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

Daniel Foshag1\*, Erik Henrich2\*, Ekkehard Hiller3§, Miriam Schäfer3‡, Christian Kerger3, Anke Burger-Kentischer3, Irene Diaz-Moreno4, Sofía M. García-Mauriño4, Volker Dötsch2, Steffen Rupp3#, Frank Bernhard2#

\*authors contributed equally

*1Institute for Interfacial Process Engineering and Plasma Technology, University of Stuttgart, Stuttgart, Germany*

*2Institute of Biophysical Chemistry, Centre for Biomolecular Magnetic Resonance, J.W. Goethe-University, Frankfurt am Main, Germany*

*3Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany*

*§present address: Chemisches und Veterinäruntersuchungsamt, Stuttgart, Germany*

‡*present address: Institute of Biochemistry, Christian-Albrechts-Universität zu Kiel, Kiel, Germany*

*4Instituto de Investigaciones Químicas (IIQ) – Centro de Investigaciones Científicas Isla de la Cartuja (cicCartuja), Universidad de Sevilla – Consejo Superior de Investigaciones Científicas (CSIC), Sevilla , Spain*

#To whom correspondence should be addressed: fbern@bpc.uni-frankfurt.de, steffen.rupp@igb.fraunhofer.de

**Running title**

S30 lysate proteome for cell-free expression

**Abbreviations**

*GeLC-MS/MS:* Pre-fractionation by 1D-SDS-PAGE coupled with liquid chromatography mass spectrometry; FDR: false discovery rate; *ABC-transporters:* ATP-binding cassette transporters; *CECF:* continuous exchange cell-free; *CFPS:* cell-free protein synthesis; *emPAI:* exponentially modified protein abundance index; *GNAT:* GCN5-related N-acetyltransferase; *GO:* gene ontology; *ICPL:* isotope coded protein label; *KEGG:* Kyoto encyclopedia of genes and genomes; *MALDI:* matrix-assisted laser desorption/ionization; *MWCO:* molecular weight cut-off; *NADH:* reduced form of nicotinamide adenine dinucleotide *NMR:* nuclear magnetic resonance; *SDS-PAGE:* sodium dodecyl sulfate poly acrylamide gel electrophoresis; *sGFP:* shifted GFP; *sfGFP:* superfolder GFP; *TF:* trigger factor; apiRBP: apicoplast RNA Binding Protein; TOF*:* time of flight

**Abstract**

Protein production using processed cell lysates is a core technology in synthetic biology and these systems are excellent to produce difficult toxins or membrane proteins. However, the composition of the central lysate of cell-free systems is still a black box. *Escherichia coli* lysates are most productive for cell-free expression, yielding several mgs of protein per ml of reaction. Their preparation implies proteome fractionation, resulting in strongly biased and yet unknown lysate compositions. Many metabolic pathways are expected to be truncated or completely removed. The lack of knowledge of basic cell-free lysate proteomes is a major bottleneck for directed lysate engineering approaches as well as for assay design using non-purified reaction mixtures.

 This study is starting to close this gap by providing a blueprint of the S30 lysate proteome derived from the commonly used *E. coli* strain A19. S30 lysates are frequently used for cell-free protein production and represent the basis of most commercial *E. coli* cell-free expression systems. A fraction of 821 proteins was identified as the core proteome in S30 lysates, representing approximately a quarter of the known *E. coli* proteome. Its classification into functional groups relevant for transcription/translation, folding, stability and metabolic processes will build the framework for tailored cell-free reactions. As an example, we show that SOS response induction during cultivation results in tuned S30 lysate with better folding capacity, and improved solubility and activity of synthesized proteins. The presented data and protocols can serve as platform for the generation of customized cell-free systems and product analysis.

**Keywords**

Cell-free expression, synthetic biology, expression engineering, chaperones, systems biology, protein production

**Introduction**

During the last two decades, cell-free expression has become a standard platform complementing well established *in vivo* protein production systems. The inherently open nature of cell-free protein synthesis offers manifold options to enable efficient production of soluble as well as membrane proteins [1-5]. Due to the lack of boundaries like a cell-wall or cell-membranes, customized artificial expression environments by adding ligands, co-factors or hydrophobic compounds [6-10] can easily be generated. Difficult proteins having problems in solubility, folding or assembly are preferred targets for cell-free expression. Combinatorial labelling of proteins [11, 12], site-specific insertion of non-natural amino acids [13, 14] and the production of toxins or membrane proteins in artificial environments [15, 16] are further frequent applications. Nevertheless, significant limitations either for production or for rapid product quality analysis still exist due to background activities present in the cell-free reaction lysates.

Very efficient and most frequently used for cell-free expression are lysates prepared from *E. coli* cells. The lysate preparation has been optimized for high translation efficiency and includes harvesting at logarithmic growth phase and fractionation by different centrifugal forces. Very frequent is centrifugation at 30,000 x *g* resulting in the S30 lysate [17-20]. Furthermore, extensive dialysis and other processing steps result in the precipitation of less stable lysate proteins. The final proteome composition of the S30 lysate as the main system component is thus unknown. However, such knowledge would be valuable for refined processes such as eliminating or suppressing critical enzymes for specific assay development directly in the reaction lysates, implementing sets of identified enzymes for directed product modification or complementing truncated pathways for supporting product formation or quality. In order to address some of these problems, the PURE cell-free expression system has previously been developed by reconstituting the *E. coli* translation machinery from individually purified components [21]. However, the low expression efficiency indicates that protein production is significantly supported by complex interactions of various lysate components. Proteomic studies of more complex but highly efficient lysates such as S30 are therefore necessary to enable preparative scale product formation in defined environments and to open up new possibilities for tuning reaction conditions and finally for the improvement of product quality [22]. This study presents the first proteome analysis of the highly productive S30 lysate based on the *E. coli* K12-derivative A19, a primary source for cell-free lysate preparation [20].

The analyzed S30 lysates were prepared by optimized protocols that have already resulted in the production of numerous proteins for structural and functional characterization [4, 18, 22]. However, we further paid attention to cultivation conditions that lead to dynamic proteome changes of *E. coli* cells in adaption to *e.g.* stress or starvation [23, 24]. Exposure of *E. coli* A19 cells to a temperature shift and ethanol induces the SOS response resulting in an increased production of folding chaperones and other potentially beneficial compounds [25]. We have exemplified such an S30 lysate tuning and present a quantitative proteome analysis of S30-S lysates derived from *E. coli* cells grown under SOS response inducing conditions [26, 27]. The modified S30-S lysate proteomes were characterized according to (*i*) regulated proteome composition, (*ii*) protein production efficiency using a standard reporter protein and (*iii*) folding efficiency of various difficult-to-express model proteins.

