Optimization of Cell-Free Protein Synthesis by Proteomics and Metabolic Engineering of *Escherichia coli* A19

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Daniel Foshag

aus Weinheim

Hauptberichter:	apl. Prof. Dr. Steffen Rupp
Mitberichter:	Prof. Dr. Roland Kontermann
Tag der mündlichen Prüfung:	23.10.2019

Institut für Grenzflächenverfahrenstechnik und Plasmatechnologie der Universität Stuttgart

2019

Abstract

Cell-free protein synthesis (CFPS) has emerged as a standard protein production system over the last two decades. Due to its open nature and various methods of directly influencing protein expression, it has replaced or complemented *in vivo* expression systems, especially for the expression of toxins, membrane proteins and other difficult-to-express proteins. Despite the widespread use of CFPS, the main component of the system, an extract derived from the centrifugation of a bacterial lysate (S30 extract), still has not been defined thoroughly. S30 extract preparation often causes changes in protein composition, altering the original proteome of exponentially growing *Escherichia coli* (*E. coli*). To optimize CFPS in a rational manner, S30 extracts from the *E. coli* K12-derivative A19 were analyzed using a GeLC-MS approach. The S30 core proteome, consisting of 821 proteins detected in several replicates, was functionally integrated and categorized using GO terms, revealing the presence of complete pathways that can be explored for energy regeneration or precursor generation. To evaluate the effects of alternative growth conditions, S30 extracts derived from cells grown at SOS response-inducing conditions were analyzed by quantitative GeLC-MS using isotope-coded protein labeling (ICPL). These modified S30-S extracts contained 3-10-fold increased folding factors and were shown to improve the solubility and folding of difficult proteins. Therefore, the manipulation of the S30 extract proteome by modifying the cultivation conditions is an effective approach for the expression of challenging proteins. A second approach to improve CFPS productivity was the engineering of specific metabolic pathways through genetic modifications. Using the previously generated proteome as a guideline, 13 genes coding for various enzymes affecting protein, amino acid and mRNA stability were either tagged or knocked out in *E. coli* strains A19 and D10. After verifying the modifications by PCR and sequencing, the viability and fitness of the strains were examined. Additionally, the transcriptional profile of a heavily modified strain was compared with the original A19 strain, revealing highly coregulated transcriptome in response to the genetic modification. The amino acid concentrations of 19 amino acids were traced during a CFPS reaction, demonstrating that amino acids can be stabilized by genetic modifications. The engineered strains showed an increase in yield for some target proteins, highlighting the relevance of metabolic engineering when optimizing CFPS. Finally, one of the metabolically engineered strains was used as an extract source and combined with purified chaperones (DsbC, Skp and FkpA) to produce different antibody fragments. DsbC was the most important chaperone for Fab folding, whereas Skp and FkpA were beneficial to produce scFab.

Zusammenfassung

Die Zellfreie Proteinsynthese wurde ein Standard für die Synthese von Proteinen. Vorallem aufgrund der offenen Natur hat die zellfreie Proteinsynthese etablierte *in vivo* Expressionssysteme teilweise ersetzt oder ergänzt, unter anderem für die Expression von Toxinen, Membranproteinen und anderen schwer darstellbaren Proteinen. Trotz der weit verbreiteten Anwendung ist das für die Synthese verwendete bakterielle S30 Extrakt nur unzureichend charakterisiert. Im Vergleich zum Ausgangsproteom der exponentiell wachsenden *Escherichia coli* (*E. coli*), resultieren Verarbeitungsschritte in einer Änderung der Proteinzusammensetzung des S30 Extrakts. Für die gezielte Optimierung der zellfreien Proteinsynthese wurde der aus dem *E. coli* K12 Derivat A19 gewonnene S30 Extrakt mittels GeLCMS analysiert und das aus 821 Proteinen bestehende Kernproteom funktionell integriert und mittels GO Zuordnungen kategorisiert. Metabolischen Pfade, die für die Energie-regeneration und die Vorläufersynthese genutzt werden können, wurden dabei nachgewiesen.

Die Auswirkungen von alternativen Wachstumsbedingungen auf die Extraktleistung wurden untersucht, indem während der Kultivierung von *E. coli* eine Stressreaktion induziert wurde und das S30 Extrakt Proteom quantitativ analysiert wurde. Die modifizierten S30-S Extrakte zeigten einen 3-10fach erhöhten Gehalt an Faltungsfaktoren und eine verbesserte Löslichkeit einzelner Proteine. Deshalb ist die Manipulation der Kultivierungsbedingungen ein wirksamer Ansatz, schwer darstellbare Proteine in löslicher Form zu exprimieren.

Eine gezielte Optimierung von Stoffwechselpfaden mittels genetischer Modifikation wurde außerdem durchgeführt, um die Produktivität der zellfreie Proteinsynthese zu verbessern. Unter Verwendung des analysierten S30 Proteoms wurden die korrespondierenden Genomabschnitte von 13 Enzymen in den Stämmen *E. coli* A19 und D10 genetisch modifiziert. Diese Enzyme beeinflussen in unterschiedlicher Weise die Protein-, Aminosäuren- und mRNA Stabilität. Nachdem die genetischen Veränderungen verifiziert waren, wurde die Vitalität und Wachstumsfähigkeit der generierten Stämme untersucht. Darüber hinaus wurde das Transkriptom vor- und nach genetischer Modifikation verglichen, wobei zahlreiche co-regulierte Gene identifiziert wurden. Die Aminosäurenkonzentrationen von 19 Aminosäuren wurden in modifizierten S30 Extrakten verfolgt, wobei eine Stabilisierung nachgewiesen werden konnte. Modifizierte Stämme zeigten eine erhöhte Ausbeute für bestimmte Zielproteine. Zuletzt wurde optimiertes S30 Extrakt mit aufgereinigten Chaperonen (DsbC, Skp und FkpA) versetzt. Im Anschluss wurden unterschiedliche Antikörperfragmente exprimiert. DsbC stellte sich als das wichtigstes Chaperone für die Faltung von Fab heraus, während Skp und FkpA sich als positiv bei der Herstellung von scFab erwiesen.

Acknowledgements

The experiments presented in this Thesis were conducted at the Institute of Interfacial Process Engineering and Plasma Technology (IGVP) of the University of Stuttgart in care of Fraunhofer IGB, Stuttgart. Without the help of numerous people, this dissertation would not have been possible.

At first, I would like to thank my supervisor apl. Prof. Dr. Rupp for his guidance, support and mentoring over the course of this thesis. My second referee Prof. Dr. Kontermann I would like to thank for accepting to evaluate this thesis.

I'm grateful to Dr. Anke Burger Kentischer for the laboratory space and funding provided. Her guidance and support were invaluable to me.

Christian Kerger, Andre Feith, Anne-Katrin Wehrmann, Michael Egger, Lisa Steiner, Sebastian Schmitz and Sarah Maier contributed as either technician or student workers to this work. It was a great pleasure to work with you.

Drs. Erik Henrich and Frank Bernhard collaborated with me on the proteomics studies, resulting in a co-authored publication. Their ideas and experiments contributed to this work and advanced my knowledge significantly. Especially to Drs. Helmut Merk and Michael Gerrits from Biotechrabbit I'm thankful for helpful discussions.

I'm thankful to my wife Débora Teixeira Duarte for her moral support, her patience and her great editing skills. This thesis is dedicated to her.

This work was funded by the German Ministry of Education and Science (BMBF) and Merck.

Contents

List of Figures

List of Tables

List of Abbreviations

Ab	antibody
ABC	ATP-binding cassette
ACN	acetonitrile
ADP	adenosine diphosphate
AhR	aryl hydrocarbon receptor
ATP	adenosine triphsophate
bp	base pairs
CAT	chloramphenicol acetyltransferase
CECF	continuous-exchange cell-free
cf-Fab	cell-free fragment antigen-binding
CFPS	cell-free protein synthesis
cf-scFab	cell-free single chain fragment antigen-binding
CHO	chinese hamster ovary (cell line)
CTP	cytidine triphosphate
ddH_2O	double distilled H_2O
DNA	deoxyribonucleic acid
DTMB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
emPAI	exponentially modified protein abundance index
EtOH	ethanol
eYFP	enhanced yellow fluorescent protein
Fab	fragment antigen binding
FDA	United States Food and Drug Administration
FDR	false discovery rate
FLPe	Flippase
FPLC	fast protein liquid chromatography
FW	forward
gDNA	genomic DNA
GeLCMS	gel-enhanced liquid chromatography mass spectrometry
GFP	green fluorescent protein
GlcN6P	D-glucosamine-6-phosphate

GNAT	GCN5-related N-acetyltransferase
GO	gene ontology
GSH	gluthathione (reduced)
GSSG	gluthathione disulfide (oxidized)
GTP	guanosine triphosphate
HC	heavy chain
HCCA	alpha-cyano-4-hydroxycinnamic acid
HEK	human embryonic kidney
HRP	horseradish peroxidase
HsGNA1	human glucosamine 6-phosphate N-acetyltransferase
IAM	iodoacetamide
ICPL	isotope coded protein label
IEA	inferred from electronic annotation
IF	initiation factor
IgG	Immunoglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISTD	internal standard
K_d	equilibrium dissociation constant
kDa	kilo dalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	lysogeny broth
LC	light chain
MALDI	matrix-assisted laser desorption/ionization
MAPKAP1	mitogen-activated protein kinase associated protein 1
MCS	multiple cloning site
MeOH	methanol
mRNA	messenger ribonucleic acid
MW	molecular weight
MWCO	molecular weight cut-off
NAD^+	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
NGS	next generation sequencing
NMR	nuclear magnetic resonance
OCPS	open cell-free protein synthesis

OD	optical density
OPA	ortho phthalaldehyde
ORF	open reading frame
PANOx-SP	PEP, amino acids, NAD, oxalic acid, spermidine and putrescine
PCR	polymerase chain reaction
PEP	phosphoenyl pyruvate
PMSF	phenylmethylsulfonyl fluoride (protease inhibitor)
PURE	protein synthesis using recombinant elements
pV	P-value
PVDF	polyvinylidene difluoride
QPPC	quantitative proteomics P-value calculator
RBP	RNA binding protein
RBS	ribosomal binding site
RF	release factor
RFU	relative fluorescence units
RIN	RNA integrity number
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	rotations per minute
RT	room temperature
RV	reverse
SBP	streptavidine binding peptide
scFab	single chain fragment antigen binding
scFv	single chain variable fragment
scIgG	single chain IgG
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
sfGFP	superfolder green fluorescent protein
sGFP	shifted green fluorescent protein
SII	Strep tag II
SIN1	synonym MAPKAP1
SRP	signal recognition particle
T7RNAP	T7 RNA polymerase

TAE	Tris-acetate-EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween 20
TCA	trichloracetic acid
TF	trigger factor
TFA	trifluoracetic acid
tHC	truncated heavy chain
TOF	time of flight
Tris-OAc	Tris-acetate
tRNA	transfer ribonucleic acid
UTP	uridine triphosphate
UV	UV/visible
wt	wild type
YTPG	yeast tryptone phosphate glucose

1 General Introduction

Between 1950 and 1960, it became clear that physical integrity of cells was not necessary for translation of mRNA into protein. Protein synthesis from endogenous mRNA could still occur after cell lysis, in a so-called cell-free extract, after addition of the necessary building blocks (amino acids and energy sources) [1-5].

Shortly after these initial observations, a breakthrough was achieved by Nirenberg and Matthaei in 1961 through a simple step – a preincubation of the cell-free extracts at 30-37 °C (known as *run-off*). This step allowed the dissociation of mRNA from the ribosome and destruction of the endogenous mRNA within extracts [6]. This process represented the foundation to translate exogeneous mRNA and resulted ultimately in the deciphering of the genetic code [7]. These early cell-free systems were dependent on the addition of mRNA until the next landmark improvement was introduced: DNA-dependent protein synthesis based on the endogenous *E. coli* RNA polymerase (RNAP). These coupled transcription-translation systems had the advantage of synchronizing both processes, avoiding the formation of secondary and tertiary mRNA structures and their inherent instabilities [8, 9].

Currently, combined transcription-translation systems remain the most popular. Despite relying on T7-specific promoters and exogenous T7 polymerase, they are very efficient. The high transcriptional activity of phage derived T7 polymerase provides an excess amount of mRNA which is spatially and temporally disconnected from the translational activity of the ribosome. Given the possibilities to optimize the ribosomal binding site (RBS) and sequence to avoid secondary and tertiary structure formation of mRNA, the advantage provided by a coupled transcription-translation system became obsolete. Furthermore, the expression of endogenous DNA can be selectively inhibited by rifampicin; as a result, all building blocks are consumed only by the expression of target genes from PCR products or plasmids.

These developments were solely carried out in simple batch format and were used only for analytical purposes. The development of continuous flow/continuous exchange reaction formats marked other milestones and paved the way for preparative peptide and protein synthesis using cell-free protein synthesis, which was pioneered by Spirin and coworkers. They separated the high molecular weight components, such as transcription-translation machinery and other proteins, from the reaction mixture by a dialysis membrane and supplied the reaction mixture with low molecular weight building blocks and energy substrates (*e.g.*, amino acids, nucleotides) by diffusion through the dialysis membrane. Simultaneously, inhibitory byproducts such as inorganic phosphate were removed, extending the reaction time (up to 10

hours) and yield (mg mL^{-1} range) substantially [10-13], thereby paving the way for cell-free protein synthesis to become a preparative protein synthesis platform.

Cell-free systems have become a standard protein production platform over the course of the last two decades, complementing or substituting well established *in vivo* production systems. Two main advantages explain their popularity for certain applications: (*i*) the workflow for the production is shortened and allows high-throughput screening, and (*ii*) the open nature of the system provides direct access to the expression environment, presenting an opportunity to improve the folding of otherwise difficult-to-express proteins.

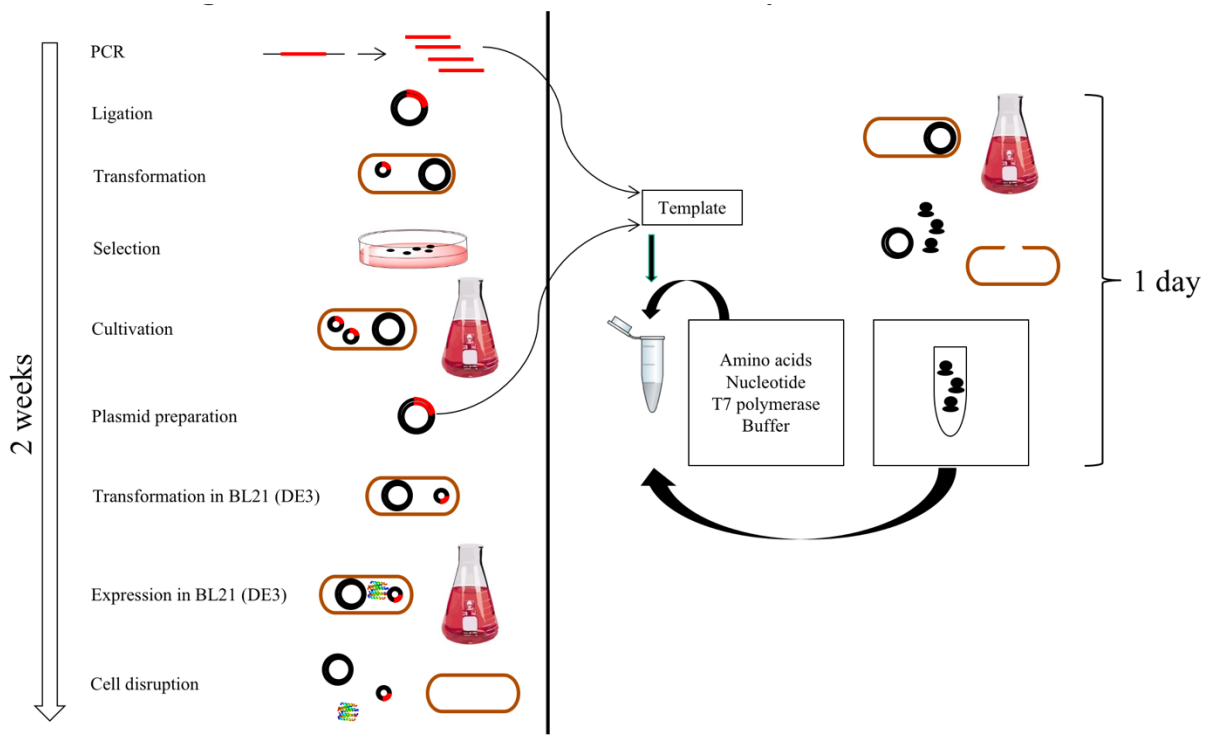


Figure 1: ***Comparison of in vivo and cell-free protein synthesis (CFPS)****. The workflow of conventional in vivo protein expression using E. coli is illustrated on the left side and the workflow for CFPS based on S30 extracts derived from E. coli is shown on the right side.*

The workflow to produce proteins *in vivo* and by cell-free protein expression with reference to (*i*) is depicted in Figure 1. *In vivo* protein production in *e.g., E. coli* comprises the following steps: amplification of the target gene and introduction of linkers for subsequent cloning steps by PCR; ligation of PCR product into the multiple cloning site (MCS) of an expression vector; transformation of the ligation product in a suitable strain such as *E. coli* DH5α; verification and selection of transformants containing the expression vector and the target gene; transformation of the expression vector in a strain suited for expression of the target gene (*e.g., E. coli* BL21 (DE3)); cultivation of the expression strain; and, finally, induction of protein production. The entire procedure, including cloning and expression, requires at least two weeks, and it is not always successful. Since protein expression occurs within the cell, the

folding process cannot be influenced and often results in the formation of large portions of incorrectly folded protein (inclusion bodies), which must then be refolded, if possible at all.

Ribosomal extracts for cell-free protein expression, on the other hand, can be prepared in large quantities, completely independent of the target protein, and frozen until needed. The preparation of extract requires, in a well-established environment, not more than 24 hours. Once supplemented with the reaction mixture containing small building blocks, polymerase and buffer, which can be stored frozen as well, the ribosomal extract can be directly used for protein expression. Linear DNA templates can be used, without laborious and time-consuming cloning steps, but circular DNA templates are equally suited. Depending on the application and protocol, the nature of the cell-free expression system allows one or more expression experiments to be performed in parallel microtiter plate-based high-throughput experiments including multiple optimized DNA templates, different expression temperatures, or expression in the presence of various additives.

The inherent open nature of cell-free systems (*ii*) allows the direct manipulation of the ribosomal extract, without any transport limitation, and without consideration for the maintenance of the metabolism and other biological functions of the living cell aside from the protein synthesis machinery. Compared to *in vivo* protein production systems such as *E. coli* or *Saccharomyces cerevisiae*, this method offers manifold possibilities to enable efficient production of difficult-to express soluble and/or membrane proteins [14-18]. The addition of chaperones, ligands, co-factors or hydrophobic compounds allows for optimization of protein folding as well as the generation of customized artificial expression environments [19-23].

Proteins containing disulfide bonds are usually not good target proteins for expression in *E. coli*, but in a cell-free system the redox environment can also be optimized, allowing for the formation of native disulfide bonds. Combinatorial labeling of proteins [24, 25], site-specific insertion of unnatural amino acids [26, 27], or the production of membrane proteins in artificial environments [28] are further frequent applications. The possibilities are summarized in Figure 2.

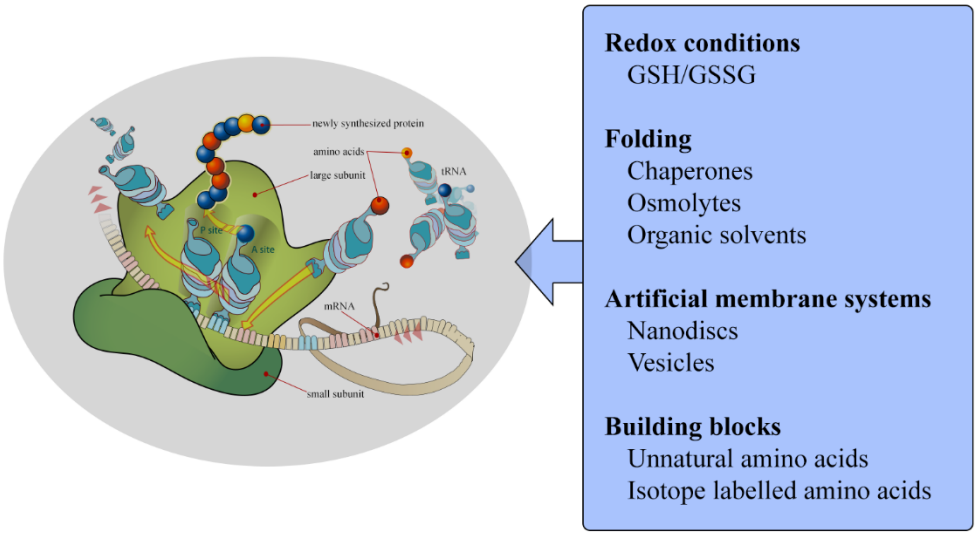


Figure 2: ***Cartoon highlighting the parameters that can be directly influenced due to the lack of a cell wall in CFPS systems****. (Illustration adapted from https://commons.wikimedia.org/wiki/File%3ARibosom_mRNA_translation_de.svg).*

1.1 *Bacterial species* for cell-free protein synthesis

Historically, the *E. coli* MRE600 and *E. coli* A19 strains are used for the preparation of ribosomal extracts for cell-free protein synthesis. The extract is prepared by centrifugation at 30,000 g and is therefore called the S30 extract. The MRE600 strain lacks RNase I, and exonuclease V is inactivated in A19, resulting in the stabilization of linear mRNA and DNA templates [29, 30]. While the MRE600 strain is rarely used for S30 extracts at present, A19 enjoys continued popularity due to the reliable and efficient S30 extracts derived from this strain [31-33]. The exact reasons for the suitability of this strain to produce S30 extracts for cell-free protein synthesis are unknown. However, it is likely that, due to the technique of UV-induced mutagenesis used to generate this strain, many uncharacterized mutations in addition to the exonuclease V mutation have been introduced, contributing to the strain's suitability for cell-free protein synthesis.

Besides *E. coli*, extremely fast-growing *Vibrio natriegens* gained some popularity in the recent years and could replace the workhorse *E. coli* at some point in the future. Several attempts to make ribosomal extract have been made with yields of up to 1.6 g/L in cell-free batch reactions. However, there have been only three reports so far showing varying results for the expression yield [34-36].

In addition to A19, other *E. coli* strains such as the commercially available B-strains have been explored and successfully employed for cell-free protein synthesis. In particular, the *E. coli* BL21 (DE3) and BL21 Star (DE3) strains have been successfully used for S30 extract preparation [37-39]. The BL21 (DE3) and BL21 Star (DE3) strains lack the protease Lon and

the outer membrane protease OmpT to stabilize the target protein and proteins resulting from proteolytic degradation. Additionally, the presence of T7 RNA-polymerase (T7RNAP) under control of the *lac*UV5 promoter allows for the expression of T7RNAP during cultivation upon induction with IPTG; therefore, supplementation of the final reaction mixture is not necessary [37]. In addition to the aforementioned modifications, the BL21 Star (DE3) strain (Thermo Fisher) lacks the C-terminal region of the RNase E (so-called *rne131* mutation) which is believed to be the main driver involved in mRNA decay [40].

While the use of these B-derived strains for cell-free protein synthesis impacts mRNA stability and target protein stability, the Swartz group took a targeted approach and genetically modified the conventionally used *E. coli* A19 (CGSC#: 5997) specifically to improve it for cell-free protein synthesis. Initially, they reversed the methionine mutation (metB1) in order to avoid the requirement of methionine in the growth medium and introduced the ΔendA knockout to stabilize DNA templates present in many cloning strains, such as *E. coli* DH5α or DH10B [41, 42]. The removal of fhuA protects the carrying strain against phage injection, which could cause genomic rearrangements and modifications of the designed genome [43]. Further knockouts such as ΔgshA, ΔsdaA/B, ΔtnaA and ΔspeA were introduced to stabilize the amino acids cysteine, serine, tryptophan and arginine, respectively. It was shown that the amino acid concentration remained stable over the course of the cell-free protein synthesis reaction [44, 45]. Interestingly, Nieß *et al.* showed that the limiting factors in cell-free protein synthesis are not necessarily the amino acids, but rather the ternary complex consisting of EFTu, tRNA and amino acid. The limited availability of the ternary complexes, accentuated by the presence of sub-optimal sequecences including repetitive codons, causes the elogantion rate to decrease and eventually the ribosome stalls [46].

A major drawback during protein production *in vivo* using *E. coli* as a host is the difficulty of producing proteins containing disulfide bonds, due to the reducing nature of the cytoplasm. Some specialty strains exist, such as the Origami strain (Novagen) lacking the thioredoxin reductase (*trxB*) and the glutathione reductase (*gor*), which are demonstrating improved cytoplasmic expression of disulfide-bonded proteins [47, 48]. Alternatively, expression in the periplasmic space has been initially described for the expression of disulfide-bonded scFv [49].

Despite these advances, *E. coli* is still considered a suboptimal host for the expression of disulfide-bonded proteins. The Swartz group also introduced the Δ*gor* knockout into their genetically modified A19 strain to stabilize the sulfhydryl redox potential in conjunction with pretreatment of the S30 extract with a low concentration of iodoacetamide (IAM) – the low

concentration of IAM was presumably still required to further inactivate *trxB*. However, the reduction of IAM from 1 mM in the presence of *gor* to 50 μM in the absence of *gor* allowed the use of glucose as an inexpensive energy source by preserving the enzymatic activity of glyceraldehyde 3-phosphate dehydrogenase [50]. The strain developed by the Swartz group is the basis for the commercial *E. coli*-based cell-free protein production platform employed by Sutro Biopharma (San Francisco, CA, USA), as implied by several of their publications [51-54]. The suitability of certain strains over others implies that the genetic background and the protein expression profile of the source strain play crucial roles with respect to the efficiency of the resulting S30 extract.

1.2 Energy regeneration in cell-free protein synthesis

Beside the genetic background of the source strain, energy regeneration is another culprit in cell-free protein synthesis. Initially, one-step phosphorylation reaction using substrate such as phosphoenolpyruvate (PEP) and creatine phosphate were used as single energy source in cell-free protein synthesis. [55]. These compounds contain a high-energy phosphate bond; the phosphate group is transferred in an enzymatic reaction *via* substrate-level phosphorylation to form ATP. While this process yields high concentrations of ATP to drive translation for a limited time (until the substrate is depleted), it also results in the accumulation of a large amount of inorganic phosphate, a known inhibitor of cell-free protein synthesis at concentrations > 30 mM [56, 57].

To overcome this limitation, Jewett *et al*. altered the reaction conditions by (*i*) avoiding unnatural compounds such as polythylen glycol, pH buffering compounds like HEPES and high concentrations of magnesium, (*ii*) reducing ionic compounds and (*iii*) adding Putrescine and Spermidine to the reaction mix. It turned out that, by providing a more natural environment, many pathways such as oxidative phosphorylation and enzymes of the TCA cycle were activated, providing a longer and steady supply of ATP at lower costs. In addition to the economic advantage of using cheap substrates like pyruvate, this method also keeps the homeostasis of inorganic phosphate [57, 58].

Until very recently it was not clear why inorganic phosphate is a limiting factor in cell-free protein synthesis. In 2016 Failmezger *et al.* showed that inorganic phosphate results in a reduction of available free magnesium by complexation, which in turn activates potential endonucleases (possibly MazF). These endonucleases can cleave 16S rRNA, inactivating the ribosome near the anti shine-dalgarno sequence [59].

Since the optimization of the reaction conditions, alternative cheap and long-lasting energy sources like polysaccharides (*e.g.* maltodextrin) and glucose were successfully used in cell-free protein synthesis employing the glycolysis pathway for energy regeneration [60-62].

Despite these advances, the most popular energy regeneration system is still the PANOxSP (PEP, Amino acids, NAD^+, Oxalic acid, Spermidine, Putrescine). This system uses a combination of high-energy phosphate donor like PEP, cofactors, spermidine/putrescine and optimized reaction conditions to facilitate integral pathways for energy regeneration [63].

1.3 S30 extract composition

Cell-free reactions contain 30-50% S30 extract, with a final protein concentration of 13-30 mg/mL. While a high degree of control exists for the low molecular weight compounds and any additives to cell-free reactions, the composition of the lysate, the main system component, remains ill-defined. The translational machinery though seems to be optimally reconstituted in S30 extract as shown by Freischmidt [64]. Extracts for cell-free protein production are routinely processed by centrifugation steps, heat treatment and extensive dialysis. Numerous proteins from the soluble lysate fraction precipitate during processing and are therefore depleted. Knowledge of the final lysate composition, especially with respect to proteases, chaperones or components of the transcription/translation machinery, is a prerequisite for a refined process design of cell-free synthetic biology.

To create a defined expression background, the PURE cell-free expression system has been developed by reconstitution of the *E. coli* translation machinery from purified individual components [65]. However, very low expression efficiency, time consuming preparation and the lack of stabilizing proteins limit its applications. Lysate proteomics studies will therefore enable new possibilities regarding the selection of additives and the creation of customized conditions to meet requirements of specific target proteins. Furthermore, the quality analysis of synthesized target proteins will be facilitated by the possession of knowledge regarding the background proteins potentially present within lysates [66].

In recent years, numerous parameters of the S30 lysate preparation have been successfully optimized to reduce time and costs while maintaining or improving the productivity of the system [9, 39, 67-71]. However, much less attention has been paid to the cultivation conditions that result in dynamic changes of the *E. coli* proteome in adaption to *e.g.,* stress starvation [72, 73]. Failmezger *et al*. investigated cultivation conditions by comparing the translational activity of S30 extracts derived exponentially growing *E. coli* vs. stressed *E. coli* (glucose limited). Strikingly, the overall stochiometry of important translational factors and

ribosomes remained optimal, and ribosomes were still efficiently translating. Extracts derived from such cultivation conditions allowed the use of alternative sigmar factors for transcription by the integral RNA-polymerase [74]. Such transient proteome modifications could be employed in S30 lysates and exert positive impacts on protein production. Exposure of *E. coli* A19 cells to temperature shifts or ethanol induces an SOS response, with increased production of chaperones, potentially leading to enrichment of these and other beneficial compounds in the resulting lysate [75].

1.4 Cell-free Protein Synthesis for Antibody Production

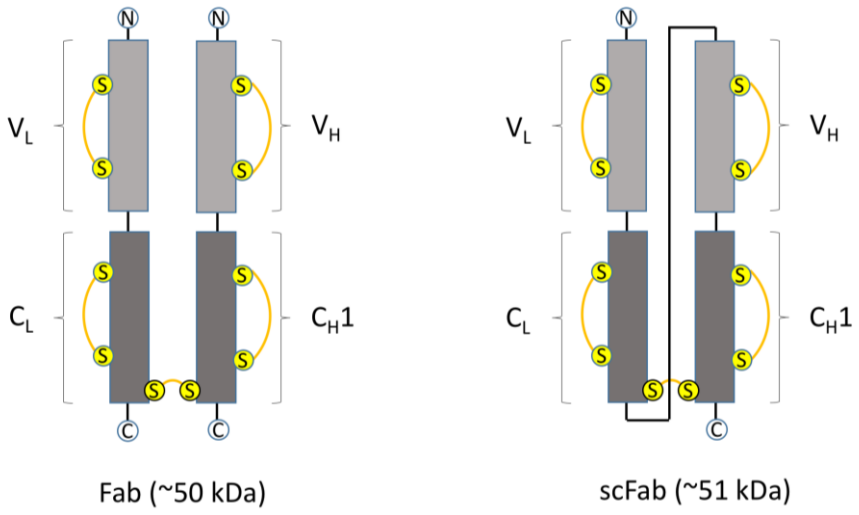
The distinct physiochemical properties of antibody fragments such as single chain variable fragments (scFvs) and antigen-binding fragments (Fabs) makes them interesting therapeutic agents. Due to their small size, antibody fragments penetrate tissue more rapidly than other antibody formats and provide means for targeted delivery of toxins or drugs to the desired site [76, 77]. Currently, Fabs and scFvs respectively account for 49% and 40% of antibody fragments in clinical development; currently, three Fabs (abciximab (Reopro), ranibizumab (Lucentis) and certolizumab pegol (Cimzia)) have been approved by the United States Food and Drug Administration (FDA) as therapeutics [78].

More importantly, antibody fragments are indispensable tools for antibody discovery techniques such as phage display [79] and ribosome display [80, 81]. Display technologies provide means for selection of antibodies against poor immunogenic epitopes and self-antigens, screening for high affinity antibodies (K_d values of up to 5×10^{-14} M have been achieved, which is a 10,000-fold improvement over natural antibodies) and antibodies with distinct biophysical properties that cannot be obtained using traditional hybridoma technology [81]. Consequently, there is considerable interest for economical and high-throughput compatible production systems for Fabs and scFvs. The most prevalent production host, *E coli*, accounted for 30% of all recombinantly produced biopharmaceuticals approved by the FDA in 2009. *E coli* is also the standard host for technologies such as phage and ribosome display. However, it is currently not considered the ideal expression host for Fabs and scFVs [82].

Problems constantly faced when expressing Fabs and scFvs in *E. coli* can be attributed to the complex structure of antibody fragments. Fabs consist of a light chain (LC) and a shortened heavy chain (HC) that lacks the hinge region and Fc region of the parental immunoglobulin. The LC and HC are joined at the C-terminus by an interchain disulfide bond. Each chain consists of two domains, namely, the variable and constant domains of LC and HC, respectively, each adopting a so-called immunoglobulin fold. Each of those four domains

consists of pairs of antiparallel beta-sheets around a hydrophobic core, which are additionally stabilized by four intrachain disulfide bonds. Thus, to adopt the native structure of Fabs, five disulfide bonds must be successfully formed, and the expression of LC and HC must be coordinated to avoid aggregation, misfolding and formation of homodimers of LC (also called Bence Jones proteins). The reducing environment in the cytoplasm, as well as the minimal or absent expression of appropriate chaperones, further result in the poor formation of disulfide linkages and aggregation when expressing Fabs in *E. coli.* A breakthrough enabled the solution to several of these issues by introducing scFvs structures, which consist of the variable domains of LC and HC joined by a flexible linker. The complexity of the scFv structure is greatly reduced, when compared to Fabs, since the coordinated expression of LC and HC can be omitted and only two intrachain disulfides must be formed to adopt their native fold; expression is therefore significantly improved in *E. coli*. Nevertheless, scFvs are prone to aggregation, are less stable and often exhibit lower affinities than their Fab counterparts; therefore, Fabs are still considered the more desirable antibody format.

To combine the advantage of scFvs and Fabs, the single chain Fab (scFab) was recently introduced [83] in which the LC is joined by linker with the HC; therefore, there is no need for coordinated expression of HC and LC, which in turn simplifies expression in *E. coli* (Figure 3).



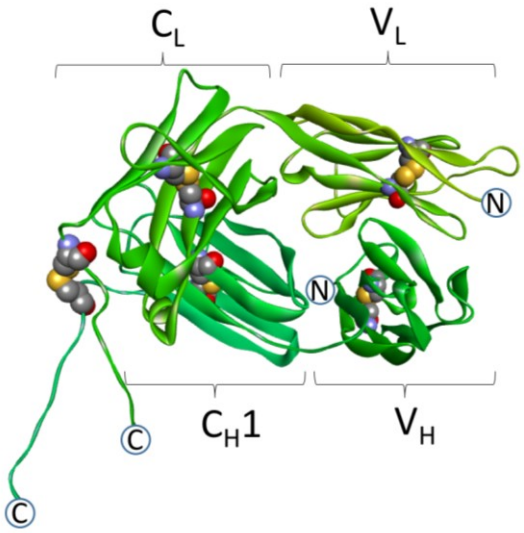


Figure 3: ***Schematic of Fab and scFab.*** *(A) Fab: tHC (tHC, truncated heavy chain consisting of V_H/C_H1) and LC (light chain) are connected by one interchain disulfide bond. Each chain forms two immunoglobulin domains stabilized by two intrachain disulfide bonds; (B) scFab: structurally equivalent to Fab except that tHC and LC are joined by 34 aa linker $(SGGG)_2(SEGGG)_4(SGGGSG)$; (C) Homology model of Fab generated with Modeller v9.13 shown as smooth ribbon. The inter- and intrachain disulfide bonds are shown as balls.*

Various scFab variants have been evaluated using different *in vivo* expression systems [83, 84]. When first describing the scFab format, Hust and colleagues evaluated scFab variants with 32 and 36 aa linkers as well as a variant without an interchain disulfide bond using a 34 aa linker. The authors found that their variant without the interchain disulfide bond and 34 aa linker exhibited high display levels; however, the products occurred mostly as multimers, while the scFab variant that retained the interchain disulfide bond resulted in reduced display levels, implying folding problems of the scFab constructs [83].

Due to the high level of aggregation of constructs without disulfides, which complicates downstream applications, Koerber and colleagues retained the interchain disulfide linkages and evaluated various linker lengths (50, 60, 70 and 80 aa) in combination with either Sec or SRP secretion systems in *E. coli* phage display. Strikingly, they achieved significantly higher display levels for all linker lengths, including the originally used 36 aa linker, when using the SRP pathway for secretion. The major difference between the Sec and SRP secretion pathways is that unfolded protein is translocated posttranslationally in the Sec pathway, while partially folded protein is translocated cotranslationally in the SRP pathway. Thus, their results imply that translocation might impose a burden with respect to the folding of scFab in *E. coli*. This notion is further supported by the fact that Koerber and colleagues did not observe significant differences when expressing scIgGs with various linker lengths in a mammalian expression system (HEK), indicating that oxidative environment, adapted chaperones and translocation systems in mammalian expression systems allow for the efficient folding of scFab independently of the linker length.

Motivated by these results, we decided to express scFab with retained interchain disulfide bonding and a short linker length of 36 aa, and we compare folding and assembly with Fab using an open cell-free protein synthesis (OCPS) system based on the ribosomal extract of *E. coli*. Since the cell wall is removed in the OCPS system, the expression of scFab can be evaluated independently of the translocation system, which seems to be rate-limiting in the folding of scFabs. The open nature of the system allows for further adjustment of the conditions such as redox potential and the titration of periplasmic chaperones of *E. coli* to achieve optimal cotranslational folding in the cytoplasmic fraction of *E. coli*.

2 Objectives

The S30 extract, the main component used in the cell-free protein synthesis reaction, is still considered a black box. The *E. coli* proteome derived from exponentially growing cells is significantly altered during processing of the lysate. The processing includes several centrifugation steps and incubation steps at higher temperatures, resulting in specific sedimentation and precipitation of proteins and membrane fractions. Despite numerous attempts to simplify and optimize the procedure, some core processing steps remain constant throughout all protocols and seem to be major determinants of the protein synthesis activity of the final S30 extract. The main goal of this thesis was to illuminate this "black box" called S30 extract, analyze its proteome composition, and then use this information as a guideline to improve cell-free protein production by removing deleterious proteins and adding relevant ones.

To achieve this goal, the S30 proteomes were analyzed by a GeLC-MS/MS approach, and identified using available *E. coli* representative proteomes. Afterwards, the proteome was functionally integrated and categorized using GO terms, to gain insights into the functional distribution and original location of S30 proteins. To evaluate the effects of alternative growth conditions, S30 lysates were prepared from cells grown at SOS response-inducing conditions (S30-S lysate), and compared by a quantitative GeLC-MS approach using isotope-coded protein labelling (ICPL). The modified S30-S lysate was characterized according to (*i*) regulated proteome composition; (*ii*) protein production efficiency, using a standard reporter protein; and (*iii*) folding efficiency of various difficult model proteins. Next, the strains *E. coli* A19 and D10 (Biotechrabbit, Berlin, Germany) were genetically modified. The modifications introduced into A19, as described by Swartz and colleagues [44, 45], were previously shown to be beneficial for cell-free protein synthesis reaction. In addition, to combine the benfits of BL21 (DE3), BL21 Star (DE3) and A19, the proteases Lon and ompT were removed in A19, and RNase E was tagged for removal during extract preparation. The modifications were verified using PCR and/or Western Blot (*i*), followed by recording their growth behavior under comparable conditions to evaluate their fitness after each genetic modification (*ii*). The effect of the amino acid stabilizing modifications was verified by following the concentration of 19 amino acids during a cell-free protein synthesis reaction (*iii*). Finally, the effect of the genetic modifications on the protein synthesis capacity of S30 extracts was evaluated using different model proteins (*iv*).

3 Materials and Methods

3.1 Materials

All chemicals were purchased from Sigma Aldrich (St. Louis, MO, USA) or Carl Roth (Karlsruhe, Germany) unless otherwise indicated. The *E. coli* strain A19 (CGSC# 5997; chromosomal markers: *rna-19*, *gdhA2*, *his-95*, *relA1*, *spoT1*, *metB1* [85]) was obtained from *E. coli* Genetic Stock Centre (Yale University, CT, USA). Serva ICPL Kit (Cat# 39230.01) was used for ICPL reaction in quantitative proteomics experiments. Cell free expression vector pIVEX2.3d and *E. coli* strain D10 were obtained from Biotechrabbit GmbH (Berlin, Germany). The plasmid pAR1219 (Sigma-Aldrich, St. Louis, MO, USA) contained the coding sequence for T7 polymerase under control of *lac*UV5 promoter and was used for overexpression of T7 polymerase in *E. coli* BL21 Star (DE3) (Thermo Scientific, Waltham, MA, USA). The coding sequence for Anti-Sin1 Antibody (clone 1C7.2) was kindly provided by Martin Zillmann and Joe Orlando from EMD Millipore Corporation (Bedford, MA, USA).

3.2 Primer Sequences

The primers used for plasmid modification (Table 1) and genome editing (Table 2) are listed as well as the primers used for verification and sequencing of knockouts are tabulated (Table 3). All primers were synthesized by Thermo Fisher Scientific (Waltham, MA, USA). Primers used for genome editing (Table 2) were additionally purified by HPLC.

*Table 1: Primers used for modification of plasmids. 5' phosphorylated termini are indicated by "*P*".*

Identifier	Sequence (5'-3')
HC_FW	*GACCACAACGGTTTCCCTCT*AGATAATAAATAAGAAGGAGATATACATA**ATG**GAAGTTCAATTGC
HC_RV	*ACTGCGGGTGGCTCCAGCTACCCTCGAG*ACAGCTTTTCGGCTCGACTTTTTT
LC_FW	*GAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTT*AGGAGGTAAAACATA**ATG**GA
LC_RV	*GGGTGGCTCCAGCTACCCTCGAG*GCACTCACCGCGGTTGAACGA
pEXTIR_MCS_FW	**P**_TCGAGGGTAGCTGGAGCCACCCGCAGTTCGAAAAATAATAAAAGGGCGAATTCCAGCACACTGGC GGCCGTTACTAGTG
pEXTIR_MCS1_RV	**P**_GATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTTTATTATTTTTCGAACTGCGGGTG GCTCCAGCTACCC
scFab_FW	*GAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTT*TAAGGAGGTAAAAA**ATG**GACA
scFab_RV	*GCGGGTGGCTCCAGCTACCCTCGAG*ACAGCTCTTAGGTTCAACTTTTTTGTCCA
T7-N6HIS_FW	**P**_GGTCATCATCACCATCACCACGGTTCTGGTAACACGATTAACATCGCTAAGAACG
T7-N6HIS_RV	**P**_CATTTAGTGCCTCTTCCAGTTAGTAAATCC

Table 2: Primers used for the generation of selection-cassettes for genome editing of E. coli A19 and D10. The respective target gene (GeneID) for modification is indicated in the identifier (bold).

Identifier	Sequence (5'-3')
cysS_FW	TGGGGATCGTGCTGGAAGATGGCCCGCAAGGGACCACCTGGCGTCGTAAGTCTTCTAAATCTAGAAG CGCG
cysS_RV	TGGCGCAGACGATAACCGGATGCGAAAACTCGCATCCGGCAATAGCGCAATAATACGACTCACTATA GGGCTC

Identifier	Sequence (5'-3')
endA_FW	CCAAAACAGCTTTCGCTACGTTGCTGGCTCGTTTTAACACGGAGTAAGTGAATTAACCCTCACTAAAG GGCG
endA_RV	GGTTGTACGCGTGGGGTAGGGGTTAACAAAAAGAATCCCGCTAGTGTAGGTAATACGACTCACTATA GGGCTC
fhuA_FW	AACAGCCAACTTGTGAAATGGGCACGGAAATCCGTGCCCCAAAAGAGAAAAATTAACCCTCACTAA AGGGCG
fhuA_RV	ATAATCATTCTCGTTTACGTTATCATTCACTTTACATCAGAGATATACCATAATACGACTCACTATAG GGCTC
gor_FW	ATTACGTCTCGCGCTACAATCGCGGTAATCAACGATAAGGACACTTTGTCAATTAACCCTCACTAAAG GGCG
gor_RV	CGGAAACGTAATTAAGGGCTAAGAGCACACTACTCTTAGCCCTTTAACATTAATACGACTCACTATA GGGCTC
gshA_FW	ACCATTACAGTTATGCTAATTAAAACGATTTTGACAGGCGGGAGGTCAATAATTAACCCTCACTAAA GGGCG
gshA_RV	AATTCAGAGATGAAATTTTGGCCACTCACGAGTGGCCTTTTTCTTTTCTGTAATACGACTCACTATAG GGCTC
lon_FW	CGAATTAGCCTGCCAGCCCTGTTTTTATTAGTGCATTTTGCGCGAGGTCAAATTAACCCTCACTAAAG GGCG
lon_RV	CAGTCGTGTCATCTGATTACCTGGCGGAAATTAAACTAAGAGAGAGCTCTTAATACGACTCACTATA GGGCTC
ompT_FW	ATATAAAAAATACATATTCAATCATTAAAACGATTGAATGGAGAACTTTTAATTAACCCTCACTAAA GGGCG
ompT_RV	AGTTATTCCCCGGGGCGATTTTCACCTCGGGGAAATTTTAGTTGGCGTTCTAATACGACTCACTATAG GGCTC
sdaA_FW	TGTTATTAGTTCGTTACTGGAAGTCCAGTCACCTTGTCAGGAGTATTATCAATTAACCCTCACTAAAG GGCG
sdaA_RV	AAAGCGGGTATAAATTCGCCCATCCGTTGCAGATGGGCGAGTAAGAAGTATAATACGACTCACTATA GGGCTC
sdaB_FW	CGCGCCGCTTTCGGGCGGCGCTTCCTCCGTTTTAACGCGATGTATTTCCTAATTAACCCTCACTAAAG GGCG
sdaB_RV	GGATGAGAAATCGGGAAGAGGCCTCGCAAAACGAGGCCTTTGGAGAGCGATAATACGACTCACTAT AGGGCTC
speA_FW	GCGATAGTCGTTAACTGTTTTACACTTAATAAAATAATTTGAGGTTCGCTAATTAACCCTCACTAAAG GGCG
speA_RV	TTGTGGCGGATTATCACCGCTAAATTAAGCGGATTCAAGTAACACAGGACTAATACGACTCACTATA GGGCTC
tnaA_FW	TGTAGGGTAAGAGAGTGGCTAACATCCTTATAGCCACTCTGTAGTATTAAAATTAACCCTCACTAAA GGGCG
tnaA_RV	TGTAATATTCACAGGGATCACTGTAATTAAAATAAATGAAGGATTATGTATAATACGACTCACTATA GGGCTC

Table 3: Primers used for verification of knockouts and sequencing. Identifiers containing the string "loc" bind approximately 500 bp up- and downstream of the locus of the target gene and identifiers containing "gene" bind within the target gene. The respective target gene (GeneID) is indicated by the identifier (bold).

Identifier	Sequence (5'-3')
cysS_loc_FW1	TGATGTCGGGCCACTATCGC
cysS_loc_RV1	CGTCTGTTATTTGCCGGAATCATCC
endA_gene_FW1	TACCGTTATTTGTCTATTGCTGCGG
endA_gene_RV1	TCTTTCGCGCCTGGCAAG
endA_loc_FW1	CCATCTGAACTATCAGGAACTTTCCTGATCT
endA_loc_FW2	CAATCTTCTGCCACTGCTGAAGC
endA_loc_RV1	AAATGCGGGGGATACGCATAGT
endA_loc_RV2	GGGCGTAACTACCCGGAGC
fhuA_gene_FW1	CCAAAACTGCTCAGCCAAAACAC
fhuA_gene_RV1	CCCCAGAAGCAGCCATAAGTG
fhuA_loc_FW1	GCGGTTTTAAGCCAACCAGC
fhuA_loc_RV1	GCTGAATCTTGTTCCGCCAGAAG
gor_gene_FW1	GGTGAATGGCGACGGTATTGTC
gor_gene_RV1	GGCTATGTACGGCCAGAAATGTG
gor_loc_FW1	GCAGAAGGTTACAAACGTTTCGCC
gor_loc_RV1	GCATGGAGAAAACGGGTGATTGATAAAGC
gor_loc_RV2	CATTGCAAGAATTAGCGATTTCTTCAGC
gshA_gene_FW1	CCGCAAACGGTTCGGTATCAG
gshA_gene_RV1	GCCTGGCTGGAAAAACATCCTC

Identifier	Sequence (5'-3')
gshA_loc_FW1	CTTCCCAGGCGAGTCTGAGTATATGA
gshA_loc_FW2	GAAGCGTTATCGCTTTTCTCGTTGTT
gshA_loc_RV1	CGTTCAGGGCATGATGTGGTG
gshA_loc_RV2	ATCAGCAACACGCTGCTCATCT
lon_gene_FW1	ACGCATTGAAATCCCCGTATTGC
lon_gene_RV1	GTCACAACCTGCATACCAGACG
lon_loc_FW1	GCTCGGTACTTTAGCAGCAGC
lon_loc_RV1	GCTGTTTAATCTGGAAGGCGTGGAT
metB_loc_FW1	GCTCAATCTATACGCAAAGAAGTTTAGATGTCC
metB_loc_RV1	CTGAGAGTACTCCGCCATAATGCC
ompT_gene_FW1	GTAACCCGATTCCATGCGCC
ompT_gene_RV1	AGTCCTGACAACCCCTATTGCG
ompT_loc_FW1	GCAGGATTCGAACCTGCGG
ompT_loc_RV1	GATTAAGGGATGAAGGAACGTCATTTACTG
rne::SBP_loc_FW1	GAGACGGCAGATATTGAAGAAGTTGTCG
rne::SBP_loc_RV1	GCGAAGCGGATAAAGAAACCGC
sdaA_gene_FW1	GACATGTTTAAGGTGGGGATTGGTC
sdaA_gene_RV1	TCACACTGGACTTTGATTGCCAGAC
sdaA_loc_FW1	AATTCATGTGAATAGTTAAGCCAGTCGCC
sdaA_loc_FW2	GCTGAAGAAGAGGTCGCTATACCG
sdaA_loc_RV1	GCCTGACGCAACAGTGGAAG
sdaA_loc_RV2	ATCGCACAGTTTGGAGTTTTGCC
sdaB_gene_FW1	ATTTTCAAAATCGGCATTGGCCC
sdaB_gene_RV1	CGCAGGCAACGATCTTCATTGC
sdaB_loc_FW1	TCTGATTGCAATCTCCGCAATCTTCT
sdaB_loc_FW2	GCTATCACCAAATCCTTCCTCGGTC
sdaB_loc_RV1	ATAACCCCCTTAAAGATCAGGCGAT
sdaB_loc_RV2	CAGTAGCCTTTATCTGTCGAAACAATCG
speA_gene_FW1	GTATAACCGTACAAACCTGCCTCG
speA_gene_RV1	TCTGACGACATGTCTATGGGTTTGC
speA_loc_FW1	AAACCCTTCCGCGTCGTTAGG
speA_loc_FW2	GCTTGTGGGTTCGTCTTAACAGG
speA_loc_RV1	GGAGCGTCAGCATTCACTGC
speA_loc_RV2	GCATACGCTTACCGGCAGC
tnaA_gene_FW1	GGTGAAGTGACGCAATACTTTCGG
tnaA_gene_RV1	CTGAACCGTTCCGCATTCGTG
tnaA_loc_FW1	GATGGTGCTTGCATATATATCTGGCG
tnaA_loc_RV1	CAGGCACCGGCAAGATCAAC
tnaA_loc_RV2	CGAACTTAACCACAATACGCTGGC

3.3 Cloning Procedures

3.3.1 apiRBP-sfGFP

The uncharacterized RNA Binding Protein (RBP) from *Plasmodium vivax* SaI-1 (NCBI Reference Sequence: XM_001616500.1, residues 33-134) was cloned into a modified pET21a+ vector in-frame with a C-terminal GFP fusion protein. Cloning was performed using the Infusion cloning kit (Clontech) following the manufacturer's instructions. The expression

construct was kindly provided by Irene Diaz-Moreno and Sofía M. García-Mauriño (Universidad de Sevilla, Sevilla, Spain).

3.3.2 pEXTIR

The expression vector for *in vitro* protein synthesis pIVEX2.3d was modified to introduce a C-terminal StrepII tag to the target gene. For this purpose, two primers, namely pEXTIR_MCS_FW and pEXTIR_MCS_RV (see Table 1), were designed and synthesized including a 5' phosphate. The primers were reconstituted in TE buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 50 mM NaCl) at a concentration of 50 µM, mixed at a ratio of 1:1, heated to 70 °C, and then allowed to cool down slowly to RT to allow hybridization of the primer pair. Upon hybridization, the resulting double stranded DNA resembles the 5' phosphorylated overhang generated after digestion with *Xho*I and *Bam*HI. This DNA fragment was ligated into the *Xho*I- and *Bam*HI-digested parental pIVEX2.3d using T4 DNA Ligase (New England Biolabs GmbH, Frankfurt) according to the standard protocol. The resulting vector, namely pEXTIR, contained multiple cloning site and an in-frame C-terminal StrepII tag, instead of the original hexahistidine affinity tag. Gene expression in pEXTIR is under control of a T7 promoter.

3.3.2.1 pEXTIR-Fab-LC, pEXTIR-Fab-HC and pEXTIR-scFab

The coding sequences for the target antibody Anti-SIN1 (Fab, scFab) were optimized according to DNA2.0's (Menlo Park, CA, USA) patented GeneGPS® optimization algorithm [86], and synthesized by order of EMD Millipore (Bedford, MA, USA). The coding sequences of Fab-light chain (LC), Fab-heavy chain (HC) and scFab including their optimized 5' UTR were each integrated into pEXTIR (3.3.2) by Transfer PCR. The following primer pairs were used for integration (Table 1): (*i*) HC_FW and HC_RV to integrate Fab-HC into pEXTIR; (*ii*) LC_FW and LC_RV for integration of Fab-LC into pEXTIR; (*iii*) scFab-FW and scFab_RV to integrate scFab coding sequence into pEXTIR. The primer's 5'-end are complementary to the recipient vector pEXTIR (shown in italic, Table 1) and 3'-end complementary to the target sequence. The forward primer also indicates the 5' UTR in which the ribosomal binding site (rbs) is underlined and the start codon (ATG), located 6-7 nucleotides downstream of the rbs, is shown in bold.

The Transfer PCR was performed according to the instructions described by Erijman *et al.* [87]. The scFab contained at first the Fab-LC, followed by a linker that translated into 34 amino acids with the sequence SGGGSGGGSEGGGSEGGGSEGGGSEGGGSGGGSG and connected the Fab-LC to Fab-HC. The inserts Fab-HC, Fab-LC and scFab were integrated in

pEXTIR in frame with a 3'-end linker (5'-CTCGAGGGTAGC-3') which translates into LEGS followed by a StrepII affinity tag. The resulting expression plasmids were verified by sequencing (GATC Biotech, Regensburg) and are hereafter referred to as pEXTIR-Fab-HC, pEXTIR-Fab-LC and pEXTIR-scFab.

3.3.3 pAR1219-NH6-T7RNAP

For expression and purification of T7 RNA polymerase (T7RNAP), the commercially available plasmid pAR1219 (Sigma Aldrich, St. Louis, MO, USA) was modified to introduce N-terminal hexahistidine affinity tag using the 5' phosphorylated T7-N6HIS_FW and 5' phosphorylated T7-N6HIS_RV (Table 1). To amplify the plasmid, we used Phusion DNA Polymerase (NEB) according to standard protocols. The PCR product was purified, *Dpn*I (NEB, R0176S) digested to remove any remaining parental vector, and blunt ends of the linear PCR product were ligated using T4 DNA Ligase (NEB). The resulting plasmid, namely pAR1219-NH6-T7RNAP, was verified by DNA sequencing (GATC Biotech, Regensburg).

3.3.4 pET19b-DSBC

The gene *dsbC* (protein disulfide isomerase II, Gene ID: 947363) including the coding sequence for the signal peptide to translocate the protein product to periplasmic space, was synthesized by Thermo Fisher Scientific (Waltham, MA, USA) including 5' UTR (5'-TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC-(ATG…)-3') that contained *Xba*I recognition site (underlined). The ORF was synthesized in-frame with a 3' linker (5'-GGTGGTAGCGGT-3') that translates into GGSG followed by hexahistidine affinity tag, double stop codon (TAATAA) and *Xho*I recognition site. The *dsbC* insert and backbone of pET19b (Novagen) were ligated after digestion with *Xba*I and *Xho*I using T7 DNA Ligase (New England Biolabs GmbH, Frankfurt) according to standard protocols. The resulting vector, namely pET19b-DSBC, was verified by DNA sequencing (GATC Biotech, Regensburg).

3.3.5 pD861-CH-SKP and pD861-CH-FKPA

The chaperones Skp (periplasmic chaperone, Gene ID: 944861) and fkpA (FKBP-type peptidyl-prolyl cis-trans isomerase, Gene ID: 947850), including their signal peptide to direct expression into the periplasm, were synthesized and cloned by DNA2.0 (Menlo Park, CA, USA) into pD861-CH in-frame with a C-terminal hexahistidine affinity tag under control of an

rhamnose-inducible promoter. The plasmids are hereafter referred to as pD861-CH-SKP and pD861-CH-FKPA, respectively.

