Protein interactions in minimal cells

Dissertation

for the award of the degree

"*Doctor rerum naturalium*"

of the Georg-August-University Göttingen

within the doctoral program

"**International Max Planck Research School for Genome Science**"

of the Georg-August University School of Science (GAUSS)

submitted by

Neil Singh

from Pune

Göttingen 2022

Thesis committee

Pröf Dr. Jörg Stülke (Supervisor and 1st Reviewer)
Institute for Microbiology and Genetics, Department of General Microbiology

Prof. Dr. Henning Urlaub (2nd Reviewer)
Max Planck Institute for Biophysical Chemistry, Bioanalytical Mass Spectrometry Group

Prof. Dr. Carsten Lüder
University Medical Centre Göttingen, Department for Medical Microbiology

Further members of examination board

Prof. Dr. Ivo Feußner
Albrecht von Haller Institute, Department of Plant Biochemistry, University of Göttingen

Prof. Dr. Rolf Daniel
Institute for Microbiology and Genetics, Department of Genomic and Applied Microbiology and G2L

Prof. Dr. Stefan Klumpp
Institute for Nonlinear Dynamics, Theoretical Biophysics Group

Date of oral examination: 21st April, 2022

Declaration of independent work

I hereby declare that this doctoral thesis titled, “Protein interactions in minimal cells” has been written independently and with no other sources or aids than those quoted.

Neil Singh

Acknowledgments

First and foremost, I would like to thank Prof. Dr. Jörg Stülke for providing me with the opportunity to carry out my PhD in his lab, under his supervision. Your open-door policy always meant that I was free to seek your inputs whenever I felt the need for them. Your guidance and re-assuring presence have taught me much about handling complex, stressful situations. Lastly, to be able to learn from your scientific knowledge and experience is something I will always consider a blessing.

I would also like to thank Prof. Dr. Henning Urlaub for your constructive feedback right from the first thesis committee meeting. Your support with the experimental setup and the follow-up collaborations is greatly appreciated.

Further, I would like to thank Prof. Dr. Carsten Lüder for your inputs, especially during the research on *M. pneumoniae* regarding the characterization of IbpM. It has been a pleasure to have you on board my thesis committee.

Finally, I would like to thank Prof. Dr. Ivo Feußner, Prof. Dr. Rolf Daniel and Prof. Dr. Stefan Klumpp for taking the time out of their busy schedules to be members of the examination committee.

My work in the laboratory was supported and enhanced by several people. I am extremely grateful to Silvia Carrillo-Castellón for her vital work in the laboratory and her cheerful nature. I would like to thank Prof. Dr. Fabian Commichau and PD Dr. Jeanine Rismondo for several scientific discussions and ideas. My work was enhanced by the pleasure I derived from teaching and supervising my students Alina Lüschen, Tayfun Kilic, Annabel Maisl and Camilo Torres. I consider it good feedback that we could work together for both your laboratory rotations and master theses, in the latter two cases. Each of your research inputs have developed these projects further and I have learnt with you. I am also indebted to PD Dr. Aleksander Chernev and Ms. Luisa Welp for their collaborations with the RNA-binding studies. I would like to especially thank PD Dr. Adam Belsom and Dr. Ana Perez for their collaboration with the protein interactome study. Lastly, a special thanks to Prof. Dr. Juri Rappsilber for the opportunity to learn more about crosslinking from your research group in Berlin.

I would like to thank Julia Busse whose experience I have drawn on for many projects. As my constant office companion, I am grateful for all our exchanges and your tolerance with my developing German-language skills. A special thanks to my predecessor, Dr. Cedric Blötz, I have learnt much from you. A special thanks also to my labmates Johannes Gibhardt, Lisa Schulz, Rica Bremenkamp and Janek Meißner for a cohesive and fun working environment, Tiago Pedreira for frank discussions, and Dennis Wicke for always

dropping by. Along with Robert Warneke for planning many lab events and Christoph Elfmann, I would finally like to thank many former and current members of the lab for a pleasant working environment. Please forgive me if I have left anyone out, it was unintentional.

Words cannot express the gratitude I have for my parents and my sister without whom I cannot imagine having come so far. That I could count on your unwavering support from two different continents, irrespective of the time difference, has always been integral to my success. Thank you. Finally, I would like to thank Inga whose support has been crucial, especially during this writing phase. A final shout out to all my friends here who are my family in this city of science, thank you.

Table of contents

Publications

Blötz C, **Singh N**, Dumke R, Stülke J. Characterization of an Immunoglobulin Binding Protein (IbpM) From *Mycoplasma pneumoniae*. Front Microbiol. (2020) **11**: 685.

O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, **Singh N**, Hagen WJH, Cramer P, Stülke J, Mahamid J, Rappsilber J. In-cell architecture of an actively transcribing-translating expressome. Science. (2020) **369**: 554-557.

Pedreira T, Elfmann C, **Singh N**, Stülke J. *Syn*Wiki: Functional annotation of the first artificial organism *Mycoplasma mycoides* JCVI-syn3A. Protein Sci. (2022) **31**: 54-62.

Abbreviations

% (v/v)	% (volume/volume) (volume percent)
% (w/v)	% (weight/volume) (mass percent)
Amp	Ampicillin
AP	Alkaline phosphatase
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CE	Crude extract
Cm	Chloramphenicol
dH_2O	Deionised water
DMEM	Dulbecco's Modified Eagle Medium
DNA	Desoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleoside triphosphate
et al.	Et altera
FBS	Foetal bovine serum
Fig.	Figure
FT	Flow through
fwd	Forward
G3P	Glycerol-3-phosphate
gDNA	Genomic DNA
Glc	Glucose
Gly	Glycerol
goi	gene of interest
HEPES	4-(2-Hydroxyethyl)-piperazin-1-ethan-sulfonic acid
IPTG	Isopropyl-1-thio-β-D-galactoside
Kan	Kanamycin
kb	Kilo base pairs
LB	Luria Bertani (Medium)
LFH	long flanking homology
MPN	*Mycoplasma pneumoniae*
mRNA	Messenger RNA
Ni^{2+}-NTA	nickel-nitrilotriacetic acid
O/N	Over night
OD_x	Optical density, measured at wavelenght λ = x nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pH	Power of hydrogen
PPLO	Pleuropneumoniae like organisms
Puro	Puromycin
rev	Reverse
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Rounds per minute
RT	Room temperature

S3A	*Syn3A*
SDS	Sodium dodecyl sulfate
TAE	Tris-acetic acid-EDTA
TEMED	N,N,N´,N´-tetramethylethylendiamine
Tn	Transposon
Tris	Tris-(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
U	Units
w/o	Without
WT	Wild type
X-Gal	5-bromo-4-chloro-indolyl-galactopyranoside

Units

°C	Degree Celsius
A	Ampere
bar	Bar
Da	Dalton
F	Farada y
g	Gram
h	Hour
l	Litre
m	Meter
min	Minute
mol	Mol
M	Molar
Pa	Pascal
sec	Second
V	Volt
Ω	Ohm

Prefixes

k	Kilo
m	Milli
µ	Mikro
n	Nano
p	Pico

Nucleosides

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

Summary

Minimal cells are used to model and understand life's complex processes. The most essential processes of life are thought to include cell replication, energy metabolism and genome maintenance. However, construction of a synthetic genome that entailed these core processes failed to yield a viable cell. This represents a gap in our knowledge regarding the functions required for life to unfold. To determine all the core processes required for life, naturally occurring minimal organisms that have undergone degenerative evolution, like *Mycoplasma pneumoniae*, are studied. Electron microscopy has shown that RNA polymerase and the ribosome form a super complex in some organisms. However, this coupling mechanism was not clear. In *M.* pneumoniae, NusA, an essential transcription factor with a disordered C-terminal region that contained crosslinks to the ribosome was investigated for this function. Truncated mutants of NusA that lacked this region were unable to grow, thus confirming the essentiality of this domain. NusA, rather than the proposed NusG, was thus proven to link the transcription and translation processes. Another approach to increase our understanding of life's core processes is to consider organisms with already small genomes such as *M. mycoides* which have been synthetically minimized to possess only genes essential for growth in rich media. The resultant strain *Syn3A*, retains only half the original genome and contains 438 proteins. Given the essentiality criteria, it was a surprise that 149 genes had no specific function and a further subset of 79 genes were completely unannotated. As gene deletions are challenging in such a minimal organism, crosslinking experiments to elucidate protein-protein and protein-RNA interactions were performed. Along with confirming several known protein interactions, 28 uncharacterized proteins were found to interact with other proteins and 80 uncharacterized proteins have self-links. Four complexes containing uncharacterized proteins have been identified for further study and efforts to characterize 3 unknown proteins, Syn3A_0439, 0440 and 0505, were initiated. Self-links were also used to validate a predicted structure of 0439. Further, the complete subset of RNA-interacting protein machinery was thought to be retained in Syn3A. A second crosslinking project found 161 RNA-binding proteins that included 122 known RNA-binding proteins, 19 previously annotated proteins with a new RNA-binding function and 20 uncharacterized proteins. 4 of these 19 proteins and 6 of the 20 uncharacterized proteins were confirmed for RNA-binding ability. All the new proteins hold the possibility of undiscovered RNA-binding motifs, mechanisms and regulatory events. Specifically, as Syn3A_0317, 0388, 0439 and 0451 are all uncharacterized RNA-binding proteins that interact with other proteins, this work provides a direction for future investigations to follow.