Two standard proteomics approaches were implemented to address different underlying questions. First, our main purpose was the qualitative proteome analysis of the standard S30 lysate by focusing on the most sensitive identification strategy to provide an as complete as possible S30 lysate blueprint. The result was the identification of 1074 proteins containing a core of 821 protein repeatedly identified in different S30 lysate preparations. The averaged values of the so-called exponentially modified protein abundance index (emPAI) were taken to provide an approximate and relative abundance estimation of each protein in the lysate, based on the number of sequenced peptides per protein (protein coverage). Second, a quantitative proteomics analysis was performed by focusing on the identification of up- and down-regulated proteins in standard S30 lysates relative to S30 lysates derived from *E. coli* cells after inducing a SOS response (S30-S). The comparison of two distinct S30 lysate samples allows the usage of a so-called isotope-coded protein labelling (ICPL) strategy resulting in a defined mass difference of peptides of the same protein derived from either S30 or S30-S lysate. As these peptides are simultaneously analyzed, the measured intensity is used to quantify proteins from different samples relative to each other. The ICPL strategy is therefore suitable to identify regulated proteins, while the emPAI value compares different proteins in the same sample based on protein coverage and is thus used to estimate the overall protein diversity and abundance [28, 29]. The general validity of the emPAI value as an approximate estimation of protein abundance is accepted and it is routinely used [30-32]. Nevertheless, protein characteristics such as ionizability or hydrophobicity can influence this value in individual cases and this should be considered upon interpretation. The ICPL strategy compares only identical peptides or proteins of different samples simultaneously and is therefore not protein dependent. However and in contrast to the emPAI values, ICPL provides no quantitative information about different proteins within a lysate sample and is therefore not suitable for an overview study of total protein composition.

This study provides, for the first time, a qualitative analysis of the cell-free S30 lysate proteome including approximate and relative quantitation of the majority of identified proteins based on the emPAI value [29]. In addition, we provide a thorough insight into up- and down-regulated proteins of the S30 lysate proteome following the SOS response during cultivation of the *E. coli* source cells relative to standard S30 lysate, using the quantitative isotope-coded protein labelling (ICPL) strategy [33]. The presented data can serve as a guideline for future protein expression strategies in view of assay development, streamlined production and fine-tuning of product quality. Addressing identified precursor degrading enzymes or proteases as well as removing apparent bottlenecks within the enzymatic network necessary for protein biosynthesis by directed optimization could result into increased productivity of cell-free expression systems. The report will help to render cell-free protein production a more reliable technique with higher precision, thus making the folding, assembly and quality control of newly synthesized proteins more predictable. Applications of systems biology such as rebuilding synthetic networks or designing artificial circuits by directed engineering will further profit from the precise knowledge of the S30 lysate proteome background.

**materials and Methods**

*Materials and Reagents -* 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 [34]*) was obtained from *E. coli* Genetic Stock Centre (Yale University, CT, USA). Serva ICPL Kit (Cat.# 39230.01) was used for ICPL reaction for quantitative proteomics experiments.

*Cloning of apiRBP-sfGFP* - An uncharacterized RNA Binding Protein (RBP) from *Plasmodium vivax* SaI-1 (NCBI Reference Sequence: XM\_001616500.1, residues 33-134) was expressed from a modified pET21a(+) vector that included a C-terminal GFP-tag. Cloning was performed using the Infusion cloning kit (Clontech) following the manufacturer’s instructions. Primer sequences are available upon request.

*Cell-Free Lysate Preparation -* All lysates were prepared from *E. coli* strain A19. For S301-4 lysates, 10 L of YPTG medium (10 g/L yeast extract, 16 g/L tryptone, 5 g/L NaCl) supplemented with 100 mM of glucose and potassium phosphate buffer (22 mM KH2PO4, 40 mM K2HPO4) were inoculated with a fresh preculture grown in LB medium in a ratio of 1:100 and incubated in a fermenter at 37°C with vigorous stirring at 500 rpm and good aeration [20, 359. As soon as the cells reached mid log growth phase (OD600 4.0 – 5.0), the culture broth was cooled down to 20-18°C. For the generation of the stress lysates (S30-S), 300 mL ethanol were added at OD600 of 4.0 and the fermentation temperature was increased to 42°C for 45 min before chilling to 20-18°C (**Figure 1A**). The chilled cells were harvested by centrifugation (6,000 x *g* for 20 min at 4°C) and cell pellets were washed three times with ice-cold buffer A (10 mM Tris-acetate, 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-acetate, 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 x *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-acetate, 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 x *g* for 30 min to remove precipitates. The clarified lysate was flash-frozen in liquid nitrogen and stored at -80°C.

*Cell-Free Protein Production -* Synthesis of proteins using cell-free lysates based on T7-polymerase transcription was carried out in a two-compartment configuration as described previously [20, 35, 36]. Protein production was conducted on an analytical scale with 55 µL of reaction mixture and 825 µL of feeding mixture. Reactions were performed in customized mini dialyzer devices [36] 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.

*GFP Fluorescence Analysis -* 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 the soluble and the insoluble fraction was analyzed as described previously [37, 38]. 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 initial 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/Zurich, Switzerland). Measurement parameters for excitation, emission and reads per well were set to 485 nm, 510 nm and 10, respectively. The GFP concentration was then determined using a reference curve generated with purified sGFP of known concentration [37, 38]. The soluble fraction [%] of sGFP fusion protein was defined as the ratio of sGFPsoluble over sGFPinsoluble.

*HsGNA1 Activity Assay -* The activity assay was performed as described previously [39, 40]. Briefly, after the synthesis 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 x *g* at4°C for 10 min) 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 MgCl2, 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 terminated 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.

*GeLC-MS/MS Analysis using MALDI-TOF/TOF -*

**Sample preparation:** 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 80 µL) and proteins were precipitated overnight at -20°C. Precipitated protein was pelleted by centrifugation (20,000 x *g* at 4°C for 15 min) and washed once with 80% ice-cold 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 GeLC-MS/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 GeLC-MS/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.

**Separation 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 mm3 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 [41]). The gel pieces were covered in trypsin solution (25 ng/µl) 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 C18 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.

**LC separation and spotting**: For LC separation, the peptide-pellets 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 NH4H2PO4) prior to depositing at a 10-sec interval directly on MTP AnchorChipTM 1536 BC 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.

**Mass spectrometry analysis**: 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 initially processed by FlexControl, FlexAnalysis 3.3 and WARP-LC software (Bruker Daltonics). Maximal 30 MS2 spectra per spot were acquired using LIFT mode based on a mass list generated from a selection of most intense peaks (signal/noise ratio ≥ 15) by WARP-LC after MS1 survey scans were completed.

Raw MS1 and MS2 spectra were re-processed as profile using Mascot Distiller version 2.6.1 (Matrix Science, UK) using default processing parameters (default.Bruker.Baf.opt). MS2 spectra were submitted to an in-house Mascot Server version 2.6 (Matrix Science, UK) and searched against reference proteome of *E. coli* MG1655 (Uniprot Proteome ID: UP000000625, 4315 sequences, as downloaded on 19th of January 2017) and collections of common contaminants (MPI Martinsried, 247 sequences, as 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 additional variable modifications for light and heavy label, respectively: ICPL (K) and ICPL 13C (6) (K). For quantitative analysis, the median peptide ratio was determined using Mascot Distiller’s Quantitation Toolbox and the individual datasets were normalized to median = 0. The single peptide ratios were exported considering quality parameters (cutoff: Fraction 0.7; Correlation 0.9; Peptide Standard Error 0.1; only peptide sequences that were unique to one protein were used for quantitation). 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 [42, 43]. 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. All assigned peptide sequences and peptide ratios of qualitative and quantitative datasets are listed in **Table S1**. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium *via* the PRIDE [44] partner repository with the dataset identifier PXD004646.