3.4 Expression and Purification Additives for CFPS reactions

3.4.1 T7RNAP

The plasmid pAR1219-NH6-T7RNAP (3.3.3) was transformed into strain BL21 Star (DE3). Expression and purification was performed according Lei Kai *et al.* [88] using HisTrap HP (GE) connected to ÄKTAexplorer 10 FPLC system. T7RNAP was eluted into T7RNAP Elution Buffer (10 mM potassium phosphate pH 8, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol). Fractions containing T7RNAP with highest purity and concentration as determined by SDS-PAGE were pooled and dialyzed against T7RNAP Dialysis Buffer (40 mM potassium phosphate pH 7.7, 10 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol).

3.4.2 DsbC, Skp and FkpA

The expression plasmids pET19b-DSBC, pD861-CH-SKP and pD861-CH-FKPA (as decribed in 3.3.4 and 3.3.5, respectively) were transformed into BL21 Star (DE3) (Thermo Fisher Scientific, Waltham, MA, USA).

Transformants were grown in LB broth with antibiotics (ampicillin 100 µg/mL or kanamycin 50 µg/mL) until $OD_{600} \sim 0.6$. Protein expression was induced by final concentration of 1 mM IPTG (pET19b-DSBC) or 4 mM rhamnose (pD861-CH-SKP, pD861-CH-FKPA), followed by incubation for additional 4 h at 37 °C and 160 rpm. Cells were pelleted by centrifugation (4424 *g* for 15 min at 4 °C) and stored at -20 °C until required.

Periplasmic proteins DsbC, Skp and fkpA were released by osmotic shock. For this purpose, the pellet of 0.5 L culture was extracted twice for 15 min on ice, first with 10 mL ice-cold Extraction Buffer (100 mM Tris-OAc, 0.5 mM EDTA, 20% (w/v) sucrose, supplemented with Protease Inhibitor Cocktail (cOmplete, EDTA-free, Roche)) and second with ice-cold ddH_2O, each time followed by centrifugation (4600 *g* for 15 min at 4 °C).

The two extracts were combined and imidazole and NaCl were added to a final concentration of 40 mM and 0.5 M, respectively. The pH of the extract was adjusted to 7.4 using NaOH and filtered using a syringe filter (PVDF, 0.45 µm pore size). The clarified extract was applied to a 1 mL Ni Sepharose High Performance column (1 mL HisTrap HP, GE Healthcare) connected to ÄKTAexplorer 10 FPLC system. The equilibration of the column, washing and elution was performed according to the column's manufacturer instruction. The elution fractions were analyzed by SDS-PAGE and fraction with highest purity and

concentration of the respective target protein were pooled and dialyzed against 2 L Dialysis Buffer (10 mM Tris-OAc, 60 mM potassium acetate) for ~16 hours at 4 °C.

The dialysate was clarified by centrifugation (17,000 *g* for 15 min at 4 °C) and directly beaded into liquid nitrogen. The purity and molecular weight distribution was analyzed by gel filtration column (TSK-GEL G3000SW$_{XL}$, TOSOH Bioscience) connected to a ÄKTAmicro chromatography system in accordance with the manufacturer's instructions as well as SDS-PAGE. The chaperones eluted as monomers as single peak in SEC and the homogeneity was estimated to be ≤ 95% by SDS-PAGE (data not shown). Protein was beaded directly into liquid nitrogen and stored at -80 °C until further use.

The concentrations of T7RNAP and chaperones were determined spectrophotometrically using extinction coefficients as determined by the primary amino acid sequence using ProtParam tool [89]. The following parameters were found: T7RNAP (MW 99.9663 kDa; ε_{280} = 141010 M^{-1} cm^{-1}), DsbC (MW 24.541 kDa; ε_{280} = 17670 M^{-1} cm^{-1}), Skp (16.5716 kDa; ε_{280} = 1490 M^{-1} cm^{-1}) and FkpA (27.1035 kDa; ε_{280} = 15930 M^{-1} cm^{-1}). The final yield was estimated spectrophotometrically based on the respective extinction coefficient to be between 3-5 mg/mL.

3.5 Genome Editing of *E. coli* A19 and D10

3.5.1 Gene knockout in *E. coli* A19 and D10

To genetically modify the source strain *E. coli* A19 we employed Quick & Easy E. coli Gene Deletion Kit (Gene Bridges GmbH, Heidelberg), which is based on Red/ET recombination as described in by Datsenko *et al.* [90]. The genes *endA*, *gshA*, *speA*, *sdaA*, *sdaB*, *tnaA*, *ompT*, *lon*, *gor*, *fhuA* were deleted. Briefly, based on the genome sequence of *E. coli* W3110 (available under identifier NC_007779.1) we designed primers with ~50 bp homologous regions upstream and downstream of the coding region (ORF) of the target gene. The 3' end of the primer was complementary to the linear selection cassette supplied by the kit. The primers were used to introduce homology arms on the linear selection cassette "FRT-PGK-gb2-neo-FRT" or "loxP-PGK-gb2-neo-loxP" (Gene Bridges, Heidelberg) using PCR. The functional cassettes are coding for a Kanamycin resistance gene flanked by site-specific recombination sites FRT and loxP, respectively.

The target strain *E. coli* A19 was, prior to recombination, transformed with the pRedET, which codes for Redα/Redβ/Gam (λ-proteins) proteins under control of an L-arabinose-inducible promoter. After pre-expression of the λ-proteins by induction with ~0.3% L-arabinose (w/v), the target strain was transformed with the purified selection cassette with homology arms

(200-600 ng) using electroporation (Gene Pulser, Bio-Rad Laboratories GmbH, operated at 1350 V and 5 ms time constant using a 0.1 cm gap-cuvette). After the recombination could proceed for 4 h at 37 °C, the positive clones were selected on LB-Agar plates containing 15 µg mL^{-1} Kanamycin.

Depending on which site-specific recombination site (FRT or loxP) was used, positive clones were transformed with the respective expression plasmid 707-FLPe or pSC101-BAD-cre (Gene Bridges, Heidelberg), respectively, to remove the Kanamycin resistance gene by recombination with Flippase or Cre recombinase. The Flippase (FLPe) was under control of a heat labile repressor cI857 (induction by temperature shift) and the Cre recombinase was under control of a L-arabinose inducible promoter. All plasmids used to genetically modify the *E. coli* A19 contained furthermore a temperature sensitive origin of replication, therefore the plasmid is abolished after extended incubation at 37 °C.

To verify the successful knockout and the subsequent removal of the selection cassette, genomic DNA (gDNA) was isolated using QIAamp DNA Mini Kit (Qiagen, Hilden) and amplified using Phusion High-Fidelity DNA Polymerase (New England Biolabs GmbH, Frankfurt). The PCR products were verified according to their expected size using Agarose gel electrophoresis as well as by sequencing (GATC Biotech, Regensburg).

3.5.2 Insertation of C-terminal affinity tags in *E. coli* A19 and D10

Affinity tags were introduced by the same protocol that was used for gene knockout (3.5.1) except for the modified selection cassettes. As affinity tags either SBP (streptavidin binding peptide) or StrepII tags were introduced on the C-terminus of RNaseE (*rne*-SBP), Release Factor I (*prfA*-StrepII) and cysteine-tRNA ligase (*cysS*-SBP∨StrepII). The SBP tag translates into 38 amino acids (MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP) and the StrepII tag translates into eight amino acids (WSHPQFEK).

The coding sequence for either SBP tag or StrepII tag was synthesized in front of the loxP site, whereas the functional elements were concatenated according to the original selection cassette (GeneBridges, Heidelberg) in the following manner: HR-SBP∨StrepII-loxP-PGK-gb2-neo-loxP-HR with HR being the homology regions. The flanking homology regions were directly synthesized on the 5' and 3' end of selection cassette to tag *rne* in *E. coli* A19 and D10 or subsequently introduced by PCR using the primers cysS_FW and cysS_RV (Table 2) to modify *cysS* in *E. coli* D10. The sequence files of the annotated selection cassettes as well as annotated modified sections of the respective *E. coli* genome (*.dna file; SnapGene software from GSL Biotech; free viewer available at snapgene.com) are referenced in Annex (0,1) and

can be used to retrace the genome editing procedure in detail. The selection cassette to modify *prfA* ("selection-cassette_prfA-SII-KAN-prmC") was kindly provided by Biotechrabbit GmbH (Berlin, Germany) and is also referenced in Annex (0,1). This cassette does not allow the recycling of the resistance gene by site-specific recombination and had to be introduced at last.

The successful introduction of tags in-frame with the respective genes (*prfA, rne, cysS*) were verified on genome-level as described in 3.5.1 (Agarose gel electrophoresis and sequencing) and by Western Blot on protein-level. For Western Blotting, the A19 wt and D10 wt as well as their derivatives were cultivated until they reached the log-phase. A sample was taken and immediately dissolved in SDS-containing sample buffer. Proteins were separated using SDS-PAGE followed by Western Blot. The correct tagging was verified by immuno-detection of the corresponding proteins using a primary antibody against either SBP-tag (α-SBP) or SII-tag (α-StrepII) followed by detection using a HRP-conjugated secondary antibody against the primary antibody.

3.5.3 Determination of Growth Rate of Genetically Modified *E. coli*

The growth rate of each genetically modified *E. coli* was determined in triplicates to ensure that no essential genes were altered and to verify the fitness of the respective strain in comparison to the original *E. coli* A19. Pre-cultures (in LB-broth (Lennox): 10 g L^{-1} peptone, 5 g L^{-1} yeast extract ,5 g L^{-1} NaCl) were inoculated directly from Glycerol Stocks the day before and incubated at 37 °C and 160 rpm in baffled shaker flasks. The next morning, 2.5 L Tunair flasks (Z710822 SIGMA) filled with 1 L 2xYTPG medium (10 g L^{-1} yeast extract, 16 g L^{-1} tryptone, 5 g L^{-1} NaCl, 100 mM glucose, 22 mM KH_2PO_4, 40 mM K_2HPO_4) were inoculated to an $OD_{600} \cong 0.05$ and incubated at 37 °C and 160 rpm. The OD_{600} was measured every 30 min, whereas the culture broth was diluted to an $OD_{600} \leq 0.4$ to avoid any artefacts of stray-light measurements. The OD_{600} were logarithmized (natural logarithm) and averaged (biological triplicates) and plotted against time [h]. A linear regression was performed and data points being not in the linear range in the beginning and the end of the growth curve (lag- and towards stationary-phase) were excluded from further analysis.

Every linear regression is based on 6 to 9 data points and the coefficient of the determination (R^2) of the linear regression was between 0.993 and 0.999 as determined by Software Package Origin (OriginLab, Northampton, MA). The slope of the regression line indicated the growth rate [$1/h^{-1}$].

3.6 Cell-Free Protein Synthesis (CFPS)

3.6.1 Preparation of $S30^{+D(1-4)}$ and $S30^{+D}$-S extract (Protocol 1) w/ dialysis

Lysates were prepared from *E. coli* strain A19. For $S30^{+D(1-4)}$ lysates, 10 L of 2xYTPG medium (10 g L^{-1} yeast extract, 16 g L^{-1} tryptone, 5 g L^{-1} NaCl) supplemented with 100 mM of glucose and potassium phosphate buffer (22 mM KH_2PO_4, 40 mM K_2HPO_4) were inoculated with a fresh pre-culture grown in LB-broth (Lennox) medium in a ratio of 1:100 and incubated in a fermenter at 37 °C with vigorous stirring at 500 rpm and good aeration [91, 92]. As soon as the cells reached mid-log growth phase (OD_{600} = 4.0-5.0), the culture broth was cooled down to 20-18 °C.

For the generation of the stress lysates ($S30^{+D}$-S), 300 mL ethanol were added at OD_{600} = 4.0 and the fermentation temperature was increased to 42 °C for 45 min before chilling to 20-18 °C. The chilled cells were harvested by centrifugation (6,000 *g* for 20 min at 4 °C) and cell pellets were washed three times with ice-cold Buffer A (10 mM Tris/OAc pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KCl, 6 mM β-mercaptoethanol). The pellets were suspended in 110% (v/w) of ice-cold Buffer B (10 mM Tris/OAc pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KCl, 1 mM DTT, 1 mM phenlymethylsulfonylfluorid (PMSF)) and disrupted by mechanical force using high-pressure homogenization. The lysate was centrifuged twice for 30 min at 30,000 *g*. The supernatant was collected and adjusted to a final concentration of 400 mM NaCl and incubated for 45 min at 42 °C with subsequent dialysis (MWCO of 12-14 kDa) against 5 L ice-cold Buffer C (10 mM Tris/OAc pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KOAc, 0.5 mM DTT (freshly prepared)). Dialysis was initially carried out for ~3 h and followed by buffer exchange (5 L Buffer C) and dialysis overnight. The lysate was finally centrifuged at 30,000 *g* for 30 min to remove precipitates. The clarified lysate was flash-frozen in liquid nitrogen and stored at -80 °C.

3.6.2 Preparation of $S30^{-D}$ extract (Protocol 2) w/o dialysis

For $S30^{-D}$ lysate preparation, a 7.5 L bioreactor (Labfors, Infors GmbH, Einsbach, Germany) filled with 2-5 L of 2xYTPG medium was inoculated to an initial OD_{600} = 0.05 with a fresh pre-culture. The bioreactor was operated at 37 °C at a stirrer speed of 500 rpm and constant gas flow of 7 vvm air to achieve sufficient aeration. When the culture reached an OD_{600} = 3-3.5 the cells were quickly cooled to 10 °C using a cooling coil and pelleted by centrifugation (6,000 *g* at 4 °C for 10 min). The pellet was washed twice with S30 Buffer (10 mM Tris/OAc pH 8.2, 14 mM $Mg(OAc)_2$, 60 mM KOAc, 2 mM DTT) followed by centrifugation (4,600 *g*,

20 min at 4 °C). After recording the wet weight, the pellet was flash frozen in liquid nitrogen and stored at -80 °C, where it was stable for several weeks.

To disrupt the cells, the pellet was thawed on ice, thoroughly suspended in 1 mL ice-cold S30 Buffer per 1 g of cell wet weight, and disrupted using high-pressure homogenization (EmulsiFlex-B15, Avestin Europe GmbH, Germany operated at 22,000 psi, 1 pass). The homogenate was centrifuged (30,000 *g* at 4 °C for 30 min) and the supernatant containing the translational active components was carefully transferred to a new centrifugation tube while avoiding any disturbance of the pellet fraction. The $S30^{-D}$ lysate was centrifuged again (30,000 *g* at 4 °C for 30 min) and further processed by the run-off reaction involving an incubation (37 °C at 160 rpm for 60 min) to dissociate endogenous mRNA and ribosomes. After this incubation, the lysate was centrifuged (30,000 *g* at 4 °C for 10 min) and the supernatant was aliquoted and flash frozen in liquid nitrogen. The frozen aliquots were stored at -80 °C until required.

3.6.3 Cell-Free Protein Synthesis Reaction (batch-mode)

Cell-free reactions contained 2 mM of each 19 proteinogenic L-amino acids excluding L-leucine, 1.2 mM L-leucine, 1.2 mM ATP, 0.86 mM of CTP/GTP/UTP, 1.5 mM Spermidine, 1 mM Putrescine, 0.27 mM Coenzyme A, 0.33 mM NAD^+, 0.07 mM folinic acid, 170 µg/mL tRNA from *E. coli* MRE600 (Roche), 4 mM sodium oxalate, 33 mM Phosphoenolpyruvate (PEP), 10 mM ammonium glutamate, template DNA (13.3 µg mL^{-1}), T7RNAP (50-100 µg mL^{-1}) and 24% (v/v) S30 extract. Each reaction contained additionally potassium glutamate and magnesium glutamate which was optimized for each batch of cell extract. For scFab and Fab expression the extract was furthermore pre-treated with various concentrations of iodoacetamide (IAM) for 30 min at RT prior to the synthesis reaction, followed by addition of oxidized and reduced gluthatione (GSSG/GSH), chaperones and 0.3 µM [^{14}C] L-leucine (300 mCi/mmol).

3.6.3.1 Optimization of potassium and magnesium glutamate

Optimization of potassium- and magnesium glutamate concentrations were performed in 50 µL-scale in 96-well microtiter plate with glass bottom sealed with breathable tape using pIVEX2.3d-SII-eYFP as template. In the expression template pIVEX2.3d-SII-eYFP the enhanced yellow fluorescent protein (eYFP) with N-terminal Strep Tag II is cloned under control of T7 promoter. The plate was incubated shaking at 30 °C in a Synergy 2 plate reader and fluorescence emission at 528 nm of newly synthesized eYFP was measured every 15 min for 3 h after excitation at 485 nm and sensitivity adjusted to 60%. Cell-free reactions to

synthesize Fab and scFab were carried out in a 15 µL-scale o/n in 1.5 mL reaction tubes under constant shaking at 600 rpm in a temperature-controlled ThermoMix (Eppendorf) set to 30 °C. The ThermoMix itself was placed in an incubator set to 35 °C to prevent condensation.

3.6.4 Cell-Free Protein Synthesis Reaction (CECF-mode)

Synthesis of proteins using cell-free lysates based on T7-polymerase transcription was carried out in a two-compartment configuration as described previously [91-93]. Protein production was conducted in analytical scale with 55 µL of reaction mixture and 825 µL of feeding mixture. Reactions were performed in customized mini dialyzer devices [93] placed into standard 24-well cell-culture plates. Reaction and feeding mixtures were separated by a dialysis membrane with a 12-14 MWCO. Reactions were allowed to proceed for 12-16 h at 30 °C at 200 rpm shaking.

3.6.5 Quantitation of Cell-Free Protein Yield (radioactive labelling)

To determine soluble and insoluble protein yield in cell-free reactions, the reaction mixture was additionally supplemented with 0.3 µM [^{14}C] L-leucine (300 mCi/mmol) as indicated before. Protein yield of TCA-precipitable was quantified after incorporation of [^{14}C] L-leucine into target protein using scintillation counting and sample was furthermore analyzed by autoradiography after non-reducing SDS-PAGE according to the method previously described by Cai *et al*. [94]. Briefly, after termination of the cell-free reaction 4 µL of the 15 µL reaction mixture was removed and directly spotted on a glass microfiber (Grade GF/C, Sigma Aldrich), hereafter referred to as C_{total} and dried on a hot plate for 2 min at 100 °C. The remaining 11 µL were centrifuged (6,000 *g*, 15 min at 4 °C) and 4 µL of the supernatant was spotted on a glass microfiber ($C_{soluble}$). The glass microfiber containing the $C_{soluble}$ were washed three times for 15 min with 5% (w/v) TCA on ice to remove any unincorporated leucine, rinsed with absolute ethanol and dried on a hot plate. The glass fibers containing the sample C_{total} and $C_{soluble}$ were each placed in a scintillation vial, coated with Quicksafe A scintillate (Zinsser Analytic GmbH, Frankfurt, Germany) and counted using a Hidex 300 SL scintillation counter. The protein concentration of soluble protein was determined using the following equation:

$$Yield_{soluble} = [leucine]_{total} \times \frac{MW}{[\#\ leucine]} \times \frac{C_{soluble}}{C_{total}} \quad (1)$$

where $Yield_{soluble}$ is the protein centration in mg/mL, $[leucine]_{total}$ the concentration of labeled und unlabeled L-leucine in the reaction in mM (1.203 mM), MW is the molecular weight in Da and [# leucine] the number of occurrences of L-leucine in the amino acid sequence of the target

protein. The fraction $C_{soluble}/C_{total}$ is the ration of counts per minute (cpm) of soluble fraction over full reaction mixture as measured by scintillation counting.

The assembly of the heavy and light chain was also analyzed by reducing and non-reducing SDS-PAGE using 4-12% Bis-Tris gels (Lifesciences Technologies, Darmstadt) followed by a Western Blot onto nitrocellulose membrane and exposure o/n on phosphor screens. The phosphor screens were scanned using FLA-5100 scanner (Fujifilm) and analyzed using ImageQuant software (GE). The ratio of assembled Fab and scFab complex over non-assembled Fab and scFab, respectively, was calculated according to equation:

$$Ratio_{\frac{assembled}{non-assembled}} = I_{non-reducing}^{assembled} / I_{non-reducing}^{non-assembled} \quad (2)$$

with $I_{non-reducing}^{assembled}$ over $I_{non-reducing}^{non-assembled}$ are the intensities of the band of correctly assembled complex in the reducing SDS-PAGE over the sum of non-assembled protein in the non-reducing SDS-PAGE as determined by ImageQuant (GE).

3.6.6 sGFP as quantitative reporter of fusion proteins

Shifted GFP (sGFP) was used as a quantitative reporter to evaluate lysate performance and to quantitate the expression-level of sGFP-fusion proteins. The concentration of reporter protein in soluble and insoluble fractions was analyzed as described previously [95, 96]. After completed protein production, reaction mixtures were centrifuged (18,000 x *g* at 4 °C for 10 min) to separate the soluble from the insoluble fraction. Following separation, samples were stored on ice for 1 h to allow complete maturation of sGFP. Pellets of reaction mixtures were resuspended in buffer C equal to the reaction mix volume. 297 µL of assay buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl) in black 96 well microtiter plates (96F Nunclon Delta Black Microwell SI, ON 137101 (Nunc, Langenselbold, Germany)) was mixed with 3 µL of either the soluble fractions or 3 µL of the resuspended pellet fractions.

The GFP fluorescence was measured utilizing a TECAN Magellan plate reader (Tecan, Maennedorf/Zuerich, Switzerland). Measurement parameters for excitation, emission and reads per well were set to 485 nm, 510 nm and 10, respectively. GFP concentration was then determined using a reference curve generated with purified sGFP of known concentration [95, 96]. The soluble fraction [%] of sGFP fusion protein was defined as the ratio of $GFP_{soluble}$ over $GFP_{insoluble}$.

3.6.7 HsGNA1 Activity Assay

The activity assay was performed as described before [97, 98]. In brief, after reaction was completed, the mini-dialyzers were taken out of the cell-culture plates, the reaction mix was diluted inside the dialyzer with 55 µL of Buffer E (50 mM Tris-HCl, pH 8.0) and then dialyzed against 2 L of Buffer E at 4 °C for 2 h. The soluble and the insoluble fractions were then separated by centrifugation (18,000 *g* for 10 min at 4 °C) and the GFP concentration was determined as described above.

The activity assay was performed in transparent 96-well flat bottom plates. 50 µL of reaction buffer (50 mM Tris-HCl, pH 8.0, 500 µM D-glucosamine 6-phosphate (GlcN6P), 500 mM acetyl-CoA, 5 mM $MgCl_2$, 10% glycerol) was mixed with a sample aliquot containing 0.4 µg GNA1-sfGFP. The reaction was incubated for 5 min at 30 °C on a heated plate shaker and stopped by the addition of 50 µL stop solution (50 mM Tris-HCl, pH 8.0, 6.4 M guanidine hydrochloride). The stop reaction was allowed to proceed for 5 min at RT. To detect the amount of produced CoA, 50 µL of color reaction buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 200 µM 5,5'-dithiobis(2-nitrobenzoic acid (DTNB)) was added and 4-nitrophenolate formation was measured at 412 nm in a microplate reader (Fisher Scientific, Schwerte, Germany). For background subtraction, equal volumes of reaction mixture containing no DNA template were used.

3.6.8 Functional Assay by Western Blotting

Recombinant antigen MAPKAP1 (Mitogen-activated protein kinase 2-associated protein, Sin1) derived from *E. coli* was purchased (MyBiosource, MBS1364264). As a positive control, mouse monoclonal Anti-MAPKAP1 (Merck Millipore, 05-1044) was used. MAPKAP1 antigen (0.25 µg) was separated using 4-12% Bis-Tris SDS-PAGE, transferred onto a nitrocellulose membrane and detected using either positive control or cell free produced antibodies.

Positive control Anti-MAPKAP-1 was used according to the manufacturer's instructions (dilution 1:500,000) and detected by Peroxidase AffiniPure Goat Anti-Mouse IgG (Jackson Immuno Research, 115-035-068) as a secondary antibody. The cell-free produced Fab (cf-Fab) and scFab (cf-scFab) were used at dilution of 1:33,000 (as cf-Fab/cf-scFab containing extracts) and detected using mouse Anti Strep Tag II Antibody (Qiagen, 34850). Anti-Strep Tag II was then detected by Peroxidase AffiniPure Goat Anti-Mouse IgG. Binding of all antibodies was done in 5% (w/v) milk-TBSt buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.05% (w/v) Tween 20), intercalated by washes in TBSt buffer (3 times, 10 min each).

Detection of the Peroxidase labeled antibody was performed using SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, 34075) in a LAS-1000 Image Analyzer (Fujifilm).

3.7 Mass Spectrometry Analysis (Proteomics)

3.7.1 Preparation of protein samples for GeLC-MS/MS

The protein concentration of S30 lysates was determined by Bradford assay to be within the range of 30-35 mg/mL. Lysate samples were diluted to a concentration of 5 mg/mL. A 5-fold excess of ice-cold acetone was added to 400 µg of protein (in a volume of 40 µL) and proteins were precipitated o/n at -20°C. Precipitated protein was pelleted by centrifugation (20,000 x *g* at 4°C for 15 min) and washed once with 80% acetone. After centrifugation (20,000 x *g* at 4°C for 5 min), the supernatant was discarded, and the pellet was air-dried for 5 min at RT to remove residual acetone.

For qualitative LC-MS analysis the protein pellet (<400 µg) was directly dissolved in 25 mM sodium-bicarbonate containing 4 M urea and the protein concentration re-determined by Bradford assay using BSA in 25 mM sodium-bicarbonate and 4 M urea as reference.

For quantitative LC-MS analysis, the protein pellet (<400 µg) was dissolved in 30 µL of lysis buffer (guanidine-HCl containing buffer as provided by Serva ICPL Kit) and the protein concentration was once again determined by Bradford assay using BSA in lysis buffer as a reference. Accordingly, 100 µg were transferred into a new reaction tube and lysis buffer was added to a final volume of 20 µL. Then, cysteine residues were reduced and alkylated followed by the labelling reaction with ICPL according to the manufacturer's instructions (Serva ICPL Kit). The labelled and combined protein samples were precipitated as described previously using 5-fold excess of acetone. The resulting protein pellet was dissolved in 25 mM sodium-bicarbonate containing 4 M urea and the protein concentration was determined by Bradford assay.

3.7.2 GeLC-MS/MS Analysis (Separation and Fractionation by SDS-PAGE)

To reduce the complexity of the sample for qualitative and quantitative analysis, 50 µg of protein (dissolved in 25 mM sodium-bicarbonate in 4 M urea) were mixed with Laemmli buffer, heated to 40 °C for 5 min and separated by 10% SDS-PAGE (migration distance ~5 cm). Each lane (Coomassie stained) was cut into twelve equal fractions and every fraction was further cut into ~1 mm^3 pieces in preparation for in-gel digestion by trypsin (Pierce trypsin protease, 90057). Destaining and washing of gel pieces and digestion was performed following

instructions of standard protocols for in-gel digestion (adapted from [99]). The gel pieces were covered in trypsin solution (25 ng μl^{-1}) in 25 mM sodium-bicarbonate. After digestion for 24 h at 37 °C the reaction was terminated by the addition of 5 µL of 10% trifluoroacetic acid (TFA) and peptides were extracted according to standard protocols. Desalting of peptides was performed using C_{18} Ziptips (Millipore Ziptips, Z720046) and peptides were subsequently eluted in 3 µL 0.1% TFA/50% acetonitrile (ACN). The solvent was removed using a vacuum concentrator and the pellet was stored at -20 °C.

3.7.3 GeLC-MS/MS Analysis (LC separation and spotting)

For LC separation, the pellets after in-gel digestion were dissolved in 20 µL 0.1% TFA/5% ACN and separated using reversed-phase Nano Liquid Chromatography (UltiMate 3000 nanoflow LC system, Dionex) equipped with pre-column (Dionex, Acclaim PepMap100, C18, 5 µm, 100 Å, 300 µm i.d. x 5 mm) in line with analytical column (Dionex, Acclaim PepMap100, C18, 5 µm, 100 Å, 75 µm i.d. x 25 cm) and Proteineer fc fraction collector (Bruker Daltonics, Germany). Peptides were initially desalted at a flow rate of 20 µL/min on the pre-column at 40°C, followed by separation at 300 nL/min and 40°C on the analytical column. The elution was conducted using linear gradient from 3-40% ACN in 0.1% TFA for 1.5 h. Column elution was mixed with α-cyano-4-hydroxycinnamic acid (HCCA) matrix solution (93.5% (v/v) 0.1% TFA/95% ACN; 4.5% saturated HCCA solution (HCCA in 0.1% TFA/90% ACN); 1% (v/v) 10% TFA; 1% (v/v) 100 mM $NH_4H_2PO_4$) prior to depositing at a 10 sec interval directly on MTP AnchorChipTM 1536 TF between 18 and 93 min.

Carryover between LC runs was minimized by running an idle circuit between runs. The calibrant in HCCA matrix solution (Peptide Calibration Standard II, Bruker Daltonics) was manually spotted afterwards.

3.7.4 GeLC-MS/MS Analysis (Protein Identification by MS/MS)

The monoisotopic mass of peptides was determined using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) (Bruker Ultraflex II TOF/TOF 200, Bruker Daltonics) using reflector mode in the mass range from 700-5,000 Da. Spectra were processed by FlexControl, FlexAnalysis 3.3 and WARP-LC software (Bruker Daltonics). Maximal 30 MS^2 measurements per spot were acquired using LIFT mode based on a mass list generated from a selection of most intense peaks (signal/noise ratio > 20) by WARP-LC after MS^1 survey scans were completed.

Raw MS^1 spectra were re-processed as profile and MS^2 as centroid using Mascot Distiller version 2.6 (Matrix Science, UK) using optimized settings (correlation threshold 0.6;

minimum signal to noise ratio 2). The Mascot Distiller *.opt file containing all the parameters for peak processing is available upon request. MS^2 spectra were submitted to an in-house Mascot Server version 2.5.1 (Matrix Science, UK) and searched against reference proteome of *E. coli* MG1655 (Uniprot Proteome ID: UP000000625, 4316 sequences, downloaded on 9th of May, 2016), FPlasmid (Uniprot Accession ID: AP001918, 106 sequences, compiled on 13th of September, 2015) and contaminants (MPI Martinsried, 247 sequences of common contaminants, downloaded on 2nd of February, 2016). The following search parameters were used for qualitative analysis: Allowing up to one missed cleavage, tolerance of 50 ppm for peptides and 0.5 Da for MS/MS, MALDI-TOF/TOF, enzyme: Trypsin, variable modifications: Oxidation (M), fixed modifications: Carbamidomethyl (C).

For quantitative analysis after ICPL labelling, the same search parameters were used with the additional parameter "ICPL duplex pre-digest [MD]". For quantitative analysis, the median peptide ratio was determined using Mascot Distiller's Quantitation Toolbox and the individual datasets were normalized to median= 0 in $\log_2$ space. The single peptide ratios were exported considering quality parameters (cutoff: Fraction 0.7; Correlation 0.9; Peptide Standard Error 0.1; Shared peptides were excluded).

The individual datasets were further evaluated by InfernoRDN (Pan-omics software) using box- and correlations plots to identify problematic or outlier datasets. Positively evaluated datasets were combined and significantly regulated peptides were identified using permutation test implemented in "Quantitative proteomics p-value calculator (QPPC)" as described in detail elsewhere [100, 101]. The following parameters were used: Stage 1 (1,000 permutations; Outlier removal activated, threshold 30; p-value correction: FDR (Benjamini-Yekutieli)); Stage 2 (p-value cut-off 0.05; Fold-change cut-off 1.5). Only proteins regulated under both criteria (p-value cut-off 0.05 and fold-change cut-off 1.5) were considered significantly regulated.

3.7.5 Experimental Design and Statistical Rationale

Qualitative MS analysis served solely to identify most prevalent proteins present in S30 lysates. For this purpose, we prepared two biological replicates (S30-1 and S30-2) according to standard protocols commonly used in the field of cell-free expression [102, 103]. Even though they are based on slightly different preparation procedures, they perform similarly in protein production reactions (3.5-4.5 mg/mL yield) and therefore provide information about the core proteins present in S30 lysates commonly used in the field of cell-free synthetic biology. The FDR (peptide-level) was estimated by searching against decoy database. Based on the number of false-positive matches in the decoy database, we determined the discrete distribution of false-positive matches within the database entries using Poisson distribution.

Considering the FDR (peptide-level) in our experiments, we found that only one entry out of 4,669 will be assigned to false-positive matches. Based on that estimation, we chose our cut-off for protein identification at two significant matches. Quantitative MS analysis was carried out to identify and quantitate up- and down regulated proteins relative to standard lysate (S30-1). The standard lysate (S30-1) and three biological replicates of heat shock lysates (S30-S), namely S30-S-1, S30-S-2 and S30-S-3, were prepared according to the same lysate preparation protocols. As ratios did not follow a normal distribution, we used a permutation test [100, 101] that is distribution independent to asses significantly changed peptide ratios.

3.8 Transcriptional Profiling (A19 wt vs. genetically modified)

3.8.1 Sample Preparation for Transcriptional Profiling

The transcriptional profile of *E. coli* A19 vs. A19 Δ(*endA/ompT/lon/gor/fhuA*) CSRW during log-phase at the time point of harvest for extract preparation was investigated. Each of the two strains were cultivated in triplicates. At first, a pre-culture in LB-broth (Lennox) was inoculated directly from Glycerol Stock and cultivated o/n at 37 °C and 160 rpm. The next day, 1 L of 2xYTPG medium in 2.5 L Tunair flasks was inoculated from pre-culture to an OD_{600}=0.05 and cultivated at 37 °C and 160 rpm until OD_{600}=3.0-3.3 at which point 1 mL of culture-broth was taken and directly beaded into liquid nitrogen. The beads were stored at -80 °C until RNA extraction was performed.

To extract RNA from the beads, the pellets were transferred frozen into appropriate Teflon-vials containing a lead-ball (pre-cooled by liquid nitrogen) and cell disruption was executed at a frequency of 30 sec^{-1} using a Retsch-mill. The frozen powder was subsequently transferred into disruption buffer as provided by RNeasy Mini Kit (Cat# 74104) and stored on ice. The RNA-extraction was performed using RNeasy Mini Kit according to manufacturers' instruction and RNA eluate was stored at -80 °C until further analysis.

3.8.2 Library Preparation for Transcriptional Profiling

The library preparation was performed by the Functional Genomics Group under supervision of Dr. Christian Grumaz (Functional Genomics, Fraunhofer IGB, Stuttgart, Germany). Before library preparation, the integrity of isolated RNA (3.8.1) was verified using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The RIN (RNA Integrity Number) was for each sample ≥ 9.20 as determined by Bioanalyzer. The library preparation was performed with TruSeq Standard Total RNA Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufactures' instructions.

3.9 Amino Acid Profiling

3.9.1 Sample Preparation for Amino Acid Profiling

Quadruplicates of $S30^{-D}$ extracts of each of the genetically modified *E. coli* A19 (A19 Δ(*endA*) and A19 rne::SBP Δ(*endA*) CSRW) were prepared according to Protocol 2 w/o dialysis (as described in detail in 3.6.2). CFPS reactions were performed in batch mode containing 2 mM of each 20 proteinogenic L-amino acids (except glutamic acid which is present at 80 mM due to its presence in the CFPS reaction mixture), 1.2 mM ATP, 0.86 mM of CTP/GTP/UTP, 1.5 mM Spermidine, 1 mM Putrescine, 0.27 mM Coenzyme A, 0.33 mM NAD^+, 0.07 mM folinic acid, 170 µg/mL tRNA from *E. coli* MRE600 (Roche), 2.7 mM sodium oxalate, 33 mM Phosphoenolpyruvate (PEP), 10 mM ammonium glutamate, 16 mM magnesium glutamate and 170 mM potassium glutamate, template DNA of pIVEX2.3d-SII-eYFP (13.3 µg mL^{-1}), T7RNAP (approximately 100 µg mL^{-1}) and 24% (v/v) $S30^{-D}$ extract. The total volume of each reaction was 220 µL.

In addition, a negative control containing all the components but with the difference that a heat-inactivated $S30^{-D}$ lysate derived of *E. coli* A19 Δ(*endA*) (5 min at 95 °C) was used where no enzymatic conversion of amino acids can occur. The reactions were set up on ice and initial samples corresponding to time point t=0 min (15 µL) were taken and stored on ice. The remainder was incubated at 37 °C and 1000 rpm using an Eppendorf incubator. Every 15 min (t=15, 30, 45, 60, 75, 90, 105, 120, 135) 15 µL of sample were taken and mixed with the same volume of 5% TCA to terminate any enzymatic conversion and to precipitate proteins. The precipitate was pelleted (17,000 *g* for 15 min at 4 °C) and the supernatant was transferred to a new reaction vial and stored at -20 °C until further analysis.

3.9.2 Quantification of Proteinogenic Amino Acids using HPLC-UV

The quantification of amino acids was performed in the department Central Analytics at Fraunhofer IGB using an automated pre-column derivatization with *ortho*-phthalaldehyde (OPA). As internal standard (ISTD) D-norvaline was used at a final concentration of 515.7 pmol $µL^{-1}$ in samples and calibration standards. To quantify amino acids, three amino acid solutions of known concentration (final concentrations after adding the ISTD: 900 pmol $µL^{-1}$; 225 pmol $µL^{-1}$; 90 pmol $µL^{-1}$) were used to generate a calibration curve. Briefly, the amino acids sample as described in 3.9.1 was diluted 1 to 10 (10 µL sample; 10 µL ISTD-solution; 80 µL 0.1 N HCl) and derivatized with *ortho*-phthalaldehyde and 3-mercaptopropionic acid on pre-column using an Agilent Technologies 1290 Infinity UHPLC-System.

The sampler program to derivatize amino acids consisted of the following steps: (1) Draw 1.3 µL borate buffer (0.4 N borate in ddH_2O pH 10.2); (2) Draw 0.5 µL sample; (3) mix (five times); (4) Draw 0.5 µL OPA (*ortho*-phthalaldehyde and 3-mercaptopropionic acid in borate buffer); (5) mix (five times); (6) Draw 0.2 µL ddH_2O; (7) mix (five times); (8) Draw 7.4 µL diluent (mobile phase A: 10 mM Na_2HPO_4,10 mM $Na_2B_4O_7$, 8 mg mL^{-1} NaN_3 pH 8.2 (adjusted with 3 mL of 30% HCl)); (9) mix (five times); (10) inject 9.9 µL sample.

The sample was separated at 40 °C on a reversed phase Zorbax Exclipse RRHD Plus-C18 (3 x 100 mm, 1.8 µm) column using a flow rate of 0.86 mL min^{-1}. A linear gradient over 13 min from 2% to 57% mobile phase B (ACN:MeOH:H_2O in a ratio of 45:45:10 (v/v)) was used to elute derivatized amino acids. The elution was followed by UV absorbance at 338 nm (band width 10 nm). The detection limit was around 10-20 pmol $µL^{-1}$ for all the amino acids and the amino acid proline was not acquired in this analysis.

4 Results

This thesis aims to provide a general framework to improve CFPS based on proteomic analysis of S30 extracts. For this purpose, two commonly used extract preparation protocols were employed, S30 extracts suitable for CFPS were prepared, and expression performance was verified. Using a GeLC-MS/MS approach, S30 extracts were analyzed to identify proteins and determine their approximate abundances. Subsequently, the proteomics data was used to optimize CFPS reactions.

Three examples of this optimization and its application on the production of difficult-to-express proteins are presented in this work: (*i*) induction of a chaperone network by heat shock during cultivation to improve solubility and enzymatic activity of target proteins, (*ii*) genetic and metabolic engineering of *E. coli* A19 to stabilize substrates and target proteins, and (*iii*) purification and addition of limiting chaperones to facilitate disulfide bond formation and enhance solubility of antibody fragments.

4.1 Protein Profiling of S30 Extracts by Mass Spectrometry

4.1.1 S30 Extract Preparation for Protein Profiling

Despite the existence of defined expression environments consisting of a highly purified and reconstituted protein synthesis system (protein synthesis using recombinant elements, "PURE"), the S30 extract based on *E. coli* is - mainly due to its productivity and simplicity - the most commonly used system for CFPS. Essential features of almost all preparation procedures are cell harvesting at mid-log growth phase, centrifugation of the lysate at 12,000-30,000 *g* and a *run-off* step, which includes incubation at 37 °C to 42 °C to dissociate endogenous mRNA and ribosomes. The centrifugation and *run-off* results in fractionation and heavy precipitation of various proteins, and thus the original *E. coli* proteome as expressed during exponential growth is significantly altered, resulting in a yet unknown proteome. While extract preparation protocols used by different groups often employ modifications of individual fermentation or processing parameters, the resulting S30 extracts are similarly productive in terms of yield.

Thus, to identify the core proteome and provide an overview of the S30 extract proteome, the S30 extract proteome was analyzed qualitatively, by using samples prepared according to two different, but frequently used protocols, named $S30^{-D}$ and $S30^{+D}$, depending on whether a dyalisis step was performed or not ("-D" for no dialysis and "+D" for dialysis) (see Figure 4, A). Quadruplicates (1-4) of $S30^{+D}$ extract were were kindly provided by Drs. Erik Henrich and Frank Bernhard (Johann Wolfgang Goethe University, Frankfurt), prepared

according to a routine protocol employed for the cell-free production of membrane proteins [91]. In contrast, the S30^{-D} was only analyzed as singlet to verify that no major proteome changes occur by slightly different extract preparation.

Despite using the same *E. coli* source strain (*E. coli* A19) for both protocols, the strain grew differently in the cultivation setting (different bioreactor; growth rate: $\mu_{S30(+D)} \approx 1.9\ h^{-1}$ versus $\mu_{S30(-D)} \approx 1.5\ h^{-1}$, which corresponds to $t_d \approx 22$ and 28 min, respectively) (see Figure 4, B). Further variations between protocols were (*i*) less dense inoculation in S30^{D+}, requiring more cell divisions to reach the desired OD_{600}; (*ii*) rapid cooling of the final broth to 10 °C using a cooling coil before harvesting in S30^{-D} versus cooling to 20-18 °C in S30^{+D}); (*iii*) cell rupture with different mechanical devices, and (*iv*) *run-off* step performed at 37 °C for 80 min for D30^{-D} versus 42 °C for 45 min for S30^{+D} [103].

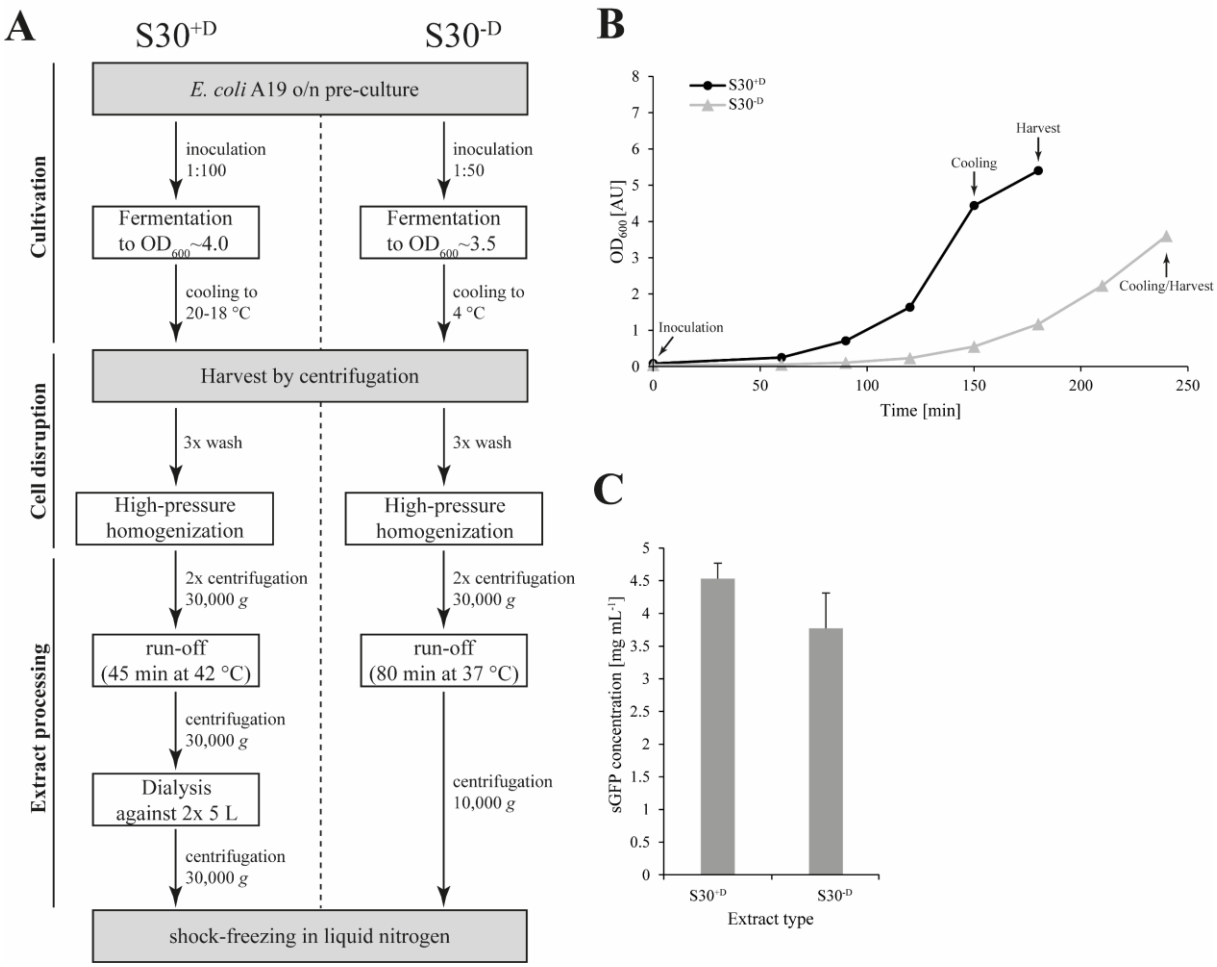


Figure 4: ***General parameters of S30 extracts used for protein profiling. Workflow of cultivation conditions, S30 extract preparation and extract performance is illustrated****. (A) Flowchart of different S30 (S30^{+D} and S30^{-D}) extract preparation procedures (as described in more detail in 3.6.1 and 3.6.2). (B) Growth curves of E. coli A19 under different fermentation conditions used for S30^{+D} and S30^{-D} extract. (C) Performance evaluation of S30^{+D} and S30^{-D} extract under standard expression conditions using sGFP as a quantitative reporter in CECF mode as described in (3.6.4). The error bars represent the standard deviation of three measurements.*

The protein production efficiency of the two lysates was evaluated in continuous exchange cell free (CECF) mode with the reporter protein sGFP (see Figure 4, C). Both lysates

performed similarly (between 3.5 to 4.5 mg mL^{-1} sGFP), with $S30^{-D}$ being slightly less productive.

4.1.2 Proteome Composition of *E. coli* A19 S30 Extracts

The S30 lysates were separated by SDS-PAGE, and subsequently each gel lane was divided into twelve slices (fractions F1-F12). Following in-gel digestion, peptides were extracted from individual gel slices (F1-F12) and analyzed by LC-MS/MS to identify proteins using bottom-up proteomics (Figure 5). The MS^2 data were searched against total *E. coli* MG1655 / K12 proteome ([104] reference proteome Uniprot ID: UP000000625 as downloaded on 19th of January 2017), which is the closest available standard to the archetypal *E. coli* K-12 strain. The database includes 4,315 predicted protein sequences assumed to be expressed based on whole genome sequencing/annotation [105]. Searching our data against the larger Swissprot database (taxonomy *Escherichia coli*, 23,014 protein sequences as downloaded on 19th of January 2017) did not result in more protein hits (data not shown), suggesting that the smaller *E. coli* reference proteome (Uniprot ID: UP000000625; 4,315 sequences) is a representative reference for the S30 extract proteome.

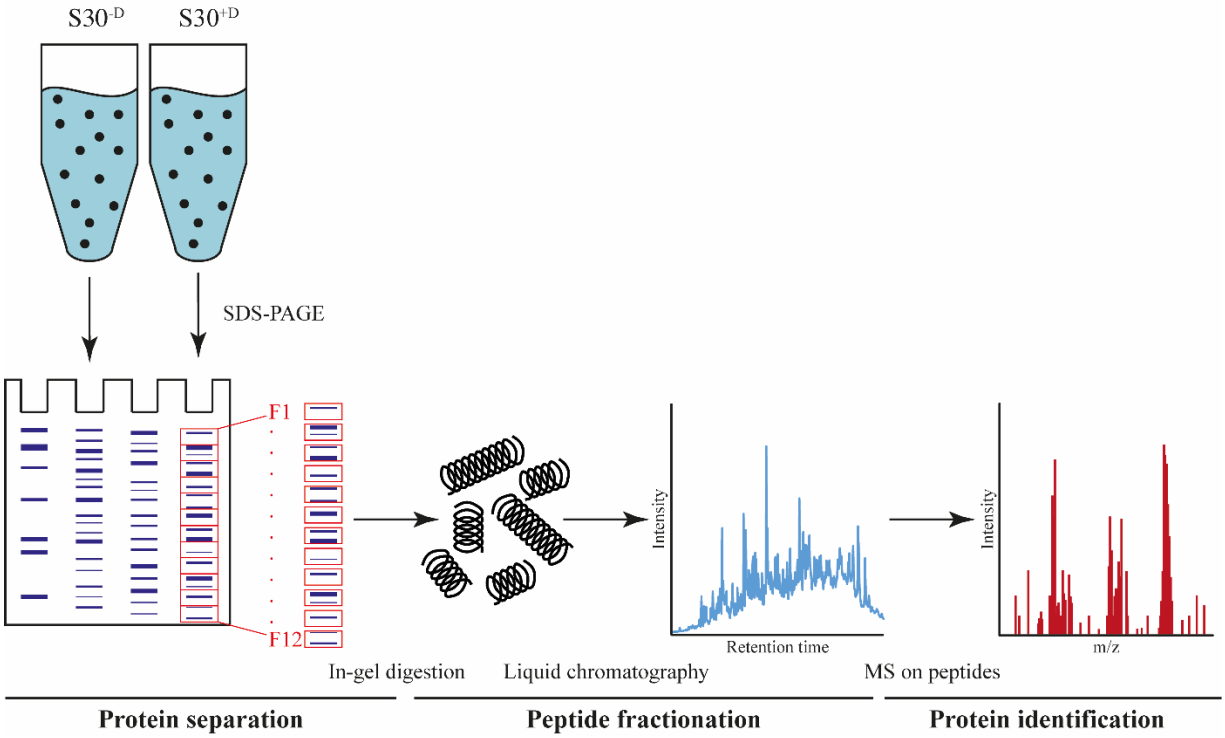


*Figure 5: **Strategy for GeLC-MS/MS analysis of $S30^{-D}$ and $S30^{+D}$ extract**. The $S30^{-D/+D}$ extracts were initially separated on protein-level by 1D-SDS-PAGE. The lane was then divided into 12 equal fractions (F1-F12) and proteins were in-gel digested (trypsin) followed by extraction. Peptides extracted from F1-F12 were further separated by nanoHPLC (Reversed Phase C18 column) and individual elution fractions (450 per gel-slice (F1-F12) were analyzed by MALDI-TOF-MS/MS to identify the $S30^{-D/+D}$ extract proteome.*

As mentioned before, the $S30^{+D}$, being the primary focus of this proteome analysis, was analyzed as quadruplicates (1-4) and the $S30^{-D}$ only as singlet, mainly to ensure that this common modification of the extract processing did not alter the proteome to a large extent.

Using the reference database (Uniprot ID: UP000000625), the proteomes of the four lysates $S30^{+D(1-4)}$ and $S30^{-D}$ were analyzed and statistical means were applied to reach an estimated false discovery rate (FDR) on peptide- and protein-level of ≤1% for each individual proteome. For the $S30^{+D(1-4)}$ extracts, 1074 unique proteins were detected (individual extract proteomes composed of 820, 902, 952 and 805 proteins, respectively), whereas in the $S30^{-D}$ extract 1056 proteins were identified. Considering the FDR on protein level, among the individual proteomes, 1 out of 100 proteins (or estimated 10 proteins in the largest proteome $S30^{-D}$) may be false positive hits, and any false positive hit, if present, is likely to be among proteins that were identified with only one or two peptides.

Next, the identified proteins in $S30^{+D(1-4)}$ were compared to $S30^{-D}$ using a Venn Diagram to identify common and unique proteins (Figure 6). The majority of proteins (865 proteins, corresponding to 68%) were common to both S30 extracts, while 209 (16%) and 191 (15%) were uniquely identified in $S30^{+D(1-4)}$ and $S30^{-D}$, respectively.

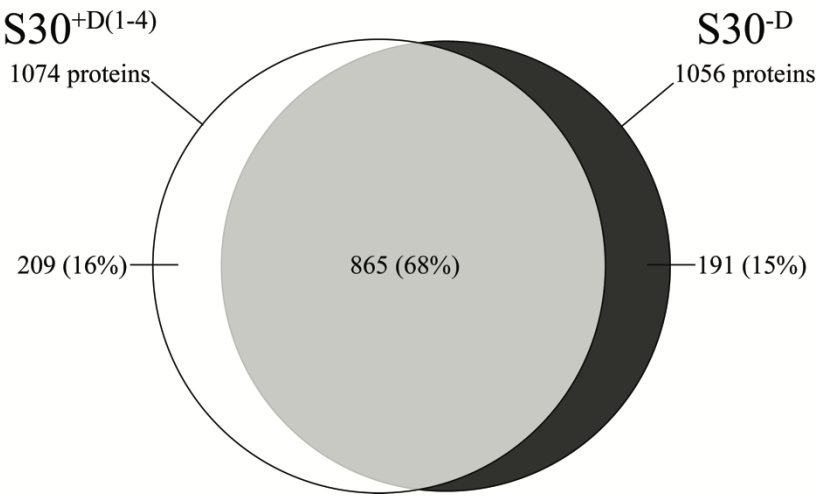


*Figure 6: **Venn Diagram to compare proteins identified in $S30^{+D(1-4)}$ and $S30^{-D}$**. The $S30^{+D(1-4)}$ with 1074 proteins is represented by the white circle and $S30^{-D}$ with 1056 proteins by the black circle. The grey area represents common proteins in $S30^{+D(1-4)}$ and $S30^{-D}$.*

Since the experimental design does not allow the distinction between uniquely identified proteins due to differences in the extract preparation and due to technical/biological variability, it is presumed that the commonly used extract preparations result into a core proteome of 865 proteins, which are mainly responsible for transcriptional/translational and other enzymatic activities in the S30 extract. However, at least 1265 (209+865+191) proteins

can be found in the S30 extract derived from *E. coli* A19 depending on preparation procedure and/or biological/technical variability. The 1265 proteins identified in S30 extracts correspond to 29.3% of the reference *E. coli* proteome MG1665 (Uniprot ID: UP000000625, 4315 predicted sequences). Based on the large number of common proteins (approx. 70%) in these two extract variants, it can be assumed that the slight modification of extract preparation results only in minor variations of the S30 extract proteome. Additionally, integration of the uniquely identified proteins of neither $S30^{+D(1-4)}$ nor $S30^{+D}$ was successful in finding functionally related gene groups or other biological patterns that might indicate biological variations (*e.g.* due to the different cultivation conditions resulting in certain pathways present in one or the other extract) (data not shown). Thus, the subsequent analysis focuses on the $S30^{+D(1-4)}$ extracts, which are representative of similar S30 extracts.

Focusing on the 1074 unique proteins identified in $S30^{+D(1-4)}$, a proteome coverage of at least approximately 25% of the theoretically predicted proteins present in the *E. coli* MG1655 reference proteome was reached. Out of 1074 detected proteins, 648 (60.3%) were common to all $S30^{+D}$ extracts, 173 (16.1%) proteins were found in three $S30^{+D}$ extracts, and 23.6% (252 proteins) were shared among two extracts or uniquely identified in one extract (Figure 7, A). Based on these data, 821 proteins (76.4%) that were identified in at least three out of four $S30^{+D}$ extracts were defined as the core S30 extract proteome, having only low variability from extract to extract. The exponentially modified protein abundance index (emPAI) of proteins identified in the core proteome were averaged to provide an approximate and relative estimation of protein abundances based on sequenced peptides per protein [106].

The Cytoscape plugin ClueGO (Marker list: *Escherichia coli* [562, 511145]; GO-source: Cellular Component as downloaded on 17th of February 2017) was used to assign Gene IDs of the identified proteins to suitable GO-terms that represent the major bacterial compartments: cytoplasm, periplasm and membrane. A subset of 2,271 proteins of the total *E. coli* MG1655 proteome (Uniprot ID: UP000000625; 4,315 sequences) is represented by the GO-term “cell part”, defined as “any constituent part of a cell, the basic structural and functional unit of all organisms”: (GO-ID 44464). As evidence for the assignment to a GO-term, only assignments based experimental data were accepted. The evidence code IEA (Inferred from Electronic Annotation) is error prone and was therefore excluded. Based on availability of experimental data, the GO-term “cell part” and its child-terms represent currently ~53% of the predicted proteins of the reference proteome.