1 INTRODUCTION

Genomic DNA stores genes which are transcribed into RNA, that ribosomes translate into amino acid sequences which fold into proteins to perform functions within a membrane-enclosed space called a cell. Somewhere along this process 'life' as we currently recognize it, comes into existence. While life, through evolution and adaptation, expresses itself in a multitude of forms, it is thought that some basic principles apply- a so called 'core set of functions'.

Efforts to search for these functions encoded in a 'core genome' began with the genomic analysis of the smallest bacterium (at the time), *Mycoplasma genitalium* (Fraser et al., 1995). While that study looked for genes that were thought to form the complete set of functions that enabled 'life'; most recently, attempts to chemically synthesize these required genes into a genome-less cell have not been completely successful (Hutchison et al., 2016). The unexpected but necessary inclusion of almost half as many genes of unknown function to the core set of genes thought to be required for life indicates that there are gaps in our knowledge.

While computational modelling of the natural minimal cell *Mycoplasma pneumoniae* (Wodke et al., 2013) and its artificial counterpart the synthetic bacterium *JCVI-Syn3A* (Wise et al., 2019) did help to match genes with some functions, significant gaps of knowledge still existed, especially outside of metabolic proteins.

The research depicted in this thesis sought to shine a light on these unknown yet essential activities, within complex protein machinery like the 'Expressome' in *M. pneumoniae* and within the entire proteome via whole cell crosslinking mass spectrometry (CLMS) in the synthetic cell. This chapter will first introduce *M. pneumoniae* and the expressome, then outline the specific issue under research. Subsequently, the second half of this introduction will deal with the synthetic cell, *Syn3A*, and in an attempt to increase the base level of information about this organism, some omics-level approaches will be described in brief.

1.1 A natural-minimal cell, *M. pneumoniae*

M. pneumoniae was first classified as 'Eaton agent' after Monroe D. Eaton who first isolated it as the causative agent of primary atypical pneumoniae in humans (Eaton et al., 1944). Initially considered a virus, only when antibiotics were found to be effective against its infections, was it classified as Mollicute, which form a class of organisms bearing a small genome with a single circular chromosome, low G+C content and cell wall permanently absent- thus rendering it falsely Gram-positive. 16S rRNA sequencing later revealed some of its closest relatives outside of the Mycoplasmas to include lactic acid bacteria (Waites & Talkington, 2004).

The Mollicutes are thought to have undergone degenerative evolution from a common ancestor to their current forms like *M. pneumoniae*, with its characteristic spindle-shaped cell, and *M. mycoides* while adapting to their particular niches in their vertebrate hosts (Miles, 1992). Simultaneously, they also lost their ability to synthesize all amino acids, have an incomplete tricarboxylic acid cycle and no electron transport chain despite possessing all ten enzymes required for glycolysis (Pollack et al., 1997). This often leads one to the observation in mycoplasmas that proteins classified in known processes tend to moonlight in other functional areas of the cell. For example, pyruvate dehydrogenase E1 β is displayed on the cell surface and facilitates fibronectin binding (Commichau & Stülke, 2008; Dallo et al., 2002).

Due to its reduced genome, small size, lack of several metabolic functions, and reliance on the host for acquisition of several bio-precursors this organism has often been looked at as a model for what life would be if it could ignore the concerns of biosynthesis, elaborate cell envelope construction and focus on adapting to only one type of environment (Krause & Balish, 2004). Thus, several systems level studies have been performed regarding metabolic modelling (Wodke et al., 2013), metabolic regulation (Yus et al., 2009), gene regulatory networks (Yus et al., 2019), transcriptome complexity (Güell et al., 2009), post-transcriptional regulation (Chen et al., 2016), proteome organization (Kühner et al., 2009) and the essentiality of small ORFs and ncRNA (Lluch-Senar et al., 2015). Focusing on the proteomic organization of *M. pneumoniae*, a complexity of protein machinery that did not match its genome minimalism was revealed. Some molecular functions could be predicted from structural genomics data for about two-thirds of the proteins (S. H. Kim et al., 2005), and of the 411 proteins identified, there were 62 homomeric and 116 heteromeric protein complexes which corresponded to 60% of the open reading frames. Excluding most of the unknown 40% of the expressible proteome, it was possible to apply structural modelling to the proteins and complexes that were captured in 26 whole-cell tomograms thus, mapping these machines to their most-likely positions in the cell (Kühner et al., 2009).

Besides proteomic modelling, *M. pneumoniae* was also used to understand and predict the basic principles of transcriptome organization (Güell et al., 2009) and transcriptional regulation (Yus et al., 2019). Again, its small genome size proved deceptive with the use of antisense, alternative transcripts, and multiple regulators per gene all describing an intricately dynamic transcriptome whose complexity approached that of a eukaryotic cell. Surprisingly, transcriptional factors were found to be used in only 20% of the cases tested. Completing this picture, studies to predict the minimal translation apparatus in *M. pneumoniae* revealed that a core of 104 genes are required for core translation activity (Grosjean et al., 2014).

Perhaps unsurprisingly, *M. pneumoniae* was chosen to develop a research and analysis proteomic pipeline that could characterize an unknown bacterium on multiple levels, simultaneously. This entailed the application of crosslinking experiments (O'Reilly &

Rappsilber, 2018) and cellular cryo-electron chromatography (cryo-ET) (Mahamid & Baumeister, 2012) to structurally analyse protein super-assemblies, in particular, the coupling of transcription to translation for which an open question still remained.

1.2 NusA and the Expressome

The lack of separate intracellular compartments in bacteria and archaea means that transcription and translation are linked in a spatio-temporal fashion (French et al., 2007; McGary & Nudler, 2013). Active ribosomes translating the nascent mRNA transcribed by the RNA polymerase (RNAP) have been found in close proximity to each other, often inferring that the ribosome just trails the RNAP (Conn et al., 2019). In *E. coli*, this cooperation between the two complexes was found to prevent spontaneous backtracking of the RNAP, and in the event of rare codon usage, this would slow down the ribosome. Further, nutrient availability as sensed by the ribosome was postulated to be a fundamental mechanism of bacterial gene regulation and thus stress adaptation (Proshkin et al., 2010).

In *E. coli*, cryo-EM showed that a sequence of about 30 nucleotides extends to the ribosomal decoding centre from the α carboxyl-terminal domain of the RNAP, but that this complex, termed the 'expressome', could only form during transcription elongation. The fact that conserved residues were involved in the super-complex implied conservation across most bacteria (Kohler et al., 2017).

NMR studies were the first to show that either the transcription factor NusE (which is identical to the ribosomal protein S10) or Rho forms a complex with the C-terminal domain of NusG. This was the first protein suggested to link the two complexes, with additional evidence being that Rho termination factor only binds when the end of the bacterial operon has been reached (Burmann et al., 2010). However different states of this RNAP-ribosome complex at the nucleic-acid-binding cleft could only suggest, not confirm which factors were responsible for coupling. More importantly, it was found that NusG could not be fit in these models because the predicted NusG-binding sites are located on opposite ends of RNAP (Demo et al., 2017). *In vivo* reporter assays coupled with the integrative cryoEM data of NusG and the 70S ribosome later demonstrated that the recruitment of NusG is heavily dependent on translation status and that linking of the two complexes only begins late in transcription (Washburn et al., 2020). Lastly, when bound to the ribosome, neither NusG nor its *E. coli* homolog RfaH, could simultaneously bind the 30 S ribosomal protein S10 because the linker arm of these two proteins was found to be too short (Kohler et al., 2017). Thus, with concrete evidence for NusG to bridge the RNAP and ribosome also lacking, NusA was left for consideration.

The *M. pneumoniae* crosslinking experiment included the ribosome, RNAP and associated factors. These links were used to decipher the topology of the expressome, as

depicted in Figure 1. The *M. pneumoniae* RNAP core consisting of the α, β and β' subunits are linked with several cofactors, namely SigA, GreA, NusG, NusA, SpxA and RNAP δ. Interactions previously reported from *E. coli*, between NusG and ribosomal protein S10 were not recorded here (Burmann et al., 2010; Washburn et al., 2020). Infact, no direct links between the RNAP core and ribosome were found. Rather NusA, an essential transcription factor with roles in elongation, termination and antitermination (Guo et al., 2018) was revealed to bridge the two complexes via its NTD contacting the RNAP and its disordered CTD interacting with the ribosomal mRNA entry site.