*Experimental Design and Statistical Rationale -* Qualitative MS analysis served solely to identify most prevalent proteins present in S30 lysates. For this purpose, we prepared four biological replicates (namely S301-4) according to standard protocol used in our lab [20]. The FDR (peptide-level) was estimated by searching the individual datasets against decoy database. The significance threshold for peptides was adjusted accordingly to reach an FDR (peptide-level) ≤ 1%. The FDR (protein level) was then estimated based on protein hits in the decoy database and the dataset was filtered using the Mascot Score to set an FDR (protein level) ≤ 1%. Quantitative MS analysis was carried out to identify and quantitate up- and down regulated proteins relative to standard lysate (S301). The standard lysate (S301) and three biological replicates of heat shock lysates (S30-S), namely S30-S1, S30-S2 and S30-S3, were prepared according to the same lysate preparation protocols. As ratios did not follow a normal distribution, we used a permutation test [42, 43] that is distribution independent to assess significantly changed peptide ratios.

*Gene Ontology (GO)-Analysis* – The GO analysis was performed using Cytoscape (version 3.4.0) and the plugin ClueGO (version 2.3.2). As marker list the set for *E. coli* (*Escherichia coli* [562, 511145]) was used. The Gene Ontology sources “Biological Process (GOA)” (as downloaded on 17th of February 2017), “Cellular Compartment (GOA)” (as download on 17th of February 2017) and KEGG pathways (as downloaded on 20th of February 2017) were used for analysis. The custom reference set containing 4315 proteins from the reference proteome (Uniprot ID: UP000000625) were used for comparison.

**Results and Discussion**

**Preparation of *E. coli* S30 Lysates for Proteome Analysis**

S30 lysates represent one of the most commonly used standards for cell-free expression reactions. Essential and therefore common features of almost all preparation procedures are cell harvesting at mid-log growth phase, centrifugation of the lysate at 12,000-30,000 x *g* and a so-called run-off step, which includes incubation at 37°C to 42°C to dissociate endogenous mRNA and ribosomes. The procedure results in sedimentation and precipitation of numerous proteins, thus fractionating the original *E. coli* proteome and resulting in a yet unknown proteome composition of the final S30 lysates. We have therefore analyzed the S30 lysate proteome composition qualitatively by using samples strictly prepared according to standard protocols for highly efficient commercial S30 lysate (e.g. CUBE Biotech) preparation [20] (**Figure 1A)**. The proteomes of four independent replicates, namely S301-4, were analyzed. The *E. coli* source cells were grown at a growth rate of 1.9 h-1 (td ≈ 22 min) (**Figure 1B**), and the total fermentation process including the cooling of the cells took approximately 3 h. The protein production efficiency of the S301-4 lysates was estimated by synthesis of the reporter protein sGFP in our routine cell-free expression protocol, and found to be comparable resulting into an average of 4.5 mg/mL sGFP (**Figure 1C**).

**Overall Proteome Coverage of *E. coli* A19 S30 Lysates**

The S30 lysates were initially fractionated on the protein-level by SDS-PAGE. Each lane was divided into twelve fractions. Following in-gel digestion, the peptides were extracted from individual gel-fractions and further analyzed by LC-MS/MS to identify the separated proteins (**Figure 2**). The MS2 data were searched against total *E. coli* MG1655 / K12 proteome ([45] 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 [46]. 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 lysate proteome.

 Using the proteome database (Uniprot ID: UP000000625), we analyzed the proteomes of the four lysates S301-4 and applied statistical means to reach an estimated FDR (false discovery rate) on peptide- and protein-level of ≤ 1% for each individual proteome. According to our analysis, the lysate proteomes are composed of 820, 902, 952 and 805 proteins for the individual S301-4, respectively. Considering the FDR on the protein level, among the individual proteomes ≤9 proteins may be false positive hits. Any false positive hit, if present, is likely to be among proteins that were identified with only one or two peptides. Overall, we found 1074 proteins in S301-4, representing approximately 25% of the theoretically predicted proteins present in the *E. coli* MG1655 reference proteome. However, many of the 4,315 proteins present in the reference proteome are regulated and might therefore not be expressed under our conditions, and some predicted proteins may not be present at all. It has been shown that exponentially growing *E. coli* cells*,* as used for S30 lysate preparation*,* express approximately 2,600 proteins of which 830 are membrane proteins [47]. This number might provide a more accurate estimation of theoretically expected proteins present directly after cultivation but before extract processing. Centrifugation steps at 30,000 x *g* as well as further S30 lysate processing resulted in the lack of detection of approximately 65% of the expected proteins. Detection is certainly limited by our implemented sample preparation methods for LC-MS/MS, by the selected ionization method and by physiochemical properties of proteins/peptides. It is important to point out that we separated the S30 lysate at the protein level (SDS-PAGE) prior to fractionation on the peptide level which allows us to cover a larger dynamic range of protein concentrations and also detect low-abundant proteins. Nevertheless, as a general limitation of such an approach, it is clearly more likely to identify highly-abundant proteins compared to very low-abundant proteins that might still be present in the S30 lysate in a few copies.

 Out of the total detected 1074 proteins, 648 (60.3%) were common to S301-4, 173 (16.1%) were found in at least three S30 lysates and 252 (23.6%) were shared between two lysates or were uniquely identified in only one lysate (**Figure 3A**). Based on these data, 76.4% or 821 proteins that were identified in at least three out of four S30 lysates were defined as a core S30 lysate proteome having only low variability. The emPAI values of proteins identified in at least three proteomes were averaged to provide an approximate estimation of protein abundances in S30 lysate based on sequenced peptides per protein [29]. Previous studies showed that these values correlate well with independently measured copy number in *E. coli* cells [48]. 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 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, we accepted only assignments, which are mostly based on experimental data. 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.

**Classification of the *E. coli* A19 S30 Lysate Proteome**

 Analyzing the 821 proteins of the S30 lysate core proteome, we found that 705 proteins (~86%) are classified by the term “cell part”, comprising 31% of all proteins within this group (**Figure 3B**). 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] [49]. Within those groups, the GO-term “cytoplasmic part” is 2-fold over-represented with 625 hits (p<0.01, = 62% of 1016 reference proteins), whereas proteins of the “cell periphery” are highly under-represented 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, **Table S1**). The compartment “Periplasmic Space” appears with 42 unbiased hits (~30% of 138 reference proteins) (**Figure 3B**). The low “cell periphery” fraction is expected as most membranes will be removed by the 30,000x*g* centrifugation steps during S30 lysate preparation. Interestingly, some membrane proteins still remain in the lysate potentially in residual small vesicles [50]. Particularly abundant components of the respiratory chain involving several subunits of the ATP-synthase and the NADH-quinone oxidoreductase were detected. This is in agreement with reports on the “Cytomim” cell-free system employing inverted inner membrane vesicles of S30 lysates for energy regeneration *via* oxidative phosphorylation [51, 52]. 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 upon the co-translational insertion of membrane proteins in supplied lipid bilayers of liposomes [53] or nanodiscs [54]. Recent studies suggest that the insertion *via* the translocon is thermodynamically similar to spontaneous insertion [55] while the major task of the translocon machinery might be rather to guarantee targeting and coordinated membrane protein assembly [56].