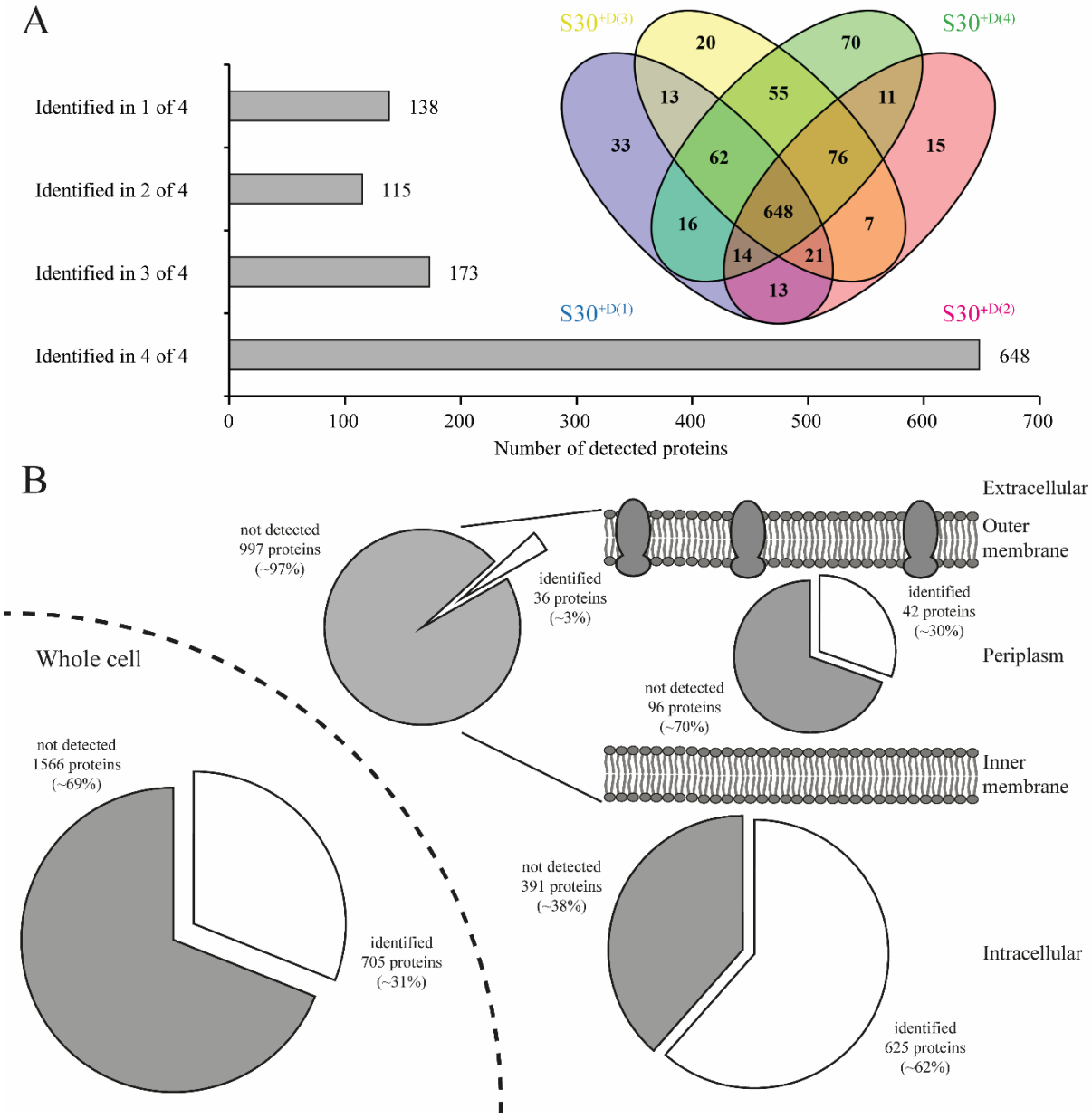


*Figure 7: **Analysis of S30$^{+D(1-4)}$ extract proteome.** (A) Number of common and unique protein identifiers in S30$^{+D(1-4)}$ illustrated by a bar as well as a Venn Diagram (821 proteins were detected in at least 3 of 4 S30 lysates). The colors of the Venn Diagram represent the four biological replicates. (B) Combined number of proteins identified in S30 extract as compared to all experimentally assigned proteins to the superordinated GO-term "Cell Part" (GO-ID 44464). The subcellular distribution of the identified S30 lysate proteins with respect to their compartmental localization in E. coli based on GO terms "Cytoplasmic Part" (GO-ID 0044444), "Periplasmic Space" (GO-ID 0042597) and "Cell Periphery" (GO-ID 0071944). Cytoplasmic Part and Cell Periphery are significantly over and underrepresented ($p \leq 0.05$) as compared to reference set of predicted proteins (Uniprot ID: UP000000625, 4315 proteins).*

4.1.3 Classification of the *E. coli* A19 S30 Extract Proteome

From the 821 proteins of the S30 core proteome, 705 proteins (~86%) were classified into the term "cell part", comprising 31% of all proteins within this group (Figure 7, B). The major subset of 685 proteins (~97.2%) is represented by the three child terms "Cytoplasmic Part" [GO-ID 0044444], "Periplasmic Space" [GO-ID 0042597] and "Cell Periphery" (= membrane) [GO-ID 0071944] [107]. Within those groups, the GO-term "Cytoplasmic Part" is

two-fold overrepresented, with 625 hits ($p<0.01$, = 62% of 1016 reference proteins); whereas proteins of the "Cell Periphery" are underrepresented, with 36 hits ($p<0.01$, = 3% of 1033 reference proteins). The majority of these hits are assigned to the subterm "plasma membrane" (~89% or 32 proteins). The compartment "Periplasmic Space", with 42 hits, appears not to be biased (= 30% of 138 reference proteins) (Figure 7, B).

To analyze how major metabolic pathways are represented in the S30 proteome, the Cytoscape plugin ClueGO (Marker list: *Escherichia coli* [562, 511145]; Source: KEGG as downloaded on 20th of February 2017) was used. The major physiological *E. coli* pathways are significantly overrepresented ($p \leq 0.05$) in the S30 extract proteome (Figure 8).

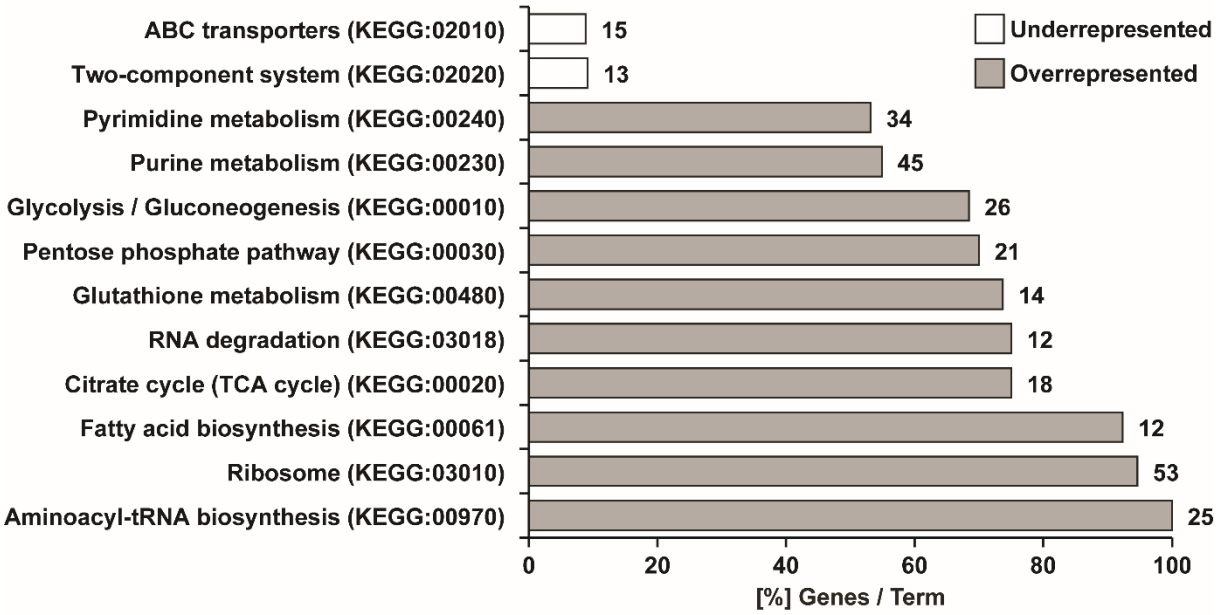


Figure 8: ***Over- and underrepresented pathways in S30^{+D} extracts (821 proteins)****. KEGG pathways are plotted against percentage of identified gene products per term using the Cytoscape plugin ClueGO. Only significantly over- (grey) and underrepresented (white) pathways are shown. The number right of each individual bar indicates the number of assigned genes per pathway.*

Important catabolic pathways involved in purine-/pyrimidine-biosynthesis and energy generation such as glycolysis/gluconeogenesis and pentose-phosphate pathway and the TCA cycle are overrepresented, as it can be expected from an extract derived from exponentially growing *E. coli*. Components central to translation, namely the ribosome and the aminoacyl tRNA synthetases (Table 4) are strongly overrepresented with high emPAI values, suggesting that they are enriched in S30 extracts. Exceptions are the glycine-tRNA ligase subunit (GlyQ) and the methionyl-tRNA formyltransferase (Fmt, required for initiation of translation), which were detected with comparatively low emPAI values, suggesting a potential shortage of these two enzymes during protein expression. Another potential limitation could be imposed by the abundant RNA degradation machinery, which is also overrepresented ($p \leq 0.05$). As a consequence of the underrepresentation of membrane proteins, pathways including membrane

proteins such as ABC transporters and two component system are also significantly underrepresented.

Table 4: Aminoacyl tRNA Synthetases identified in S30+D(1-4) extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
		S30 [1]	S30 [2]	S30 [2]	S30 [4]	
P00957	Alanine--tRNA ligase (alaS)	42	28	31	28	5.5 ± 1.7
P11875	Arginine--tRNA ligase (argS)	15	15	15	13	2.0 ± 0.3
P0A8M0	Asparagine--tRNA ligase (asnS)	19	21	21	23	6.4 ± 1.7
P21889	Aspartate--tRNA ligase (aspS)	26	27	26	27	7.0 ± 0.9
P21888	Cysteine--tRNA ligase (cysS)	17	15	12	13	3.1 ± 0.5
P04805	Glutamate--tRNA ligase (gltX)	17	20	15	20	4.4 ± 1.4
P00962	Glutamine--tRNA ligase (glnS)	15	16	17	13	2.0 ± 0.4
P27305	Glutamyl - Q tRNA(Asp) synthetase (gluQ)	1	1	n/d	1	0.1 ± 0.0
P00960	Glycine--tRNA ligase alpha subunit (glyQ)	7	6	4	5	1.3 ± 0.3
P00961	Glycine--tRNA ligase beta subunit (glyS)	39	32	34	33	8.0 ± 0.7
P60906	Histidine--tRNA ligase (hisS)	n/d	19	19	18	8.6 ± 1.2
P00956	Isoleucine--tRNA ligase (ileS)	43	38	39	34	5.9 ± 0.8
P07813	Leucine--tRNA ligase (leuS)	33	28	29	26	4.0 ± 0.3
P0A8N3	Lysine--tRNA ligase (lysS)	20	22	23	23	15.2 ± 3.4
P0A8N5	Lysine--tRNA ligase, heat inducible (lysU)	19	22	24	19	14.4 ± 4.8
P00959	Methionine--tRNA ligase (metG)	28	21	22	20	3.2 ± 0.6
P08312	Phenylalanine--tRNA ligase alpha subunit (pheS)	15	16	17	16	8.5 ± 1.6
P07395	Phenylalanine--tRNA ligase beta subunit (pheT)	27	24	23	19	2.6 ± 0.4
P16659	Proline--tRNA ligase (proS)	21	22	22	21	4.8 ± 0.7
P0A8L1	Serine--tRNA ligase (serS)	12	17	17	15	4.4 ± 1.2
P0A8M3	Threonine--tRNA ligase (thrS)	34	22	27	25	5.6 ± 1.2
P00954	Tryptophan--tRNA ligase (trpS)	9	9	6	9	2.2 ± 0.3
P0AGJ9	Tyrosine--tRNA ligase (tyrS)	5	11	9	9	1.7 ± 0.7
P07118	Valine--tRNA ligase (valS)	42	46	46	34	7.0 ± 2.2

4.1.4 Relevant Factors for Transcription/Translation

Efficient translation is the core application of S30 extract in cell-free synthetic biology. Protein production is usually operated as coupled transcription/translation system by supplying T7-RNA-polymerase. Alternatively, the more complex endogenous RNA polymerase can be employed for transcription [108, 109].

The endogenous transcription pathway appears to be completely present in S30 lysates. All four constituents (RpoABCZ), corresponding to the subunits α, β, β' and ϖ building the *E. coli* RNA polymerase core enzyme could be identified, and overall high emPAI values indicate abundant copies of these proteins within the extract (Table 5). Four out of the seven *E. coli* sigma factors are present, and highest emPAI value was obtained for sigma70 (RpoD), necessary for recognition of general housekeeping promoters.

The S30$^{+D(1-4)}$ extracts exclusively contained the cold-shock transcription factors cspA and cspG, while cspC is present in both S30 extracts (S30$^{+D(1-4)}$ and S30^{-D}) at similar levels (data not shown).

Table 5: Transcription-related proteins identified in S30$^{+D(1-4)}$ extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

	UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
			S30 [1]	S30 [2]	S30 [2]	S30 [4]	
RNAP	P0A7Z4	DNA-directed RNA polymerase subunit alpha (rpoA)	17	19	16	18	10.6 ± 2.3
	P0A8V2	DNA-directed RNA polymerase subunit beta (rpoB)	74	67	73	67	13.4 ± 1.3
	P0A8T7	DNA-directed RNA polymerase subunit beta' (rpoC)	74	68	74	63	12.4 ± 2.1
	P0A800	DNA-directed RNA polymerase subunit omega (rpoZ)	4	5	6	4	7.9 ± 3.9
	P60240	RNA polymerase -associated protein (rapA)	3	7	18	n/d	0.6 ± 0.5
σ-factors	P00579	RNA polymerase sigma factor (rpoD)	17	13	13	10	1.5 ± 0.4
	P24255	RNA polymerase sigma-54 factor (rpoN)	n/d	1	4	2	0.2 ± 0.1
	P24251	Sigma factor-binding protein (crl)	3	4	5	4	2.3 ± 0.8
termination	P0A6W5	Transcription elongation factor (greA)	9	9	7	9	8.1 ± 1.9
	P0AG30	Transcription termination factor (rho)	n/d	9	9	2	1.0 ± 0.6
	P0AFF6	Transcription termination/antitermination protein (nusA)	16	19	20	20	4.8 ± 1.1
	P0AFG0	Transcription termination/antitermination protein (nusG)	6	7	8	6	5.0 ± 1.1
cold-shock	P0A9X9	Cold shock protein (cspA)	1	3	2	2	5.8 ± 3.1
	P0A9Y6	Cold shock-like protein (cspC)	3	4	3	4	7.7 ± 3.0
	P0A972	Cold shock-like protein (cspE)	n/d	3	3	3	5.3 ± 0.0
	P0A978	Cold shock-like protein (cspG)	1	3	2	2	5.3 ± 2.8

In view of the translation machinery, over 90% of the ribosome associated proteins could be identified, suggesting a high abundance of this key complex for cell-free protein synthesis (Figure 8). All subunits necessary for ribosome assembly are present in sufficient amounts, given the expression efficiency of both extracts (Figure 4). However, there are other essential factors required for efficient initiation, elongation and termination during translation. The initiation of prokaryotic translation involves formylation of the start methionine and the initiation factors IF1-3 (InfABC). These central factors were all detected repeatedly with comparable emPAI values (Table 6, Initiation), whereas the ratio of InfA:InfB:InfC based on averaged emPAI values was determined to be approx. 5:1:2. The expression of these factors is often highly regulated, to maintain an optimal ratio of these factors to each other and to the amount of ribosomes present, ensuring optimal interaction.

Three out of four elongation factors are essential for the elongation cycle during translation. The GTP bound EF-Tu (coded by *tufA*) and GTP exchange factor EF-Ts (coded by *tsf*) mediate the entry of loaded tRNA into the A-site of the ribosome. This process is catalyzed by GTP hydrolysis. Then, the elongation factor EF-G (coded by *fusA*) promotes translocation of tRNA from the A to the P site, which is also catalyzed by GTP hydrolysis.

The essential elongation factors EF-Tu, EF-Ts, and EF-G have all been detected. Particularly, the factors EF-Tu and EF-Ts were detected with comparably high emPAI values of approx. 27, and 18-24 unique significant peptide sequences, suggesting a high abundance of these factors. Another important ratio repeatedly mentioned in literature is the ratio of *efg*:*tsf* and was determined based on averaged emPAI values to be approx. 1:2.

The release factors 1-3 (RF1-3) (coded by *prfABC*), and the ribosome recycling factor are all present at comparable levels with exception of RF-1 (coded by *prfA*) which appears to be less abundant. RF-1 recognizes specifically the amber stop codon. This finding might become interesting for site-specific labelling approaches using the amber stop suppression strategy [110].

Table 6: Translation-related proteins identified in S30$^{+D(1-4)}$ extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

	UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
			S30 [1]	S30 [2]	S30 [2]	S30 [4]	
initiation	P69222	Translation initiation factor IF-1 (infA)	2	3	4	3	13.2 ± 9.5
	P0A705	Translation initiation factor IF-2 (infB)	30	29	29	20	2.8 ± 0.6
	P0A707	Translation initiation factor IF-3 (infC)	7	6	6	4	4.6 ± 1.0
elongation	P60785	Elongation factor 4 (lepA)	10	14	14	9	1.3 ± 0.5
	P0A6M8	Elongation factor G (fusA)	34	33	33	33	12.6 ± 1.2
	P0A6N4	Elongation factor P (efp)	3	4	3	3	1.1 ± 0.2
	P0A6N8	Elongation factor P-like protein (yeiP)	2	2	2	1	0.4 ± 0.1
	P0A6P1	Elongation factor Ts (tsf)	18	19	17	20	27.9 ± 11.2
	P0CE47	Elongation factor Tu 1 (tufA)	21	24	24	24	27.2 ± 9.2
	P0A9W3	Energy-dependent translational throttle protein (ettA)	1	8	10	n/d	0.7 ± 0.4
	P0AFX0	Ribosome-hibernation promoting factor (hpf)	2	n/d	2	2	2.5 ± 0.2
	P0A832	SsrA-binding protein (smpB)	6	4	4	1	1.7 ± 0.9
	P45748	Threonylcarbamoyl-AMP synthase (tsaC)	5	5	4	n/d	2.0 ± 0.6
termination	P0A7I0	Peptide chain release factor 1 (prfA)	3	4	4	5	0.6 ± 0.1
	P07012	Peptide chain release factor 2 (prfB)	n/d	12	13	11	2.9 ± 0.4
	P0A7I4	Peptide chain release factor 3 (prfC)	11	14	14	14	2.0 ± 0.5
	P0A805	Ribosome-recycling factor (frr)	4	4	4	3	1.9 ± 0.3

4.1.5 Protein Folding and Stability of Precursors/Target Proteins

During and after translation, the correct folding of synthesized recombinant proteins is of pivotal importance for applications of cell-free expression systems. Chaperones assisting protein folding are often necessary to obtain high quality samples of functional protein. In addition, chaperones may be beneficial in supporting the stability and lifetime of proteins involved in the expression machinery of the extract.

A variety of chaperones originating from the cytoplasm as well as from the periplasm could be identified in the S30 extracts (Table 7). Judging by number as well as their emPAI

values, the cytoplasmic chaperones appear more abundant if compared to the periplasmic ones. The major cytoplasmic chaperone systems such as trigger factor (TF), DnaK/DnaJ and GroEL/GroES (major determinants of nascent protein folding) are present in S30 extracts (Table 7). Especially TF, DnaK and GroEL were identified with high emPAI values, suggesting high abundance. The corresponding co-chaperones DnaJ and GroES are less abundant. Also, several types of peptidyl cis-trans-isomerases either from cytoplasmic as well as from periplasmic origin have been detected (Table 7).

One of the major advantages of cell-free protein synthesis is its open nature that allows direct access to the translational environment. This access is especially advantageous for the expression of disulfide-bonded proteins requiring optimized redox conditions to form their native disulfide bonds. The formation of disulfide bonds is *in vivo* catalyzed by a disulfide bond (Dsb) formation system consisting of the enzymes DsbA, B, C, D, and G. This proteome analysis identified only DsbA and DsbC with averaged emPAI values of 2.3 and 0.6, respectively (Table S1). Especially in $S30^{-D}$ these proteins were always detected with emPAI values <0.22 (data not shown), suggesting a potential shortage of these enzymes in S30 extract depending on the preparation procedure. The membrane embedded DsbB was not identified at all.

Table 7: Chaperones identified in $S30^{+D(1-4)}$ extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

	UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
			S30 [1]	S30 [2]	S30 [2]	S30 [4]	
cytoplasmic	P0A850	Trigger factor (tig)	22	31	27	28	23.2 ± 8.9
	P0A6F9	10 kDa chaperonin (groS)	4	5	4	6	7.5 ± 3.7
	P23869	Peptidyl -prolyl cis -trans isomerase B (ppiB)	8	8	8	6	7.1 ± 2.2
	P0A6Y8	Chaperone protein (dnaK)	34	25	29	29	6.5 ± 1.2
	P0A6F5	60 kDa chaperonin (groL)	20	23	23	23	6.4 ± 1.5
	P0A9L5	Peptidyl -prolyl cis -trans isomerase C (ppiC)	n/d	4	4	3	4.3 ± 1.0
	P0A9K9	FKBP -type peptidyl -prolyl cis -trans isomerase (slyD)	4	4	4	4	2.6 ± 0.2
	P0A9L3	FKBP -type 22 kDa peptidyl -prolyl cis -trans isomerase (fklB)	6	6	6	6	2.4 ± 0.2
	P25522	tRNA modification GTPase (mnmE)	n/d	12	9	7	1.5 ± 0.5
	P77395	Uncharacterized protein (ybbN)	4	4	3	2	0.6 ± 0.2
	P10408	Protein translocase subunit (secA)	3	6	14	n/d	0.5 ± 0.3
	P0AEM0	FKBP -type 16 kDa peptidyl -prolyl cis -trans isomerase (fkpB)	1	2	2	1	0.5 ± 0.2
periplasmic	P45523	FKBP -type peptidyl -prolyl cis -trans isomerase (fkpA)	4	6	5	8	1.6 ± 0.7
	P31697	Chaperone protein (fimC)	6	4	6	4	1.5 ± 0.3
	P0ABZ6	Chaperone (surA)	2	10	10	7	1.3 ± 0.7
	P0AFL3	Peptidyl -prolyl cis -trans isomerase A (ppiA)	4	3	2	1	0.8 ± 0.4
	P0AEU7	Chaperone protein (skp)	1	2	2	2	0.7 ± 0.3

The major cost factor for cell-free synthetic biology are the amounts of supplied precursor such as NTPs, amino acids or the energy source phosphoenolpyruvate. Energy consuming enzymes, which are redirecting valuable energy sources to non-productive

metabolic pathways instead of protein synthesis, are frequent in the S30 extract proteome (Table S1). Many precursor consuming proteins involved in metabolic pathways including glycolysis (approximately 70% identified), fatty acid biosynthesis (92% identified) and pyruvate metabolism (67% identified) were detected (Figure 8).

Precursor stability is further affected by degrading or modifying enzymes, and such processes are major problems in applications like protein labelling. Several amino acid metabolic pathways, including those of aspartate, glutamate and tryptophan are remarkably overrepresented, with 75% or more of identified proteins (Figure 8). As a consequence, numerous amino acid modifying enzymes like TnaA (tryptophanase), AspC (aspartate aminotransferase), GlnA (glutamine synthetase), GshA (glutamate--cysteine ligase) and SpeA (arginine decarbocylase) remain in S30 lysates (Table 8), resulting in significant consumption and destabilization of amino acid or even causing severe scrambling of labelled amino acids in *e.g.* NMR experiments [24].

Besides numerous amino acid modifying enzymes, 5 out of 8 subunits of the *E. coli* ATP synthase were continuously found in all biological replicates, with a considerable number of unique significant peptides and parts of the closely-linked electron transport chain, albeit with low emPAI values. While these important membrane-embedded proteins of the plasma membrane are detected repeatedly, genuine outer membrane are scarce in S30 extracts. Examples of outer membrane proteins that were detected repeatedly in biological replicates were OmpACX, but with low emPAI values.

Besides folding, the stability and integrity of the target as well as S30 extract proteins are central to CFPS. Peptidases are vital to all living cells fulfilling important roles in regulation, metabolism or differentiation. However, for cell-free protein synthesis, as well as the subsequent purification of protein products, such enzymes are usually a threat. A comprehensive set of proteases was identified within the S30 extracts, mostly with relatively low emPAI values (Table 9). The most abundant protease with respect to its emPAI value is a subunit of the ATP dependent ClpP, which is mostly responsible for degradation of SsrA-tagged proteins [111]. Albeit with lower emPAI values, protease Lon was detected repeatedly with the highest number of unique peptides. In the class of membrane integrated proteases, only the outer membrane protease OmpT was identified - albeit with low abundance values and only occasionally – suggesting that OmpT is mostly depleted during extract preparation (data not shown).

In summary, 1074 proteins were identified in $S30^{+D(1-4)}$ combined, including proteins that were identified uniquely in one extract. The S30 core proteome, defined as proteins that

were identified repeatedly in at least three biological replicates, was comprised of 821 proteins, of which relative quantitative and approximate emPAI values were provided to

Table 8: Miscellaneous proteins identified in S30$^{+D(1-4)}$ extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

	UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
			S30^{1}	S30^{2}	S30^{3}	S30^{4}	
amino acid modifying	P22106	Asparagine synthetase B [glutamine-hydrolyzing] (asnB)	1	4	7	n/d	0.4 ± 0.3
	P00509	Aspartate aminotransferase (aspC)	11	10	12	13	2.6 ± 0.6
	P21170	Biosynthetic arginine decarboxylase (speA)	17	10	11	12	1.3 ± 0.3
	P0ABK5	Cysteine synthase A (cysK)	17	15	16	16	9.9 ± 1.4
	P16703	Cysteine synthase B (cysM)	5	5	4	4	1.0 ± 0.3
	P09831	Glutamate synthase [NADPH] large chain (gltB)	58	49	40	31	2.9 ± 1.1
	P09832	Glutamate synthase [NADPH] small chain (gltD)	n/d	4	3	3	0.4 ± 0.1
	P0A6W9	Glutamate--cysteine ligase (gshA)	6	8	8	9	0.9 ± 0.2
	P0A9C5	Glutamine synthetase (glnA)	17	20	19	18	6.0 ± 1.6
	P0A9C5	Glutamine synthetase (glnA)	17	20	19	18	6.0 ± 1.6
	P06715	Glutathione reductase (gor)	2	7	8	7	1.0 ± 0.5
	P0A962	L-asparaginase 1 (ansA)	4	3	2	1	0.4 ± 0.2
	P16095	L-serine dehydratase 1 (sdaA)	4	5	4	4	0.6 ± 0.1
	P23721	Phosphoserine aminotransferase (serC)	9	8	5	10	1.9 ± 0.6
	P0A817	S-adenosylmethionine synthase (metK)	12	14	13	14	3.8 ± 1.0
	P0A9D4	Serine acetyltransferase (cysE)	5	4	4	4	0.9 ± 0.1
	P0A825	Serine hydroxymethyltransferase (glyA)	8	17	18	21	7.3 ± 4.2
	P0A877	Tryptophan synthase alpha chain (trpA)	2	2	1	n/d	0.3 ± 0.1
	P0A853	Tryptophanase (tnaA)	n/d	11	9	4	1.1 ± 0.5
respiration / ATP-synthase	P0ABB0	ATP synthase subunit alpha (atpA)	12	13	11	14	1.8 ± 0.3
	P0A6E6	ATP synthase epsilon chain (atpC)	2	3	4	3	1.6 ± 0.6
	P0ABB4	ATP synthase subunit beta (atpD)	9	13	11	14	2.0 ± 0.6
	P0ABA6	ATP synthase gamma chain (atpG)	5	4	4	4	0.8 ± 0.1
	P0ABA4	ATP synthase subunit delta (atpH)	n/d	2	2	2	0.6 ± 0.0
	P33599	NADH-quinone oxidoreductase subunit C/D (nuoC)	3	4	6	3	0.3 ± 0.1
	P0AFD1	NADH-quinone oxidoreductase subunit E (nuoE)	2	4	4	2	1.7 ± 0.7
	P33602	NADH-quinone oxidoreductase subunit G (nuoG)	7	n/d	6	5	0.3 ± 0.0
translocation	P10408	Protein translocase subunit (secA)	3	6	14	n/d	0.5 ± 0.3
	P0AG86	Protein-export protein (secB)	2	3	3	3	1.4 ± 0.5
	P0A917	Outer membrane protein X (ompX)	3	2	2	1	0.6 ± 0.2
	P0A910	Outer membrane protein A (ompA)	6	4	2	3	0.6 ± 0.3
	P06996	Outer membrane protein C (ompC)	8	6	6	6	1.1 ± 0.1
	P0AE18	Methionine aminopeptidase (map)	9	7	8	8	2.7 ± 0.3

estimate how abundant the respective protein in S30$^{+D(1-4)}$ is. The S30 proteome was furthermore classified according to the original localization of the respective proteins in the living cell. As expected, most proteins were originally located in the cytoplasm and to a lesser extend in the periplasm. Membrane proteins, except for proteins of the plasma membrane, are in contrast mostly depleted. Due to the open nature of CFPS systems, they are considered

advantageous for the expression of disulfide bonded proteins. The *dsb* systems located in the periplasm and central for the native expression of disulfide bonded proteins, however, appears to be incomplete, and with low abundance values, therefore revealing a possible weak point of CFPS systems based on S30 extracts.

Table 9: Proteases identified in S30$^{+D(1-4)}$ extract. The emPAI values are described as mean ± SD of proteins detected in at least three biological replicates.

UniProt #	Description (Gene ID)	# significant unique sequences				emPAI
		S30 [1]	S30 [2]	S30 [2]	S30 [4]	
P0A6G7	ATP-dependent Clp protease proteolytic subunit (clpP)	5	5	5	4	2.8 ± 0.9
P0ACF4	DNA-binding protein HU-beta (hupB)	n/d	3	2	2	2.3 ± 0.8
P27298	Oligopeptidase A (prlC)	23	13	15	15	1.7 ± 0.5
P0A7B8	ATP-dependent protease subunit (hslV)	3	2	2	2	1.2 ± 0.1
P0A9M0	Lon protease (lon)	21	5	20	10	1.2 ± 0.7
P0A7C2	LexA repressor (lexA)	4	2	3	n/d	0.8 ± 0.2
P12008	Chorismate synthase (aroC)	3	3	3	3	0.4 ± 0.0
P23865	Tail-specific protease (prc)	7	6	5	3	0.4 ± 0.1
P0C0V0	Periplasmic serine endoprotease (degP)	1	2	3	5	0.3 ± 0.2
P0AAI3	ATP-dependent zinc metalloprotease (ftsH)	4	n/d	3	4	0.3 ± 0.0
P05458	Protease 3 (ptrA)	9	2	6	2	0.3 ± 0.1
P0A898	Endoribonuclease (ybeY)	1	1	1	1	0.3 ± 0.0
P24171	Peptidyl -dipeptidase (dcp)	7	2	2	n/d	0.2 ± 0.2
P0AEB2	D-alanyl -D-alanine carboxypeptidase (dacA)	1	1	1	1	0.1 ± 0.0

Furthermore, groups of proteins central to transcription, translation, folding and protein degradation were summarized in tables, to provide a framework for future approaches in cell-free metabolic engineering, genome editing to improve protein stability/folding (*e.g.* by knockout of proteases or knock in of chaperones) or precursor stability (*e.g.* by knockout of substrate-degrading enzymes).

4.2 Tuning Folding of Proteins in Cell-Free Extracts

While the S30 extract preparation procedure has been optimized over the years, only little attention has been paid to the cultivation conditions of the source strain *E. coli* A19 before extract preparation. It was therefore hypothesized that the induction of a SOS response during cultivation and prior to harvesting results in dynamic proteome changes, and in turn results in improved S30 extracts composition, due to the up-regulation of chaperone networks in response to stressful environmental conditions (suboptimal high temperatures and the presence of ethanol). The extracts after induction of a SOS response were compared to S30^{+D} extracts, and changes in proteome were determined using a quantitative approach based on ICPL labelling. Subsequently, the productivity of standard S30^{D+} and SOS extracts was compared using difficult-to-express proteins.

4.2.1 Comparative Analysis of Heat Shock Extracts vs. Standard Extracts

To investigate the effects of altered cultivation conditions on the composition of S30 extracts, the cultivation and extract preparation protocol for $S30^{+D}$ as described in 3.6.1 was employed. For the comparison, the $S30^{+D(1-4)}$ from the previous chapter was used and three biological replicates of $S30^{+D}$-S, hereafter referred to as $S30^{+D}$-S^{1-3}, were prepared according to the scheme illustrated in Figure 9, A and are described in detail (3.6.1). The cultivation for $S30^{+D}$-S^{1-3} included a heat shock at 42 °C for 30 min and simultaneous exposure to 3% (v/v) EtOH (SOS response*).

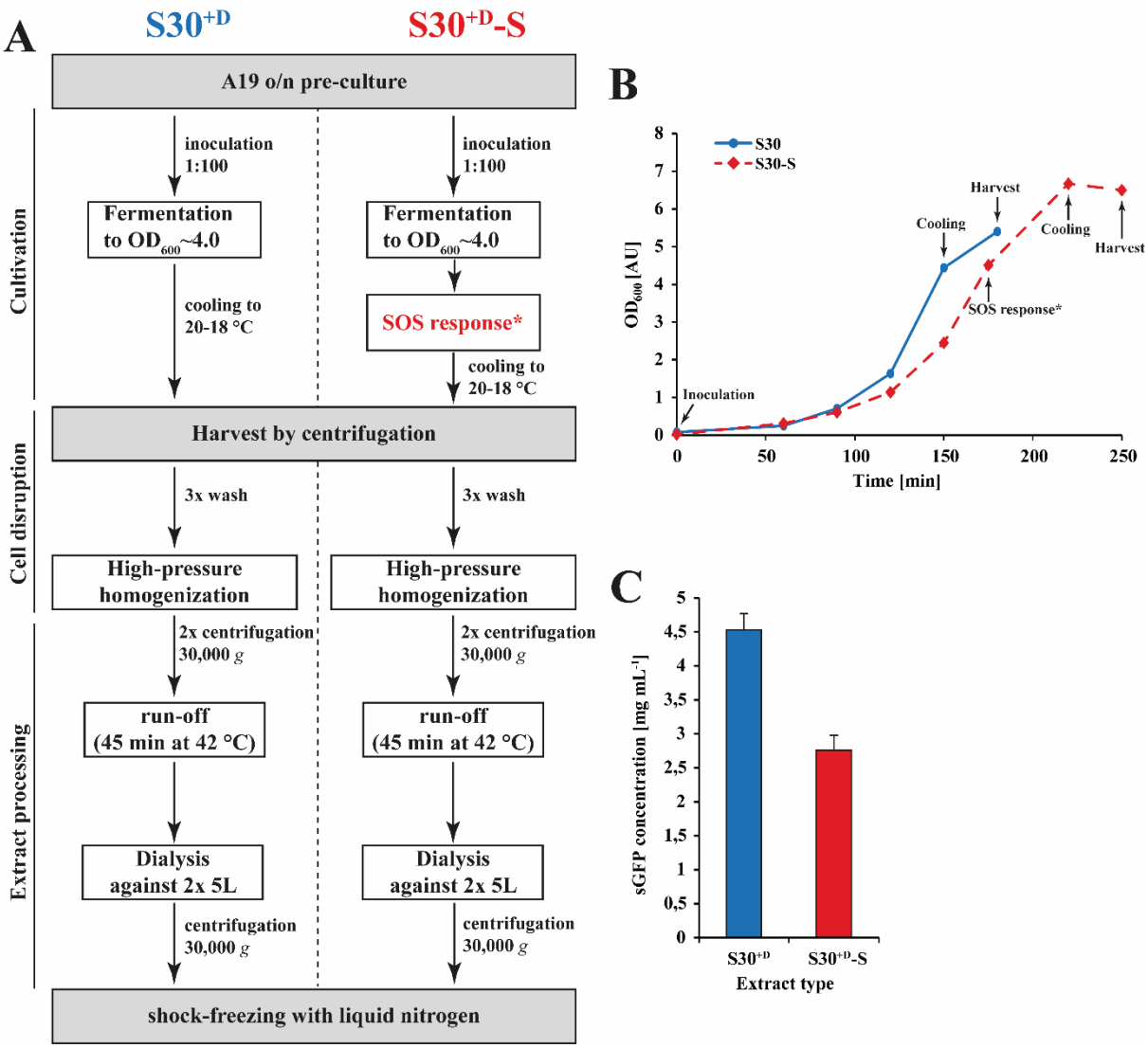


*Figure 9: **General parameters of S30 extracts used comparative analysis of $S30^{+D}$ and $S30^{+D}$-S extracts. Workflow of cultivation conditions, S30 extract preparation and extract performance is illustrated**. (A) Flowchart of different S30 ($S30^{+D}$ and $S30^{+D}$-S) extract preparation procedures (as described in more detail in 3.6.1). (B) Growth curves of E. coli A19 as observed under different fermentation conditions used for $S30^{+D}$ and $S30^{+D}$-S extract, respectively. (C) Performance evaluation of $S30^{+D}$ and $S30^{+D}$-S extract as determined under standard expression conditions using sGFP as a quantitative reporter and was carried out in CECF mode as described in (3.6.4). The error bars represent the standard deviation of three measurements. *The SOS-response was induced by addition of 3% (v/v) EtOH to the culture broth and simultaneous heat shock at 42 °C for 30 min.*

Inclusion of the SOS response at the tail of the fermentation led to a slightly different growth curve and longer fermentation time; the growth curve is depicted in Figure 9, B. The

expression efficiency of S30^{+D}-S$^{1-3}$ extracts is reduced compared to S30$^{+D(1-4)}$, with a yield of 2.75 mg mL^{-1} and 4.5 mg mL^{-1} for S30^{+D}-S$^{1-3}$ and S30$^{+D(1-4)}$, respectively, for sGFP (Figure 9, C). Despite the lower expression efficiency of S30^{+D}-S$^{1-3}$, it was hypothesized that these extracts may exhibit beneficial properties for the expression of proteins that have been shown to be difficult-to-express due to *e.g.* higher chaperone dependency.

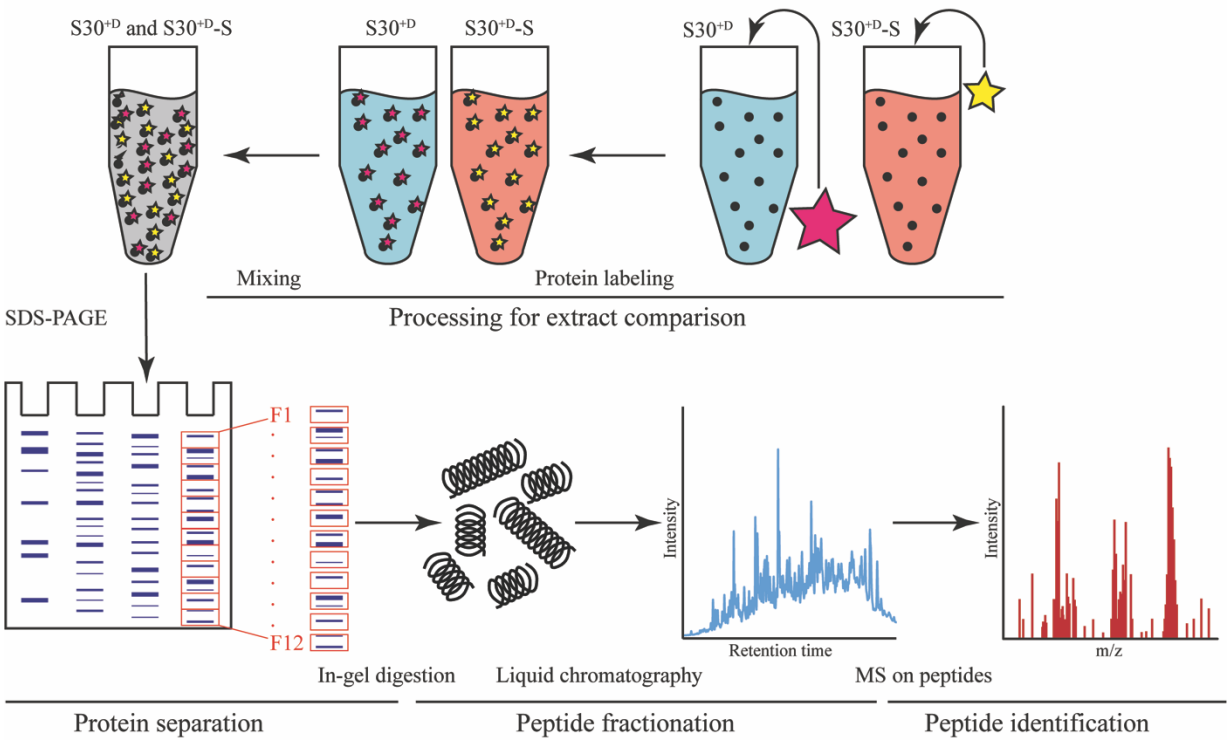


Figure 10: ***Strategy for quantitative GeLC-MS/MS analysis of S30^{+D} vs. S30^{+D}-S extract****. The S30$^{+D(1)}$ and S30^{+D}-S$^{1-3}$ extracts were individually labelled using ICPL strategy by heavy and light label. The individual biological replicates S30^{+D}-S^{1}, S30^{+D}-S^{2} and S30^{+D}-S^{3} were each mixed with the respective sample of S30$^{+D(1)}$ and then separated on protein-level by 1D-SDS-PAGE. The lane was then divided into 12 equal fractions (F1-F12) and proteins were in-gel digested (trypsin) followed by extraction. Peptides extracted from F1-F12 were further separated by nanoHPLC (Reversed Phase C18 column) and individual elution fractions (450 per gel-slice (F1-F12) were analyzed by MALDI-TOF-MS/MS to identify and quantify protein of both sample relative to each other.*

In order to quantify the changes of the S30^{+D}-S$^{1-3}$ proteome relative to the standard S30$^{+D(1)}$ extract proteome, an ICPL labelling strategy was employed (see Figure 10). In this process, the lysine residues of all denatured proteins of S30^{+D} extract were labelled with a heavy label, while the proteins of S30^{+D}-S$^{1-3}$ were labelled with a light label. Subsequently, the labelled extracts were mixed and separated using 1D-SDS-PAGE and analyzed as described previously using bottom-up proteomics. The light and heavy labelled peptides co-eluting during LC separation were analyzed simultaneously in one MS^1 spectrum as duplets with 6 Da difference in mass, whereas their relative intensities to each other were compared to identify up-, down- and non-regulated peptides. Protein identification was based on MS^2 data acquired in a second step.

The MS^2 data was then searched against the *E. coli* K12 reference proteome (Uniprot ID: UP000000625 as downloaded on 19th of January 2017). Statistical means were applied to reach a FDR on protein- and peptide level of ≤1%. The dataset was analyzed in triplicate

(biological replicates: $S30^{+D}$-S^1 vs. $S30^{+D(1)}$ (=B01_T01); $S30^{+D}$-S^2 vs. $S30^{+D(1)}$ (=B02_T01); $S30^{+D}$-S^3 vs. $S30^{+D(1)}$ (=B03_T01), whereas $S30^{+D}$-S^1 vs. $S30^{+D(1)}$ was additionally analyzed as technical replicate by swapping the labels (=B01_T02). Peptide ratios were determined by Quantitation Toolbox (Mascot Distiller v2.6, Matrix Science, UK), using the quality thresholds as described in detail in 3.7.4.

With an FDR on protein-level of ≤1% in in the individual datasets B01_T01, B01_T02, B02_T01 and B03_T01, a number of 724, 724, 806 and 761 proteins were found. The technical replicates B01_T01 and B01_T02 shared 655 proteins (82.6%), and 69 proteins (8.7%) were uniquely identified in each technical replicate. The high number of shared proteins in the two technical replicates suggests good technical reproducibility of the quantitative proteome analysis.

Next, the determined peptide ratios were normalized to median=0 in log_2-space and visualized using a box plot (Figure 11, A). The plot shows that the median of all technical and biological replicate datasets is on a line around log_2=0, suggesting that the datasets are not skewed and the distribution of data points, including the outliers, is comparable in all datasets. The correlation plot in Figure 11, B confirmed the overall high correlation between datasets, with Pearson coefficients between 0.738 ≤ x ≤ 0.826 for biological replicates, and 0.874 for the two technical replicates.

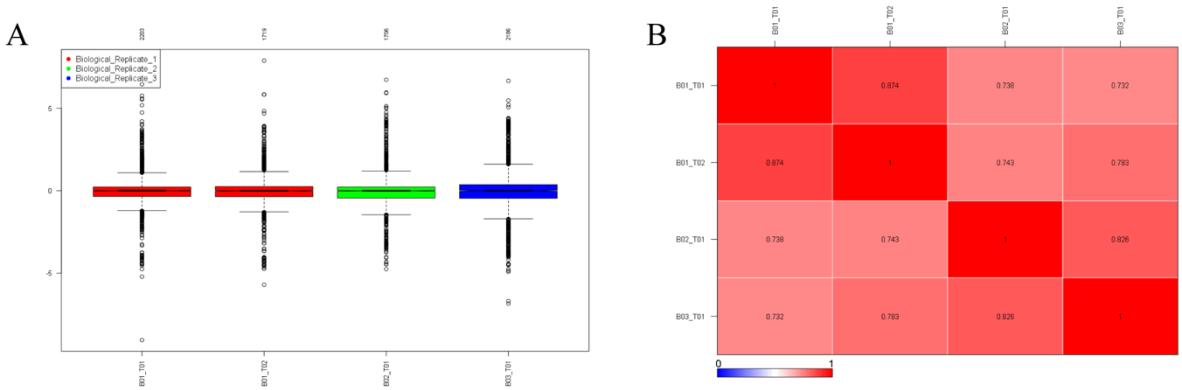


Figure 11: ***Evaluation of technical and biological replicates by Box plot and correlation plot****. (A) Box plot of technical replicates (T01-T02) and biological replicates (B01-B03) after normalization to median=0 in log_2 space. (B) Correlation plot of biological and technical replicates to identify problematic and outlier datasets. The Pearson correlation factor between each dataset is indicated.*

Having confirmed the technical reproducibility, the technical replicate B01_T02 (using swapped labels in contrast to the other biological replicates) was omitted from further analysis to avoid any bias towards the biological replicate #1.

In the datasets B01_T01, B02_T02 and B03_T01 a total of 901 proteins were identified. Among these datasets, 622 (69%) proteins were found in all three replicates, and 146 (16.3%) were found in at least two datasets. 25 (2.8%), 63 (7%) and 45 (5%) proteins were uniquely identified in B01_T01, B02_T01 and B03_T01, respectively (Figure 12, A). The fact that most

proteins (85.3%) were repeatedly detected in at least two biological replicates reinforces the high technical and biological reproducibility.

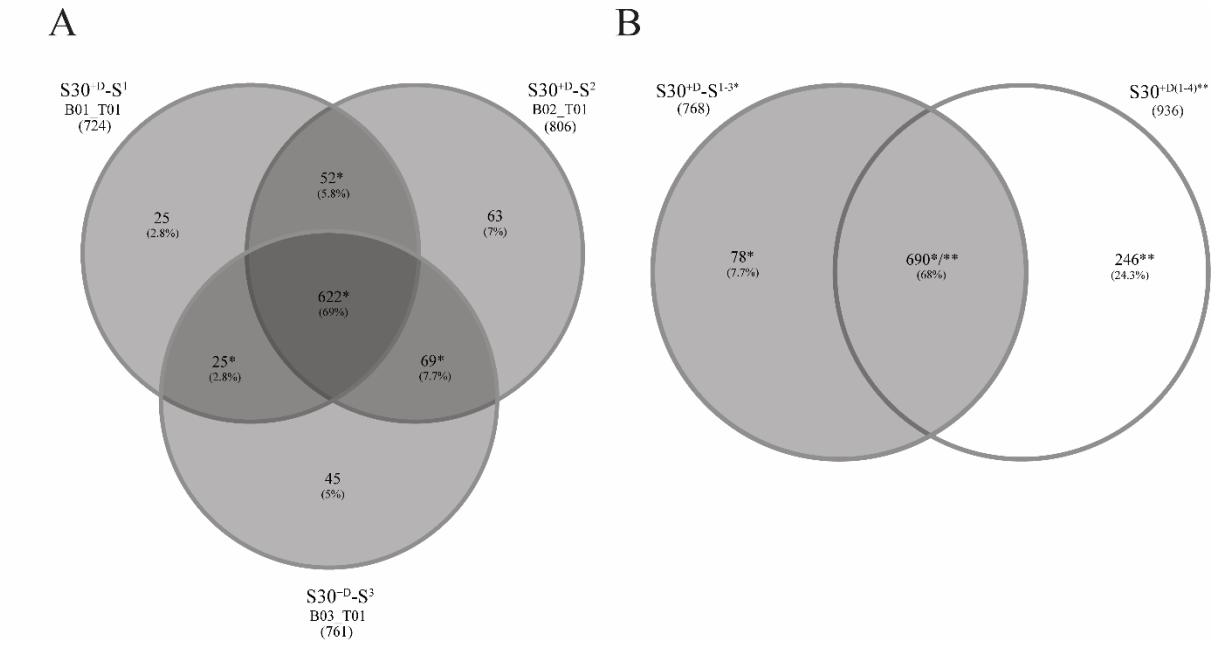


*Figure 12: **Venn Diagram to identify common and unique proteins in different samples.** (A) Number and percentage of common und unique proteins identified in three biological replicates B01_T01, B02_T01 and B03_T01 ($S30^{+D}$-S^{1-3} dataset). Proteins that were identified in at least two biological replicates are marked with an asterisk (*). (B) Number and percentage of common and unique proteins as identified in S30-S extracts and standard $S30^{+D}$ extracts. As basis for the comparison proteins were chosen that were found in at least two biological replicates in either the $S30^{+D}$-S^{1-3} dataset (indicated with one asterisk *) or in dataset $S30^{+D(1-4)}$ (indicated with two asterisks **).*

Proteins that were identified at least twice in $S30^{+D}$-S^{1-3} (referred to as $S30^{+D}$-S^{1-3}*) were then compared to proteins that were identified at least twice the previous dataset $S30^{+D(1-4)}$ (referred to as $S30^{+D(1-4)}$**) using a Venn diagram (Figure 12, B). The analysis shows that most proteins (690 or 68%) appeared in both analyses. Since the $S30^{+D(1-4)}$ analysis was performed without labelling, resulting in a less complex sample and thus higher sensitivity in the proteome analysis, it is not surprising that more proteins were identified.

The proteins identified uniquely either in $S30^{+D}$-S^{1-3}* or $S30^{+D(1-4)}$** were further analyzed using the Cytoscape plugin ClueGO (Marker List: *Escherichia coli* [562, 511145]; Ontology: Biological Process-GOA-3023 as download on 18th of November, 2016), to test if any biological pattern would arise when integrating the uniquely identified genes to GO terms. The 246** proteins identified in $S30^{+D(1-4)}$** did not result in any significantly enriched terms (pV≤0.01), thus supporting the idea that they were present due to technical differences in the analysis process. In contrast, within the 78* uniquely identified proteins in $S30^{+D}$-S^{1-3}*, the GO terms response to temperature stimulus (GOID 9266) and organonitrogen compound metabolic process (GOID 1901564) were significantly enriched (pV≤0.01). In particular, the enrichment of the GO term response to temperature stimulus in $S30^{+D}$-S^{1-3}* implies that some proteins are not detectable in standard $S30^{+D(1-4)}$ extract and are only expressed after the induction of the

SOS response for S30^{+D}-S$^{1-3}$* extract. These uniquely identified proteins are potentially of high interest, as they are highly up-regulated but probably not covered by the quantitative proteomics approach, as no corresponding labelled peptide from S30$^{+D(1-4)}$ may be present to compare peak intensities of the two samples. These uniquely identified proteins were therefore be matched after analyzing the quantitative data.

After individual analysis, the datasets B01_T01, B02_T01 and B03_T01 were combined and further analyzed using Quantitative Proteomics p-Value Calculator (QPPC) to identify significantly regulated proteins based on an implementation of a permutation test [101, 112]. The calculated pVs were transformed using the negative logarithm to the base 10 and plotted against fold change transformed using natural logarithm to visualize regulated proteins as Volcano plot (Figure 13, A).

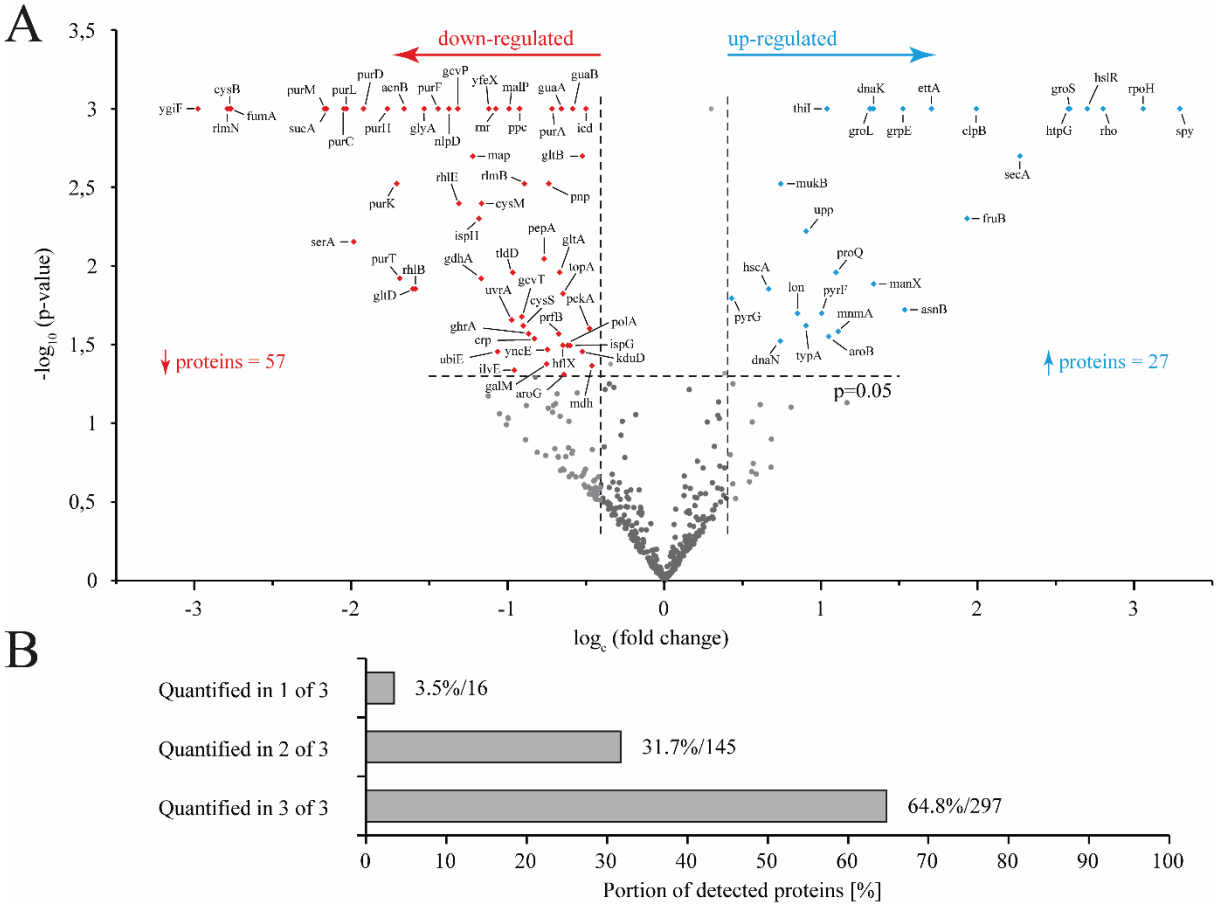


Figure 13: ***Quantitative proteome analysis of S30$^{+D(1)}$ vs. S30^{+D}-S$^{1-3}$ extract.*** *(A) Volcano plot of quantitative proteome analysis of S30$^{+D(1)}$ vs. S30^{+D}-S$^{1-3}$. 57 proteins were found down-regulated (left, red) and 27 proteins were up regulated (right, blue) in S30^{+D}-S$^{1-3}$ relative to S30$^{+D(1)}$ extract, whereas a total of 458 proteins of 901 identified proteins were quantified with at least two peptides. Only proteins that were showing a fold change ≥1.5 (dashed vertical line) and were above significance level (pV≤0.05) (dashed horizontal line) as determined by Quantitative Proteomics p-Value Calculator (QPPC) [101, 112] were considered significantly regulated. (B) Bar diagram illustrating reproducibility of three biological replicates (S30^{+D}-S$^{1-3}$). The majority of proteins (>96%) were quantified in at least two biological replicates. Only 3.2% were uniquely quantified in one biological replicate. Numbers besides the bars represent the percentage and the number of quantified proteins.*

Regulated proteins appear, due to the nature of data transformation, in the left and right upper corner of the plot, and the proteins that were regulated under both criteria (pV cut-off of 0.05 and regulated more than 1.5-fold) were labelled with the corresponding GeneID.

According to this analysis, 57 proteins were down-regulated (Table 10) and 27 proteins were up-regulated (Table 11) in $S30^{+D}$-S^{1-3} relative to $S30^{+D(1)}$. Only proteins that were quantified in at least one biological replicate and with at least two peptides were incorporated in the Volcano plot. Therefore, the Volcano plot covers all quantified proteins, where 64.8% (=297 proteins) were found in all three replicates, 31.7% (=145 proteins) in two replicates, and 3.5% (=16 proteins) in one replicate.

As mentioned before, the 78 proteins exclusively identified in $S30^{+D}$-S^{1-3} (Figure 12, B) are potentially highly regulated proteins, but may be missed in the quantitative analysis due to the lack of labelled counterpart in $S30^{+D(1-4)}$ extract. The proteins rpoH, spy, hslR, fruB and aroB, which are among the most up-regulated proteins in the quantitative analysis (see Figure 13), are borderline cases. The availability of quantitative data for these proteins suggests that, even though a labelled counterpart of these proteins must be present in $S30^{+D(1)}$, the amount was not sufficient for identification based on MS^2 spectra. Therefore, their identity was inferred from the labelled peptide present in $S30^{+D}$-S^{1-3} extract. In order to cover potentially highly regulated proteins as well, uniquely identified but not quantified proteins that were assigned to GO term response to temperature stimulus (GOID 9266) (namely dtd, hflC, hflK, pspA, raiA, and yajL) were included in further analyses, as their expression was likely induced during the heat shock applied in $S30^{+D}$-S^{1-3} extract preparation.

In order to get an impression of their function, up- and down-regulated proteins listed in Table 10 and Table 11 were integrated using the Cytoscape plugin ClueGO to assign Gene ID to GO terms and KEGG pathways. This analysis could determine whether certain proteins were significantly enriched in comparison to a reference set of proteins (Uniprot ID: UP000000625, 4315 sequences as downloaded on 19th of January 2017) that represents all proteins that could theoretically be detected in the sample.