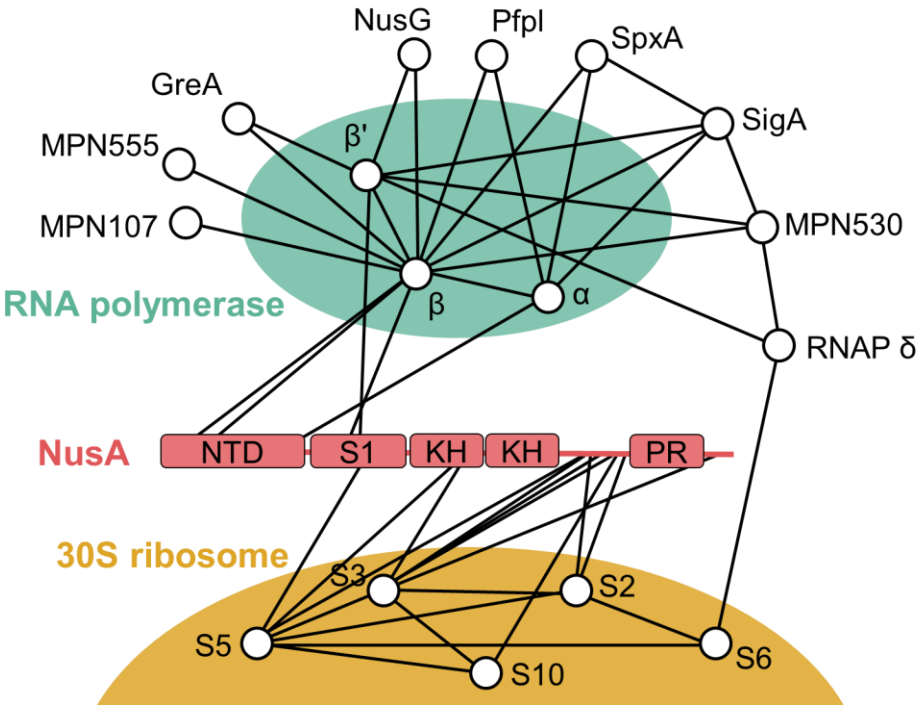


Figure 1: Interaction of RNAP and the ribosome. The RNAP and ribosomal complexes are connected indirectly via NusA and directly via the RNAP δ subunit and the 30S ribosomal S6 protein. NusA N-terminal domain (NTD) contains links to the RNAP while its S1, first KH domain and disordered C-terminal domain (CTD) connect to the 30S ribosome. (modified from O'Reilly et al., 2020).

Given the essential nature of NusA, the function this disordered region of NusA must play was of great interest. Due to this protein not being reported as often as NusG as a super-complex bridge and that the linker arm of RNAP-bound-NusG was too short to interact with the ribosome, we searched for more proof to confirm this aspect of expressome architecture in *M. pneumoniae*. A transposon insertion library that was constructed as part of an effort to determine all the essential genomic components required for minimal artificial life appeared to shed more light on NusA.

This transposon insertion map confirmed the essentiality of NusA (Lluch-Senar et al., 2015) but also indicated that the first three quarters of NusA is structurally crucial as indicated by relatively few mutants surviving until the first passaging event. However, the final quarter or

CTD region, showed many more instances of insertion events in passage one. Given that cryo-EM studies had detected a protein of NusA's electron density, but subsequent integrative modelling could not predict the disordered CTD region, it was decided to isolate truncation mutants that would lack different sections of this disordered CTD. In this way, information regarding the disordered region could be gleamed via truncation experiments, thus covering a part of the protein that cryo-EM and integrative modelling of the complexes could not predict.

1.3 A synthetic-minimal cell, Syn3A

A minimal cell consists of a defined set of annotated and characterized genes, where *in-silico* modelling matches *in vivo* observations. Efforts to develop such a cell have been diverse and varied (Chi et al., 2019; Michalik et al., 2021; Rees-Garbutt et al., 2020). Work on a bottom up approach using Mycoplasma genomes as a starting point commenced as far back as the 1990s (Hutchison et al., 1999) and has yielded the synthetic bacterium JCVI Syn3A based on the genome of *Mycoplasma mycoides* capri LC GM12.

The motivation behind this project was to use bacterial genome sequencing data, as the technique became more widespread, to design, then build and test a cell from the bottom up. Thus, this approach heavily featured chemical gene synthesis and genome assembly using a yeast based cloning biology platform. This multi-faceted research gave rise first to a 1.08 Mb pair *M. mycoides* JCVI-Syn1.0.genome that was transplanted into a *M. capricolum* recipient cell (Gibson et al., 2010). This cell contained all the genes of the wild type *M. mycoides* and some watermarking genes along with polymorphisms introduced during the genome assembly phase in yeast. Subsequently, this genome was reduced further based on a transposon insertion map that classified genes as essential, quasi-essential or non-essential. Thus, the final cell Syn3 was created (Hutchison et al., 2016), and later its derivative Syn3A (depicted in Figure 2).

Syn3 showed vesicle formation and extensive filamentation while in log phase and a slower doubling time of 2-3 hours compared to the spheroidal cell morphology and 1 hr doubling time seen with Syn1 or the wild type. As a means of increasing cell fitness and reproducing the original cell morphology, segment 6 was rearranged, deleting 2 genes, and incorporating 19 others. Despite such a high degree of control of the construction process and the rational design behind its creation, of the 493 genes present, 91 have no known function (Wise et al., 2019).

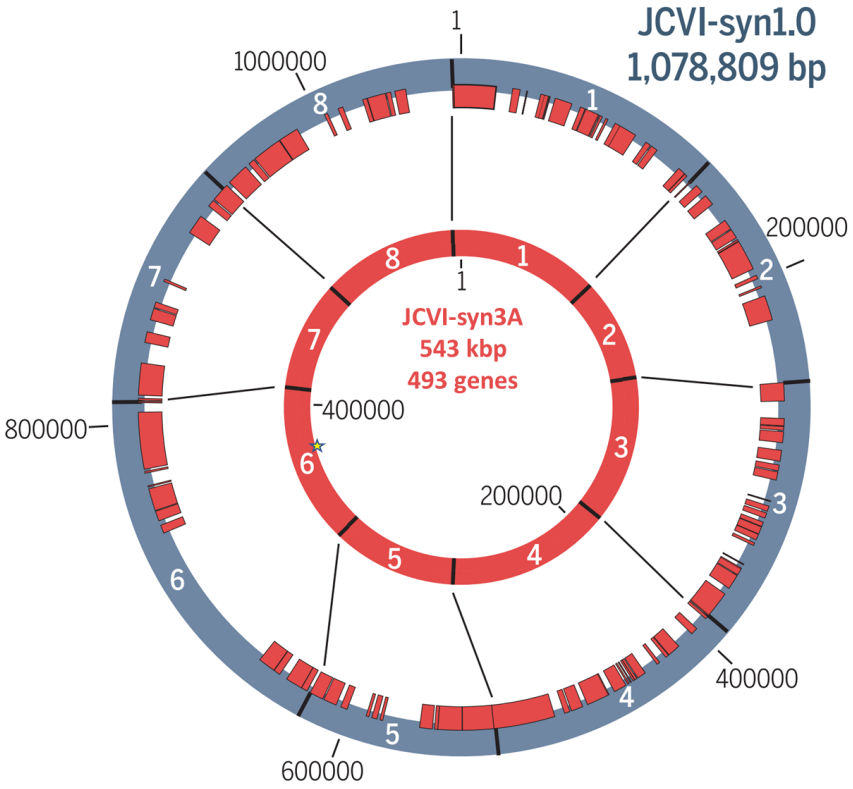


Figure 2: Genome comparison of JCVI-syn1.o (outer grey circle) with Syn3A (inner red circle). The genome was assembled from 8 synthesized segments, the red bars on the outer circle indicate the regions that were retained in Syn3. The yellow star indicates the rearranged segment six which was used to create Syn3A. Adapted from Hutchison et al., 2016.

While the original number of unclear function-genes in Syn3.0 came down from 149 genes (65 unknown and 84 specific substrate/function unclear) due to intense bioinformatic research (Danchin & Fang, 2016; Yang & Tsui, 2018), such methods have their limits due to less than 1% of the protein sequences of Uniprot being annotated with experimentally confirmed gene ontology functions (Antczak et al., 2019). Thus, there clearly was a need for the characterization of unknown, yet essential genes from this model minimal cell. To address this need, we pursued construction of protein-protein and protein-RNA interaction maps.

1.4 Crosslinking whole cells to visualize protein-protein interactions

Whole cell crosslinking is an analytical technique whose power has increased in lockstep with the development of mass spectrometric capabilities. To gain a quick understanding of CLMS it is necessary to describe in brief its enabling component, a crosslinker.

1.4.1 A brief foray into the world of crosslinkers

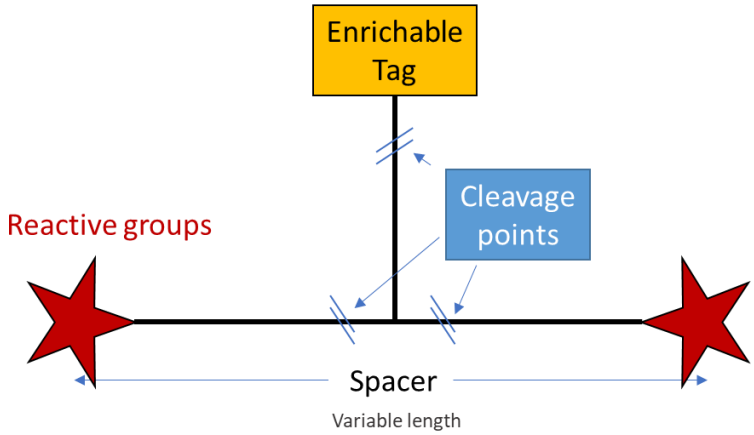


Figure 3: Characteristics of a crosslinker. At the very minimum, a crosslinker contains a spacer and two reactive groups. It may also be cleavable in the mass spectrometer at three locations. Finally, an enrichable tag is useful for purifying only crosslinked peptides, especially when crosslinking whole cells.

