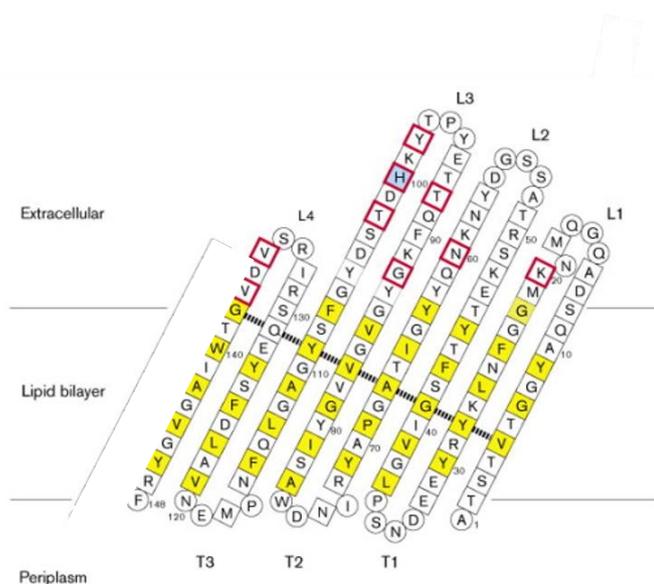
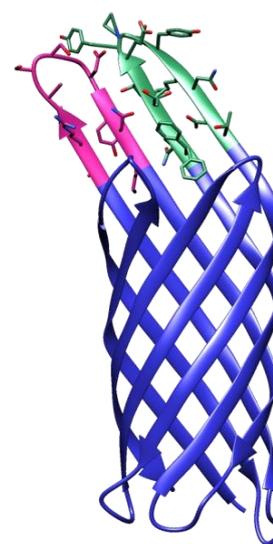
A**B**

Figure 3.1: Topology and Structure of OmpX. **A.** Topological depiction of OmpX. Extracellular loops are labelled L1 –L4. Periplasmic end turns are labelled T1 –T3. Residues pointing to the nonpolar membrane moiety are coloured in yellow. Residues in β -strands are shown in squares; other residues are in circles. Residue positions involved in virulence and defence of Rck of *S. typhimurium* (Lys20 and Thr88) and of Ail of *Y. enterocolitica* (Asn60, Gly92, Tyr98, His100, Thr102, Val135 and Val137) are marked in red. The dashed black line indicates the hydrogen-bonding register within the β -barrel (Modified from Vogt & Schulz, 1999). **B.** Crystal structure of OmpX (pdb: 1QL8) with extracellular loop 2 coloured in purple and extracellular loop 3 in green.

Another outer membrane protein belonging to the outer membrane family is OmpA, which is also expressed in the cell wall of Gram-negative bacteria. Whereas the primary sequence differs from OmpX, both exhibit a similar eight-stranded transmembrane domain structure, which is highly conserved throughout evolution [228]. OmpA additionally contains a periplasmic domain, which is not present in OmpX. Previous studies showed that OmpA is involved in virulence by binding and selectively activation antigen presenting cells [229]. Based on this structural homology, OmpX may interact in a similar fashion with immune cells and exhibits similar biological properties.

Photo-crosslinking agents such as benzophenones, aryl azides, diazoesters and diazaarenes have been extensively used as photophysical probes to identify and map protein–protein or protein-nucleic acid interactions *in vitro* and living cells [38, 39, 131, 230]. Among these, benzophenones are the most useful since they are chemically stable, can be routinely manipulated under ambient light and do not photo-dissociate [231, 232]. Upon excitation at 350–360 nm, the reactive groups interact with other adjacent carbon-hydrogen bonds. In 1986, DeGrado and coworkers were the first to demonstrate that the photo-crosslinking amino acid *para*-benzoyl-L-phenylalanine (BPA) could be site-specifically incorporated into synthetic proteins via solid-phase peptide synthesis [233]. Although this method has been used intensively since then for protein-peptide interaction analysis [231], the method was limited to small proteins or peptides and amino acids on the surface proteins cacheable by solid-phase peptide synthesis. In 2002, Schultz and coworkers developed a method based on orthogonal aminoacyl-tRNA synthetase/tRNA pairs for the *in vivo* incorporation of BPA into proteins in response to the “amber” codon, which is not limited to protein size or accessibility of the amino acid [94].

Understanding the function, biochemical characteristics and stabilizing agents of the OmpX structure, as well as factors determining protein interaction with its surrounding lipid environment, is critical to decipher its role in virulence and pathogenicity. Hence, the light-activatable crosslink probe BPA was used to analyse the function of OmpX and identify possible protein-protein-interactions in *in vivo* experiments. Since BPA is an analogue of tyrosine (Figure 3.2) both amino acids share high structural similarity making tyrosine residues easily replaceable by BPA. Crucial tyrosine residues in the extracellular loops were selected and replaced by BPA to identify possible interaction partners and elucidate the function of OmpX.

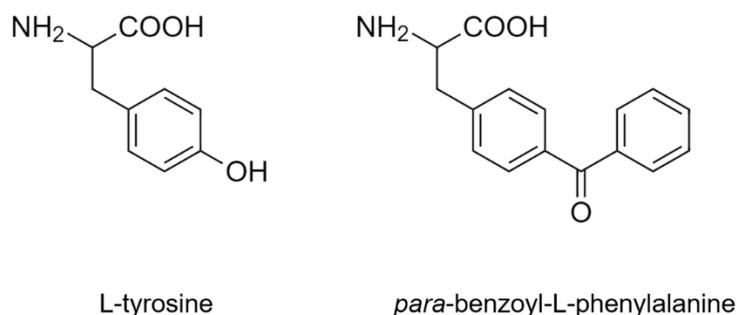


Figure 3.2: L-Tyrosine and its analogue *para*-benzoyl-L-phenylalanine (BPA).

3.2 Materials and Methods

3.2.1 Cloning and expression of OmpX variants

The plasmid pET21a-OmpX Δ S-6His was a generous gift from Franz Hagn (TU München, Germany). Cloning of the *ompX* gene containing the signal peptide for insertion in the membrane was achieved by the generation of pET21a-OmpXSP via site-directed PCR mutagenesis using primers containing the signal peptide motif. This plasmid was cotransformed into *E. coli* BL21 (DE3) Gold cells together with the plasmid pET-sfGFP-His6, containing the gene for sfGFP. The cotransformation and expression of sfGFP allows localisation of bacterial cells in fluorescence microscopy. Proteins were overexpressed in *E. coli* strain BL21(DE3) Gold in terrific broth (TB) medium. A single colony of the cells, grown on a LB-agar plate with ampicillin (100 μ g/ml) and kanamycin (25 μ g/ml) overnight at 37 °C, was used to inoculate a 10 ml pre-culture with LB-medium containing the same antibiotics. The pre-culture was grown for 16 hours at 37 °C at 180 rpm and subsequently used to inoculate the expression culture (50 ml TB-medium at 37 °C und 180 rpm, supplemented with the same antibiotics). Gene expression was induced at an OD₆₀₀ of 0.6-0.8 by addition of 1 mM isopropyl thiogalactopyranoside (IPTG) and the expression was conducted for 5 hours. Purity was assessed by NMR.

For incorporation of *para*-(benzoyl)-phenylalanine (BPA) at two specific extracellular loop positions, the residues Y57 or Y95 were mutated to the “amber” codon, TAG, via site-directed mutagenesis, to generate pET21a-OmpXSP-Y57BPA and pET21a-OmpXSP-Y95BPA. This plasmid was cotransformed into *E. coli* BL21 (DE3) Gold cells together with the plasmid pET-sfGFP-His6, containing the gene for sfGFP, and pEVOL-pBpF, containing the orthogonal tRNA / aaRS pair for incorporation of BPA. Proteins

were overexpressed in *E. coli* strain BL21(DE3) Gold in terrific broth (TB) medium. A single colony of the cells, grown on a LB-agar plate with ampicillin (100 µg/ml), kanamycin (25 µg/ml) and chloramphenicol (34 µg/ml) overnight at 37 °C, was used to inoculate a 10 ml pre-culture with LB-medium containing the same antibiotics. The pre-culture was grown for 16 hours at 37 °C at 180 rpm and subsequently used to inoculate the expression culture (50 ml TB-medium at 37 °C und 180 rpm, supplemented with the same antibiotics). BPA was added to a final concentration of 1 mM and arabinose to a final concentration of 0.02 % for induction of the tRNA/aaRS pair, and protein production was induced by the addition of 1 mM IPTG with a delay of 60 min and the expression was conducted for 5 hours.

Truncated extracellular loop variants were generated by site-directed mutagenesis changing the motif of the extracellular loops to a “GSSG” (glycine-serine-serine-glycine) motif, yielding pET21a-OmpX-SP-Y95BPA-EL2, pET21a-OmpX-SP-Y57BPA-EL3 and pET21a_OmpX-SP-EL2-EL3. Sequences of all isolated plasmids were verified by sequencing. Plasmids were cotransformed into *E. coli* BL21 (DE3) Gold cells together with the plasmid pET-sfGFP-His6, containing the gene for sfGFP, and pEVOL-pBpF, containing the orthogonal tRNA / aaRS pair for incorporation of BPA. Proteins were overexpressed in *E. coli* strain BL21(DE3) Gold in terrific broth (TB) medium. A single colony of the cells, grown on a LB-agar plate with ampicillin (100 µg/ml), kanamycin (25 µg/ml) and chloramphenicol (34 µg/ml) overnight at 37 °C, was used to inoculate a 10 ml pre-culture with LB-medium containing the same antibiotics. The pre-culture was grown for 16 hours at 37 °C at 180 rpm and subsequently used to inoculate the expression culture (50 ml TB-medium at 37 °C und 180 rpm, supplemented with the same antibiotics). BPA was added to a final concentration of 1 mM and arabinose to a final concentration of 0.02 % for induction of the tRNA/aaRS pair, and protein production was induced by the addition of 1 mM IPTG with a delay of 60 min and the expression was conducted for 5 hours.

3.2.2 Human cell culture

Human embryonic kidney 293 cells (HEK 293) were cultivated in DMEM GlutaMAX media (Gibco) supplemented with 10 % fetal calf serum (FCS) and 1 % Penicillin / Streptomycin at 37 °C and 5 % CO₂. HeLa cells were cultivated in DMEM GlutaMAX media (Gibco) supplemented with 10 % FCS and 1 % Penicillin / Streptomycin at 37 °C and 5 % CO₂. THP-1 cells were cultivated in RPMI1640 GlutaMAX media (Gibco) supplemented with 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 0.05 mM 2-mercaptoethanol, 10 % FCS and 1 % Penicillin / Streptomycin at 37 °C and 5 % CO₂. Jurkat cells were cultivated in RPMI1640 GlutaMAX media (Gibco) supplemented with 25 mM HEPES, 7.5 %

newborn calf serum and 1 % Penicillin / Streptomycin at 37 °C and 5 % CO₂.

3.2.3 Adhesion assay

Liquid adhesion tests: To determine bacterial adhesion to human cells, adhesion assays were performed as described previously with minor alterations [234]. Briefly, bacteria were prepared by transferring sufficient inoculum from overnight cultures into fresh, prewarmed TB media to reach an OD₆₀₀ of 0.1, followed by incubation at 37°C with constant shaking until OD₆₀₀ of 0.6 – 0.8 and gene expression for 5 hours. Bacterial cells were washed three times with PBS and resuspended in corresponding human cell growth medium without antibiotics. Human suspension cells were collected by centrifugation at 200 × g, washed three times with phosphate-buffered saline (PBS) at pH 7.4 (Gibco), and resuspended in corresponding growth medium without antibiotics for adhesion assays. Human adhesion cells were treated with trypsin for detachment, collected by centrifugation at 200 × g, washed three times with PBS (pH 7.4), and resuspended in corresponding growth medium without antibiotics for adhesion assays. Human cell cultures were infected with bacteria at a multiplicity of infection (MOI) of 10, 20 and 50 and incubated for 1 h at 37 °C with mild agitation.

Adhesive adhesion test: To determine bacterial adhesion to human cells, adhesion assays were performed as described previously with minor alterations [234]. Briefly, bacteria were prepared by transferring sufficient inoculum from overnight cultures into fresh, prewarmed TB media to reach an OD₆₀₀ of 0.1, followed by incubation at 37°C with constant shaking until OD₆₀₀ of 0.6 – 0.8 and gene expression for 5 hours. Bacterial cells were washed three times with PBS and resuspended in corresponding human cell growth medium without antibiotics. Human suspension cells were collected by centrifugation at 200 × g, washed three times with PBS (pH 7.4), and resuspended in corresponding growth medium. Cells were plated in appropriate cell concentrations according to the American Type Culture Collection's manuals on fibronectin-coated [50 µg/ml] 35 mm MatTEK glass bottom dishes (14 mm glass diameter) and incubated for 45 min at 37 °C and 5 % CO₂. Cells were washed three times with PBS and growth medium without antibiotics was added. Monolayer cells were infected with bacteria at a multiplicity of infection (MOI) of 50 and incubated for 1 h. Cell viability was determined by trypan blue staining. Human adhesion cells were treated with trypsin for detachment, collected by centrifugation at 200 × g, washed three times with PBS (pH 7.4), resuspended in corresponding growth medium. Cells were plated in appropriate cell concentrations according to the American Type Culture Collection's manuals on fibronectin-coated 35 mm MatTEK glass bottom dishes (14 mm glass diameter) and incubated for 16 hours at 37 °C and 5 % CO₂. The cells formed monolayers after 1 day and cell

viability was determined by trypan blue staining. Cells were washed three times with PBS and growth medium without antibiotics was added. Monolayer cells were infected with bacteria at a multiplicity of infection (MOI) of 50 and incubated for 1 h. Cell viability was determined by trypan blue staining.

3.2.4 Photo-crosslinking

To perform *in vivo* crosslinking, liquid samples and dishes containing cells were subjected to UV irradiation at 365 nm light (Spectroline® E-Series 6 W UV lamp) for 1 h. To remove unbound bacterial cells, liquid samples were centrifuged at 200 x g and washed with PBS and the procedure was repeated twice. Afterwards crosslinked cells were transferred to a microscope slide, covered with a cover slip and sealed before imaging. To remove unbound bacterial cells from adherent cells, dishes were washed three times with PBS, followed by adding PBS before imaging. As control, bacterial cells expressing only green fluorescent protein were crosslinked with *para*-formaldehyde (3.7% in PBS; 15 min) and washed three times with PBS.

3.2.5 Confocal laser fluorescence microscopy

To monitor bacterial cell location, all bacterial cells were cotransformed with green fluorescent protein (GFP) by electroporation. The presence of the plasmid and cytosolic GFP did not affect the adhesion capabilities. Fluorescence signals were recorded using an inverted confocal laser scanning Leica DM IRBE microscope combined with fluorescence correlation spectroscopy (Leica, Germany) with a 40 x oil immersion objective. Bright field and fluorescence images were analysed and merged with Fiji [235].

3.2.6 Statistical analysis

All experiments were performed at least in triplicate. Data from cell culture experiments were analysed using GraphPad Prism (version 7.04 for Windows, GraphPad Software, La Jolla California USA).

3.2.7 Membrane preparations

Crosslinked cells were collected from 35 mm MatTEK glass bottom dishes (14 mm glass diameter), centrifuged at 4,000 x g for 20 min at 4°C and resuspended in 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 1 mM PMSF. Lysozyme was added to the concentration of 300 µg/ml and the cells incubated for 4h at 4° C. 1 mM MgCl₂ and 1 µl of DNase I (1mg/ml) were added and incubated for 30

min at 4° C. Disrupted cells were centrifuged at 16,000 x g for 30 mins, and the supernatant centrifuged at 40,000 x g for 90 min in an ultracentrifuge to retrieve the membranes.

3.2.8 Western Blot

SDS-PAGE was performed as described in the section Appendix – General Sample Preparation Techniques. Proteins were transferred to a polyvinylidene difluoride membrane (activated before usage in methanol) by western blotting at 50 V for 1.5 h at RT with ice-cold transfer buffer (20 mM Tris, 150 mM glycine, 10 % SDS and 20 % methanol). Blotted membranes were carefully rinsed with ddH₂O. The membrane was blocked for unspecific binding with 5 ml 5 % milk powder in PBS with 0.1 % Tween 20 for 1 hours at RT. The blocking buffer was removed and membranes washed three times for 10 min with PBS. The membrane was incubated with 1:5,000 diluted Penta-His-HRP conjugate antibody (Qiagen) in 5 ml and 3 % for 16 hours at 4 °C. The antibody was removed and the membranes washed three times for 10 min with PBS. 1 ml of ECL solution (Supersignal West His Probe Kit, Thermo Fisher Scientific) was added to the membrane and incubated for an appropriate time. The picture was acquired with a Bio-Rad ChemiDoc MP imaging system.

3.3 Results and Discussion

3.3.1 Adhesion assay and imaging

Understanding the function and biochemical interactions of OmpX is critical to decipher its role in virulence and pathogenicity. Based on this lack of knowledge, OmpX mutant variants were designed for protein-protein-interaction studies with the photo-crosslinker BPA in human cell lines in order to elucidate possible interaction partners indicating the function of OmpX. A sequence analysis of the *ompX* gene revealed three tyrosine residues in the extracellular loops: Y57 in extracellular loop 2 (EL2) and Y95 and Y98 in extracellular loop 3 (EL3) with Y95 stretching away from the protein and Y98 pointing towards the protein. Based on these findings, Y57 and Y95 were chosen for possible positions for BPA incorporation, and OmpX variants containing the “amber” codon at corresponding position were generated. After mutagenesis, the mutant variants were coexpressed with an orthogonal tRNA and aminoacyl-tRNA-synthetase pair recombinantly in *E. coli* BL21(DE3) Gold cells.

For protein-protein interactions studies, the human cell lines HEK293, HeLa, THP-1 and Jurkat were

chosen for a high cell lines diversity, whereof HEK293 and HeLa cells are adherent cells and THP-1 and Jurkat cells are suspension cells. HEK293 cells are epithelial cells originally derived from human embryonic kidney [236]. Characteristic for these cells is the expression of an unusual cell surface receptor for vitronectin composed of the integrin β -1 subunit and the vitronectin receptor α subunit. HeLa cells are also epithelial cells and were originally derived from cervical cancer cells. It is the oldest and most commonly used human cell line [237]. THP-1 is a human monocytic cell line derived from the peripheral blood an acute monocytic leukaemia patient [238]. This cell line expresses the cytokine signalling molecule interleukin-2 (IL2) of the human immune system as well as the antigen cluster of differentiation 3 (CD3) T cell co-receptor involved in T cell and T helper cell activation. Additionally it harbours the T cell antigen receptor. Jurkat cells are lymphoblasts derived from T lymphocyte cells also from the peripheral blood. This cell line produces several proteins involved in the human immune system such as HLA A2, A9, B5, DRw1 and DRw2. Moreover, it expresses several antigens and features the complement receptor 3 as well as the Fc receptor. The cells are phagocytic and lack surface and cytoplasmic immunoglobulin [239].

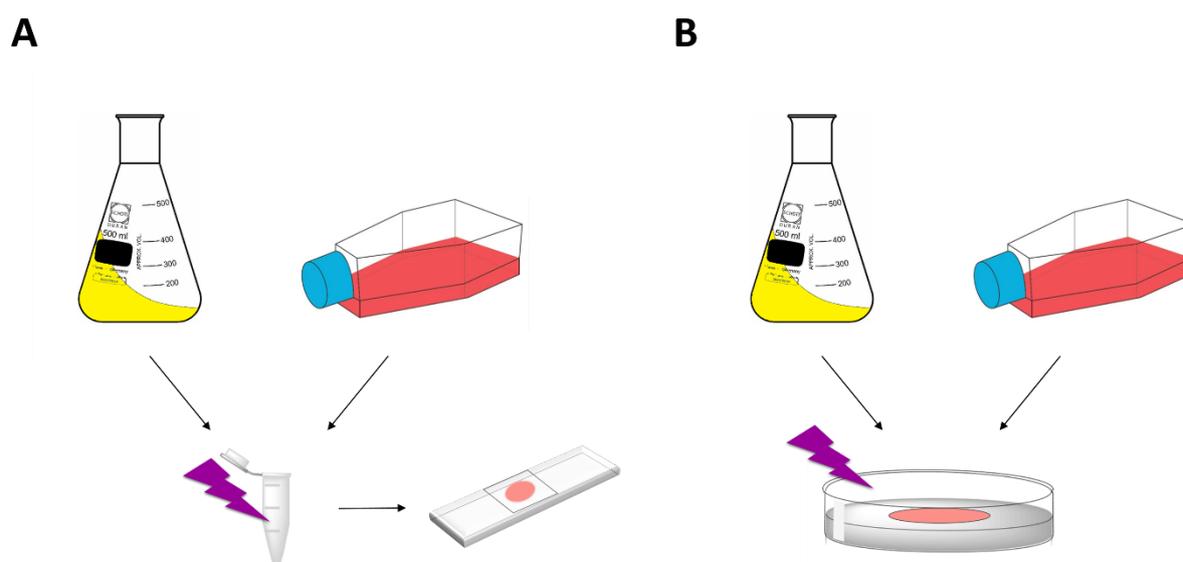


Figure 3.3: Schematic overview of experimental setup for adhesion assays. A. For liquid adhesion assays, human cells in suspension were infected with an excess of bacterial cells overexpressing *ompX*, crosslinked with light of 365 nm, excessively washed to remove unbound bacterial cells and transferred to microscope slides for imaging. **B.** For adherent adhesion assays, human cells growing on glass bottom dishes were infected with an excess of bacterial cells overexpressing *ompX*, crosslinked with light of 365 nm, excessively washed to remove unbound bacterial cells, followed by direct imaging.

For *in vivo* liquid adhesion assays, all cell lines were used in suspension. In contrast to the suspension cells THP-1 and Jurkat, the adherent HEK293 and HeLa cells therefore required a trypsin treatment resulting in suspended cell. Suspension cells were infected with bacterial cells expressing the different

OmpX variants at a multiplicity of infection (MOI) of 10, 20 and 50 and crosslinked and imaged on microscope slides with a Leica DM IRBE inverted microscope with a 40 x oil immersion objective (Figure 3.3A). *E. coli* expressing only endogenous OmpX level were not crosslinked, whereas *E. coli* overexpressing OmpX were crosslinked with 3.7 % *para*-formaldehyde for 15 min and *E. coli* expressing the both OmpX-Y57BPA and OmpX-Y95BPA variants were crosslinked by irradiation of 365 nm light for 60 min. All bacterial cells coexpressed *gfp* as probe for easy and quick visualization and localization. The cells were first imaged using conventional light microscopy allowing identification of the large eukaryotic cells in bright field illumination. Eukaryotic cells had excellent spreading and good dispersions allowing single cell imaging. To identify bacterial cells, fluorescence was induced by irradiation of 355 nm light. Images were analysed by Fiji software and bright field and fluorescence images were merged. Imaged eukaryotic cells with an MOI of 50 are exemplary shown in Figure 3.4. For all three experiments with different MOI of 10, 20 and 50 similar observations could be made: For HEK293 and HeLa nearly no adhesion of *E. coli* cells and only scattered *E. coli* could be overserved after infection and crosslinking with *E. coli* expressing the different OmpX variants. For Jurkat and THP-1 cells a slightly higher adhesion could be observed for *E. coli* overexpressing OmpX variants crosslinked either with *para*-formaldehyde or with 365 m light and no adhesion for *E. coli* with endogenous levels of OmpX (Figure 3.4). The fact that no attachment of bacterial cells could be observed might have several reasons. First, in the most obvious case OmpX might not bind specifically to any of the selected eukaryotic cells. Since HEK293 and HeLa are epithelial cells from internal organs, these might be no direct targets in cell adhesion of OmpX. THP-1 cells are monocytes found in the peripheral blood and are a key player in the human innate immune system. Similar to THP-1 cells, Jurkat cells are also involved in the immune system: as T lymphocytes they play a central role in cell-mediated immunity. Furthermore, the crosslinking experiments might have not been successful since no difference between conventional crosslinking with *para*-formaldehyde or BPA can be detected. Additionally, the excessive washing steps after crosslinking might have been too harsh and destroyed the crosslinks. Apart from this a liquid test may not be suitable for adhesion assays since it complicates the binding due to continuous agitation during the experiment.