The presence of certain porins of the outer membrane such as OmpA or OmpC in S30 lysates was previously speculated based on detected background activities during single channel electrophysiology measurements [57]. Accordingly, these genuine membrane-spanning proteins were repeatedly identified with high scores in our proteome analysis. However, most other proteins identified in the class of membrane proteins constitute soluble subunits of membrane spanning complexes such as the group of ABC-transporters. None of the identified 15 ABC transporter related gene products are considered to be multi-spanning transmembrane proteins, but rather solute binding subunits, such as ArtI [58], ProX [59] or MlaC [60]. The deficiency of integral membrane protein detection could furthermore be partially caused by a technical bias towards more hydrophilic peptides during LC-MS/MS analysis. However, since we used a GeLC-MS/MS approach based on 1D-SDS-PAGE, we expect membrane proteins to be efficiently solubilized and fractionated, as previously demonstrated [61]. Therefore, it is unlikely that this bias has primarily technical reasons, but is rather caused by membrane depletion during S30 lysate preparation.

In order to analyze how major metabolic pathways are represented in the S30 lysate proteome, we used the Cytoscape plugin ClueGO (Marker list: *Escherichia coli* [562, 511145]; Source: KEGG as downloaded on 20th of February 2017). We found that major physiological *E. coli* pathways are significantly over-represented (p≤0.05) in the S30 lysate proteome (**Figure 4**). In particular, important catabolic pathways involved in purine-/pyrimidine-biosynthesis and energy generation such as glycolysis/gluconeogenesis and pentose-phosphate pathway are over-represented as expected for a lysate derived from exponentially growing *E. coli*. Reassuringly, components central to translation, namely the ribosome and the aminoacyl tRNA synthetases (**Table 1**) are also strongly over-represented with high emPAI values, suggesting that they are enriched in S30 lysate and highly abundant. Exceptions are the glycine-tRNA ligase subunit GlyQ, the glutamyl-Q tRNA synthetase GluQ and the methionyl-tRNA formyltransferase Fmt required for initiation of translation, respectively, 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 over-represented (p≤0.05). S30 lysates are mostly used in a coupled transcription/translation reaction, based on the highly processive bacteriophage derived T7 polymerase (**Figure 4**). This leads to a rapid accumulation of free mRNA, that in turn is prone to degradation by the RNA degradation machinery. Supplying RNase inhibitors to cell-free reactions is therefore highly recommended. As a consequence of the noted membrane protein deficiency, pathways including membrane proteins such as ABC transporters and two component systems are significantly under-represented (**Figure 4**).

**Proteins Relevant for Recombinant Protein Production**

Efficient translation is the core application of S30 lysates in cell-free expression. Protein production is usually operated as a coupled transcription/translation system by either using T7-promoters and supplying T7-RNA-polymerase or by using the more complex endogenous RNA polymerase [62, 63]. Degradation and formation of unfavorable mRNA tertiary structures could be reduced by using the endogenous transcription system comprising RNA polymerase as well as compatible sigma factors [64]. All four subunits of the *E. coli* RNA polymerase core enzyme could be identified and overall high emPAI values indicate abundant copies of these proteins within the lysate (**Table 2**). Three 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 lysate also contains the cold-shock transcription factors CspA, CspC, CspE and CspG. It is tempting to speculate that these transcription factors are induced during the cooling procedure of the *E. coli* cell fermentation broth.

In respect of the translation machinery, approximately 95% of the ribosome associated proteins could be identified, suggesting a high abundance of this key complex necessary for cell-free protein synthesis (**Figure 4**). It is evident that all subunits necessary for ribosome assembly are present to maintain the observed protein expression efficiency (**Figure 1C**). However, further factors are required for efficient initiation, elongation and translation termination. The initiation of prokaryotic translation involves formylation of the start methionine and the initiation factors IF1-3. The obtained ratio of averaged emPAI values for InfA:InfB:InfC of approximately 5:1:2 (**Table 3**) matches nicely with the optimized ratio of 6.75:1:3.75 of the three proteins in the reconstituted PURE system [21, 65].

Three out of four elongation factors are essential for the elongation cycle during translation. The GTP bound EF-Tu and GTP exchange factor EF-Ts mediate the entry of the loaded tRNA into the A-site of the ribosome. This process is catalyzed by GTP hydrolysis. The elongation factor EF-G then promotes the translocation of tRNA from the A to the P site and is also catalyzed by GTP hydrolysis. During exponential growth the elongation factor EF-Tu is the most abundant protein reaching an EF-Tu:ribosome ratio of 8:1 [66]. Consistent with this statement, the emPAI values for EF-Tu belongs to the group of most abundant proteins in the S30 lysate. Unfortunately, the emPAI values of the ribosomal proteins in our analysis did not allow to estimate the ratio of EF-Tu:ribosome, as the emPAI values for the different ribosomal subunits were strongly scattered. Notably, the scattering of the most abundant ribosomal proteins has been reported previously and is probably due to saturation effects and artefacts caused by small ribosomal proteins [48]. Therefore, we omitted all emPAI values of ribosomal proteins from the analysis. It has further been shown that there is approximately 1 molecule of EF-G and 0.2 molecules of EF-Ts per ribosome. This ratio appears not to be preserved in the S30 lysate based on the approximate estimation of the averaged emPAI values. The concentration of EF-Ts seems to be even slightly higher when compared with that of EF-G. Accordingly, supplying the purified elongation factors at the above mentioned ratios into cell-free expression reactions increases elongation rate and product yield, as well as pre-expression of EF-Tu proved to significantly enhance yield and specific activity of β-lactamase [67, 68]. Our data further support the finding that the elongation factors may not be present in optimal ratios in S30 lysates. The fourth elongation factor EF-P is required to alleviate ribosome stalling at three or more consecutive proline residues and stimulates peptidyl transferase activity and is involved in synthesis of the initial peptide bond [69]. This factor appears to be less abundant compared to others. Interestingly, the pre-expression of EF-P prior to synthesis of the target protein β-lactamase was identified as a positive effector, thus supporting the notion of a potential shortage of this factor in S30 lysate extracts [68]. The additional elongation factors LepA and YeiP, postulated paralogues of EF-G and EF-P, are assumed not to be essential and might increase translation fidelity.

The release factors RF2 and RF3 and the ribosome recycling factor are present at comparable levels. Release factor RF-1 recognizing the amber stop codon appears to be less abundant. This finding could explain the drastically increased incorporation of non-natural amino acids *in vitro* using cell-free expression compared to *in vivo* systems [70] and is interesting for site-specific labelling approaches using the amber stop suppression strategy [71].

**Proteins Relevant for Protein Folding and for Precursor/Product Stability**

Folding and overall quality of synthesized proteins is of pivotal importance for applications of cell-free expression systems. Chaperone systems assisting protein folding and supporting the stability and lifetime of proteins involved in the expression machinery are often indispensable. Based on number as well as on their emPAI values, the cytoplasmic chaperones appear generally more abundant in the S30 lysate if compared to periplasmic chaperones (**Table 4**). Reassuringly, the major cytoplasmic chaperone systems such as trigger factor (TF), DnaK/DnaJ and GroEL/GroES as the major determinants of nascent protein folding are present in the S30 lysate. In particular, TF, DnaK and GroEL/GroES were identified with high emPAI values suggesting high abundance. The co-chaperone DnaJ seems to be less abundant. In particular DnaJ was only identified with very low emPAI values in lysates S302 and S303, indicating a potential shortage of this important co-chaperone. In addition, several types of peptidyl cis-trans-isomerases both from cytoplasmic as well as from periplasmic origin have been detected.