As shown in Figure 3 crosslinkers are modular and hence in their long history over 268 crosslinkers have been created. Disuccinimidyl suberate (DSS) and disuccinimidyl sulfoxide (DSSO) have been extensively used for biological studies, often in combination due to their simplicity, specificity, and stability (Iacobucci et al., 2019). In addition, DSSO is cleavable at the C-S bonds which are adjacent to the sulfoxide group. These bonds (indicated in Figure 4) are weaker than the peptide backbone and can hence be selectively destroyed upon collision induced dissociation, a mass spectrometric fragmentation technique (Stieger et al., 2019).

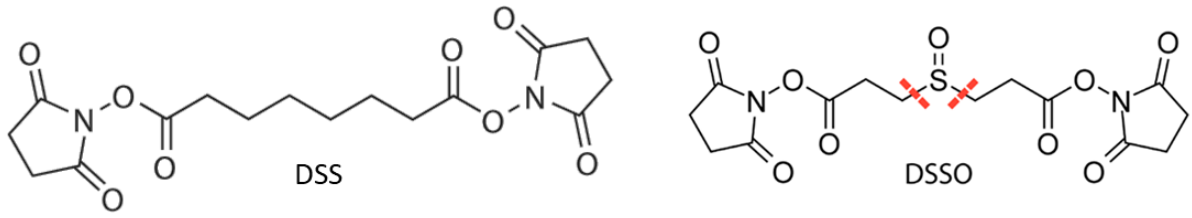


Figure 4: Chemical structures of the two most used crosslinkers. DSS and DSSO share their reactive groups but differ in how they are processed in the mass spectrophotometer (due to the cleavable bonds marked in red), and therefore the information they provide.

These two crosslinkers were used for the previous crosslinking study in *M. pneumoniae* and required a substantial amount of cellular material. This was largely because there was no way to separate the crosslinked peptides from non-crosslinked peptides. To remedy this situation, a new experimental crosslinker termed 'StageCL' was used for the Syn3A experiments. This crosslinker contains an affinity tag grafted to its backbone, that enables the enrichment of only crosslinked peptides. It also contains a shorter spacer length of 7.1 Å (compared to 11.4 Å for DSS and 10.3 Å for DSSO), which enables the deeper penetration of

protein complexes. Lastly, two NHS-ester reactive groups at either end of the spacer, react predominantly with lysine and protein N-termini, but also with serine, threonine, and tyrosine.

1.4.2 Crosslinking workflow for characterization of unknown proteins

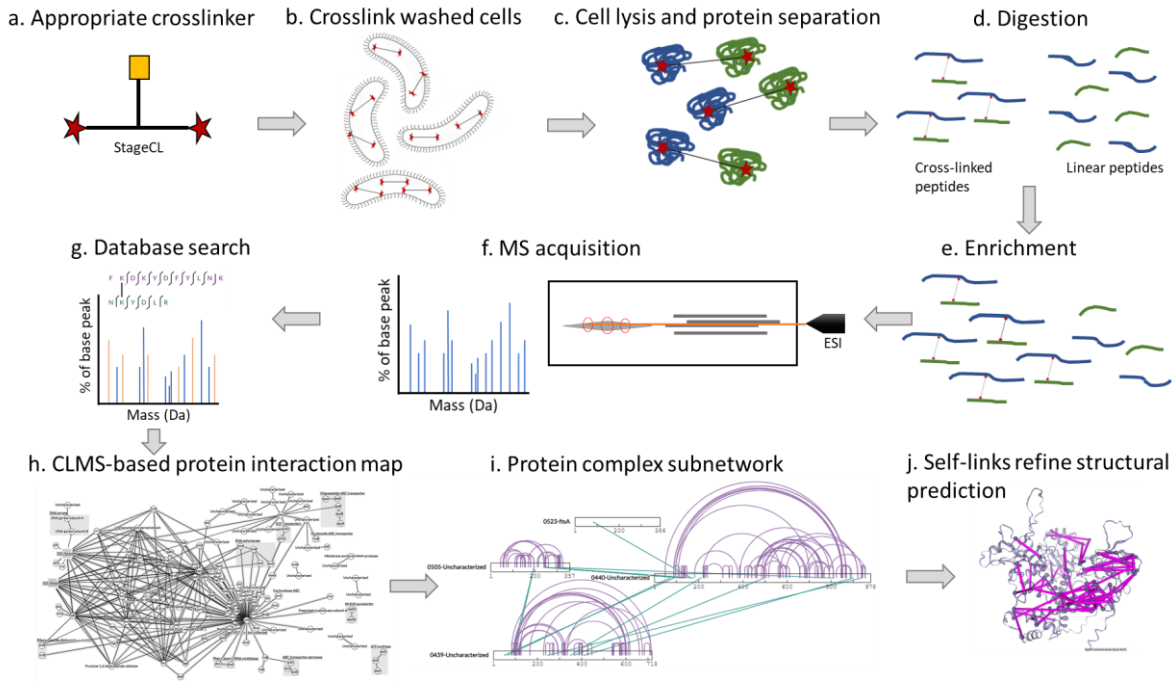


Figure 5: CLMS workflow. a. StageCl is used for the first time in whole cell cross-linking. It is enrichable, has a small spacer arm and NHS reactive groups. **b.** Each bacterial species must be optimized for crosslinker to cell concentration ratios and reaction temperature. **c.** Cell lysis is mechanical or chemical, and if often followed by fractionation to eliminate non-proteinaceous matter. **d.** Proteins are digested after immobilization in gels or in solution and produce a mixture of cross-linked and linear peptides. **e.** Crosslinked peptides are enriched significantly via affinity and size exclusion chromatography (SEC). **f.** Peptides are fed into standardized MS/MS acquisition pipelines which have been optimized to select crosslinked peptides for fragmentation. **g.** Data based software solutions are used to identify the two linked peptides from the spectra and match them to proteins. **h.** The whole proteome in terms of interacting proteins is then visualized based on crosslinked peptides fitted to proteins. **i.** Subnetwork complexes reveal proteins of unknown or unexpected interaction. **j.** Self-links or crosslinks within the same protein can be used to validate predicted structures in the absence of reported structures.

In Syn3A, all the genes have a predefined status regarding their essentiality. This makes the study using traditional methods of gene deletion and phenotype characterization challenging as most likely, the deletion will be lethal or impair growth severely. Hence, a protein-protein interaction experiment such as the one described in Figure 5 is a useful way to gain knowledge on the functional role of uncharacterized proteins in cellular complexes. The benefit of selecting a crosslinker like StageCL with a small spacer arm is that the density of self-links is likely to be higher than with DSS or DSSO. Hence, this could enable validation of structures and later inform on the modelling of protein-complex topology (O'Reilly et al., 2020).

1.5 UV crosslinking to study RNA-binding proteins

Research on the annotated genes of Syn3A genome had revealed that the overall RNA component of the cell like ribosomal proteins, ribonucleases and transcription factors had remained the same compared to Syn1.0. In light of several non-essential proteins being deleted, it was hypothesized that the control of several cellular processes might now be governed more by RNA related mechanisms. Hence, Syn3A would be an ideal platform on which to undertake a whole cell RNA-binding protein study. Such a study is arranged in two main stages, the first of which is briefly described in Figure 6.

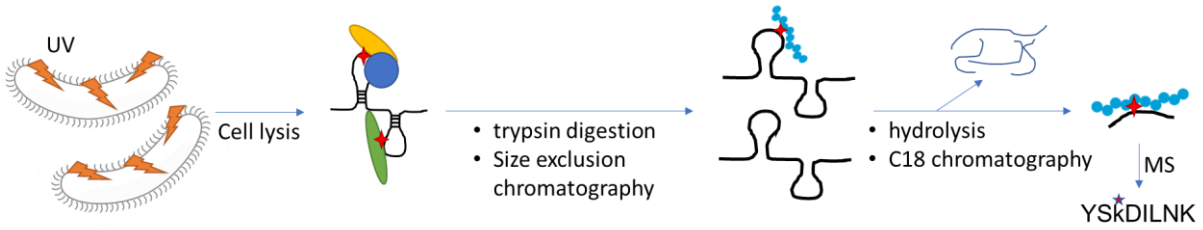


Figure 6: Overview of an UV based RNA-binding protein localization study. Cells are UV crosslinked and then lysed using denaturing conditions. Proteins are trypsinized and crosslinked protein-RNA complexes are enriched by SEC. Hydrolysis separates RNA with or without crosslinked peptides and C18 chromatography removes RNA oligonucleutides from crosslinked heteroconjugates. These RNA-peptide species are analyzed by MS and resulting modified peptides matched to a library of pre-calculated peptide fragments.