Based on the experiences gained with liquid adhesion assays, a fixed-cell adhesion assay was performed. In this regard, MATtek glass bottom dishes with 14 mm glass windows for direct imaging were coated with fibronectin for eukaryotic cell adhesion. The adherent cell lines HEK293 and HeLa formed monolayers after 1 day. The suspension cell lines Jurkat and THP-1 were fixed 1h before the experiment to maintain high cell viability. Eukaryotic cells were infected with a MOI of 50 to exclude too low infection rates in the former experiment, followed by crosslinking and imaging with a Leica DM

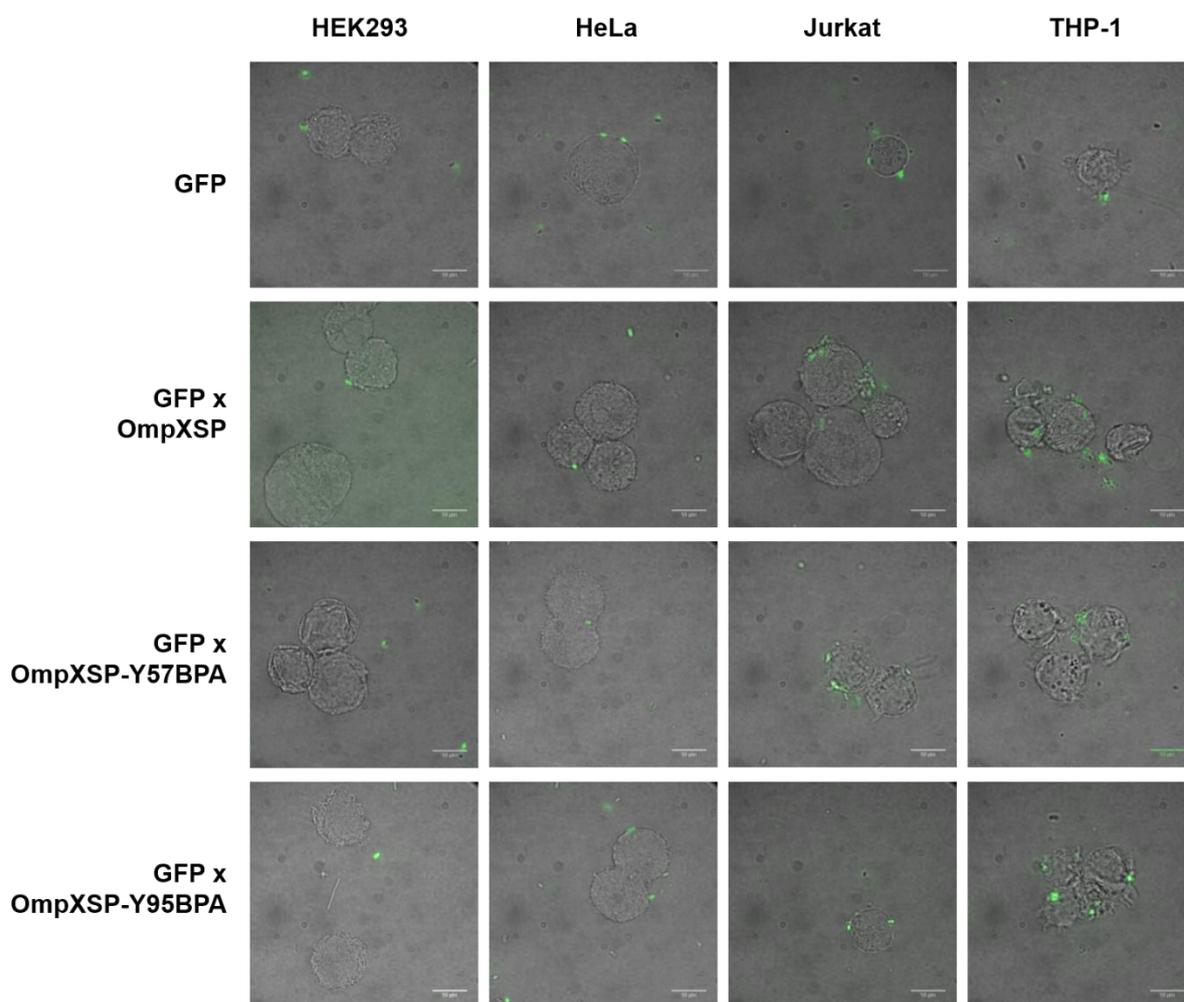


Figure 3.4: Confocal fluorescence micrographs of liquid adhesion assay. HEK293, HeLa, Jurkat and THP-1 cells were infected with bacterial cells with a MOI of 50. All bacterial cells coexpress *gfp* for visualisation. Bacterial cells overexpressing *ompXSP* were crosslinked with *para*-formaldehyde. Bacterial cells overexpressing *ompXSP* containing BPA were crosslinked with 365 nm light for 60 min. Scale bars represent 10 μm in all images. A Leica DM IRBE inverted microscope with a 40 x oil immersion objective were used. Abbreviations: GFP, bacterial cells expressing only endogenous *ompX*; OmpXSP, bacterial cells overexpressing *ompXSP* crosslinked with *para*-formaldehyde, OmpXSP-Y57BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 57 crosslinked with 365 nm light, OmpXSP-Y95BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 95 crosslinked with 365 nm light.

IRBE inverted microscope with a 40 x oil immersion objective (Figure 3.3B). As previously, *E. coli* expressing only endogenous OmpX level were not crosslinked, whereas *E. coli* overexpressing *ompX* were crosslinked with 3.7 % *para*-formaldehyde for 15 min and *E. coli* expressing the both OmpX-Y57 and OmpX-Y95 variants were crosslinked by irradiation of 365 nm light for 60 min. All bacterial cells coexpressed *gfp* as probe for easy and quick visualization and localization. The cells were first imaged using conventional light microscopy allowing identification of the large eukaryotic cells in bright field IRBE inverted microscope with a 40 x oil immersion objective (Figure 3.3B). As previously, *E. coli* ex-

pressing only endogenous OmpX level were not crosslinked, whereas *E. coli* overexpressing *ompX* were crosslinked with 3.7 % *para*-formaldehyde for 15 min and *E. coli* expressing the both OmpX-Y57 and OmpX-Y95 variants were crosslinked by irradiation of 365 nm light for 60 min. All bacterial cells coexpressed *gfp* as probe for easy and quick visualization and localization. The cells were first imaged using conventional light microscopy allowing identification of the large eukaryotic cells in bright field

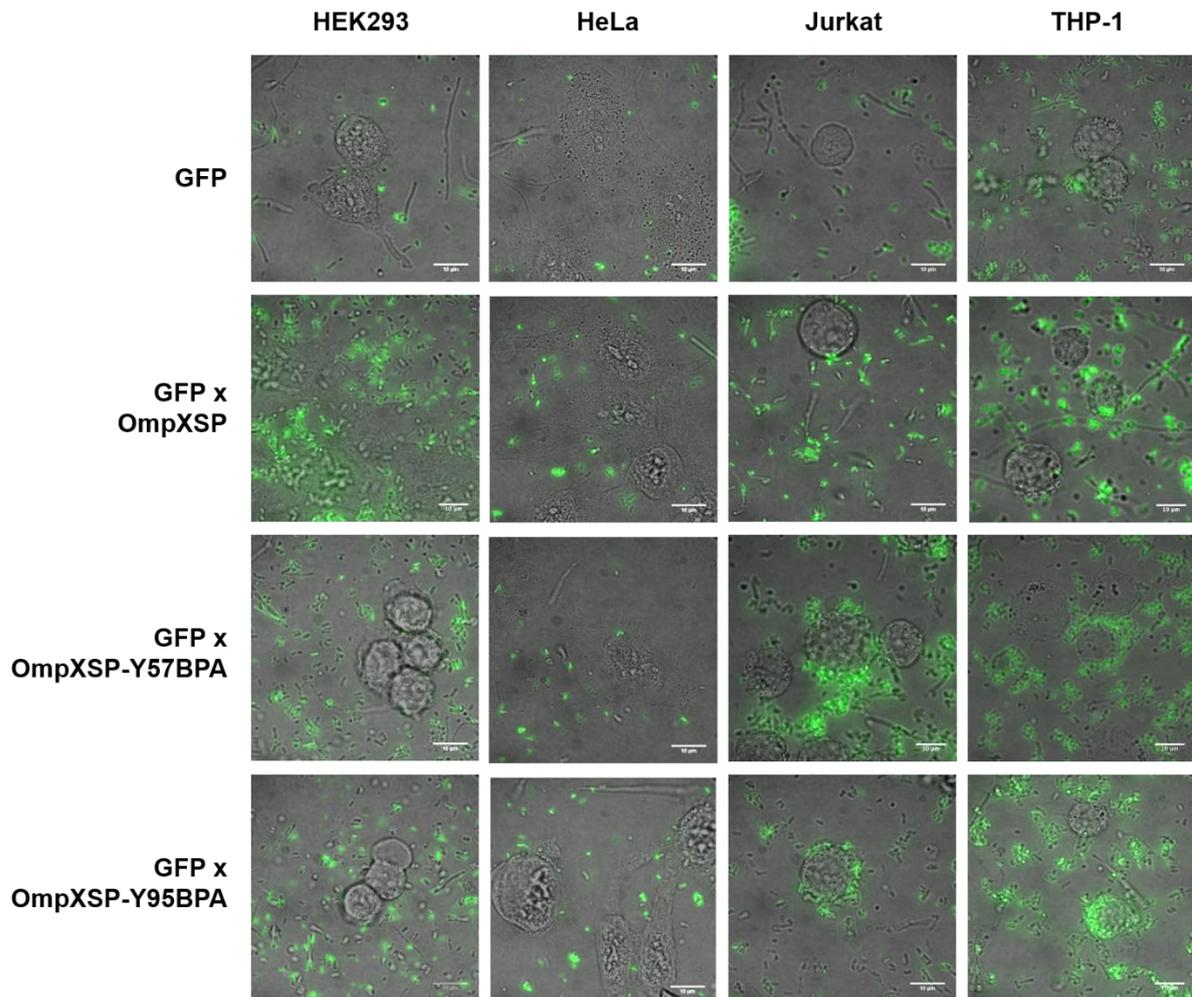


Figure 3.5: Confocal fluorescence micrographs of fixed-cell adhesion assay. HEK293, HeLa, Jurkat and THP-1 cells were infected with bacteria with a MOI of 50. All bacterial cells coexpress *gfp* for visualisation. Bacterial cells overexpressing *ompXSP* were crosslinked with *para*-formaldehyde. Bacterial cells overexpressing *ompXSP* containing BPA were crosslinked with 365 nm light for 60 min. Scale bars represent 10 μ m in all images. A Leica DM IRBE inverted microscope with a 40 x oil immersion objective were used. Abbreviations: GFP, bacterial cells expressing only endogenous *ompX*; OmpXSP, bacterial cells overexpressing *ompXSP* crosslinked with *para*-formaldehyde, OmpXSP-Y57BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 57 crosslinked with 365 nm light, OmpXSP-Y95BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 95 crosslinked with 365 nm light.

illumination. Again, good cell spreading was achieved to allow single cell imaging. To identify bacterial cells, fluorescence was induced by irradiation of 355 nm light. Images were analysed by Fiji software and bright field and fluorescence images merged. Imaged infected eukaryotic cells with an MOI of 50 are shown in Figure 3.5. For HEK293 cells no bacterial cell adhesion could be observed for all four bacterial variants, although a higher number of unbound, randomly moving bacteria is present compared to the liquid adhesion assay. Similar to HEK293, no adhesion of bacterial variants to HeLa cells can be observed and a higher number of unbound, randomly moving bacteria compared to the liquid adhesion assay. In case of the Jurkat cells, no interaction could be seen for bacterial cells with endogenous OmpX and some adhesion for bacteria overexpressing *ompX* and crosslinked via para-formaldehyde. Interestingly, the BPA-crosslinked bacteria cluster around the eukaryotic cells indicating a distinct adhesion visible as green circles and crescents in intersections as well as spheres in focus stacking. In the presence of THP-1 cells, the bacteria show a similar behaviour as in presence of Jurkat cells: *para*-formaldehyde- as well as BPA-crosslinked bacteria appear as bright green clusters around the eukaryotic cells. In all experiments a higher number of randomly moving bacteria can be observed compared to the liquid adhesion assay indicating either insufficient washing steps or already detached bacteria due to excessive washing, with the first one more likely.

Additionally to the qualitative observations by confocal fluorescence microscopy, the ability of the bacteria with endogenous OmpX levels, overexpressed *ompX* levels and the BPA-crosslinking-variants to attach to eukaryotic cells have been analysed quantitatively (Figure 3.6). Experiments were performed in triplicate and for each experiment at least 20 cells were analysed with respect to adhesion. For HEK293 and HeLa cells, there was no significant difference between the different bacteria and no adhesion could be observed at all. In case of the Jurkat cells, the differences in adhesion ability between the *para*-formaldehyde crosslinked and between the BPA-crosslinked variants were statistically significant, with the number for the BPA-crosslinked attached bacterial cells at least twice as high. A similar high number of attached bacterial cells was reached for THP-1 cells either crosslinked with *para*-formaldehyde or BPA. Both Jurkat and THP-1 cells show no adhesion to bacterial cells with non-crosslinked endogenous OmpX levels. These results suggest that OmpX specifically binds to Jurkat and THP-1 cells upon higher expression levels. Both EL2 and EL3 seem to play a crucial role since similar ratios of adherent bacterial cells could be observed for both BPA-containing OmpX variant proteins. Additionally, BPA shows a slightly higher crosslinking compared to *para*-formaldehyde.

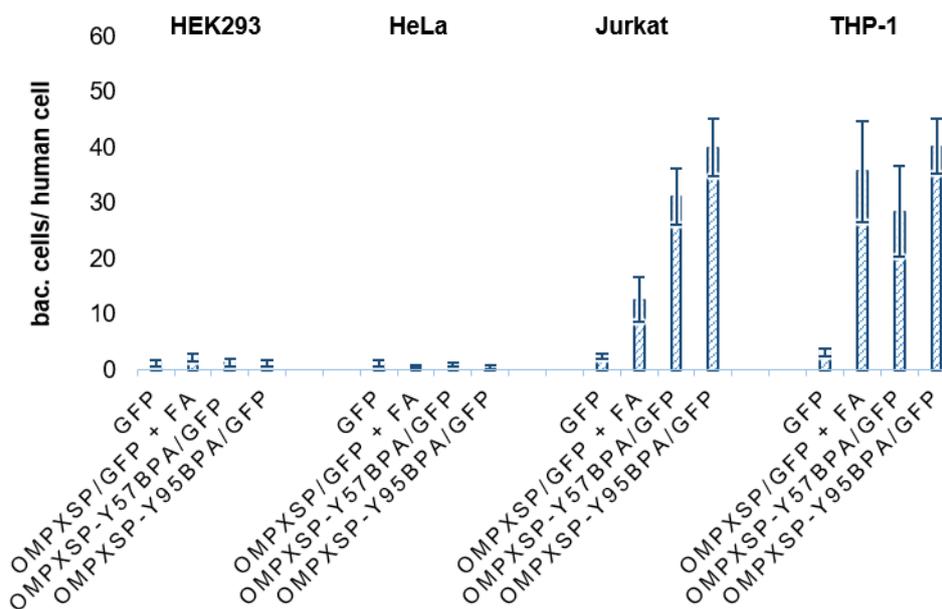


Figure 3.6: Statistical analysis of adhesion by bacterial cells expressing different *ompX* variants. Eukaryotic cells grown on 35 mm glass bottom dishes (14 mm glass diameter) were infected with *E. coli* expressing *ompX* variants at an MOI of 50 and incubated for 1 h, followed by crosslinking for 1 h. The bars represent means plus standard deviations from three independent performed experiments. Abbreviations: GFP, bacterial cells expressing only endogenous *ompX*; OmpXSP, bacterial cells overexpressing *ompXSP* crosslinked with *para*-formaldehyde, OmpXSP-Y57BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 57 crosslinked with 365 nm light, OmpXSP-Y95BPA, bacterial cells overexpressing *ompXSP* containing BPA at position 95 crosslinked with 365 nm light.

This coincides with work done by Delneste and coworkers using fluorescence-activated cell sorting (FACS) showing also that OmpX from *E. coli* bind various human and murine antigen presenting cells such as monocytes, macrophages and T cells and is internalized by them, but surprisingly does not activate them [240]. It further might suggest that OmpX binds similar as OmpA to cells involved in the immune reaction but not to the same structures. This indicates that OmpX-binding molecules or structures may exist on Jurkat (T lymphocytes) as well as THP-1 (monocytes) cells.

In order to investigate the role of both extracellular loops EL2 and EL3, truncated extracellular loop variants were generated with either one or both extracellular loops 2 and 3 being shortened (Figure 3.7) and additional replacing the loop regions by a glycine-serine-serine-glycine (“GSSG”) motif, yielding OmpXSP-Y57BPA-EL3 (EL3 truncation), OmpXSP-Y95BPA-EL2 (EL2 truncation) and OmpXSP-EL2-EL3 (EL2 and EL3 truncation). Extracellular loop truncations does not seem to impact gene expression or protein folding since normal cell growth was observed during gene expression and protein sequence was confirmed with mass spectrometry afterwards.

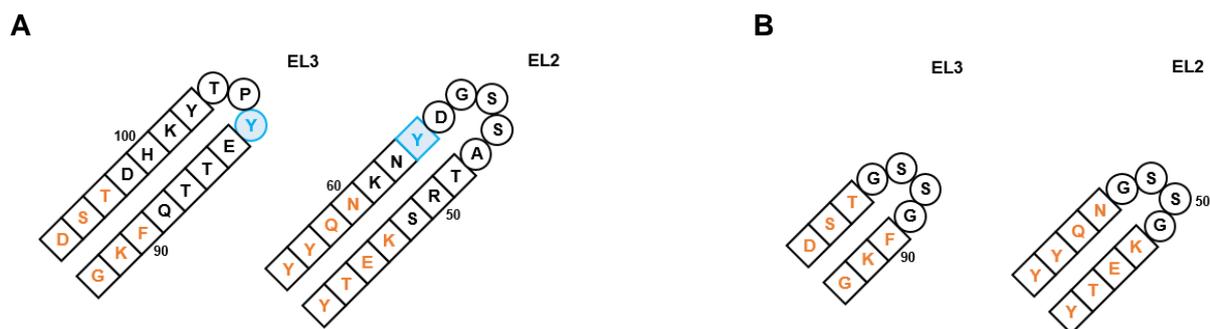


Figure 3.7: Construction of extracellular loop 2 (EL2) and 3 (EL3) mutants. Residues in β -strands are shown in squares; other residues are in circles. Residues from the wild-type OmpX also used for mutants are coloured in orange for better comparison **A**. Arrangement of EL2 and EL3 in wild-type OmpX. Position Y57 in EL2 and Y97 in EL3 for BPA incorporation are coloured in blue. **B**. Arrangement of EL2 and EL3 in OmpX mutant variants. Extracellular loops were truncated and replaced by the “GSSG” motif.

The abilities of the extracellular truncation mutants to attach to HEK293 cells, HeLa cells, Jurkat cells and THP-1 cells were compared using adhesion assays. For OmpXSP-Y57BPA-EL3, OmpXSP-Y95BPA-EL2 and OmpXSP-EL2-EL3 the fixed-cell adhesion assay was repeated as described above. Eukaryotic cells were infected with a MOI of 50 with the extracellular loop truncation variants and OmpXSP-Y57BPA-EL3 and OmpXSP-Y95BPA-EL2 were again photo-crosslinked with 365 nm light and OmpXSP-EL2-EL3 was crosslinked with 3.7 % *para*-formaldehyde, followed by direct imaging with a Leica DM IRBE inverted microscope with a 40 x oil immersion objective (Figure 3.3B). The resulting micrographs are shown in Figure 3.7. Again no adhesion could be observed for any of the variants with HEK293 or HeLa cells. The adhesion ability of the truncated extracellular versions decreased drastically, with basically no adhesion for the double truncated version for both Jurkat and THP-1 cells, whereas the single truncated variants show a weak ability to adhere to Jurkat and THP-1 cells (Figure 3.8 + Figure 3.9) with a slightly higher ability of OmpXSP-Y95BPA-EL2. These findings suggest that both EL2 and EL3 are essential for binding to Jurkat and THP-1 cells and deletion of at least one of the extracellular loops leads to decreased adhesion ability.

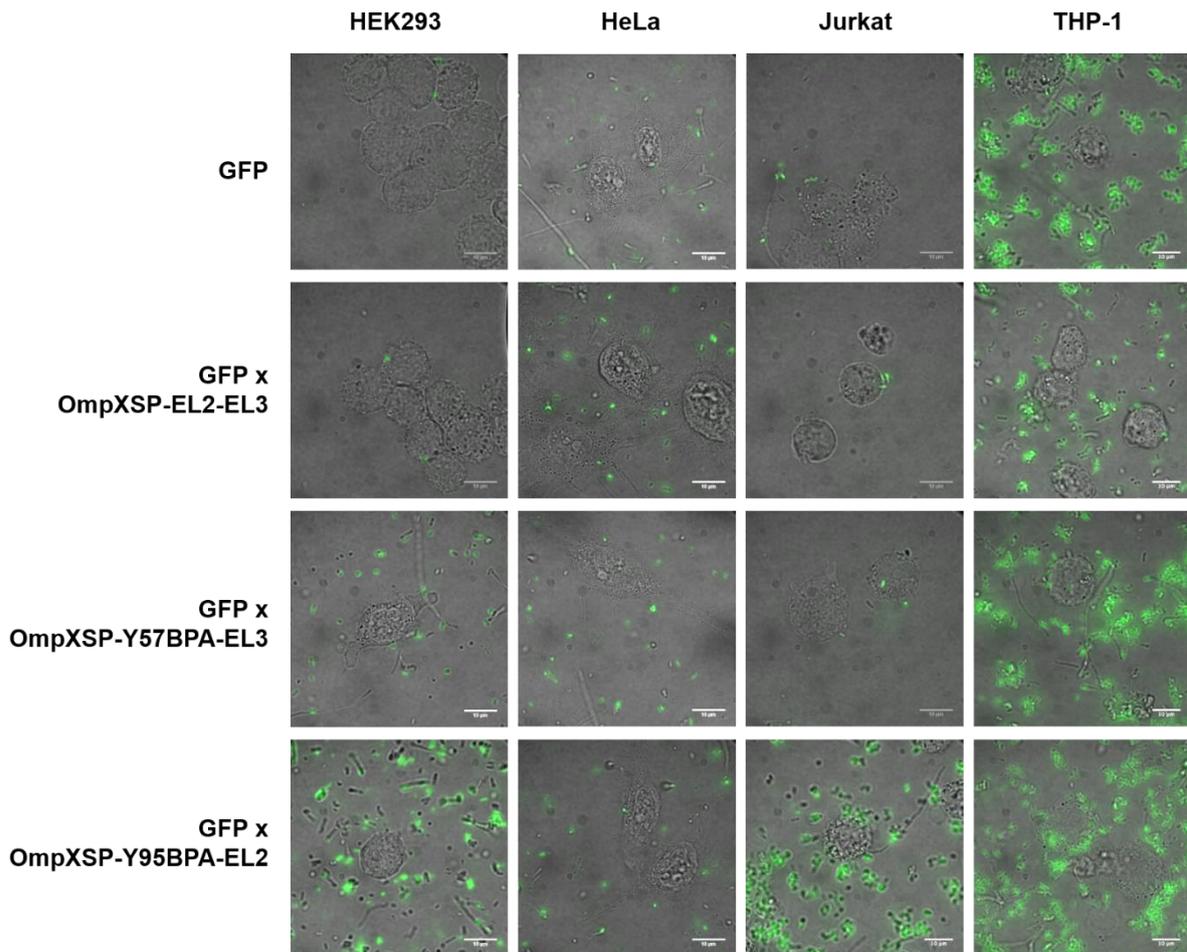


Figure 3.8: Confocal fluorescence micrographs of fixed-cell adhesion assay of OmpX extracellular loop variants. HEK293, HeLa, Jurkat and THP-1 cells were infected with bacteria with a MOI of 50. All bacterial cells coexpress *gfp* for visualisation. Bacterial cells overexpressing *ompXSP* were crosslinked with *para*-formaldehyde. Bacterial cells overexpressing *ompXSP* containing BPA were crosslinked with 365 nm light for 60 min. Scale bars represent 10 μ m in all images. A Leica DM IRBE inverted microscope with a 40x oil immersion objective were used. Abbreviations: GFP, bacterial cells expressing only endogenous *ompX*; OmpXSP-EL2-EL3, bacterial cells overexpressing *ompXSP* with EL2 and EL3 truncated crosslinked with *para*-formaldehyde, OmpXSP-Y57BPA-EL3, bacterial cells overexpressing *ompXSP* containing BPA at position 57 and a truncated EL3 crosslinked with 365 nm light, OmpXSP-Y95BPA-EL2, bacterial cells overexpressing *ompXSP* containing BPA at position 95 and an truncated EL2 crosslinked with 365 nm light.