The 57 down-regulated proteins were analyzed using the ontology sets BiologicalProcess-GOA (as downloaded on 18th of November 2016) and KEGG (as downloaded 21st of November 2016). Out of 57 proteins, 53 (=92.98%) were functionally annotated in the selected ontology sets, and after applying the selection criteria (GO terms/pathways with $pV \leq 0.01$; kappa score > 0.4; minimum number of genes 3 and minimum percentage 4%; GO level 3-8; statistical test: Fishers Exact Test for enrichment (right-sided hypergeometric test); correction for multiple test: Bonferroni step down), 47 proteins (=82.46%) were associated with representative GO-terms/pathways, where the assignment was based on all evidence levels except inferred from electronic annotation (IEA).

Table 10: Down-regulated proteins in $S30^{+D}$-S^{I-3} relative to $S30^{+D(1)}$ as determined by QPPC.

UniProt #	Description (Gene ID)	Ratio (log_e)	$Ratio^{-1}$ (L/H)	# obs.	p-Value
P30871	Inorganic triphosphatase (ygiF)	-3.0	19.6	4	1.0E-03
P36979	Dual-specificity RNA methyltransferase (rlmN)	-2.8	16.3	5	1.0E-03
P0A9F3	HTH-type transcriptional regulator (cysB)	-2.8	16.0	2	1.0E-03
P0AC33	Fumarate hydratase class I, aerobic (fumA)	-2.8	15.9	2	1.0E-03
P08178	Phosphoribosylformylglycinamidine cyclo-ligase (purM)	-2.2	8.8	6	1.0E-03
P0AFG3	2-oxoglutarate dehydrogenase E1 component (sucA)	-2.2	8.6	52	1.0E-03
P0A7D7	Phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)	-2.0	7.8	6	1.0E-03
P15254	Phosphoribosylformylglycinamidine synthase (purL)	-2.0	7.6	57	1.0E-03
P0A9T0	D-3-phosphoglycerate dehydrogenase (serA)	-2.0	7.3	2	7.0E-03
P15640	Phosphoribosylamine--glycine ligase (purD)	-1.9	6.8	12	1.0E-03
P15639	Bifunctional purine biosynthesis protein (purH)	-1.8	5.9	31	1.0E-03
P09029	N5-carboxyaminoimidazole ribonucleotide synthase (purK)	-1.7	5.5	3	3.0E-03
P33221	Phosphoribosylglycinamide formyltransferase 2 (purT)	-1.7	5.4	2	1.2E-02
P36683	Aconitate hydratase B (acnB)	-1.7	5.3	22	1.0E-03
P09832	Glutamate synthase [NADPH] small chain (gltD)	-1.6	5.0	2	1.4E-02
P0A8J8	ATP-dependent RNA helicase (rhlB)	-1.6	4.9	2	1.4E-02
P0A825	Serine hydroxymethyltransferase (glyA)	-1.5	4.6	27	1.0E-03
P0AG16	Amidophosphoribosyltransferase (purF)	-1.4	4.2	16	1.0E-03
P0ADA3	Murein hydrolase activator (nlpD)	-1.4	4.0	4	1.0E-03
P33195	Glycine dehydrogenase (decarboxylating) (gcvP)	-1.3	3.7	18	1.0E-03
P25888	ATP-dependent RNA helicase (rhlE)	-1.3	3.7	5	4.0E-03
P0AE18	Methionine aminopeptidase (map)	-1.2	3.4	6	2.0E-03
P62623	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (ispH)	-1.2	3.3	4	5.0E-03
P00370	NADP-specific glutamate dehydrogenase (gdhA)	-1.2	3.2	3	1.2E-02
P16703	Cysteine synthase B (cysM)	-1.2	3.2	5	4.0E-03
P76536	Probable deferrochelatase/peroxidase (yfeX)	-1.1	3.1	7	1.0E-03
P21499	Ribonuclease R (rnr)	-1.1	2.9	8	1.0E-03
P0A887	Ubiquinone/menaquinone biosynthesis C-methyltransferase (ubiE)	-1.1	2.9	3	3.5E-02
P00490	Maltodextrin phosphorylase (malP)	-1.0	2.7	28	1.0E-03
P0A698	UvrABC system protein A (uvrA)	-1.0	2.7	4	2.2E-02
P0AGG8	Metalloprotease (tldD)	-1.0	2.6	6	1.1E-02
P0AB80	Branched-chain-amino-acid aminotransferase (ilvE)	-1.0	2.6	3	4.6E-02
P00864	Phosphoenolpyruvate carboxylase (ppc)	-0.9	2.5	29	1.0E-03
P27248	Aminomethyltransferase (gcvT)	-0.9	2.5	5	2.1E-02
P21888	Cysteine--tRNA ligase (cysS)	-0.9	2.5	4	2.4E-02
P63177	23S rRNA (guanosine-2-O)-methyltransferase (rlmB)	-0.9	2.4	9	3.0E-03
P75913	Glyoxylate/ hydroxypyruvate reductase A (ghrA)	-0.9	2.4	5	2.7E-02
P0ACJ8	cAMP-activated global transcriptional regulator (crp)	-0.8	2.3	5	2.9E-02
P68767	Cytosol aminopeptidase (pepA)	-0.8	2.2	7	9.0E-03
P0A9C3	Aldose 1-epimerase (galM)	-0.8	2.1	5	4.2E-02
P76116	Uncharacterized protein (yncE)	-0.7	2.1	6	3.4E-02
P05055	Polyribonucleotide nucleotidyltransferase (pnp)	-0.7	2.1	18	3.0E-03
P0A7D4	Adenylosuccinate synthetase (purA)	-0.7	2.0	47	1.0E-03

Table 10 (continued): Down-regulated proteins in S30^{+D}-S$^{1-3}$ relative to S30$^{+D(1)}$ as determined by QPPC.

UniProt #	Description (Gene ID)	Ratio (log_e)	Ratio^{-1} (L/H)	# obs.	p-Value
P07012	Peptide chain release factor 2 (prfB)	-0.7	2.0	7	2.7E -02
P0ABH7	Citrate synthase (gltA)	-0.7	2.0	11	1.1E -02
P04079	GMP synthase [glutamine-hydrolyzing] (guaA)	-0.7	1.9	35	1.0E -03
P25519	GTPase (hflX)	-0.6	1.9	8	3.2E -02
P06612	DNA topoisomerase 1 (topA)	-0.6	1.9	10	1.5E -02
P0AB91	Phospho-2-dehydro-3-deoxyheptonate aldolase, Phe-sensitive (aroG)	-0.6	1.9	6	4.9E -02
P00582	DNA polymerase I (polA)	-0.6	1.9	7	3.2E -02
P62620	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin) (ispG)	-0.6	1.8	9	3.2E -02
P0ADG7	Inosine-5-monophosphate dehydrogenase (guaB)	-0.6	1.8	79	1.0E -03
P37769	2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (kduD)	-0.5	1.7	10	3.5E -02
P09831	Glutamate synthase [NADPH] large chain (gltB)	-0.5	1.7	22	2.0E -03
P08200	Isocitrate dehydrogenase [NADP] (icd)	-0.5	1.6	37	1.0E -03
P22259	Phosphoenolpyruvate carboxykinase [ATP] (pckA)	-0.5	1.6	14	2.5E -02
P61889	Malate dehydrogenase (mdh)	-0.5	1.6	12	4.3E -02

The down-regulated proteins were assigned to several GO terms/KEGG pathways with oversignificant pV, as depicted in Figure 14, A, and were functionally grouped based on their kappa scores to five major groups (pie chart in Figure 14, B). Overview terms were chosen based on the most significant GO term assigned in Figure 14, A.

The network visualization in Figure 14, C shows the individual gene products assigned to the individual terms, wherein only the most significant terms are labelled (compare Figure 14, B). The edges demonstrate furthermore the high connectivity and high number of shared proteins between the terms. Only the term RNA degradation is separated from the other terms.

In summary, it appears that mostly biosynthetic pathways for nucleotides (IMP metabolic process) and amino acids (cellular amino acid metabolic process) as well as energy generative pathways such as citrate cycle, glyoxylate and dicarboxylate metabolism are down-regulated in S30^{+D}-S$^{1-3}$ in comparison to S30$^{+D(1)}$. These pathways may negatively affect the supply of energy during CFPS reactions. In contrast, the downregulation of the RNA degradation machinery in these extracts should enhance stability of *e.g.* T7 generated mRNA during CFPS.

The 27 up-regulated proteins listed in Table 11 and the proteins uniquely identified in S30^{+D}-S$^{1-3}$ previously assigned to the term response to temperature stimulus, (dtd, hflC, hflK, pspA, raiA, and yajL), were also functionally integrated. The same functional analysis, including the parameters for selection for down-regulated proteins was performed, except that only the ontology source BiologicalProcess-GOA (as downloaded on 18th of November 2016) was used.

Table 11: Up-regulated proteins in $S30^{+D}$-S^{I-3} relative to $S30^{+D(I)}$ as determined by QPPC.

UniProt #	Description (Gene ID)	Ratio (log_e)	Ratio (L/II)	# obs.	p-Value
P0A7E5	CTP synthase (pyrG)	0.4	1.5	20	1.6E-02
P0A6Z1	Chaperone protein (hscA)	0.7	1.9	10	1.4E-02
P0A988	DNA polymerase III subunit beta (dnaN)	0.7	2.1	6	3.0E-02
P22523	Chromosome partition protein (mukB)	0.7	2.1	12	3.0E-03
P0A9M0	Lon protease (lon)	0.9	2.3	6	2.0E-02
P32132	GTP-binding protein TypA/BipA (typA)	0.9	2.5	4	2.4E-02
P0A8F0	Uracil phosphoribosyltransferase (upp)	0.9	2.5	7	6.0E-03
P08244	Orotidine 5-phosphate decarboxylase (pyrF)	1.0	2.7	4	2.0E-02
P77718	tRNA sulfurtransferase (thiI)	1.0	2.8	10	1.0E-03
P07639	3-dehydroquinate synthase (aroB)	1.0	2.9	3	2.8E-02
P45577	RNA chaperone (proQ)	1.1	3.0	4	1.1E-02
P25745	tRNA-specific 2-thiouridylase (mnmA)	1.1	3.0	3	2.6E-02
P0A6F5	60 kDa chaperonin (groL)	1.3	3.7	110	1.0E-03
P0A6Y8	Chaperone protein (dnaK)	1.3	3.8	104	1.0E-03
P69797	PTS system mannose-specific EIIAB component (manX)	1.3	3.8	3	1.3E-02
P09372	Protein (grpE)	1.5	4.6	10	1.0E-03
P22106	Asparagine synthetase B [glutamine-hydrolyzing] (asnB)	1.5	4.6	2	1.9E-02
P0A9W3	Energy-dependent translational throttle protein (ettA)	1.7	5.5	6	1.0E-03
P69811	Multiphosphoryl transfer protein (fruB)	1.9	6.9	2	5.0E-03
P63284	Chaperone protein (clpB)	2.0	7.3	45	1.0E-03
P10408	Protein translocase subunit (secA)	2.3	9.7	2	2.0E-03
P0A6Z3	Chaperone protein (htpG)	2.6	13.2	44	1.0E-03
P0A6F9	10 kDa chaperonin (groS)	2.6	13.3	3	1.0E-03
P0ACG8	Heat shock protein 15 (hslR)	2.7	14.9	6	1.0E-03
P0AG30	Transcription termination factor (rho)	2.8	16.5	14	1.0E-03
P0AGB3	RNA polymerase sigma factor (rpoH)	3.1	21.3	5	1.0E-03
P77754	Periplasmic chaperone (spy)	3.3	26.9	2	1.0E-03

All 33 proteins were functionally annotated in the selected ontology source, and 25 (=75.76%) of them were represented by GO-terms after applying the selection criteria. The representative GO terms including the number/percentage of assigned Gene IDs per term are illustrated as bar diagram in Figure 15, A, and the major GO terms are grouped in the pie chart (Figure 15, B) with the most significant GO term used as label for the group term. The assigned Gene IDs to individual GO terms including their connectivity are illustrated as network in Figure 15, C. According to the analysis (and as expected after induction of a heat shock response), mostly proteins associated with response to temperature stimulus and the closely associated GO term chaperone mediated proteins folding were among the most significant. In particular, the chaperones GroS, GroL, DnaK and Spy

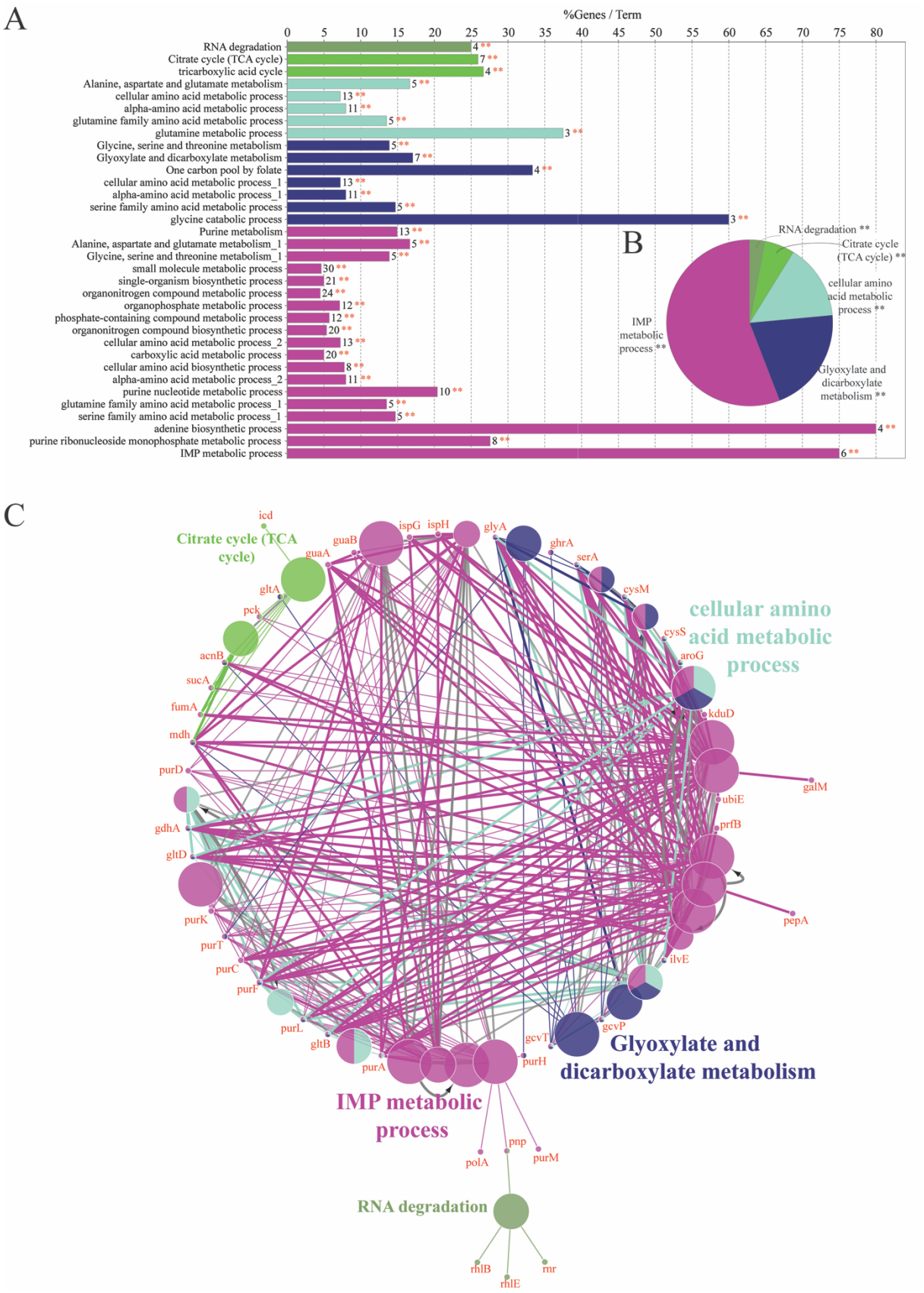


Figure 14: ***Functional analysis of down-regulated proteins in S30^{+D}-S^{I-3} relative to S30$^{+D(1)}$ using the Cytoscape plugin ClueGO based on the proteins listed in Table 10.*** *(A) Assigned GO terms and KEGG terms including the percentage of identified proteins per term. The number behind the bar represents the absolute number of proteins assigned per term and the ** indicates an oversignificant term pV (<0.001). (B) Summary of terms to functional groups, where the most significant GO term defines the group term. The double asterisk ** indicates an oversignificant group pV (<0.001). (C) GO terms and assigned gene products with most significant GO terms as nodes and the edges represent the kappa score or connectivity between pathways (≥0.4). The node size represents significance and the node is colored according to the functional grouping as shown in (A and B) only the label of the most significant term is shown.*

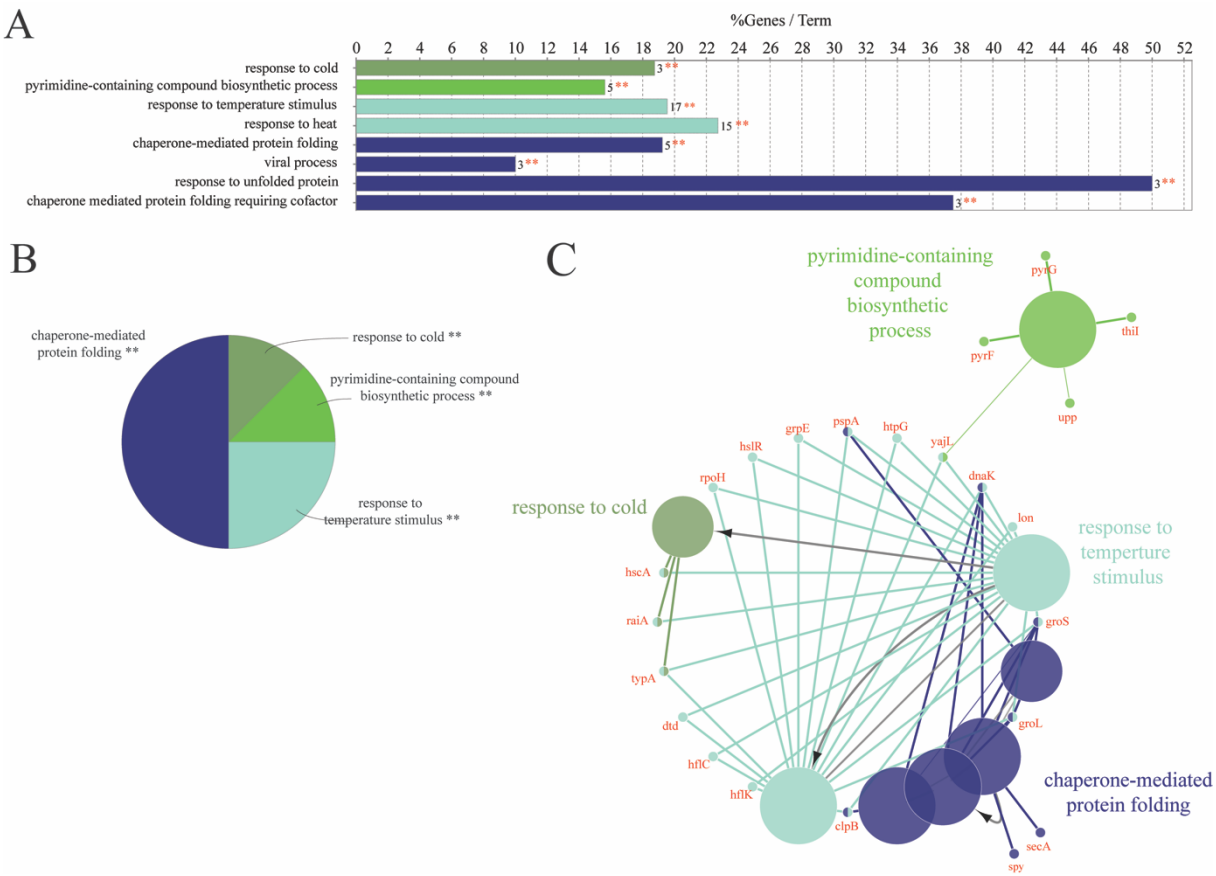


Figure 15: ***Functional analysis of up-regulated proteins in S30^{+D}-S$^{1-3}$ relative to S30$^{+D(1)}$ using the Cytoscape plugin ClueGO based on the proteins listed in Table 11.*** *(A) Assigned GO terms and including the percentage of identified proteins per term. The number behind the bar represents the absolute number of proteins assigned per term and the ** indicates an oversignificant term pV (<0.001). (B) Summary of terms to functional groups, where the most significant GO term defines the group term. The double asterisk ** indicates an oversignificant group pV (<0.001). (C) GO terms and assigned gene products with GO terms as nodes and the edges represent the kappa score or connectivity between pathways (≥0.4). The node size represents significance and the node is colored according to the functional grouping as shown in (A and B) and only the label of the most significant term is shown.*

were upregulated in S30^{+D}-S$^{1-3}$ relative to S30$^{+D(1)}$ between 3-30 fold (compare Table 11). Thus, chaperones and other proteins induced in response to unfolded proteins (*e.g.* proteases) are significantly more abundant in S30^{+D}-S$^{1-3}$ extracts and may provide a beneficial folding environment for the expression of difficult-to-express proteins.

4.2.2 Performance of Heatshock Extracts vs. Standard Extracts

Having now a precise picture of the regulated proteins in S30^{+D}-S$^{1-3}$ extracts, the impact of changes in the proteome was evaluated in CFPS reactions with difficult-to-express proteins. As shown in previously in Figure 9, C, the yield in S30^{+D}-S$^{1-3}$ was reduced by 30% as compared to standard S30$^{+D(1-4)}$, when using the standard reporter sGFP. However, the increase in chaperones, among other proteome changes in S30^{+D}-S$^{1-3}$ extract, might be beneficial to the solubility and folding of certain chaperone-dependent target proteins.

In order to test that, we analyzed the solubility and specific activity of two challenging target proteins, namely apiRBP and GNA-1. Both proteins were synthesized in frame with a C-terminal fusion protein (superfolderGFP, sGFP) to be used as a quantitative reporter of gene

expression. The apiRBP protein is a 15 kDa protein with putative RNA binding domain and located inside the apicoplast, an essential organelle surrounded by 4 membranes, crucial for the viability of the malaria-causing agent *Plasmodium vivax*. The protein was synthesized in high amounts (1-1.5 mg mL^{-1}) in cell-free systems using standard S30 extracts.

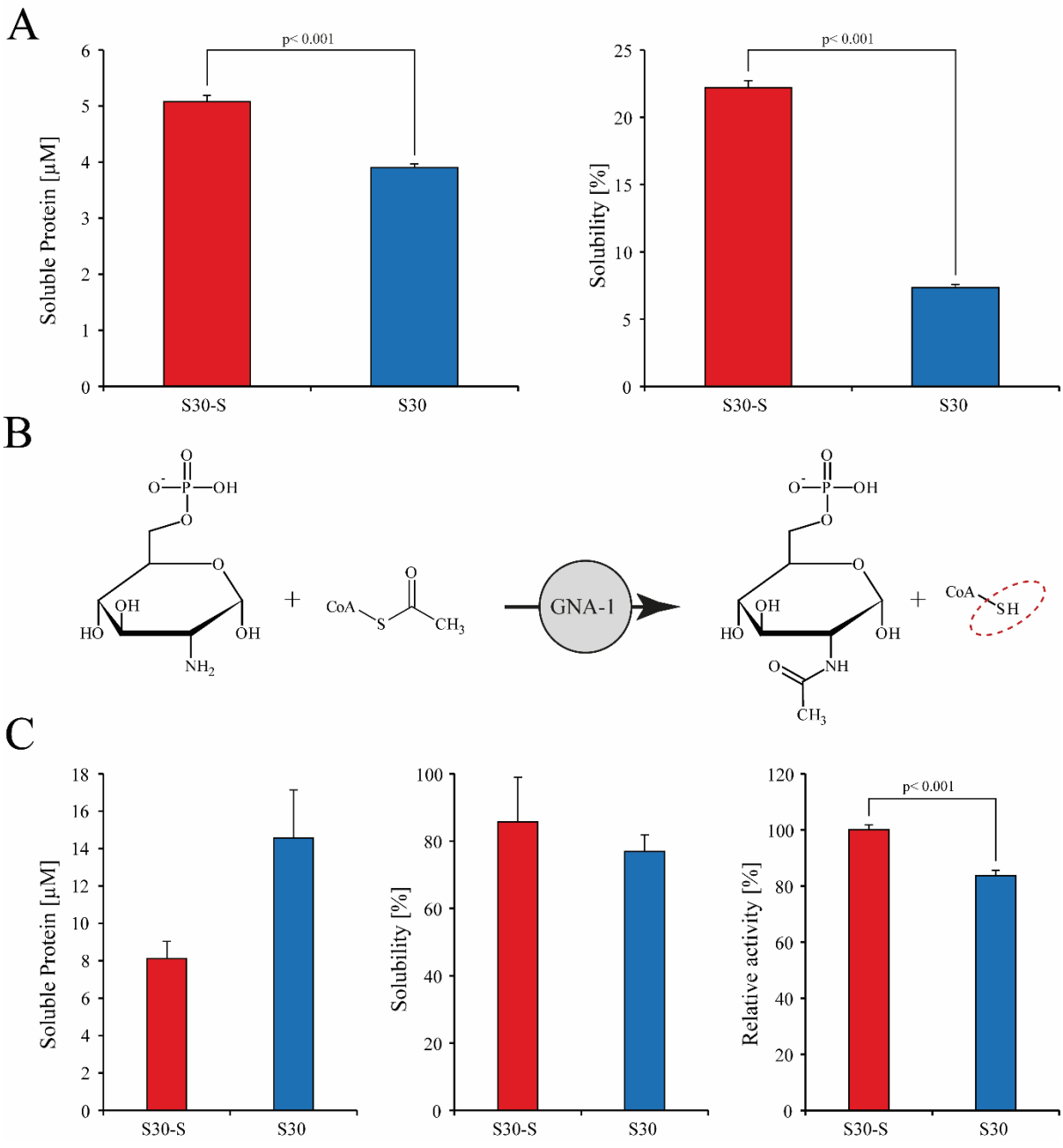


Figure 16: ***Production of difficult-to-express proteins in chaperone enriched S30^{+D}-S$^{1-3}$ (short S30-S) vs. standard S30$^{+D(1-4)}$ (short S30) extracts.*** *(A) Expression of ApiRBP-sGFP. Left: Soluble ApiRBP-sGFP in reaction supernatants as determined using sGFP as quantitative reporter; right: Soluble yield of synthesized ApiRBP-sGFP in S30-S and standard S30 extracts. (B) Reaction catalyzed by hGNA-1. The acetyl group of AcCoA is transferred to GlcN6P to form GlcNAc-6P. Released free thiol groups are quantified by DTNB. (C) Expression of hGNA-1-sGFP. Left: Soluble hGNA-1-sGFP in reaction supernatants as determined using sGFP as a reporter; middle: Total solubility of synthesized hGNA-1-sGFP in S30-S and standard S30 extracts; right: Normalized enzymatic activity of hGNA-1 synthesized either in standard S30 or in S30-S extract. Values of the S30-S sample are set to 100%. The error bar in (A and C) represent the standard deviation of at least three independent measurements (experiments performed by Dr. Erik Henrich at Goethe University Frankfurt).*

However, it remained mostly unfolded and precipitated, leaving only about 100 µg mL^{-1} (sGFP) or 3.7 µM of presumably native protein in the supernatant. Synthesis of apiRBP in $S30^{+D}$-S^{1-3} extracts increased the yield of soluble protein by approximately 30% , reaching 140 µg mL^{-1} (sGFP) or 5 µM. The fraction of soluble apiRBP compared to the total protein yield (soluble and insoluble) was increased 3-fold from 7% in standard S30 extracts to 22% in S30-S extracts (Figure 16, A).

As a second example, we expressed the human enzyme GNA-1, a member of the GNAT superfamily. This enzyme catalyzes the transfer of the acetyl moiety from acetyl coenzyme A to the primary amine group of D-glucosamine-6-phosphate (Figure 16, B) [113]. This protein exhibits a stable expression of around 0.5 mg/mL in a standard CFPS system. In contrast to apiRBP, hGNA-1 is approximately 70-80% soluble when expressed cell-free with standard S30 extract [98]. The yield of hGNA-1 in $S30^{+D}$-S^{1-3} extracts was approximately half of that in standard $S30^{+D(1-4)}$ extracts, and the soluble fraction remained unaltered when expressed in $S30^{+D}$-S^{1-3} extract. Strikingly, the specific enzymatic activity of hGNA-1 was significantly increased in $S30^{+D}$-S^{1-3} lysates, suggesting that larger fraction of hGNA-1 was natively folded in $S30^{+D}$-S^{1-3} extract (Figure 16, C). This implies that at least 50% of hGNA-1 expressed in standard $S30^{+D(1-4)}$ extract, despite being soluble, was actually inactive. Thus, chaperone-enriched lysate can provide excellent means to improve product quality of particularly challenging proteins.

4.3 Genome Editing of *Escherichia coli* A19

The comprehensive blueprint of S30 extract derived from *Escherichia coli* A19 served as a base to design an optimized *E. coli* A19 for cell-free protein synthesis. This work started in collaboration with RiNA GmbH (Berlin, Germany) and later continued with Biotechrabbit GmbH by (Berlin, Germany, with Dr. Micheal Gerrits and Dr. Helmut Merk) in the course of the BMBF funded project “Optimierung von *E. coli*-Stämmen zur Lysatherstellung für die zellfreie Proteinsynthese (OPTEC)”. RiNA and Biotechrabbit offered cell-free protein expression services and have substantial experience in synthesizing challenging proteins for their customers. Based on their previous experience with challenging proteins, and using the proteomics framework in conjunction with information published in current literature, several target proteins were chosen to rationally design and genetically engineer an improved *E. coli* A19, better suited for cell-free protein synthesis. The target proteins as well as the rationale for choosing them are described below.

4.3.1 Selection of Target Genes and Target Strains

E. coli strains A19 and D10 were chosen as target strains. *E. coli* A19 is commonly used for S30 extracts and was also the basis for the S30 extract proteome analysis described in 4.1, and 4.2. *E. coli* D10 was provided by Biotechrabbit GmbH and is a derivative of *E. coli* D10 (CGSC#: 6587) from "The Coli Genetic Stock Center" (Yale University, New Haven, USA).

As shown in 4.1 and 4.2, several proteins were identified that are considered to negatively affect CFPS reaction in a variety of ways. These unwanted effects can be categorized as follows: (*i*) effects on precursor stability (*e.g.* unproductive pathways degrading amino acids or ATP), (*ii*) effects on template DNA and mRNA transcripts (*e.g.* endonucleases, endoribonucleases), (*iii*) effects on the integrity of target protein as well as the proteins of the S30 extract (*e.g.* proteases). From each category, one or more genes were chosen to be addressed in the genome editing approach.

For (*i*), many enzymes present in the extract affect precursor stability, including all components for energy regeneration (*e.g.* NTPs, Phosphoenolpyruvate), and amino acids that are quickly degraded by non-productive pathways. It has been shown that the amino acids arginine, serine, cysteine and to a lesser extend tryptophan are quickly degraded by SpeA, SdaAB, GshA and TnaA, respectively [44, 45]. Due to degradation, these amino acids are no longer available as a building blocks for target proteins and thus limit the synthesis yield. The enzymes SpeA, SdaA, TnaA and GshA were found repeatedly in S30 extracts based on *E. coli* A19 (Table 8), albeit with relatively low emPAI values of 1.3, 0.6, 0.9 and 1.1, respectively. However, depending on their catalytic activity, a relatively small amount of these enzymes may result in quick degradation of precursors and therefore limit the productivity.

In category (*ii*), the endonuclease A (EndA) was selected for threatening the integrity of template DNA and the major RNase E (*rne*) was selected for degrading mRNA derived from endogenous and T7 RNAP. EndA has been shown to nick DNA plasmid during plasmid preparation procedures [114], and RNase E initiates mRNA decay in bacteria [115]. In the proteome analysis, only RNase E was repeatedly identified, albeit with a low emPAI value of 0.5 (Table S1), whereas EndA has not been identified at all.

In category (*iii*), the proteases Lon and OmpT were selected as targets for genome editing as well as gluthatione reductase (gor) was selected. OmpT is an outer membrane protease that has been knocked out in many *E. coli* strains (*e.g.* BL21(DE3)) to improve heterologous protein expression. Even though OmpT was not identified in S30 extract, suggesting that this outer membrane protein is depleted during extract preparation, it was chosen because of numerous positive reports on the use of ompT-deficient *E. coli* strains such as BL21(DE3) for CFPS applications. The beneficial properties of this strain may in part be

attributed to the ompT-deficiency, which may affect both the integrity of the extract during preparation and the integrity of the target protein during cell-free protein synthesis. In contrast, the protease Lon was found repeatedly with an emPAI value of approximately 1.2, and since this protease is often deleted in conjunction with ompT, it was decided to delete this gene as well, hypothesizing that this may result in a synergic effect.

The removal of the enzyme glutathione reductase (*gor*) has been described to be beneficial in CFPS reactions if disulfide bonds need to be formed. For the formation of disulfide bonds, oxidized and reduced glutathione are empirically optimized and added to the CFPS reaction to facilitate the disulfide bond shuffling and formation in target proteins. Due to the enzymatic activity of the glutathione reductase, however, oxidized glutathione is rapidly transformed into reduced glutathione and is therefore not available for disulfide bond formation anymore. For potential disulfide bonded target proteins, the *gor* gene was therefore chosen as a target for genetic modification. The outer membrane receptor FhuA (also known as TonA) was removed to make the genetically modified strain resistant to phage infections (T1, T5, φ80, and UC-1) and therefore avoid potential unwanted genome modifications [116].

The cysteinyl-tRNA synthetase (cysS) and *release factor I* (abbrev. RF1, gene: *prfA*) were modified by a C-terminal in-frame SBP- and SII-tags, respectively, to permit the removal of those enzymes. This allows the usage of amber stop codons (which would normally cause interruption of translation upon binding of RF1 [117, 118]) and codons coding for cysteine for introduction of non-natural amino acids [Helmut Merk and Michael Gerrits, personal communication in 2015].

The strain A19 wt as obtained from *E. coli* Genetic Stock Centre (CGSC# 5997, Yale University, CT, USA) was initially modified by introducing an C-terminal SBP-tag (SBP, streptavidin binding peptide) in-frame with the RNase E (rne) resulting in the source strain named A19 rne::SBP. The presence of an SBP-tag on RNase E does not modify the extract, as long as no removal of RNase E is performed during extract preparation. The A19 rne::SBP served as the source strain for consecutive genetic modification of two lineages of A19 resulting in the final strains A19 rne::SBP Δ(endA/ompT) CSRW and A19 rne::SBP Δ(ompT/lon/gor/fhuA).

The strain D10 as received from RiNA GmbH was the base for one lineage resulting in the final strain D10 rne::SBP/RF1::SII Δ(ompT/lon) and two singly modified strains namely D10 CysS::SBP and D10 CysS::SII.

All genes were knocked out with the Quick & Easy *E. coli* Gene Deletion Kit, based on Red/ET recombination. As described in detail 3.5, cassette overhangs were designed based on available genome information from *E. coli* W3110 (NC_007779.1) and genes were edited

according to the kits' manufacturer instructions. Since the selection marker used to replace the genes was flanked by loxP or FRT sites for site-specific recombination, the selection maker was removed and recycled for consecutive genome modification in the same strain. In order to verify the successful genome modification, the genomic DNA (gDNA) was extracted after each editing cycle and checked for the most recent modification as well as all previous modifications. The previous modifications were each time re-evaluated as they still contained recombination sites (loxP- or FRT-site) after the removal of the selection cassette and were therefore potentially subject to modification by subsequent recombination events.

4.3.2 Genetic modifications in A19 rne::SBPΔ(endA/ompT/lon/gor/fhuA)CSRW

In Figure 17 the analysis of each individual strain is depicted during each genomic modification cycle starting with the original strain A19 wt and ending with the final A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW. On the left-hand side or in the case of original A19 wt (Figure 17, A), the individual gene loci (of endA, rne, gshA a.s.o.) were amplified using primers binding approximately 500 bp down- and upstream of each individual target gene and the resulting PCR products were analysed using agarose gel electrophoresis. In contrast, on the right-hand side or in case of the original A19 wt (Figure 17, B) each individual gene (endA, rne, gshA a.s.o.) was amplified from gDNA, meaning that the primer bound within the target gene. Again, the PCR products were analyzed using agarose gelelectrophoresis. Therefore, the amplification up- and downstream of the target gene ensures that the target gene is indeed removed from the target locus after each modification cycle, whereas the amplification of the target gene itself confirms that there are no other copies of this very gene present in the genome or that is was transferred to another locus due to the unwanted recombination events.

All target genes (endA, rne, gshA, sdaA, sdaB, speA, tnaA, ompT, lon, gor and fhuA) were present in the original strain (see fig 17 A and B) Figure 17, A and B show the original state of the target locus (A) and also suggests that all target genes are present in the original genome (B) with respect to the target genes (endA, rne, gshA, sdaA, sdaB, speA, tnaA, ompT, lon, gor and fhuA). The PCR amplification of all target genes and loci of A19 wt strain was in agreement with the expected product sizes based on the genome of *E. coli* W3110 (see Table 12). This suggests the targeted loci are well conserved between these two strains and therefore the genome of A19 is well represented by W3110.

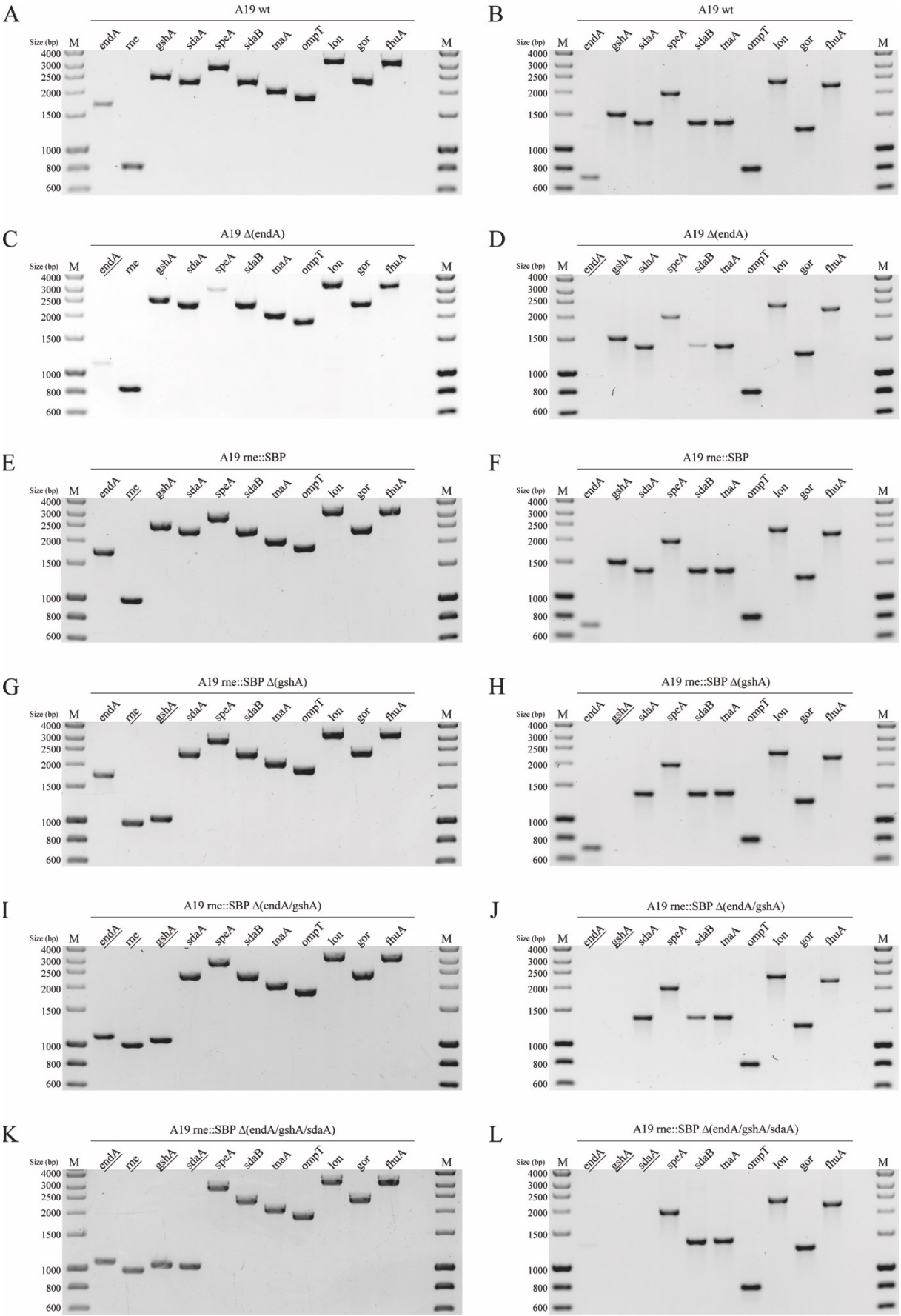


Figure 17: ***Verification of knockouts by agarose gel electrophoresis****. (A,C,E,G,I,K) Genomic DNA (gDNA) isolated from each strain was used as template. Primer binding up- and downstream (~500 bp) of the target locus were used to verify that target gene was successfully removed from original locus. (B,D,F,H,J,L) gDNA from each strain was amplified using primer binding within the target gene to ensure the complete removal of the target gene from the E. coli genome. Edited target genes in each strain are underlined.*

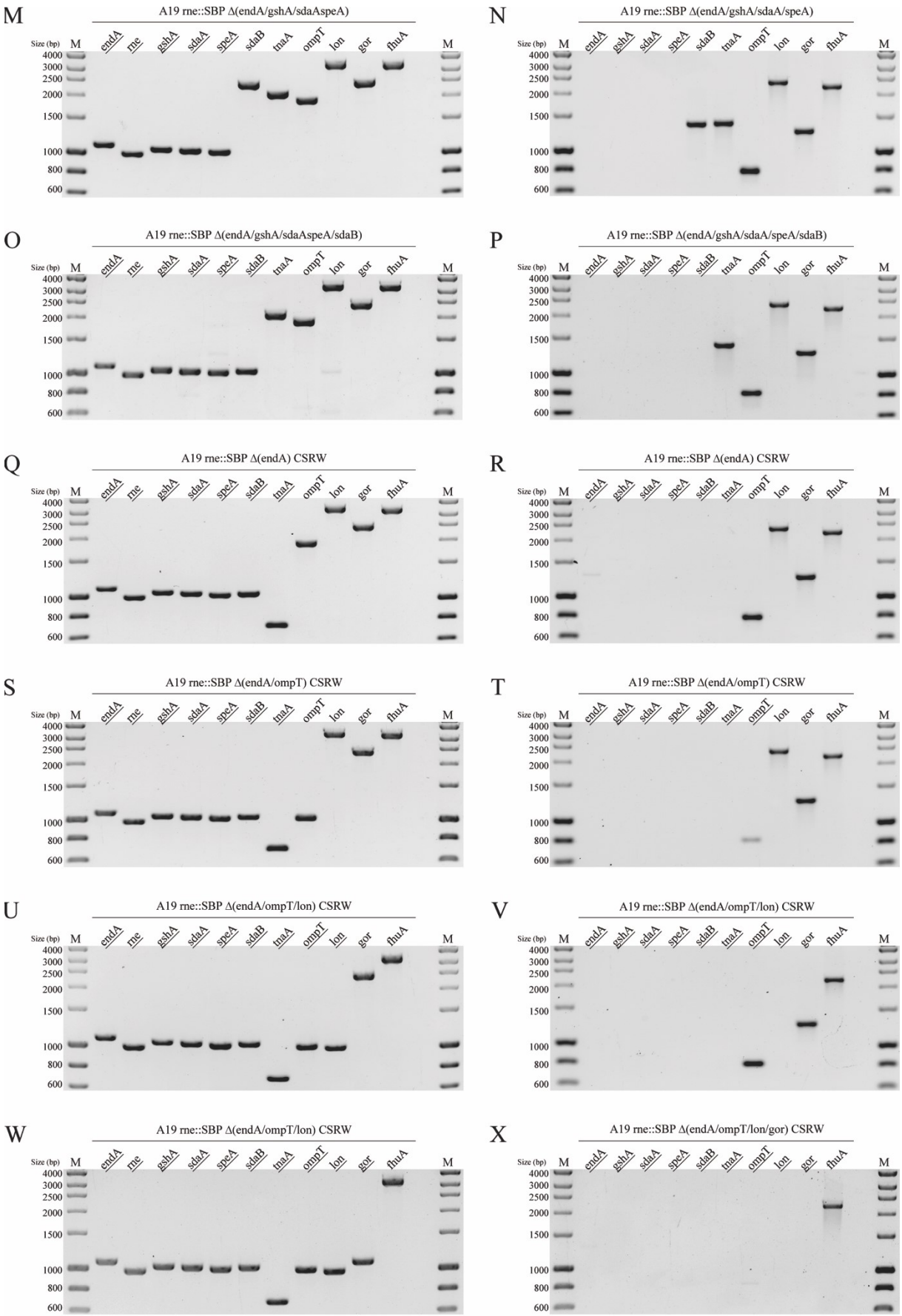


Figure 8 (continued): ***Verification of knockouts by agarose gel electrophoresis****. (M,O,Q,S,U,W) Genomic DNA (gDNA) isolated from each strain was used as template. Primer binding up- and downstream (~500 bp) of the target locus were used to verify that target gene was successfully removed from original locus. (N,P,R,T,V,X) gDNA from each strain was amplified using primer binding within the target gene to ensure the complete removal of the target gene from the E. coli genome. Target genes in each strain are underlined.*

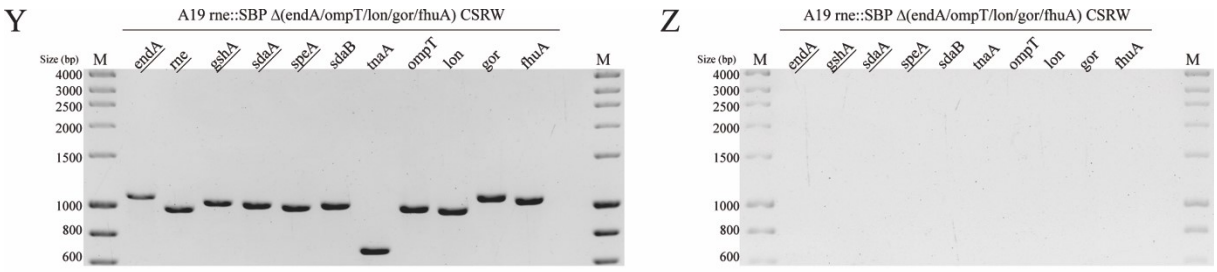


Figure 8 (continued): ***Verification of knockouts by agarose gel electrophoresis****. (Y) Genomic DNA (gDNA) isolated from each strain was used as template. Primer binding up- and downstream (~500 bp) of the target locus were used to verify that target gene was successfully removed from original locus. (Z) gDNA from each strain was amplified using primer binding within the target gene to ensure the complete removal of the target gene from the E. coli genome. Target genes in each strain are underlined.*

In the first modification cycle, endA was knocked out, resulting in the strain A19 Δ(endA). The verification by PCR (see Figure 17, C and D) shows that the band for endA locus (17 C), previously located at 1700 bp in A19 wt shifted to approx. 1050 bp. This suggests that a sequence of approx. 650 bp was removed from this locus, indicating that the knockout was successful. In addition, the amplification of the endA gene results in a 700 bp band in A19 (B), whereas no band in visible for A19 (endA) (D)as seen in D, there is no band visible for the endA gene, whereas in A19 wt the amplification of endA (B) resulted in a band at around 700 bp, thus confirming that there is no other copy present in the genome. The removal of a sequence of 700 bp agrees with size of endA as amplified in B as well as the theoretical size of endA according to *E. coli* W3110 (NC_007779.1) (compare Table 12). The other amplified regions remained unaltered compared to A19 wt, as expected.

Table 12: Theoretically expected PCR product size based on available genome data from E. coli W3110 (NC_007779.1) and self-designed primers listed in Table 3.

Gene Identifier	Amplified Locus	ORF (predicted)	Amplified ORF
endA	1695	708	700
gshA	2507	1557	1507
sdaA	2300	1365	1364
speA	2896	1977	1955
sdaB	2285	1368	1345
tnaA	2036	1416	1373
ompT	1896	954	807
lon	3243	2355	2325
gor	2341	1353	1250
fhuA	3205	2244	2191

The strain A19 rne::SBP (Figure 17, E and F) is based on A19 wt (Figure 17, A and B) and shows the same amplification pattern with regard to endA, gshA, sdaA, sdaB, speA, tnaA, ompT, lon, gor and fhuA, confirming that those genes are unaltered except for rne locus. Since the RNase E (rne) is an essential enzyme for maturation of ribosomal RNA, this enzyme cannot simply be removed. However, in order to be able to remove this enzyme during extract preparation, a SBP (streptavidin binding peptide) was introduced in-frame on the C-terminus before the stop codon extending the original enzyme by 135 bp (SBP tag: 114 bp (38 aminoacids) + 21 bp (7 aminoacids) before the SBP tag). Additionally, 70 bp are introduced

after the stop-codon due to technical circumstances with respect to Red/ET recombination. The tagging is apparent by the migration pattern of the band resulting from amplification of the rne locus. While the band in A19 wt is visible close to the 800 bp marker band (A), the same band migrates slightly below 1000 bp marker in A19 rne::SBP (E) suggesting that locus became approx. 200 bp larger in A19 rne::SBP. This extension is in agreement with the introduction of 134 bp (38 aminoacids) by the SBP tag and 7 additional amino acids as well as 700 bp. Since the exact modification is of pivotal importance the correct modification was furthermore confirmed by sequencing of the PCR product, which confirmed the exact modification of RNase E by an in-frame SBP tag. Additionally, the correct expression of the SBP tag was verified in subsequent experiments.

All consecutive strains are based on A19 rne::SBP, and the target genes were removed in the following order: gshA, endA, sdaA, speA, sdaB, tnaA, ompT, lon, gor and fhuA. The analyses for each strain consistently confirmed the removal of the target gene (left-hand side) and also confirmed for all strains except ompT knockouts that no other copy is present in the genome. The ompT locus was in previous strains (e.g. A19 rne::SBP Δ(endA) CSRW (Figure 17, Q and R) as well as A19 wt (Figure 17, A and B) amplified with a product size of approx. 1900 bp. This agrees with the theoretically expected size of 1896 bp (compare Table 12, identifier ompT).

After modification of the ompT locus (beginning with the strain A19 rne::SBP Δ(endA/ompT) CSRW (see Figure 17, S and T) as well as in all consecutive strains), the corresponding band migrated at around 1000 bp, consistent with the removal of approx. 900 bp and with the ORF-size of 954 bp as predicted in reference genome of *E. coli* W3110 (NC_007779.1). These results indicate that the genome recombination occurred exactly as expected (Figure 17, S). However, after PCR amplification using primers binding within the gene, a product with the size of 800 bp was still detectable (Figure 17, T), suggesting that ompT is still present in another location in the genome. This band was also detectable in consecutive strains, with varying intensity. Sequencing of this PCR product confirmed that it was indeed *ompT* (data not shown) and not a close homolog (*e.g.* another omp gene). It is not clear where the ompT copy is located in the A19 genome and if it was present even in the source strain even before any recombination event, which promoter controls it and whether it is expressed at all. This question will be addressed in subsequent experiments.

4.3.3 Genetic modifications in A19 rne::SBPΔ(ompT/lon/gor/fhuA)

The strain A19 rne::SBP Δ(ompT/lon/gor/fhuA), also based on the previously verified A19 rne::SBP, was similarly analyzed after each individual genetic modification using the same method as described in 4.3.2 by amplifying the locus containing the gene as well as the gene itself to ensure that no other copies are present in the genome. This strain contains only the knockouts of ompT, lon, gor and fhuA and was created to have an additional control to investigate the effects of the removal of ompT, lon and gor in the absence of the multiple amino acid stabilizing genetic modifications.

The verification of the strain A19 rne::SBP Δ(ompT/lon/gor/fhuA) is depicted in Figure 18, whereas A shows the amplification of each individual target locus and B (always exactly below corresponding locus verification for each individual strain) the amplification using primers binding internally of each gene. The first modification that was introduced into A19 rne::SBP was the ompT knockout shown in Figure 18, A under label A19 rne::SBP Δ(ompT). The band representing the ompT locus migrates at approx. 1000 bp consistent with the removal of 954 bp (compare Table 12, ompT) from this locus which was before the removal amplified with approx. 1900 bp (compare Figure 17, A). The corresponding verification of the removal from the genome by amplification using internal primers compare shown in Figure 18, B confirms for the A19 rne::SBP Δ(ompT) and all subsequent strains the complete removal of ompT from the genome.

As opposed to the previously described A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW (Figure 17, T) there is at no point a PCR-product of the internal amplification of ompT detectable, suggesting that there is no additional copy in the genome of A19 wt present and therefore implying that the ompT copy observed in an unidentified location in the genome A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW is due to an recombination event that occurred in the course of the genome modification procedure.

The subsequent introduction of the lon, gor and fhuA knockout in this strain were accompanied by the expected reduction of the PCR product size consistent with corresponding ORF length of lon, gor and fhuA, respectively (compare Figure 18, A and Table 12). The internal amplification of lon, gor and fhuA yielded no detectable PCR product after successful removal of each respective gene from its locus confirming that no additional copies of the genes are present in the genome (compare corresponding analytical gels shown in Figure 18, B). In conclusion, the A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW and all preceding intermediates have been successfully verified.

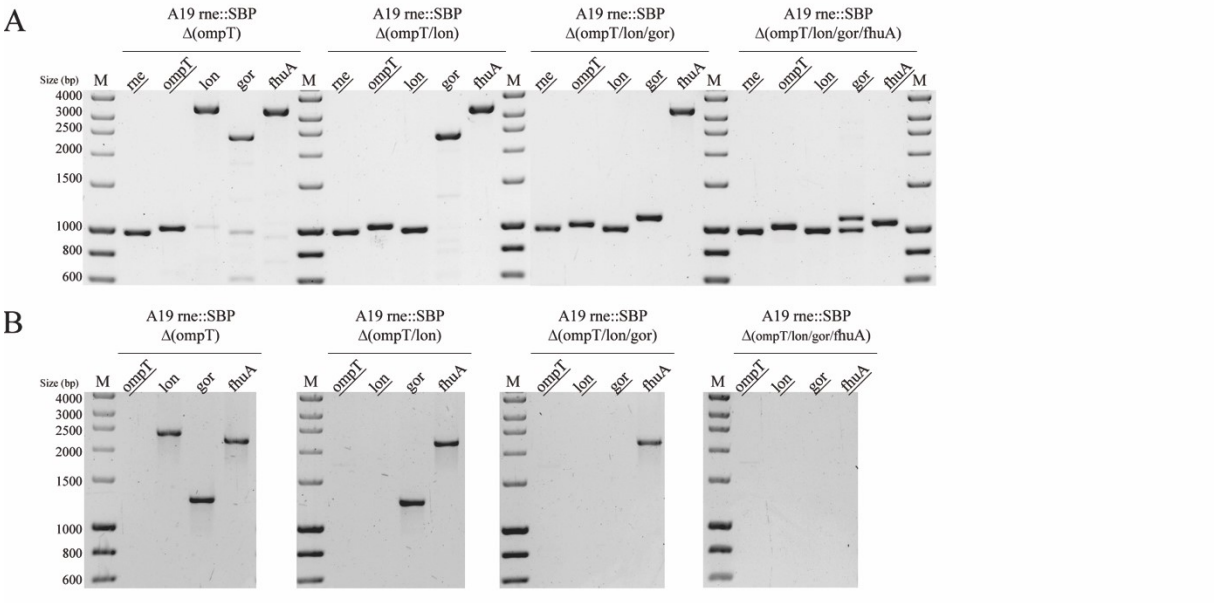


Figure 18: ***Verification of knockouts by agarose gel electrophoresis.*** *(A) Genomic DNA (gDNA) isolated from each strain was used as template. Primer binding up- and downstream (~500 bp) of the target locus were used to verify that target gene was successfully removed from original locus. (B) gDNA from each strain was amplified using primer binding within the target gene to ensure the complete removal of the target gene from the E. coli genome. Target genes in each strain are underlined.*

4.3.4 Genetic modifications in D10 derivatives

As mentioned previously, the strain *E. coli* D10 was kindly provided by RiNA GmbH (Berlin, Germany). Initially, the genes ompT and lon were one after the other removed from the *E. coli* D10 wt. Then, the gene *rne* was again tagged by introduction of an C-terminal SBP-tag (as described in detail in preceding chapters). Upon request by RiNA GmbH this strain was additionally equipped with an in-frame C-terminal SII-tag on the *prfA* gene to allow the removal of RF1 during S30 extract preparation for special applications such as the introduction of non-natural amino acids at the amber stop codon. Also, for the purpose of introducing non-natural amino acids, based on D10 Δ(ompT/lon), the CysS (cysteinyl-tRNA-synthetase, *cysS*) was modified to carry either an C-terminal SBP- or SII-tag on genome level to allow the removal during extract preparation.