UV light of 254 nm is used to covalently couple, or crosslink proteins directly bound to RNA *in vivo*, due to the naturally photoreactive nucleotide bases, specifically pyrimidines and certain amino acids like Phe, Trp, Tyr, Cys, and Lys (Brimacombe et al., 1988). UV crosslinking requires direct contact i.e., zero distance between the RNA species and protein. It also has been shown to not lead to protein-protein crosslinking (Pashev et al., 1991) and has been used before to study purified RNA-protein complexes *in vitro* (Kramer et al., 2014). Thus, such an experiment would provide accurate data on the RNA-binding proteins as well as the exact RNA-interacting amino acid residue

In the second stage, again UV crosslinking is used on a sample of cells, but also an identical control sample is not UV-treated. These samples are treated the as depicted in Figure 7, with main aim being to analyse all the peptide fragments. This experiment does not give any data on the crosslinking site in the protein but can provide data on the abundance of potential RNA-binding proteins.

Hence, results of the second stage provide a level of confirmation to the direct UV crosslinks between RNA and amino acid residues from the first stage of experiments.

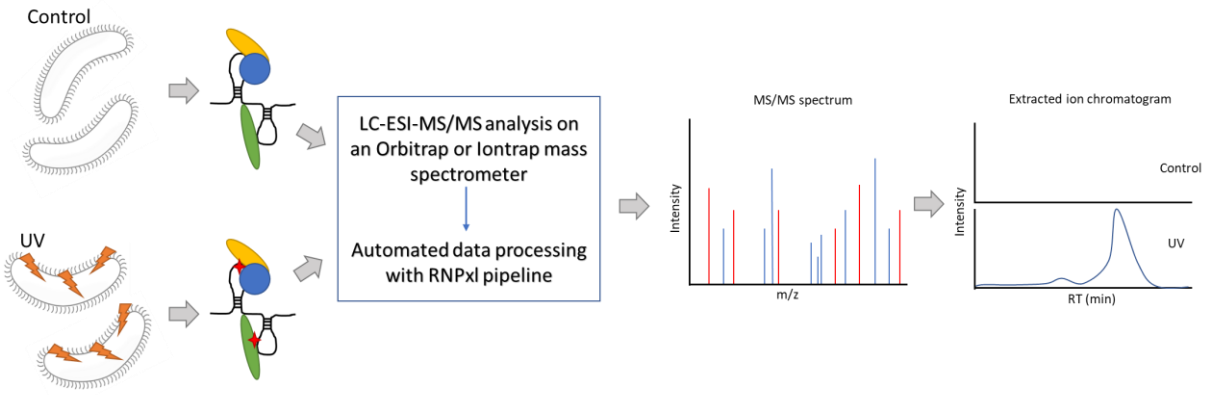


Figure 7: Experimental workflow of RNA-binding protein enrichment study. Whole cell samples are UV irradiated along with a control. Cells are resuspended and lysed in TRIzol™ reagent. Irradiated protein-RNA complexes are processed for MS/MS analysis along with a non-irradiated control. MS data analysis workflow with RNPxl pipeline results in potential cross-linked peptides. Results are validated by checking of their extracted ion chromatogram intensities in the UV-irradiated sample vs the control. Further, the expected and observed fragmentation patterns in the MS/MS spectrum are also evaluated.

1.6 Aims of this work

The coupling of RNA polymerase and the ribosome has been confirmed predominantly in studies with *E. coli* (Kohler et al., 2017). However, the question of how these complexes were connected has not yet been resolved. CLMS and cryo-EM studies both showed that this 'Expressome' structure exists in *M. pneumoniae* as well. NusG, an oft mentioned candidate for bridging the two complexes was not found to be the linking protein (Demo et al., 2017), but rather NusA was a more likely candidate. As integrative modelling could not predict the structure of the NusA disordered CTD, which was crosslinked to the 30S ribosome, one aim of this dissertation is to validate the essentiality of this region by attempting to construct truncation mutants lacking the C-terminal region.

The CLMS study demonstrated the usefulness of this technique in characterizing proteins with unknown functions. Hence, the second aim of this thesis is to undertake a series of crosslinking experiments in Syn3A, an artificial minimal cell with the smallest genome of a free-living organism. As almost one third of its genome is poorly annotated or uncharacterized (Hutchison et al., 2016), the aim is to provide context to some of these unknown proteins and begin characterizing them. Lastly, due to half of its genome being deleted, regulatory events and mechanisms are thought to be shifted towards RNA-based regulation. Hence, UV-crosslinking studies are to be undertaken to identify the complete set of RNA-binding proteins in this minimal cell. These results are to be partially confirmed with RNA-binding protein enrichment studies later.

2 MATERIALS AND METHODS

2.1 Materials

Materials including antibodies, chemicals, commercial services, enzymes, equipment, oligonucleotides, proteins, software, utilities, and websites are listed in the Appendix.

2.1.1 Bacterial strains and plasmids

A list of all bacterial plasmids and strains used in this work can be found in the appendix.

2.1.2 Growth media and facultative additives

All buffers, media, and solutions were prepared using deionized water (unless stated otherwise) and autoclaved for 20 minutes at 121°C and 2 bar. Thermolabile substances were dissolved and sterile filtered through a 0.2µm sterile filter. For solid media, 15 g/L agar was added.

Table 1: Formulations of growth media, stock colutions and additives

Type	Name	Component/Instruction	Amount
Media for cultivation of *E. coli*	LB Medium (1L) (modified after Bertani 1951)	Tryptone	10 g
		Yeast extract	5 g
		NaCl	10 g
		dH_2O	*ad* 1 l
	SOB medium	Tryptone	20 g
		Yeast extract	5 g
		NaCl	0.58 g
		KCl	0.186 g
		ad 1 l dH_2O, autoclave	
		$MgCl_2$ (1 M)	10 ml
		$MgSO_4$ (1 M)	10 ml
Media for cultivation of *M. pneumoniae*	MP Medium (500 ml) Modified Hayflick's medium (Chanock *et al.*, 1962)	PPLO broth	7.35 g
		HEPES	11.92 g
		Phenol red (0.5%)	2 ml
		NaOH (2 M)	14 ml
		dH_2O	Ad 400 ml
		Autoclave, cool down to ~50°C and then add:	
		Horse serum (heat inactivated)	100 ml
		Penicillin (100,000 U/ml)	5 ml
		Carbon source (glycerol/glucose)	10 ml
		Difco PPLO broth	3.5 g

Media for cultivation of *Syn1, Syn3,* and *Syn3A*	SP4 medium (modified after Tully, Whitcomb *et al.* 1977) Part I	Bacto tryptone Bacto peptone dH_2O pH 7.5, autoclave 15 min, 121°C	10 g 5.3 g Ad 600 ml
	Part II	20% Glucose CMRL 1066 7.5% $NaHCO_3$ 200 mm L-glutamine 25% Gibco Yeast extract solution 2% TC yeastolate (autoclaved) FBS, heat inactivated Penicillin G (400,000 U/ml) 0.5% phenol red Sterile filter (0.2 µm) Mix part I and II, store at 4°C.	20 ml 50 ml 15.6 ml 5 ml 35 ml 100 ml 170 ml 2.5 ml 3 ml
Additives (1000x)	IPTG X-Gal	Stock solution in H_2O Stock solution in DMSO	1 M 80 mg/ml

Table 2: Formulations of buffers, stock solutions, and staining solutions

Type	Name	Component/Instruction	Amount
Blotting solutions	TBST	Tris-Base NaCl pH 7.6 dH_2O Tween-20	6 g 9 g *ad* 1 l 1 ml
	Blotto	skimmed milk powder TBST	25 g *ad* 1 l
	Buffer III	Tris-base NaCl pH 9.5 dH_2O	12 g 5.8 g *ad* 1 l
ssDNA preparation	Washing buffer	5M NaCl 10x TE buffer dH_2O	40 ml 10 ml 50 ml
	Melting buffer	10 M NaOH dH_2O	125 µl 9,875 ml
	Neutralization solution	3 M NaAc 1x TE buffer	50 µl 2.45 ml
Protein purification	Buffer W	Tris-HCl pH 8.0 NaCl EDTA	100 mM 150 mM 1 mM
	Buffer E	Tris-HCl pH 8.0 NaCl EDTA	100 mM 150 mM 1 mM

		d-Desthiobiotin	2.5 mM
	ZAP buffer	NaCl	200 mM
		Tris-HCl pH 7.5	10 mM
	HEPES buffer (adapted from (Y. Kim et al., 2009))	HEPES	100 mM
		NaCl	500 mM
		Glycerol	50 ml
		pH	8.0
		dH_2O	*ad* 1 l
		Before use add:	
		β-mercaptoethanol	10 mM
		Imidazole	10 mM
SDS-PAGE	Stacking gel (5%)	Rotiphorese® Gel 30%	1.3 ml
		1.5 M Tris-HCl pH 6.8	0.87 ml
		SDS (10%)	100 µl
		APS (10%)	100 µl
		TEMED	30 µl
		dH_2O	6.83 ml
	Running gel (12%)	dH_2O	3.3 ml
		Rotiphorese® Gel 30%	4.0 ml
		1.5 M Tris-HCl pH 8.8	2.5 ml
		SDS (10%)	0.1 ml
		APS (10%)	0.1 ml
		TEMED	30 µl
Staining solutions	Fixation solution (Collodial Coomassie)	Methanol	450 ml
		Acetic acid	100 ml
		dH_2O	Ad 1 l
	Staining solution (Colloidal coomassie)	dH_2O	100 ml
		H_3PO_4	100 ml
		Ammonium sulfate	100 g
		when ammonium sulfate has dissolved, add:	
		Coomassie Brilliant blue G-250	1.2 g
		Methanol	200 ml
		dH_2O	Ad 1 l
	Fixation solution (Silver stain)	Methanol	50 ml
		Acetic acid	20 ml
		Formaldehyde (37%)	100 µl
		dH_2O	Ad 100 ml
	Silver-Developer	Na_2CO_3	6 g
		Thiosulfate solution	2 ml
		Formaldehyde (37%)	50 µl
		dH_2O	Ad 100 ml
	Silver-Impregnator	$AgNO_3$	0.2 g
		Formaldehyde	37 µl