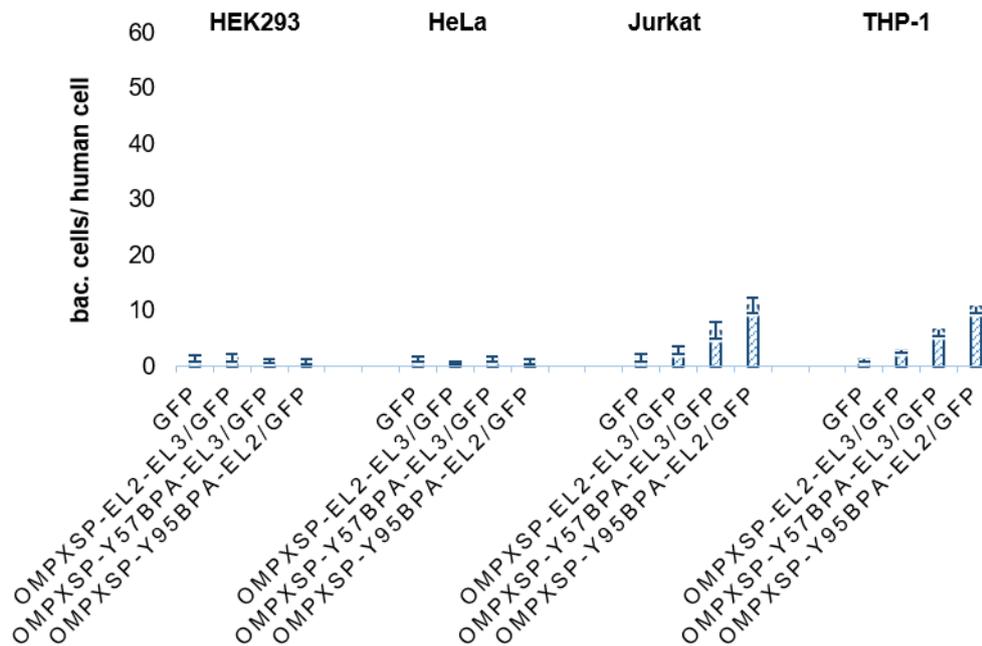


Figure 3.9: Statistical analysis of adhesion by bacterial cells expressing different *ompX* extracellular loop variants. Eukaryotic cells grown on 35 mm glass bottom dishes (14 mm glass diameter) were infected with *E. coli* expressing *ompX* variants at an MOI of 50 and incubated for 1 h, followed by crosslinking for 1 h. The bars represent means plus standard deviations from three independent performed experiments. Abbreviations: GFP, bacterial cells expressing only endogenous *ompX*; OmpXSP-EL2-EL3, bacterial cells overexpressing *ompXSP* with EL2 and EL3 truncated, OmpXSP-Y57BPA-EL3, bacterial cells overexpressing *ompXSP* containing BPA at position 57 and a truncated EL3, OmpXSP-Y95BPA-EL2, bacterial cells overexpressing *ompXSP* containing BPA at position 95 and a truncated EL2.

3.3.2 Membrane preparations

In order to identify possible interaction and binding partners of OmpX, the fixed-cell adhesion assay was performed as previously describe with bacterial cells overexpressing *ompXSP* crosslinked with *para*-formaldehyde, bacterial cells overexpressing *ompXSP* containing BPA at position 57 crosslinked with 365 nm light, and bacterial cells overexpressing *ompXSP* containing BPA at position 95 crosslinked with 365 nm light, as well as the eukaryotic cell lines Jurkat and THP-1 and Hela cells as negative controls. Crosslinked cells were imaged as above to control successful crosslinking and adhesion, followed my membrane preparation. Membranes were isolated by ultracentrifugation and crude membrane extracts were used to perform a SDS-PAGE as well as western blot analysis (Figure 3.10). A faint band appears around 20 kDa corresponding to OmpX. OmpX can be clearly identified by western blot, but no additional bands for possible OmpX crosslinked proteins appear. Subsequently, the gel was cut into small appropriate fractions for each crosslinking experiment and analysed by mass

spectrometry to identify possible interactions partners (data not shown). The received list of possible interaction partners was not conclusive and revealed only impossible interaction partners. This might be due to two possible reasons: either the crosslinked interactions were not stable enough and broke during membrane preparation or OmpX does not bind to any protein or peptide structures but rather binds to oligo- or polysaccharide structures present on the surface of eukaryotic cells involved in immune reactions. To exclude disruption of the crosslinks, crude crosslinked cell extracts were also analysed by SDS-PAGE and mass spectrometry yielding the same results indicating that OmpX might indeed bind to oligo- or polysaccharide structures present on the human antigen presenting cells.

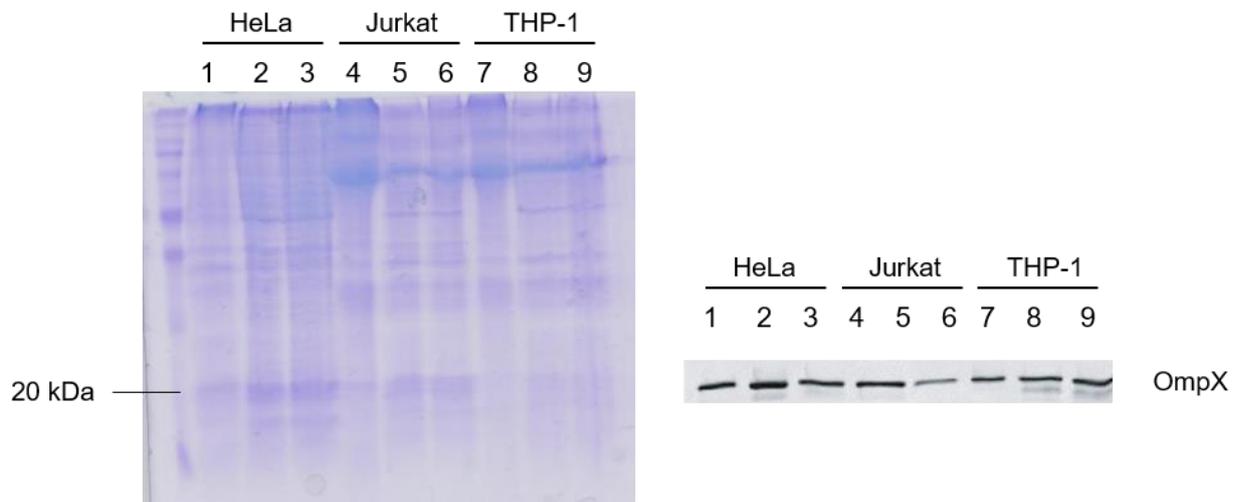


Figure 3.10: SDS-PAGE and western blot analysis of OmpX variant membrane preparations. A faint band corresponding to OmpX becomes visible in the sample after membrane and a strong band in the western blot appears (1 = OmpXSP, 2 = OmpXSP-Y57BPA, 3 = OmpXSP-Y95BPA, 4 = OmpXSP, 5 = OmpXSP-Y57BPA, 6 = OmpXSP-Y95BPA, 7 = OmpXSP, 8 = OmpXSP-Y57BPA, 9 = OmpXSP-Y95BPA).

3.4 Conclusion

Fixed-cells adhesion assays were the key to allow attachment of *E. coli* cells to human antigen presenting cells. Crosslinking with BPA improved stabilization and fixation of crosslinks between bacterial and eukaryotic cells compared to conventional *para*-formaldehyde crosslinking. Consistent with previous studies, OmpX binds to monocytes and T lymphoblasts of the human immune system but not to human epithelial cells. Overexpression of *ompX* is required for establishment of binding indicating that upon recognition and binding of antigen presenting cells a complex signal cascade needs to be activated. The adhesion of bacterial cells is established by the extracellular loops 2 and 3 of OmpX since extracellular loop truncations lead to much less effective binding. Analysis of membrane fractions of crosslinked cells could not reveal any peptide-interactions but might indicate a possible interaction

with distinct oligo- or polysaccharide structures found on the cell surface. In the future, highly interesting information about the still poorly understood binding mechanism as well as possible interaction partners might be obtained from saccharide microarrays or peptide microarrays for high-throughput screenings. Further experiments might also provide valuable insights into possible binding motifs and mechanism.

SECTION C

RESIDUE-SPECIFIC PROTEIN LABELLING

CHAPTER 4

4. THIOCYANATE LABELLING IN LYSOZYME

Parts of this work were accomplished in collaboration with Dr. Stephan Niebling (Huse Lab, University Hamburg, Germany).

4.1. Introduction

Vibrational spectroscopy utilizes the interaction of light with matter on the femtosecond timescale, making it a perfect tool for the study of structure and dynamics in a wide variety of molecular systems and biological applications [241-244]. Each molecule absorbs specific frequencies in dependence on its structural characteristics including spatial arrangements, constitution of atoms as well as the multiplicity of the related chemical bonds. As a result the frequency of the absorbed radiation of a molecule coincides with its vibrational frequency which occurs in different so-called vibrational modes that can be classified according to symmetry and type of mode: symmetric and antisymmetric stretching, scissoring, rocking, wagging and twisting. The activation of a vibrational mode in the context of infrared spectroscopy requires a change in the dipole moment [245]. The endogenous vibrational modes of the protein backbone including the amide I mode (1600 – 1700 cm^{-1}), amide II mode (1480 – 1575 cm^{-1}), amide III (1230 – 1330 cm^{-1}) and amide A mode (3250 – 3300 cm^{-1}) has been widely used in the analysis of changes in local structure, solvation status, electrostatics, and dynamics of proteins [246-249]. However, in many cases it is difficult, if not even impossible, to investigate the local structural changes in proteins based on the intrinsic vibrational modes as these represent the global structural state of a protein leading to spectral crowding and difficult signal assignment. To overcome these limitations, recent research focused on the development and application of a variety of methods to incorporate exogenous vibrational that can be used for site-specific studies [244]. An alternative way to study local conformational changes in proteins with vibrational spectroscopy uses specific unnatural amino acids as infrared probes, which can be either site-specifically incorporated for instance *in vivo* via “amber”-codon suppression [48] or residue-specifically by *in vitro* side chain modifications such as cysteine alkylation [250] or cysteine thio-cyanylation [251]. These infrared

probes are often rather small causing minimal structural perturbations, have a strong transition dipole moment allowing low sample concentrations, and show a strong dependence on the physical property of interest, such as the local electric field [252]. In addition, an ideal probe features a simple and localized transition that is only sensitive to its microenvironment, since vibrational couplings and spectral overlap might significantly render the interpretation, as well as a spectrally separated and intense absorption in the infrared spectrum. Commonly used side chain-based infrared probes contain nitrile or azide groups, such as 4-cyano-L-phenylalanine [252, 253], 4-azido-L-phenylalanine [254], β -cyano-L-alanine [255] and S-cyano-L-cysteine [251, 256], which provide a strong infrared signal in an uncongested region of the infrared spectrum of proteins and water, and have been used for diverse applications, ranging from protein folding to enzymatic reactions (Figure 4.1). Their absorption is highly characteristic in shape and spectral position to the local chemical environment, making them sensitive probes of the surrounding structure and dynamics and providing highly localized information on structural dynamics. The spectroscopic behaviour of the infrared active functional groups has been analysed by several groups in order to draw a more quantitative conclusion from spectral positions and shapes [256-261]. However, a broader application is impaired by the lack of computational tools assessing promising labelling positions and evaluating experimental infrared spectra. In particular, a detailed study of the effect of infrared probes on native structure and dynamics is still missing and only a few publications discuss the structural impact of labelling [241, 262].

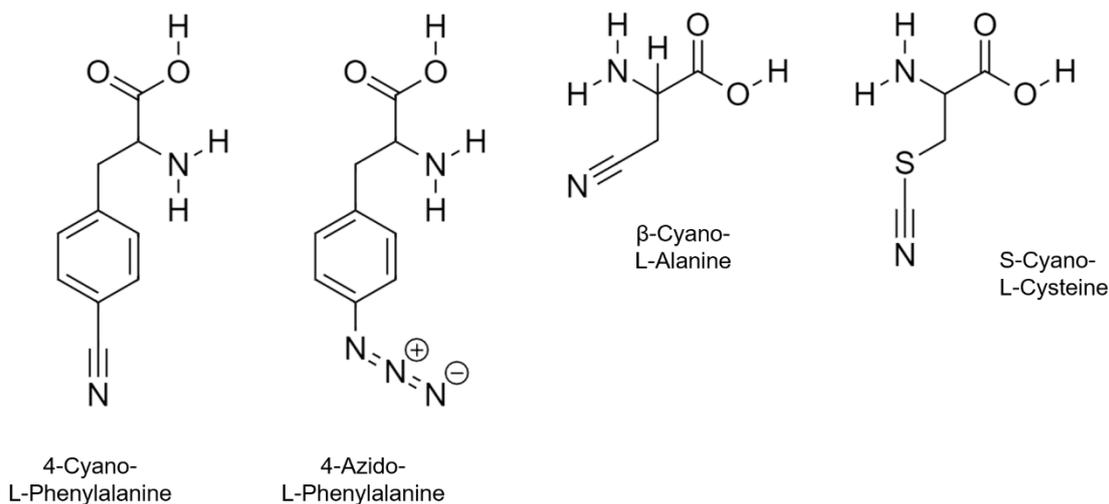


Figure 4.1: Overview of selected site-specific infrared probes.

This chapter is hence reporting on the investigation of structural and dynamic effect of thiocyanate labelling in proteins with infrared spectroscopy, MD simulations and neutron spectroscopy. To assess the structural and dynamic effect of thiocyanate labelling, S-cyano-L-cysteine (Figure 4.1.) was selected as unnatural amino acids candidate due to its extremely small size and its vibrational frequency location in an isolated region of the spectra. T4 lysozyme [263] was chosen as a model system since it is a small, stable protein, contains two intrinsic unpaired cysteine residues, is one of the best-characterized model proteins [264] and could be easily produced in sufficient amounts. As mentioned, T4 lysozyme contains only two unpaired cysteine residues at position 54 and position 97, of which the cysteine at position 54 is slightly more buried inside the protein and not solvent accessible, whereas the cysteine at position 97 is more solvent accessible (Figure 4.2.A). These distinct characteristics of the two cysteines allow the investigation of the sensitivity of the infrared label with respect to differences in the chemical protein microenvironment. Vibrational spectroscopy was complemented by molecular dynamics (MD) simulations as well as neutron spectroscopy allowing conclusions about global changes in protein structure. Whereas vibrational spectroscopy only allows the analysis of the local microenvironment of the probes, neutron spectroscopy offers the possibility to measure the atomic and magnetic motions of atoms in a complete molecular system by observing the changes in the energy of the neutron as it scatters from the sample and can be used to probe a wide variety of different physical phenomena such as the motions of atoms (diffusional or hopping), the rotational modes of molecules, molecular vibrations, or even electronic transitions.

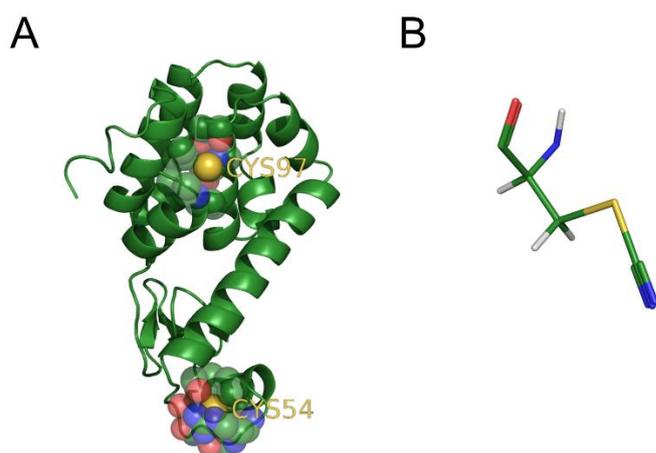


Figure 4.2: T4 Lysozyme as model system to investigate the influence of infrared label in local structural changes. **A.** Structure of wild-type T4 lysozyme (pdb: 256L). Wild type T4 lysozyme contains two cysteines: one of which is more buried inside the protein (C54) whereas the other is more solvent accessible (C97). **B.** Cyanate labelled cysteine yielding S-cyano-L-cysteine.

4.2. Materials and Methods

4.2.1. Cloning, expression, purification and cyanylation of T4 lysozyme variants

The gene for wild-type T4 Lysozyme was a gift from Brian Matthews (Addgene plasmid #18110) in the pHS1403 expression vector. The gene was sequenced to confirm the location of cysteine residues. The T4 lysozyme WT gene was cloned into a pET28a plasmid with a C-terminal His₆-tag. Single cysteine-containing mutant variants of the T4 lysozyme WT protein were generated by site-directed mutagenesis. Proteins were overexpressed in *E. coli* strain BL21(DE3) Gold in terrific broth (TB) medium. A single colony of the cells, grown on a LB-agar plate with ampicillin (100 µg/ml) overnight at 37 °C, was used to inoculate a 20 ml pre-culture with LB-medium containing the same antibiotic. The pre-culture was grown for 16 hours at 37 °C at 180 rpm and subsequently used to inoculate the expression culture (1L TB-medium at 37 °C und 180 rpm, supplemented with the same antibiotic). Gene expression was induced at an OD₆₀₀ of 0.6-0.8 by addition of 1 mM isopropyl thiogalactopyranoside (IPTG), the temperature was reduced to 20 °C and the expression was conducted for 16 hours. The cells were harvested after 16 hours by centrifugation at 4000 x g for 20 minutes at 4 °C. Cells were disrupted by sonication twice for 3 min on ice (30 % duty cycle). The lysate was pelleted at 20,000 x g for 30 min to remove cell debris and the protein was obtained in the supernatant. Proteins were purified by a Ni-NTA resin purification system. Metal ion affinity chromatography using a Ni-NTA column with gravity flow was applied as a first purification step. All steps were performed at room temperature with ice-cold buffers. The column was equilibrated with lysis buffer (40 mM Tris pH 8.5, 300 mM NaCl, 5 % glycerol) and the cleared lysate was loaded on the column. The flow-through was collected and the resin was washed twice with 20 ml lysis buffer supplemented with 20 mM imidazole, to remove unspecific bound protein. The protein was eluted in several steps with 10 ml elution buffer containing 40 mM Tris pH 8.5, 300 mM NaCl, 5% glycerol, 200 mM imidazole. The elution fractions were analysed by SDS-PAGE (see Appendix – General Sample Preparation Techniques) and suitable fractions were combined. The combined fractions were dialyzed overnight against imidazole free elution buffer at 4 °C. Subsequently, the sample was concentrated to suitable volumes using a Vivaspin 6, 5-kDa MWCO concentrator (Sartorius, Germany).

To obtain single-labelled T4 lysozyme variants, the purified single cysteine containing mutant variants were incubated at room temperature for 15 min with 1.1 molar equivalents of 5,5'-dithiobis-(2-nitrobenzoic acid) [251]. When the reaction was initiated, a rapid rise ($t_{1/2} < 1$ min) in TNB concentration was observed by monitoring the absorbance at 412 nm. 20 equivalents of potassium cyanide were

added and incubated at room temperature for 30 min before applying it to a size exclusion column (PD-10, GE Bioscience), equilibrated with 40 mM TrisHCl, pH 8.5. The protein was concentrated using an Amicon Ultra-4, 5-kDa MWCO concentrator. To obtain double-labelled T4 lysozyme, the purified wild-type protein was incubated at room temperature for 15 min with 2.2 molar equivalents of 5,5'-dithiobis-(2-nitrobenzoic acid). Afterwards, 40 equivalents of potassium cyanide were added and incubated at room temperature for 30 min before applying it to a size exclusion column (PD-10, GE Bioscience), equilibrated with 40 mM TrisHCl, pH 8. The protein was concentrated using an Amicon Ultra-4, 5-kDa MWCO concentrator. For infrared measurements at pH 6, the thiocyanate-labelled proteins were dialyzed for 16 hours against 40 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6 and concentrated using an Amicon Ultra4, 5-kDa MWCO concentrator.

4.2.2. Infrared spectroscopy

Infrared spectra were acquired on a Bruker Vertex 70 FT-IR spectrometer. Each measurement was performed using a Harrick cell with 2 mm CaF₂ windows and a 25 μm teflon spacer as sample cell loaded with 30 μl 3 – 4 mM protein solution. Experimental spectra were smoothed in the spectral window between 2100 cm⁻¹ and 2250 cm⁻¹. To estimate the number of spectral components, the second derivate of the smoothed spectrum was determined. Before fitting the data, a baseline (6th order polynomial) taking into account all points except the peak region of the thiocyanate signal (2145 cm⁻¹ – 2190 cm⁻¹ for C54A_C97CN (pH 8) and 2145–2175 cm⁻¹ for the other measurements) was defined. Based on the number of zero-crossings in the second derivative spectrum, two Gaussian curves were fitted to the baseline corrected spectrum. All spectra processing and plotting was done with self-written Python scripts.

4.2.3. Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed by Dr. Stephan Niebling (Huse Lab, University Hamburg, Germany). MD simulations were performed with GROMACS 4.5.5 [265] using the Amber03 force field [266] with parameters as published by the Webb Group (University of Austin, Texas). Initial coordinates were taken from pdb 256L [267]. Single cysteine mutants and its labelled derivatives were created with PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). The peptides were solvated in a cubic box with periodic boundary conditions and a side length of ~78 Å (10 Å initial minimum distance of solute to all boundaries) comprising the peptide and ~15000 H₂O molecules and 8 chlorine ions to neutralize the protein charge. For all systems, the same molecular

dynamics protocol was used. After a steepest descent energy minimisation (convergence criteria 500000 steps or maximum force $< 10 \text{ kJ mol}^{-1} \text{ nm}^{-1}$) two 100 ps equilibration MD runs were performed. The first one in the constant particle number, volume, temperature ensemble (NVT; with modified Berendsen thermostat with velocity rescaling at 300 K and a 0.1 ps time step; separate heat baths for peptide and solvent/ions); the second one in the constant particle number, pressure, temperature ensemble (NPT; Parrinello-Rahman pressure coupling at 1 bar with a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$ and a 2 ps time constant). During both equilibration runs, a position restraint potential with a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ was added to all peptide atoms. For all MD simulations the leap-frog integrator was used with a time step of 2 fs. Coordinates were saved every 2 ps. The same temperature and pressure coupling schemes as applied for the equilibration runs were used for the subsequent MD simulations. All bonds to hydrogen atoms were constrained using the Linear Constrained Solver [268] with an order of 4 and one iteration. A grid-based neighbour list with a threshold of 10 \AA was used and updated every 5 steps (10 fs). The particle-mesh Ewald method [269] was used for long-range electrostatic interactions above 10 \AA with a fourth order interpolation and a maximum spacing for the FFT grid of 1.6 \AA . Lennard-Jones interactions were cut-off above 10 \AA . A long-range dispersion correction for energy and pressure was used to compensate for the Lennard-Jones interaction cut-off. In total six independent trajectories were generated: wild type (WT), WT control, C97A, C97A-C54CN, C54A, C54A-C97CN. For analysis and visualisation of MD trajectories self-written Python scripts using the modules MDAnalysis [270, 271], NumPy [272] and Matplotlib [273] were used.