A challenging task in recombinant protein production is the synthesis of disulfide-bonded proteins. The cytoplasmic environment in *E. coli* is highly reducing and has not evolved to support the formation of disulfide bonds. In contrast, the periplasmic space maintains an oxidizing environment and contains specialized enzymes such as DsbA, DsbB and DsbC that catalyze the formation of native disulfide bonds. DsbA and DsbC are present in the S30 lysate with only relatively low averaged emPAI values of ~1 and ~0.6 (**Table S1**), respectively, while the membrane embedded DsbB was not identified. The oxidoreductase DsbA and its regenerating counterpart DsbB are usually mimicked in cell-free reactions by chemical redox systems such as GSH/GSSG. The disulfide isomerase DsbC is essential for disulfide bond shuffling and needs to be supplemented *e.g.* during cell-free synthesis of antibodies [72, 73]. Our analysis confirms the previous notion of DsbC being a limiting factor in S30 lysates.

A major cost factor for cell-free synthetic biology is the stability of supplied precursors such as NTPs, amino acids or the energy source phosphoenolpyruvate. It is evident that non-productive energy consuming enzymes from residual metabolic pathways or pathway fragments are abundant in the S30 lysate (**Table S1**). Prominent are precursor consuming enzymes from the glycolysis pathway (63% complete) or from fatty acid biosynthesis (92% complete) (**Figure 4**). Particularly problematic are precursor degrading or modifying enzymes as they cause significant problems upon protein labelling approaches. Numerous amino acid modifying enzymes, such as TnaA (tryptophanase), AspC (aspartate aminotransferase) or GlnA (glutamine synthetase), remain in S30 lysates (**Table 5**), resulting in significant consumption and destabilization of amino acids or even causing severe label scrambling in NMR applications [11]. Also constantly identified in all S30 lysates were glycerol kinase (glpK), arginine decarboxylase (speA) and malate dehydrogenase (mdh) with emPAI values of 0.2±0.1, 3.2±0.7 and 1.3±0.3, respectively. Even though the emPAI values appear relatively low, these enzymes resulted in a significant reduction of yield when pre-expressed prior to cell-free expression reaction at concentration of 2.9 µM (glpK), 1.4 µM (speA) and 1.6 µM (mdh) [68]. Notably speA, identified with the highest emPAI of 3.2, has been shown previously to degrade arginine at concentrations as high as 2.5 mM within 30 min and the knockout of speA in the source strain completely eliminated this activity, thus stabilizing the amino acid in cell-free protein synthesis reactions [74]. Specific elimination or inhibition of these enzymes will become a future approach of S30 lysate tuning in order to generate labeled proteins more suitable for subsequent NMR studies or to generally improve yields in cell-free protein synthesis.

Integrity of the synthesized target protein is of fundamental importance for basically all applications. Numerous proteases belong to the essential standard repertoire in any cell, while in cell-free expression systems they are rather problematic enzymes. A comprehensive set of proteases remains within the S30 lysate, although mostly showing relatively low emPAI values (**Table 6**). Most abundant is the non-specific protease Lon. The relative high protease content of S30 lysates is in accordance with the required supply of protease inhibitors into cell-free expression reactions in continuous exchange format in order to maintain product integrity [20, 75].

**S30 Lysate Tuning for Chaperone Enrichment**

The physiology and thus the proteome of *E. coli* cells changes rapidly in response to external stimuli, thus providing numerous options to tune the composition of resulting S30 lysates for particular applications [76]. Approximately 10% of all proteins contain at least one chaperone-dependent folding step [77]. Furthermore, certain protein folding pathways even require the synergistic interplay of different chaperones. Depletion or insufficient copy numbers of individual chaperones in S30 lysates could result in the formation of improperly folded or unfolded and aggregated proteins concomitant with low specific activity or even precipitation. Some endogenous chaperones are present at relatively high concentrations in the standard S30 lysate (**Table 4**), but concentrations of others might become rapidly limiting upon preparative scale expression of chaperone-dependent target proteins. Exposure to elevated temperature or to chemicals are well known inducers of so-called heat shock responses in *E. coli*, resulting into the increased production of chaperones and other rescue or SOS-response proteins [25, 78]. To demonstrate an example of S30 lysate tuning, we exposed the *E. coli* A19 source cells during fermentation to a heat shock at 42°C for 45 min in combination with the addition of 3% (v/v) ethanol (**Figure 1A**).

The proteome of tuned S30-S lysates was analyzed relative to standard S30 lysates using the ICPL labelling strategy. Three independent S30-S lysates were analyzed and in total approximately 400 proteins were quantified in each replicate (**Figure 5**). From this protein group, a core of 64.8% was found in all three, 31.7% in at least two and only a minor fraction of 3.5% in only one biological replicate (**Figure 5B**). By quantitative evaluation relative to the standard S30 lysate, a total of 27 proteins was up- and another 57 proteins were down-regulated (**Table S2, S3**) (**Figure 5A**). As predicted, most regulated proteins belong to the general stress response induced by heat shock or ethanol stress [23, 25, 79]. The 27 upregulated proteins are mainly chaperones or factors related to transcription/translation. In particular, GroEL, GroES, DnaK, GrpE and SecA are 3-10-fold increased in S30-S lysates. We also detected the chaperone HslR acting as recycling factor for ribosomal subunits and the periplasmic chaperone Spy exclusively in S30-S lysates. Some detected upregulated proteins such as RpoH and ManX match those previously found after growth of *E. coli* under ethanol stress [79]. However, the alcohol and aldehyde dehydrogenases YqhG and AldB as well as other known ethanol induced proteins were not found or were not regulated. These proteins might therefore be 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 still 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 lysate preparation from *E. coli* cells grown continuously at 42oC for chaperone enrichment has been reported [80]. However, in our hands 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.

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 degradation of protein aggregates and in addition supporting chaperone activity by complex formation [81, 82]. Not surprisingly, the heat shock specific sigma factor 32 (RpoH) was also found in this group [83, 84] (**Figure 5C**). Interestingly, the transcriptional termination factor Rho is upregulated, potentially providing more efficient Rho-dependent termination activity. In contrast to the upregulated protein fraction, the group of 57 downregulated proteins is more diverse (**Figure 5A**), 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. Besides central metabolic proteins, connected pathways comprising 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 5D**). Here, proteins such as the down-regulated CysS (cysteine tRNA ligase) should be pointed out as one possible limiting factor for cysteine containing proteins **(Figure 5D**). This might contribute to the generally lower protein production efficiency of S30-S lysates. However, in view of amino acid scrambling, the reduced protein content of amino acid modifying enzymes such as GlyA could be beneficial [11]. Signal intensities of several peptides per protein are compared simultaneously in the same run when using the quanitative approach based on ICPL labelling. Thus, this quantitative method is less prone to technical bias due to *e.g.* physiochemical properties of individual peptides. Nevertheless, the possibility that not all regulated proteins are detected and quantified cannot be excluded with this approach and this limitation is inherent in all proteomics analyses.