The verification of each individual modification is shown in Figure 19. The amplification of the target loci is shown in Figure 19, A each time labelled with the name of the strain and below the targeted gene name. Following the order of modification, the corresponding analysis of the internal amplification is depicted in Figure 19, B, confirming that the gene is not present in another location in the genome. In contrast to the A19 wt, the amplification of the loci in D10 wt showed an unexpected pattern for *lon* while all the other amplification yielded products as expected based on genomic information of *E. coli* W3110 (compare Table 12 for *ompT* and Figure 17, A for *rne*). The amplification of the *lon* locus,

however, yielded a product of about 4700 bp which is substantially larger than the expected size of 3243 bp based on *E. coli* W3110 primer design and confirmed in A19 wt. Regretably, the attempt to sequence the PCR product was not successful, in part due to repetitive sequences present in this PCR product. Despite the difference to the expected amplification pattern, it was decided to proceed with the knockout of *lon* and *ompT*. As seen in Figure 19, A D10 Δ(ompT), the amplified locus of ompT migrates at around 1000 bp suggesting the removal of approx. 900 bp from this locus in comparison with Figure 19, A D10 wt and in agreement with the size of the ORF of ompT (compare Table 12) as well as comparable to the removal of *ompT* in A19 (Figure 17, Q vs. S). While ompT could be amplified using the internal primers in D10 wt there is no product for ompT detectable in D10 Δ(ompT) (Figure 19, B), suggesting the absence of ompT in the genome of D10 Δ(ompT) after genome editing. While the amplified lon locus migrated rather unexpected at around 4700 bp, the removal of lon was nevertheless possible. Upon removal of lon shown in Figure 19, A D10 Δ(ompT/lon), the amplification of the lon locus using the external primers resulted in a PCR product migrating at 2400 bp corresponding to the removal of approx. 2300 bp. This is consistent with the predicted ORF size for lon of 2325 bp (Table 12), suggesting that lon was removed from the target locus.

Additionally, the amplification using the internal primer in the corresponding analysis yielded no product Figure 19, B D10 Δ(ompT/lon) as compared to D10 Δ(ompT), where the amplification products is detectable, thus, confirming that lon was removed from the target locus as well as no other copy is present in the genome. The tagging, namely rne::SBP, RF1::SII, CysS::SBP and CysS::SII, was initially analyzed by amplification of the target locus.

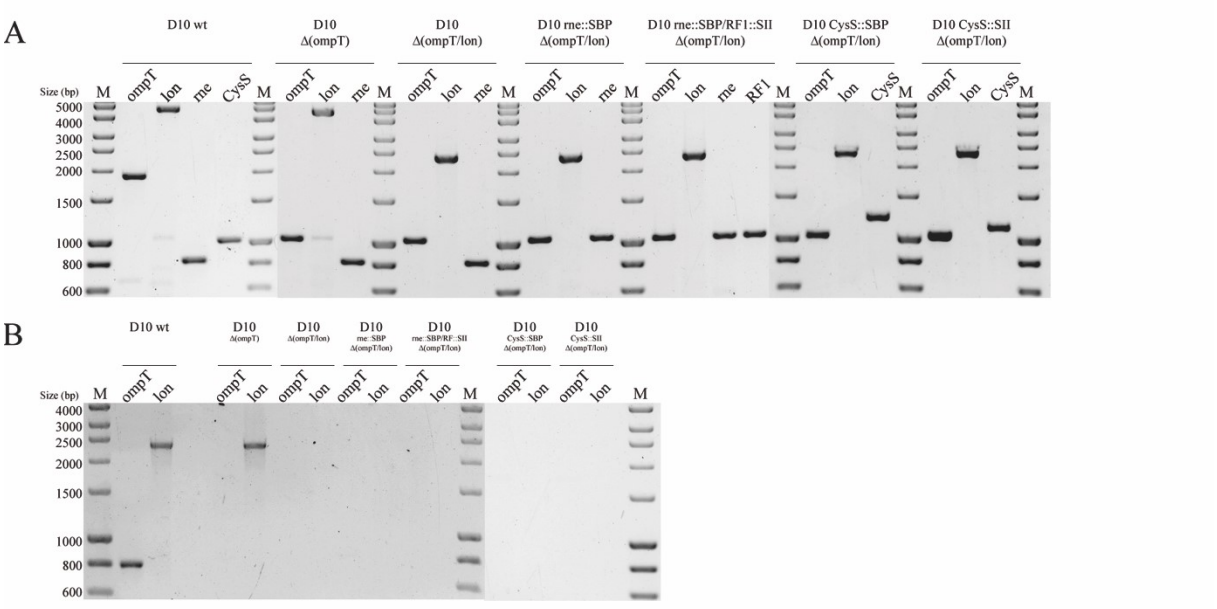


Figure 19: ***Verification of knockouts by agarose gel electrophoresis.*** *(A) Genomic DNA (gDNA) isolated from each strain was used as template. Primer binding up- and downstream (~500 bp) of the target locus were used to verify that target gene was successfully removed from original locus. (B) gDNA from each strain was amplified using primer binding within the target gene to ensure the complete removal of the target gene from the E. coli genome. Target genes in each strain are underlined.*

The tagging of RNase E with SBP was performed as previously described for A19 derivates, whereas the successful integration of the tag is indicated by shift of the amplified target locus from approx. 800 bp (Figure 19, A D10 wt) to approx. 1000 bp (Figure 19, A D10 rne::SBP Δ(ompT/lon)) consistent with the introduction of approx. 200 bp. Similarly, the amplified locus of cysS after introduction of either SII or SBP tag are shifted to approx. 1250 bp (Figure 19, A D10 CysS::SBP Δ(ompT/lon)) and 1100 bp (Figure 19, A D10 CysS::SBP Δ(ompT/lon)), respectively, as compared to D10 wt (Figure 19, A D10 wt). The increase in locus size is in agreement with theoretically expected increase after introduction of either a SBP-tag (total of 207 bp: 7 amino acids in-frame linker (21 bp), 38 amino acids in-frame SBP-tag (114 bp) and additional 69 bp due to technical circumstances by Red/ET recombination) or a SII-tag (total of 114 bp: 7 amino acids in-frame linker (21 bp), 8 amino acids in-frame SII-tag (24 bp) and additional 69 bp due to technical circumstances by Red/ET recombination). For the introduction of the SII tag in-frame onto RF1, there is no reference amplification in D10 wt available. However, the amplified locus of migrating at approx. 1000 bp (Figure 19, A D10 rne::SBP/RF1::SII Δ(ompT/lon) is consistent with the theoretically expected size of 1048 bp as predicted based on reference genome of *E. coli* W3110 and therefore suggesting that the introduction of the SII tag on RF1 was successful.

Since the correct integration of in-frame linker and tags are of pivotal importance to avoid any frame-shifts resulting in incorrect expression, the amplified products were additionally sequenced. In the course of sequencing the correct integration of tags was successfully verified for rne::SBP, RF1::SII, CysS::SBP and CysS::SII modifications (data not shown).

4.3.5 Verification of protein Tagging by Western Blot

The integration of tags and their correct expression on protein-level was verified by Western Blot. The Western Blot analysis of strain A19 rne::SBP (#2) and the final A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW (#3) is shown in Figure 20, A. As expected, the non-tagged RNase E shown in lane #1 is not detected by the α-SBP antibody, whereas the RNase E is identified in lane #2 and #3 migrating between 98 kDa and 188 kDa (compare lane #1, marker) consistent with the translated ORF (1106 amino acids, 123,2 kDa) as predicted based on reference genome of *E. coli* W3110 (NC_007779.1) after integration of the additional linker and SBP-tag in-frame with RNase E. Since the RNase E could be detected in the source strain A19 rne::SBP and the final A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW, it was assumed that the SBP-tag is also in all intermediate strains correctly expressed.

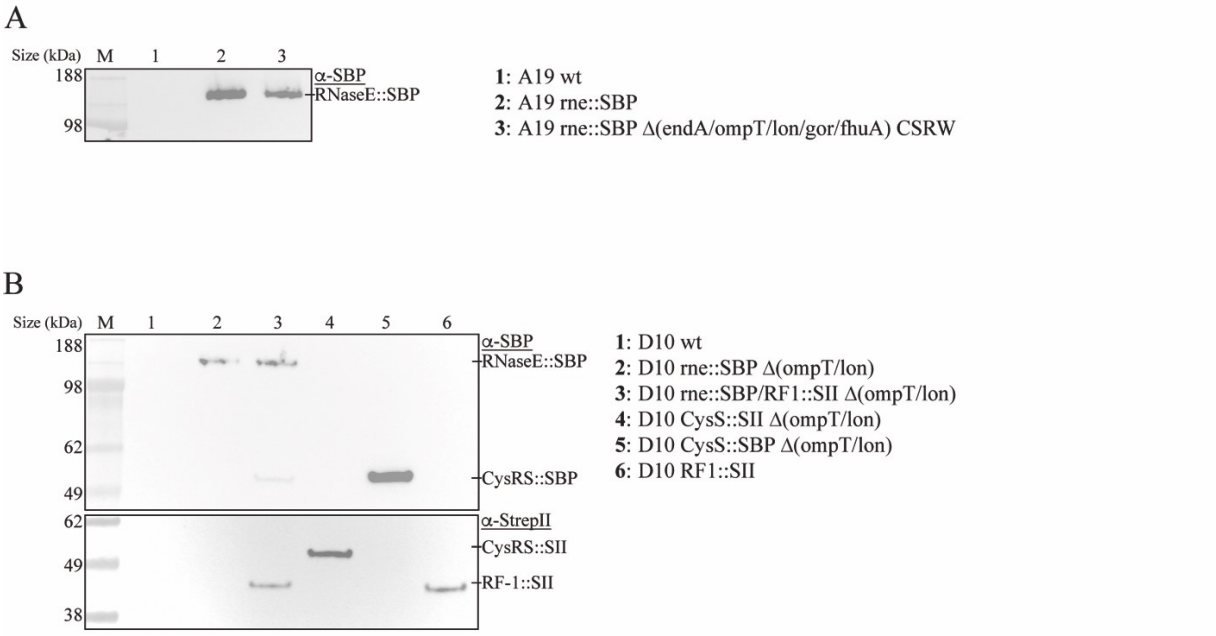


Figure 20: ***Western Blot to verify expression of tags (SBP, Streptavidin Binding Peptide; SII, Strep Tag II) introduced on the C-terminus of target proteins (gene product of rne; cysS; prfA (RF1)).*** *(A) The SBP tag on the C-terminus of RNaseE (rne) was verified using anti-SBP antibody (mouse) followed by detection with a secondary HRP-conjugated anti-mouse antibody in E. coli A19 lysate. (B) SBP tag (upper blot) and SII tag (lower blot) were detected anti-SBP antibody (mouse) and anti-StrepII antibody (mouse), respectively. The primary antibody was detected with secondary HRP-conjugated anti-mouse antibody.*

As for the D10 wt (Figure 20, B #1) and its descendants (#2-6), the same analysis was performed using the α-StrepII antibody. As seen in lane #1, no SII and SBP-tagged proteins are identified in D10 wt lysate, confirming the specificity of the antibodies used. In D10 rne::SBP Δ(ompT/lon) only RNase E was tagged by an SBP-tag, and appears between 98 and 188 kDa, consistent with the predicted molecular weight of 123,2 kDa and with the previous analysis in A19 (Figure 20, A).

After the tagging of RF1 with SII (#3), it was detected with an approx. molecular weight of 45 kDa (predicted 41,8 kDa) by the α-StrepII antibody. The D10 RF1::SII shown in lane #6 was provided by RiNA GmbH (Berlin Germany) and served as a reference for the integration of RF1 and shows the same migration pattern for RF1 in this analysis.

Contrary to the previous analysis of RNase E in #2, there was also a C-terminal breakdown product of RNase E with an approx. molecular weight of 50 kDa, suggesting that RNase E is cut by some site-specific protease present in the lysate. The D10 derivates containing either an SBP- or SII-tagged CysS were analyzed in lane #4 and #5. The band for CysS::SII (#4) and CysS::SBP (#5) migrated between 49 and 62 kDa (compare #1) with CysS::SBP migrating slightly higher. This is in agreement with the predicted molecular weight for CysS::SBP and CysS:SII with 57,2 kDa and 53,9 kDa, respectively.

In summary, the Western Blot analysis confirms that all tags were integrated into the genome as targeted, are expressed correctly, and can be detected by tag-specific antibodies.

There is no indication of frame-shifts. Since RNase E, CysS and PrfA are essential proteins, the fact that the strains can be grown suggests there is no significant impairment due to tagging or knockout.

4.3.6 Influence of Knockouts on Growth

In the next step, the fitness of the genetically modified strains was investigated by growing A19 wt and D10 wt and their derivates under defined conditions and by comparing their growth rates as described in detail in the methods (3.5.3. The growth rate for each strain is plotted as bar diagram for direct comparison of the strains with each other (Figure 21, A, B and C).

In figure Figure 21, A, the strain A19 wt, its intermediates and the final strain A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW were compared. It is apparent that the knockout ΔendA (#1) and the tagging of RNase E (rne::SBP, #3) do not have a significant impact on the growth rate of the carrying strain as compared to A19 wt (#2). The growth rates measured for modified strains (#1 and #3) were all around μ=1.5 h^{-1} and comparable to the growth rate A19 wt (#1). Due to the relationship $\mu=\ln(2)/t_d$ (with t_d, doubling time), the corresponding $t_d \approx 28$ min. That means that during the log-phase A19 wt doubles approx. every 28 min. Only after introduction of the gshA-(glutamate—cysteine ligase-) knockout, the growth rate drops markedly from 1.5 to approx. 1.3 h^{-1}, resulting in an increase of the doubling time of 4 min to 32 min.

The successively introduced knockouts endA, sdaA, speA, sdaB, tnaA and ompT exhibit rather small influence on the growth rate of A19, remaining always above 1.2 h^{-1} , with some small recoveries of the growth rate after knocking out speA and ompT.

A more severe impact on growth rate is apparent after the Lon protease (lon, #11) and Glutathione reductase (gor) were knocked out, with a drop of the growth rate from 1.3 to 1.2 h^{-1} and 1.2 to 1.1 h^{-1}, respectively. In total, the growth rate dropped from 1.5 h^{-1} (A19 wt) to 1.1 h^{-1} (A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW), corresponding to an increase of the doubling time from 28 to 38 min. Most severe effects however, were observed in knockouts affecting the glutathione metabolism (gshA and gor) and the Lon-protease (lon).

In Figure 21, B the A19 rne::SBP lineage and the final strain A19 rne::SBP Δ(ompT/lon/gor/fhuA) without the amino acid stabilizing knockouts “CSRW” were analyzed, providing an opportunity to evaluate the effects of Lon- and Gor-knockouts without the preceding amino acid stabilizing knockouts and their potentially synergic effect on growth rate. While the A19 rne::SBP Δ(ompT) (#14) with ompT-knockout on its own shows a growth rate

comparable to A19 wt of 1.5 h^{-1}, the growth rate falls below 1.3 h^{-1} following knockout of Lon protease (#15), suggesting a major impact of this protease on growth of *E. coli*.

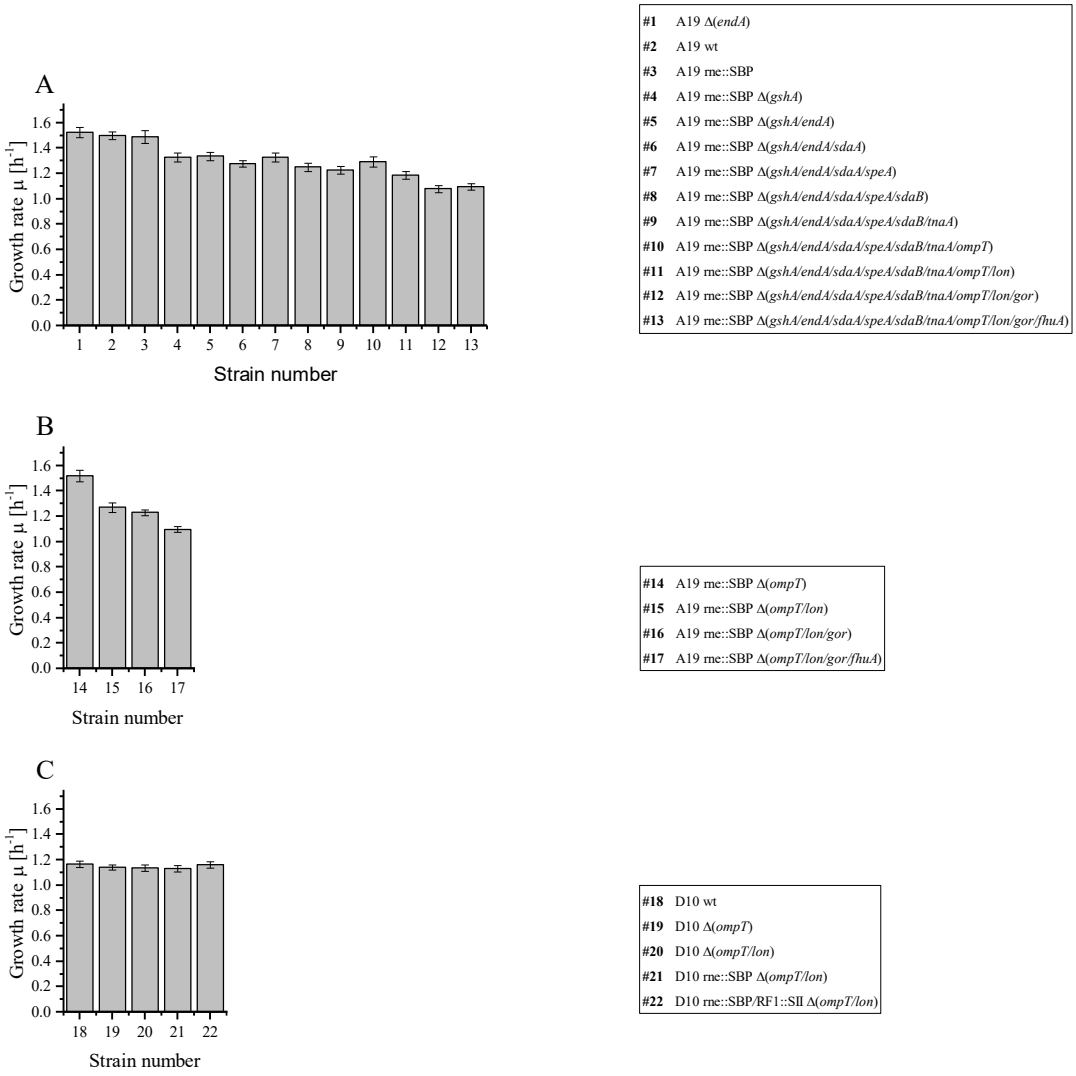


Figure 21: ***Growth rate of various genome edited E. coli A19 and D10 strains cultivated in 2xYTPG medium****. (A and B) growth rate of E. coli A19 after various gene knockouts and modifications. (C) growth rate of E. coli D10 after various gene knockouts and modifications.*

Interestingly, after the following knockout of Gor (#16) the growth rate falls slightly as compared to Figure 21, A (#12), suggesting a synergic effect of the Gor knockout with one of the preceding genetic modifications, causing this larger drop. After introduction of the FhuA knockout (#17), the growth rate drops to 1.1 h^{-1}. The FhuA knockout in strain #13 did not have any effect and seemed to be caused by preceding genetic modifications in this strain.

E. coli D10 (D10 wt #18 in Figure 21, C) showed a growth rate below 1.2 h^{-1} even before any genetic modification was implemented. During successive tagging of RF1 and RNase E, as well as introduction of the ompT- and lon-knockout, no change in the specific growth rate is observed. Overall, the analysis show that all strains are viable and no severe effects on fitness of the corresponding strain is observed for any individual modification or any combination of modifications.

4.3.7 Transcriptional Profiling of Genome Edited *Escherichia coli* A19

To examine the effect of multiple genetic modifications on the overall expression profile and to verify the absence of any expression of the knocked-out genes, a differential expression analysis using Next Generation Sequencing (NGS) was performed. In doing so, the source strain A19 wt and the modified A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW were cultivated under identical conditions in triplicates. At an OD_{600}=3.0-3.3 a sample was taken and flash-frozen using liquid nitrogen. Subsequently, the mRNA was extracted and prepared for NGS analysis and differential expression analysis was performed.

The differential expression analysis is shown Figure 22 as a volcano plot. Genes were considered significantly regulated with a FDR adjusted p-value (q-value) ≤ 0.01 and more that 1.5-fold up- or downregulated. Based on those cut-off, 153 genes are significantly upregulated and 309 genes are downregulated in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt (shown as green dot, partially labelled).

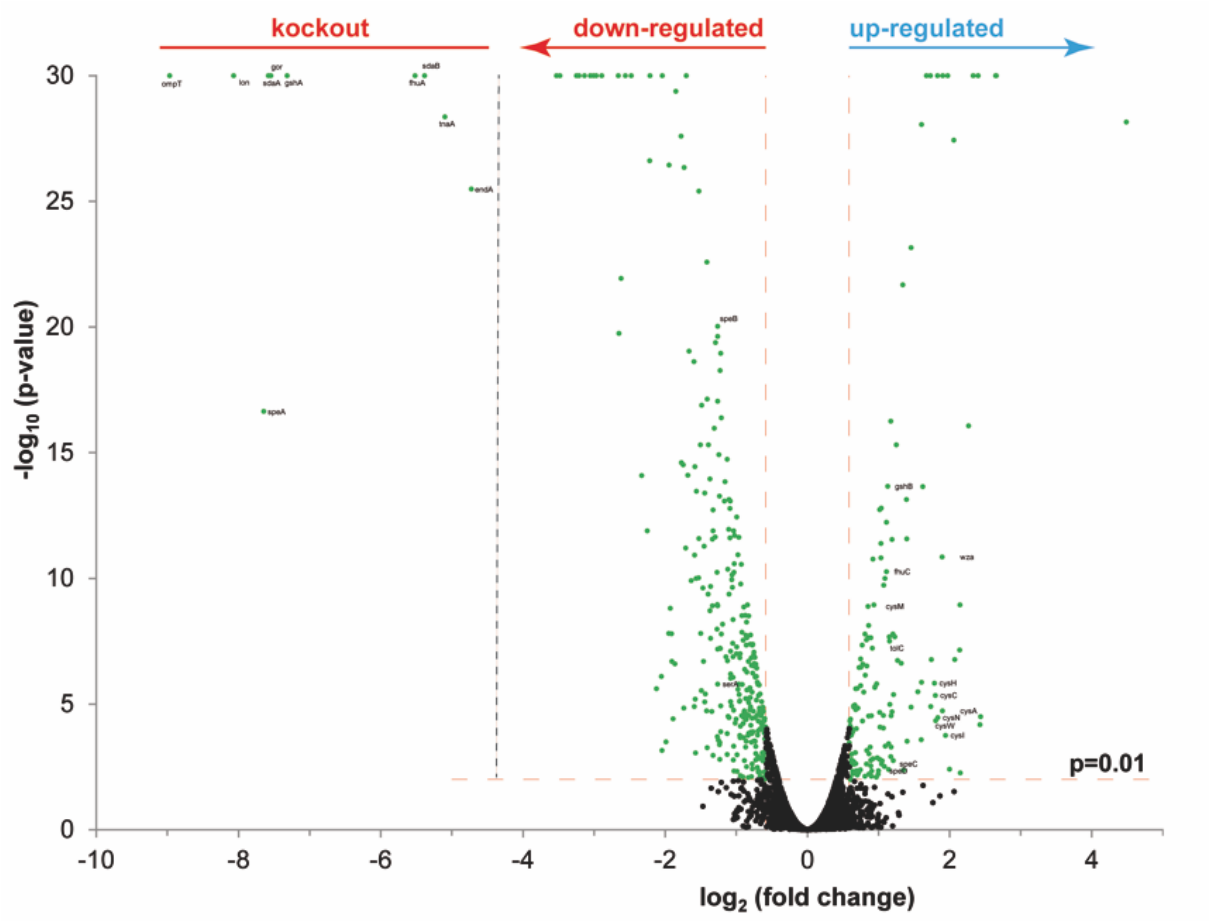


Figure 22: ***Differential Expression Analysis of A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt illustrated as volcano plot****. Only proteins that were showing a fold change ≥1.5 (dashed vertical line, red) and were above significance level (pV≤0.01) (dashed horizontal line) were considered significantly regulated. The dashed vertical line (black) on the left hand side separates genes that were intentionally knocked out in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW and appear therefore as highly downregulated in this analysis.*

Within the analysis, the target genes gshA, endA, sdaA, speA, sdaB, tnaA, ompT, lon, gor and fhuA all appear on the very left side of the volcano plot as highly downregulated genes separated by the dashed vertical line, thus confirming that the removal of all target genes was successful. OmpT also appears in this area, confirming despite being still present in the genome, it is not expressed. This suggests that ompT is in an area that is not transcribed and/or is not under control of a promoter allowing transcription of the gene.

The 153 and 309 co-regulated genes - regulated in response to the numerous genetic modifications - make clear that the knockouts are accompanied by several differences with respect to S30 extract composition. The list of all identified up- and downregulated genes is provided in the Annex (up- and down-regulated genes are listed in Table 15 and Table 16, respectively). The gene listed in those tables, excluding the deliberately introduced genetic modifications (gshA, endA, sdaA, speA, sdaB, tnaA, ompT, lon, gor and fhuA), were functionally integrated using ClueGO v2.5.0 [119] to examine the direct consequences of the genetic modification.

The up- and down-regulated genes were analyzed separately. First, the upregulated genes (153) were analyzed based on the ontology sets GO_BiologicalProcess and and KEGG as downloaded on 21st of February 2018, since those sets resulted in most functional assigments. The following selection criteria were applied: evidence code: all_without_IEA; p-Value cutoff: 0.05 using right-sided hypergeometric test (enrichment combined with Benjamini-Hochberg algorithm for correction of p-values for multiple testing). Of these 153 genes, 105 (68,6%) were functionally annotated in the selected ontologies and 78 (74.3%) were assigned to representative GO-terms and pathways after applying the above-mentioned selection criteria.

The significantly enriched functional groups are shown in Figure 23, A where the 78 genes are categorized into 26 GO-terms or pathways. The number of genes [%] associated with a specific term is illustrated as a bar graph. Functionally related terms (based on kappa score of ≥0.4) are shown in the same color. Based on the functional relationship between those terms, the genes are categorized into twelve major groups using the most significant term as the leading term (Figure 23, B). All associated 78 genes are assigned to their main group.

In Figure 23 C, the functionally grouped network is shown with links based on their kappa score (≥0.4). The node size indicates the enrichment significance (the larger the node, the more significantly enriched is the associated term). The most significant terms of each functional group (leading term) is labelled in color and the subordinated terms are labelled in black. The percentage of genes per term are illustrated within the nodes as pie

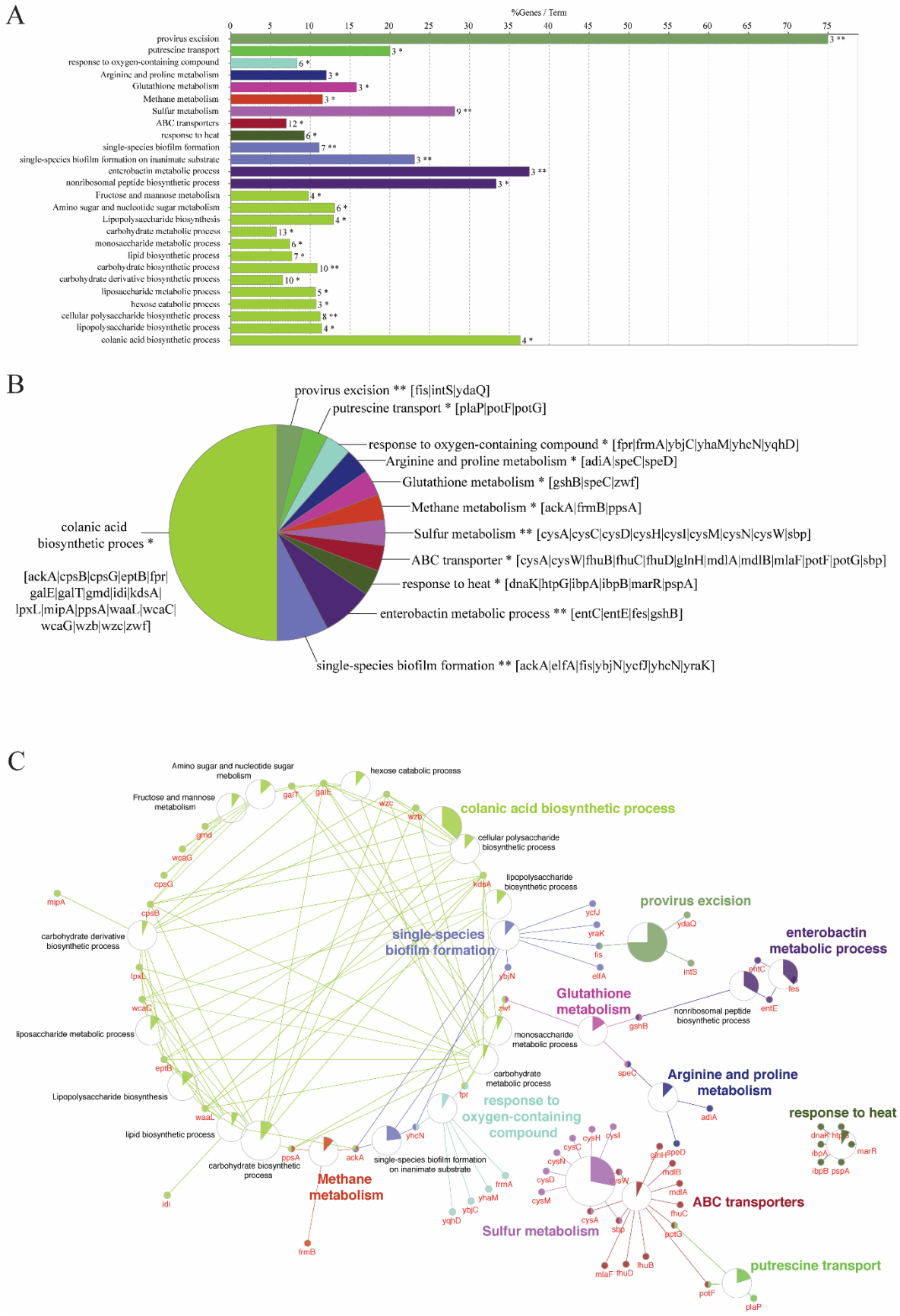


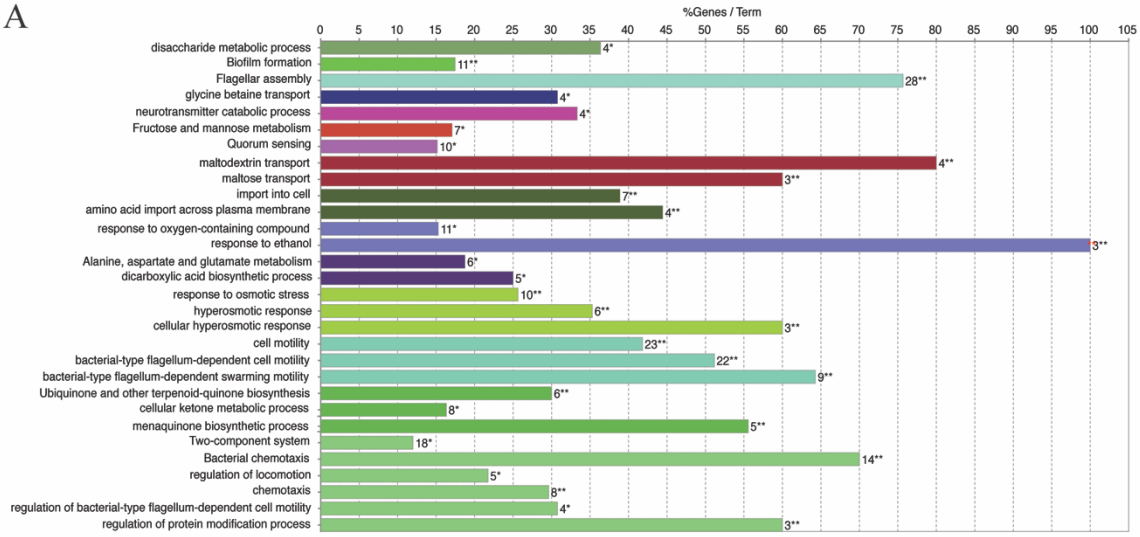
Figure 23: ***Functional analysis of up-regulated genes in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt using the software tool ClueGO.*** *(A) significantly enriched GO-terms and pathways categorized in 26 groups indicating the percentage of genes present per term. (B) 26 categories from (A) summarized into 12 terms based on their functional relationship based on kappa score (≥0.4). (C) Functionally integrated network, where node size represents the term significance linked based on their kappa score. Functionally related groups partially overlap.*

chart. Genes that were functionally grouped were also listed with their repsctive UniProt #, description and regulation factor in Table 13 (below).

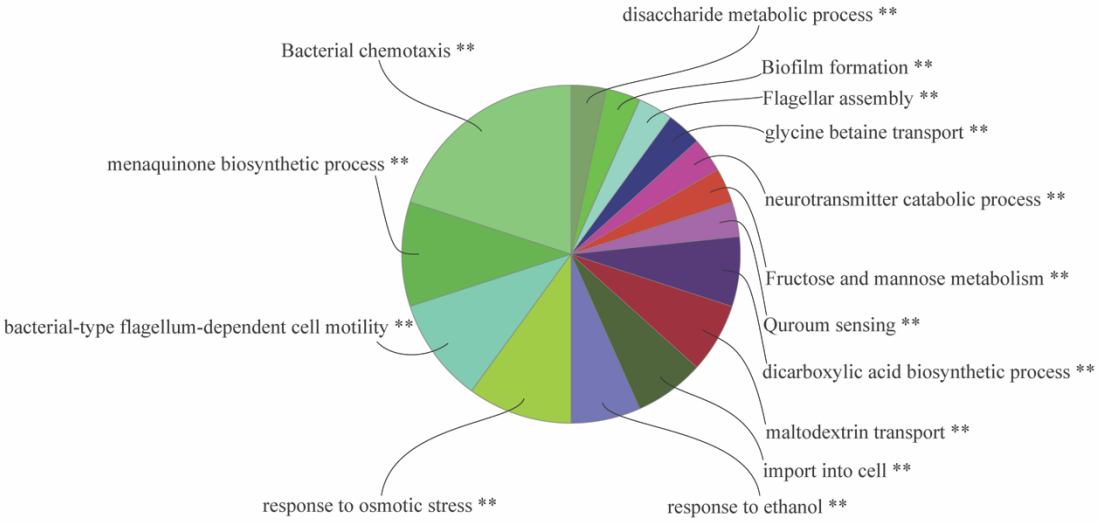
Table 13: Up-regulated genes in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt as determined by differential expression analysis. Only by ClueGO functionally assigned genes are listed, the complete list of upregulated genes is tabulated in Table 15 (Annex).

UniProt #	Description (Gene ID)	Fold change
P46119	Uncharacterized protein (ybjC)	2.646552
P76057	Uncharacterized protein (ydaQ)	2.434538
P0AAB2	Low molecular weight protein-tyrosine-phosphatase (wzb)	2.425544
P0C058	Small heat shock protein (ibpB)	2.266310
Q46822	Isopentenyl-diphosphate Delta-isomerase (idi)	2.058646
P21156	Sulfate adenylyltransferase subunit 2 (cysD)	1.998895
P17846	Sulfite reductase [NADPH] hemoprotein beta-component (cysI)	1.941382
P16676	Sulfate/thiosulfate import ATP-binding protein (cysA)	1.897836
P23845	Sulfate adenylyltransferase subunit 1 (cysN)	1.835026
P0AEB0	Sulfate transport system permease protein (cysW)	1.805111
P0A6J1	Adenylyl-sulfate kinase (cysC)	1.797465
P17854	Phosphoadenosine phosphosulfate reductase (cysH)	1.785610
P32055	GDP-L-fucose synthase (fcl)	1.744886
P64614	Uncharacterized protein (yhcN)	1.728328
P0AC88	GDP-mannose 4,6-dehydratase (gmd)	1.623178
P0AG78	Sulfate-binding protein (sbp)	1.601416
P0AEJ2	Isochorismate synthase (entC)	1.551492
P0AAY6	Uncharacterized protein (ybjN)	1.456755
P24175	Phosphomannomutase (manB)	1.268114
P0C054	Small heat shock protein (ibpA)	1.227245
P24174	Mannose-1-phosphate guanylyltransferase (manC)	1.208838
P13039	Enterochelin esterase (fes)	1.208134
P0ACV0	Lipid A biosynthesis lauroyltransferase (lpxL)	1.201600
P0AB35	Uncharacterized protein (ycfJ)	1.190928
P28629	Biodegradative arginine decarboxylase (adiA)	1.172362
P76387	Tyrosine-protein kinase (wzc)	1.148766
P71237	Putative colanic acid biosynthesis glycosyl transferase (wcaC)	1.142148
P04425	Glutathione synthetase (gshB)	1.127602
P07821	Iron(3+)-hydroxamate import ATP-binding protein (fhuC)	1.112371
P21169	Ornithine decarboxylase, constitutive (speC)	1.042976
P0AC53	Glucose-6-phosphate 1-dehydrogenase (zwf)	1.033368
P0AA47	Low-affinity putrescine importer (plaP)	1.017674
P77265	Multidrug resistance-like ATP-binding protein (mdlA)	1.015572
P07822	Iron(3+)-hydroxamate-binding protein (fhuD)	0.951725
P27245	Multiple antibiotic resistance protein (marR)	0.919561
P0A7F6	S-adenosylmethionine decarboxylase proenzyme (speD)	0.901667
P75855	Fimbrial subunit (elfA)	0.898458
P10378	Enterobactin synthase component E (entE)	0.897432
P06972	Iron(3+)-hydroxamate import system permease protein (fhuB)	0.890981
P23538	Phosphoenolpyruvate synthase (ppsA)	0.864752
P28861	Ferredoxin--NADP reductase (fpr)	0.862201
P16703	Cysteine synthase B (cysM)	0.851683
P0A6R3	DNA-binding protein (fis)	0.833322
P0AAG5	Multidrug resistance-like ATP-binding protein (mdlB)	0.814136
P0AEQ3	Glutamine-binding periplasmic protein (glnH)	0.807150
P37661	Kdo(2)-lipid A phosphoethanolamine 7"-transferase (eptB)	0.770001
P09148	Galactose-1-phosphate uridylyltransferase (galT)	0.751936
P37326	Putative prophage CPS-53 integrase (intS)	0.742175
P51025	S-formylglutathione hydrolase (frmB)	0.726757
P0A6Z3	Chaperone protein (htpG)	0.711308
P0A6A3	Acetate kinase (ackA)	0.709088
P31133	Putrescine-binding periplasmic protein (potF)	0.699763
P0A6Y8	Chaperone protein (dnaK)	0.699652
P31134	Putrescine transport ATP-binding protein (potG)	0.688690
P43319	Uncharacterized fimbrial-like protein (yraK)	0.679398
P0AFM6	Phage shock protein A (pspA)	0.679294
P09147	UDP-glucose 4-epimerase (galE)	0.666788
P63386	Probable phospholipid import ATP-binding protein (mlaF)	0.646657
P42626	UPF0597 protein (yhaM)	0.635398
P0A715	2-dehydro-3-deoxyphosphooctonate aldolase (kdsA)	0.627075
P27243	O-antigen ligase (rfaL)	0.617871
P0A908	MltA-interacting protein (mipA)	0.605244
Q46856	Alcohol dehydrogenase (yqhD)	0.597109
P25437	S-(hydroxymethyl)glutathione dehydrogenase (frmA)	0.594772

A



B



C

GO-term/pathway	Assigned Gene IDs
Bacterial chemotaxis	arnB|cheB|cheR|cheW|cheY|cheZ|dppA|evgA|flgM|flhC|flhD|fliA|fliC|fliG|fliM|fliZ|malE|motA|motB|narH|narW|narZ|tap|tar|tsr|ycgR|yhjH
Biofilm formation	flgM|flhC|flhD|fliA|fliZ|gmr|malP|mlrA|ycgR|ydaM|yhjH
Flagellar assembly	flgB|flgC|flgD|flgE|flgF|flgG|flgH|flgI|flgK|flgL|flgM|flgN|flhC|flhD|fliA|fliC|fliD|fliF|fliG|fliH|fliJ|fliK|fliM|fliO|fliS|fliT|motA|motB
Fructose and mannose metabolism	fbaA|fbaB|fruA|manX|manZ|mtlD|yedP
Quorum sensing	flhC|flhD|gadB|gmr|oppB|oppC|oppD|ydcS|ydcT|ydcV
bacterial-type flagellum-dependent cell motility	cheY|cheZ|flgB|flgC|flgD|flgE|flgF|flgG|flgI|flgJ|fliD|fliG|fliK|fliL|fliS|fliT|fliZ|malE|motA|tsr|ycgR|yciG|yhjH
dicarboxylic acid biosynthetic process	aroA|aroE|gabD|gabT|gadB|glsA|gltB|gltD
disaccharide metabolic process	malQ|otsA|otsB|treA
glycine betaine transport	osmF|yehW|yehX|yehY
import into cell	artM|artP|artQ|manX|manZ|proP|ydcS
maltodextrin transport	lamB|malE|malF|malK
menaquinone biosynthetic process	dkgA|gpr|menC|menD|menE|menF|menI|rob|wrbA
neurotransmitter catabolic process	gabD|gabT|gcvH|gcvT
response to ethanol	adhP|aldB|malE|osmC|rclA|tar|tsr|uspB|ydeI|yfcF|ygiW
response to osmotic stress	flgH|katE|osmC|osmE|osmF|osmY|otsA|otsB|proP|treA

Figure 24: ***Functional analysis of down-regulated genes in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt using the software tool ClueGO.*** *(A) significantly enriched GO-terms and pathways categorized in 30 groups indicating the percentage of genes present per term. (B) 30 categories from (A) summarized into 15 terms based on their functional relationship determined by their kappa score (≥0.4). The most significant term from (A) was used as the leading term. (C) All to the leading term (compare B) assigned genes are tabulated. (D) Functionally integrated network, where node size represents the term significance and the percentage of genes associated to each individual term is indicated by pie chart within the node. The genes assigned to each individual term are linked, whereas functionally related groups partially overlap.*

D

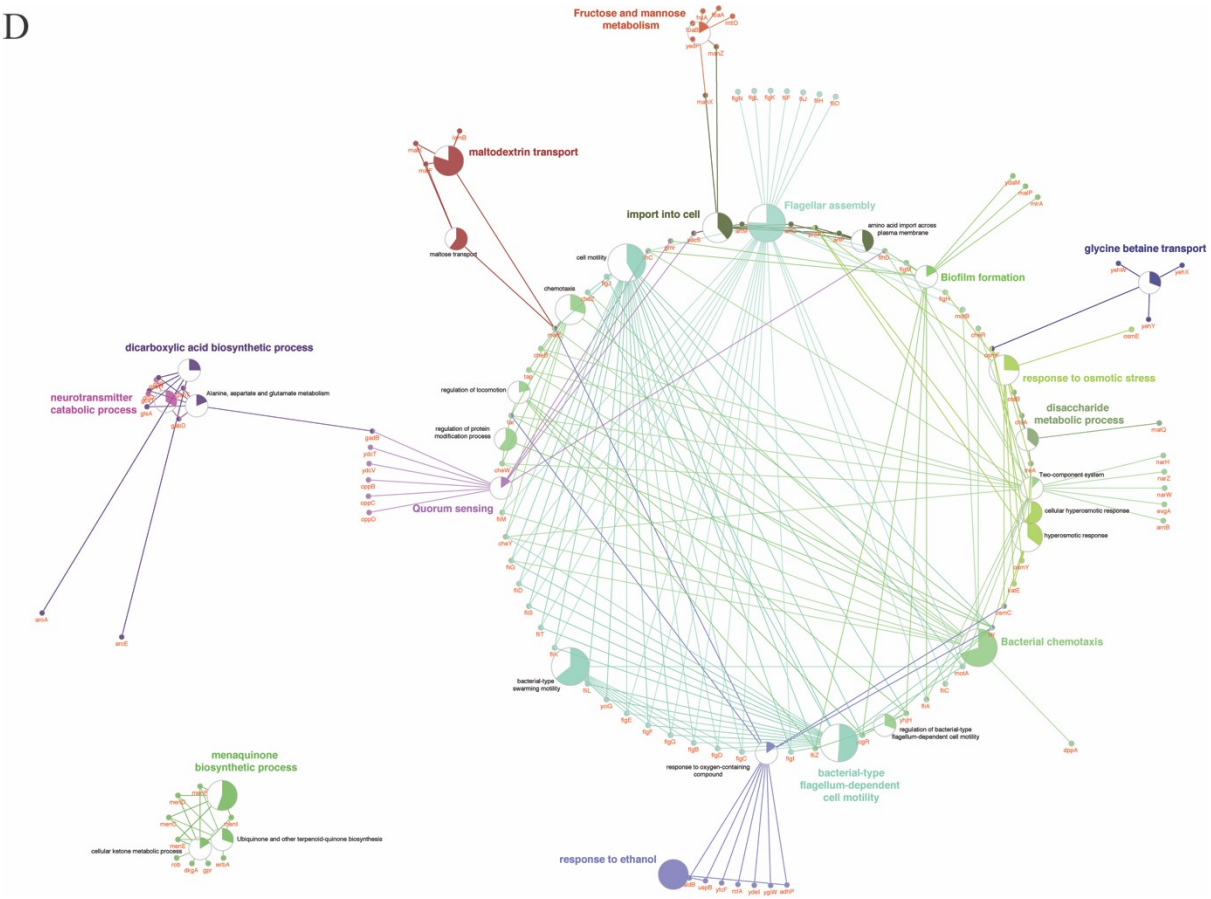


Figure 24 (continued): Functional analysis of down-regulated genes in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt using the software tool ClueGO. (A) significantly enriched GO-terms and pathways categorized in 30 groups indicating the percentage of genes present per term. (B) 30 categories from (A) summarized into 15 terms based on their functional relationship determined by their kappa score (≥0.4). The most significant term from (A) was used as the leading term. (C) All to the leading term (compare B) assigned genes are tabulated. (D) Functionally integrated network, where node size represents the term significance and the percentage of genes associated to each individual term is indicated by pie chart within the node. The genes assigned to each individual term are linked, whereas functionally related groups partially overlap.

Next, the downregulated genes (exluding knockouts gshA, endA, sdaA, speA, sdaB, tnaA, ompT, lon, gor and fhuA, 299 out of 309 downregulated genes) were functionally analyzed using ClueGO. The knockouts were excluded in order to focus the analysis on biological effect as a consequence of the knockouts and exclude any bias due to deliberately introduced genetic modifications.

For the analysis, the same ontology sets, namely KEGG and GO_BiologicalProcess as downloaded on 21st of February 2018, were used. Out of 299 genes, 204 (68.2%) were recognized and functionally annotated in the selected ontology and 111 individual genes (56.6%) are associated to representative terms and pathways after applying the selection criteria (evidence code: all_without_IEA; p-Value cutoff: 0.05 using right-sided hypergeometric test (enrichment combined with Benjamini-Hochberg algorithm for correction of p-values for multiple testing). These genes were assigned to 30 GO terms and KEGG pathways (compare Figure 24. A). The percentage of genes associated per GO term or pathway are represented as bar graph indicates the completeness of the term relative to all associated genes of the respective

term. The bars representing individual GO terms and pathways are colored in the same color in case their kappa score exceeded 0.4 indicating a high connectivity and therefore functional relationship of the corresponding terms. On this basis, the functionally related terms were grouped in 15 major groups using the GO terms or pathways with the smallest p-value of each individual functional group as the leading term (Figure 24, B). The genes associated to the 15 major groups labeled after the most significantly enriched term are tabulated in Figure 24, C. The associated genes partially overlap and amount to 111 after excluding double counted genes. All significantly enriched GO-terms and pathways as well as the individually assigned genes are shown as network.

All 111 genes assigned to the 15 major groups as well as the deliberately removed genes are listed in Table 14 and sorted by regulation factor.

Table 14: Down-regulated genes in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt as determined by differential expression analysis. Only by ClueGO functionally assigned genes are listed, the complete list of upregulated genes is abulated in Table 16: All 309 down-regulated genes found in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt as determined by differential expression analysis.

UniProt #	Description (Gene ID)	Fold change
P09169	Protease 7 (ompT)	-8.970902
P0A9M0	Lon protease (lon)	-8.069021
P21170	Biosynthetic arginine decarboxylase (speA)	-7.645337
P16095	L-serine dehydratase 1 (sdaA)	-7.583413
P06715	Glutathione reductase (gor)	-7.544772
P0A6W9	Glutamate--cysteine ligase (gshA)	-7.316966
P06971	Ferrichrome-iron receptor (fhuA)	-5.517509
P30744	L-serine dehydratase 2 (sdaB)	-5.383357
P0A853	Tryptophanase (tnaA)	-5.099459
P25736	Endonuclease-1 (endA)	-4.728869
P07017	Methyl-accepting chemotaxis protein II (tar)	-3.526871
P04949	Flagellin (fliC)	-3.481082
P37646	Cyclic di-GMP phosphodiesterase (yhjH)	-3.246441
P09348	Motility protein A (motA)	-3.21432
P07018	Methyl-accepting chemotaxis protein IV (tap)	-3.134911
P07330	Chemotaxis response regulator protein-glutamate methylesterase (cheB)	-3.05296
P24216	Flagellar hook-associated protein 2 (fliD)	-3.009602
P0AF06	Motility protein B (motB)	-2.97171
P0A964	Chemotaxis protein (cheW)	-2.894882
P26608	Flagellar protein (fliS)	-2.659458
P0AEM6	RNA polymerase sigma factor (fliA)	-2.651567
P07364	Chemotaxis protein methyltransferase (cheR)	-2.62206
P0AE67	Chemotaxis protein (cheY)	-2.561267
P33235	Flagellar hook-associated protein 1 (flgK)	-2.479062
P0ABY2	Flagellar protein (fliT)	-2.332172
P68187	Maltose/maltodextrin import ATP-binding protein (malK)	-2.252193
P0A9H9	Protein phosphatase (cheZ)	-2.213499
P0ABX2	Flagellar basal-body rod protein (flgC)	-2.124454
P75937	Flagellar hook protein (flgE)	-2.052378
P75936	Basal-body rod modification protein (flgD)	-2.04667
P29744	Flagellar hook-associated protein 3 (flgL)	-2.043006
P0ABW9	Flagellar basal body rod protein (flgB)	-1.991373
P76010	Flagellar brake protein (ycgR)	-1.950398
P0AEX9	Maltose-binding periplasmic protein (malE)	-1.948084
P0A6S0	Flagellar L-ring protein (flgH)	-1.907782
P21179	Catalase HPII (katE)	-1.852451
P0AEM4	Negative regulator of flagellin synthesis (flgM)	-1.778174
P75938	Flagellar basal-body rod protein (flgF)	-1.739406
P43533	Flagella synthesis protein (flgN)	-1.733642
P0A8G6	NAD(P)H dehydrogenase (quinone) (wrbA)	-1.712834
P52627	Regulator of sigma S factor (fliZ)	-1.59618
P0ABX5	Flagellar basal-body rod protein (flgG)	-1.590508
P37685	Aldehyde dehydrogenase B (aldB)	-1.581132
P75942	Peptidoglycan hydrolase (flgJ)	-1.579898
P25798	Flagellar M-ring protein (fliF)	-1.5737
P0ADB1	Osmotically-inducible lipoprotein E (osmE)	-1.559353
P21361	Uncharacterized protein (yciG)	-1.499132

UniProt #	Description (Gene ID)	Fold change
P06974	Flagellar motor switch protein (fliM)	-1.463427
P31678	Trehalose-6-phosphate phosphatase (otsB)	-1.452872
P19319	Respiratory nitrate reductase 2 alpha chain (narZ)	-1.44411
P0ABZ1	Flagellar motor switch protein (fliG)	-1.443547
P0A6S3	Flagellar P-ring protein (flgI)	-1.438007
P31130	Uncharacterized protein (ydeI)	-1.417368
P00490	Maltodextrin phosphorylase (malP)	-1.412605
P11349	Respiratory nitrate reductase 1 beta chain (narH)	-1.398603
P31677	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] (otsA)	-1.369355
P22256	4-aminobutyrate aminotransferase (gabT)	-1.363628
P39451	Alcohol dehydrogenase, propanol-preferring (adhP)	-1.345162
P0AFH8	Osmotically-inducible protein Y (osmY)	-1.336501
P25526	Succinate-semialdehyde dehydrogenase [NADP(+)] (gabD)	-1.326093
P0A991	Fructose-bisphosphate aldolase class 1 (fbaB)	-1.325711
P0C0L2	Peroxiredoxin (osmC)	-1.270691
P0ABX8	Flagellar protein (fliL)	-1.252672
P15977	4-alpha-glucanotransferase (malQ)	-1.244318
P77302	Probable diguanylate cyclase (ydaM)	-1.237686
P02943	Maltoporin (lamB)	-1.23653
P02916	Maltose transport system permease protein (malF)	-1.234442
P33362	Glycine betaine-binding protein (yehZ)	-1.230016
P19317	Probable nitrate reductase molybdenum cofactor assembly chaperone (narW)	-1.224531
P52613	Flagellar protein (fliJ)	-1.218742
P02942	Methyl-accepting chemotaxis protein I (tsr)	-1.208979
P0ADU5	Protein (ygiW)	-1.169852
P27248	Aminomethyltransferase (gcvT)	-1.145338
P31068	Flagellar assembly protein (fliH)	-1.128452
P0AFH2	Oligopeptide transport system permease protein (oppB)	-1.117231
P33361	Glycine betaine uptake system permease protein (yehY)	-1.103408
P33360	Glycine betaine uptake system ATP-binding protein (yehX)	-1.090925
Q46857	2,5-diketo-D-gluconic acid reductase A (dkgA)	-1.085676
P76329	Mannosyl-3-phosphoglycerate phosphatase (yedP)	-1.082416
P76108	Putative ABC transporter periplasmic-binding protein (ydcS)	-1.046998
P22586	Flagellar protein (fliO)	-1.033393
P0AFH6	Oligopeptide transport system permease protein (oppC)	-1.015139
P09831	Glutamate synthase [NADPH] large chain (gltB)	-1.004202
P33359	Glycine betaine uptake system permease protein (yehW)	-0.993738
P77795	Uncharacterized ABC transporter ATP-binding protein (ydcT)	-0.990094
P13482	Periplasmic trehalase (treA)	-0.962487
P0A6T9	Glycine cleavage system H protein (gcvH)	-0.959388
P52614	Flagellar hook-length control protein (fliK)	-0.953386
P77454	Glutaminase 1 (glsA1)	-0.92661
P77690	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase (arnB)	-0.922502
P20966	PTS system fructose-specific EIIBC component (fruA)	-0.917105
P0ABY7	Flagellar transcriptional regulator (flhC)	-0.89154
Q46851	L-glyceraldehyde 3-phosphate reductase (gpr)	-0.890774
P76027	Oligopeptide transport ATP-binding protein (oppD)	-0.88053
P0AFR9	Inner membrane ABC transporter permease protein (ydcV)	-0.867493
P09832	Glutamate synthase [NADPH] small chain (gltD)	-0.859303
P38051	Isochorismate synthase (menF)	-0.858779
P0ACZ4	Positive transcription regulator (evgA)	-0.855253
P33358	HTH-type transcriptional regulator (mlrA)	-0.851328
P69910	Glutamate decarboxylase beta (gadB)	-0.832214
P69805	Mannose permease IID component (manZ)	-0.827851
P0AE30	Arginine ABC transporter permease protein (artM)	-0.827117
P77781	1,4-dihydroxy-2-naphthoyl-CoA hydrolase (menI)	-0.80911
P69797	PTS system mannose-specific EIIAB component (manX)	-0.79702
P0C0L7	Proline/betaine transporter (proP)	-0.795181
P77212	Probable pyridine nucleotide-disulfide oxidoreductase (rclA)	-0.780315
P23847	Periplasmic dipeptide transport protein (dppA)	-0.746441
P17109	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase (menD)	-0.738309
P0A8S5	Universal stress protein B (uspB)	-0.735734
P0AE34	Arginine ABC transporter permease protein (artQ)	-0.715468
P0A6D3	3-phosphoshikimate 1-carboxyvinyltransferase (aroA)	-0.705841
P77544	Glutathione S-transferase (yfcF)	-0.694611
P09424	Mannitol-1-phosphate 5-dehydrogenase (mtlD)	-0.692888
P77334	Cyclic di-GMP phosphodiesterase (gmr)	-0.679334
P0AB71	Fructose-bisphosphate aldolase class 2 (fbaA)	-0.642063
P15770	Shikimate dehydrogenase (NADP(+)) (aroE)	-0.637019
P0ACI0	Right origin-binding protein (rob)	-0.635975
P0A8S9	Flagellar transcriptional regulator (flhD)	-0.608896
P0AAF6	Arginine transport ATP-binding protein (artP)	-0.608397
P37353	2-succinylbenzoate--CoA ligase (menE)	-0.606305
P29208	o-succinylbenzoate synthase (menC)	-0.599185

4.3.8 Amino Acid Profiling of amino acid stabilized *E. coli* A19

Within this chapter, the effect of amino acid stabilizing knockouts ΔgshA ΔsdaA ΔspeA ΔsdaB and ΔtnaA, are examined. For this purpose, four independently prepared S30 extracts were prepared from A19 ΔendA (positive control) and the derivative strain A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA, A19 CSWR). Heat-inactivated extract from A19 ΔendA was used as a negative control. The removed gene products were the glutamate-cysteine ligase (gshA), L-serine deaminase I and II (sdaA and sdaB), arginine decarboxylase (speA) and tryptophanase (tnaA) in order to stabilize the amino acids cysteine, serine, arginine and tryptophan, respectively.

The concentration profile over time for each individual amino acid except cysteine and proline are illustrated in Figure 25, A-R. The concentration profile of alanine (A) shows that the concentration in S30 extract is independent of any genetic modification and remains relatively stable over the course of 105 min. The same is true for histidine (H), isoleucine (I), leucine (J), phenylalanine (M), tryptophan (P), tyrosine (Q) and valine (R). These amino acids can be considered non-critical during CFPS and there are apparently no enzymatic activities present in the extract that would result in a degradation of these amino acids. The initial concentration of tryptophan was calculated to be 2 mM, however, the initial concentration measured was only 1 mM, which was most likely due to a mistake during preparation of the amino acids mixture.