4.2.4. Neutron spectroscopy

Neutron spectroscopy has been performed in collaboration with Dr. Stephan Niebling (Huse Lab, University Hamburg, Germany). For neutron spectroscopy protein samples were dialyzed against ammonium acetate to exchange out the bulk of residual nonvolatile counterions. To this end, lysozyme was dissolved in 2.0 mM ammonium acetate buffer of pH 9.5 (at 4°C) and to concentrations of 100 mg/ml. Protein samples then were dialyzed against 20 volumes of the same solution for 4 h at 4°C . Followed by flash-freezing in liquid N_2 , the samples were lyophilized on a Modulyo freeze drier (Edwards, United Kingdom) overnight. Samples were resuspended in D_2O pH 8.5 to concentrations of 100 mg/ml to exchange internal protons, flash-frozen again in liquid N_2 and lyophilized using a freeze dryer overnight. For liquid neutron spectroscopy measurements at IN5 and IN16B at Institute Laue Langevin (ILL, Grenoble, France), samples were resuspended to a protein concentration of 150 mg/ml in 40 mM TrisHCl pH 8.5 in D_2O . For each measurement 1.5 ml protein solution were used. All samples were measured in round aluminium containers (Tübingen-type containers).

IN16b is a sub-micro-eV energy resolution backscattering spectrometer with a very high count rate and wide dynamic range. The flux at the sample position is enhanced by a new 'quasi-ballistic' neutron guide and its focusing to a fast rotating Phase Space Transformation chopper at the end-of-guide position. The chopper together with a linear motor Doppler drive enables to double the energy transfer range at sub-micro-eV energy resolution. The vertical increase of the analyser surface together with a new vertically position sensitive multi-tube-detector will nearly double the count rate for the large angles. The background is optimized by placing the analysers and the entire flight path in vacuum and by using a background chopper. The beam size at the sample is 30 x 30 mm². At IN16B T4 Lysozyme_C54CN_C97A, the unlabelled variant T4 Lysozyme_C97A as well as buffer only and vanadium references were measured. Samples were measured at 295 K with 10 Å for 7.5 hours, buffer only was measured for 6.5 hours and the vanadium reference for 2 hours. IN5 is a high precision direct geometry time-of-flight spectrometer to study low-energy transfer processes as a function of momentum transfer. The time regime is between a few ps – few 10s of ns. Typically this instrument is used for measurements in the small energy and momentum transfer region with values of about 1 % for the energy resolution (e.g. quasi-elastic scattering in liquids and inelastic scattering with small energy transfers in the range 10 µeV - 100 meV). Samples were measured with a chopper velocity of 7100 rpm. The beam size at the sample is 15 x 50 mm², elastic energy resolution at 5.0 Å, 8500rpm is ~ 100 µeV and flux at the sample at 5.0 Å is 6.83×10^5 [n/cm²/s]. At IN5 T4 Lysozyme_C54CN_C97A, the unlabelled variant T4 Lysozyme_C97A as well as buffer only and vanadium references were measured. Samples were measured at 295 K with 10 Å, 295 K with 5 Å, 283 K with 5 Å and 150 K with 5 Å for each 1 hour. Data were processed using LAMP (Large Array Manipulation Program, ILL, France) and Mantid software [274, 275]. TOSCA is an indirect geometry spectrometer optimised for the study of molecular vibrations in the solid state with a beam size at sample of 40 mm x 40 mm and an energy resolution of ~1% ΔE/E. For dry protein measurements at TOSCA at ISIS (Didcot, United Kingdom), between 0.2 – 0.5 g protein samples were used for each measurement. All samples were prepared in a 4 x 4.8 cm aluminium foil sachet and measured in 1 mm flat aluminium containers sealed with indium wires. At TOSCA T4 Lysozyme_C54CN_C97A, the unlabelled variant T4 Lysozyme_C97A, T4 Lysozyme_C54A_C97CN, the unlabelled variant T4 Lysozyme_C54A, as well as aluminium can with an empty foil sachet as reference were measured. Samples were measured at 10 K for at least 8.5 hours. Data were processed by Mantid software [274, 275].

4.3. Results and Discussion

4.3.1. Recombinant expression, purification and cyanylation

A broad application of vibrational probes to study structure and dynamics in a wide variety of molecular systems is impaired by the lack of detailed studies investigating the effect of infrared probes on native structure and dynamics. Based on this lack of knowledge, single-cysteine T4 lysozyme mutant variants were designed for cysteine-specific thiocyanate labelling in order to analyse the effect of the vibrational probe on local and global structural changes. The two single-cysteine T4 lysozyme mutant variants (T4 Lysozyme_C54A and T4 Lysozyme_C97A) were designed by site-directed mutagenesis. After mutagenesis, the mutant variants were expressed recombinantly in *E. coli* BL21(DE3) Gold for 16 hours. The mutant variant proteins were purified from the cleared lysate by Ni-NTA affinity chromatography. To obtain single S-cyano-L-cysteine labelled T4 lysozyme variants, the purified mutant variants T4 Lysozyme_C54A and T4 Lysozyme_C97A were post-translationally modified via cysteine cyanylation in a two-step chemical reaction [251] yielding T4 Lysozyme_C54A_C97CN and T4 Lysozyme_C54A_C97CN, respectively (Figure 4.3.). Subsequently, size exclusion column was performed as a second purification step to remove remaining impurities. SDS-PAGE was used to analyse the purity of the proteins during purification, labelling and concentration.

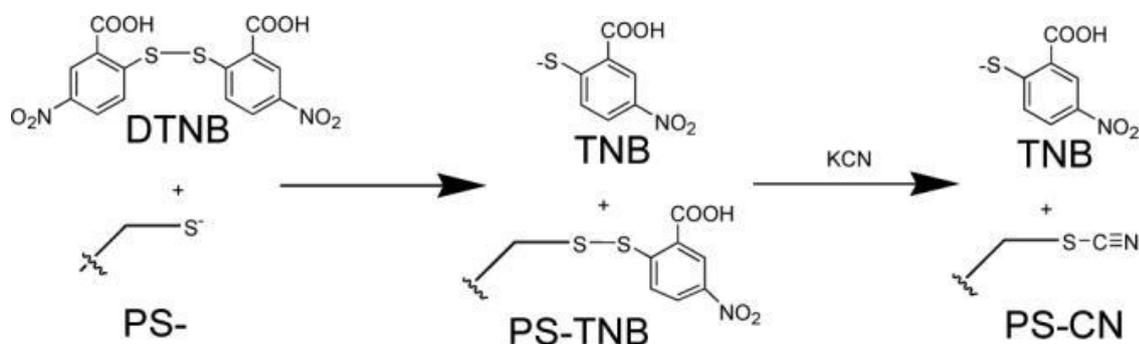


Figure 4.3: Cysteine cyanylation. Purified protein cysteine thiols (PS) were reacted with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) to form the mixed protein-thionitrobenzoic acid disulfide (PS-TNB) followed by displacement by cyanide (CN^-), to form the protein-thiocyanate (PS-CN) after addition of potassium cyanide (KCN) (Adopted from Fafarman et al., 2006 [251]).

4.3.2. Infrared spectroscopy

Infrared spectroscopy of single S-cyano-L-cysteine labelled T4 lysozyme mutant proteins were acquired to investigate the sensitivity of the vibrational label towards differences in its microenvironment. For both variant proteins T4 lysozyme_C54A_C97CN and T4 lysozyme_C54CN_C97A the infrared spectra at pH 6 and pH 8 were recorded in a frequency range of 800 cm^{-1} to 4000 cm^{-1} . The full infrared spectrum for T4 lysozyme_C54CN-C97A at pH 8 is exemplary shown in Figure 4.4. The signal of the thiocyanate label appears around 2165 cm^{-1} in the spectra (Figure 4.5.). This signal can be assigned to the distinct nitrile $\text{C}\equiv\text{N}$ stretching usually occurring between 2153 and 2164 cm^{-1} in proteins [255, 256, 258]. T4 lysozyme_C54A_C97CN at pH 8 shows a broad signal for the label in the spectra between 2150 cm^{-1} and 2185 cm^{-1} with its maximum at 2165 cm^{-1} , whereas the signal for T4 lysozyme_C54CN_C97A at pH 8 is slightly shifted and rather narrow between 2155 cm^{-1} and 2175 cm^{-1} with its maximum at 2165 cm^{-1} . In contrast to the measurements at pH 8, the signals for the labels at pH 6 are shifted to lower wavenumbers: for T4 lysozyme_C54A_C97CN between 2155 cm^{-1} and 2175 cm^{-1} with its maximum at 2163 cm^{-1} and T4 lysozyme_C54CN_C97A between 2150 cm^{-1} and 2175 cm^{-1} with its maximum at 2162 cm^{-1} . The discrepancy in signal width at pH 8 indicates that the proposed different microenvironments of the two cysteines (of which the cysteine at position 54 is more buried inside the protein, hence less solvent-exposed, and the cysteine at position 97 is more solvent-exposed) might have an influence on the spectral behaviour of the probe. These might directly reporting on changes in the probes' local environment which have been reported previously such as local electrostatic changes as well as hydrogen-bonding effects [241, 276]. An effect of the introduced cysteine to alanine mutations (C54A / C97A) on the structure of lysozyme leading to an alternated infrared spectra can be excluded as shown by Brian W. Matthews [277]. The frequency-shifting of the infrared signals of the thiocyanate label at pH 6 suggests a dependency of the infrared signal on pH indirectly indicating the use of this specific vibrational probe as indicator for local changes in pH. Interestingly, at pH 6 the signal of the probe appears rather similar in spectra, compared to the notable shift at pH 8.

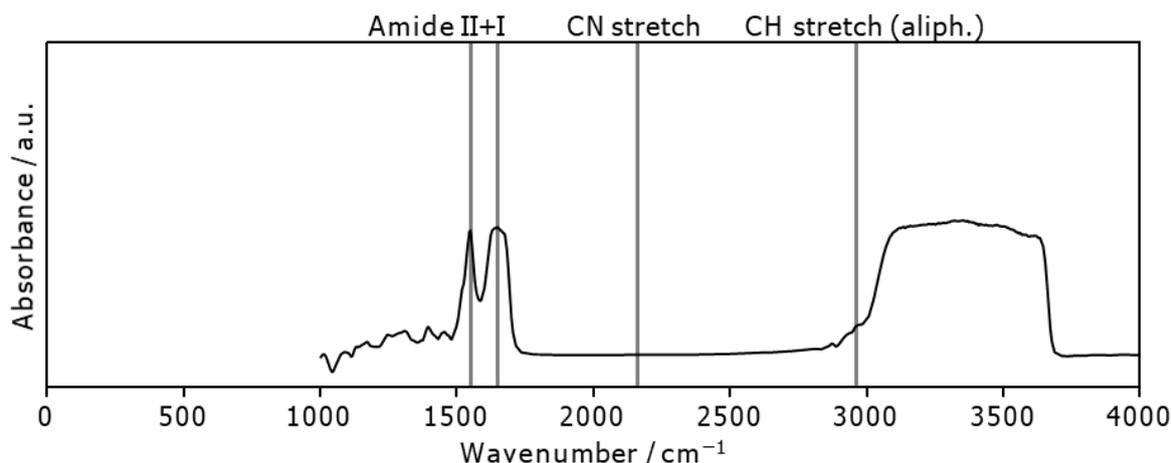


Figure 4.4: Full spectra of T4 lysozyme-C54CN-C97A at pH 8. Amide I and II, $C\equiv N$ stretching as well as aliphatic C–H stretching are visible.

A good initial approximation to estimate the number of components of the signal is to count the zero-crossings in the second derivative. The second derivative of the smoothed spectra in this region shows two zero-crossings suggesting at least two components. After smoothing and baseline correction, two Gaussian curves were fitted whose sum is in good agreement with the experimental data. It is important to note that the choice of Gaussian curves for the fitting is motivated by practical aspects such as better descriptions of the experimental data. The purpose of these fits is to measure the absorption band positions as well as widths in order to quantitatively compare these parameters of the label in different positions. The infrared signal of the label can be described by the sum of two Gaussians. The first Gaussian can be assigned to the free label (Free), whereas the second Gaussian can be assigned to a hydrogen bonded (HB) label. By integrating the two Gaussians, the ratio between the two species can be estimated in a first approximation (Table 3). The fitted Gaussian curves of T4 lysozyme_C54CN_C97A show maxima at 2162 cm^{-1} for the free label and 2166 cm^{-1} for the hydrogen-bonded label with a width of 6 cm^{-1} and 5 cm^{-1} , respectively and for T4 lysozyme_C54A_C97CN maxima at 2162 cm^{-1} for the free label and 2171 cm^{-1} for the hydrogen-bonded label with a width of 7 cm^{-1} and 8 cm^{-1} , respectively. Both cyanylated cysteines do not only exhibit strong differences on the width of the curves, but also show strong differences in the ratio in free label to hydrogen-bonded label: For T4 lysozyme_C54A_C97CN a much higher band can be observed for the free label compared to the hydrogen-bonded label and, whereas for T4 lysozyme_C54CN_C97A exactly the opposite can be observed with a higher band for the hydrogen-bonded label compared to the free label. This is of particular interest as both cysteines have a proposed different local environment.

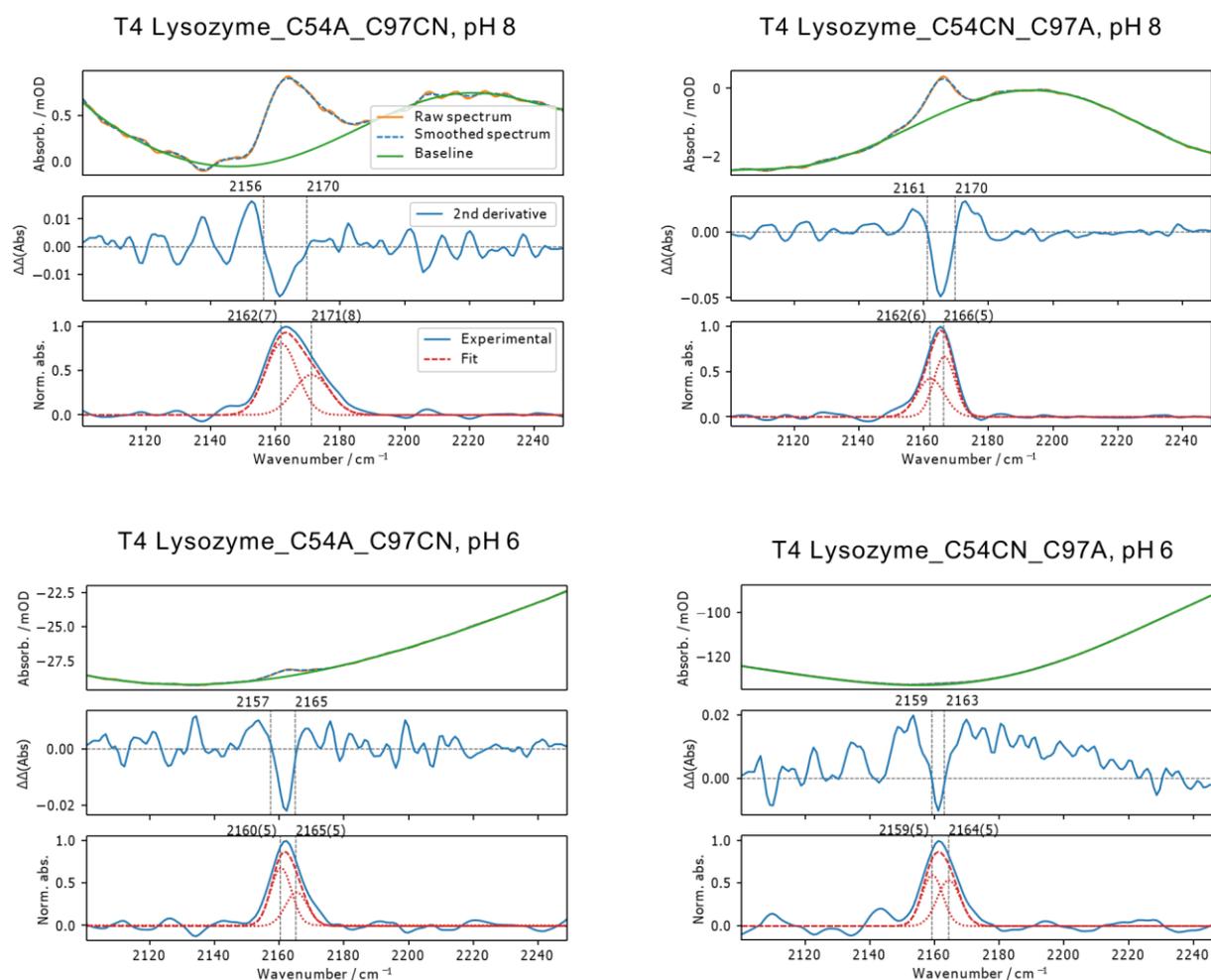


Figure 4.5: Infrared spectra of the vibrational active thiocyanate label from single S-cyano-L-cysteine labelled T4 lysozyme variant proteins at pH 8 and pH 6. The signal of the thiocyanate label appears around 2165 cm^{-1} in the experimental data. The second derivative of the smoothed spectra in this region shows two zero-crossings suggesting at least two components. Zero-crossing values are given above. After baseline correction, experimental data was fitted with two Gaussian curves. Maxima of the Gaussian curves are indicated above as well as the width in parentheses. Infrared signal of the thiocyanate label at pH 8 show slight alternations in linewidth indicating that the microenvironment might have an influence on the behaviour of the probe. Infrared signals of the thiocyanate label at pH 6 are blue-shifted suggesting a dependency of the infrared signal on pH.

These findings of a slightly shifted spectral position to lower frequencies of T4 lysozyme_C54A_C97CN as well as the an alternate broader spectral shape of the signal compared to T4 lysozyme_C54CN_C97A coincides with previous findings showing also a shift to lower wavenumbers as well as an alternate spectral shape for 4-cyano-L-phenylalanine (Figure 4.6) [252, 278]. Interestingly, this indicates that although the cysteine at position 97 is more solvent-accessible the label itself is located in more hydrophobic environment and the cysteine at position 54 buried deep inside the protein is situated in a more aqueous environment.

Table 3: Comparison of the infrared signal of the thiocyanate label of single S-cyano-L-cysteine labelled T4 lysozyme variant proteins at pH 8. Baseline corrected, experimental data was fitted with two Gaussian curves. The first Gaussian curve can be assigned to the free label, whereas the second Gaussian can be assigned to a hydrogen bonded label. Infrared signals of the thiocyanate label at pH 8 show slight alternations in linewidth for free and hydrogen-bonded labels. Abbreviations: Free, free infrared label; HB, hydrogen-bonded label.

	Type	Position	Width
C54CN	Free	2162 cm^{-1}	6 cm^{-1}
	HB	2166 cm^{-1}	5 cm^{-1}
C97CN	Free	2162 cm^{-1}	7 cm^{-1}
	HB	2171 cm^{-1}	8 cm^{-1}

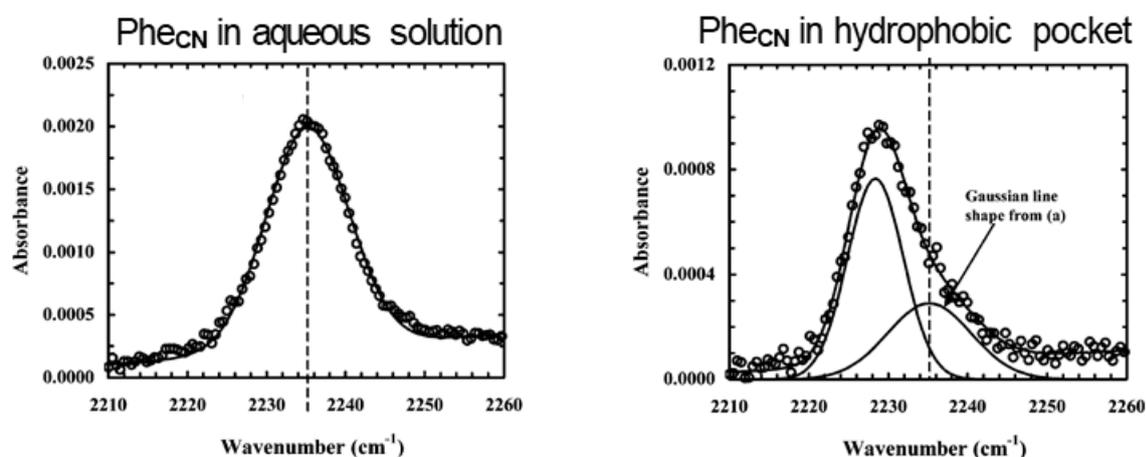


Figure 4.6: Spectral behaviour of 4-cyano-L-phenylalanine in different local environments. In aqueous solutions the spectral signal of 4-cyano-L-phenylalanine is located at 2235cm^{-1} . In hydrophobic environments the signal shifts to lower wavenumbers and shows a different spectral shape (Adopted from Lindquist et al., 2009 [278]).

4.3.3. Hydrogen bond analysis by molecular dynamics simulations

As static structures can explain the difference in linewidth of the infrared signals only to a limited extent for the two different labelled cysteine systems, 1000 ns molecular dynamics (MD) simulations have been performed to structurally explain the differences observed in experimental infrared signals and screen for possible hydrogen bond contacts between the infrared label and hydrogen bond donors in its vicinity such as solvent molecules or polar groups of the protein itself. Hence, a distance threshold

of $d < 3 \text{ \AA}$ and an angle threshold of $>120^\circ$ ($\text{CN}\cdots\text{H}$) was chosen to obtain hydrogen bond ratios that are in good agreement with the experimental infrared data. The smoothed curves for the two systems show distinct differences in the hydrogen bond modes (Figure 4.7). Whereas T4 lysozyme_C54CN_C97A shows a relatively stable ratio between hydrogen bond modes with simultaneous hydrogen bonds with the solvent and the proteins, T4 lysozyme_C54A_C97CN shows two populations with a slow exchange resulting in hydrogen bonds either with the solvent or the protein suggesting hydrogen bond heterogeneity around the infrared label. This might explain the larger linewidth of the infrared signal of C97CN.