The chaperone-tuned S30-S lysate was evaluated in cell-free expression reactions for the production of difficult proteins. The expression efficiency of the reporter protein sGFP was reduced by 30% compared with standard S30 lysate (**Figure 1C**). This was expected as the fermentation conditions for the standard S30 lysate were optimized for high expression yield. However, increased chaperone content in S30-S lysates could be more suitable for the solubility and folding of certain difficult and potentially chaperone-dependent target proteins. ApiRBP is a 15 kDa protein with a putative RNA-binding domain and located inside the apicoplast, an essential plastid organelle surrounded by 4 membranes. This organelle is crucial for the viability of *Plasmodium vivax* and offers an exciting line of attack for malarial infection. The protein is synthesized in high amounts (1-1.5 mg/mL) in cell-free systems using standard S30 lysates. However, it remains mostly unfolded and the majority of the synthesized ApiRBP precipitates, leaving only about 100 µg/mL in the supernatant. Synthesis of ApiRBP in S30-S lysates increased the yield of total soluble protein by 30% to 140 µg/mL. Considering the ratio of soluble to insoluble protein, the fraction of soluble ApiRBP was even increased 3-fold from 7% in standard S30 lysates to 22% in S30-S lysates (**Figure 6A**).

As a second example, we analyzed the human enzyme GNA-1, which is 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 6B**) [85]. It is expressed at approximately 0.5 mg/mL with standard S30 lysate and with approximately 70-80% solubility [39]. The yield of hGNA-1 in S30-S lysates was approximately half of that in standard S30 lysates and its solubility was apparently unaffected. However, the specific enzymatic activity of hGNA-1 was significantly increased in S30-S lysates, suggesting that a larger fraction of hGNA-1 was natively folded (**Figure 6C**). This implies that at least 50% of hGNA-1 synthesized in standard S30 lysate, despite being soluble, was actually inactive. The two examples show that chaperone-tuned S30-S lysate can provide an interesting choice to improve product quality of particularly challenging proteins. 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 [86, 87]. However, additional workload for the production and purification of the chaperones would be necessary and the effect of one or a few chaperones might be different from that of a complex chaperone network. Furthermore, selected chaperones could be overexpressed in the cells prior to lysate preparation, resulting into increased solubility or activity of target proteins [88].

**Conclusion**

Our study shows that at least approximately 25% of the predicted or 40% of the expressed *E. coli* proteome remains in standard S30 lysates prepared for cell-free protein expression. The S30 lysate proteins have been classified and numerous data of enriched or deficient proteins match with published observations on cell-free expression studies. The S30 lysate proteome composition exhibits significant variations and may be divided into a constant core of approximately 60% of the proteins, some 16% with low and the residual 24% with high variability. The variation is most likely caused by certain precipitation steps during S30 lysate preparation, but also potentially by sensitivity limitations in detection of low abundant proteins. The proteomes of other lysate types occasionally used in cell-free expression such as S12 or S15 lysates [22] might therefore differ from that of the S30 lysate. The presented data help to move conventional cell-free expression based on S30 lysates towards better defined systems. The study further serves as a platform for directed lysate optimization and tailored protein production. The presented lysate proteome can be used as an initial checklist to identify or to rule out problems that could possibly occur during the production and analysis of a given target protein. Problematic proteins could be evaluated and after positive verification selectively removed from lysates prior to expression studies. *Vice versa*, potentially limiting components could be supplied and enriched in optimal concentrations. In previous studies, the removal of release factor-1 (RF1) from S30 lysates considerably increased the efficiency of non-natural amino acid incorporation into proteins by the amber stop codon technology [13, 71]. Alternatively, lysates from engineered strains containing selected deficiencies have been prepared [14, 89, 90]. Identified proteins causing difficulties for the structural analysis of labelled proteins could furthermore be addressed by adding selective inhibitors as exemplified for amino acid scrambling enzymes [11, 91]. A potential tool for the elimination of non-desired proteins could be the modification with degradation tags such as the ssrA tag [92]. Elimination will then be promoted by the presence of specific proteases like ClpX and ClpP in the S30 lysate (**Table S1**). The modular portfolio of cell-free synthetic biology could further be expanded by complementing fragmented biosynthetic pathways *e.g.* for the synthesis of stable isotope labeled amino acids from cheap precursors for the production of NMR-suitable labeled proteins. However, as an intrinsic limitation of proteomics studies, it should be considered that the proteome analysis provides no information on the quality of the identified proteins.

**Acknowledgements:**

The authors thank Debora Teixeira Duarte for helpful discussions regarding this manuscript. This work was funded by the Collaborative Research Centre (SFB) 807 of the German Research Foundation (DFG) and by the German Ministry of Education and Science (BMBF). The software tool InfernoRDN from Pacific Northwest National Laboratory (OMICS.PNNL.GOV) is gratefully acknowledged. Further support was provided by the Andalusian Government (P11-CVI-7216 and BIO198) and the Spanish Ministry of Science and Innovation (BFU2015-71017-P).

**Author contributions:**

DF, EH, FB, SR, ABK contributed to conception of experimental design. EH and FB prepared the S30 lysates. EH performed protein production and biochemical analysis of expressed proteins. DF, EkH, MS performed experiments and did the data acquisition of proteomics studies. DF, EH, FB did the data analysis and interpretation. IDM and SMGM provided the apiRBP construct. EH, DF and FB wrote the manuscript. SR IDM, SMGM and CK revised the manuscript.

**Figure Legends**

**Figure 1. Preparation and efficiency of S30 lysates.** Workflow of lysate preparation, culture conditions and performance of the analyzed lysates. (A) Flowchart of S30 lysate preparation procedures. (B) Growth curves of *E. coli* A19 as observed under different fermentation conditions for S30 and S30-S lysate. (C) Performance of S30 and S30-S lysates as determined under standard expression conditions using sGFP as a reporter. The error bars represent the standard deviation of multiple measurements (n=3).

**Figure 2. Strategies for S30 lysate analysis by 2D-LC-MS analysis.** The general workflow was as follows: (*i*) S30 lysates were separated by SDS-PAGE, (*ii*) the gel slice with the separated proteins was divided into 12 fractions (F1-F12), (*iii*) the proteins were in gel digested (trypsin) and then extracted, (*iv*) the peptides of each fraction (F1-12) were separated by nanoHPLC (Reversed Phase C18), (*v*) the elution fractions were spotted and analyzed by mass spectrometry. (A) Qualitative proteome analysis was carried out for S30 lysates. (B) Quantitative proteome analysis for S30-S lysate relative to S30 was performed using ICPL labelling strategy. One sample is labelled with light ICPL tags, the other with heavy ICPL tags and the mixture of both is analyzed using the general workflow as described previously.

**Figure 3. Analysis of the S30 lysateproteome.** (A) Number of common and unique protein identifiers in S301-4 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 different replicates. (B) Combined number of proteins identified in S30 lysate 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≤0.05) as compared to a reference set of predicted proteins (Uniprot ID: UP000000625, 4315 proteins).

**Figure 4. Over- and underrepresented pathways in S30 lysate (821 proteins).** Pathways are plotted against the percentage of identified gene products using the Cytoscape plugin ClueGO. Overrepresented and underrepresented pathways (p≤0.05), tested against KEGG reference set, are shown in grey and white, respectively. The number of proteins identified per pathway are indicated.

**Figure 5. Proteome comparison of S301 vs. S30-S lysate.** (A) Volcano plot of quantitative proteome data from S301 vs. S30-S. 57 proteins were found to be downregulated (left, red) and 27 proteins were upregulated (right, blue) in S30-S relative to S30 lysates. (B) Bar diagram illustrating reproducibility of three biological replicates (S30-S). The majority of proteins (>96%) were found in at least two biological replicates. Only 3.5% were uniquely detected in one biological replicate. Numbers besides the bars represent the percentage and the number of identified proteins. (C) and (D) show the networks (generated by Cytoscape in combination with ClueGO). Functionally grouped network with terms as nodes linked based on their kappa score level (≥0.4) using GO reference sets Biological Process and KEGG for up- and downregulated proteins, respectively. The node size represents the significance (p≤0.05) and filled node the share of detected proteins per GO term.