Beside these non-critical amino acids, the concentration profile of aspartic acid (D), glutamic acid (E), glycine (G), lysine (K), methionine (L) and threonine (O) change during the CFPS reaction. For lysine (K), methionine (L) and threonine (O) there was a decrease of the concentration observed, resulting in a final concentration of approx. 1.25, 1.25 and 0.75 mM, respectively, which means that of those amino acids 35-65% of the initial amount were consumed by potentially non-productive pathways present in the S30 extract. The consumption was independent of whether the strain was genetically modified or not. Glutamic acid was initially present at a high concentration (75 mM) and dropped by approx. 10 mM, whereas there seems to be a higher decrease in the A19 ΔendA as compared to A19 CSRW, suggesting that the pathways degrading or transforming glutamic acid are less active in the amino-stabilized *E. coli*.

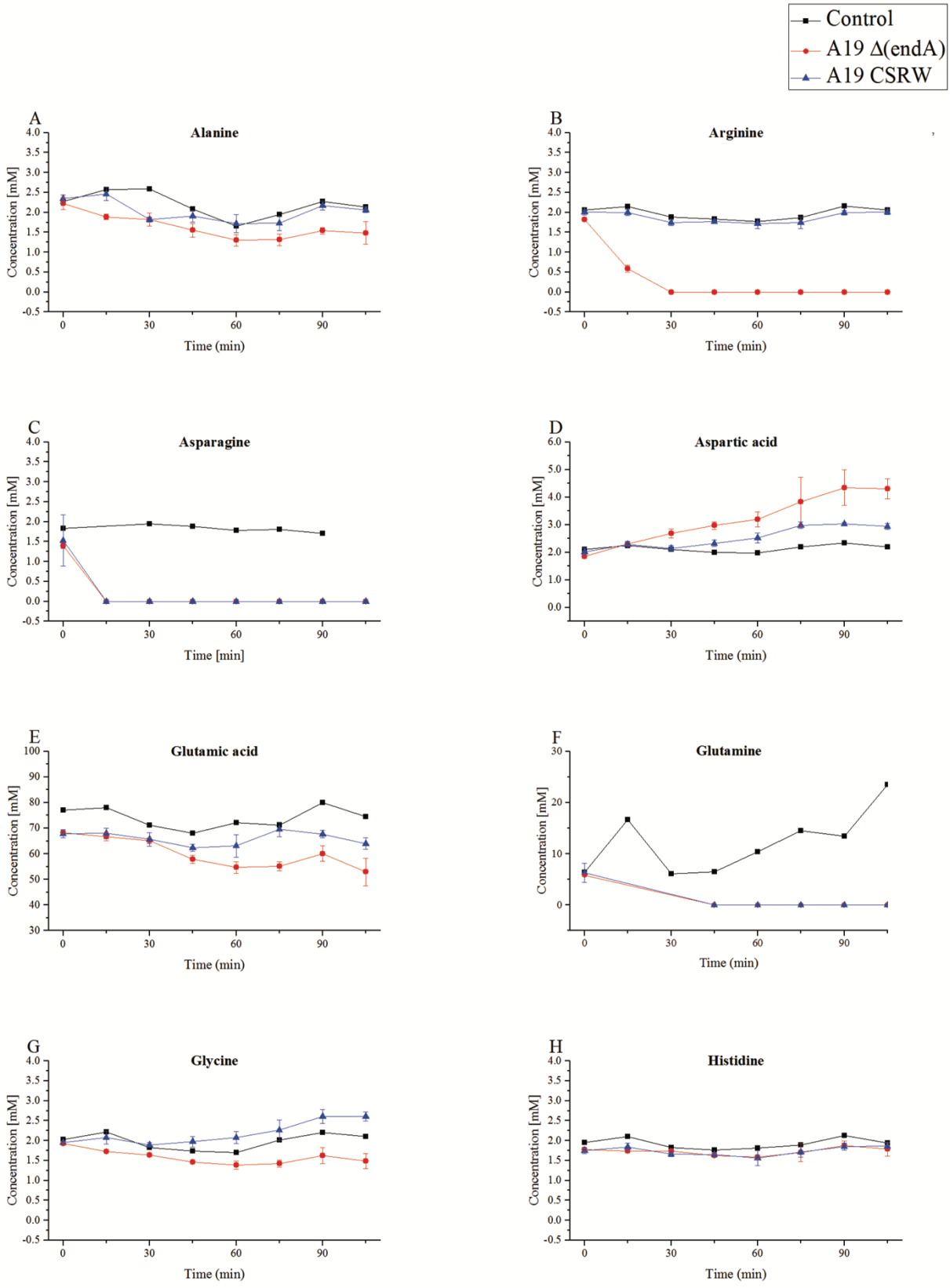


Figure 25: ***Course of amino acid concentration in S30 extracts over time****. The progression of amino acid concentration over time in heat-inactivated S30 extract (Control; black, square) derived from E. coli A19 Δ(endA) and in S30 extracts derived from A19 Δ(endA) (red, round) and A19 Δ(endA) CSRW (blue, triangle) are illustrated. Each graph shows the progression of one individual amino acid. Error bars indicate standard deviation from three biological replicates (n=3).*

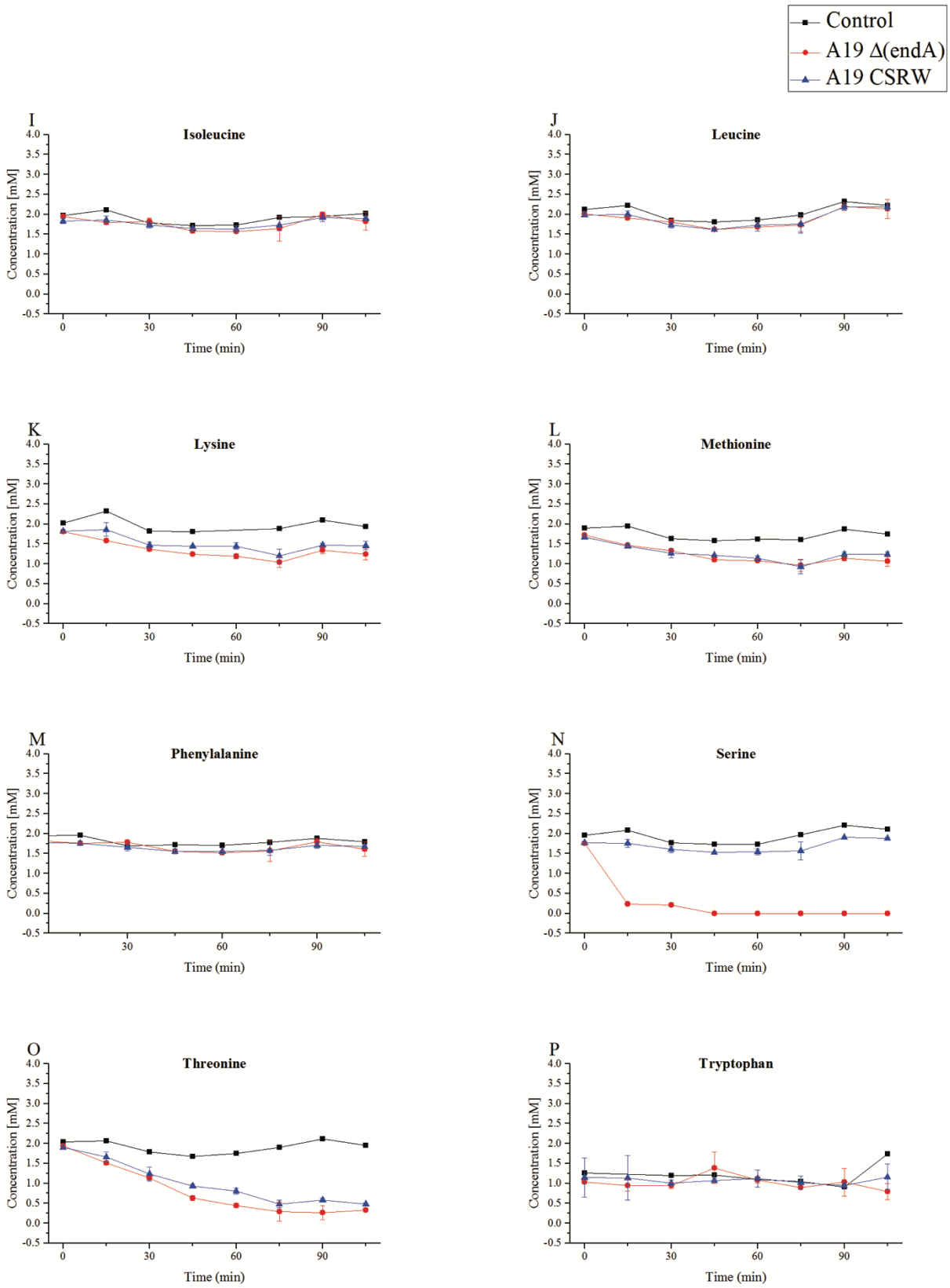


Figure 12 (continued): ***Course of amino acid concentration in S30 extracts over time****. The progression of amino acid concentration over time in heat-inactivated S30 extract (Control; black, square) derived from E. coli A19 Δ(endA) and in S30 extracts derived from A19 Δ(endA) (red, round) and A19 Δ(endA) CSRW (blue, triangle) are illustrated. Each graph shows the progression of one individual amino acid. Error bars indicate standard deviation from three biological replicates (n=3).*

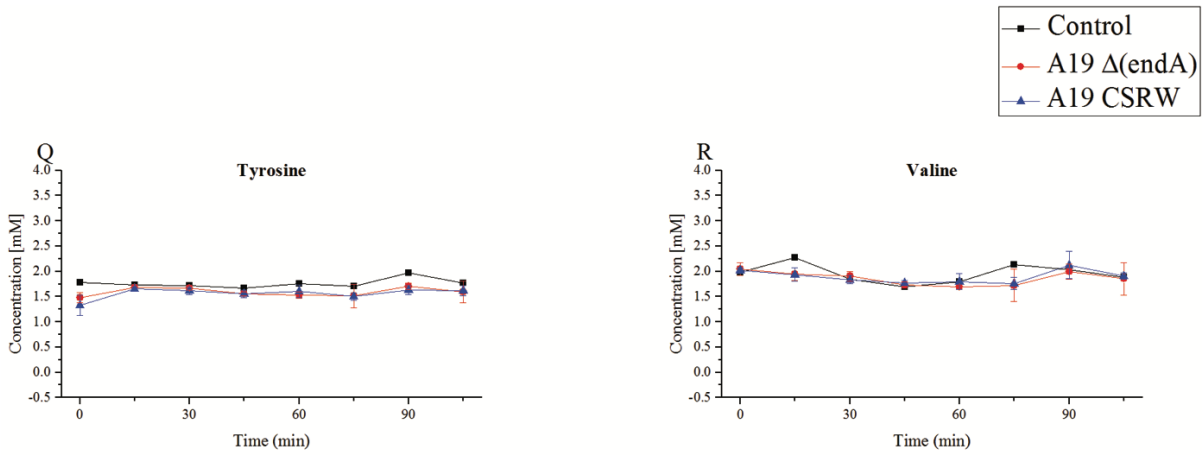


Figure 12 (continued): ***Course of amino acid concentration in S30 extracts over time****. The progression of amino acid concentration over time in heat-inactivated S30 extract (Control; black, square) derived from E. coli A19 Δ(endA) and in S30 extracts derived from A19 Δ(endA) (red, round) and A19 Δ(endA) CSRW (blue, triangle) are illustrated. Each graph shows the progression of one individual amino acid. Error bars indicate standard deviation from three biological replicates (n=3).*

Interestingly, the concentration of aspartic acid increases by approx. 100% for the A19 ΔendA and only by approx. 40% in the amino acid stabilized strain thus, showing the same trend but suggesting a less active enzymatic conversion in the strain A19 CSRW. For glycine, the strains show opposing trends. While the concentration increases in the A19 CSWR strain by 0.5 mM within 105 min, the concentration in A19 ΔendA decreases by the same amount.

As potentially limiting amino acid even after the genetic modifications, asparagine (C) and glutamine (F) were identified. Asparagine's initial concentration of approx. 1.5 mM could not be detected in S30 extracts after 15 min, while remaining stable in the heat-inactivated negative control-extract. This suggests that it is rapidly transformed by non-productive pathways present in S30 extract. Similarly, glutamine is not detectable after 45 min (t=15 min and t=30 min could not be evaluated due to technical difficulties). Technical issues are also apparent when evaluating the heat-inactivated negative control extract. The concentration profile, despite being detectable, is highly fluctuating, which raises the question about the reliability of this dataset. Nevertheless, further investigation needs to be performed to investigate the role of asparagine and glutamine as potentially limiting protein building blocks.

The stabilizing effect by removal of sdaA/B and speA on serine and arginine, respectively, were clearly observable in this analysis. As seen in (B), arginine is rapidly depleted within 30 min to non-detectable concentration, suggesting a potential shortage of this amino acid. However, the concentration profile measured in A19 CSRW resembles the profile observed in the heat-inactivated negative control extracts, suggesting that the source of degradation of arginine present in A19 ΔendA is completely eliminated in A19 CSRW. Comparable effects were observed for serine, which is also rapidly depleted in A19 ΔendA. The concentration drops rapidly and not detectable anymore at t=45 min. This concentration

profile is similar to the heat-inactivated negative control extracts. This implies that, in A19 CSRW, the source of degradation of serine was completely eliminated.

In summary, the analysis showed that serine and arginine are potentially limiting building blocks that can be completely stabilized within the *E. coli* derived extract by genetic modification of the source strain. Limiting effects of the tryptophan concentration could not be confirmed in this analysis and, due to technical difficulties, cysteine stability, presumably affected by the gene product of gshA, could not be captured in this experiment and remains to be evaluated by other means. Besides arginine and serine, the amino acids glutamine and asparagine were identified as potential limiting building blocks and should be addressed by additional experiments.

4.3.9 Performance of modified *E. coli* A19 during *in vitro* Protein Synthesis

Next, the performance of the S30 extracts derived from A19 wt and selected genetically modified strains was evaluated. For the comparison, the strains A19 wt and its derivatives A19 rne::SBP Δ(gshA/endA), A19 rne::SBP Δ(endA) CSRW, A19 rne::SBP Δ(endA/ompT/lon) CSRW, A19 rne::SBP Δ(endA/ompT/lon/gor) CSRW were selected. In addition, A19 derivatives without amino-acid stabilizing knockouts were included, namely A19 rne::SBP Δ(ompT), A19 rne::SBP Δ(ompT/lon), A19 rne::SBP Δ(ompT/lon/gor) and A19 rne::SBP Δ(ompT/lon/gor/fhuA).

From all those strains, three independent S30 extracts were generated and Chloramphenicol Acetyl Transferase (CAT) was expressed in each extract as triplicates using the PANOx-SP energy regeneration system. The concentration of TCA-precipitable CAT was measured after incorporation of ^{14}C leucine. As the values scattered strongly, the measurements (three indepently prepared S30 extracts of each strain, each extract measured five times) are shown as box plot, whereas the complete dataset is shown in the upper box plot and the data set after data curation including obvious outliers and values far off the median value are shown in the lower box plot. The plots before and after removal are showing the same trend.

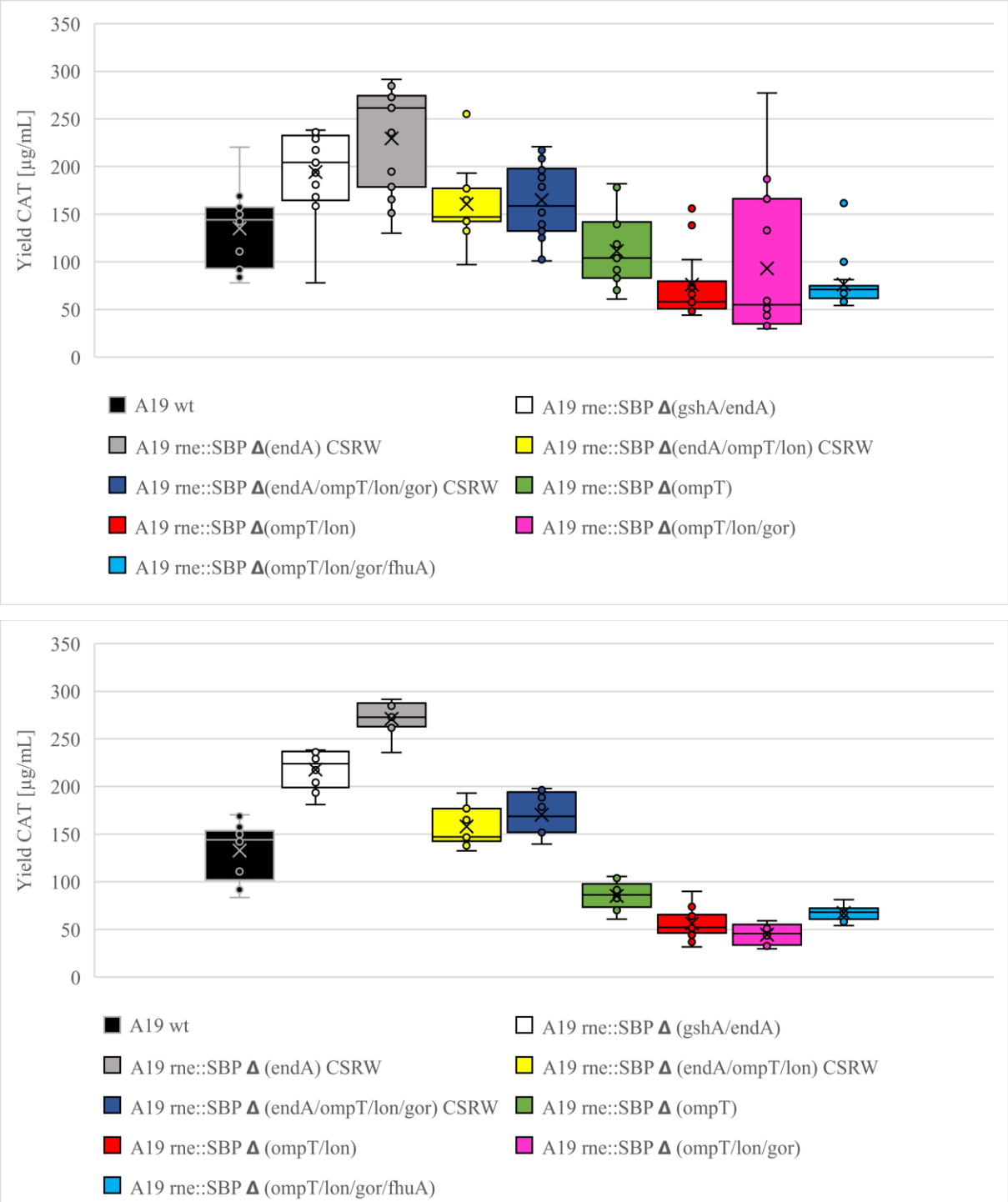


Figure 26: ***TCA-precipitable yield of CAT expressed in various S30 extracts derived from A19 wt and genetically modified derivatives.*** *The upper box plot show the original data. The lower box plot shows the same data without outliers and data points far off the median.*

As shown in Figure 26, extracts derived from A19 wt yield between 92 and 169 µg/mL of CAT (median and average yield 144 and 133 µg/mL, repectively). Removing the endA and gshA gene resulted in markedly higher yield (median 224 µg/mL), suggesting that the stabilization of cysteine by removal of gshA and/or the stabilization of the template DNA by removal of endA is responsible for increased yield. The additional removal of sdaA/B (to stabilize serine), speA (to stabilize arginine) and tnaA (to stabilize tryptophan) resulting in

strain A19 rne::SBP Δ(endA) CSRW had likewise a beneficial effect on CAT yields resulting in a median yield of 273 µg/mL. In total, the amino acids stabilizing genetic modification resulted in increase of the yield by approx. 90%, suggesting that the amino acids are indeed a limiting factor in CFPS. However, the additional removal of the major proteases ompT and lon decreased the yield markedly and reversed the beneficial effects of the amino acid stabilizing knockouts, resulting in yields comparable to A19 wt (median 147 µg/mL). The additional knockout of gor results in a slight but not markedly increased median yield of 167 µg/mL.

The analysis of effects of ompT, lon, gor and fhuA in the absence of the amino acid stabilizing knockouts gshA, sdaA/B, speA, tnaA and in conjunction with the removal of endA showed a rather negative effect compared to extracts derived from A19 wt. The yields achieved with extracts derived from those strains were between 45 and 100 µg/mL.

The genetic modifications were introduced during the project "Optimierung von *E. coli*-Stämmen zur Lysatherstellung für die zellfreie Proteinsynthese" in collaboration with Biotechrabbit (Berlin, Germany). Therefore, the strains were also tested by Dr. Hemut Merk and Dr. Michael Gerrits in Biotechrabbits' commercially available CECF expression system. Two target proteins, namely arylhydrocarbon receptor (AhR) and EF-Ts (elongation factor thermos stable) were expressed mode and is commercially available from. These proteins tested in this experiment as shown in Figure 27 have been shown previously to be difficult-to-express proteins in standard expression systems.

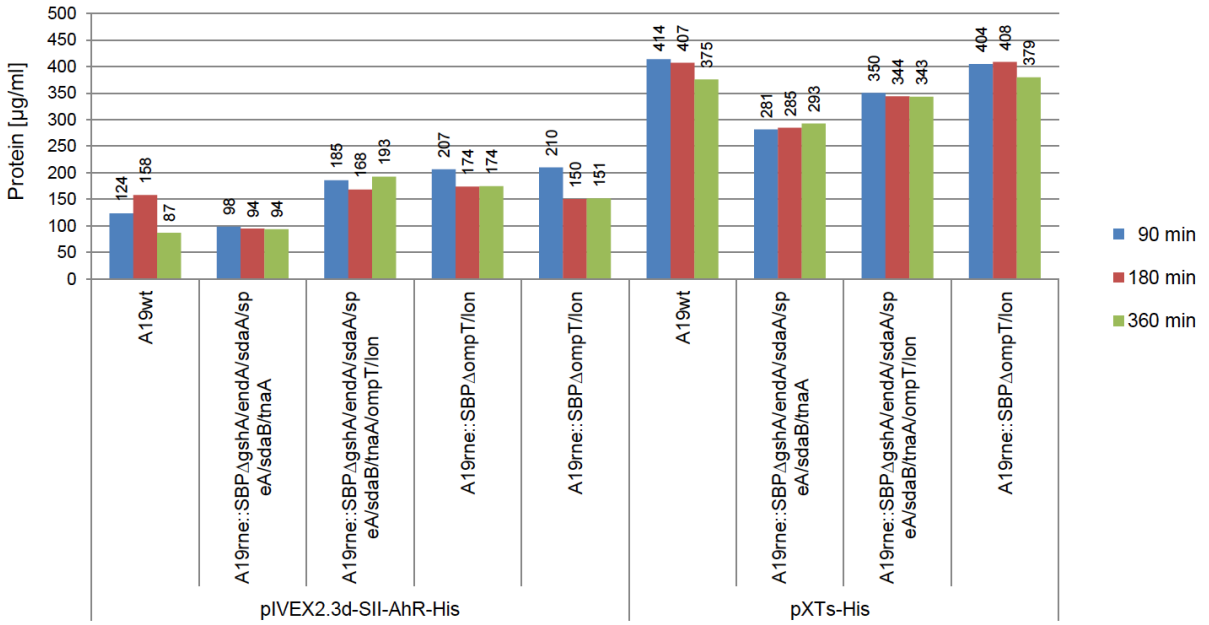


Figure 27: ***Comparison of A19 wt and selected geneticically modified strains (experiments performed by Dr. Micheal Gerrits and Dr. Helmut at Merk, Berlin)..***

The analysis shows that the target protein AhR reached maximum expression level in A19 wt-derived extract after 180 min expression time with a concentration of 158 µg/mL per

reaction. The extract derived from genetically modified A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA) on the other hand resulted, in contrast to the previous results obtained during expression of CAT in our lab (compare Figure 26), in rather low yield of 98 µg/mL, which did not improve with longer expression times of up to 360 min. While the genetically modified strains after removing of ompT and lon performed sub-optimally when expressing CAT in the batch system, the yield of AhR was increased in the presence and absence of the amino-acid stabilizing knockouts of up to a maximum concentration of 210 µg/mL per reaction.

For the expression of the target protein pXTs-His, no beneficial effect of knockouts could be observed. The expression level went down by 100 µg/mL when using amino-acid stabilized extracts as compared to the expression of pXTs-His in A19 wt extracts. The expression level recovered by 50 µg/mL with additional knockout of ompT and lon in the presence of amino-acid stabilized knockouts, but was still below A19 wt. Also, the knockout of ompT and lon only did not improve expression of pXTs-His when compared to A19 wt. This shows that effect of the knockouts are highly target protein dependent and no universal improvement of expression could be achieved by genetic modification of the S30 extract source strain.

4.4 Expression of Fab/scFab with optimized S30 extracts

4.4.1 Purification of Chaperone DsbC/FkpA and Skp

In order to supplement the S30 extract with chaperones present in the periplasmic space which might aid folding of Fab and scFab during *in vivo* production, the major periplasmic chaperones namely protein disulfide isomerase (DsbC), prolyl isomerase (FkpA) and holdase (skp) were overexpressed and purified as described in the methods section (3.4).

4.4.2 Genetic Modification of *E. coli* A19

As decribed in detail in the previous chapters, the parental strain *E. coli* strain A19 (genotype: *Rna-19gdhA2his-95relA1spoT1metB1*) served as source strain for genetic modifications [29]. Several endogenous enzymes of *E. coli* have been previously shown to deplete, modify or degrade substrates, template DNA and target protein during CFPS reactions and recombinant protein expression *in vivo* [44, 45, 48, 120, 121]. The gene *endA* coding for endonuclease A was removed to prevent degradation of template DNA. To stabilize amino acids, particularly arginine, cysteine, serine and tryptophan which have been shown to be

rapidly depleted in *E. coli* extract due to endogenous enzymes [44, 45], the genes coding for arginine decarboxylase (*speA*), glutamate cysteine ligase (*gshA*), L-serine dehydratase 1 and 2 (*sdaA* and *sdaB*) and tryptophanase (*tnaA*) were removed. The genes *ompT* and *lon* coding for outer membrane protease 7 and Lon protease, respectively, were knocked out to prevent proteolytic degradation of foreign proteins as well as added T7 RNAP needed to drive expression of target proteins in CFPS reactions [122-124]. Most importantly in the context of antibody expression, the gene *gor* coding for glutathione reductase was removed to maintain an oxidizing environment using oxidized glutathione which is otherwise rapidly reduced in the presence of glutathione reductase [50].

Additionally, the gene *fhuA* (coding for ferrichrome-iron receptor) was knocked out to protect our strain from phage infection [125]. As previously described in detail, the resulting strain *E. coli* A19 Δ*endA* Δ*gshA* Δ*speA* Δ*sdaA* Δ*sdaB* Δ*tnaA* Δ*ompT* Δ*lon* Δ*gor* Δ*fhuA* grew similarly to parental strain *E. coli* A19 with a growth rate of 1.1 h^{-1} (parental strain showed growth rate of 1.3 h^{-1}) in 2xYTPG media at 37 °C (Figure 21), suggesting that the introduced knockout were not vital. We verified the removal of the respective genes in the target locus by isolating genomic DNA of parental strain *E. coli* A19 and genetically modified strain and amplification of target locus using PCR. The resulting PCR products were separated using agarose gel electrophoresis and compared to theoretical expected band size based on genome sequence of *E. coli* W3110 (NC_007779.1) before and after knockout (Figure 17). Additionally, the PCR products were verified by sequencing. The genetically modified A19 strain, hereafter referred to as A19 rne::SBP Δ(endA/ompT/lon/gor/fhuA) CSRW, was cultivated and a single batch of S30 extract was prepared and used for all following experiments.

4.4.3 Optimization of Magnesium and Potassium Glutamate

Besides the ribosomal extracts, the energy regeneration system is a crucial component in CFPS systems. Due to its robustness, it was decided to employ the so-called PANOx-SP energy regeneration system which has been described in detail elsewhere [126]. Two important constituents, magnesium- and potassium-glutamate, were optimized to obtain high yield of target protein in our batch of ribosomal extract. Free divalent magnesium cations are essential to CFPS reactions involving nucleotides, while potassium and glutamate are representing the predominant intracellular cation and anion, respectively, thus closely mimicking the intracellular milieu in CFPS systems.

Concentrations of 12-19 mM magnesium-glutamate and 90-360 mM potassium-glutamate were evaluated in a correlated screen using while following fluorescence emission of the indicator protein eYFP. The eYFP was chosen a due to its folding properties and easy

detectability within the synthesis reaction. We observed highest fluorescence emission using a combination of 15 mM magnesium glutamate and 280 mM potassium glutamate (Figure 28). The combination of magnesium and potassium glutamate proving optimal in this experiment were used in all subsequent experiments.

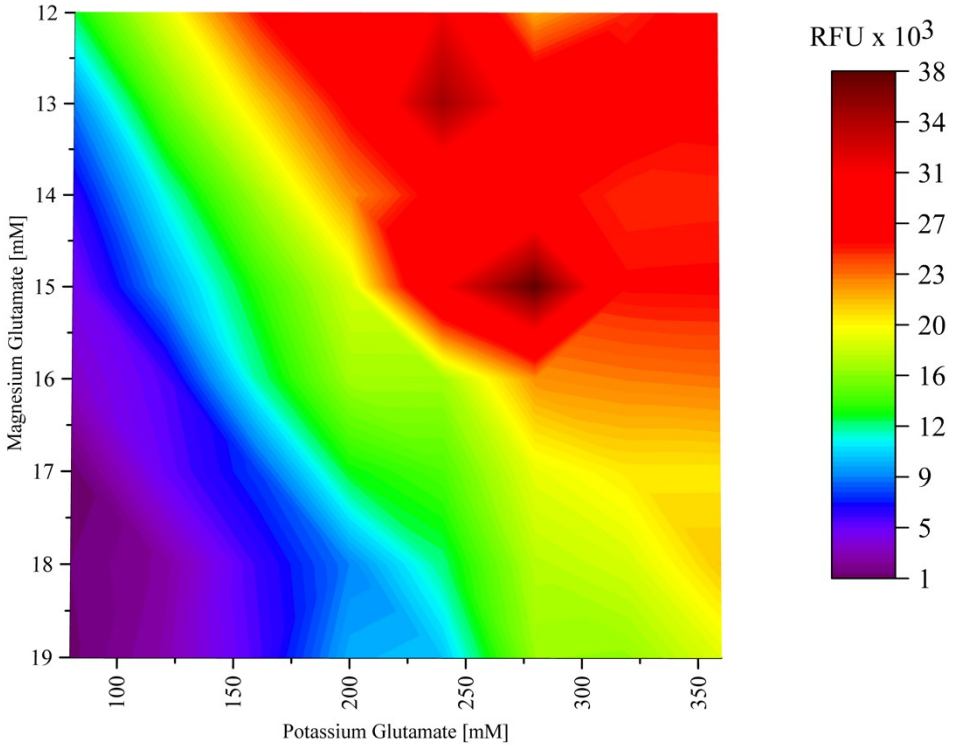


Figure 28: ***Correlated screen of potassium- and magnesium-glutamate visualized as 2D-plot.*** *The formation of eYFP in the presence of various concentrations of potassium- and magnesium-glutamate was followed by fluorescence (λ_{ex} = 485 nm; λ_{em} = 525 nm) measured every 10 min at 60% sensitivity. When the fluorescence leveled-off after ~2 h, the relative fluorescence units (RFU) were plotted against potassium- and magnesium-glutamate concentrations. Highest RFU-values are plotted in red and lowest in purple (compare legend).*

4.4.4 Optimization of Redox Conditions

Maintaining favorable redox conditions to form disulfides bridges during CFPS is crucial for functional folding of antibody fragments. Therefore, the redox environment in OCPS systems is commonly adjusted by using oxidized glutathione (GSSG) to provide an oxidizing environment that allows formation of disulfides bonds. While the formation of disulfide bonds is highly dependent on oxidizing environment, the disulfide bond isomerase (DsbC), essential for disulfide bond shuffling and proofreading, is only active if the cysteines in the active site are maintained in a reduced state. Thus, to provide optimal conditions for the formation of disulfide bonds and their shuffling by DsbC, the optimal ratio of oxidized and reduced

glutathione (GSSG and GSH) in the presence of 13 µM DsbC after treatment of the extract with iodoacetamide (IAM) was determined.

Initially, the extract was treated with various concentrations of the alkylating agent IAM to irreversibly deactivate active site cysteine residues of S-S reducing enzymes. As our strain carries a Δ*gor* mutation and therefore the glutathione reductase is not present in the resulting extracts, it was possible to use a low concentration of IAM ranging from 0-425 µM to deactivate the major remaining enzyme thioredoxin reductase (*trxB*) responsible for reducing disulfides during CFPS reactions. After treatment of the extract for 30 min at RT, 13 µM of DsbC and 2 mM of oxidized glutathione along with other constituents of the CFPS reaction were added and Anti-Sin1 Fab was expressed by addition of 1:1 ratio of HC- and LC- coding template DNA. After termination, equal amounts of the various IAM-treated reactions were analyzed using Western Blot under reducing and non-reducing conditions detecting fusion StrepII tag of HC and LC (data not shown). The treatment with various IAM concentrations had only little effect on final yield of correctly assembled Fab. Using densitometry, we found that the yield of correctly assembled Fab increased slightly (~10-20%) up to a concentration of 100 µM as compared to non-treated extract. On the contrary, concentration >100 µM decreased yield of Fab slightly. Based on that experiment, we decided to employ pre-treatment of our extract with 100 µM IAM for all subsequent experiments.

Next, different ratios (1:4; 2:3; 3:2; 4:1) of oxidized glutathione (GSSG) to reduced glutathione (GSH) at varying total concentrations of glutathione were evaluated. Extracts were prior to expression of Fab pretreated with 100 µM IAM and afterwards supplemented with 13 µM DsbC. The CFPS reaction was carried out in the presence of radioactively labelled leucine which leads to incorporation of radioactively labelled leucine into newly synthesized protein which in turn allows quantitation of newly synthesized protein. Assuming that only correctly folded HC and LC and correctly folded Fab remains in the supernatant. The soluble yield after centrifugation was determined by liquid scintillation counting. The soluble protein yield was determined to be 48 µg ml^{-1} in our reference sample w/o gluthathione in the presence 13 µM of DsbC. In the presence of 2 mM total glutathione at a ratio of GSSG to GSH of 1:1 we achieved soluble yields of 213 µg ml^{-1} (Figure 29).

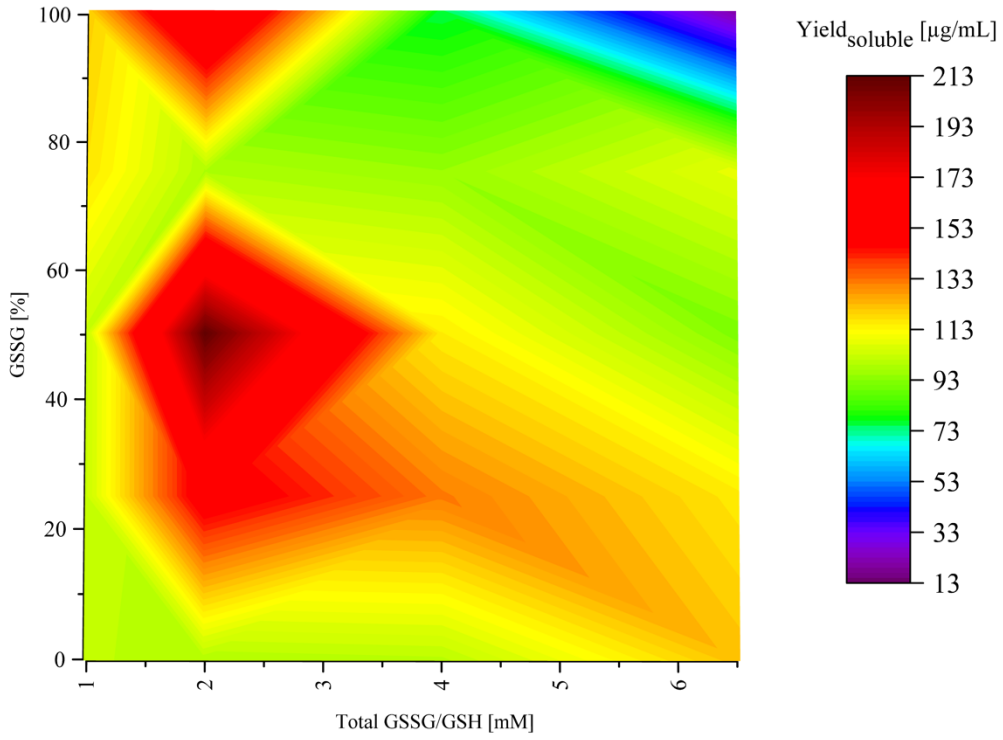


Figure 29: ***Correlated screen of the formation of Fab at various total concentrations and ratios of GSSG/GSH visualized as 2D-plot.*** *Radioactively labelled, TCA-precipitable protein in the soluble fraction was quantified using scintillation counting. The yield determined by scintillation counting was plotted against respective total concentration and ratio of GSSG/GSH to determine the optimal GSSG/GSH content.*

Compared to the reference w/o glutathione, this represents an increase in soluble yield of 300% and emphasizes the importance GSSG/GSH to folding of Fab. Based on optimization described in this chapter, we pretreated our extract with 100 µM IAM and used 2 mM GSSG/GSH at a ratio of 1:1.

4.4.5 Tuning functional folding during CFPS using chaperones

Fab and scFab were expressed using parameters (magnesium-glutamate; potassium-glutamate; IAM and GSSG/GSH) as determined previously. The antibody fragments were expressed in the absence of any chaperone (control reaction) and increasing amounts of either DsbC, FkpA or Skp were titrated to evaluate the effects of these chaperones on folding of Fab and scFab, respectively. The soluble fraction was analyzed using non-reducing autoradiography to quantify assembled, non-assembled and aggregates each normalized to the respective control reaction without any chaperones (Figure 30) and were then analyzed using ImageQuant software (GE-healthcare) and expressed as fold change as depicted in Figure 31.

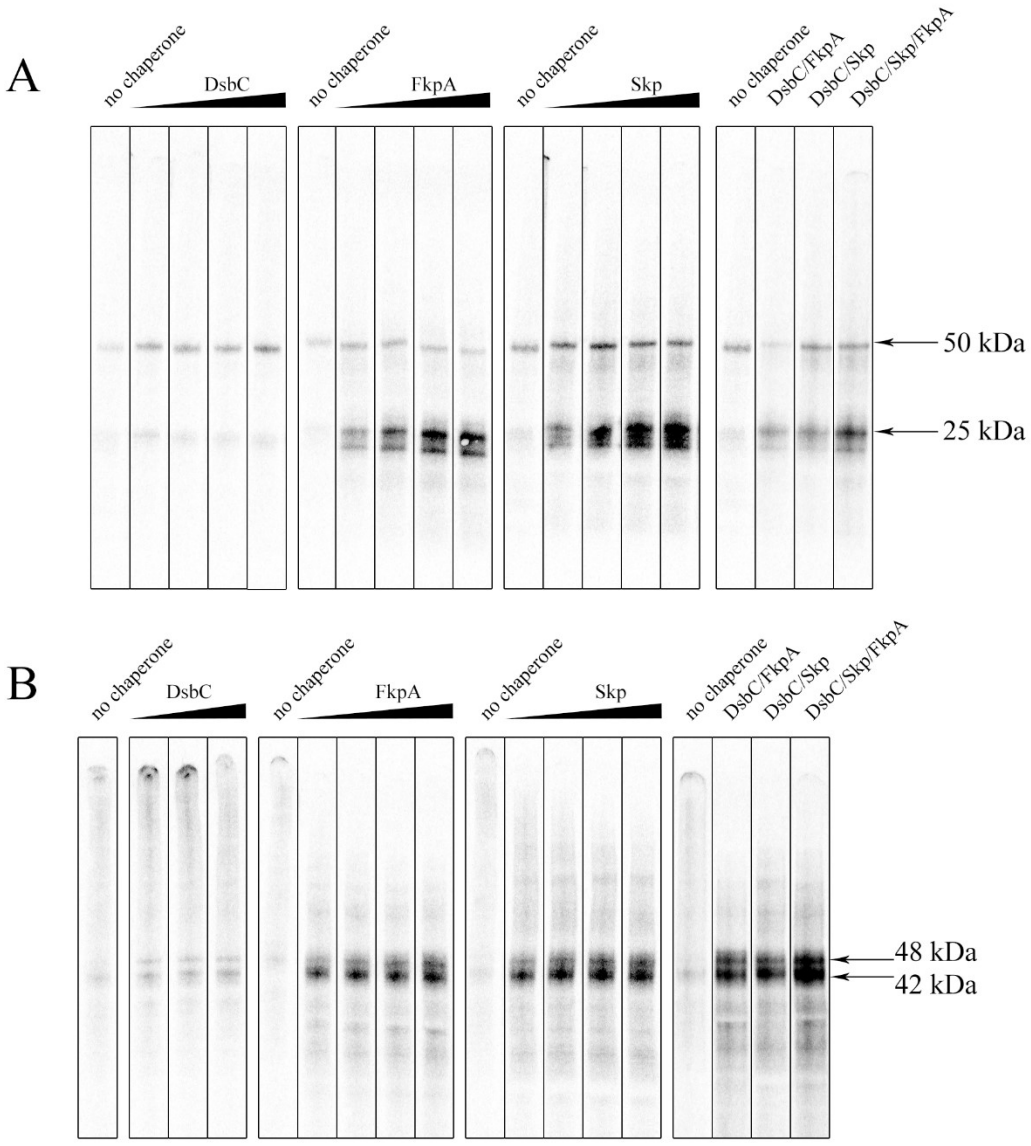


Figure 30: ***Analysis of soluble fraction after titration of DsbC/FkpA/Skp using autoradiography.*** *The soluble fractions after expression of Fab (A) and scFab (B) were analyzed after non-reducing SDS-PAGE and exposed onto phosphor screen (o/n). The phosphor screen was scanned using FLA-5100 and further analyzed using ImageQuant (GE-Healtcare)*

Correctly assembled Fab is composed of LC connected *via* interchain disulfide bond to HC (~50 kDa) and can therefore be easily distinguished from single HC (~25 kDa) and LC (~25 kDa) by molecular weight under non-reducing conditions (Figure 30, A). On the contrary, in scFab the LC and HC are joined by flexible $(SGGG)_4$ linker resulting in slightly higher molecular weight of 51 kDa as compared to Fab (50 kDa). Since scFab forms the same intra- and interchain disulfide bonds between LC and HC, it was expected to have a slightly higher apparent molecular weight due to the linker, but otherwise similar migration pattern after denaturation with SDS under non-reducing conditions as compared to Fab (Figure 30, B).

However, as shown in the autoradiogram (Figure 30, B) there were two bands observed under non-reducing conditions exhibiting an apparent molecular weight between 49 and 38

kDa. It was presumed that the upper band with an apparent molecular weight close to that observed for Fab (~48 kDa) represents correctly folded scFab, while the lower band with an apparent molecular weight of ~42 kDa contains non-native disulfide bonds resulting in a more compact amino acid chain and therefore has lower apparent molecular weight. Therefore, the upper band with an apparent mass of 48 kDa was quantified as natively folded scFab, while the lower band with an apparent mass of 42 kDa was quantified as non-natively folded scFab.

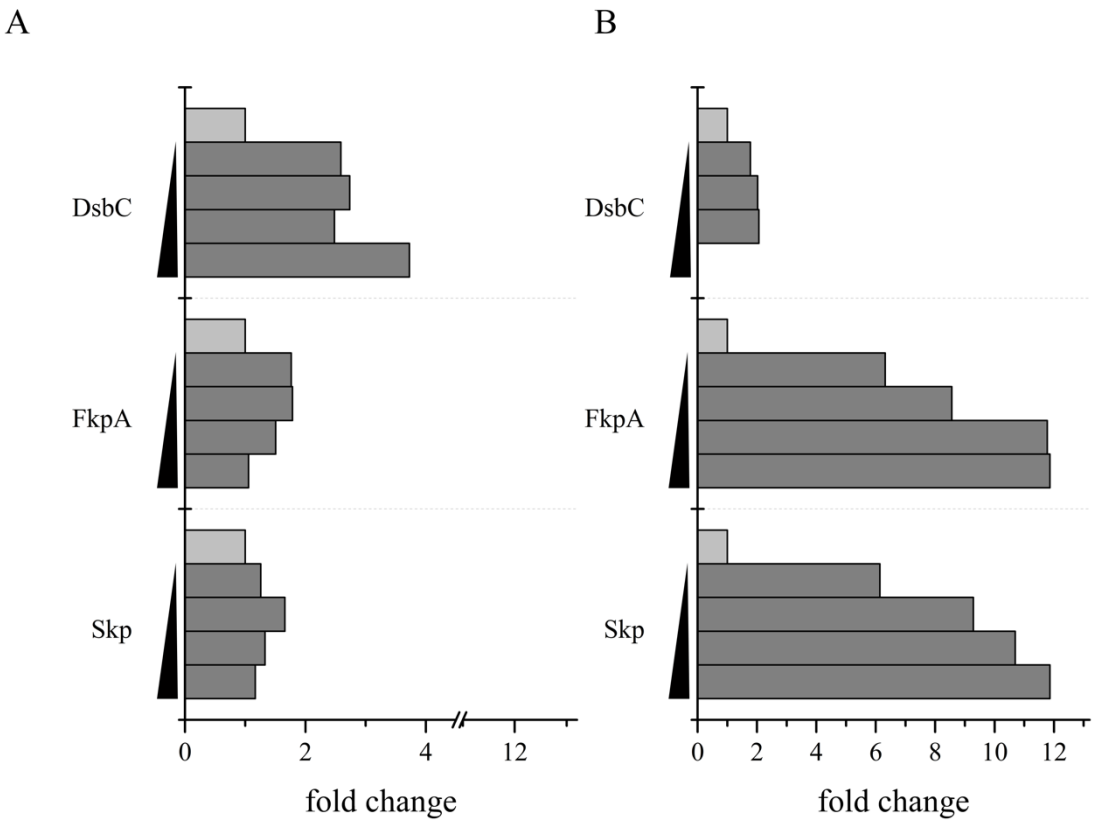


Figure 31: ***Fold-change of Fab(A) and scFab (B) as determined by autoradiography.*** *The intensity of the corresponding band on the autoradiogram was measured using ImageQuant (GE-Healtcare)* ***and normalized against corresponding negative control w/o chaperones to determine fold-change.***

The effect of purified DsbC on Fab was evaluated in concentration range of 13-73.5 µM on expression of Fab (Figure 32, A "DsbC"). Titration of DsbC resulted successively in 3-fold increase of assembled Fab correlating with the concentrations of DsbC. Increasing amount of DsbC during expression led to decreasing amounts of non-assembled HC/LC as well as aggregates, while the amount of assembled Fab successively increased (Figure 30). The amount of assembled Fab and non-assembled HC/LC was quantified using ImageQuant and was expressed as the binary logarithm of folded (assembled Fab) over unfolded (non-assembled HC/LC). The ratio of folded/unfolded resulted only for expression of Fab in the presence of DsbC in positive values (more assembled Fab than non-assembled HC/LC) that successively

increased with titration of DsbC (Figure 32 Fab/DsbC), meaning that titration of DsbC shifted the ratio of folded over unfolded more towards assembled Fab. The chaperones FkpA or Skp led in lower titration range up to ~70 µM and ~55 µM, respectively, to 1.5-1.7-fold change of assembled Fab that was reversed with higher concentrations of the respective chaperones (Figure 31 A "FkpA/Skp"). However, the ratio of folded over unfolded gave negative values in the presence of FkpA and Skp and shifted towards unfolded with titration of these chaperones, suggesting that these chaperones prevent proper assembling of HC/LC (Figure 32).

Even though yield of assembled Fab remained constant with titration of chaperones, the overall protein yield increased significantly in the presence of FkpA and Skp up to 400% as determined by liquid scintillation counting (data not shown) and indicative by the overall intensities of the autoradiogram (Figure 30).

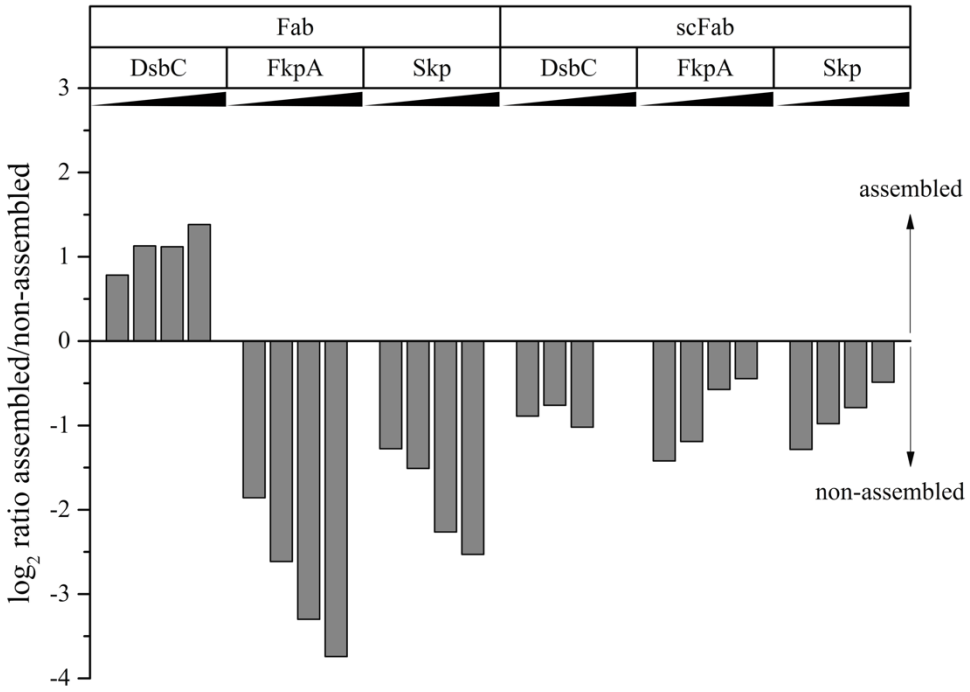


Figure 32: ***Logarithmized ratio of natively disulfide bonded protein over non-natively disulfide bonded protein as determined by autoradiography.*** *The intensity of the corresponding band on the autoradiogram was determined by ImageQuant software (GE-Healtcare) and the ratio was reported to assess the quality of the Fab and scFab.*

DsbC exhibited also positive effects on expression of scFab and resulted in up to 2-fold increase of natively folded scFab as compared to the control reaction w/o chaperone. The presumed natively folded scFab was only visible in the autoradiogram after titration 13 µM DsbC (Figure 30). The ratio of folded over unfolded scFab was negative and remained constant with titration of DsbC, suggesting that the joined expression of HC and LC is prone to form

non-native disulfide bonds that can hardly be resolved by DsbC (Figure 32 “DsbC”). Unlike Fab, the expression of natively folded scFab was greatly enhanced in the presence of FkpA and Skp exhibiting fold changes of up to 12 (Figure 31, B). The ratio of folded over unfolded remained negative, however contrary to Fab, it shifted towards folded with titration of these chaperones (Figure 32).

Taken together, the titration experiments imply that DsbC is most important and sufficient to improve assembling of Fab. Skp and FkpA increase overall yield of soluble protein, however, prevent proper assembling of HC and LC. DsbC has only little effect on folding of scFab as compared to Skp and FkpA. Titration of Skp and FkpA increase overall yield and amount of natively folded scFab significantly. Nevertheless, scFab is prone to form non-native disulfide bonds that can hardly be resolved. Combination of DsbC/FkpA, DsbC/Skp and DsbC/Skp/FkpA were also evaluated but had no additional or further positive effects on folding of Fab or scFab (Figure 30). For future experiments, based on the previously presented results it was determined that 13 µM DsbC for expression of cf-Fab and 55 µM of FkpA for expression of cf-scFab are optimal.

4.4.6 Functional Assays to verify binding of cf-Fab and cf-scFab

To verify that the cell-free produced antibody fragments are functionally folded and can bind their antigen MAPKAP1, a Western Blot using cf-Fab and cf-scFab was performed (Figure 33). MAPKAP1 was detected after transfer onto nitrocellulose membrane by either Fab- or scFab-containing cell-free extract as primary antibody source.

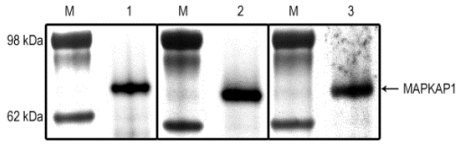


Figure 33: ***Western Blot analysis to assess the quality of the cell-free expressed Fab and scFab in comparison to CHO-derived IgG****. The antigen MAPKAP1 (72 kDa) was detected with either CHO-derived anti-MAPKAP1 IgG (mouse) (lane 1), cf-Fab (lane 2) and cf-scFab (lane 3). Anti-MAPKAP1 IgG was directly detected using HRP-labelled anti-mouse Ab, whereas cf-Fab and cf-scFab were detected with an anti-StrepII Ab.*

After several washing steps, StrepII-fused Fab and scFab were then detected by anti-StrepII mouse antibody, which in turn was detected using HRP-labelled anti mouse antibody. As negative control extracts w/o template DNA were used, and as positive control served the CHO-derived anti-MAPKAP1 antibody. The antigen MAPKAP1 was not detectable in the negative control using extract w/o template showing that our secondary/tertiary antibodies are specifically binding the antigen MAPKAP1 (data not shown). The positive control, cf-Fab and cf-scFab could be detected and resulted in a clear band between 98 kDa and 62 kDa as compared

to the molecular weight marker, consistent with the predicted molecular weight of the antigen MAPKAP1 (72 kDa). This result shows that cf-Fab and cf-scFab bind the antigen MAPKAP1 comparable to CHO-derived full-length antibody in Western Blot analysis.

5 Discussion

5.1 Protein Profiling and Systems Approach to Optimize CFPS

Initially developed as a research tool in 1961 by Nirenberg and Matthaei to decipher the genetic code, cell-free expression systems have become a standard protein expression system by either complementing or substituting bacterial and eukaryotic expression system. Due to their open nature and the use of the protein synthesis machinery without a vital cell, cell-free expression systems are especially suited to express membrane protein, toxic proteins and other difficult-to-express proteins, including antibodies, by providing direct access to the folding machinery and allowing the addition and optimization of folding factors using high throughput methods.

Theoretically, the PURE system derived from *E. coli* - being a defined platform made from purified individual components - should be an ideal expression system, since unnecessary metabolic pathways directing energy resources away from the protein synthesis machinery are absent, and resources are focused on energy regeneration for protein synthesis. Despite this advantage of providing a highly defined protein synthesis and energy regeneration environment, the productivity of this system is relatively low and the preparation quite costly and labor intensive, which is therefore limiting its application. The most productive and easy to use cell-free expression systems are based on prokaryotic extracts derived from *E. coli* after centrifugation at 30,000 *g* (S30 extract). In addition to S30 extract, cell-free expression of proteins based on S12 or S15 extracts have been reported [68]. The S30 lysate represents the best compromise between complexity and efficiency, however and was thus selected for further analysis and is the subject matter of this dissertation. The *E. coli* A19, a strain commonly and historically utilized for cell-free protein expression, was used for S30 extract preparation.

These S30 extracts are more efficient when compared to the PURE system. The preparation of S30 extracts comprises the growing of cells in a rich and complex medium, harvesting the cells in mid-log phase and several washing steps followed by cell disruption *via* high pressure homogenization. For the protein synthesis activity of the resulting extract, it is crucial that extract is fractionated using centrifugation and to perform a *run off* step by incubation at 37-42 °C. The latter steps result in the separation of a membrane fraction and heavy precipitation of proteins, and therefore a significant alteration of the proteome from exponentially growing source cells, resulting in a yet unknown proteome. Cell-free reactions contain 30-50% of these S30 extracts with a final protein concentration of 13-30 mg mL^{-1}. Despite being a core technology within the field of synthetic biology, the lack of information about the proteome background still poses a restriction when addressing common problems

such as low protein quality, high assay backgrounds, proteolytic degradation or label scrambling. This dissertation is closing the gap by providing a comprehensive blueprint of the S30 extract proteome as derived from the most commonly applied extract preparation procedure. Hereby, approximately one-fourth of the total *E. coli* proteome building the core S30 extract proteome was identified. It was also demonstrated that certain common modifications during S30 lysate preparation do not result in major changes in proteome composition, showing that the lysate proteome is representative of the most commonly used S30 extract preparation procedures. The initial analysis of two slightly different but commonly used preparation procedures is described in detail in 4.1.1. The most abundant proteins, including proteins needed for energy regeneration and protein synthesis, represent the core proteome (68%) and were common in $S30^{+D(1-4)}$ and $S30^{-D}$ extract. The 209 and 191 proteins exclusively found in $S30^{+D(1-4)}$ and $S30^{-D}$, respectively, are mostly low abundant proteins that were identified with very few peptides. It is not clear whether those differences arise due to the slightly different preparation procedures and/or technical/biological variability. It is assumed, however, that the technical variability is responsible for the differences observed as the variability of the proteome overlap was similar for quadruplicates derived from the same preparation procedure. More important determinants of the S30 lysate proteome could be the growth phase of the cells at the time of harvest.

Having concluded that slight modifications of the preparation procedure have probably minor effects on the proteome of S30 extract, the subsequent analysis focused on the analysis and classification of proteins present in $S30^{+D(1-4)}$ due to the availability of multiple biological replicates resulting a more reliable and comprehensive blueprint of the S30 extract proteome as well as its classification. In total, 1074 proteins were uniquely identified in either one or multiple S30 extract replicates, representing approximately 25% of the known *E. coli* proteome (4315 proteins, as predicted in the high-quality reference proteome UniProt ID: UP000000625) based on sequencing, ORF prediction and experimental evidence. However, many of the 4,315 proteins present in the reference proteome are regulated and might therefore not be expressed at our conditions, and some predicted proteins may not be present at all. It has been shown by transcriptomics and LC-MS/MS analyses employing extremely long gradients that exponentially growing *E. coli* cells, as they are used for S30 extract preparation, are expressing approximately 2,600 proteins, of which 830 proteins are membrane proteins [127]; this number might provide the most accurate estimation of theoretically expected proteins present directly after cultivation but before extract processing. However, extract processing includes a centrifugation step at 30,000 *g* that is central to S30 extracts and represents a subcellular fractionation resulting in the sedimentation of mainly membrane and membrane associated

proteins. Thus, it was assumed that a fraction of at least 32% (830 membrane proteins out of 2600 expressed proteins corresponds to approx. 32%) of proteins is removed during extract processing, not even considering the *run-off*, which results in heavy precipitation of soluble proteins. Thus, a conservative estimation of theoretically detectable proteins in S30 lysate would be approximately 1770 proteins. This would imply that at least 60% of the theoretical S30 extract proteome was identified, and a fraction of 40% is currently unaccounted for; these proteins may be removed during extract processing or may be undetectable. Notably, the detected number of proteins is also determined by the implemented sample preparation methods for LC-MS/MS, by the selected ionization method and by physiochemical properties of proteins/peptides. Therefore, there can only be speculation about the fraction of proteins that are removed during extract processing, but it can be reliably stated that at least 1074 proteins in S30 extract can be present.