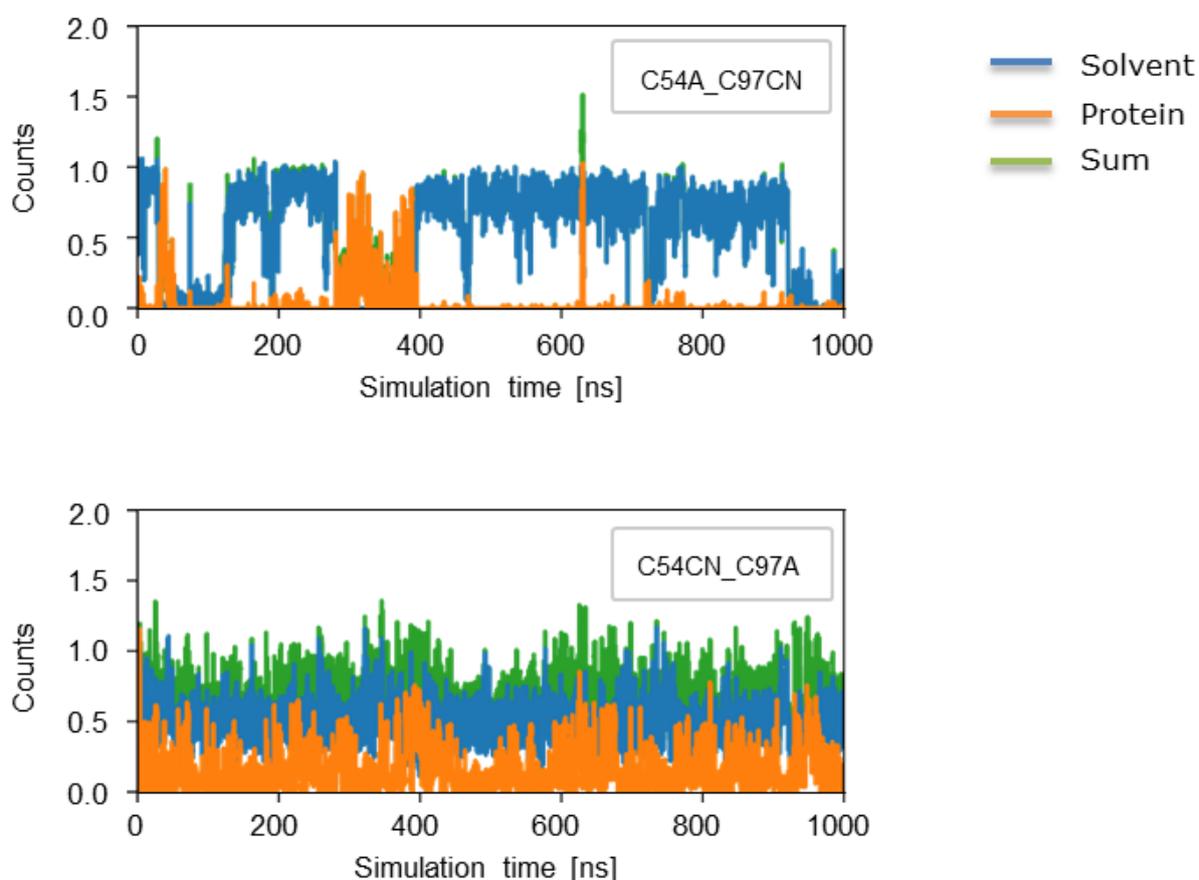


Figure 4.7: Hydrogen bond analysis of the two different cyanylated cysteine systems. Whereas C54CN shows a relatively stable ratio between hydrogen bond modes, C97CN shows two populations with a slow exchange. Hydrogen bond interaction with the solvent are shown in blue, with the polar groups of the protein in orange and the total in green (*At the date of this thesis, the trajectories with the Amber03 parameters, as described in the methods section, were still being processed. Instead older data for the Amber99SB force field is shown here*).

4.3.4. Neutron Spectroscopy

Infrared spectroscopy allows us to investigate the local microenvironment around the infrared labels. Neutron spectroscopy complements these experiments by investigating the effects of labelling on global dynamics. In particular we want to study the influence of the attached infrared probe on low-frequency protein modes by acquiring density of states curves for pairs of the labelled and unlabelled single cysteine T4 lysozyme variant proteins with inelastic neutron scattering techniques provided by the IN5 and IN16B beamlines of Institute Laue-Langevin (ILL, Grenoble, France) as well as the TOSCA beamline at ISIS (Didcot, United Kingdom). While experiments at the ILL allow us to analyse low-frequency modes of the protein around 3-10 THz, experiments at the TOSCA beamline allow us to study the weak label signal at 2160 cm^{-1} and the low THz regime (2-10 THz, approx. 12-40 meV) at the same time enabling the analysis of both effects in the microenvironment around the label and the global low-frequency modes of the protein simultaneously. In advance to the neutron spectroscopy experiments, long MD simulations (1 μs simulation time) of the two labelled T4 lysozyme variants and their unlabelled counterparts (Figure 4.8 A+B) were performed as computational tool. The analysis scheme based on internal coordinates can be used to assess the effect of labelling on the structure and dynamics. These predict pronounced notable effects of labelling in position C54 both for global structure and dynamics, whereas for position C97 only minor effects were predicted (Figure 4.8 B).

For experiments at IN5 and IN16B beamlines at ILL (Grenoble, France) protein solutions in D_2O were measured. In a simple first observation, the measurements provided sufficient strong signals with a count rate of 7800 c/s for the unlabelled proteins and 12000 c/s for the labelled protein in contrast to 430 c/s for the buffer control at IN5 and a count rate of 4500 c/s for the unlabelled proteins and 8000 c/s for the labelled protein in contrast to 13 c/s for the buffer control at IN5. The analysis and evaluation of the neutron spectroscopy measurements is very complex and has not yet been finally completed. For experiments at TOSCA beamline at ISIS (Didcot, United Kingdom) dry protein samples were measured, whereby for each measurement 0.2 – 0.5 g protein were used. Although all samples were lyophilized, the labelled proteins did not dry properly yielding a yellow honey-like state indicating that the labelling might influence drying capacity. The preliminary results of the inelastic neutron spectroscopy are summarized in Figure 4.9. For comparison, additionally the spectra of D_2O and H_2O are shown. At first glance, it is obvious that the strong signal of the $\text{C}\equiv\text{N}$ stretching at 2165 cm^{-1} is not visible. This might be due to the limited sensitivity of the TOSCA beamline to only 2000 cm^{-1} wavenumbers or to the poor neutron scattering cross section for vibrations not involving hydrogens.

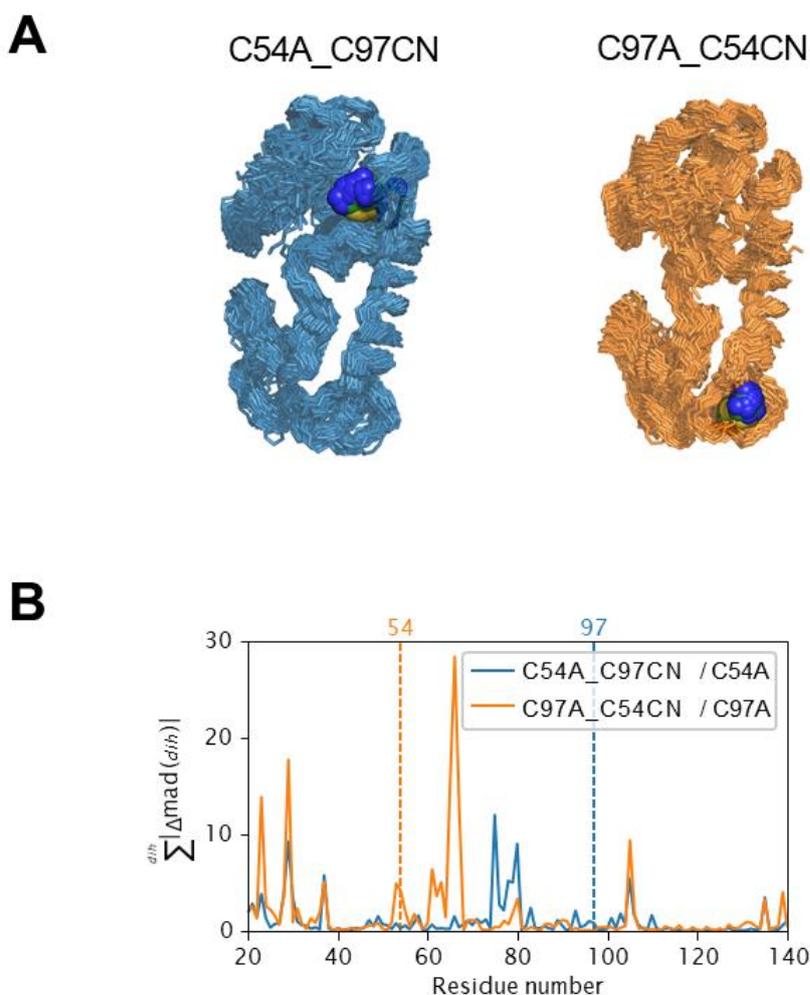


Figure 4.8: Influence of the cyanate-labelling of cysteines on the protein backbone structure/dynamics. **A.** 1 μ s trajectories of T4 lysozyme cyanylated at C54 (right, orange) and C97 (left, blue). The label atoms are shown as spheres in dark blue. **B.** The influence of labelling on backbone structure/dynamics can be assessed by comparing trajectories of labelled and unlabelled proteins with dihedral statistics. The differences in the median absolute deviations (mad) can be used to detect differences in dynamics. The differences of the flexible termini are omitted for clarity (*At the date of this thesis, the trajectories with the AmberO3 parameters, as described in the methods section, were still being processed. Instead older data for the Amber99SB force field is shown here*).

Unlabelled and labelled protein variants show distinct features in the spectra between 1100 cm^{-1} and 1600 cm^{-1} below the amide II signal indicating that upon labelling changes in both systems appear. For all samples the aliphatic C–H stretching is clearly visible (2900 cm^{-1} – 3000 cm^{-1}). For the unlabelled protein variants additional spectral signals appear around 1300 cm^{-1} , whereas the labelled variants show a pronounced triplet signal between 1200 cm^{-1} – 1550 cm^{-1} as well as several bands in the lower frequency range ($< 300\text{ cm}^{-1}$). To exclude that the signals are an effect of uneven water or D_2O content, the spectra were compared to $\text{D}_2\text{O}/\text{H}_2\text{O}$ data. Only the spectra of the unlabelled T4 lysozyme_C54A variant seems to have minor contributions from water or D_2O spectra. To higher frequencies (>1700

cm^{-1}) no strong differences between labelled and unlabelled species could be detected. Interestingly, the both infrared and neutron spectroscopy are very complementary as the infrared spectroscopy features the amide I and II modes, and the $\text{C}\equiv\text{N}$ stretching of the label, whereas in neutron scattering neither the NH/OH stretching nor the amide I or II are visible and the aliphatic C–H stretching as well as several notable spectral bands below the amide II frequency appear, for instance between 1200 cm^{-1} – 1550 cm^{-1} , at 700 cm^{-1} and 900 cm^{-1} . The challenging part will be the assignment of the spectral bands to corresponding vibrational modes. This is still ongoing work.

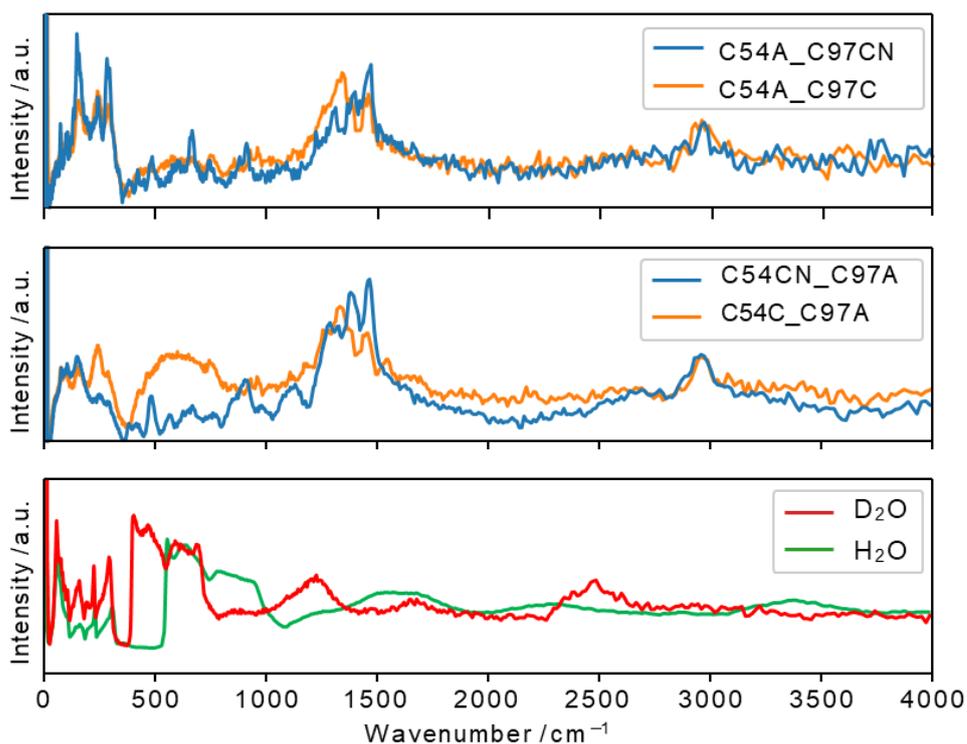


Figure 4.9: Inelastic neutron scattering spectra of the labelled and unlabelled T4 protein variant pairs. The inelastic neutron scattering of T4 lysozyme_C54A_C97CN (blue) and its unlabelled complement (orange) are shown in the upper panel, T4 lysozyme_C54CN_C97A (blue) and its unlabelled complement (orange) are shown in the middle panel and inelastic neutron scattering of D₂O (red) and H₂O (green) in the lower panel.

4.4. Conclusion

Small infrared labels can be used to study local environments in proteins in a site-selective manner. Unlike Förster resonance energy transfer and electron paramagnetic resonance labels, the use of infrared labels is not limited to the protein surface. For instance, they can also be placed inside a protein to study the local environment around its active site. Neutron scattering experiments complement Raman and infrared spectroscopy by delivering information about delocalized vibrational modes independent of spectroscopic selection rules that might be vital for enzyme activity. Combining

these experimental methods opens up new possibilities of a better understanding on how proteins and enzymes work. For a broader application of infrared labels, the structural change upon infrared labelling needs to be assessed. Hence, for the first time the influence of the infrared label S-cyano-L-cysteine on the global protein dynamics of T4 lysozyme was investigated. Molecular dynamics simulations predict a small but significant influence of labelling on dynamics. Additionally, the preliminary neutron scattering experiments supports these findings, but further effort towards the analysis of the data is required, especially the assignment on the low frequency modes. Infrared spectroscopy of the same system at different pH values suggested the use of this specific vibrational probe as indicator for local changes in pH as well changes in solvation-accessibility. Furthermore, it would be also interesting to investigate whether global and local changes due to the label also occur in the crystal structure of the T4 lysozyme variants and could provide further evidence for the previous findings.

APPENDIX

GENERAL SAMPLE PREPARATION TECHNIQUES

Buffer and solution preparation

Buffers and solutions were prepared using de-ionized water from a Milli-Q ultrapure lab water system (Merck Millipore, Germany). Buffer pH was measured using a peqMeter 1.14 (peqLab via VWR, United Kingdom) and monitored while adjusting. All buffers and solutions were filtered through a 0.22 µm Millex-GP filter (Merck Millipore, Germany) prior to usage.

Maintenance and growth of bacteria

All bacterial cells were grown on LB media (Lennox) (Carl Roth, Germany) or Terrific Broth (TB) media (recipe provided in Table 4). All media were autoclaved. For overexpression 900 ml of TB media was supplemented with 100 ml TB phosphate buffer. All cells were grown at 37 °C unless otherwise specified. For antibiotic selection ampicillin was used at a concentration of 100 µg/ml, kanamycin 25 µg/ml and chloramphenicol 34 µg/ml.

Table 4: Growth media

Media	Composition
TB media	24 g/l yeast extract
	12 g/l tryptone
	3.8 ml/l glycerol
TB Phosphate buffer	0.017 M KH_2PO_4
	0.072 M K_2HPO_4

Determination of bacterial growth by optical density

Bacterial growth was determined by optical density using a SmartSpec Plus spectrophotometer (Bio-Rad, USA).

Preparation of electrocompetent *E. coli* cells

E. coli cells were grown in LB media at 37 °C until OD_{600} reached 0.4-0.6. Bacterial culture was chilled on ice or in the cold room for 20-60 min, followed by centrifugation at 4000 x g and 4 °C for 15 min.

The pellet was washed with 800 ml ice-cold ddH₂O, followed by centrifugation at 4000 x g and 4 °C for 15 min. The pellet was washed with 500 ml ice-cold ddH₂O, followed by centrifugation at 4000 x g and 4 °C for 15 min. pellets were resuspended in 20 ml 10 % glycerol and transferred to 50 ml tube and centrifuged for 15 min at 4 °C and 3800 x g. The pellet was resuspended to final volume of 2-4 ml 10 % glycerol. Resuspended cells were aliquoted into 50 µl fractions, frozen in liquid nitrogen and stored in -80 °C until further use.

Protein sample concentration

Protein samples were concentrated using an Amicon Ultra-15 (Merck, Germany) centrifugal filter unit with a molecular weight cut-off (MWCO) of 10 kDa or a Vivaspin 6 (Sartorius, Germany) centrifugal filter unit with a molecular weight cut-off (MWCO) of 5 kDa as appropriate. The protein solutions were then centrifuged at 4 °C at 3,500 x g until the desired concentration was reached. Protein concentrations were determined photometrically using a Thermo Fisher Nanodrop 2000 device, which uses the specific absorption of aromatic amino acids at a wavelength of 280 nm to calculate the concentration according to the equation:

$$c = \frac{A_{280} Mw}{\epsilon d}$$

In the equation above A_{280} is the measured specific absorption at 280 nm, Mw the molecular weight of the protein, d the sample depth and ϵ the extinction coefficient of the protein at 280 nm. A list of the parameters used for the individual proteins can be found in Table 5.

Table 5: Protein properties used for concentration determination protein molecular weights and extinction coefficients used for the protein concentration determination.

Protein	Molecular Weight [kDa]	Extinction Coefficient [$M^{-1} \text{ cm}^{-1}$]
Lysozyme	14	27,000
GFP	27	55,000
OmpX	18	35,000

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a method separating biological macromolecules according to their molecular weight and electrical charge. Protein samples are pipetted into pockets of a vertically orientated polyacrylamide gel and then an electrical current is applied. Due to the varying sizes of individual molecules, the velocity at which they traverse the gel varies, and therefore they travel different distances along the gel in a specified amount of time. Adding

SDS to the gel results in all molecules acquiring a negative charge so that they are now only separated according to the respective molecular weight. The gel itself comprises two layers, the first of which is the stacking gel where the proteins are applied. This smaller layer is used to focus the proteins and the subsequent separating gel through which the proteins are then separated according to their respective sizes. A list of the buffers used for SDS-PAGE can be found in Table 6. Prior to being applied to the gel, protein sample were mixed with an appropriate amount of 2x sample buffer and heated to 96°C for 5 min. The Roti-Mark STANDARD (Carl Roth) was used for assigning the protein bands to corresponding molecular weights.

Table 6: List of buffers and solutions used for SDS-PAGE.

Buffer/Solution	Composition
1x Electrode buffer	190 mM Glycine
	24.8 mM Tris-HCl
	0.1% (w/v) SDS
Stacking Buffer	0.5 M Tris-HCl pH 6.8
	0.4% (w/v) SDS
Separating Buffer	1.5 M Tris-HCl pH 8.8
	0.4% (w/v) SDS
5x Sample Buffer	50 mM Tris-HCl pH 6.8
	10% (w/v) SDS
	50% (v/v) Glycerine
	125 mM DTT
Coomassie Staining Solution	0.1% (w/v) Bromophenol blue
	2 tablets PhastGel Blue R in 95 % EtOH (40% (v/v))
	25% (v/v) 2-Propanol
	10% (v/v) Acetic Acid
De-staining Solution	10% (v/v) Acetic Acid
	25% (v/v) 2-Propanol

Polymerase chain reaction (PCR)

PCR was used to amplify DNA fragments of interest. A typical amplification PCR reaction consisted of 1 µl Phu DNA polymerase (Thermo Fisher, USA), 10 µl of 5x Phu HF buffer, 100 pmol of forward and reverse primers with appropriate restriction sites, 0.05 - 0.1 µg of template DNA and 0.2 mmol of each nucleotide from the dNTP mix (NEB, United Kingdom) and made up to a final volume of 50 µl. The

reactions were carried out in the peqlab peqSTAR 2x Gradient Thermocycler. The amplification program consisted of an initial denaturation step at 95 °C for 120 s, followed by 30 cycles of denaturation (95 °C, 30 s), annealing (58 °C, 30 s) and elongation (72 °C, 180 s) with a final extension step (72°C, 600 s). Amplified products were separated using agarose gel electrophoresis. Polymerase chain reaction was also used for whole plasmid site directed mutagenesis as described above. After reaction was complete, products were digested with 0.5 µl of the restriction enzyme DpnI at 37 °C for 1 hour (Thermo Fisher, USA). 5 µl of the digested product were used for transformation into electrocompetent *E. coli* cells.

Bacterial strains & Plasmids

Bacterial strains and plasmids used are shown in Table 7 and 8, respectively. For plasmid sequences see Appendix – plasmid sequences.

Table 7: Bacterial strains.

Strain	Genotype and description	Remarks
<i>Escherichia coli</i> XL1 blue	Blue/White screening, EndA-endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F' [::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK- mK+)	Used for plasmid preparations
<i>Escherichia coli</i> BL21(DE3) Gold	EndA-, F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Used for gene expression

Table 8: Table of used plasmids.

Plasmid	Description	Source
pET-His6-GFP-TEV-LIC	Superfolder GFP; N-terminal hexahistidine-tag; resistance: Kan	Scott Gradia Lab, via Addgene #29663
pET21a-OmpX Δ S-6His	OmpX without signal peptide; C-terminal hexahistidine-tag; resistance: Amp	Franz Hagn Lab, TU München, Germany
pET28a-His	Protein overexpression plasmid; C-terminal hexahistidine-tag; resistance: Kan; promoter: T7	Via Invitrogen
pEVOL-pBpF	<i>M.j.</i> p-benzoylphenylalanine RS (2 copies +tRNA); resistance: Chlor; promoter: araBAD, p15A vector	Peter Schultz Lab, via Addgene #31190
pEVOL-ONBY	<i>M.j.</i> o-nitrobenzyl tyrosine RS (2 copies +tRNA); resistance: Chlor; promoter: araBAD, p15A vector	Peter Schultz Lab, Scripps Research Institute, USA
pHS1403-T4-lysozyme-WT	T4 lysozyme WT; N-terminal hexahistidine-tag; resistance: Amp	Brian Matthews Lab, via Addgene #18110
pET-His6-GFP66ONBY	Superfolder GFP with amber codon at position Y66; N-terminal hexahistidine-tag; resistance: Kan	This study
pET-His6-GFP66ONBY-F145A	Superfolder GFP with “amber” codon at position Y66 and F145A pointmutation; N-terminal hexahistidine-tag; resistance: Ka;	This study
pET21a-OmpXSP	OmpX with signal peptide; C-terminal hexahistidine-tag; resistance: Amp	This study
pET21a-OmpXSP-Y57BPA	OmpX with signal peptide and “amber” codon at position Y57; C-terminal hexahistidine-tag; resistance: Amp	This study

pET21a-OmpXSP-Y95BPA	OmpX with signal peptide and “amber” codon at position Y95; C-terminal hexahistidine-tag; resistance: Amp	This study
pET21a-OmpXSP-Y57BPA-EL2	OmpX with signal peptide and “amber” codon at position Y57 and truncated EL2; C-terminal hexahistidine-tag; resistance: Amp	This study
pET21a-OmpXSP-Y95BPA-EL3	OmpX with signal peptide and “amber” codon at position Y95 and truncated EL3; C-terminal hexahistidine-tag; resistance: Amp	This study
pET21a-OmpXSP-EL2-EL3	OmpX with signal peptide and truncated EL2 and EL3; C-terminal hexahistidine-tag; resistance: Amp	This study
pET28a-T4-lysozyme-WT	T4 lysozyme WT; N-terminal hexahistidine-tag; resistance: Kan	This study
pET28a-T4-lysozyme-C54A	T4 lysozyme with C54A point mutation; N-terminal hexahistidine-tag; resistance: Kan	This study
pET28a-T4-lysozyme-C97A	T4 lysozyme with C97A point mutation; N-terminal hexahistidine-tag; resistance: Kan	This study

Oligonucleotide primers

All oligonucleotide primers used in this study were synthesized by Sigma-Aldrich

Table 9: Table of used primers.