**Figure 6. Production of challenging proteins in chaperone enriched S30-S lysate.** (A) Expression of ApiRBP-sfGFP. Left: Soluble ApiRBP-sfGFP in reaction supernatants as determined by sfGFP as reporter; right: Total solubility of synthesized ApiRBP-sfGFP in S30-S and standard S30-1 lysates. (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-sfGFP. Left: Soluble hGNA-1-sfGFP in reaction supernatants as determined by sfGFP as a reporter; middle: Total solubility of synthesized hGNA-1-sfGFP in S30-S and standard S30-1 lysates; right: Normalized enzymatic activity of hGNA-1 synthesized either in standard S30-1 or in S30-S lysate. Values of the S30-S sample are set to 100%.

**References**

[1] Busso D, Kim R, Kim SH. Using an Escherichia coli cell-free extract to screen for soluble expression of recombinant proteins. J Struct Funct Genomics 2004;5(1-2):69-74.

[2] Katzen F, Chang G, Kudlicki W. The past, present and future of cell-free protein synthesis. Trends Biotechnol 2005;23(3):150-6.

[3] Rosenblum G, Cooperman BS. Engine out of the chassis: cell-free protein synthesis and its uses. FEBS Lett 2014;588(2):261-8.

[4] Henrich E, Hein C, Dötsch V, Bernhard F. Membrane protein production in Escherichia coli cell-free lysates. FEBS Lett 2015;589(15):1713-22.

[5] Salehi AS, Earl CC, Muhlestein C, Bundy BC. Escherichia coli-based Cell-free Extract Development for Protein-based Cancer Therapeutic Production. Int J Dev Biol 2016;60:237-43.

[6] Matsuda T, Kigawa T, Koshiba S, Inoue M, Aoki M, Yamasaki K, et al. Cell-free synthesis of zinc-binding proteins. J Struct Funct Genomics 2006;7(2):93-100.

[7] Hein C, Henrich E, Orbán E, Dötsch V, Bernhard F. Hydrophobic supplements in cell‐free systems: Designing artificial environments for membrane proteins. Eng Life Sci 2014;14(4):365-79.

[8] Henrich E, Ma Y, Engels I, Münch D, Otten C, Schneider T, 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):2535-46.

[9] Rues RB, Dötsch V, Bernhard F. Co-translational formation and pharmacological characterization of beta1-adrenergic receptor/nanodisc complexes with different lipid environments. Biochim Biophys Acta 2016;1858(6):1306-16.

[10] Casteleijn MG, Urtti A, Sarkhel S. Expression without boundaries: cell-free protein synthesis in pharmaceutical research. Int J Pharm 2013;440(1):39-47.

[11] LaGuerre A, Löhr F, Bernhard F, Dötsch V. Labeling of Membrane Proteins by Cell-Free Expression. Methods Enzymol 2015;565:367-88.

[12] Löhr F, Tumulka F, Bock C, Abele R, Dötsch V. An extended combinatorial 15N, 13Cα, and 13C' labeling approach to protein backbone resonance assignment. J Biomol NMR 2015;62(3):263-79.

[13] Peuker S, Andersson H, Gustavsson E, Maiti KS, Kania R, Karim A, et al. Efficient Isotope Editing of Proteins for Site-Directed Vibrational Spectroscopy. J Am Chem Soc 2016;138(7):2312-8.

[14] Hong SH, Ntai I, Haimovich AD, Kelleher NL, Isaacs FJ, Jewett MC. Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth Biol 2014;3(6):398-409.

[15] Proverbio D, Henrich E, Orbán E, Dötsch V, Bernhard F. Membrane Protein Quality Control in Cell-Free Expression Systems: Tools, Strategies and Case Studies. In: Mus-Veteau I, editor. Membrane Proteins Production for Structural Analysis. New York, NY: Springer New York; 2014. p. 45-70.

[16] Salehi AS, Smith MT, Bennett AM, Williams JB, Pitt WG, Bundy BC. Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. Biotechnol J 2016;11(2):274-81.

[17] Zubay G. In vitro synthesis of protein in microbial systems. Annu Rev Genet 1973;7:267-87.

[18] Kigawa T, Yabuki T, Matsuda N, Matsuda T, Nakajima R, Tanaka A, et al. Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. J Struct Funct Genomics 2004;5(1-2):63-8.

[19] Yang WC, Patel KG, Wong HE, Swartz JR. Simplifying and streamlining Escherichia coli-based cell-free protein synthesis. Biotechnol Prog 2012;28(2):413-20.

[20] Schwarz D, Junge F, Durst F, Frölich N, Schneider B, Reckel S, et al. Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. Nat Protoc 2007;2(11):2945-57.

[21] Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, et al. Cell-free translation reconstituted with purified components. Nat Biotechnol 2001;19(8):751-5.

[22] Kwon YC, Jewett MC. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Sci Rep 2015;5:8663.

[23] Soufi B, Krug K, Harst A, Macek B. Characterization of the E. coli proteome and its modifications during growth and ethanol stress. Front Microbiol 2015;6:103.

[24] Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, et al. The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol 2016;34(1):104-10.

[25] Arsène F, Tomoyasu T, Bukau B. The heat shock response of Escherichia coli. Int J Food Microbiol 2000;55(1-3):3-9.

[26] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422(6928):198-207.

[27] Aebersold R. Quantitative proteome analysis: methods and applications. J Infect Dis 2003;187 Suppl 2:315-20.

[28] Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. Genome Res 2002;12(8):1231-45.

[29] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, 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):1265-72.

[30] Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, et al. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 2005;122(2):209-20.

[31] Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Mol Syst Biol 2011;7:549.

[32] Li C, Tan XF, Lim TK, Lin Q, Gong Z. Comprehensive and quantitative proteomic analyses of zebrafish plasma reveals conserved protein profiles between genders and between zebrafish and human. Sci Rep 2016;6:24329.

[33] Schmidt A, Kellermann J, Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics 2005;5(1):4-15.

[34] Gesteland RF. Isolation and characterization of ribonuclease I mutants of Escherichia coli. J Mol Biol 1966;16(1):67-84.

[35] Henrich E, Dotsch V, Bernhard F. Screening for lipid requirements of membrane proteins by combining cell-free expression with nanodiscs. Methods Enzymol 2015;556:351-69.

[36] Reckel S, Sobhanifar S, Durst F, Lohr F, Shirokov VA, Dötsch V, et al. Strategies for the cell-free expression of membrane proteins. Methods Mol Biol 2010;607:187-212.

[37] Roos C, Zocher M, Müller D, Münch D, Schneider T, Sahl HG, 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):3098-106.

[38] Schwarz D, Daley D, Beckhaus T, Dötsch V, Bernhard F. Cell-free expression profiling of E. coli inner membrane proteins. Proteomics 2010;10(9):1762-79.

[39] Ma Y, Ghoshdastider U, Wang J, Ye W, Dötsch V, Filipek S, et al. Cell-free expression of human glucosamine 6-phosphate N-acetyltransferase (HsGNA1) for inhibitor screening. Protein Expr Purif 2012;86(2):120-6.