		dH_2O	Ad 100 ml
	Thiosulfate solution	$Na_2S_2O_3 \cdot 5\ H_2O$	20 mg
		dH_2O	Ad 100 ml
	Silver-Stop solution	EDTA	1.86 g
		dH_2O	Ad 100 ml
	Transfer buffer	Tris-base	15.1 g
		glycine	14.4 g
		methanol	200 ml
		dH2O	*ad* 1 l
Miscellaneous	Electroporation Buffer	HEPES	8 mM
		Sucrose	272 mM
		Adjust to pH 7.4	
	Phosphate Buffered Saline (PBS)	KCl	0.2 g
		KH_2PO_4	0.2 g
		$Na_2HPO_4 \cdot 2H_2O$	1.4 g
		NaCl	8 g
		pH	7.4
		dH_2O	*ad* 1 l
	Buffered formalin	Formaldehyde (37%)	10 ml
		PBS pH 7.4	90 ml
	Crystal violet solution	Crystal violet	0.1 g
		dH_20	100 ml
	TAE buffer (50x)	Tris	242 g
		Acetic acid	57.1 ml
		0.5 M EDTA pH 8.0	100 ml
		dH_2O	Ad 1 l
	DNA loading dye (5x)	100% glycerol	5 ml
		50x TAE	200 µl
		Bromophenol blue	10 mg
		Xylene cyanol	10 mg
		dH_2O	4.5 ml
	Lysis Buffer (adapted from Wise et al., 2019)	1 M NaF	5 µl
		1 M BGP	5 µl
		200 mM sodium orthovanadate	25 µl
		1 M sodium pyrophosphate	50 µl
		1 M Tris-HCl, pH 7	125 µl
		10% CHAPS	500 µl
		5 M NaCl	100 µl
		50 % glycerol	100 µl
		1 M Pefablock	5 µl
		dH_2O	*ad* 5 ml

2.1.3 Antibiotics

All antibiotics were prepared in 1000x concentrated solutions. The solutions were sterilised by filtration and stored at -20°C. Prior to use, they were thawed on ice and added to fresh, autoclaved medium at 45-50°C. (Commichau et al., 2015)

Table 3: Antibiotics and selective concentrations

Organism	Antibiotic	Selective concentration
E. coli	Ampicillin	100 µg/ml (distilled water)
	Carbenicillin	100 µg/ml (distilled water)
	Kanamycin	50 µg/ml (distilled water)
	Chloramphenicol	30 µg/ml (ethanol)
	Streptomycin	100 µg/ml (distilled water)
M. pneumoniae	Chloramphenicol	20 µg/ml (ethanol)
	Gentamycin	80 µg/ml (distilled water)
	Penicillin	1,000 U/ml (distilled water)
	Puromycin	3.5 µg/ml (distilled water)

2.2 Methods

Original publications of general methods employed in this work are described in Table 4.

Table 4: General methods.

Method	Reference
Chain termination DNA sequencing	Sanger *et al.*, 1977
Coomassie staining of protein gels	Fazekas de St Groth *et al.*, 1963
Determination of optical density	Sambrook *et al.*, 1989
Determination of protein amounts	Bradford, 1976
Gel electrophoresis of DNA	Sambrook *et al.*, 1989
Gel electrophoresis of proteins (SDS-PAGE)	Laemmli, 1970
Ligation of DNA fragments	Sambrook *et al.*, 1989
Plasmid preparation from E. coli	Sambrook *et al.*, 1989
Precipitation of nucleic acids	Sambrook *et al.*, 1989

2.2.1 Cultivation techniques

Growth and preservation of E. coli

E. coli strains were grown over night at 37°C in LB medium or on plates containing 1.5% agarose in LB medium, supplemented with appropriate antibiotics and selective substrates.

In order to preserve certain strains over longer periods of time, cryo stocks were created. One colony of bacteria was picked from a plate and cultivated in LB medium with the appropriate antibiotics at 37°C over night. 900 µl of the culture was mixed with 100 µl of DMSO

and transferred into a sterile screw cap tube. Alternatively, 800 µl of the culture was mixed with 200 µl of glycerol. The tubes were frozen in liquid nitrogen and stored at -80°C.

Growth and preservation of M. pneumoniae

M. pneumoniae was grown at 37°C in modified Hayflick medium (MP medium) for 96 hours in adherent tissue culture flasks. Cell culture flasks with 12.5, 75, 150 and 300 cm² surface area corresponding to a volume of 10, 50, 80, 150 ml modified MP medium were used. Fresh cultures were grown by pre-warming the appropriate volume of medium, inoculating at a ratio of 1:100 stock culture to sterile medium.

In order to preserve cultures over a longer period of time, the supernatant of a freshly grown culture was discarded after 96 hours or upon formation of a visible growth layer, depending on the strain. Cells were scraped in 20 ml fresh modified Hayflick medium, and then aliquoted in 10 ml falcon tubes- Aliquots were frozen in liquid nitrogen and stored at -80°C.

Growth and preservation of Syn1, Syn3, and Syn3A

Cells were grown as static cultures or under agitation at a maximum of 180 rpm at 37°C. Growth rates under both conditions are equal for *Syn3* and *Syn3A*, whereas *Syn1* can only be grown under static conditions. In liquid cultures cells grow as planktonic aggregates that form cell pellets in stationary cultures, which must be shaken prior to cell passaging. SP4 medium (modified from Tully, Whitcomb *et al.* 1977) used to grow these cells contains a pH indicator allowing for colour change of the media after around 50 hours of growth. This change is best monitored by keeping sterile medium filled tubes for visual comparison. Culture volumes are to be scaled 1:100 into fresh prewarmed medium for 2-3 day incubation periods. Final cell yields of *Syn3* and *Syn3A* are lower than *Syn1*.

Cells may be plated on agar medium containing SP4 medium and 1.5% agar, incubated at 37°C for upto 5 days. Longer incubation times are required for individual cells to display typical mycoplasmal, fried egg like morphology.

For long term storage, log phase cultures are resuspended by shaking and then aliquoted into 5-10 ml aliquots to be frozen in liquid nitrogen and stored at -80°C. Viability of stocks is best when cultures that are just beginning to turn orange are used for preservation. This indicates majority of cells are in pre-stationary growth phase.

2.2.2 Genetic modification of *E. coli*

Preparation of competent E. coli

Two methods were used to create competent *E. coli* cells.

Preparation of chemically competent E. coli cells (Inoue *et al*., 1990)

20 ml LB medium was inoculated with an isolated colony of freshly streaked *E. coli* DH5α or XL1-Blue and incubated for 20-24 hours at 28°C and 220 rpm. 6 ml of this liquid preculture was then used to inoculate a scaled-up culture of 250 ml SOB medium which was incubated for 20-24 h at 16°C and 220 rpm or until an OD_{600} of 0.5–0.9. The temperature of the culture was reduced via cooling on ice for 10 mins followed by centrifugation for 10 min at 5000 rpm, 4°C thus harvesting the cells. 80 ml ice-chilled TB buffer was used to resuspend the cell pellet after discarding the supernatant. The cells were incubated on ice for 10 mins and then harvested once again, as before. Finally, 20 ml ice-cold TB was used to resuspend the cells and then DMSO was added to a final concentration of 7% (v/v). 200 µl aliquots of chemically competent *E. coli* cells were flash frozen using liquid nitrogen and stored until further use at - 80°C.

Preparation of fresh chemically competent *E. coli* cells (Lederberg and Cohen, 1974)

Fresh competent *E. coli* cells were prepared using the calcium chloride method. This involved the cultivation of DH5α, XL1-Blue, BL21 (DE3) or Rosetta overnight at 28°C and 220 rpm. This was used to inoculate a 10 ml shake-flask culture to an OD_{600} of 0.1. After incubation at 37°C until OD_{600} 0.3, the culture was transferred to a 15 ml falcon tube and centrifuged for 6 minutes, using 4,000 rpm at 4°C. The supernatant was discarded, the cell pellet resuspended in ice cold $CaCl_2$ (50 mM) and incubated on ice for 30 minutes. This suspension was further centrifuged (using the same parameters), the supernatant discarded, and the cell pellet resuspended in 1 ml ice cold $CaCl_2$ (50 mM). An aliquot of 200 µl was used for subsequent transformation.

Transformation of chemically competent E. coli cells

200 µl of fresh competent cells was mixed in a 1.5 ml Eppendorf tube with 10-100 ng of DNA and incubated for 30 minutes on ice. This mixture was heat shocked for 90 seconds at 42°C and incubated on ice again for 5 minutes. 500 µl LB-medium (without antibiotics) was then added and cells were incubated for 2 hours at 37°C, 220 rpm (this incubation step may be shortened to 10 minutes when ampicillin is used as the selectivity marker). The entire cell

suspension was finally plated on selective media (2 plates, different dilutions) and incubated overnight at 37°C.