Primer	Additional feature	Sequence
OmpX_SP_fwd	Signal peptide sequence insert	GCCGCAGTTCTGGCTTTCACCGCAGGTA ACTTCCGT AGCTGCGACTTCTACTGTA ACTG
OmpX_SP_rev	Signal peptide sequence insert	CAGAACTGCGGCCAGTGCTGAAAGACATGCAATT TTTTTCATATGTATATCTCCTTC
EGFP_Y66TAG_fwd	Stop codon insert	GTGACCACCCTGACCTAGGGCGTGCA GTGC
EGFP_Y66TAG_rev	Stop codon insert	GCACTGCACGCCCTAGGTCAGGGTGGT CAC
OmpX_Y57TAG_fwd	Stop codon insert	AGCTCTGGTGACTAGAACAAAAAC CAG
OmpX_Y57TAG_rev	Stop codon insert	CTGGTTTTTGTCTAGTACCAGAGCT
OmpX_Y95TAG_fwd	Stop codon insert	CAGACCACTGAATAGCCGACCTAC
OmpX_Y95TAG_rev	Stop codon insert	GTAGGTCGGCTATTCAGTGGTCTG
T4Lys_Cys54Ala_fwd	C54A point mutation	TTAGATAAAGCTATTGGGCGTAATGCG
T4Lys_Cys54Ala_rev	C54A point mutation	CGCATTACGCCCAATAGCTTTATCTAA
T4Lys_Cys97Ala_fwd	C97A point mutation	ATGCGGTTCGTCGcgGCATTGATT
T4Lys_Cys97Ala_rev	C97A point mutation	AATCAATGCcgGCGACGAACCGCAT
T4Lys_C-ter_His_fwd	C-terminal hexahistidine-tag insertion	ACTTGGGACGCGTATAAAAATCTACATCAT CATCA TCATCATTA
T4Lys_C-ter_His_rev	C-terminal hexahistidine-tag insertion	TTAATGATGATGATGATGATGTAGATTTT TATACG CGTCCCAAGT
OmpX_EL2_GSSG_fwd	EL2 truncation and "GSSG" motif insertion	ATCGGTTCTTTCACCTACACCGAGAAAG GATCTTC TGGAACCACTACTACGGCATCACTGC
OmpX_EL2_GSSG_rev	EL2 truncation and "GSSG" motif insertion	GCAGTGATGCCGTAGTACTGGTTTCCAGA AAGATCC TTTCTCGGTGTAAGTGAAAGAACCGAT
OmpX_EL3_GSSG_fwd	EL3 truncation and "GSSG" motif insertion	AGTGGGTGTGGGTTATGGTAAATTCGGAT CTTCTG GAACCAGCGACTACGGTTTCTCCTACG
OmpX_EL3_GSSG_rev	EL3 truncation and "GSSG" motif insertion	CGTAGGAGAAACCGTAGTCGCTGGTCCAG AAGA TCCGAATTTACCATAACCCACACCCACT

Restriction enzyme digest

All restriction enzymes were purchased from New England Biolabs (NEB), United Kingdom or Thermo Fisher, USA. Typically, 10 - 20 µg of DNA was mixed with 1.5 µl of 10x restriction enzyme buffer (Thermo Fisher, USA) and appropriate restriction enzymes, made up to 15 µl and incubated at 37 °C for 3 - 4 hours. Digested DNA was separated by agarose gel electrophoresis

Agarose gel electrophoresis

Agarose gels were made using 0.8% (w/v) agarose and TAE buffer (0.17 M TrisHCl, 1 mM EDTA, pH 8, 20 mM acetic acid) followed heating in a microwave until the agarose was dissolved. The mixture was allowed to cool to approximately 40 - 50 °C before Roti-GelStain (Carl Roth, Germany) was added (1:25,000) and poured into gel tray (Thermo Fisher, USA). Once the gel was set, TAE buffer was added to cover the gel entirely. All DNA samples were made up with 5x DNA loading dye. 1 kb GeneRuler (Thermo Fisher, USA) was loaded as a molecular size reference. Gels were run at 80 V for 40 - 60 min. Gels were documented using a Bio-Rad GelDoc 200 system (Bio-rad, United Kingdom).

Agarose gel DNA extraction

DNA bands of interest were excised from agarose gels under a transilluminator and extracted using a Qiagen gel extraction kit (Qiagen, United Kingdom) following the manufacturer's instructions. DNA was eluted in 25 - 50 µl of elution buffer (10 mM TrisHCl, pH 8.0) and stored at -20 °C.

DNA fragment ligation

0.1 - 1 µg of digested plasmid DNA was mixed with a 5-fold excess of digested ligation fragment, 1.5 µl of 10x T4 DNA ligase buffer (Thermo Fisher, USA) and 1 µl of T4 DNA ligase (Thermo Fisher, USA). The ligation reaction volume had a final volume of 15 µl and incubated overnight at RT. 5 µl of the reaction was used for the transformation into electrocompetent *E. coli* cells.

Transformation of electrocompetent *E. coli* cells

Competent *E. coli* cells were thawed at room temperature and 0.5 µl of pure plasmid (50 - 100 ng/ml) was added to 50 µl of cells. In the case of a ligation reaction, 5 µl of the ligation mixture was added to the cells. Cells were electro-shocked using a Bio-Rad MicroPulser (Bio-Rad, USA). 0.3 ml of LB media was added to the cells. Cells were then placed at 37 °C for 45 - 60 min for recovery and plated out on agar plates with appropriate selection antibiotics; transformed cells were grown overnight at 37 °C.

INSTRUMENTATION AND CHEMICALS

INSTRUMENTATION

Table 10: Instrumentation (listed alphabetically).

Instrument	Manufacturer
Äktapure	GE Healthcare, USA
Analogue tube rollers: SRT6	Stuart Equipment, United Kingdom
Analytical balance: ABJ	Kern, Germany
Autoclave: VX-120	Systec, Germany
Centrifuge: 5415 R	Eppendorf, Germany
Centrifuge: 5810 R	Eppendorf, Germany
Centrifuge: Avanti J-26SXP	Beckman & Coulter, USA
Centrifuge rotor: 70 Ti UZ	Beckman & Coulter, USA
Centrifuge rotor: JA 25.50	Beckman & Coulter, USA
Centrifuge rotor: JLA-8.1	Beckman & Coulter, USA
Confocal fluorescence Microscope: Leica DM IRBE inverted microscope & 40 x oil immersion objective	Leica, Germany
Electrophoresis system: Mini Protetra Cell	Bio-Rad, USA
Electroporation system: MicroPulser	Bio-Rad, USA
Freeze drier: Modulyo	Edwards, United Kingdom
Gel documentation: Gel Doc 200 & ChemiDoc MP	Bio-Rad, USA
Gel tray for SDS-PAGE	Precision mechanics, University Hamburg,

Germany	
Heat block: MKR13	HLC, Germany
High-pressure homogeniser: Emulsiflex-C3	Avestin, Canada
Incubator: Ecotron	Infors HT, Germany
Incubator: Multitron Standard	Infors HT, Germany
Infrared spectrometer: Vertex 70 FT-IR	Bruker, USA
Laminar airflow cabinet: BDK	Weiss Technik, Germany
Laser: Q1TH Q-switched diode-pumped Nd:YAG laser	Standa, Lithuania
Magnetic stirrer: Hei-Mix L	Heidolph Instruments, Germany
Mass spectrometer: QToF2 modified for high mass experiments	Waters, USA MS Vision, Germany
micropipette puller: P-1000 equipped with a squared box filament (2.5 x 2.5 mm)	Sutter Instruments, USA
Microscope: SZX12 with camera DP10	Olympus, Japan
Microwave	Bosch (via Carl Roth), Germany
PCR instrument: Thermocycler peqSTAR2x Gradient	Peqlab (via VWR International), USA
pH-Meter: peqMeter 1,14	Peqlab (via VWR International), USA
Pipetting aid: accu-jet pro	Brand, Germany
Platform shaker: Polymax 1040	Heidolph Instruments, Germany
Power supply unit: peqPower	Peqlab (via VWR International), USA
Precision balance: EG	Kern & Sohn, Germany
Scanner: LiDE 110	Canon, Japan
Spectrophotometer: Nanodrop 2000	Thermo Fisher Scientific, USA
Spectrophotometer: SmartSpec Plus	Bio-Rad, USA
Spectrophotometer: QE Pro spectrometer and a DH-2000 light source	Ocean Optics, United Kingdom

sputter coater: Q150R	Quorum Technologies, United Kingdom
Ultracentrifuge: Optima XE-90	Beckman & Coulter, USA
Ultrasonic cell disruptor: Sonopuls	Bandelin, Germany
UV hand lamp: H466.1	Carl Roth, Germany
Vacuum pump: BVC control	Vacuubrand, Germany
Vortexer: Vortex-Genie 2	Scientific Industries, USA

CHEMICALS USED (GHS CLASSIFICATION)

Table 11: Chemicals (listed alphabetically).

Compound	CAS-No.	Supplier	GHS hazard	Hazard Statements	Precautionary Statements
2-Nitrobenzylbromid	3958-60-9	Th. Geyer	GHS05, GHS07	H314, H335	P280, P304+P340 P305+P351+P338 P310
4-Benzoyl-L-phenylalanine	104504-45-2	Bachem	-	-	-
5,5'-Dithiobis(2-nitrobenzoic acid)	69-78-3	Sigma-Aldrich	GHS07	H315, H319, H335	P261, P305+P351+P338
Acetic acid, 96 %	64-19-7	Carl Roth	GHS02, GHS05	H226, H314	P280, P305+351+338, P310
Acetone	67-64-1	Carl Roth	GHS02, GHS07	H225, H319, H336	P210, P280, P304+P340+P312, P305+P351+P338, P337+P313, P403+P235
Active carbon	7440-44-0	Carl Roth	-	-	-
Acrylamide 37 %	79-06-1	Carl Roth	GHS06, GHS08	H301, H312, H315, H317, H319, H332, H340, H350, H361f, H372	P201, P280, P301+310, P305+351+338, P308+313
Agar-Agar	9002-18-0	Sigma-Aldrich	-	-	-

Agarose	9012-36-6	Sigma-Aldrich	-	-	-
Albumin Fraktion V BSA	90604-29-8	Carl Roth	-	-	-
Ammonium formate	540-69-2	Sigma Aldrich	GH202	H315, H319, H335	P261, P305+351+338
Ammonium acetate	631-61-8	Carl Roth	-	-	-
Ampicillin	69-52-3	Carl Roth	GHS08	H334, H317	P280, P261, P302+352, P342+311
APS	7727-54-0	Carl Roth	GHS03 GHS07 GHS08	H272, H302, H315, H317, H319, H334, H335	P280, P305+351+338, P302+352, P304+341, P342+311
Arabinose	5328-37-0	Carl Roth	-	-	-
beta-Mercaptoethanol	60-24-2	Sigma-Aldrich	GHS05, GHS06, GHS08, GHS09	H301 + H331-H310- H315-H317- H318-H373- H410	P260, P262, P273, P280, P301+P310+P330, P302+P352+P310, P305+P351+P338+P310, P391, P403+P233
Brilliant blue G 250	6104-58-1	Carl Roth	-	-	-
Bromphenol blue	115-39-9	Sigma-Aldrich	-	-	-
CaCl₂	10043-52-4	Sigma-Aldrich	GHS07	H319	P305+P351+P338
Chloramphenicol	56-75-7	Sigma-Aldrich	GHS08	H351	P280
Citric acid	77-92-9	Carl Roth	GHS05	H318	P305+P351+P338, P311
Coomassie Brilliant Blue R250	6104-59-2	Sigma-Aldrich	-	-	-
Copper (II) sulphate	7758-98-7	Sigma-Aldrich	GHS05, GHS09	H302-H315- H319-H410	P264, P273, P280, P337+P313, P391, P501
D₂O	7789-20-0	Sigma-Aldrich	-	-	-
Di-Potassium hydrogen phosphate	16788-57-1	Carl Roth	-	-	-
Dimethylformamid	68-12-2	Carl Roth	GHS02, GHS07, GHS08	H226- H312+H332- H319-H360D	P201-P210-P261- P280-P308+P313- P370+P378

				P	
Dnase I	-	Sigma- Aldrich	-	-	-
DTT	578517	Carl Roth	GHS07	H302, H315, H319, H335	P302+352, P305+351+338
EDTA	60-00-4	Carl Roth	GHS07	H319	P305+351+338
Ethanol	64-17-5	Carl Roth	GHS02	H225	P210
Fibronectin	-	Sigma- Aldrich	-	-	-
Formaldehyde	50-00-0	Carl Roth	GHS02, GHS05, GHS06, GHS08	H226- H301+H311+ H331-H314- H317-H335- H341-H350- H370	P201-P210-P260- P280- P301+P310+P330- P303+P361+P353- P304+P340+P310- P305+P351+P338+P3 10-P308+P311- P370+P378- P403+P233
Glycerol	56-81-5	Carl Roth	-	-	-
Hydrochloric acid, 6 N	7647-01-0	Carl Roth	GHS05 GHS07	H290 H314, H335	P260-P280- P303+P361 +P353- P304+P340+P310- P305+P351+P338
HEPES	7365-45-9	Carl Roth	-	-	-
Imidazole	288-32-4	Carl Roth	GHS05 GHS06 GHS08	H301, H314, H361	P260, P281, P303+361+353, P301+330+331, P305+351+338, P308+313
IPTG	367-93-1	Carl Roth	-	-	-
Isopropanol	67-63-0	Carl Roth	GHS02 GHS07	H225, H319, H336	P210, P233, P305+351+338
Kanamycin sulfate	25389-94- 0	Carl Roth	GHS08	H360	P201-P308+P313
LB-medium Lennox	-	Carl Roth	-	-	-
L-Tyrosine	60-18-4	Carl Roth	-	-	-
Lysozyme (<i>Gallus gallus</i>)	12650-88- 3	Sigma- Aldrich	-	-	-
2- Mercaptoethanol	60-24-2	Sigma- Aldrich	GHS06 GHS09	H302, H411, H315, H335, H311, H319	P280, P312, P302+350, P261, P273, P301+312, P305+351+338

MES	4432-3-9	Carl Roth	-	-	-
Methanol	67-56-1	Sigma-Aldrich	GHS02, GHS06, GHS08	H225- H301+H311+ H331-H370	P210-P280- P302+P352+P312- P304+P340+P312- P370+P378- P403+P235
MgCl₂	7786-30-3	Carl Roth	-	-	-
Milk powder	68514-61-4	Carl Roth	-	-	-
NaOAc	127-09-3	Sigma-Aldrich	GHS07	H319	P305+P351+P338
NaCl	7647-14-5	Carl Roth	-	-	-
NaOH	1310-73-2	Carl Roth	GHS05	H314	P280, P310, P305+351+338
Ni-NTA-Agarose		Sigma-Aldrich	GHS02 GHS07 GHS08	H226-H317- H350i- H360D- H373-H412	P201-P273-P280- P308+P313- P333+P313- P370+P378
Ni(II)SO₄	10101-97-0	Carl Roth	GHS07 GHS08 GHS09	H302+332, H315, H317, H334, H341, H350i, H360d, H372 H410	P201, P261, P273, P280, P284, P304+340+312
Potassium Carbonate	584-08-7	Sigma-Aldrich	GHS07	H315, H319, H335	P302+P352, P305+P351+P338
Penicillin-Streptomycin	-	Gibco	GHS07 GHS07	H315, H317, H334, H335	P280, P261, P264, P284, P271, P302+P352, P333+P313, P304+P340, P342+P311, P312, P403+P233, P501
Potassium cyanide	151-50-8	Sigma-Aldrich	GHS05 GHS06 GHS08 GHS09	H290- H300+H310+ H330-H370- H371-H372- H410	P260, P262, P264, P273, P280, P284, P301+P310,P301+P310 +P330, P302+P352+P310, P304+P340+P310, P391, P403+P233
Potassium di-hydrogen phosphate	7778-77-0	Carl Roth	-	-	-

Roti-GelStain	-	Carl Roth	-	-	-
SDS	151-21-3	Carl Roth	GHS02 GHS06	H228, H302, H311, H315, H319, H335	P210, P261, P280, P312, P305+351+338
Sodium citrate	1545832	Sigma- Aldrich	-	-	-
TEMED	1185-53-1	Carl Roth	GHS07	H315, H319, H335	P261, P305+351+338
Tris hydrochloride	1185-53-1	Carl Roth	-	-	-
Triton X 100	9002-93-1	Carl Roth	GHS05 GHS07 GHS09	H302-H315- H318-H410	P280, P301+P312+P330, P305+P351+P338+P3 10
Tryptone/Peptone	8952.4	Carl Roth	-	-	-
Trypsin	9002-07-7	Carl Roth	GHS07 GHS08	H315-H319- H334-H335	P261, P280, P284, P304+P340, P337+P313, P342+P311
Yeast Extract	8013-01-2	Carl Roth	-	-	-

PROTEIN CRYSTALLIZATION SCREENS

Table 12: Crystallisation screens with HP-statements.

Compound	Supplier	GHS hazard	Hazard Statements	Precautionary Statements
JCSG Core Suite I	Qiagen	GHS02, GHS06, GHS05	H225, H315, H319, H350, H360, H335, H412	P201, P210, P280, P308+P313
JCSG Core Suite II	Qiagen	GHS02, GHS07, GHS08, GHS09	H225, H302, H315, H317, H319, H350, H360, H335, H373, H411	P201, P210, P280, P308+P313
JCSG Core Suite III	Qiagen	GHS02, GHS05, GHS06, GHS08, GHS09	H225, H302, H314, H317, H318, H340, H350, H360, H336, H335, H373, H400, H410	P201, P210, P260, P264, P280, P305+P351+P338+P310 P308+P313 P403+P233
JCSG Core Suite IV	Qiagen	GHS02, GHS07, GHS08, GHS09	H225, H302, H315, H319, H335, H336, H350, H360, H373, H410	P201, P210, P260, P280, P308+P313
JCSG+ Suite	Qiagen	GHS02 GHS05 GHS06 GHS07 GHS08	H225, H301, H312, H315, H318, H331, H335, H350, H411	P101, P201, P270, P280, P305+351+338, P309+311, P313
Pi-minimal_HTS screen	Jena Bioscience	GHS07	H302, H312, H332, H315, H319, H317, H335	P260, P280, P301+P310+P330 P303+P361+P353 P304+P340+P310 P305+P351+P338 P403+P233
Additive Screen HT	Hampton Research	GHS02 GHS05 GHS06 GHS07 GHS08	H225, H290, H301, H302, H311, H312, H314, H315, H317, H318, H319, H330, H331, H332, H334, H335, H336, H340, H341, H350,	P201, P210, P233, P260, P261, P264, P270, P273, P281, P280, P284, P309, P310, P311, P312, P321, P330, P362, P391, P405, P501, P301+P310, P301+P312, P301+P330+P331,

H351, H360,
H370, H372,
H373, H400,
H410, H411,
H412

P302+P352,
P303+P361+P353,
P304+P341,
P305+P351+P338,
P308+P313,
P309+P311,
P332+P313,
P333+P313,
P337+P313,
P342+P311,
P370+P378,
P370+P378,
P370+P378,
P403+P233,
P403+P235,

GHS AND RISK SYMBOLS



Figure Appendix 1: GHS pictograms (from <https://www.ohsa.gov>)

GHS HAZARD STATEMENTS

H225	Highly flammable liquid and vapour
H226	Flammable liquid and vapour
H228	Flammable solid
H272	May intensify fire; oxidizer
H290	May be corrosive to metals
H301	Toxic if swallowed
H302	Harmful if swallowed
H303	May be harmful if swallowed
H311	Toxic in contact with skin
H312	Harmful in contact with skin
H313	May be harmful in contact with skin
H314	Causes severe skin burns and eye damage
H315	Causes skin irritation
H316	Causes mild skin irritation
H317	May cause an allergic skin reaction
H318	Causes serious eye damage
H319	Causes serious eye irritation
H330	Fatal if inhaled
H331	Toxic if inhaled
H332	Harmful if inhaled
H333	May be harmful if inhaled
H334	May cause allergy or asthma symptoms or breathing difficulties if inhaled
H335	May cause respiratory irritation
H336	May cause drowsiness or dizziness
H340	May cause genetic defects

H341	Suspected of causing genetic defects
H350	May cause cancer
H350i	May cause cancer by inhalation
H360	May damage fertility or the unborn child
H360D	May damage the unborn child
H360F	May damage fertility
H360Fd	May damage fertility. Suspected of damaging the unborn child
H360FD	May damage fertility. May damage the unborn child
H361	Suspected of damaging fertility or the unborn child
H361d	Suspected of damaging the unborn child.
H361f	Suspected of damaging fertility
H370	Cause damage to organs
H372	Causes damage to organs through prolonged or repeated exposure
H373	May cause damage to organs through prolonged or repeated exposure.
H400	Very toxic to aquatic life with long-lasting effects
H410	Very toxic to aquatic life with long lasting effects
H411	Toxic to aquatic life with long lasting effects
H412	Harmful to aquatic life with long lasting effects.
EUH032	Contact with acids liberates very toxic gas
EUH066	Repeated exposure may cause skin dryness or cracking

GHS PRECAUTIONARY STATEMENTS

P101	If medical advice is needed, have product container or label at hand
P201	Obtain special instructions before use
P210	Keep away from heat/sparks/open flames/hot surfaces – No smoking
P233	Keep container tightly closed
P260	Do not breathe dust/fume/gas/mist/vapours/spray
P261	Avoid breathing dust/fume/gas/mist/vapours/spray
P264	Wash thoroughly after handling
P270	Do not eat, drink or smoke when using this product
P273	Avoid release to the environment
P281	Use personal protective equipment as required
P280	Wear protective gloves/protective clothing/eye protection/face protection
P284	Wear respiratory protection
P309	IF exposed or you feel unwell
P310	Immediately call a POISON CENTER or doctor/physician
P311	Call a POISON CENTER or doctor/physician
P312	Call a POISON CENTER or doctor/physician if you feel unwell
P321	Specific treatment (see respective MSDS)
P330	Rinse mouth
P362	Take off contaminated clothing and wash before reuse
P391	Collect spillage
P405	Store locked up
P501	Dispose of contents/container in accordance with local/regional/national/international regulations
P301+P310	IF SWALLOWED: Immediately call a POISON CENTER or doctor/physician
P301+P312	IF SWALLOWED: Call a POISON CENTER or doctor/physician if you feel unwell
P301+P330+P331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting
P302+P352	IF ON SKIN: Wash with soap and water
P303+P361+P353	IF ON SKIN (or hair): Remove/Take off immediately all contaminated clothing. Rinse skin with water/shower
P304+P341	IF INHALED: If breathing is difficult, remove victim to fresh air and keep at rest in a position comfortable for breathing
P305+P351+P338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do - continue rinsing

P308+P313	IF exposed or concerned: Get medical advice/attention
P309+P311	IF exposed or you feel unwell: Call a POISON CENTER or doctor/physician
P332+P313	If skin irritation occurs: Get medical advice/attention
P333+P313	If skin irritation or rash occurs: Get medical advice/attention
P337+P313	If eye irritation persists: Get medical advice/attention
P342+P311	Call a POISON CENTER or doctor/physician
P370+P378	In case of fire: Use for extinction: Alcohol resistant foam
P370+P378	In case of fire: Use for extinction: Fire-extinguishing powder
P370+P378	In case of fire: Use for extinction: Carbon dioxide
P403+P233	Store in a well-ventilated place. Keep container tightly closed
P403+P235	Store in a well-ventilated place. Keep cool

DISPOSAL

All chemicals used during the experimental phase were handled and disposed according to their H- and P-Sentences. Solvents and contaminated waste was stored in the specific boxes and disposed according to the safety instructions. Genetically modified organisms and related waste were autoclaved according to the "Gentechnikgesetz" before disposal for 20 min at 121°C and 5 bar.

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Dr. Henning Tidow for the exciting topic, the supervision and the appraisal of my thesis. Thank you for the freedom of the independent scientific work in your group, participation in various conferences and workshops, as well as for the continuous support of my Ph.D study and related research, which have given me excellent opportunities to continue my professional education and enriched me personally.

Besides my supervisor, I would like to thank my advisor Prof. Dr. Arwen Pearson for her very motivating and inspiring words and ideas that helped me to overcome every hurdle and access to her laboratory and research facilities and Prof. Dr. Uli Hahn for the great support during my time as a doctoral student. My sincere thanks also goes to Prof. Dr. Michael Kolbe for the preparation of the second opinion, as well as the examiners of my disputation.