[40] Kai L, Dötsch V, Kaldenhoff R, Bernhard F. Artificial environments for the co-translational stabilization of cell-free expressed proteins. PLoS One 2013;8(2):e56637.

[41] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 1996;68(5):850-8.

[42] Chen D, Shah A, Nguyen H, Loo D, Inder KL, Hill MM. Online quantitative proteomics p-value calculator for permutation-based statistical testing of peptide ratios. J Proteome Res 2014;13(9):4184-91.

[43] Nguyen HD, Wood IA, Hill MM. A robust permutation test for quantitative SILAC proteomics experiments. Journal of Integrated OMICS 2012;2(2):80-93.

[44] Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 2016;44(D1):D447-56.

[45] Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science 1997;277(5331):1453-62.

[46] Besemer J, Lomsadze A, Borodovsky M. 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):2607-18.

[47] Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. 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):2616-20.

[48] Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, et al. Protein abundance profiling of the Escherichia coli cytosol. BMC Genomics 2008;9:102.

[49] Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 2009;25(22):3045-6.

[50] Boland C, Li D, Shah ST, Haberstock S, Dötsch V, Bernhard F, et al. Cell-free expression and in meso crystallisation of an integral membrane kinase for structure determination. Cell Mol Life Sci 2014;71(24):4895-910.

[51] Jewett MC, Swartz JR. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng 2004;86(1):19-26.

[52] Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR. An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 2008;4:220.

[53] Matsubayashi H, Kuruma Y, Ueda T. In vitro synthesis of the E. coli Sec translocon from DNA. Angew Chem Int Ed Engl 2014;53(29):7535-8.

[54] Frauenfeld J, Gumbart J, Sluis EO, Funes S, Gartmann M, Beatrix B, et al. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nat Struct Mol Biol 2011;18(5):614-21.

[55] Ulmschneider MB, Ulmschneider JP, Schiller N, Wallace BA, von Heijne G, White SH. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. Nat Commun 2014;5:4863.

[56] Cymer F, von Heijne G, White SH. Mechanisms of integral membrane protein insertion and folding. J Mol Biol 2015;427(5):999-1022.

[57] Berrier C, Guilvout I, Bayan N, Park KH, Mesneau A, Chami M, 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. Biochim Biophys Acta 2011;1808(1):41-6.

[58] Wissenbach U, Six S, Bongaerts J, Ternes D, Steinwachs S, Unden G. 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):675-86.

[59] Gul N, Poolman B. Functional reconstitution and osmoregulatory properties of the ProU ABC transporter from Escherichia coli. Mol Membr Biol 2013;30(2):138-48.

[60] Malinverni JC, Silhavy TJ. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. Proc Natl Acad Sci U S A 2009;106(19):8009-14.

[61] Schirle M, Heurtier MA, Kuster B. Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics 2003;2(12):1297-305.

[62] Matthies D, Haberstock S, Joos F, Dötsch V, Vonck J, Bernhard F, et al. Cell-free expression and assembly of ATP synthase. J Mol Biol 2011;413(3):593-603.

[63] Shin J, Jardine P, Noireaux V. Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth Biol 2012;1(9):408-13.

[64] Shin J, Noireaux V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. J Biol Eng 2010;4(1):8.

[65] Ueda T. The Constructive Approach for Cell-Free Translation. In: Spirin AS, Swartz JR, editors. Cell-Free Protein Synthesis. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2007. p. 35-50.

[66] Defeu Soufo HJ, Reimold C, Linne U, Knust T, Gescher J, Graumann PL. Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein. Proc Natl Acad Sci U S A 2010;107(7):3163-8.

[67] Underwood KA, Swartz JR, Puglisi JD. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. Biotechnol Bioeng 2005;91(4):425-35.

[68] Woodrow KA, Swartz JR. A sequential expression system for high-throughput functional genomic analysis. Proteomics 2007;7(21):3870-9.

[69] Aoki H, Dekany K, Adams SL, Ganoza MC. The gene encoding the elongation factor P protein is essential for viability and is required for protein synthesis. J Biol Chem 1997;272(51):32254-9.

[70] Bundy BC, Swartz JR. Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. Bioconjug Chem 2010;21(2):255-63.

[71] Loscha KV, Herlt AJ, Qi R, Huber T, Ozawa K, Otting G. Multiple-site labeling of proteins with unnatural amino acids. Angew Chem Int Ed Engl 2012;51(9):2243-6.

[72] Goerke AR, Swartz JR. Development of cell-free protein synthesis platforms for disulfide bonded proteins. Biotechnol Bioeng 2008;99(2):351-67.

[73 ]Knapp KG, Goerke AR, Swartz JR. Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source. Biotechnol Bioeng 2007;97(4):901-8.

[74] Calhoun KA, Swartz JR. Total amino acid stabilization during cell-free protein synthesis reactions. J Biotechnol 2006;123(2):193-203.

[75] Ali M, Suzuki H, Fukuba T, Jiang X, Nakano H, Yamane T. Improvements in the cell-free production of functional antibodies using cell extract from protease-deficient Escherichia coli mutant. J Biosci Bioeng 2005;99(2):181-6.

[76] Smith MT, Hawes AK, Shrestha P, Rainsdon JM, Wu JW, Bundy BC. Alternative fermentation conditions for improved Escherichia coli-based cell-free protein synthesis for proteins requiring supplemental components for proper synthesis. Process Biochemistry 2014;49(2):217 - 22.

[77] Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002;295(5561):1852-8.

[78] Lüders S, Fallet C, Franco-Lara E. Proteome analysis of the Escherichia coli heat shock response under steady-state conditions. Proteome Sci 2009;7:36.

[79] Chaudhuri S, Jana B, Basu T. Why does ethanol induce cellular heat-shock response? Cell Biol Toxicol 2006;22(1):29-37.

[80] Yamane T, Ikeda Y, Nagasaka T, Nakano H. 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):608-13.

[81] Wickner S, Maurizi MR, Gottesman S. Posttranslational quality control: folding, refolding, and degrading proteins. Science 1999;286(5446):1888-93.

[82] Lee I, Suzuki CK. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. Biochim Biophys Acta 2008;1784(5):727-35.

[83] Guisbert E, Herman C, Lu CZ, Gross CA. A chaperone network controls the heat shock response in E. coli. Genes Dev 2004;18(22):2812-21.

[84] Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. 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):1776-89.

[85] Selitrennikoff CP, Sonneborn DR. 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):408-16.

[86] Niwa T, Kanamori T, Ueda T, Taguchi H. Global analysis of chaperone effects using a reconstituted cell-free translation system. Proc Natl Acad Sci U S A 2012;109(23):8937-42.

[87] Ryabova LA, Desplancq D, Spirin AS, Plückthun A. Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. Nat Biotechnol 1997;15(1):79-84.

[88] Kang SH, Kim DM, Kim HJ, Jun SY, Lee KY. Cell-free production of aggregation-prone proteins in soluble and active forms. Biotechnol Prog 2005;21(5):1412-9.

[89] Hong SH, Kwon YC, Martin RW, Des Soye BJ, de Paz AM, Swonger KN, et al. Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1. Chembiochem 2015;16(5):844-53.

[90] Kim DM, Swartz JR. Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. Biotechnol Bioeng 2004;85(2):122-9.

[91] Su XC, Loh CT, Qi R, Otting G. 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. J Biomol NMR 2011;50(1):35-42.

[92] Shin J, Noireaux V. Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. J Biol Eng 2010;4:9.