It is also important to point out that the S30 extract was separated on the protein-level (SDS-PAGE) prior to fractionation on peptide-level, which allows for covering a larger dynamic range of protein concentrations and detects low-abundant proteins. Nevertheless, it is more likely that highly abundant proteins will be identified compared to less abundant proteins and more abundant proteins may therefore be overrepresented in our analysis. To provide an estimation of the abundance of identified proteins, the emPAI value was reported. This value is based on the number of identified observable peptides. Previous studies showed that these values correlate well with the independently measured copy number in *E. coli* cells [128].

For the classification of the S30 proteome, the analysis focused on repeatedly detected proteins, present in at least three out of four replicates, which counted 821 proteins or 76.4% of the uniquely identified proteins. The low “Cell Periphery” fraction is expected as most membranes will be removed by the 30,000 g centrifugation steps during S30 extract preparation. Interestingly, some membrane proteins still remain in the lysate potentially in residual small vesicles [129]. Particularly abundant components of the respiratory chain involving several subunits of the ATP-synthase and the NADH-quinone oxidoreductase were detected. This agrees with reports on the cytomin cell-free system employing inverted inner membrane vesicles of S30 lysates for energy regeneration *via* oxidative phosphorylation [57, 130]. The identified ATP-synthase complex and respiratory chain constituents are therefore likely to be present in active conformation and may be used for energy regeneration. A further interesting observation is the presence of parts of the Sec translocon machinery, which could be beneficial for the cotranslational insertion of membrane proteins in supplied lipid bilayers of liposomes or nanodiscs [92]. Recent studies suggest that insertion *via* the translocon is thermodynamically similar to spontaneous insertion [131], while the major task of the

translocon machinery might rather be to guarantee targeting and coordinated membrane protein assembly [132].

In addition to proteins integrated into the inner bacterial membrane, several membrane proteins of the outer membrane were also identified. The presence of certain porins of the outer membrane such as OmpA or OmpF was previously speculated based on detected background activities in S30 lysate during single channel electrophysiology measurements [66]. It is worth noting that the GO term classification of the 74 detected proteins as membrane proteins should be viewed with caution as the GO term also includes membrane associated proteins such as the enzyme MurG that behaves similarly to soluble proteins [21, 133]. Moreover, soluble subunits of complexes that are part of membrane spanning proteins such as the group of ABC-transporters are included. None of the identified 15 ABC transporter related gene products are considered to be a multispanning transmembrane proteins. Mostly periplasmic solute binding subunits, such as ArtI [134] and ProX [135], as well as the MlaC subunit of the MlaFEDB transporter complex, were detected [136].

The lack of integral membrane proteins may be caused by a technical bias towards more hydrophilic peptides in the course of GeLC-MS/MS analysis. It is therefore impossible to distinguish if detected membrane complexes are likely reconstituted natively in vesicles or if the complex is destroyed during S30 lysate preparation and only soluble subunits remain in the supernatant. The example of the respiratory chain/ATP synthase, however, provides strong evidence that it is basically possible that membrane embedded complexes remain active in S30 lysate [130]. Accordingly, these genuine membrane-spanning proteins were repeatedly identified with high scores in our proteome analysis. Because we used the GeLC-MS/MS approach based on 1D-SDS-PAGE, however, it was expected that membrane protein are efficiently solubilized and fractionated, as has been demonstrated previously [137]. Therefore, it is unlikely that this bias has primarily technical reasons but is rather caused by membrane depletion during S30 extract preparation.

The initiation of proteins synthesis is a highly regulated process to ensure the correct interplay of ribosomal subunits and accuracy of the translation. Besides the ribosomes and the mRNA itself, the presence and the interplay of initiation factors 1-3 (IF1-3 represented by the proteins InfABC) is crucial for efficient translation initiation [138]. Considering the ratio of averaged emPAI values of $S30^{+D(1-4)}$ and $S30^{-D}$ for infA:infB:infC of 4.2:1:3.2 (Table 6), it is actually surprisingly close to the ratio of initiation factors used in the reconstituted PURE system. Here, the concentrations were extensively optimized and resulted in an optimized ratio of 6.75:1:3.75 for infA:infB:infC [65]. It should be noted, however, that the PURE system exhibits significantly lower productivity compared to S30 extract-based systems. In addition,

based on the MS approach, a conclusion cannot be drawn with respect to the absolute concentration of these factors in S30 extract, which might be important, as the concentration of initiation must be in balance with the number of ribosomes present in the S30 extract.

During exponential growth, the elongation factor EF-Tu is the most abundant protein, reaching an EF-Tu:ribosome ratio of 8:1 [139]. Consistent with that statement, the emPAI values for EF-Tu belongs to the group of most abundant proteins in S30 extract and the relatively high abundance of this factor during S30 extract preparation appears to be preserved. Unfortunately, the emPAI values of the ribosomal proteins in the analysis did not allow the estimation of the ratio of ribosome:EFs, as the emPAI values for the ribosome scattered strongly; however, it has been shown that there is approximately 1 molecule of EF-G and 0.2 molecules EF-Ts per ribosome. This ratio does not appear to be preserved based on the approximate estimation by averaged emPAI of elongation factors. The concentration of EF-Ts seems to be even slightly higher compared to EF-G.

Interestingly in this context, Swartz and coworkers added three purified elongation factors at the abovementioned ratios and were able to increase the elongation rate and yield [140]. The data at hand may support the notion that the elongation factors are not optimally reconstituted in cell-free expression systems. The fourth elongation factor, namely, EF-P, is required to alleviate ribosome stalling at three or more consecutive proline residues. This elongation factor appears to be less abundant compared to others, which may, however, still be within physiological range. When expressing poly-proline containing target proteins, the effect of supplementing EF-P on synthesis rate and yield could be interesting to evaluate. The additional elongation factors LepA and YeiP, postulated paralogs of EF-G and EF-P, are assumed not to be essential and might increase translation fidelity. Even though the elongation factors may be overall abundant, they interact as a ternary complex consisting of tRNA, amino acids and elongation factors. A limitation of one of these factors (*e.g.* during translation of repetitive codons) can result in delay or ribosome stalling [141].

S30 extracts are mostly used in a combined transcription/translation reaction based on the bacteriophage derived T7 polymerase. This leads to a rapid accumulation of free mRNA, which is prone to degradation by the RNA degradation machinery as well as tertiary structure formation inhibiting efficient translation. Supplying RNase inhibitors to cell-free reactions is therefore a common measure. Degradation and mRNA folding issues could be reduced by using endogenous RNA polymerase for transcription. The latter strategy depends on the presence of endogenous transcription systems comprising DNA dependent RNA polymerase as well as compatible sigma factors in the S30 lysate [142]. The analysis show that all four subunits of the *E. coli* RNA polymerase core enzyme were identified, and, overall, high emPAI values

indicate abundant copies of these proteins within the lysate (Table 5). Three out of the seven *E. coli* sigma factors are present, and the highest emPAI value was obtained for sigma70 (RpoD), which is necessary for recognition of general housekeeping promoters. The lysate also contains the cold-shock transcription factors CspA, CspC, CspE and CspG. It is tempting to speculate that these transcription factors are induced due to the slower cooling of the broth during $S30^{+D(1-4)}$ preparation compared to $S30^{-D}$, which is rapidly cooled with a cooling coil.

Chaperones are important modulators of protein quality and approximately 10% of proteins contain at least one chaperone-dependent folding step [143]. Furthermore, individual stages of certain protein folding pathways do require the synergetic interplay of different chaperones. Depletion or an insufficient copy number of individual chaperones in S30 extract could result into the formation of improperly folded or high levels of aggregated proteins concomitant with lower specific activity or precipitation. Many endogenous chaperones have been identified at high concentrations in the S30 extract (Table 7), but concentrations might become rapidly limiting, especially in the preparative scale expression of chaperone-dependent target proteins. Since the cytoplasmic environment in *E. coli* is highly reducing, it does not allow the formation of disulfide-bonded proteins. In contrast, the periplasmic space maintains an oxidizing environment and contains important enzymes such as DsbA/B/D/E/G that catalyze in cooperation the formation of native disulfide bonds.

Nevertheless, given the low abundance of disulfide-linked proteins in *E. coli*, it can be assumed that the machinery for disulfide bond formation has not evolved for high-level expression of disulfide bonded proteins *in vivo* and is therefore also suboptimally represented in S30 lysate, partially due to the separation of membrane fraction and membrane associated factors during extract preparation. While the oxidoreductase DsbA and its regenerating counterpart DsbB can easily be replaced by a chemically maintained redox potential using various ratios of GSH/GSSG, the disulfide isomerase enzyme is essential for the formation of most native disulfide bonds. Especially during the production of antibodies in CFPS systems, DsbC or eukaryotic counterparts are often supplemented to improve disulfide bond formation in cell-free protein synthesis reactions [144, 145]. The proteome profiling at hand confirms the notion of DsbC as a limiting factor in S30 lysates.

The detection of a significant variety of proteases in S30 lysates is in accordance with the reported requirement for protease inhibiters in cell-free expression reactions in order to maintain the integrity of S30 lysate as the production platform, as well as the target proteins [91, 146].

To date, little attention has been given to the growth conditions of *E. coli* prior to extract preparation. It is anticipated that extract derived from exponentially growing cells provide the

most adequate S30 extract to achieve a high protein synthesis rate and in turn a high protein yield. However, Failmegzer *et al.* have showed that a large portion of the ribosomes in S30 extract derived from fast growing *E. coli* are not actively involved in the translation process. The limited number of ribosomes in S30 extracts derived from slow-growing and stressed *E. coli* can have a synthesis rate comparable to ribosome derived from fast-growing *E. coli,* and it was also shown that the translation apparatus is optimally reconstituted in such extracts [74]. While it is certainly true that the protein synthesis rate during exponential growth is at its maximum, it was hypothesized that modified growth conditions may add benefits that balance or outweigh the potential loss of high protein synthesis capacity. Exposure to elevated temperature or chemicals are well-known inducers of so-called heat shock responses in *E. coli*, resulting in the increased production of chaperones and other rescue or SOS-response proteins [75, 147].

To take advantage of the increased chaperone content following such a SOS-responses, a S30-S lysate after exposing *E. coli* A19 to a heat shock at 42 °C for 45 min in the presence of 3% (v/v) ethanol before the harvest was prepared. The proteome was quantitatively compared to the standard S30 extract, and the effect of the altered proteome on protein synthesis was evaluated using different difficult-to-express model proteins. In total, 26 proteins were upregulated and another 56 proteins were downregulated (Table 10, Table 11) in S30-S lysates relative to the standard $S30^{+D(1)}$ extract (Figure 10).

As predicted, most regulated proteins belong to the general stress response induced by heat shock or ethanol stress [72, 75, 148]. The 26 upregulated proteins are mainly chaperones or factors related to transcription/translation. In particular, GroEL, GroES, DnaK, GrpE and SecA are 3-10-fold upregulated in S30-S lysates. The chaperone HslR acting as recycling factor for ribosomal subunits and the periplasmic chaperone Spy were exclusively detected in S30-S lysates. Some upregulated proteins detected such as RpoH and ManX match with those previously found after growth of *E. coli* at ethanol stress [148]. However, the alcohol and aldehyde dehydrogenases YqhG and AldB, as well as other known ethanol induced proteins, were either not found or were not regulated.

These proteins could have been removed during S30 lysate preparation or they might accumulate only after extended SOS induction periods. The SOS response negatively affects the translation efficiency and the ethanol and heat treatment of the cell culture was therefore limited to 45 min in order to maintain sufficient protein expression efficiency in subsequent cell-free expression reactions. Other or modified procedures could certainly result in a different chaperone portfolio in the resulting lysate but also in different expression efficiencies. S30 extract preparation from *E. coli* cells grown continuously at 42 °C for chaperone enrichment

has been reported [149]. In our hands, however, the protein production efficiency of such lysates was reduced to only 30% of comparable standard S30 lysates (data not shown). A future strategy could thus be to apply adjusted mixtures of different lysate preparations in order to generate optimal synergies of high expression efficiency and suitable folding background. A comparable investigation to determine actively translating risbosome, as shown by Failmezger *et al.,* would be interesting to perform [74].

The GO term response to temperature covered almost all chaperones mentioned above but also further members of the SOS network such as the Lon protease mainly responsible for the degradation of protein aggregates, as well as supporting chaperone activity by complex formation [150, 151]. Not surprisingly, the heat shock specific sigma factor 32 (RpoH) was also found in this group [152, 153] (Figure 15). Interestingly, the transcriptional termination factor Rho is also upregulated, potentially providing more efficient Rho-dependent termination activity. In contrast to the upregulated protein fraction, the group of 56 downregulated proteins is more diverse (Figure 14), but includes mostly metabolic enzymes, which were integrated using the source KEGG pathway. We found that most enzymes in the downregulated group belong to the central metabolism such as pyruvate metabolism, TCA cycle and aerobic respiration. Since the growth rate is significantly reduced during heat/shock/ethanol exposure, this effect could be related to lower energy demand during this period.

In addition to central metabolic proteins, connected pathways such as enzymes involved in cellular amino acid metabolic processes are also downregulated. In particular, several amino acid manipulating proteins such as glyA, cysM, and gcvT, as well as components necessary for transcription and translation are less abundant (Figure 14). Here, proteins such as the downregulated CysS (cysteine tRNAligase) should be pointed out as one possible limiting factor for cysteine containing proteins (Figure 14). This might contribute to the generally lower protein production efficiency of S30-S lysates. In view of amino acid scrambling, however, the reduced protein content of amino acid modifying enzymes, especially GlyA, could be beneficial [24].

Rational alternatives could be the addition of selected chaperones such as the GroL/S or the DnaK/J systems into the cell-free reaction mixtures [154, 155]. However, additional workload for the production and purification of the chaperones would be necessary and the effect of one or few chaperones might be different from that of a complex chaperone network.

5.2 Gene Editing to Improve CFPS

After generating a comprehensive framework of the S30 extract, this information was used to rationally design an optimized *E. coli* strain. A project around this topic was kindly funded by BMBF/KMU-innovativ, in a partnership with Biotechrabbit GmbH (Drs. Merck and Gerrits). Despite not being included in the proteome profiling, a derivative of strain *E. coli* D10 (CGSC#: 6587) was also used as a target for genetic modification, since it is also a well-established source strain used for S30 extract preparation at Biotechrabbit. Due to the similar genetic background, major changes on the proteome profile were not expected. The rationale for choosing the target genes was previously described (see Section (4.3.1)

For *E coli* A19, the modifications comprised the addition of an SBP tag to rne, followed by the successive deletion of gshA, endA, sdaA, speA, sdaB, tnaA, ompT, lon, gor and fhuA. As shown in the fitness evaluation (Figure 21), tagging did not have any effect on growth rate, as expected, and the first drop in growth from 1.5 to approx. 1.3 h^{-1} was observed after removal of *gshA*, implying the important role of this enzyme during growth. GshA catalyzes the conversion of L-glutamate and L-cysteine under consumption of ATP to L-γ-glutamyl-cysteine. Due to the removal of *gshA*, the cell is no longer capable of producing glutathione, which is an important factor in addressing reactive oxygen species. Oxidative stress and/or problems with proteins containing cysteines and/or requiring native disulfide bonds may be the reason for the decreased growth rate observed. L-glutamate is used in many cell-free reactions in order to mimic the physicochemical environment of the living cell; it is present in large quantities (80-200 mM) and is therefore not limiting. The consumption of ATP and L-cysteine by unproductive pathways, however, is counterproductive. L-cysteine is present at a concentration of only 2 mM and is known to be rapidly depleted by gshA [44]. Regrettably, the depletion could not be reproduced due to problems during the analysis of the cysteine dataset. The course of the L-glutamate concentration fell to approximately 60 mM after 30 min of incubation in *E. coli* A19 ΔendA extract and remained more stable in the *E. coli* rne::SBP ΔgshA/endA/sdaA/speA/sdaB/tnaA (referred to as A19 rne:SBP CSRW) at approximately 70 mM (Figure 25, L-glutamic acid). The stabilization may be partially attributed to the removal of *gshA*; however, other factors also seem to play a role since an increase by maximum 2 mM would be expected by removal of *gshA* alone, as there are only 2 mM of L-cysteine in the reaction mixture.

Experiments with labeled L-glutamate showed that L-glutamate is converted to succinate, acetate, malate, aspartate and oxalacetate in S30 extract by enzymes of the central metabolism [130]. With the decrease in consumption, it was also observed that the formation

of aspartic acid is decreased, which suggested that the activity of the central metabolism is decreased due to the combined elimination of gshA, endA, sdaA, speA, sdaB and tnaA. In addition to cysteine, during the reaction catalyzed by gshA, precious ATP is converted to ADP and inorganic phosphate, limiting the available energy, and resulting in the accumulation of inorganic phosphate, which has also been shown to inhibit cell-free protein synthesis [58].

The *E. coli* A19 rne::SBP ΔgshA was used as a source to produce S30 extract and was compared to extract derived from A19 wt (Figure 26). The production of CAT (chloramphenicol acetyltransferase) containing five cysteines resulted in an increase of median yield from 146 µg/mL to 223 µg/mL, an increase of approximately 50%. The increase may be attributed to the increased availability of ATP and cysteine in the reaction mixture. Following these encouraging results, the genes sdaA, speA, sdaB and tnaA in conjunction with endA were knocked out, and their removal has been shown to stabilize the amino acids L-serine, L-arginine and L-tryptophan. The evaluation of the fitness after each knockout proved that in contrast to gshA, the removal of sdaA/B, speA and tnaA had only minor effects on the growth rate, leading to a combined decrease from 1.3 to 1.2 h^{-1}.

From the resulting strain A19 rne:SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA), S30 extracts were compared and the course of the amino acid concentration was followed and compared to the amino acid profile of A19 Δ(endA). As described by Calhoun et al. [44], the results are in agreement with the depletion of the amino acids L-arginine and L-serine within 30 min and 15 min, respectively. The knockout of SpeA (Arginine decarboxylase), which catalyzes the conversion of arginine to agmatine and CO_2, shows a highly stable arginine concentration over the whole incubation time and therefore provides evidence that SpeA is responsible for the rapid depletion of arginine (Figure 25).

The amino acid L-serine was similarly depleted in A19 Δ(endA) and remained constant in A19 rne:SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA), where the stabilization is caused by removing sdaA/B. As opposed to Calhou *et al.* in the analysis at hand, no depletion of tryptophan was observed and therefore no specific effect of tryptophanase can be expected. The combined removal of the gshA/endA/sdaA/speA/speB/tnaA in A19 resulted in a median yield of approx. 275 g/mL of CAT.

Overall, an increase of more than 85% yield was achieved by stabilizing the amino acids arginine, serine and cysteine. The knockout of tnaA probably does not contribute to this effect. Interestingly, in the hands of Biotechrabbit, who expressed the large protein AhR (Arylhydrocarbon-receptor) and the EFTs (elongation factor temperature sensitive), the opposite effect on protein expression was observed, resulting in a slight decrease of expression

efficiency for AhR in comparison to A19 wt and a substantial decrease by approximately 30% for EFTs. This suggests that the effects are highly protein-dependent and supports the observation of Nieß *et al.* that the amino acid sequence is crucial, resulting in limitation of the ternary complex (consisting of the amino acids, tRNA and elongation factor). The stabilization of the free amino acids is therefore only one part of the problem, and the availability of all constituents of the ternary complex needs to be ensured to address the underlying limitation [141, 156].Furthermore, the energy regeneration system could play a role since in the experiment, the energy regeneration system PANOx-SP was used, whereas Biotechrabbit used a different system (not revealed by Biotechrabbit). It is likely, however, that Biotechrabbit's system is not using spermidine and putrescine in their energy mix, which has been shown to activate intrinsic pathways. Instead, they probably used PEG (polyethylene glycol) and possibly acetate salts which has been shown to inhibit some pathways that can be employed for energy regernation [57].

While most of the remaining amino acids were stabilized after genetic modification, the amino acids asparagine and glutamine were also depleted within 15 min and at least 45 min (Figure 25) and provide an opportunity to further improve CFPS reactions by stabilizing building blocks. The first two datapoints of glutamine could not be considered due to high scattering and therefore it is unclear when exactly glutamine was depleted in S30 extract. The decrease of glutamine and asparagine was also described by Calhoun et al.; however, the concentration remained above 0.5 mM and 1 mM for 180 min for glutamine and asparagine, respectively. The concentration of threonine also decreases substantially over time from 2 mM to 0.5 mM. Albeit not limiting, the initial concentration of threonine should not be decreased for CFPS to avoid any limiting effect of this building block or the corresponding ternary complex.

The strain BL21(DE3) Star, commercially available from ThermoFisher Scientific, is also a popular *E. coli* strain for protein production *in vivo* and *in vitro*. The genetic modifications that are held responsible for the high productivity of this strain are mainly the deletion of the outer membrane protease *ompT* and the major protease *lon*. In addition, this strain carries a truncated version of *rne*, also know as *rne131*, lacking the part of RNaseE responsible for degradation of mRNA [157]. It is assumed that the truncation results in an increased mRNA stability and therefore higher protein yield. It was hypothesized that this would also be beneficial for cell-free protein synthesis. Since the truncation is protected by patents (*e.g.*, US20080153128A1) and therefore cannot be used in a commercial context, it was decided to circumvent this patent by tagging the essential RNase E. This way RNase E can be completely removed during extract preparation using an affinity column. The SBP tag was therefore chosen

because of its high affinity to streptavidin of 2.5 nM. The method for removing tagged proteins during extract preparation was well established in Biotechrabbit's Lab and used on a daily basis; however, no positive effect on yield and stability of CFPS could be shown (data not shown). Therefore, this approach was not further investigated.

The knockout of ompT, lon and gor and fhuA were investigated in the amino acid stabilizing genetic background where gshA/endA/sdaA/speA/sdaB were knocked out. Furthermore, to investigate the effect of those knockouts apart from the amino acid stabilizing knockout, it was decided to introduce also the knockout into the A19 rne::SBP background. The fitness analysis showed that the removal of ompT had absolutely no effect on growth rate in the A19 rne::SBP background and a rather positive effect on the amino acid stabilized genetic background. In contrast, the knockout of the highly conserved protease lon, which plays a major role in protein degradation and therefore protein quality control, lead to a decrease of growth rate in the amino acids stabilized genetic background and slightly higher but similar results in the A19 rne::SBP background (Figure 21). Nevertheless, in agreement with being a popular work horse for protein production, the strain remains vital and fit after removing *ompT* and *lon*.

When using the strain A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tna/ompT/lon) to express CAT, a drop in the amount of synthesized soluble protein from 275 µg/mL to approximately 150 µg/mL was observed, a level similar to A19 wt reversing the effects of the amino acid stabilization. Again, in the hands of Biotechrabbit, different effects were observed. In case of the expression of AhR, the introduction of ompT/lon in the amino acid stabilized genetic background increased from 124 to 185 µg/mL and even reached 210 µg/mL when using extract derived from A19 rne::SBP Δ(ompT/lon) after 30 min of incubation. This corresponds to an increase of 70% due to the effect of ompT and lon alone.

When expressing EF-Ts, a yield of 350 and 404 µg/mL was observed for extract derived from A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA/ompT/lon) and A19 rne::SBP Δ(ompT/lon), respectively. Compared to A19 wt, this corresponds to a decrease of 60 µg/mL in the amino acid stabilized background and approximately the same level compared to A19 rne::SBP Δ(ompT/lon) derived extract. It should be considered in this context that the analysis described in Figure 26 only considered soluble protein after a centrifugation step, whereas in the analysis depicted in Figure 27, there was no centrifugation and therefore an insoluble protein was measured. A possible explanation is that the presence of insoluble proteins lead to the aggregation of folding intermediates and was therefore separated due to the centrifugation step.

The insoluble aggregated protein, usually detected and degraded by the proteases lon and ompT, remained in the supernatant and resulted in a higher total yield. It must be

determined, however, if this protein is soluble and active in order to determine the benefits of knocking out ompT and lon. The knockout of ompT and lon may increase the total yield but may decrease the yield of desired natively folded and enzymatically active protein.

The knockout of gor and fhuA resulted in a slight decrease in the growth rate from 1.2 to 1.1 h^{-1} in the amino acid stabilized background. The decrease seemed to be mostly caused by the removal of gor, whereas the removal of fhuA did not decrease the growth rate further (Figure 21). In contrast, the growth rate was more decreased by removal of fhuA in the A19 rne::SBP Δ(ompT/lon) background, resulting in the same growth rate of 1.1 h^{-1} (Figure 21). These strains were evaluated with respect to their effect on protein synthesis in a CFPS system. It was observed that the removal of both gor and fhuA had a significant impact on the protein synthesis rate, resulting in a CAT yield below the level of A19 wt. It should be noted, however, that the knockout of gor was not expected to have a positive effect on CAT synthesis since it was removed to improve the expression and folding of the disulfide bond containing proteins such as antibodies by stabilizing the content of added oxidized GSH and prevent its reduction to GSSH. This strain is further evaluated in chapter 4.4 expressing otherwise difficult-to-express proteins in *E. coli*, namely, Fab and scFab. It is not clear why the outer membrane protein fhuA also had such a negative impact on protein expression. It may be related to the coregulation of genes in response to the knockout, which will be discussed later in this chapter.

As required by the project partner Biotechrabbit, the D10 strain was modified by removing of ompT and lon in conjunction with tagging CysS (Cysteinyl Synthetase) and the RF1 (release factor 1). This modified strain is used to introduce unnatural amino acids into proteins during cell-free protein synthesis by two methods. The first method includes the removal of CysS and RF1 during extract preparation, whereas a modified CysS was modified by site-directed mutagenesis to accept a certain nonnatural amino acid. Since RF1 and CysS are tagged with the same affinity tag, both factors can be removed in one processing step. Alternatively, only RF1 is removed to introduce nonnatural amino acids by the conventional method. RF1 must be removed to suppress the amber stop codon. Interestingly, despite being a close derivative of *E. coli* A19, the strain showed a low growth rate of 1.1 h^{-1} before any genetic modification was performed and remained constant; even the knockout of lon had no influence on the growth rate. This suggests that other modifications are present in *E. coli* D10 that were limiting the growth rate of this strain.

In addition, the results imply that the effects of genetic modifications cannot always be evaluated individually as there are synergic effects of certain knockout with other genetic modifications, encoding unknown modifications.

The impact of knockouts on the overall expression profile is often neglected. The transcription profiling of the highly genetically modified *E. coli* A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA/ompT/lon/gor/fhuA) in comparison with the source strain *E. coli* A19 wt was therefore analyzed. While the growth rate was still in a good range to grow *E. coli* for CFPS and the strain appeared vital, the transcription profiling revealed a large number of genes that were differentially regulated in response to the introduced knockouts. Regrettably, transcriptional profiling could not be performed after the introduction of every individual knockout due to cost and time constraints; therefore, the regulation of particular genes or gene groups cannot be attributed to an individual knockout and the exact cause of coregulation can only be speculated. The genes that were targeted for knockout, namely, gshA, endA, sdaA, sdaB, speA, tnaA, ompT, lon, gor and fhuA, were all highly downregulated in comparison to A19 wt with $\log_2$ values smaller than -4. This confirms, in addition to the previous analyses, that no copy of any of the target genes exists in the genome any more (Figure 22).

In total, 309 (299 excluding the knockouts) genes were downregulated (Table 16) and 153 genes were upregulated (Table 15). To find patterns, the genes excluding the knocked-out genes were functionally analyzed by GO terms as determined by the software tool ClueGO. The functionally grouped genes and significantly enriched terms are shown in Figure 23 and Figure 24 for the upregulated genes and downregulated genes, respectively. Evaluating the functional groups, it is apparent that the genetically modified *E. coli* may suffer under nutrient deficiency and stress and regulates several genes groups trying to overcome this deficiency.

The nonribosomal peptide biosynthetic process is possibly upregulated in response to the fhuA knockout as well as being part of the ABC transporters. In addition to being an entry for phages, the fhuA receptor (also known as TonA) is also involved in the import of nutrients such as iron by the siderophore ferrichrome [116]. Since this siderophore can no longer be imported by fhuA, the *E. coli* switches to the alternative iron acquisition system based on the siderophore enterobactin and upregulates its synthesis. Concomitantly, the esterase *fes* is also upregulated to allow the degradation of iron-loaded siderophore enterobactin upon reimport. In addition, the siderophore hydroxamate seems to be facilitated as implied by the upregulation of ABC transporters fhuBCD. The receptor FepA [158] is responsible for the import of enterobactin and it is therefore likely that the *E. coli* can easily overcome the impaired iron import system based on fhuA. The supplementation of an alternative iron source such as ferric citrate, which can directly diffuse trough the porins, could support growth and stabilize the growth rate of fhuA mutants [159, 160].

Another deficiency that presents itself in the upregulation of genes is probably caused by spermidine and putrescine, which cannot be synthesized from Arginine by *speA* (coding for Arginine-Decarboxylase). A total lack of spermidine/putrescine, however, would result in a decrease of the growth rate by approximately 40% [161], which is not the case (compare to Figure 21). This indicates that putrescine and spermidine are efficiently synthesized by L-ornithine and S-adenosyl-L-methionine, which is supported by the observation that speC (ornithine decarboxylase) and speD (S-adenosylmethionine Decarboxylase), respectively, are upregulated (compare Figure 23 and Table 13).

In addition, the biodegradative arginine decarboxylase (adiA) is probably upregulated in response to speA removal [162]. This is an alternative to arginine decarboxylase [163] catalyzing the same reaction as speA. Since the L-arginine concentration is highly stable in extract derived from a speA knockout strain (Figure 25), however, the enzymatic activity has either already diminished during extract preparation, or the catalytic activity is low compared to speA.

Despite the presence of alternative pathways, the receptors potG/F and plaP, which are involved in the uptake of spermidine and putrescine, are upregulated, suggesting further demand for these important polyamines [162]. Future experiments could test if the addition of spermidine and putrescine in the growth medium helps alleviate any shortage of spermidine and putrescine to produce better S30 extracts derived from the genetically modified A19 strain.

The gshB codes for the glutathione synthase located downstream of the gshA (glutamate-cysteine ligase). By the upregulation, the cell tries to counteract the lack of GSH and GSSH, which is of course impossible since the substrate (L-γ-glutamyl-cysteine) for GshB is lacking. Conspicuously, the cysteine synthetic pathway is also upregulated. This may indicate that due to the lack of GSH and GSSH, the cell also responds with the synthesis of cysteine to provide the substrate of gshA, as this is most likely the limiting substrate (in contrast to glutamate).

Another interesting observation is the upregulation of genes involved in colonic acid as well as its multiple subgroups and single species biofilm formation in the genetically modified strain compared to A19 wt (Figure 23 and Table 15). The exopolysaccharide colanic acid consists of glucose, galactose, fucose and glucuronic acid, acetate and pyruvate and can be found on the outer membrane of several strains, including *E. coli* [164]. The overproduction of colonic acid was observed before and can be attributed to the Lon knockout. The positive transcriptional regular RcsA (a transcriptional regulator of capsule synthesis) has been shown to have a half-life of approximately 5 min in lon^+ cells, which was increased to 20 min in the

absence of the protease lon, suggesting that the increased half-life increased the production of capsule-related enzymes controlled by RcsA [165].

Consistent with this statement is the observation that the colony form changed after the introduction of the lon knockout, showing a more outspread colony formation, whereas before the introduction the colonies were always more localized and sharper (data not shown). Other regulated genes/groups, such as response to heat or provirus excision, may show primary or secondary effects of the genetic modification, but no apparent link could be found.

The number of downregulated genes was almost twice as high at 299 genes compared to the number of upregulated genes. These genes were also functionally integrated, and the classification is depicted in Figure 24. A list of all regulated genes can be found in Table 16. It is striking that highly enriched groups include related GO terms such as “bacterial chemotaxis,” “bacterial-type flagellum-dependent cell motility,” “biofilm formation” and “flagellar assembly,” as well as the term “quorum sensing,” which is known to regulate the expression of genes related to chemotaxis, flagellum assembly and biofilm formation [166]. There is no apparent reason based on the removed genes that would obviously result in the downregulation of genes related to these GO terms. It should be considered that flagellum assembly is already a highly regulated and finely coordinated process involving more than 50 genes. Approximately 25 genes are coding for structural components of the flagellum and the other 25 proteins are responsible for either the regulation of flagellum assembly or the detection of environmental factors to regulate locomotility [167-169].

Structural genes and regulatory genes as well as genes involved in the detection of environmental factors (quorum sensing) are downregulated in this strain. Two of the main regulators of the flagellum assembly are the factors flhCD and the alternative σ^{28} factor (aka fliA). It has been suggested that flhDC is regulated on mRNA- as well as on the protein-level, possibly by the protease *lon* in *E. coli* [170]. Additionally, it has been suggested that the sigma factor fliA and its counterpart fliM are regulated by *lon* [171]. Since these factors are positive regulators, however, an upregulation would be expected in the absence of *lon*. There may be repressing factors that cannot be degraded in the absence of *lon*, however, and the quorum sensing/motility apparatus in *E. coli* may therefore be downregulated. This could be determined in additional experiments. Alternatively, the downregulation may simply be a secondary effect caused by the decreased growth rate. In fact, a recent study has shown a correlation between flagellum expression/cell motility and the growth rate in *E. coli* strain RP437 controlled by flhDC [172]. Similarly, the lower growth rate may be responsible for the downregulation of other genes involved in the transport and biosynthetic processes of nutrients, as well as the

downregulation of stress responses (ethanol and osmotic stress). However, to determine this in detail, further and detailed experiments are necessary.

5.3 Expression of Fab/scFab with optimized S30 extracts

The antibody scaffold scFab entails the potential to combine the easy producibility of scFvs in *E. coli* with the high stability of the Fab-scaffold. Previous studies, however, exposed several problems of scFab's expression *in vivo* [173]. When first describing the scFab scaffold, the authors evaluated various constructs including scFabΔC in which the cysteines forming the disulfide bond between HC and LC were removed [83]. Since scFabΔC expression was accompanied with high levels of aggregation, however, the cysteines were retained in another study, and instead various linker lengths, as well as the translocation systems SEC and SRP, were evaluated in *E. coli.* Their results showed that longer linker lengths and the use of the cotranslational secretory pathway SRP improved display levels, indicating that scFab folds cotranslationally in the cytoplasm prior to secretion to the periplasm [173]. Interestingly, the linker length had no impact on the expression of scIgG in mammalian expression systems, suggesting that their secretory pathways are better suited for the folding of scIgG [174].

In contrast to previous studies, the new scFab-scaffold was expressed for the first time in a CFPS system based on the ribosomal extract of *E. coli*, which opens up possibility of investigating scFab expression in the absence of any translocation limitation, while the reaction conditions can be adjusted and supplemented to mimic the periplasmic environment that allows efficient folding of disulfide-bonded proteins in *E. coli*.

As discussed previously, the genetically modified *E. coli* A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA/ompT/lon/gor/fhuA) was evaluated for this rather difficult-to-express protein. Especially advantageous when expressing Fabs and scFabs should be the absence of *gor* to stabilize oxidized GSH in the extract and create a stable redox environment, therefore allowing native disulfide bond formation [175]. In addition, the *ompT* and *lon* knockouts are expected to stabilize folding intermediates by avoiding proteolytic degradation during folding. The simultaneous addition of chaperones aids folding and avoids the formation of large amounts of insoluble aggregates and useless product in the absence of these two major proteases. A similar genetic background with the exception of *ompT* and *lon* is routinely used for the expression of disulfide bonded proteins, even in the commercial setting at Sutro Biopharma (San Francisco, CA, USA) [33, 176, 177].

When characterizing the A19-derived S30 extracts by LC-MS (compare 4.1.5), significant amounts of ompT and lon protease were found in the extracts, as expected. The Lon-

protease is known to degrade many foreign proteins *in vivo* [124] and the outer membrane endoprotease OmpT has been shown to degrade T7RNAP and other foreign proteins upon cell disruption as performed during S30 extract preparation [122]. These proteases are also absent in the popular expression host and source strain for ribosomal extract based on BL21(DE3). As shown in the previous chapter (compare to 5.2), the performance of S30 extract derived from the genetically modified strain is highly target protein specific. Initial experiments showed that the expression of our constructs for scFab and Fab were highly enhanced in extracts derived from the genetically modified strain (data not shown).

The expression of Fab and scFab in ribosomal extracts derived from genetically modified *E. coli* A19 rne::SBP Δ(gshA/endA/sdaA/speA/sdaB/tnaA/ompT/lon/gor/fhuA) was evaluated in the presence of various concentrations of the chaperones DsbC, FkpA and Skp, which represent the most important chaperones of the periplasm in *E. coli* and have been shown to improve yields of antibodies and antibody fragments when expressed in *E. coli* [33]. These chaperones were chosen to aid the functional folding of antibodies and antibody fragments in our CFPS system. For the optimization of the concentration of glutathione (reduced and oxidized), the disulfide isomerase DsbC was used at 13 µM. The results indicated that 2 mM total glutathione at a ratio of 1:1 of oxidized:reduced glutathione resulted in the highest yield, but higher ratios of oxidized glutathione also supported high yields of >150 µg/mL, which is in agreement with other studies [33, 176].

It is noteworthy that we determined only the soluble yield after centrifugation without further analysis to determine amount of assembled Fab, HC/LC and soluble aggregates. It was assumed, however, that 300% increase in soluble yield is caused by natively folded HC or LC and assembled Fab, which is in turn less susceptible to degradation and aggregation and therefore remains in the soluble fraction. This notion is further supported by DsbC-titration experiments in which assembled Fab was formed under comparable conditions (Figure 30).

Next, DsbC, FkpA and Skp were titrated while expressing Fab and scFab (compare Figure 30, Figure 31 and Figure 32). A 3.5-fold change of correctly assembled Fab upon use of the optimal concentration of DsbC (73 µM) was observed in the experiment at hand. Because the concentrations of DsbC were increased, it was expected that positive effects of DsbC would level off at some concentration. The data at hand, however, suggest that higher concentrations of DsbC might even further enhance assembling of Fab. Notably, only DsbC improved the ratio of non-assembled HC/LC to assembled Fab, providing evidence for the importance of disulfide isomerization and proofreading for the expression of antibodies and antibody fragments in *E. coli* (Figure 32) and DsbC.

To our surprise, FkpA and Skp had little effect on the amount of assembled Fab (up to 1.5-fold increase) when compared to DsbC. This minor effect was even reversed with higher concentrations of FkpA and Skp. With FkpA primarily reported as peptidyl-prolyl cis/trans isomerase [178-180] and Skp as a holdase [181, 182], we expected distinct effects on the assembling of Fab; however, the patterns of fold-change observed for assembled Fab (Figure 31) and the folded/unfolded ratio (Figure 32) were similar. The data showed a strong shift towards unfolded with increasing concentration of both FkpA and Skp, whereas the shift towards unfolded was more pronounced for FkpA, suggesting that FkpA and Skp perform similar functions *in vitro*.

The large amount of non-assembled HC/LC observed in the presence of FkpA and Skp suggests these chaperones bind LC and/or HC tightly, and therefore impose an impediment to the assembly of Fab. In addition, the combination of chaperones DsbC/Skp, DsbC/FkpA and DsbC/SkpA/Skp resulted in a large amount of non-assembled HC/LC, showing that the addition of DsbC to either Skp or FkpA cannot reverse this effect. This implies that HC and/or LC are indeed not accessible for assembly in the presence of Skp and FkpA. Skp has been crystallized (PDB code: 1U2M) as a trimer in a jellyfish-like form, whereas the target protein is bound in a cavity and surrounded by tentacles [183].

While this arrangement is useful to protect folding intermediates from aggregating, it could be a steric hindrance preventing the correct assembly of Fab. The chaperone FkpA, being primarily a peptidyl-prolyl isomerase, was suggested to perform additional chaperone-like activity that could bind hydrophobic or proline containing patches and therefore prohibit the assembly of Fab [179]. Since FkpA did not exhibit any additional benefits compared to DsbC- or Skp-titration experiment, it was assumed that the peptidyl-prolyl isomeric activity plays a minor role during the expression of Fab in the cell-free system.

Since our Fab contains two proline in the trans position, however, it shall be assumed that other prolyl-isomerases present in the S30 extract can completely substitute FkpA. It is also noteworthy that the overall yield (soluble and insoluble) increased in the presence of FkpA and Skp, possibly since high molecular weight aggregates were kept in solution rather than being degraded. This could be caused either by a protective effect of these chaperones against proteolytic degradation of synthesized protein, or the fact that protein synthesis is enhanced in the presence of these chaperones, for example, due to direct interaction of these chaperones with the ribosome or the emerging polypeptide chain.

In conclusion, DsbC is most important for assembly of Fab *in vitro*. While FkpA and Skp increased the soluble yield significantly, the solubilization mechanism seems to impose a hindrance to proper assembly of HC and LC.

Upon titration of DsbC, the detectable amount of scFab increased similarly to Fab, approximately 2-fold, whereas the ratio of folded/unfolded remained constantly negative. In contrast to titration during Fab-expression, the titration of FkpA and Skp during the expression of scFab increased the amount of folded scFab up to 12-fold. Interestingly, the ratio of folded/unfolded shifted more towards folded with increasing concentrations of FkpA and Skp as opposed to Fab expression.

It is important to point out that in the case of scFab expression, we observed two bands under nonreducing conditions. In the case of native disulfide bond formation in scFab, a similar migration pattern to Fab under nonreducing, denaturing conditions was expected. Therefore, we assumed that the 48 kDa band represents scFab with native disulfide bonds. Since we observed only one band (data not shown) under reducing conditions, we presumed that the 43 kDa band represents more likely nonnative disulfide bond formation, resulting in a more compact structure rather than a degradation product.

Interestingly, Koerber and colleagues observed in a thermal denaturation experiment two different species of scFab when using the short linker of 36 aa (T_m= 81 °C (scFab); T_m=68 °C (a less stable species) with similar molecular weights, according to SEC (Size Exclusion Chromatography). It is tempting to speculate that the two species observed on the autoradiogram represent those two species observed in thermal denaturation experiments by Koerber *et al.* [184].

In contrast to Fab, the chaperones FkpA and Skp had a high impact on the fold-change of scFab, suggesting that FkpA and Skp can directly bind the emerging polypeptide chain and protect it from aggregation and other interactions. Since HC and LC are already joined in scFab, no further interactions are required for correct assembly; therefore, FkpA and Skp show positive effects on expression of scFab. This notion is further supported by the fact that these chaperones are mostly reported to be beneficial in the context of scFvs and other single chain membrane protein [182, 185, 186].

Despite those positive effects on folding of scFab, the two species present imposed a problem for further analysis. The two scFab species were copurified, along with large amounts of His-tagged chaperones, which were apparently still associated with one or both of the scFab species. The association of scFab with chaperones, which was not observed in the course of purification of Fab, suggests that scFab or at least a fraction of scFab cannot overcome intermediary folding states and therefore remains associated with chaperones due to the exposure of, *e.g.*, hydrophobic patches of the protein. Alternatively, proteases may be needed to release finished proteins. Nevertheless, the presented approach allowed the expression of

scFab in the absence of translocation systems, which have been suggested to limit expression of scFab *in vivo*.

6 Conclusion

The initial objective of this thesis was to characterize commonly used S30 extracts for cell-free proteins synthesis. This study has revealed that approximately 40% of the proteome of exponentially growing *E. coli* remains in the final S30 lysate after standard preparation procedures. It should be considered that the growth phase of the cells at harvest as well as the applied centrifugation force during lysate preparation could have a much higher impact on the lysate proteome composition. In addition to the most common S30 (centrifugation at 30,000 *g*) extract, cell-free expression of proteins based on S12 or S15 lysates has been reported [68]. The S30 extract represents the best compromise between complexity and efficiency, however, and was thus selected for analysis.

The presented data help to move S30 lysates towards a better-defined system suitable for directed lysate optimization and fine-tuned tailored protein production. In general, it should be considered that the presence of a protein in the lysate does not automatically implicate its correct folding and activity. The presented lysate proteome can serve as an initial checklist to identify or to rule out problems that could possibly occur in the production and analysis of a given target protein. In previous work, the removal of release factor-1 from S30 lysates considerably increased the efficiency of nonnatural amino acid incorporation into proteins by the amber stop codon technology [26, 110]. Alternatively, lysates from engineered strains containing selected deficiencies have been prepared [27, 187, 188]. Identified proteins causing problems for the structural analysis of labeled proteins could be further addressed by adding selective inhibitors as exemplified for amino acid scrambling enzymes [24, 189].

A potential tool for the elimination of nondesired proteins could be modification with degradation tags such as the ssrA tag [190]. Elimination will then be promoted by the presence of specific proteases such as ClpX and ClpP in the S30 lysate. The modular portfolio of cell-free synthetic biology could further be expanded by complementing selected partially present biosynthetic pathways, *e.g.*, for the synthesis of stable isotope labeled amino acids from cheap precursors as an alternative to the addition of expensive synthetic labeled amino acids. Problematic proteins could be evaluated and after positive verification, selectively removed from lysates prior to expression studies. Conversely, potentially limiting components could be supplied and enriched for optimal concentrations as partially applied in the following applications.

After the proteome profiling was concluded, several approaches were undertaken to improve the efficiency of cell-free protein synthesis that can be categorized as follows: (*i*) system approach by applying a heatshock during growth before S30 extract preparation to change the proteome composition beneficially; (*ii*) targeted approach by genetically modifying the *E. coli* A19 strain to improve target protein- and precursor stability; and (*iii*) purification of chaperones and their targeted supplementation to cell-free protein synthesis reaction to improve antibody folding.

The ultimate objective was to improve cell-free protein synthesis in terms of yield while avoiding incorrectly folded protein/aggregates and improving natively folded protein yield instead. By the systems approach (*i*), the induction of a heatshock/ethanol stress during cultivation after the exponential growth phase was analyzed using a quantitative MS-based method based on the labeling agent ICPL. While the expression efficiency was generally lower with standard proteins such as GFP, two proteins, namely, apiRBP and hGNA-1, were identified whose expression greatly benefited from the altered proteome composition of S30-S extracts. The protein quality (as measured by enzymatic activity and soluble yield) was substantially improved, showing that S30-S lysate can be beneficial for the expression of soluble chaperone-dependent proteins and the lower protein synthesis capacity of S30-S extracts can be advantageous for improving overall protein quality.

Next, several enzymes and proteases identified in the S30 extract proteome were targeted (approach ii) and either knocked out or modified by a tag. To our regret, the genetic modifications were not universally beneficial but highly protein specific. While the amino acid stabilization, which was also proven independently by analyzing the time course of the amino acid concentration in S30 extract, showed improvement in yield in experiments with CAT, the stabilization also showed rather negative effects during expression of EF-Ts and AhR as performed by Biotechrabbit. In contrast, the knockout of the proteases ompT and lon showed improvements in AhR only, but no effect of EF-Ts and even a negative effect on expression of CAT. These experiments show that the effects of genetic modifications are highly target protein specific and must be evaluated for every individual protein. Even the stabilization of amino acids, which was expected to have a rather universal effect on protein synthesis in cell free system, was also shown to be highly target protein specific.

For approach (iii), the highly genetically modified derivative of A19 was used, which included the modifications to stabilize amino acids and knockout of major proteases ompT, lon and gor (glutathione reductase) to stabilize oxidized glutathione for efficient disulfide formation. The disulfide isomerase DsbC was chosen as a chaperone, a peptidyl-prolyl-isomerase including chaperone activity, namely, FkpA, and the chaperone Skp. For the

assembly of two chain Fab, the DsbC enzyme proved to be the most important chaperone, while the other chaperones seem to prevent the assembly of the heavy chain fragment and light chain. The scFab, on the other hand benefitted from the chaperone activity of FkpA and Skp, while DsbC did not substantially improve scFab yield. Both antibodies were at least partially natively folded and bound to their antigen MAPKAP1. In conclusion, this experiment also showed that universal improvement of cell-free expression system by addition of chaperones is not possible and the effects must be evaluated target protein-specific; however, the open nature of cell-free systems makes the high throughput optimization on a target protein basis possible. The positive effect may then be applied to a target strain to generate an optimized expression strain for specific target proteins.

7 References

1. Gale, E.F. and J.P. Folkes, *Effect of nucleic acids on protein synthesis and amino-acid incorporation in disrupted staphylococcal cells.* Nature, 1954. **173**(4417): p. 1223-7.
2. Keller, E.B. and P.C. Zamecnik, *The effect of guanosine diphosphate and triphosphate on the incorporation of labeled amino acids into proteins.* J Biol Chem, 1956. **221**(1): p. 45-59.
3. Lamborg, M.R. and P.C. Zamecnik, *Amino acid incorporation into protein by extracts of E. coli.* Biochim Biophys Acta, 1960. **42**: p. 206-11.
4. McQuillen, K., R.B. Roberts, and R.J. Britten, *SYNTHESIS OF NASCENT PROTEIN BY RIBOSOMES IN ESCHERICHIA COLI.* Proc Natl Acad Sci U S A, 1959. **45**(9): p. 1437-47.
5. Tissieres, A., D. Schlessinger, and F. Gros, *AMINO ACID INCORPORATION INTO PROTEINS BY ESCHERICHIA COLI RIBOSOMES.* Proc Natl Acad Sci U S A, 1960. **46**(11): p. 1450-63.
6. Nirenberg, M.W.W. and J.H.H. Matthaei, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides.* Proceedings of the National Academy of Sciences of the United States of America, 1961. **47**(10): p. 1588-602.
7. Nirenberg, M., *Historical review: Deciphering the genetic code--a personal account.* Trends Biochem Sci, 2004. **29**(1): p. 46-54.
8. Chen, H.Z. and G. Zubay, *Prokaryotic coupled transcription-translation.* Methods Enzymol, 1983. **101**: p. 674-90.
9. Zubay, G., *In vitro synthesis of protein in microbial systems.* Annu Rev Genet, 1973. **7**: p. 267-87.
10. Ryabova, L.A., I. Morozov, and A.S. Spirin, *Continuous-flow cell-free translation, transcription-translation, and replication-translation systems.* Methods Mol Biol, 1998. **77**: p. 179-93.
11. Spirin, A.S., et al., *A continuous cell-free translation system capable of producing polypeptides in high yield.* Science (New York, N.Y.), 1988. **242**(4882): p. 1162-4.
12. Shirokov, V.A., et al., *Continuous-exchange protein-synthesizing systems.* Methods Mol Biol, 2007. **375**: p. 19-55.
13. Chong, S., *Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications.* Curr Protoc Mol Biol, 2014. **108**: p. 16 30 1-16 30 11.
14. Busso, D., R. Kim, and S.H. Kim, *Using an Escherichia coli cell-free extract to screen for soluble expression of recombinant proteins.* J Struct Funct Genomics, 2004. **5**(1-2): p. 69-74.
15. Katzen, F., G. Chang, and W. Kudlicki, *The past, present and future of cell-free protein synthesis.* Trends Biotechnol, 2005. **23**(3): p. 150-6.
16. Rosenblum, G. and B.S. Cooperman, *Engine out of the chassis: cell-free protein synthesis and its uses.* FEBS Lett, 2014. **588**(2): p. 261-8.
17. Henrich, E., et al., *Membrane protein production in Escherichia coli cell-free lysates.* FEBS Lett, 2015. **589**(15): p. 1713-22.
18. Salehi, A.S., et al., *Escherichia coli-based Cell-free Extract Development for Protein-based Cancer Therapeutic Production.* Int J Dev Biol, 2016.
19. Matsuda, T., et al., *Cell-free synthesis of zinc-binding proteins.* J Struct Funct Genomics, 2006. **7**(2): p. 93-100.
20. Hein, C., et al., *Hydrophobic supplements in cell-free systems: Designing artificial environments for membrane proteins.* Engineering in Life Sciences, 2014.

21. Henrich, E., et al., *Lipid Requirements for the Enzymatic Activity of MraY Translocases and in Vitro Reconstitution of the Lipid II Synthesis Pathway.* J Biol Chem, 2016. **291**(5): p. 2535-46.
22. Rues, R.B., V. Dötsch, and F. Bernhard, *Co-translational formation and pharmacological characterization of beta1-adrenergic receptor/nanodisc complexes with different lipid environments.* Biochim Biophys Acta, 2016. **1858**(6): p. 1306-16.
23. Casteleijn, M.G., A. Urtti, and S. Sarkhel, *Expression without boundaries: cell-free protein synthesis in pharmaceutical research.* Int J Pharm, 2013. **440**(1): p. 39-47.
24. LaGuerre, A., et al., *Labeling of Membrane Proteins by Cell-Free Expression.* Methods Enzymol, 2015. **565**: p. 367-88.
25. Löhr, F., et al., *An extended combinatorial 15N, 13Cα, and 13C' labeling approach to protein backbone resonance assignment.* J Biomol NMR, 2015. **62**(3): p. 263-79.
26. Peuker, S., et al., *Efficient Isotope Editing of Proteins for Site-Directed Vibrational Spectroscopy.* J Am Chem Soc, 2016. **138**(7): p. 2312-8.
27. Hong, S.H., et al., *Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation.* ACS synthetic biology, 2014. **3**(6): p. 398-409.
28. Proverbio, D., et al., *Membrane Protein Quality Control in Cell-Free Expression Systems: Tools, Strategies and Case Studies.* 2014: p. 45-70.
29. Gesteland, R.F., *Isolation and characterization of ribonuclease I mutants of Escherichia coli.* Journal of Molecular Biology, 1966. **16**(1): p. 67-84.
30. Yang, H.L., et al., *Cell-free coupled transcription-translation system for investigation of linear DNA segments.* Proc Natl Acad Sci U S A, 1980. **77**(12): p. 7029-33.
31. Rues, R.B., et al., *Cell-Free Production of Membrane Proteins in Escherichia coli Lysates for Functional and Structural Studies.* Methods Mol Biol, 2016. **1432**: p. 1-21.
32. Liu, D.V., J.F. Zawada, and J.R. Swartz, *Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis.* Biotechnology progress, 2005. **21**(2): p. 460-5.
33. Groff, D., et al., *Engineering toward a bacterial "endoplasmic reticulum" for the rapid expression of immunoglobulin proteins.* mAbs, 2014. **6**(3): p. 671-678.
34. Failmezger, J., et al., *Cell-Free Protein Synthesis From Fast-Growing Vibrio natriegens.* Front Microbiol, 2018. **9**: p. 1146.
35. Wiegand, D.J., et al., *Establishing a Cell-Free Vibrio natriegens Expression System.* ACS Synth Biol, 2018. **7**(10): p. 2475-2479.
36. Des Soye, B.J., et al., *Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens.* ACS Synth Biol, 2018. **7**(9): p. 2245-2255.
37. Ahn, J.H., C.Y. Choi, and D.M. Kim, *Effect of energy source on the efficiency of translational termination during cell-free protein synthesis.* Biochem Biophys Res Commun, 2005. **337**(1): p. 325-9.
38. Krinsky, N., et al., *A Simple and Rapid Method for Preparing a Cell-Free Bacterial Lysate for Protein Synthesis.* PLoS One, 2016. **11**(10): p. e0165137.
39. Kwon, Y.-C. and M.C. Jewett, *High-throughput preparation methods of crude extract for robust cell-free protein synthesis.* Scientific Reports, 2015. **5**: p. 8663-8663.
40. Lill, R., et al., *Specific recognition of the 3'-terminal adenosine of tRNAPhe in the exit site of Escherichia coli ribosomes.* J Mol Biol, 1988. **203**(3): p. 699-705.
41. Grant, S.G., et al., *Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4645-9.
42. Lin, J.J., *Endonuclease A degrades chromosomal and plasmid DNA of Escherichia coli present in most preparations of single stranded DNA from phagemids.* Proc Natl Sci Counc Repub China B, 1992. **16**(1): p. 1-5.