2.2.3 Genetic modification of *M. pneumoniae*

Electroporation was used to transform *M. pneumoniae* as it lacks the genetic components required to develop competence. The requisite strain was grown over 4-5 days in MP medium at 37°C, in T75 flasks. The spent media was then discarded, the adhered cells washed twice with electroporation buffer and then scrapped off into 1.5 ml electroporation buffer. This suspension was centrifuged for 10 mins at 4000 rpm, 4°C. The resulting pellet was finally resuspended in 150 µl ice cold electroporation buffer.

50 µl of this cell suspension was further mixed with 3.5 µg of single stranded DNA and 1 µl yeast tRNA in a 1.5 ml tube. This reaction was made up to a volume of 80 µl by adding electroporation buffer. It was transferred to an electroporation cuvette and incubated for 15 minutes on ice. The samples were electroporated using 2.5 kV, 25 µF, and 100 Ω after which they were immediately transferred to a T25 flask containing 10 ml MP medium without additional antibiotics. This was incubated overnight at 37°C. The following day, the cells were scrapped off and centrifuged for 10 mins at 4000 rpm, 4°C. The supernatant was discarded, and the cell pellet resuspended in 500 µl PBS using a fine tipped needle. This solution was finally used to create serial dilutions till 10^{-5}, which were plated on MP agar plates containing the appropriate antibiotic as selection marker. The plates were incubated at 37°C under humid conditions until colonies appear, but not longer than 3 weeks.

2.2.4 Preparation and detection of DNA

All work with DNA was performed using DNase-free solutions, and materials. Unless specifically mentioned, experiments were performed at room temperature, and DNA was solubilized in sterilized, deionized water from a Merck Millipore water purification system.

Isolation of chromosomal DNA from M. pneumoniae and Syn3A

Cells grown for 4-5 days in a T75-flask were washed twice with PBS (pH 7.4) and scrapped off in 1.5 ml PBS. The cells were centrifuged (1 min, 13,000 rpm, 4°C) and the pellet was resuspended in 50 µl lysis buffer and 10 µl RNaseA (20 mg/ml). The sample was then incubated for at least 25 min at 37°C and DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen®) as per the manufacturer's instructions. (Buffers used were part of the kit.)

Preparation of plasmid DNA from E. coli

Plasmid DNA was isolated using the NucleoSpin Plasmid Mini-prep Kit (Macherey-Nagel®). *E. coli* carrying the desired plasmid was grown overnight, at 37°C and 180 rpm in 4 ml LB medium containing the appropriate antibiotic as selective agent. Cells were harvested by centrifugation of 2 ml culture at 11,000 rpm, 2 min at room temperature. The pellet was treated further as per the manufacturer's instructions. This isolation is based on alkaline lysis followed by chromatographic separation.

Quantification of nucleic acids

NanoDrop® was used to determine the concentration of DNA in 1.25 µl of sample. The samples were analysed using 260-280nm spectra in the UV-visible range with the pure solution used to suspend the DNA as blank (Warburg & Christian, 1942). Purity of the nucleic acid samples was verified ascertained by analysis of the UV graph, 260/280 and 260/230 nm absorbance ratios.

DNA sequencing

Sanger sequencing was utilized for all DNA sequencing reactions. This method relies on the chain termination method and fluorescence labelled dideoxynucleotides (Sanger et al., 1977). Commercial services were provided by Microsynth Seqlab (Göttingen) and alignment was performed using the Geneious Prime© Software package.

Gel electrophoresis

To visualize or separate DNA of differing size or conformation, gel electrophoresis was performed. Gels containing 1% agarose in TAE buffer and 0.001% HD Green (Intas®) were prepared. DNA samples premixed with 5x DNA loading dye along with a DNA ladder (containing fragments of known size) were loaded on the gel and separated using 120 V for 30 minutes. Gels were photographed under UV-light (λ = 254 nm) using a Molecular Imager Gel Doc XR+ system (Bio-Rad Laboratories). DNA ladder was self-made consisting of λ-phage DNA digested with *BamHI* and *EcoRI*.

Polymerase chain reaction (PCR)

For the polymerase chain reaction, chromosomal or plasmid DNA was used as template and reactions were performed in a thermocycler. Either DreamTaq® or Fusion® high-fidelity

polymerase was used for sequencing purposes or check PCR and cloning (Fusion only). Oligonucleotides were designed using the Geneious Software 19 (Biomatters). The oligonucleotides created or used in this work are listed in the Appendix. PCR reactions were performed in 50 µl aliquots (Commichau et al., 2015).

Long-flanking homology PCR or Gibson cloning

This is a method based on homologous recombination used to generate directed and clean gene deletions. LFH PCR (Wach, 1996) enables one to create a gene deletion cassette which is flanked by homology regions. Thus, it combines two PCR steps- first, the amplification of a marker gene and the amplification of a ~500 bp long up and downstream region of the gene of interest. It is imperative that the inner primers used for these two flanking regions have a sequence complementary to the ends of the marker gene. This would enable the integration of the marker gene in the second PCR reaction step. The three sequences are prepared separately and then mixed in a reaction tube without oligonucleotides. The terminal oligoes were added after an initial run which allowed for better mixing.

Table 5: LFH-PCR pipetting scheme

Volume (µL)	Compound
20	5x Fusion® HF buffer
4	dNTPs (12.5 µmol/ml)
8	Forward primer (5 µmol)
8	Reverse primer (5 µmol)
X	100 ng upstream fragment
X	100 ng downstream fragment
X	150 ng deletion cassette
2	DNA Fusion® polymerase (2 U/µl)
ad 100	dH_2O

Table 6: LFH-PCR Protocol

Reaction	Temperature(°C)	Duration	Number of cycles
Initial denaturation	98	1 min	
Denaturation	98	15 sec	10
Annealing	52	30 sec	
Elongation	72	30 sec / kbp	
Hold	8	∞	
Addition of primers			
Denaturation	98	15 sec	25
Annealing	52	30 sec	
Elongation	72	30 sec / kbp	
Final elongation	72	10 min	
Hold	8	∞	

Combined chain reaction PCR (CCR)

This method is used to induce site directed mutations in PCR products (Hames et al., 2005). One of the primers is mutagenic inducing one or more base pair changes as it binds more strongly to the template than external primers due to its +10°C higher melting temperature. These mutagenic primers are also phosphorylated at their 5' ends so that they may be phosphorylated to the 3'-OH groups of the sequence from the upstream primer. Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) is used as it does not exhibit 5'-3' exonuclease activity.

This method is especially relevant when dealing with genes from *M. pneumoniae* expressed in organisms like *E. coli* as they differ in the reading of their genetic code. *M. pneumoniae* uses TGA to code for tryptophan which is a stop codon in *E. coli* (Inamine et al., 1990). Therefore, prior to any application all instances of this codon must be converted to TGG.

The reaction setup and program are identical to a normal PCR except for the addition of 4 µl mutagenesis primer, 3 µl Ampligase (Epicentre. Madison, USA) and 2 µl BSA (New England Biolabs, Ipswich, USA). CCR-PCR was always performed using a plasmid harbouring the gene of interest rather than chromosomal DNA, to reduce unspecific amplification.

Table 7: CCR Reaction Mix

Volume	Compound
2 µl	Forward primer (5 µm)
2 µl	Reverse primer (5 µm)
4 µl	Mutagenic primer (5 µm)
1 µl	Plasmid template DNA (1 ng/µl)
5 µl	CCR buffer (10×)
1 µl	Phusion DNA polymerase (2 U/µl)
3 µl	Ampligase (5 U/µl)
2 µl	dNTPs (12.5 mm)
2 µl	BSA (20 µg/µl)
	ad 50 µl dH2O

Table 8: Thermo Cycler Program

Reaction	Temperature	Duration	Number of cycles
Initial denaturation	95°C	5 min	
Denaturation	95°C	1 min	30
Annealing	T_m-5°C	1 min	
Elongation	68°C	1 min / 1 kb	
Final elongation	68°C	10 min	
Cool down	8°C	∞	

Digestion and dephosphorylation of DNA

Vector DNA and linear DNA fragments (inserts) were separately digested using FastDigest restriction enzymes and FastDigest buffer (Thermo Fisher Scientific) according to the manufacturers protocol. To prevent re-ligation, digested vectors were dephosphorylated at their 5' end via addition of 1 µl FastAP thermosensitive alkaline phosphatase (Thermo Fisher Scientific) and incubation for 10 mins at 37°C. This step was repeated once more and then samples were purified using QIAquick PCR purification kit (Qiagen®).

Ligation of DNA

Digested and purified inserts were ligated to dephosphorylated vectors using a three to 10-fold excess of insert. Ligation was performed using T4 DNA ligase (Thermo Fisher Scientific) in its associated ligase buffer. As per the manufacturers protocol, the ligation reaction was carried out in the dark, at 16°C overnight or for 10 minutes at 22°C (adapted from Thermo Fisher Scientific Inc. 2012). A re-ligation control to measure the efficiency of the dephosphorylation was also used per reaction.

2.2.5 Targeted gene deletion in *M. pneumoniae*

To delete a gene from the genome of *M. pneumoniae*, the gene of interest is replaced with an antibiotic selection marker, using a recombineering approach (Piñero-Lambea et al., 2020). This marker is later excised leaving minor *lox* scars.