Moreover, I would like to express my sincere gratitude to my lab fellows: Dr. Inokentij Josts, Dr. Katharina Jungnickel, Tina Bohstedt, Anne Creon, Ellen Gattkowsky, Yasser Almeida Hernandez, Julius Nitsche and Katharina Veith. Thank you for your support and for all the advice and data discussions. I wish you all the best!

For this research many collaborations were created. So I would like to use this opportunity to thank everybody for supporting me and my project. A very special thanks to Prof. Dr. Andreas Guse for giving me access to his cell culture and imaging laboratories and facilities and Dr. Ralf Fliegert, not only for always putting so much of your equipment at my disposal, but also for helpful discussions and ideas. Further, I would like to thank Prof. Dr Nils Huse for access to his laboratories and facilities and Dr. Stephan Niebling for assisting me in infrared spectroscopy and the fabulous neutron spectroscopy beamtimes! My gratitude also to John Joseph Zaitsev-Doyle and Dr. Marta Sans Valls for your patience with me in my attempt to do organic chemistry in your red-light lab. Thank you to Dr. Briony Yorke and Anke Puchert for helping me with spectroscopic measurements.

This work would not have been possible without the doctoral fellowship of the Centre for Ultrafast Imaging Graduate School and funding of the Centre for Ultrafast Imaging (CUI). I would also like to

thank all the staff of the CUI Cluster Office and the CUI Graduate School for their outstanding support in all organizational matters and concerns, as well as their constant efforts to create an optimal environment for doctoral training.

Ein ganz besonderer Dank gilt meinen Eltern Monika Schäfer und Carsten Schäfer, die mich immer unterstützt haben meinen Weg zu gehen während des Universitätsstudiums (und davor), sowie meinem Bruder Björn Schäfer und meinen Freunden für Ihr Interesse an meiner Arbeit, den vielen Ermutigungen und die tolle Unterstützung in allen Lebenslagen. Mein größter Dank geht an Firat Kesgin, weil du mich auch in stressigen Phasen immer unterstützt und Kraft gegeben hast. Danke, dass Du besonders in der Endphase dieser Arbeit immer für mich da warst. *Teşekkür ederim, sen olmasaydın ben yapamazdım. Gelecekte ne olursa olsun, sadece seninle yaşamayı dört gözle bekliyorum!*

CURRICULUM VITAE

Education /Professional Career

- 05/2015 – 01/2019 Research Associate (PhD student)
Excellence Cluster: Hamburg Center for Ultrafast Imaging (CUI) &
Department of Chemistry
University of Hamburg, Germany
Project title: „New probes for (membrane) protein structure, function
and analysis“, Prof. Dr. Henning Tidow
- 10/2012 – 03/2015 Study of Microbiology & Biochemistry (Master of Science),
Georg-August-University Göttingen, Germany
Master Thesis: „GTPase activity of elongation factor G“, Prof. Dr.
Marina Rodnina, Max-Planck-Institute for Biophysical Chemistry
- 10/2009 – 03/2012 Study of Biology, focus on Molecular Biology (Bachelor of Science),
Georg-August-University Göttingen, Germany
Bachelor Thesis: „Development of a semi-quantitative ESI-MS²
method for the analysis of the wax ester content in transgenic oil
seeds“, Prof. Dr. Feußner

EIDESSTÄTTLICHE ERKLÄRUNG

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde. Ferner versichere ich, dass ich noch keine Promotionsversuche an anderen Universitäten unternommen habe.

Hamburg, 16.01.2019

REFERENCES

1. Crick, F.H.C., et al., *General Nature of the Genetic Code for Proteins*. Nature, 1961. **192**: p. 1227.
2. Nirenberg, M.W. and J.H. Matthaei, *The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides*. 1961. **47**(10): p. 1588-1602.
3. Crick, F.H.C., *On protein synthesis*. Symposia of the Society of Experimental Biology 1958. **12**: p. 138-63.
4. Lehman, I.R., et al., *Enzymatic Synthesis of Deoxyribonucleic Acid: I. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM ESCHERICHIA COLI*. 1958. **233**(1): p. 163-170.
5. Bessman, M.J., et al., *Enzymatic Synthesis of Deoxyribonucleic Acid: II. GENERAL PROPERTIES OF THE REACTION*. 1958. **233**(1): p. 171-177.
6. Holley, R.W., et al., *Structure of a Ribonucleic Acid*. 1965. **147**(3664): p. 1462-1465.
7. Brown, C., *Translation: DNA to mRNA to Protein*. Nature Education, 2008. **1**(1): p. 101.
8. Schmeing, T.M. and V. Ramakrishnan, *What recent ribosome structures have revealed about the mechanism of translation*. Nature, 2009. **461**: p. 1234.
9. Burbaum, J.J. and P. Schimmel, *Structural relationships and the classification of aminoacyl-tRNA synthetases*. 1991. **266**(26): p. 16965-16968.
10. Woese, C.R., et al., *Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process*. Microbiology and molecular biology reviews : MMBR, 2000. **64**(1): p. 202-236.
11. J Lucas-Lenard, a. and F. Lipmann, *Protein Biosynthesis*. 1971. **40**(1): p. 409-448.
12. Berk, V. and J.H.D. Cate, *Insights into protein biosynthesis from structures of bacterial ribosomes*. Current Opinion in Structural Biology, 2007. **17**(3): p. 302-309.
13. Rodnina, M.V., M. Beringer, and W. Wintermeyer, *How ribosomes make peptide bonds*. Trends in Biochemical Sciences, 2007. **32**(1): p. 20-26.
14. Bashan, A., et al., *Structural Basis of the Ribosomal Machinery for Peptide Bond Formation, Translocation, and Nascent Chain Progression*. Molecular Cell, 2003. **11**(1): p. 91-102.
15. Peske, F., M.V. Rodnina, and W. Wintermeyer, *Sequence of Steps in Ribosome Recycling as Defined by Kinetic Analysis*. Molecular Cell, 2005. **18**(4): p. 403-412.
16. Rodnina, Marina V. and W. Wintermeyer, *The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation*. Biochemical Society Transactions, 2011. **39**(2): p. 658.
17. Ogle, J.M. and V. Ramakrishnan, *STRUCTURAL INSIGHTS INTO TRANSLATIONAL FIDELITY*. 2005. **74**(1): p. 129-177.
18. Ling, J., N. Reynolds, and M. Ibba, *Aminoacyl-tRNA Synthesis and Translational Quality Control*. 2009. **63**(1): p. 61-78.
19. Söll, D., U.L. RajBhandary, and T.L. RajBhandary, *tRNA: Structure, Biosynthesis, and Fznction*. 1995.
20. Parker, J., *Errors and alternatives in reading the universal genetic code*. Microbiological reviews, 1989. **53**(3): p. 273-298.
21. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, *Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications*. 2005. **44**(45): p. 7342-7372.
22. Ryšlavá, H., et al., *Effect of posttranslational modifications on enzyme function and assembly*. Journal of Proteomics, 2013. **92**: p. 80-109.

23. Cesaro, L., L.A. Pinna, and M. Salvi, *A Comparative Analysis and Review of lysyl Residues Affected by Posttranslational Modifications*. Current genomics, 2015. **16**(2): p. 128-138.
24. Bischoff, R. and H. Schlüter, *Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications*. Journal of Proteomics, 2012. **75**(8): p. 2275-2296.
25. Namy, O., et al., *Reprogrammed Genetic Decoding in Cellular Gene Expression*. Molecular Cell, 2004. **13**(2): p. 157-168.
26. Bullwinkle, T., B. Lazazzera, and M. Ibba, *Quality Control and Infiltration of Translation by Amino Acids Outside of the Genetic Code*. 2014. **48**(1): p. 149-166.
27. Richmond, M.H., *The effect of amino acid analogues on growth and protein synthesis in microorganisms*. Bacteriological reviews, 1962. **26**(4): p. 398-420.
28. Rosenthal, G.A., *The biochemical basis for the deleterious effects of l-canavanine*. Phytochemistry, 1991. **30**(4): p. 1055-1058.
29. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA*. Science, 2002. **296**(5572): p. 1459.
30. Zinoni, F., et al., *Cotranslational insertion of selenocysteine into formate dehydrogenase from Escherichia coli directed by a UGA codon*. Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(10): p. 3156-3160.
31. Atkins, J.F. and P.V. Baranov, *Translation: Duality in the genetic code*. Vol. 448. 2007, Nature. 2.
32. Ambrogelly, A., et al., *Pyrrolysine is not hardwired for cotranslational insertion at UAG codons*. Proceedings of the National Academy of Sciences, 2007. **104**(9): p. 3141.
33. Krzycki, J.A., *The direct genetic encoding of pyrrolysine*. Current Opinion in Microbiology, 2005. **8**(6): p. 706-712.
34. Ibba, M. and D. Söll, *Genetic Code: Introducing Pyrrolysine*. Current Biology, 2002. **12**(13): p. R464-R466.
35. Hatfield, D.L. and V.N. Gladyshev, *How selenium has altered our understanding of the genetic code*. Molecular and cellular biology, 2002. **22**(11): p. 3565-3576.
36. Budisa, N., *Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire*. 2004. **43**(47): p. 6426-6463.
37. Bilgiçer, B. and K. Kumar, *Protein Design Using Unnatural Amino Acids*. Journal of Chemical Education, 2003. **80**(11): p. 1275.
38. Liu, C.C. and P.G. Schultz, *Adding New Chemistries to the Genetic Code*. 2010. **79**(1): p. 413-444.
39. Chin, J.W., *Expanding and reprogramming the genetic code*. Nature, 2017. **550**: p. 53.
40. Cohen, S.N., et al., *Construction of biologically functional bacterial plasmids in vitro*. Proceedings of the National Academy of Sciences of the United States of America, 1973. **70**(11): p. 3240-3244.
41. Cowie, D.B. and G.N. Cohen, *Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur*. Biochimica et Biophysica Acta, 1957. **26**(2): p. 252-261.
42. Hendrickson, W.A., *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation*. Science, 1991. **254**(5028): p. 51.
43. Hendrickson, W.A., *Maturation of MAD phasing for the determination of macromolecular structures*. 1999. **6**(4): p. 845-851.
44. Xie, J. and P.G. Schultz, *A chemical toolkit for proteins — an expanded genetic code*. Nature Reviews Molecular Cell Biology, 2006. **7**: p. 775.
45. Connor, R.E. and D.A. Tirrell, *Non-Canonical Amino Acids in Protein Polymer Design*. Polymer Reviews, 2007. **47**(1): p. 9-28.
46. Wang, L., *Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts*. Accounts of chemical research, 2017. **50**(11): p. 2767-2775.
47. Xie, J. and P.G. Schultz, *An expanding genetic code*. Methods, 2005. **36**(3): p. 227-238.

48. Schultz, K.C., et al., *A Genetically Encoded Infrared Probe*. Journal of the American Chemical Society, 2006. **128**(43): p. 13984-13985.
49. Ye, J.C., et al., *NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy*. NeuroImage, 2009. **44**(2): p. 428-447.
50. Bose, M., et al., *The Incorporation of a Photoisomerizable Amino Acid into Proteins in E. coli*. Journal of the American Chemical Society, 2006. **128**(2): p. 388-389.
51. Alfonta, L., et al., *Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins*. Journal of the American Chemical Society, 2003. **125**(48): p. 14662-14663.
52. Tang, Y. and D.A. Tirrell, *Biosynthesis of a Highly Stable Coiled-Coil Protein Containing Hexafluoroisoleucine in an Engineered Bacterial Host*. Journal of the American Chemical Society, 2001. **123**(44): p. 11089-11090.
53. Sakamoto, K., et al., *Genetic Encoding of 3-Iodo-Tyrosine in *Escherichia coli* for Single-Wavelength Anomalous Dispersion Phasing in Protein Crystallography*. Structure, 2009. **17**(3): p. 335-344.
54. Meng, H. and K. Kumar, *Antimicrobial Activity and Protease Stability of Peptides Containing Fluorinated Amino Acids*. Journal of the American Chemical Society, 2007. **129**(50): p. 15615-15622.
55. Johnson, J.A., et al., *Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications*. Current opinion in chemical biology, 2010. **14**(6): p. 774-780.
56. Young, T.S., et al., *An enhanced system for unnatural amino acid mutagenesis in E. coli*. J Mol Biol, 2010. **395**(2): p. 361-74.
57. Voloshchuk, N. and J.K. Montclare, *Incorporation of unnatural amino acids for synthetic biology*. Molecular BioSystems, 2010. **6**(1): p. 65-80.
58. de Graaf, A.J., et al., *Nonnatural Amino Acids for Site-Specific Protein Conjugation*. Bioconjugate Chemistry, 2009. **20**(7): p. 1281-1295.
59. Beatty, K.E. and D.A. Tirrell, *Two-color labeling of temporally defined protein populations in mammalian cells*. Bioorganic & medicinal chemistry letters, 2008. **18**(22): p. 5995-5999.
60. Budisa, N., et al., *Atomic mutations in annexin V*. 1998. **253**(1): p. 1-9.
61. Pine, M.J., *Comparative physiological effects of incorporated amino acid analogs in Escherichia coli*. Antimicrobial agents and chemotherapy, 1978. **13**(4): p. 676-685.
62. Budisa, N., et al., *Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids*. Protein science : a publication of the Protein Society, 2001. **10**(7): p. 1281-1292.
63. Xiu, X., et al., *Nicotine binding to brain receptors requires a strong cation-pi interaction*. Nature, 2009. **458**(7237): p. 534-537.
64. Banerjee, S., et al., *Structure of a heterogeneous, glycosylated, lipid-bound, in vivo-grown protein crystal at atomic resolution from the viviparous cockroach *Diploptera punctata**. IUCrJ, 2016. **3**(4): p. 282-293.
65. Dumas, A., et al., *Designing logical codon reassignment – Expanding the chemistry in biology*. Chemical Science, 2015. **6**(1): p. 50-69.
66. Wang, K., et al., *Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET*. Nature chemistry, 2014. **6**(5): p. 393-403.
67. Fekner, T. and M.K. Chan, *The pyrrolysine translational machinery as a genetic-code expansion tool*. Current opinion in chemical biology, 2011. **15**(3): p. 387-391.
68. Wang, L., et al., *Expanding the Genetic Code of *Escherichia coli**. Science, 2001. **292**(5516): p. 498.
69. Wang, L., J. Xie, and P.G. Schultz, *EXPANDING THE GENETIC CODE*. 2006. **35**(1): p. 225-249.
70. Melançon, C.E., 3rd and P.G. Schultz, *One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases*. Bioorganic & medicinal chemistry letters, 2009. **19**(14): p. 3845-3847.

71. Krishnakumar, R. and J. Ling, *Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion*. 2014. **588**(3): p. 383-388.
72. O'Donoghue, P., et al., *Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion*. FEBS letters, 2012. **586**(21): p. 3931-3937.
73. *A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli*. Angewandte Chemie International Edition, 2010. **49**(18): p. 3211-3214.
74. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome*. Nature, 2010. **464**: p. 441.
75. Longstaff, D.G., et al., *A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(3): p. 1021-1026.
76. Martin, G.W., 3rd, J.W. Harney, and M.J. Berry, *Selenocysteine incorporation in eukaryotes: insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element*. RNA (New York, N.Y.), 1996. **2**(2): p. 171-182.
77. Pott, M., M.J. Schmidt, and D. Summerer, *Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids*. ACS Chemical Biology, 2014. **9**(12): p. 2815-2822.
78. Link, A.J. and D.A. Tirrell, *Reassignment of sense codons in vivo*. Methods, 2005. **36**(3): p. 291-298.
79. Kiick, K.L. and D.A. Tirrell, *Protein Engineering by In Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-tRNA Synthetase*. Tetrahedron, 2000. **56**(48): p. 9487-9493.
80. Kiick, K.L., J.C.M. van Hest, and D.A. Tirrell, *Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host*. 2000. **39**(12): p. 2148-2152.
81. Kast, P. and H. Hennecke, *Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations*. Journal of Molecular Biology, 1991. **222**(1): p. 99-124.
82. Ibba, M., P. Kast, and H. Hennecke, *Substrate Specificity Is Determined by Amino Acid Binding Pocket Size in Escherichia coli Phenylalanyl-tRNA Synthetase*. Biochemistry, 1994. **33**(23): p. 7107-7112.
83. Datta, D., et al., *Selectivity and specificity of substrate binding in methionyl-tRNA synthetase*. Protein science : a publication of the Protein Society, 2004. **13**(10): p. 2693-2705.
84. Beatty, K.E., et al., *Selective Dye-Labeling of Newly Synthesized Proteins in Bacterial Cells*. Journal of the American Chemical Society, 2005. **127**(41): p. 14150-14151.
85. Tang, Y., et al., *Introduction of an aliphatic ketone into recombinant proteins in a bacterial strain that overexpresses an editing-impaired leucyl-tRNA synthetase*. Chembiochem : a European journal of chemical biology, 2009. **10**(13): p. 2188-2190.
86. van Hest, J.C.M. and D.A. Tirrell, *Efficient introduction of alkene functionality into proteins in vivo*. 1998. **428**(1-2): p. 68-70.
87. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(1): p. 19-24.
88. Wang, L., et al., *Evolution of new nonantibody proteins via iterative somatic hypermutation*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(48): p. 16745-16749.
89. Wang, P., Y. Tang, and D.A. Tirrell, *Incorporation of Trifluoroisoleucine into Proteins in Vivo*. Journal of the American Chemical Society, 2003. **125**(23): p. 6900-6906.
90. Minks, C., et al., *Towards New Protein Engineering: In Vivo Building and Folding of Protein Shuttles for Drug Delivery and Targeting by the Selective Pressure Incorporation (SPI) Method*. Tetrahedron, 2000. **56**(48): p. 9431-9442.

91. Dieterich, D.C., et al., *Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)*. Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(25): p. 9482-9487.
92. Aerni, H.R., et al., *Revealing the amino acid composition of proteins within an expanded genetic code*. Nucleic acids research, 2015. **43**(2): p. e8-e8.
93. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins*. Science, 1989. **244**(4901): p. 182.
94. Chin, J.W., et al., *Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(17): p. 11020-11024.
95. Jackson, J.C., J.T. Hammill, and R.A. Mehl, *Site-Specific Incorporation of a 19F-Amino Acid into Proteins as an NMR Probe for Characterizing Protein Structure and Reactivity*. Journal of the American Chemical Society, 2007. **129**(5): p. 1160-1166.
96. Zhang, Z., et al., *A New Strategy for the Synthesis of Glycoproteins*. Science, 2004. **303**(5656): p. 371.
97. Mehl, R.A., et al., *Generation of a Bacterium with a 21 Amino Acid Genetic Code*. Journal of the American Chemical Society, 2003. **125**(4): p. 935-939.
98. Furter, R., *Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli*. Protein science : a publication of the Protein Society, 1998. **7**(2): p. 419-426.
99. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the Degeneracy of the Genetic Code*. Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.
100. Kwon, I. and D.A. Tirrell, *Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells*. Journal of the American Chemical Society, 2007. **129**(34): p. 10431-10437.
101. Kwon, I., P. Wang, and D.A. Tirrell, *Design of a Bacterial Host for Site-Specific Incorporation of p-Bromophenylalanine into Recombinant Proteins*. Journal of the American Chemical Society, 2006. **128**(36): p. 11778-11783.
102. Chen, P.R., et al., *A facile system for encoding unnatural amino acids in mammalian cells*. Angewandte Chemie (International ed. in English), 2009. **48**(22): p. 4052-4055.
103. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding Ne-acetylysine in recombinant proteins*. Nature Chemical Biology, 2008. **4**: p. 232.
104. Chin, J.W., et al., *An Expanded Eukaryotic Genetic Code*. Science, 2003. **301**(5635): p. 964.
105. Chen, M., et al., *Site-specific incorporation of unnatural amino acids into urate oxidase in Escherichia coli*. Protein science : a publication of the Protein Society, 2008. **17**(10): p. 1827-1833.
106. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells*. Nucleic acids research, 2002. **30**(21): p. 4692-4699.
107. Jackson, S.E., T.D. Craggs, and J.R. Huang, *Understanding the folding of GFP using biophysical techniques*. Expert Rev Proteomics, 2006. **3**(5): p. 545-59.
108. Santoro, S.W., et al., *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity*. Nature Biotechnology, 2002. **20**: p. 1044.
109. Brustad, E.M., et al., *A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer*. Journal of the American Chemical Society, 2008. **130**(52): p. 17664-17665.
110. Young, T.S., et al., *Expanding the Genetic Repertoire of the Methylotrophic Yeast Pichia pastoris*. Biochemistry, 2009. **48**(12): p. 2643-2653.
111. Wang, W., et al., *Genetically encoding unnatural amino acids for cellular and neuronal studies*. Nature neuroscience, 2007. **10**(8): p. 1063-1072.
112. Ye, S., et al., *Site-specific Incorporation of Keto Amino Acids into Functional G Protein-coupled Receptors Using Unnatural Amino Acid Mutagenesis*. 2008. **283**(3): p. 1525-1533.
113. Mukai, T., et al., *Adding l-lysine derivatives to the genetic code of mammalian cells with*

- engineered pyrrolysyl-tRNA synthetases*. Biochemical and Biophysical Research Communications, 2008. **371**(4): p. 818-822.
114. Greiss, S. and J.W. Chin, *Expanding the genetic code of an animal*. Journal of the American Chemical Society, 2011. **133**(36): p. 14196-14199.
115. Chen, S., P.G. Schultz, and A. Brock, *An Improved System for the Generation and Analysis of Mutant Proteins Containing Unnatural Amino Acids in Saccharomyces cerevisiae*. Journal of Molecular Biology, 2007. **371**(1): p. 112-122.
116. Deiters, A., et al., *A genetically encoded photocaged tyrosine*. Angew Chem Int Ed Engl, 2006. **45**(17): p. 2728-31.
117. Hohsaka, T., et al., *Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an Escherichia coli in Vitro Translation System*. Biochemistry, 2001. **40**(37): p. 11060-11064.
118. Muranaka, N., T. Hohsaka, and M. Sisido, *Four-base codon mediated mRNA display to construct peptide libraries that contain multiple nonnatural amino acids*. Nucleic acids research, 2006. **34**(1): p. e7-e7.
119. Taira, H., et al., *Four-base codon-mediated incorporation of nonnatural amino acids into proteins in a eukaryotic cell-free translation system*. Journal of Bioscience and Bioengineering, 2005. **99**(5): p. 473-476.
120. Taki, M., J. Matsushita, and M. Sisido, *Expanding the Genetic Code in a Mammalian Cell Line by the Introduction of Four-Base Codon/Anticodon Pairs*. 2006. **7**(3): p. 425-428.
121. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(20): p. 7566-7571.
122. Rackham, O. and J.W. Chin, *A network of orthogonal ribosome-mRNA pairs*. Nature Chemical Biology, 2005. **1**: p. 159.
123. Wang, K., et al., *Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion*. Nature Biotechnology, 2007. **25**: p. 770.
124. Chen, I.A. and M. Schindlinger, *Quadruplet codons: one small step for a ribosome, one giant leap for proteins: an expanded genetic code could address fundamental questions about algorithmic information, biological function, and the origins of life*. BioEssays : news and reviews in molecular, cellular and developmental biology, 2010. **32**(8): p. 650-654.
125. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool*. Biochimica et biophysica acta, 2014. **1844**(6): p. 1059-1070.
126. Perona, J.J. and A. Hadd, *Structural Diversity and Protein Engineering of the Aminoacyl-tRNA Synthetases*. Biochemistry, 2012. **51**(44): p. 8705-8729.
127. Lacey, V.K., et al., *Expanding the library and substrate diversity of the pyrrolysyl-tRNA synthetase to incorporate unnatural amino acids containing conjugated rings*. ChemBiochem : a European journal of chemical biology, 2013. **14**(16): p. 2100-2105.
128. Xiao, H., et al., *Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase*. ACS chemical biology, 2014. **9**(5): p. 1092-1096.
129. Nguyen, D.P., et al., *Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Specific Protein Labeling via a Bio-orthogonal Cyanobenzothiazole Condensation*. Journal of the American Chemical Society, 2011. **133**(30): p. 11418-11421.
130. Myers, R.M., Z. Larin, and T. Maniatis, *Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes*. Science, 1985. **230**(4731): p. 1242.
131. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology*. Nature Reviews Molecular Cell Biology, 2012. **13**: p. 168.
132. Minnihan, E.C., et al., *Incorporation of fluorotyrosines into ribonucleotide reductase using an evolved, polyspecific aminoacyl-tRNA synthetase*. Journal of the American Chemical Society, 2011. **133**(40): p. 15942-15945.