43. Carmel, G. and J.W. Coulton, *Internal deletions in the FhuA receptor of Escherichia coli K-12 define domains of ligand interactions.* J Bacteriol, 1991. **173**(14): p. 4394-403.
44. Calhoun, K.a. and J.R. Swartz, *Total amino acid stabilization during cell-free protein synthesis reactions.* Journal of biotechnology, 2006. **123**(2): p. 193-203.
45. Michel-Reydellet, N., K. Calhoun, and J. Swartz, *Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome.* Metabolic engineering, 2004. **6**(3): p. 197-203.
46. Nieß, A., M. Siemann-Herzberg, and R. Takors, *Protein production in Escherichia coli is guided by the trade-off between intracellular substrate availability and energy cost.* Microbial Cell Factories, 2019. **18**(1): p. 1-10.
47. Derman, A.I., et al., *Mutations that allow disulfide bond formation in the cytoplasm of Escherichia coli.* Science, 1993. **262**(5140): p. 1744-7.
48. Rosano, G.L. and E.a. Ceccarelli, *Recombinant protein expression in Escherichia coli: Advances and challenges.* Frontiers in Microbiology, 2014. **5**(April): p. 1-17.
49. Skerra, A. and A. Pluckthun, *Assembly of a functional immunoglobulin Fv fragment in Escherichia coli.* Science, 1988. **240**(4855): p. 1038-41.
50. Knapp, K.G., A.R. Goerke, and J.R. Swartz, *Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source.* Biotechnology and bioengineering, 2007. **97**(4): p. 901-8.
51. Cai, Q., et al., *A simplified and robust protocol for immunoglobulin expression in Escherichia coli cell-free protein synthesis systems.* Biotechnol Prog, 2015. **31**(3): p. 823-31.
52. Groff, D., et al., *Engineering toward a bacterial "endoplasmic reticulum" for the rapid expression of immunoglobulin proteins*, in *MAbs*. 2014. p. 671-8.
53. Zawada, J.F., S. Biopharma, and S. Francisco, *Cell-Free Production of Pharmaceutical Proteins*. Second Edi ed. Vol. 1. 2011: Elsevier B.V. 391-398.
54. Kline, T., et al., *Methods to Make Homogenous Antibody Drug Conjugates.* Pharm Res, 2015. **32**(11): p. 3480-93.
55. Swartz, J., *Developing cell-free biology for industrial applications.* Journal of industrial microbiology & biotechnology, 2006. **33**(7): p. 476-85.
56. Kim, D.M. and J.R. Swartz, *Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis.* Biotechnology and Bioengineering, 2001. **74**(4): p. 309-316.
57. Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnol Bioeng, 2004. **86**(1): p. 19-26.
58. Jewett, M.C., et al., *An integrated cell-free metabolic platform for protein production and synthetic biology.* Molecular systems biology, 2008. **4**: p. 220-220.
59. Failmezger, J., et al., *Site-Specific Cleavage of Ribosomal RNA in Escherichia coli-Based Cell-Free Protein Synthesis Systems*, in *PLoS One*. 2016.
60. Caschera, F. and V. Noireaux, *A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system.* Metabolic Engineering, 2015. **27**: p. 29-37.
61. Calhoun, K.a. and J.R. Swartz, *Energizing cell-free protein synthesis with glucose metabolism.* Biotechnology and bioengineering, 2005. **90**(5): p. 606-13.
62. Kim, H.-C., T.-W. Kim, and D.-M. Kim, *Prolonged production of proteins in a cell-free protein synthesis system using polymeric carbohydrates as an energy source.* Process Biochemistry, 2011. **46**(6): p. 1366-1369.
63. Voloshin, A.M. and J.R. Swartz, *Efficient and scalable method for scaling up cell free protein synthesis in batch mode.* Biotechnology and bioengineering, 2005. **91**(4): p. 516-21.

64. Freischmidt, A., *Untersuchungen an einem zellfreien Proteinsynthesesystem basierend auf S30 Extrakten von Escherichia coli*. 2012.
65. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components*, in *Nat Biotechnol*. 2001: United States. p. 751-5.
66. Berrier, C., et al., *Coupled cell-free synthesis and lipid vesicle insertion of a functional oligomeric channel MscL MscL does not need the insertase YidC for insertion in vitro*, in *Biochim Biophys Acta*. 2011, 2010 Elsevier B.V: Netherlands. p. 41-6.
67. Kigawa, T., et al., *Preparation of Escherichia coli cell extract for highly productive cell-free protein expression*, in *J Struct Funct Genomics*. 2004: Netherlands. p. 63-8.
68. Kwon, Y.C. and M.C. Jewett, *High-throughput preparation methods of crude extract for robust cell-free protein synthesis.* Sci Rep, 2015. **5**: p. 8663.
69. Levine, M.Z., et al., *Redesigned upstream processing enables a 24-hour workflow from E. coli cells to cell-free protein synthesis.* 2019.
70. Gregorio, N.E., M.Z. Levine, and J.P. Oza, *A User's Guide to Cell-Free Protein Synthesis*, in *Methods Protoc*. 2019.
71. Au - Levine, M.Z., et al., *Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology.*
72. Soufi, B., et al., *Characterization of the E. coli proteome and its modifications during growth and ethanol stress.* Front Microbiol, 2015. **6**: p. 103.
73. Schmidt, A., et al., *The quantitative and condition-dependent Escherichia coli proteome.* Nat Biotechnol, 2016. **34**(1): p. 104-10.
74. Failmezger, J., et al., *Cell-free protein synthesis from non-growing, stressed Escherichia coli.* Scientific Reports, 2017. **7**(1): p. 1-10.
75. Arsène, F., T. Tomoyasu, and B. Bukau, *The heat shock response of Escherichia coli.* Int J Food Microbiol, 2000. **55**(1-3): p. 3-9.
76. Yokota, T., et al., *Rapid Tumor Penetration of a Single-Chain Fv and Comparison with Other Immunoglobulin Forms Rapid Tumor Penetration of a Single-Chain Fv and Comparison with Other Immunoglobulin Forms.* 1992: p. 3402-3408.
77. Version, U. and C. Articles, *Physiological Barriers to Delivery of Monoclonal Antibodies and Other Macromolecules in Tumors.* 1990.
78. Nelson, A.L., *Antibody fragments: hope and hype.* mAbs, 2010. **2**(February): p. 77-83.
79. Hammers, C.M. and J.R. Stanley, *Antibody Phage Display: Technique and Applications.* Journal of Investigative Dermatology, 2014. **134**(2): p. e17-e17.
80. Zawada, J.F., *Ribosome Display and Related Technologies.* 2012. **805**: p. 31-42.
81. He, M. and F. Khan, *Ribosome display: next-generation display technologies for production of antibodies in vitro.* Expert review of proteomics, 2005. **2**(3): p. 421-430.
82. Ferrer-Miralles, N., et al., *Microbial factories for recombinant pharmaceuticals.* Microbial cell factories, 2009. **8**: p. 17-17.
83. Hust, M., et al., *Single chain Fab (scFab) fragment.* BMC biotechnology, 2007. **7**: p. 14-14.
84. Koerber, J.T., M.J. Hornsby, and J.A. Wells, *An improved single-chain Fab platform for efficient display and recombinant expression.* Journal of molecular biology, 2015. **427**(2): p. 576-86.
85. Gesteland, R.F., *Isolation and characterization of ribonuclease I mutants of Escherichia coli.* J Mol Biol, 1966. **16**(1): p. 67-84.
86. Welch, M., et al., *Design Parameters to Control Synthetic Gene Expression in <italic>Escherichia coli</italic>.* PLoS ONE, 2009. **4**(9): p. e7002.
87. Erijman, A., et al., *Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering.* Journal of Structural Biology, 2011. **175**(2): p. 171-177.
88. Kai, L., et al., *Chemical Genomics and Proteomics.* 2012. **800**: p. 201-225.
89. Walker, J.M., *The proteomics protocols handbook*. 2005, Totowa, N.J.: Humana Press. xviii, 988 p.

90. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences, 2000. **97**(12): p. 6640-6645.
91. Schwarz, D., et al., *Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems.* Nat Protoc, 2007. **2**(11): p. 2945-57.
92. Henrich, E., V. Dotsch, and F. Bernhard, *Screening for lipid requirements of membrane proteins by combining cell-free expression with nanodiscs.* Methods Enzymol, 2015. **556**: p. 351-69.
93. Schneider, B., et al., *Membrane protein expression in cell-free systems.* Methods Mol Biol, 2010. **601**: p. 165-86.
94. Cai, Q., et al., *A simplified and robust protocol for immunoglobulin expression in Escherichia coli cell-free protein synthesis systems.* Biotechnology Progress, 2015: p. n/a-n/a.
95. Roos, C., et al., *Characterization of co-translationally formed nanodisc complexes with small multidrug transporters, proteorhodopsin and with the E. coli MraY translocase.* Biochim Biophys Acta, 2012. **1818**(12): p. 3098-106.
96. Schwarz, D., et al., *Cell-free expression profiling of E. coli inner membrane proteins.* Proteomics, 2010. **10**(9): p. 1762-79.
97. Kai, L., et al., *Artificial environments for the co-translational stabilization of cell-free expressed proteins.* PLoS One, 2013. **8**(2): p. e56637.
98. Ma, Y., et al., *Cell-free expression of human glucosamine 6-phosphate N-acetyltransferase (HsGNA1) for inhibitor screening.* Protein Expr Purif, 2012. **86**(2): p. 120-6.
99. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.* Anal Chem, 1996. **68**(5): p. 850-8.
100. Chen, D., et al., *Online quantitative proteomics p-value calculator for permutation-based statistical testing of peptide ratios.* J Proteome Res, 2014. **13**(9): p. 4184-91.
101. Nguyen, H.D., I.A. Wood, and M.M. Hill, *A robust permutation test for quantitative SILAC proteomics experiments.* Journal of Integrated OMICS, 2012. **2**(2): p. 80-93.
102. Klammt, C., et al., *Cell-free production of integral membrane proteins on a preparative scale.* Methods Mol Biol, 2007. **375**: p. 57-78.
103. Yang, W.C., et al., *Simplifying and streamlining Escherichia coli-based cell-free protein synthesis.* Biotechnol Prog, 2012. **28**(2): p. 413-20.
104. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-62.
105. Besemer, J., A. Lomsadze, and M. Borodovsky, *GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions.* Nucleic Acids Res, 2001. **29**(12): p. 2607-18.
106. Ishihama, Y., et al., *Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.* Mol Cell Proteomics, 2005. **4**(9): p. 1265-72.
107. Binns, D., et al., *QuickGO: a web-based tool for Gene Ontology searching.* Bioinformatics, 2009. **25**(22): p. 3045-6.
108. Matthies, D., et al., *Cell-free expression and assembly of ATP synthase.* J Mol Biol, 2011. **413**(3): p. 593-603.
109. Shin, J., P. Jardine, and V. Noireaux, *Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction.* ACS Synth Biol, 2012. **1**(9): p. 408-13.
110. Loscha, K.V., et al., *Multiple-site labeling of proteins with unnatural amino acids.* Angew Chem Int Ed Engl, 2012. **51**(9): p. 2243-6.

111. Gottesman, S., et al., *The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system.* Genes Dev, 1998. **12**(9): p. 1338-47.
112. Chen, D., et al., *Online Quantitative Proteomics p-value Calculator for Permutation-Based Statistical Testing of Peptide Ratios.* Journal of Proteome Research, 2014: p. 140724045651009-140724045651009.
113. Selitrennikoff, C.P. and D.R. Sonneborn, *The last two pathway-specific enzyme activities of hexosamine biosynthesis are present in Blastocladiella emersonii zoospores prior to germination.* Biochim Biophys Acta, 1976. **451**(2): p. 408-16.
114. Taylor, R.G., D.C. Walker, and R.R. McInnes, *E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing.* Nucleic Acids Res, 1993. **21**(7): p. 1677-8.
115. Clarke, J.E., et al., *Direct entry by RNase E is a major pathway for the degradation and processing of RNA in Escherichia coli.* Nucleic Acids Res, 2014. **42**(18): p. 11733-51.
116. Braun, V., *FhuA (TonA), the career of a protein.* J Bacteriol, 2009. **191**(11): p. 3431-6.
117. Albayrak, C. and J.R. Swartz, *Direct polymerization of proteins.* ACS Synth Biol, 2014. **3**(6): p. 353-62.
118. Martin, R.W., et al., *Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids.* Nat Commun, 2018. **9**(1): p. 1203.
119. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.
120. Oh, I.-S., et al., *Cell-free production of functional antibody fragments.* Bioprocess and biosystems engineering, 2010. **33**(1): p. 127-32.
121. Riesenberg, D., et al., *High cell density fermentation of recombinant Escherichia coli expressing human interferon alpha 1.* Appl Microbiol Biotechnol, 1990. **34**(1): p. 77-82.
122. Grodberg, J. and J.J. Dunn, *ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification.* Journal of bacteriology, 1988. **170**(3): p. 1245-53.
123. Stumpe, S., et al., *Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of Escherichia coli.* Journal of Bacteriology, 1998. **180**(15): p. 4002-4006.
124. Gottesman, S., *PROTEASES AND THEIR TARGETS IN ESCHERICHIA COLI 1.* Annual Review of Genetics, 1996. **30**(1): p. 465-506.
125. Bonhivers, M., et al., *FhuA, a transporter of the Escherichia coli outer membrane, is converted into a channel upon binding of bacteriophage T5.* The EMBO journal, 1996. **15**(8): p. 1850-1856.
126. Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnology and bioengineering, 2004. **86**(1): p. 19-26.
127. Iwasaki, M., et al., *One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the Escherichia coli proteome on a microarray scale.* Anal Chem, 2010. **82**(7): p. 2616-20.
128. Ishihama, Y., et al., *Protein abundance profiling of the Escherichia coli cytosol.* BMC Genomics, 2008. **9**: p. 102.
129. Boland, C., et al., *Cell-free expression and in meso crystallisation of an integral membrane kinase for structure determination.* Cell Mol Life Sci, 2014. **71**(24): p. 4895-4910.
130. Jewett, M.C., et al., *An integrated cell-free metabolic platform for protein production and synthetic biology*, in *Mol Syst Biol*. 2008: England. p. 220.

131. Ulmschneider, M.B., et al., *Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion.* Nat Commun, 2014. **5**: p. 4863.
132. Cymer, F., G. von Heijne, and S.H. White, *Mechanisms of integral membrane protein insertion and folding.* J Mol Biol, 2015. **427**(5): p. 999-1022.
133. Ha, S., et al., *The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.* Protein Sci, 2000. **9**(6): p. 1045-52.
134. Wissenbach, U., et al., *A third periplasmic transport system for L-arginine in Escherichia coli: molecular characterization of the artPIQMJ genes, arginine binding and transport.* Mol Microbiol, 1995. **17**(4): p. 675-86.
135. Gul, N. and B. Poolman, *Functional reconstitution and osmoregulatory properties of the ProU ABC transporter from Escherichia coli.* Mol Membr Biol, 2013. **30**(2): p. 138-48.
136. Malinverni, J.C. and T.J. Silhavy, *An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane.* Proc Natl Acad Sci U S A, 2009. **106**(19): p. 8009-14.
137. Schirle, M., M.A. Heurtier, and B. Kuster, *Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry.* Mol Cell Proteomics, 2003. **2**(12): p. 1297-305.
138. Gualerzi, C.O. and C.L. Pon, *Initiation of mRNA translation in bacteria: structural and dynamic aspects*, in *Cell Mol Life Sci*. 2015. p. 4341-67.
139. Defeu Soufo, H.J., et al., *Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 3163-8.
140. Underwood, K.A., J.R. Swartz, and J.D. Puglisi, *Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis.* Biotechnol Bioeng, 2005. **91**(4): p. 425-35.
141. Niess, A., et al., *Experimentally Validated Model Enables Debottlenecking of in Vitro Protein Synthesis and Identifies a Control Shift under in Vivo Conditions.* ACS Synth Biol, 2017. **6**(10): p. 1913-1921.
142. Shin, J. and V. Noireaux, *Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70*, in *J Biol Eng*. 2010: England. p. 8.
143. Hartl, F.U. and M. Hayer-Hartl, *Molecular chaperones in the cytosol: from nascent chain to folded protein.* Science, 2002. **295**(5561): p. 1852-8.
144. Goerke, A.R. and J.R. Swartz, *Development of cell-free protein synthesis platforms for disulfide bonded proteins.* Biotechnology and bioengineering, 2008. **99**(2): p. 351-67.
145. Knapp, K.G., A.R. Goerke, and J.R. Swartz, *Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source.* Biotechnol Bioeng, 2007. **97**(4): p. 901-8.
146. Ali, M., et al., *Improvements in the cell-free production of functional antibodies using cell extract from protease-deficient Escherichia coli mutant.* J Biosci Bioeng, 2005. **99**(2): p. 181-6.
147. Lüders, S., C. Fallet, and E. Franco-Lara, *Proteome analysis of the Escherichia coli heat shock response under steady-state conditions.* Proteome Sci, 2009. **7**: p. 36.
148. Chaudhuri, S., B. Jana, and T. Basu, *Why does ethanol induce cellular heat-shock response?* Cell Biol Toxicol, 2006. **22**(1): p. 29-37.
149. Yamane, T., et al., *Enhanced cell-free protein synthesis using a S30 extract from Escherichia coli grown rapidly at 42 degrees C in an amino acid enriched medium.* Biotechnol Prog, 2005. **21**(2): p. 608-13.
150. Wickner, S., M.R. Maurizi, and S. Gottesman, *Posttranslational quality control: folding, refolding, and degrading proteins.* Science, 1999. **286**(5446): p. 1888-93.

151. Lee, I. and C.K. Suzuki, *Functional mechanics of the ATP-dependent Lon protease-lessons from endogenous protein and synthetic peptide substrates.* Biochim Biophys Acta, 2008. **1784**(5): p. 727-35.
152. Guisbert, E., et al., *A chaperone network controls the heat shock response in E. coli.* Genes Dev, 2004. **18**(22): p. 2812-21.
153. Nonaka, G., et al., *Regulon and promoter analysis of the E. coli heat-shock factor, sigma32, reveals a multifaceted cellular response to heat stress.* Genes Dev, 2006. **20**(13): p. 1776-89.
154. Niwa, T., et al., *Global analysis of chaperone effects using a reconstituted cell-free translation system.* Proc Natl Acad Sci U S A, 2012. **109**(23): p. 8937-42.
155. Ryabova, L.A., et al., *Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones.* Nat Biotechnol, 1997. **15**(1): p. 79-84.
156. Niess, A., M. Siemann-Herzberg, and R. Takors, *Protein production in Escherichia coli is guided by the trade-off between intracellular substrate availability and energy cost.* Microb Cell Fact, 2019. **18**(1): p. 8.
157. Lopez, P.J., et al., *The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo.* Mol Microbiol, 1999. **33**(1): p. 188-99.
158. Fischbach, M.A., et al., *How pathogenic bacteria evade mammalian sabotage in the battle for iron.* Nat Chem Biol, 2006. **2**(3): p. 132-138.
159. Braun, V., *Iron uptake by Escherichia coli.* Front Biosci, 2003. **8**: p. s1409-21.
160. Pressler, U., et al., *Genetics of the iron dicitrate transport system of Escherichia coli.* J Bacteriol, 1988. **170**(6): p. 2716-24.
161. Michael, A.J., *Biosynthesis of polyamines and polyamine-containing molecules.* Biochem J, 2016. **473**(15): p. 2315-29.
162. Igarashi, K. and K. Kashiwagi, *Polyamine transport in bacteria and yeast.* Biochem J, 1999. **344 Pt 3**: p. 633-42.
163. Blethen, S.L., E.A. Boeker, and E.E. Snell, *Argenine decarboxylase from Escherichia coli. I. Purification and specificity for substrates and coenzyme.* J Biol Chem, 1968. **243**(8): p. 1671-7.
164. Sutherland, I.W., *Structural studies on colanic acid, the common exopolysaccharide found in the enterobacteriaceae, by partial acid hydrolysis. Oligosaccharides from colanic acid.* Biochem J, 1969. **115**(5): p. 935-45.
165. Torres-Cabassa, A.S. and S. Gottesman, *Capsule synthesis in Escherichia coli K-12 is regulated by proteolysis.* J Bacteriol, 1987. **169**(3): p. 981-9.
166. Miller, M.B. and B.L. Bassler, *Quorum sensing in bacteria.* Annu Rev Microbiol, 2001. **55**: p. 165-99.
167. Liu, R. and H. Ochman, *Stepwise formation of the bacterial flagellar system.* 2007.
168. Fitzgerald, D.M., R.P. Bonocora, and J.T. Wade, *Comprehensive Mapping of the Escherichia coli Flagellar Regulatory Network*, in *PLoS Genet*. 2014.
169. Patrick, J.E. and D.B. Kearns, *Swarming motility and the control of master regulators of flagellar biosynthesis.* Mol Microbiol, 2012. **83**(1): p. 14-23.
170. Claret, L. and C. Hughes, *Rapid Turnover of FlhD and FlhC, the Flagellar Regulon Transcriptional Activator Proteins, during Proteus Swarming*, in *J Bacteriol*. 2000. p. 833-6.
171. Barembruch, C. and R. Hengge, *Cellular levels and activity of the flagellar sigma factor FliA of Escherichia coli are controlled by FlgM-modulated proteolysis.* Mol Microbiol, 2007. **65**(1): p. 76-89.
172. Sim, M., et al., *Growth rate control of flagellar assembly in Escherichia coli strain RP437*, in *Sci Rep*. 2017.
173. Koerber, J.T., M.J. Hornsby, and J.A. Wells, *An improved single-chain Fab platform for efficient display and recombinant expression.* J Mol Biol, 2015. **427**(2): p. 576-86.

174. Schirrmann, T., et al., *Oligomeric forms of single chain immunoglobulin (scIgG)*, in *MAbs*. 2010. p. 73-6.
175. Lee, Y.J. and K.J. Jeong, *Challenges to production of antibodies in bacteria and yeast.* J Biosci Bioeng, 2015. **120**(5): p. 483-90.
176. Yin, G., et al., *Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system.* mAbs, 2012. **4**(2).
177. Zawada, J.F., et al., *Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines.* Biotechnology and bioengineering, 2011. **108**(7): p. 1570-8.
178. Saul, F.A., et al., *Structural and Functional Studies of FkpA from Escherichia coli, a cis/trans Peptidyl-prolyl Isomerase with Chaperone Activity.* Journal of Molecular Biology, 2004. **335**(2): p. 595-608.
179. Arié, J.P., N. Sassoon, and J.M. Betton, *Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of Escherichia coli.* Molecular Microbiology, 2001. **39**(1): p. 199-210.
180. Ramm, K. and A. Plueckthun, *The periplasmic E. coli peptidyl-prolyl-isomerase FkpA: molecular mechanism of isomerase and chaperone activity.* Nova Acta Leopoldina, Supplementum, 2001. **16**(Structure, Self-Organization and Stability of Proteins: Experiments and Models): p. 171-172.
181. Entzminger, K.C., et al., *The Skp chaperone helps fold soluble proteins in vitro by inhibiting aggregation.* Biochemistry, 2012. **51**(24): p. 4822-34.
182. Hayhurst, A. and W.J. Harris, *Escherichia coli skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments.* Protein expression and purification, 1999. **15**(3): p. 336-43.
183. Walton, T.A. and M.C. Sousa, *Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation.* Molecular Cell, 2004. **15**(3): p. 367-374.
184. Koerber, J.T., M.J. Hornsby, and J.A. Wells, *An Improved Single-Chain Fab Platform for Efficient Display and Recombinant Expression.* J Mol Biol, 2014. **427**(2): p. 576-86.
185. Sonoda, H., et al., *Effects of cytoplasmic and periplasmic chaperones on secretory production of single-chain Fv antibody in Escherichia coli.* Journal of bioscience and bioengineering, 2011. **111**(4): p. 465-70.
186. Wang, R., et al., *Engineering production of functional scFv antibody in E. coli by co-expressing the molecule chaperone Skp.* Frontiers in cellular and infection microbiology, 2013. **3**: p. 72-72.
187. Hong, S.H., et al., *Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1.* Chembiochem, 2015. **16**(5): p. 844-53.
188. Kim, D.M. and J.R. Swartz, *Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli.* Biotechnol Bioeng, 2004. **85**(2): p. 122-9.
189. Su, X.C., et al., *Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enymes for selective 15N-labelling and production of perdeuterated proteins in H2O.* Journal of biomolecular NMR, 2011. **50**(1): p. 35-42.
190. Shin, J. and V. Noireaux, *Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system.* J Biol Eng, 2010. **4**: p. 9.

A Annex

Additional data generated during this work is referenced here and available on supplemental media. A free viewer for *.dna files is available at snapgene.com (SnapGene software from GSL Biotech).

1. Folder “Genome Editing of E. coli A19 and D10”

   cysS-AS-SBP-cassette.dna – *selection cassette to introduce SBP tag on the C-terminus of cysS.*

   NC_007779.1_CysS-SBP_(553,166...555,980).dna – *E. coli genome section after applying selection cassette cysS-AS-SBP-cassette to introduce SBP tag on cysS.*

   cysS-AS-SII-cassette.dna – *selection cassette to introduce StrepII tag on the C-terminus of cysS.*

   NC_007779.1_CysS-SII_(553,166...555,890).dna – *E. coli genome section after applying selection cassette cysS-AS-SII-cassette to introduce StrepII tag on cysS.*

   selection-cassette_prfA-SII-KAN-prmC.dna – *selection cassette to separate prfA (release factor I, RF1) and prmC and to introduce StrepII tag on prfA followed by kanamycin resistance gene.*

Table 15: All 152 down-regulated genes found in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt as determined by differential expression analysis.

UniProt #	Description (Gene ID)	Fold change
P37662	Uncharacterized MFS-type transporter (yhjX)	4.486728
P17117	Oxygen-insensitive NADPH nitroreductase (nfsA)	2.654024
P46119	Uncharacterized protein (ybjC)	2.646552
P76057	Uncharacterized protein (ydaQ)	2.434538
P0AAB2	Low molecular weight protein-tyrosine-phosphatase (wzb)	2.425544
P0C0U4	Ribosomal protein S6--L-glutamate ligase (rimK)	2.399991
P27294	Protein (inaA)	2.330002
P0C058	Small heat shock protein (ibpB)	2.266310
P33015	UPF0394 inner membrane protein (yeeE)	2.149057
P0AB46	Uncharacterized protein (ymgD)	2.146830
P33014	UPF0033 protein (yeeD)	2.142356
Q7DFV3	UPF0757 protein (ymgG)	2.069974
Q46822	Isopentenyl-diphosphate Delta-isomerase (idi)	2.058646
P21156	**Sulfate adenylyltransferase subunit 2 (cysD)**	**1.998895**
P0AEY5	Modulator of drug activity B (mdaB)	1.972932
P17846	Sulfite reductase [NADPH] hemoprotein beta-component (cysI)	1.941382
P0ACB2	Delta-aminolevulinic acid dehydratase (hemB)	1.903546
P16676	**Sulfate/thiosulfate import ATP-binding protein (cysA)**	**1.897836**
P0A930	Putative polysaccharide export protein (wza)	1.895862
P23845	**Sulfate adenylyltransferase subunit 1 (cysN)**	**1.835026**
P38489	Oxygen-insensitive NAD(P)H nitroreductase (nfsB)	1.828932
P0AEB0	**Sulfate transport system permease protein (cysW)**	**1.805111**
P0A6J1	Adenylyl-sulfate kinase (cysC)	1.797465
P17854	**Phosphoadenosine phosphosulfate reductase (cysH)**	**1.785610**
P32055	GDP-L-fucose synthase (fcl)	1.744886
P0ACD2	Putative colanic acid biosynthesis acetyltransferase (wcaF)	1.735364
P64614	Uncharacterized protein (yhcN)	1.728328
P33030	Zinc-binding GTPase (yeiR)	1.674750
P0AC88	GDP-mannose 4,6-dehydratase (gmd)	1.623178
P76445	Lipid A 1-diphosphate synthase (lpxT)	1.604774
P76115	Probable TonB-dependent receptor (yncD)	1.603764
P0AG78	Sulfate-binding protein (sbp)	1.601416
P0AEJ2	Isochorismate synthase (entC)	1.551492
P0AAY6	Uncharacterized protein (ybjN)	1.456755
P77414	Putative colanic acid biosynthesis glycosyl transferase (wcaA)	1.456636
A5A611	Uncharacterized protein (ymgI)	1.399751
P36879	Uncharacterized ABC transporter ATP-binding protein (yadG)	1.396914
P0AE58	Transcriptional activatory protein (caiF)	1.393343
P32056	GDP-mannose mannosyl hydrolase (gmm)	1.349554
P0ADV5	Uncharacterized protein (yhbW)	1.339174
P39375	Anti-adapter protein (iraD)	1.318783
P24175	Phosphomannomutase (manB)	1.268114
P0AFN6	Inner membrane transport permease (yadH)	1.248179
P0C054	Small heat shock protein (ibpA)	1.227245
P24174	Mannose-1-phosphate guanylyltransferase (manC)	1.208838
P13039	Enterochelin esterase (fes)	1.208134
P0ACV0	Lipid A biosynthesis lauroyltransferase (lpxL)	1.201600
P0AB35	Uncharacterized protein (ycfJ)	1.190928
P75616	Uncharacterized protein (yaaX)	1.189743
P32057	Putative colanic acid biosynthesis glycosyl transferase (wcaI)	1.176835
P0DMC9	Transcriptional regulatory protein (rcsA)	1.175531
P28629	Biodegradative arginine decarboxylase (adiA)	1.172362
P0AGL7	Ribosomal RNA small subunit methyltransferase E (rsmE)	1.171480
P71239	Putative colanic acid biosynthesis glycosyl transferase (wcaE)	1.158452
Q47083	HTH-type transcriptional regulator (cbl)	1.153285
P76387	Tyrosine-protein kinase (wzc)	1.148766
P71237	Putative colanic acid biosynthesis glycosyl transferase (wcaC)	1.142148
P0AAW9	Multidrug efflux pump accessory protein (acrZ)	1.140596
P04425	**Glutathione synthetase (gshB)**	**1.127602**
P07821	Iron(3+)-hydroxamate import ATP-binding protein (fhuC)	1.112371
P0ADE8	tRNA-modifying protein (ygfZ)	1.111558
P0AA10	50S ribosomal protein L13 (rplM)	1.093969
P52647	Probable pyruvate-flavodoxin oxidoreductase (ydbK)	1.089234
P39336	Uncharacterized protein (yjgL)	1.080104
P0ACJ5	Uncharacterized HTH-type transcriptional regulator (ybaO)	1.074414
P0ABB8	Magnesium-transporting ATPase, P-type 1 (mgtA)	1.069607
P0AEB5	Low conductance mechanosensitive channel (ynaI)	1.068731
P0A7X3	30S ribosomal protein S9 (rpsI)	1.052185
P21169	Ornithine decarboxylase, constitutive (speC)	1.042976
P0A7I7	GTP cyclohydrolase-2 (ribA)	1.038707
Q2MB16	Uncharacterized protein (yobH)	1.034854
P0AC53	**Glucose-6-phosphate 1-dehydrogenase (zwf)**	**1.033368**

UniProt #	Description (Gene ID)	Fold change
P32696	Phage shock protein G (pspG)	1.021475
P0AA47	Low-affinity putrescine importer (plaP)	1.017674
P77265	**Multidrug resistance-like ATP-binding protein (mdlA)**	**1.015572**
P0AE06	Multidrug efflux pump subunit (acrA)	1.015029
P0A8R4	Protein (slyX)	0.997496
P0ADN2	UPF0438 protein (yifE)	0.993595
P23890	Transcriptional activator (cadC)	0.980997
P31224	Multidrug efflux pump subunit (acrB)	0.973971
P0AD33	UPF0381 protein (yfcZ)	0.968684
P07822	Iron(3+)-hydroxamate-binding protein (fhuD)	0.951725
P76116	Uncharacterized protein (yncE)	0.946756
P25744	Multidrug resistance protein (mdtG)	0.936067
P0AFI9	Putative permease (perM)	0.919937
P27245	Multiple antibiotic resistance protein (marR)	0.919561
P02930	Outer membrane protein (tolC)	0.911654
P0A7F6	S-adenosylmethionine decarboxylase proenzyme (speD)	0.901667
P75855	Fimbrial subunit (elfA)	0.898458
P10378	Enterobactin synthase component E (entE)	0.897432
P0AAU2	Uncharacterized protein (ybfA)	0.891789
P77567	N-hydroxyarylamine O-acetyltransferase (nhoA)	0.891483
P06972	**Iron(3+)-hydroxamate import system permease protein (fhuB)**	**0.890981**
P21829	Pyridoxal phosphate phosphatase (ybhA)	0.880941
P23538	**Phosphoenolpyruvate synthase (ppsA)**	**0.864752**
P76471	Uncharacterized protein (yfaZ)	0.863520
P28861	Ferredoxin--NADP reductase (fpr)	0.862201
P39830	Inner membrane protein (ybaL)	0.860697
P0ADR0	Inner membrane protein (yqaA)	0.855330
P16703	Cysteine synthase B (cysM)	0.851683
P77733	Probable formate transporter 2 (focB)	0.835006
P0A6R3	DNA-binding protein (fis)	0.833322
A5A616	Uncharacterized protein (yneM)	0.828173
P15005	5-methylcytosine-specific restriction enzyme B (mcrB)	0.828148
P0AAG5	**Multidrug resistance-like ATP-binding protein (mdlB)**	**0.814136**
P23849	Trk system potassium uptake protein (trkG)	0.810289
P0AEQ3	Glutamine-binding periplasmic protein (glnH)	0.807150
P0A6N8	Elongation factor P-like protein (yeiP)	0.789392
P77306	Inner membrane protein (yqiK)	0.777757
P37147	UPF0716 protein (fxsA)	0.773494
P37661	Kdo(2)-lipid A phosphoethanolamine 7"-transferase (eptB)	0.770001
P09148	Galactose-1-phosphate uridylyltransferase (galT)	0.751936
P05824	DNA repair protein (recN)	0.747832
P0A6P7	Probable GTP-binding protein (engB)	0.747820
P37326	Putative prophage CPS-53 integrase (intS)	0.742175
P51025	S-formylglutathione hydrolase (frmB)	0.726757
P31802	Nitrate/nitrite response regulator protein (narP)	0.716343
P0A6Z3	Chaperone protein (htpG)	0.711308
P0A6A3	**Acetate kinase (ackA)**	**0.709088**
P31133	Putrescine-binding periplasmic protein (potF)	0.699763
P0A6Y8	Chaperone protein (dnaK)	0.699652
P77377	Lipopolysaccharide biosynthesis protein (wzxC)	0.690775
P31134	Putrescine transport ATP-binding protein (potG)	0.688690
P0ABR1	DNA-damage-inducible protein I (dinI)	0.688038
P65556	Uncharacterized Nudix hydrolase (yfcD)	0.687461
P43319	Uncharacterized fimbrial-like protein (yraK)	0.679398
P0AFM6	Phage shock protein A (pspA)	0.679294
P23842	Uncharacterized protein (yfeA)	0.678673
P36881	Putative phosphotransferase enzyme IIA component (yadI)	0.674895
P09147	UDP-glucose 4-epimerase (galE)	0.666788
P0AGC5	Membrane-bound lytic murein transglycosylase F (mltF)	0.660969
P77182	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein (mnmC)	0.654690
P0AEP1	Galactose-proton symporter (galP)	0.654514
P0AB61	Protein (yciN)	0.650061
P76071	Transposase InsH for insertion sequence element IS5Y (insH5)	0.647631
P63386	**Probable phospholipid import ATP-binding protein (mlaF)**	**0.646657**
P0ADL1	Purine ribonucleoside efflux pump (nepI)	0.637070
P0AE18	Methionine aminopeptidase (map)	0.635512
P42626	UPF0597 protein (yhaM)	0.635398
P0A715	2-dehydro-3-deoxyphosphooctonate aldolase (kdsA)	0.627075
P0AFP0	UPF0126 inner membrane protein (yadS)	0.622038
P27243	O-antigen ligase (rfaL)	0.617871
P0AAA5	Uncharacterized protein (ymcE)	0.617211
P32700	Uncharacterized protein (yjcB)	0.616439
P15286	Flagellar regulator (flk)	0.614675
P43337	Uncharacterized Nudix hydrolase (nudL)	0.607411
P37180	Probable Ni/Fe-hydrogenase 2 b-type cytochrome subunit (hybB)	0.607216
P0A908	MltA-interacting protein (mipA)	0.605244
P0ABY4	Flavodoxin-2 (fldB)	0.601268

UniProt #	Description (Gene ID)	Fold change
Q46856	Alcohol dehydrogenase (yqhD)	0.597109
P25437	S-(hydroxymethyl)glutathione dehydrogenase (frmA)	0.594772
P37767	Uncharacterized HTH-type transcriptional regulator (yfhH)	0.591761
P38392	Superinfection exclusion protein B (sieB)	0.590011

Table 16: All 309 down-regulated genes found in A19 rne::SBP Δ(ompT/lon/gor/fhuA) CSRW relative to A19 wt as determined by differential expression analysis.

UniProt #	Description (Gene ID)	Fold change
P09169	Protease 7 (ompT)	-8.970902
P0A9M0	Lon protease (lon)	-8.069021
P21170	Biosynthetic arginine decarboxylase (speA)	-7.645337
P16095	L-serine dehydratase 1 (sdaA)	-7.583413
P06715	Glutathione reductase (gor)	-7.544772
P0A6W9	Glutamate--cysteine ligase (gshA)	-7.316966
P06971	Ferrichrome-iron receptor (fhuA)	-5.517509
P30744	L-serine dehydratase 2 (sdaB)	-5.383357
P0A853	Tryptophanase (tnaA)	-5.099459
P25736	Endonuclease-1 (endA)	-4.728869
P07017	Methyl-accepting chemotaxis protein II (tar)	-3.526871
P04949	Flagellin (fliC)	-3.481082
P37646	Cyclic di-GMP phosphodiesterase (yhjH)	-3.246441
P09348	Motility protein A (motA)	-3.214320
P07018	Methyl-accepting chemotaxis protein IV (tap)	-3.134911
P07330	Chemotaxis response regulator protein-glutamate methylesterase (cheB)	-3.052960
P24216	Flagellar hook-associated protein 2 (fliD)	-3.009602
P0AF06	Motility protein B (motB)	-2.971710
P0A964	Chemotaxis protein (cheW)	-2.894882
P26608	Flagellar protein (fliS)	-2.659458
P0AEM6	RNA polymerase sigma factor (fliA)	-2.651567
P07364	Chemotaxis protein methyltransferase (cheR)	-2.622060
P0AE67	Chemotaxis protein (cheY)	-2.561267
P33235	Flagellar hook-associated protein 1 (flgK)	-2.479062
P0ABY2	Flagellar protein (fliT)	-2.332172
P68187	Maltose/maltodextrin import ATP-binding protein (malK)	-2.252193
P77609	Protein (flxA)	-2.219552
P0A9H9	Protein phosphatase (cheZ)	-2.213499
P0ABX2	Flagellar basal-body rod protein (flgC)	-2.124454
P75937	Flagellar hook protein (flgE)	-2.052378
P75936	Basal-body rod modification protein (flgD)	-2.046670
P29744	Flagellar hook-associated protein 3 (flgL)	-2.043006
P0ABW9	Flagellar basal body rod protein (flgB)	-1.991373
P76010	Flagellar brake protein (ycgR)	-1.950398
P0AEX9	Maltose-binding periplasmic protein (malE)	-1.948084
P39267	Uncharacterized protein (yjcZ)	-1.931014
P0DM85	Clamp-binding protein (crfC)	-1.913306
P0A6S0	Flagellar L-ring protein (flgH)	-1.907782
P0A9S3	Galactitol-1-phosphate 5-dehydrogenase (gatD)	-1.889904
P56614	Uncharacterized protein (ymdF)	-1.864568
P21179	Catalase HPII (katE)	-1.852451
P0AEM4	Negative regulator of flagellin synthesis (flgM)	-1.778174
P75691	Aldehyde reductase (yahK)	-1.773201
P33570	Transketolase 2 (tktB)	-1.745754
P75938	Flagellar basal-body rod protein (flgF)	-1.739406
P43533	Flagella synthesis protein (flgN)	-1.733642
P0A8G6	NAD(P)H dehydrogenase (quinone) (wrbA)	-1.712834
P75694	Uncharacterized protein (yahO)	-1.704066
P76402	UPF0339 protein (yegP)	-1.682244
P0A867	Transaldolase A (talA)	-1.663981
P76235	UPF0229 protein (yeaH)	-1.633866
P52627	Regulator of sigma S factor (fliZ)	-1.596180
P0ABX5	Flagellar basal-body rod protein (flgG)	-1.590508
P0AEH5	Protein (elaB)	-1.586165
P37685	Aldehyde dehydrogenase B (aldB)	-1.581132
P75942	Peptidoglycan hydrolase (flgJ)	-1.579898
P25798	Flagellar M-ring protein (fliF)	-1.573700
P0AB14	Uncharacterized protein (yccJ)	-1.563962
P0ADB1	Osmotically-inducible lipoprotein E (osmE)	-1.559353
P0A9V5	Uncharacterized HTH-type transcriptional regulator (yiaG)	-1.529653
P0AG84	Uncharacterized oxidoreductase (yghA)	-1.528804
P0ACY3	Uncharacterized protein (yeaG)	-1.527983
P75717	Putative uncharacterized protein (exoD)	-1.503757
P21361	Uncharacterized protein (yciG)	-1.499132

UniProt #	Description (Gene ID)	Fold change
P13035	Aerobic glycerol-3-phosphate dehydrogenase (glpD)	-1.493304
P77148	Uncharacterized protein (ydhS)	-1.487493
P0ADQ7	Uncharacterized protein (ygaM)	-1.470804
P06974	Flagellar motor switch protein (fliM)	-1.463427
P31678	Trehalose-6-phosphate phosphatase (otsB)	-1.452872
P19319	Respiratory nitrate reductase 2 alpha chain (narZ)	-1.444110
P0ABZ1	Flagellar motor switch protein (fliG)	-1.443547
P0A6S3	Flagellar P-ring protein (flgI)	-1.438007
P31130	Uncharacterized protein (ydeI)	-1.417368
P00490	Maltodextrin phosphorylase (malP)	-1.412605
P21362	Protein (yciF)	-1.410345
P77717	Uncharacterized lipoprotein (ybaY)	-1.409586
P11349	Respiratory nitrate reductase 1 beta chain (narH)	-1.398603
P0AE63	Cation transport regulator (chaB)	-1.392967
P31677	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] (otsA)	-1.369355
P21367	Uncharacterized protein (ycaC)	-1.369276
P22256	4-aminobutyrate aminotransferase (gabT)	-1.363628
P33368	Uncharacterized oxidoreductase (yohF)	-1.362339
P39451	Alcohol dehydrogenase, propanol-preferring (adhP)	-1.345162
P64567	Uncharacterized protein (yqgB)	-1.338143
P0AFH8	Osmotically-inducible protein Y (osmY)	-1.336501
P64503	Uncharacterized protein (yebV)	-1.329167
P25526	Succinate-semialdehyde dehydrogenase [NADP(+)] (gabD)	-1.326093
P0A991	Fructose-bisphosphate aldolase class 1 (fbaB)	-1.325711
P03841	Maltose operon periplasmic protein (malM)	-1.309363
P0ADX5	Uncharacterized protein (yhfG)	-1.299586
P37327	Inner membrane protein (yfdC)	-1.293789
P0C0L2	Peroxiredoxin (osmC)	-1.270691
P0ADE6	Uncharacterized protein (ygaU)	-1.269913
P64585	Inner membrane protein (yqjE)	-1.268559
P68191	Stationary-phase-induced ribosome-associated protein (sra)	-1.268135
P0A9T0	D-3-phosphoglycerate dehydrogenase (serA)	-1.264081
P37645	Uncharacterized protein (yhjG)	-1.263447
P37339	L-2-hydroxyglutarate oxidase (lhgO)	-1.263225
P60651	Agmatinase (speB)	-1.262171
P64581	Uncharacterized protein (yqjD)	-1.261914
P0ADB7	Entericidin B (ecnB)	-1.261402
P0ABX8	Flagellar protein (fliL)	-1.252672
P15977	4-alpha-glucanotransferase (malQ)	-1.244318
P77302	Probable diguanylate cyclase (ydaM)	-1.237686
P02943	Maltoporin (lamB)	-1.236530
P02916	Maltose transport system permease protein (malF)	-1.234442
P0AFW2	Ribosome modulation factor (rmf)	-1.233591
P29013	Uncharacterized protein (ycgB)	-1.230388
P33362	Glycine betaine-binding protein (yehZ)	-1.230016
P19317	Probable nitrate reductase molybdenum cofactor assembly chaperone (narW)	-1.224531
P64474	Uncharacterized protein (ydhL)	-1.219372
P52613	Flagellar protein (fliJ)	-1.218742
P75839	UPF0702 transmembrane protein (ycaP)	-1.218045
P33195	Glycine dehydrogenase (decarboxylating) (gcvP)	-1.212703
P02942	Methyl-accepting chemotaxis protein I (tsr)	-1.208979
P42620	Glutathionyl-hydroquinone reductase (yqjG)	-1.193429
P0ADU5	Protein (ygiW)	-1.169852
P41039	Uncharacterized protein (ybiI)	-1.157823
P27248	Aminomethyltransferase (gcvT)	-1.145338
P45470	Protein deglycase 2 (yhbO)	-1.128577
P31068	Flagellar assembly protein (fliH)	-1.128452
P18956	Gamma-glutamyltranspeptidase (ggt)	-1.126825
P0AET8	7-alpha-hydroxysteroid dehydrogenase (hdhA)	-1.122963
P76097	Uncharacterized protein (ydcJ)	-1.120692
P0AFH2	Oligopeptide transport system permease protein (oppB)	-1.117231
P0ABD3	Bacterioferritin (bfr)	-1.107781
P33361	Glycine betaine uptake system permease protein (yehY)	-1.103408
P26612	Cytoplasmic alpha-amylase (amyA)	-1.103188
P25527	GABA permease (gabP)	-1.101015
P33360	Glycine betaine uptake system ATP-binding protein (yehX)	-1.090925
P56262	Putative carboxymethylenebutenolidase (ysgA)	-1.089067
Q46857	2,5-diketo-D-gluconic acid reductase A (dkgA)	-1.085676
P76329	Mannosyl-3-phosphoglycerate phosphatase (yedP)	-1.082416
P76621	Protein (csiD)	-1.081632
P0A6J5	D-amino acid dehydrogenase (dadA)	-1.080846
P64517	Uncharacterized protein (yodC)	-1.079829
P0ADJ3	Uncharacterized protein (yhjR)	-1.067763
P29012	Alanine racemase, catabolic (dadX)	-1.064904
P25906	Putative oxidoreductase (ydbC)	-1.058936
P42616	Protein (yqjC)	-1.053505
P52037	Uncharacterized oxidoreductase (ygfF)	-1.051200

UniProt #	Description (Gene ID)	Fold change
P0AAV6	Uncharacterized protein (ybgS)	-1.047794
P21693	ATP-dependent RNA helicase (dbpA)	-1.047637
P76108	Putative ABC transporter periplasmic-binding protein (ydcS)	-1.046998
P77444	Cysteine desulfurase (sufS)	-1.043059
P68206	UPF0337 protein (yjbJ)	-1.039160
P77526	Disulfide-bond oxidoreductase (yfcG)	-1.038218
P0ABT2	DNA protection during starvation protein (dps)	-1.036592
P77804	Protein (ydgA)	-1.034732
P22586	Flagellar protein (fliO)	-1.033393
P25738	Acidic protein (msyB)	-1.030715
P64479	Uncharacterized protein (ydiZ)	-1.024766
P0AFH6	Oligopeptide transport system permease protein (oppC)	-1.015139
P09831	Glutamate synthase [NADPH] large chain (gltB)	-1.004202
P0AAW1	Uncharacterized protein (ybhP)	-0.999293
P76113	NADPH-dependent curcumin reductase (curA)	-0.996124
P33359	Glycine betaine uptake system permease protein (yehW)	-0.993738
P76214	Protein (ves)	-0.992564
P77795	Uncharacterized ABC transporter ATP-binding protein (ydcT)	-0.990094
P24252	Uncharacterized protein (ybgA)	-0.978409
P0AAX6	Uncharacterized protein (mcbA)	-0.978116
P0AE28	Protein (aroM)	-0.968885
P13482	Periplasmic trehalase (treA)	-0.962487
P0A6T9	Glycine cleavage system H protein (gcvH)	-0.959388
P77748	Uncharacterized protein (ydiJ)	-0.954760
P52614	Flagellar hook-length control protein (fliK)	-0.953386
P16681	Protein (phnB)	-0.951208
P0A901	Outer membrane lipoprotein (blc)	-0.949921
Q46866	Probable transcriptional regulator (ygiV)	-0.947979
P64485	UPF0410 protein (yeaQ)	-0.947188
P63235	Probable glutamate/gamma-aminobutyrate antiporter (gadC)	-0.944512
P76193	Probable L,D-transpeptidase (ynhG)	-0.944167
P20605	Probable adenosine monophosphate-protein transferase (fic)	-0.940192
P76407	Lipid kinase (yegS)	-0.928273
P77454	Glutaminase 1 (glsA1)	-0.926610
P77400	Inner membrane transport protein (ybaT)	-0.926463
P77690	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase (arnB)	-0.922502
P0AF59	Uncharacterized protein (yjdI)	-0.920389
P20966	PTS system fructose-specific EIIBC component (fruA)	-0.917105
P77528	Putative uncharacterized protein (peaD)	-0.907332
P28903	Anaerobic ribonucleoside-triphosphate reductase (nrdD)	-0.904757
P77398	Bifunctional polymyxin resistance protein (arnA)	-0.903269
P77213	Putative glutamate--cysteine ligase 2 (ybdK)	-0.902253
P0AD17	Inner membrane protein (yohC)	-0.899432
P0AFM4	Phosphate starvation-inducible protein (psiF)	-0.896799
P76155	Tail fiber assembly protein homolog from lambdoid prophage Qin (tfaQ)	-0.894890
P31131	Protein (ydeJ)	-0.894505
P0ABY7	Flagellar transcriptional regulator (flhC)	-0.891540
Q46851	L-glyceraldehyde 3-phosphate reductase (gpr)	-0.890774
P04982	D-ribose pyranase (rbsD)	-0.886732
P77757	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase (arnC)	-0.884815
P76027	Oligopeptide transport ATP-binding protein (oppD)	-0.880530
P0AEV1	Regulator of RpoS (rssB)	-0.873726
P65292	Uncharacterized lipoprotein (ygdI)	-0.870417
P0AFR9	Inner membrane ABC transporter permease protein (ydcV)	-0.867493
P0A8M0	Asparagine--tRNA ligase (asnS)	-0.863756
P0AB31	Uncharacterized protein (yceK)	-0.860371
P31062	DNA-packaging protein NU1 homolog (nohD)	-0.859988
P09832	Glutamate synthase [NADPH] small chain (gltD)	-0.859303
P38051	Isochorismate synthase (menF)	-0.858779
P0ACZ4	Positive transcription regulator (evgA)	-0.855253
P0AAQ6	Uncharacterized protein (ybaA)	-0.854877
P0A799	Phosphoglycerate kinase (pgk)	-0.852794
P00914	Deoxyribodipyrimidine photo-lyase (phrB)	-0.852469
P33358	HTH-type transcriptional regulator (mlrA)	-0.851328
P08997	Malate synthase A (aceB)	-0.842864
P64616	Uncharacterized protein (yhcO)	-0.832596
P69910	Glutamate decarboxylase beta (gadB)	-0.832214
P07650	Thymidine phosphorylase (deoA)	-0.831125
P69805	Mannose permease IID component (manZ)	-0.827851
P0AE30	Arginine ABC transporter permease protein (artM)	-0.827117
P0A9G6	Isocitrate lyase (aceA)	-0.826427
P75783	Moderate conductance mechanosensitive channel (ybiO)	-0.825223
P0AF86	Uncharacterized protein (yjfY)	-0.809712
P76100	Uncharacterized acetyltransferase (ydcK)	-0.809298
P77781	1,4-dihydroxy-2-naphthoyl-CoA hydrolase (menI)	-0.809110
P0AA84	Cardiolipin synthase B (clsB)	-0.807705
P0AC59	Glutaredoxin-2 (grxB)	-0.805890

UniProt #	Description (Gene ID)	Fold change
P39385	Uncharacterized protein (yjiN)	-0.804080
P0AFR0	NTE family protein (rssA)	-0.801070
P39274	Uncharacterized protein (yjdJ)	-0.800822
P69797	PTS system mannose-specific EIIAB component (manX)	-0.797020
P0C0L7	Proline/betaine transporter (proP)	-0.795181
P06993	HTH-type transcriptional regulator (malT)	-0.789519
Q2M5U1	UPF0391 membrane protein (ytjA)	-0.788781
P37348	UPF0759 protein (yecE)	-0.788022
P77165	Putative xanthine dehydrogenase iron-sulfur-binding subunit (yagT)	-0.787108
P76472	Probable 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase (arnD)	-0.780908
P77212	Probable pyridine nucleotide-disulfide oxidoreductase (rclA)	-0.780315
P75818	Uncharacterized lipoprotein (ybjP)	-0.779196
P0ACP7	HTH-type transcriptional repressor (purR)	-0.778661
P46125	Inner membrane protein (yedI)	-0.774862
P76121	Uncharacterized protein (yddH)	-0.774455
P37663	Uncharacterized protein (yhjY)	-0.771247
P19323	Formate hydrogenlyase transcriptional activator (fhlA)	-0.770585
P45425	N-acetylmannosamine kinase (nanK)	-0.769548
P0AGD1	Superoxide dismutase [Cu-Zn] (sodC)	-0.767380
P27250	Aldehyde reductase (ahr)	-0.762624
P0AGM0	UPF0118 inner membrane protein (yhhT)	-0.758213
P52095	Lysine decarboxylase, constitutive (ldcC)	-0.753609
P64550	L-alanine exporter (alaE)	-0.751459
P76145	Trans-aconitate 2-methyltransferase (tam)	-0.749286
P23847	Periplasmic dipeptide transport protein (dppA)	-0.746441
P77649	UPF0061 protein (ydiU)	-0.739693
P17109	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase (menD)	-0.738309
P0A8S5	Universal stress protein B (uspB)	-0.735734
P77735	Uncharacterized oxidoreductase (yajO)	-0.733105
P58094	Uncharacterized protein (yciX)	-0.728740
P42588	Putrescine aminotransferase (patA)	-0.728151
P76011	UPF0410 protein (ymgE)	-0.727685
P0ADI7	Isochorismatase family protein (yecD)	-0.724656
P76250	HTH-type transcriptional regulator (dmlR)	-0.716444
P30859	Putative ABC transporter arginine-binding protein 2 (artI)	-0.715983
P0AE34	Arginine ABC transporter permease protein (artQ)	-0.715468
P0A6D3	3-phosphoshikimate 1-carboxyvinyltransferase (aroA)	-0.705841
P76172	Uncharacterized protein (ynfD)	-0.705748
P37642	Inner membrane protein (yhjD)	-0.704695
P75713	(S)-ureidoglycine aminohydrolase (allE)	-0.699017
P12994	UPF0098 protein (ybhB)	-0.695890
P77544	Glutathione S-transferase (yfcF)	-0.694611
P0ACX9	Uncharacterized protein (ydiE)	-0.694027
P09424	Mannitol-1-phosphate 5-dehydrogenase (mtlD)	-0.692888
P0AB65	Acylphosphatase (yccX)	-0.691293
P19926	Glucose-1-phosphatase (agp)	-0.691230
P0AA99	Putative L,D-transpeptidase (yafK)	-0.688274
P76114	HTH-type transcriptional regulator (mcbR)	-0.685283
P45748	Threonylcarbamoyl-AMP synthase (tsaC)	-0.683619
P77334	Cyclic di-GMP phosphodiesterase (gmr)	-0.679334
P12758	Uridine phosphorylase (udp)	-0.675011
P18843	NH(3)-dependent NAD(+) synthetase (nadE)	-0.672648
P21889	Aspartate--tRNA ligase (aspS)	-0.672513
P03024	HTH-type transcriptional regulator (galR)	-0.665949
P0C8J6	D-tagatose-1,6-bisphosphate aldolase subunit (gatY)	-0.665326
P11289	Uncharacterized protein (yfiL)	-0.664572
P0AER0	Glycerol uptake facilitator protein (glpF)	-0.660152
P76318	Putative SOS response-associated peptidase (yedK)	-0.658720
P77489	Putative xanthine dehydrogenase molybdenum-binding subunit (yagR)	-0.655155
P0AD47	Inner membrane protein (yphA)	-0.652597
P62707	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (gpmA)	-0.645018
P76176	Uncharacterized serine protease (ydgD)	-0.644365
P32701	Putative cyclic-di-GMP phosphodiesterase (yjcC)	-0.644186
P0AB71	Fructose-bisphosphate aldolase class 2 (fbaA)	-0.642063
P37338	HTH-type transcriptional repressor (csiR)	-0.637937
P15770	Shikimate dehydrogenase (NADP(+)) (aroE)	-0.637019
P39835	High-affinity gluconate transporter (gntT)	-0.636712
P75804	Soluble aldose sugar dehydrogenase (yliI)	-0.636471
P0ACI0	Right origin-binding protein (rob)	-0.635975
P33011	Inner membrane protein (yeeA)	-0.632588
P33012	DNA gyrase inhibitor (sbmC)	-0.627461
P51981	L-Ala-D/L-Glu epimerase (ycjG)	-0.626046
P37659	Protein homolog (bcsG)	-0.626029
P64451	Uncharacterized lipoprotein (ydcL)	-0.618746
P0ACA1	Uncharacterized GST-like protein (yibF)	-0.618355
P21437	Fructose-1,6-bisphosphatase 2 class 2 (yggF)	-0.616238
P42619	Inner membrane protein (yqjF)	-0.611901

UniProt #	Description (Gene ID)	Fold change
P0A8S9	Flagellar transcriptional regulator (flhD)	-0.608896
P0AAF6	Arginine transport ATP-binding protein (artP)	-0.608397
P45424	Uncharacterized protein (yhcH)	-0.607492
P04983	Ribose import ATP-binding protein (rbsA)	-0.607421
P37353	2-succinylbenzoate--CoA ligase (menE)	-0.606305
P76223	Protein (ynjB)	-0.603729
P0AFD4	NADH-quinone oxidoreductase subunit H (nuoH)	-0.600277
P52697	6-phosphogluconolactonase (pgl)	-0.599612
P29208	o-succinylbenzoate synthase (menC)	-0.599185
P0AAN5	Uncharacterized protein (yaiA)	-0.597835
P63201	HTH-type transcriptional regulator (gadW)	-0.592388
P0AFE8	NADH-quinone oxidoreductase subunit M (nuoM)	-0.589928
A5A614	UPF0509 protein (yciZ)	-0.589511
P60240	RNA polymerase-associated protein (rapA)	-0.586654

Publication

Daniel Foshag, Erik Henrich, Ekkehard Hiller, Miriam Schäfer, Christian Kerger, Anke Burger-Kentischer, Irene Diaz-Moreno, Sofía M. García-Mauriño, Volker Dötsch, Steffen Rupp, Frank Bernhard.

The E. coli S30 lysate proteome: A prototype for cell-free protein production

New Biotechnology, Volume 40, Part B, 2018, Pages 245-260.

Declaration of Authorship

I hereby certify that the dissertation entitled: “Optimization of Cell-Free Protein Synthesis by Proteomics and Metabolic Engineering of *Escherichia coli* A19“ is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly quoted.

Erklärung über die Eigenständigkeit der Dissertation

Ich versichere, dass ich die vorliegende Arbeit mit dem Titel: “Optimization of Cell-Free Protein Synthesis by Proteomics and Metabolic Engineering of *Escherichia coli* A19“ selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe; aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht.

Name/Name: Daniel Foshag

Signed/Unterschrift: _____

Date/Datum: 23rd of October 2019