133. Mitchell, A.L., et al., *A Unique Genetically Encoded FRET Pair in Mammalian Cells*. 2017. **18**(6): p. 511-514.
134. Xiao, H. and P.G. Schultz, *At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code*. 2016.
135. Tian, F., et al., *A general approach to site-specific antibody drug conjugates*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(5): p. 1766-1771.
136. VanBrunst, M.P., et al., *Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody–Drug Conjugates Using Click Cycloaddition Chemistry*. Bioconjugate Chemistry, 2015. **26**(11): p. 2249-2260.
137. Love, K.R., et al., *Enabling Glycosyltransferase Evolution: A Facile Substrate-Attachment Strategy for Phage-Display Enzyme Evolution*. 2006. **7**(5): p. 753-756.
138. Liu, C.C., et al., *Protein evolution with an expanded genetic code*. Proceedings of the National Academy of Sciences, 2008. **105**(46): p. 17688.
139. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea*. 1962. **59**(3): p. 223-239.
140. Enterina, J.R., L. Wu, and R.E. Campbell, *Emerging fluorescent protein technologies*. Current Opinion in Chemical Biology, 2015. **27**: p. 10-17.
141. Mishin, A.S., et al., *Novel uses of fluorescent proteins*. Current Opinion in Chemical Biology, 2015. **27**: p. 1-9.
142. Lippincott-Schwartz, J. and G.H. Patterson, *Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging*. Trends Cell Biol, 2009. **19**(11): p. 555-65.
143. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein*. Biochemistry, 1997. **36**(22): p. 6786-91.
144. Cubitt, A.B., et al., *Understanding, improving and using green fluorescent proteins*. Trends Biochem Sci, 1995. **20**(11): p. 448-55.
145. Heim, R., D.C. Prasher, and R.Y. Tsien, *Wavelength mutations and posttranslational autoxidation of green fluorescent protein*. Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12501-4.
146. Tsien, R.Y., *The green fluorescent protein*. Annu Rev Biochem, 1998. **67**: p. 509-44.
147. Levantino, M., et al., *Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules*. Curr Opin Struct Biol, 2015. **35**: p. 41-8.
148. Burgess, N.K., et al., *Beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro*. The Journal of biological chemistry, 2008. **283**(39): p. 26748-26758.
149. Meccas, J., et al., *Identification and characterization of an outer membrane protein, OmpX, in Escherichia coli that is homologous to a family of outer membrane proteins including Ail of Yersinia enterocolitica*. Journal of Bacteriology, 1995. **177**(3): p. 799.
150. Vogt, J. and G.E. Schulz, *The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence*. Structure, 1999. **7**(10): p. 1301-1309.
151. Fleming, A. and V.D. Allison, *Observations on a Bacteriolytic Substance (“Lysozyme”) Found in Secretions and Tissues*. British journal of experimental pathology, 1922. **3**(5): p. 252-260.
152. Ragland, S.A. and A.K. Criss, *From bacterial killing to immune modulation: Recent insights into the functions of lysozyme*. PLoS pathogens, 2017. **13**(9): p. e1006512-e1006512.
153. Lukyanov, K.A., et al., *Innovation: Photoactivatable fluorescent proteins*. Nat Rev Mol Cell Biol, 2005. **6**(11): p. 885-91.
154. Fernandez-Suarez, M. and A.Y. Ting, *Fluorescent probes for super-resolution imaging in living cells*. Nat Rev Mol Cell Biol, 2008. **9**(12): p. 929-43.
155. Patterson, G.H. and J. Lippincott-Schwartz, *A photoactivatable GFP for selective photolabeling of proteins and cells*. Science, 2002. **297**(5588): p. 1873-7.
156. Ando, R., H. Mizuno, and A. Miyawaki, *Regulated fast nucleocytoplasmic shuttling observed by*

- reversible protein highlighting*. Science, 2004. **306**(5700): p. 1370-3.
157. Wiedenmann, J., et al., *EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion*. Proc Natl Acad Sci U S A, 2004. **101**(45): p. 15905-10.
158. Tsutsui, H., et al., *Semi-rational engineering of a coral fluorescent protein into an efficient highlighter*. EMBO Rep, 2005. **6**(3): p. 233-8.
159. Ando, R., et al., *An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12651-6.
160. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, Aequorea*. J Cell Comp Physiol, 1962. **59**: p. 223-39.
161. Ormo, M., et al., *Crystal structure of the Aequorea victoria green fluorescent protein*. Science, 1996. **273**(5280): p. 1392-5.
162. Yang, F., L.G. Moss, and G.N. Phillips, Jr., *The molecular structure of green fluorescent protein*. Nat Biotechnol, 1996. **14**(10): p. 1246-51.
163. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression*. Science, 1994. **263**(5148): p. 802.
164. Kneen, M., et al., *Green fluorescent protein as a noninvasive intracellular pH indicator*. Biophysical journal, 1998. **74**(3): p. 1591-1599.
165. Romoser, V.A., P.M. Hinkle, and A. Persechini, *Detection in Living Cells of Ca²⁺-dependent Changes in the Fluorescence Emission of an Indicator Composed of Two Green Fluorescent Protein Variants Linked by a Calmodulin-binding Sequence: A NEW CLASS OF FLUORESCENT INDICATORS*. 1997. **272**(20): p. 13270-13274.
166. Miyawaki, A., et al., *Fluorescent indicators for Ca²⁺-based on green fluorescent proteins and calmodulin*. Nature, 1997. **388**: p. 882.
167. Chapman, H.N., et al., *Femtosecond X-ray protein nanocrystallography*. Nature, 2011. **470**(7332): p. 73-77.
168. Schlichting, I., *Serial femtosecond crystallography: the first five years*. IUCrJ, 2015. **2**(Pt 2): p. 246-255.
169. Martin-Garcia, J.M., et al., *Serial femtosecond crystallography: A revolution in structural biology*. Archives of biochemistry and biophysics, 2016. **602**: p. 32-47.
170. Kubelka, J., *Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics*. Photochem Photobiol Sci, 2009. **8**(4): p. 499-512.
171. Lee, H.M., D.R. Larson, and D.S. Lawrence, *Illuminating the chemistry of life: design, synthesis, and applications of "caged" and related photoresponsive compounds*. ACS Chem Biol, 2009. **4**(6): p. 409-27.
172. Trincao, J., et al., *Dynamic structural science: recent developments in time-resolved spectroscopy and X-ray crystallography*. Biochem Soc Trans, 2013. **41**(5): p. 1260-4.
173. Groff, D., et al., *A new strategy to photoactivate green fluorescent protein*. Angew Chem Int Ed Engl, 2010. **49**(42): p. 7677-9.
174. Krafft, G.A., W.R. Sutton, and R.T. Cummings, *Photoactivable fluorophores. 3. Synthesis and photoactivation of fluorogenic difunctionalized fluoresceins*. J Am Chem Soc, 1988. **110**(1): p. 301-303.
175. Adams, S.R. and R.Y. Tsien, *Controlling cell chemistry with caged compounds*. Annu Rev Physiol, 1993. **55**: p. 755-84.
176. Givens, R.S., et al., *New photoprotecting groups: desyl and p-hydroxyphenacyl phosphate and carboxylate esters*. Methods Enzymol, 1998. **291**: p. 1-29.
177. Miller, J.C., et al., *Flash Decaging of Tyrosine Sidechains in an Ion Channel*. Neuron, 1998. **20**(4): p. 619-624.
178. Pedelacq, J.D., et al., *Engineering and characterization of a superfolder green fluorescent protein*. Nat Biotechnol, 2006. **24**(1): p. 79-88.
179. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

180. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 282-92.
181. McCoy, A.J., et al., *Phaser crystallographic software*. J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
182. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation*. Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.
183. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
184. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine*. Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.
185. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
186. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry*. Anal Chem, 2006. **78**(21): p. 7473-83.
187. Brejc, K., et al., *Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein*. Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2306-11.
188. Chatteraj, M., et al., *Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer*. Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8362-7.
189. Fang, C., et al., *Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy*. Nature, 2009. **462**: p. 200.
190. Ueno, T., et al., *Rational principles for modulating fluorescence properties of fluorescein*. J Am Chem Soc, 2004. **126**(43): p. 14079-85.
191. Kobayashi, T., et al., *Highly activatable and rapidly releasable caged fluorescein derivatives*. J Am Chem Soc, 2007. **129**(21): p. 6696-7.
192. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
193. Rousseau, F., J.W. Schymkowitz, and L.S. Itzhaki, *The unfolding story of three-dimensional domain swapping*. Structure, 2003. **11**(3): p. 243-51.
194. Liu, Y. and D. Eisenberg, *3D domain swapping: as domains continue to swap*. Protein Sci, 2002. **11**(6): p. 1285-99.
195. Shameer, K., et al., *3DSwap: curated knowledgebase of proteins involved in 3D domain swapping*. Database (Oxford), 2011. **2011**: p. bar042.
196. Bennett, M.J., M.P. Schlunegger, and D. Eisenberg, *3D domain swapping: a mechanism for oligomer assembly*. Protein Sci, 1995. **4**(12): p. 2455-68.
197. Bennett, M.J., S. Choe, and D. Eisenberg, *Domain swapping: entangling alliances between proteins*. Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3127-31.
198. Liu, Y., et al., *Structures of the two 3D domain-swapped RNase A trimers*. Protein Sci, 2002. **11**(2): p. 371-80.
199. Penna, T.C., et al., *Thermal stability of recombinant green fluorescent protein (GFPuv) at various pH values*. Appl Biochem Biotechnol, 2004. **113-116**: p. 469-83.
200. Pitman, D.J., et al., *Exploring the folding pathway of green fluorescent protein through disulfide engineering*. Protein Sci, 2015. **24**(3): p. 341-53.
201. Hsu, S.T., G. Blaser, and S.E. Jackson, *The folding, stability and conformational dynamics of beta-barrel fluorescent proteins*. Chem Soc Rev, 2009. **38**(10): p. 2951-65.
202. Huang, J.R., et al., *Stable intermediate states and high energy barriers in the unfolding of GFP*. J Mol Biol, 2007. **370**(2): p. 356-71.
203. Andrews, B.T., et al., *The rough energy landscape of superfolder GFP is linked to the chromophore*. J Mol Biol, 2007. **373**(2): p. 476-90.
204. Andrews, B.T., et al., *The dual-basin landscape in GFP folding*. Proc Natl Acad Sci U S A, 2008. **105**(34): p. 12283-8.

205. Orte, A., et al., *Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence*. J Am Chem Soc, 2008. **130**(25): p. 7898-907.
206. Enoki, S., et al., *The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein*. J Mol Biol, 2006. **361**(5): p. 969-82.
207. Dietz, H. and M. Rief, *Exploring the energy landscape of GFP by single-molecule mechanical experiments*. Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16192-7.
208. Reddy, G., Z. Liu, and D. Thirumalai, *Denaturant-dependent folding of GFP*. Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17832-8.
209. Chandler, S.A. and J.L. Benesch, *Mass spectrometry beyond the native state*. Curr Opin Chem Biol, 2018. **42**: p. 130-137.
210. Koebnik, R., K.P. Locher, and P. Van Gelder, *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*. Molecular Microbiology, 2000. **37**(2): p. 239-253.
211. Blattner, F.R., et al., *The Complete Genome Sequence of Escherichia coli K-12*. Science, 1997. **277**(5331): p. 1453.
212. Rosenbusch, J.P., *Secondary and tertiary structure of membrane proteins*. Zentralblatt fuer Bakteriologie, Mikrobiologie und Hygiene, 1988. **17**: p. 259 - 266.
213. Pautsch, A. and G.E. Schulz, *Structure of the outer membrane protein A transmembrane domain*. Nature Structural Biology, 1998. **5**: p. 1013.
214. Vogt, J. and G.E. Schulz, *The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence*. Structure, 1999. **7**(10): p. 1301-1309.
215. Snijder, H.J., et al., *Structural evidence for dimerization-regulated activation of an integral membrane phospholipase*. Nature, 1999. **401**: p. 717.
216. Garavito, R.M. and J.P. Rosenbusch, *Three-dimensional crystals of an integral membrane protein: an initial x-ray analysis*. The Journal of Cell Biology, 1980. **86**(1): p. 327.
217. Cowan, S.W., et al., *Crystal structures explain functional properties of two E. coli porins*. Nature, 1992. **358**: p. 727.
218. Schirmer, T., et al., *Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution*. Science, 1995. **267**(5197): p. 512.
219. Forst, D., et al., *Structure of the sucrose-specific porin ScrY from Salmonella typhimurium and its complex with sucrose*. Nature Structural Biology, 1998. **5**: p. 37.
220. Ferguson, A.D., et al., *Siderophore-Mediated Iron Transport: Crystal Structure of FhuA with Bound Lipopolysaccharide*. Science, 1998. **282**(5397): p. 2215.
221. Locher, K.P., et al., *Transmembrane Signaling across the Ligand-Gated FhuA Receptor: Crystal Structures of Free and Ferrichrome-Bound States Reveal Allosteric Changes*. Cell, 1998. **95**(6): p. 771-778.
222. Buchanan, S.K., et al., *Crystal structure of the outer membrane active transporter FepA from Escherichia coli*. Nature Structural Biology, 1999. **6**: p. 56.
223. Schulz, G.E., *The structure of bacterial outer membrane proteins*. Biochimica et Biophysica Acta (BBA) - Biomembranes, 2002. **1565**(2): p. 308-317.
224. Heffernan, E.J., et al., *Specificity of the complement resistance and cell association phenotypes encoded by the outer membrane protein genes rck from Salmonella typhimurium and ail from Yersinia enterocolitica*. Infection and immunity, 1994. **62**(11): p. 5183-5186.
225. de Kort, G., et al., *Invasion of rabbit ileal tissue by Enterobacter cloacae varies with the concentration of OmpX in the outer membrane*. Infection and immunity, 1994. **62**(11): p. 4722-4726.
226. Kolodziejek, A.M., C.J. Hovde, and S.A. Minnich, *Yersinia pestis Ail: multiple roles of a single protein*. Frontiers in cellular and infection microbiology, 2012. **2**: p. 103-103.
227. Kim, K., et al., *Outer membrane proteins A (OmpA) and X (OmpX) are essential for basolateral invasion of Cronobacter sakazakii*. Applied and environmental microbiology, 2010. **76**(15): p. 5188-5198.

228. Confer, A.W. and S. Ayalew, *The OmpA family of proteins: Roles in bacterial pathogenesis and immunity*. Veterinary Microbiology, 2013. **163**(3): p. 207-222.
229. Jeannin, P., et al., *OmpA targets dendritic cells, induces their maturation and delivers antigen into the MHC class I presentation pathway*. Nature Immunology, 2000. **1**: p. 502.
230. Turro, N.J., et al., *CONVENIENT AND SIMPLE METHODS FOR THE OBSERVATION OF PHOSPHORESCENCE IN FLUID SOLUTIONS. INTERNAL AND EXTERNAL HEAVY ATOM AND MICELLAR EFFECTS*. Photochemistry and Photobiology, 1978. **27**(5): p. 523-529.
231. Dorman, G. and G.D. Prestwich, *Benzophenone Photophores in Biochemistry*. Biochemistry, 1994. **33**(19): p. 5661-5673.
232. Galardy, R.E., et al., *Photoaffinity Labeling of Peptide Hormone Binding Sites*. 1974. **249**(11): p. 3510-3518.
233. Kauer, J.C., et al., *p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide*. 1986. **261**(23): p. 10695-700.
234. Mange, J.-P., et al., *Adhesive properties of Enterobacter sakazakii to human epithelial and brain microvascular endothelial cells*. BMC microbiology, 2006. **6**: p. 58-58.
235. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis*. Nature methods, 2012. **9**(7): p. 676-682.
236. van der Eb, A., *USA FDA CTR For Biologics Evaluation and Research Vaccines and Related Biological Products Advisory Committee Meeting*. 2012: p. 14-22.
237. Masters, J.R., *HeLa cells 50 years on: the good, the bad and the ugly*. Nature Reviews Cancer, 2002. **2**: p. 315.
238. Tsuchiya, S., et al., *Establishment and characterization of a human acute monocytic leukemia cell line (THP-1)*. International Journal of Cancer, 1980. **26**(2): p. 171-176.
239. Abraham, R.T. and A. Weiss, *Jurkat T cells and development of the T-cell receptor signalling paradigm*. Nature Reviews Immunology, 2004. **4**: p. 301.
240. Maisnier-Patin, K., et al., *The outer membrane protein X from Escherichia coli exhibits immune properties*. Vaccine, 2003. **21**(25): p. 3765-3774.
241. Fafarman, A.T., et al., *Decomposition of Vibrational Shifts of Nitriles into Electrostatic and Hydrogen-Bonding Effects*. Journal of the American Chemical Society, 2010. **132**(37): p. 12811-12813.
242. Cho, M., *Coherent Two-Dimensional Optical Spectroscopy*. Chemical Reviews, 2008. **108**(4): p. 1331-1418.
243. Kim, H. and M. Cho, *Infrared Probes for Studying the Structure and Dynamics of Biomolecules*. Chemical Reviews, 2013. **113**(8): p. 5817-5847.
244. Ma, J., et al., *Site-specific infrared probes of proteins*. Annual review of physical chemistry, 2015. **66**: p. 357-377.
245. Atkins, P. and J. De Paula, *Atkins' Physical Chemistry*. 2006. 783-827.
246. Garidel, P. and H. Schott, *Fourier-transform midinfrared spectroscopy for analysis and screening of liquid protein formulations. Part 1: understanding infrared spectroscopy of proteins*. Vol. 4. 2006. 40-46.
247. Ganim, Z., et al., *Amide I Two-Dimensional Infrared Spectroscopy of Proteins*. Accounts of Chemical Research, 2008. **41**(3): p. 432-441.
248. Kim, Y.S. and R.M. Hochstrasser, *Applications of 2D IR Spectroscopy to Peptides, Proteins, and Hydrogen-Bond Dynamics*. The Journal of Physical Chemistry B, 2009. **113**(24): p. 8231-8251.
249. Zhao, Y., et al., *Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling*. Nature Nanotechnology, 2014. **9**: p. 466.
250. Jo, H., et al., *Selective incorporation of nitrile-based infrared probes into proteins via cysteine alkylation*. Biochemistry, 2010. **49**(49): p. 10354-10356.
251. Fafarman, A.T., et al., *Site-specific conversion of cysteine thiols into thiocyanate creates an IR probe for electric fields in proteins*. Journal of the American Chemical Society, 2006. **128**(41): p. 13356-13357.

252. Getahun, Z., et al., *Using Nitrile-Derivatized Amino Acids as Infrared Probes of Local Environment*. Journal of the American Chemical Society, 2003. **125**(2): p. 405-411.
253. Bazewicz, C.G., et al., *Expanding the Utility of 4-Cyano-L-Phenylalanine As a Vibrational Reporter of Protein Environments*. The Journal of Physical Chemistry B, 2012. **116**(35): p. 10824-10831.
254. Bazewicz, C.G., et al., *Sensitive, Site-Specific, and Stable Vibrational Probe of Local Protein Environments: 4-Azidomethyl-L-Phenylalanine*. The Journal of Physical Chemistry B, 2013. **117**(30): p. 8987-8993.
255. Oh, K.-I., et al., *β -Azidoalanine as an IR Probe: Application to Amyloid A β (16-22) Aggregation*. The Journal of Physical Chemistry B, 2008. **112**(33): p. 10352-10357.
256. Maienschein-Cline, M.G. and C.H. Londergan, *The CN Stretching Band of Aliphatic Thiocyanate is Sensitive to Solvent Dynamics and Specific Solvation*. The Journal of Physical Chemistry A, 2007. **111**(40): p. 10020-10025.
257. Alía, J.M. and H.G.M. Edwards, *Vibrational Spectroscopic Properties of Hydrogen Bonded Acetonitrile Studied by DFT*. The Journal of Physical Chemistry A, 2005. **109**(35): p. 7977-7987.
258. Choi, J.-H., et al., *Nitrile and thiocyanate IR probes: Quantum chemistry calculation studies and multivariate least-square fitting analysis*. The Journal of Chemical Physics, 2008. **128**(13): p. 134506.
259. van Wilderen, L.J.G.W., et al., *Correction: Vibrational dynamics and solvatochromism of the label SCN in various solvents and hemoglobin by time dependent IR and 2D-IR spectroscopy*. Physical Chemistry Chemical Physics, 2017. **19**(14): p. 9676-9678.
260. Wolfshorndl, M.P., et al., *Covalently Bound Azido Groups Are Very Specific Water Sensors, Even in Hydrogen-Bonding Environments*. The Journal of Physical Chemistry B, 2012. **116**(3): p. 1172-1179.
261. Creon, A., et al., *Conformation-specific detection of calmodulin binding using the unnatural amino acid p-azido-phenylalanine (AzF) as an IR-sensor*. Structural dynamics (Melville, N.Y.), 2018. **5**(6): p. 064701-064701.
262. Reiner, A., P. Henklein, and T. Kiefhaber, *An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain*. Proceedings of the National Academy of Sciences, 2010. **107**(11): p. 4955.
263. R Rose, D., et al., *Crystal structure of T4-lysozyme generated from synthetic coding DNA expressed in Escherichia coli*. Vol. 2. 1988. 277-82.
264. Jollès, P. and J. Jollès, *What's new in lysozyme research?* Molecular and Cellular Biochemistry, 1984. **63**(2): p. 165-189.
265. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1-2**: p. 19-25.
266. Duan, Y., et al., *A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations*. Vol. 24. 2003. 1999-2012.
267. Faber, H.R. and B.W. Matthews, *A mutant T4 lysozyme displays five different crystal conformations*. Nature, 1990. **348**: p. 263.
268. Hess, B., et al., *LINCS: A Linear Constraint Solver for molecular simulations*. Vol. 18. 1998.
269. Essmann, U., et al., *A smooth particle mesh Ewald method*. The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.
270. J Gowers, R., et al., *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*. 2016.
271. Michaud-Agrawal, N., et al., *MDAnalysis: a toolkit for the analysis of molecular dynamics simulations*. Journal of computational chemistry, 2011. **32**(10): p. 2319-2327.
272. Oliphant, T.E., *Python for Scientific Computing*. Computing in Science & Engineering, 2007. **9**(3): p. 10-20.
273. Hunter, J.D., *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 2007. **9**(3): p. 90-95.

274. Taylor, J., *Mantid, A high performance framework for reduction and analysis of neutron scattering data*. Vol. 57. 2012.
275. Arnold, O., et al., *Mantid—Data analysis and visualization package for neutron scattering and μ SR experiments*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2014. **764**: p. 156-166.
276. Bagchi, S., S.D. Fried, and S.G. Boxer, *A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields*. Journal of the American Chemical Society, 2012. **134**(25): p. 10373-10376.
277. Matthews, B.W., *Studies on Protein Stability With T4 Lysozyme*, in *Advances in Protein Chemistry*, C.B. Anfinsen, et al., Editors. 1995, Academic Press. p. 249-278.
278. Lindquist, B.A., K.E. Furse, and S.A. Corcelli, *Nitrile groups as vibrational probes of biomolecular structure and dynamics: an overview*. Physical Chemistry Chemical Physics, 2009. **11**(37): p. 8119-8132.