A plasmid must first be constructed that contains the 3' and 5' homologous regions (500 bp each) upstream and downstream of the gene of interest. The backbone used for our purpose was pGP2727 where the chloramphenicol resistance marker is flanked by lox sites (lox71 and lox66). These lox sites are required to get rid of the Cm^R cassette at the end. LFH-PCR was used to create an insert that resembles the regions of the gene of interest in the *M. pneumoniae* genome, with the actual gene replaced by lox71-Cm^R-lox66. The correct orientation of the gene in this experimental method is crucial.

PCR amplification and clean up

Using this plasmid as template along with protected (for the targeted strand) and biotinylated oligonucleotides, PCR products of the desired region on the plasmid were produced and purified together with the QIAquick PCR purification Kit (Qiagen®).

Single stranded (ss) DNA purification

Magnetic beads coated with Streptavidin (Dynabeads™, ThermoFisher Scientific) are used to separate double-stranded (ds)-DNA wherein one strand contains a biotin-tag and the other is protected. 80 µl of magnetic beads were used per purified PCR product. The desired amount of beads was transferred to an Eppendorf tube, washed with 500 µl washing buffer and placed on a magnet. This led to accumulation of beads near the magnet and the supernatant could be discarded. This washing step was repeated two more times. The beads were then resuspended in 75 µl washing buffer and mixed with the purified PCR product. After vortexing, the tubes were incubated for 20 min at RT under rotation. This allowed binding of the biotin-tagged DNA fragments to the magnetic beads. Following this, the mixture was transferred to a new Eppendorf tube and the supernatant was discarded again via magnetic separation. Addition of 50 µl melting buffer led to denaturation of the dsDNA molecule. As the ssDNA harboring the biotin-tag still binds to the beads, the other strand is now in solution. The tube was placed on the magnet again and the supernatant was removed and mixed with 500 µl neutralization solution. 50 µl melting buffer was added once more and the supernatant added to the mixture created in the previous step. Afterwards, 3 µl glycogen (20 µg/µl) and 600 µl isopropanol were added to the neutralization mixture and incubated over night at -20°C. On the next day, samples were centrifuged (30 min, 13,000 rpm, 4°C). The supernatant was discarded, and the DNA pellet was washed once with 70% ethanol. The pellet was dried and afterwards resuspended in 30 µl electroporation buffer. DNA concentration was estimated using Nanodrop®.

Screening

After transforming *M. pneumoniae* as described earlier, plates were screened for positive clones. Using a sterilized toothpick, colonies are picked and used to inoculate 200 µl of MP-medium (supplemented with puromycin and chloramphenicol) in a 96 well plate. A well containing only medium served as negative control (no color change) while another well inoculated with wild type cells (no antibiotics added) served as a positive control. These plates were incubated at 37°C for 7-10 days, continuously checking for color change. Upon change in color, 4 µl of the supernatant was taken for PCR analysis. After boiling the sample to release the genetic material, PCR was performed with primers amplifying the Cm^R cassette (CB85/CB86). If the PCR revealed positive clones, these clones were then checked for the presence of the homologous regions and the absence of the gene (again via PCR). Finally, using internal (on the deletion cassette) and external (outside the homologous region, on the genome) primers, insertion of the cassette at the correct location was verified. Only, when all PCR analysis steps showed the desired results, were clones accepted as deletion mutants.

Further, the antibiotic marker is excised using a Cre recombinase. A transformation step identical to that described before was to be performed with 3 µg of pBSKP438Cre (harboring gentamycin resistance). Transformants would then be selected for using MP-medium (supplemented with puromycin and gentamycin). Finally, a mutant strain lacking the gene of interest but containing a slight lox scar would be created.

2.2.6 Preparation and analysis of proteins

Unless mentioned otherwise, proteins were stored for upto 48 hours in aqueous solution at 4°C, or for longer terms in appropriate buffers containing 40% glycerol and DTT at -20°C.

Over-expression of recombinant proteins in E. coli

A protein specific strain of *E. coli* (DH5α / BL21(DE3) / Rosetta / C43 / CLG190) transformed with the desired plasmid was inoculated into 50-100 ml LB as preculture and incubated at 37°C, at 180 rpm, overnight. Preculture was used to inoculate 1 l 2x LB medium with the selective antibiotic, to an OD_{600} of 0.1. The culture was grown to an OD_{600} of 0.8-1.0 at 37 °C, 180 rpm. At this OD, over production of the recombinant protein was induced via addition of isopropyl-β-d-thio-galactopyranoside (IPTG) to a final concentration of 1 mM. After incubation for 3 hours more at 37°C or 16 hours more at 25°C, the cells were harvested via centrifugation at 5,000 rpm for 15 minutes and 4°C. The cell pellet was resuspended in the respective down-stream buffer (Strep: Buffer W, His_6: ZAP with 10 mm imidazole). After transfer to 50 ml falcon tubes, the cell suspension was once again centrifuged at 8500 rpm, for 15 min, at 4°C. The supernatant was discarded, and the pellets stored at -20°C until further processing.

Disruption of bacterial cells

For disruption of the cells, cell pellets of the 1 l cultures were pooled and resuspended in 15 ml 1x buffer (the same to be used for later purification steps). Disruption was performed with SLM Aminco 2-FA-078-E1 French Press Cell (SLM Aminco) at 18,000 psi twice per sample. The cell lysate was ultracentrifuged (1 hr, 35,000 rpm, 4°C) to clear the lysate. The supernatant (crude extract) was used for further protein purification.

For whole cell protein extraction in *Mycoplasmal* cells, sonication was also used as per the manufacturer's instruction. 1ml of freezer-stored JCVI-syn3A was thawed and pelleted at 6,000 rpm for 5 min. The supernatant was discarded, and the pellet washed twice in PBS. Cells were then resuspended in 2ml of lysis buffer and passed through a 21-gauge needle 20

times. The cells were disrupted by sonication (3 × 10 impulses, amplitude 60, 0.6 cycles). Then, the cell debris was removed by spinning at 14,000 rpm for 5min. This supernatant was used for pull-down experiments.

Purification of His_6-tagged proteins by affinity chromatography

Proteins carrying a 6x His-tag bind selectively to a matrix of Ni-NTA®. 2.5 ml of Ni-NTA® (IBA Lifesciences) matrix per 1 l cell culture was pipetted into the column (BioRad Poly-Prep® Chromatography Column) and equilibrated with 12.5 ml ZAP buffer containing 10 mM imidazole. The crude extract was then loaded on the column and the flow through collected. Column washing was performed with 12.5 ml 10 mM Imidazole solution. The elution steps progressively increased the concentration of imidazole to 500 mM (50, 100, 200 mM). All elution fractions were collected and stored for SDS-PAGE analysis.

Purification of Strep-tagged proteins by affinity chromatography

Strep-purification is an affinity-based purification system where the desired protein is tagged with the Strep-tag II (WSHPQFEK). The frozen cell pellet is resuspended in Buffer W, lysed via passage through French press twice, ultracentrifuged at 35,000 rpm for 30 mins and then the supernatant taken as crude extract. 1 ml of a 50% Strep-Tactin sepharose resin (IBA Lifesciences) is loaded onto a Poly-Prep chromatography column (Bio-Rad Laboratories) and equilibrated with 5 ml of buffer W. After all the crude extract has passed through the column matrix, the resin is washed five times with 2.5 ml Buffer W. Strep-tagged proteins are finally eluted with Buffer E and collected in four 250 µl fractions. The eluted fractions are later analyzed by SDS-PAGE along with the pellet from the ultracentrifuge tubes, flowthrough, initial and final wash fractions.

Dialysis

To get rid of imidazole or desthiobiotin from the previous purification steps, the desired eluent fractions (based on SDS-PAGE analysis) were pooled and pipetted into Viva Spin® columns contains 5 or 30 kDa pore size dialysis membranes. Final volume was made to 15 ml with original buffer and the suspension spun down to ~2 ml. Twice more the suspension was made to 15 ml and centrifuged at 5000 rpm until 2 ml was left. This allowed for the imidazole or desthiobiotin to pass through the membrane but not the protein of interest. Eventually, purified protein in the original buffer was obtained. Alternatively, MEMBRA-CEL dialysis tubing

(SERVA) of appropriate pore size was used, against 1,000-fold Buffer W/ZAP (v/v) at 4°C, overnight with gentle stirring.

Size exclusion chromatography

To remove unwanted protein contamination and imidazole from the elutions of the recombinant proteins in appropriate buffer, the elution fractions were cleaned up by gel filtration using an Äktaprime plus system with a *HiLoad™ 16/600* and *26/600 Superdex™ 200* prep grade column attached. Elution from the column was measured via an inline absorption spectrometer and collected in 4 ml fractions.

Determination of protein concentration

This was done via Bradford assay wherein the formation of a complex between Coomassie brilliant blue G-250 and the proteins is measured at OD_{595} (Bradford, 1976). The protein concentration is determined relative to the slope of a standard BSA curve. Alternatively, for highly purified protein extracts, Nanodrop 1000® was used in conjunction with pre-determined molecular weight and extinction coefficient of the protein.

Pulldowns

A protein pull-down experiment is applied to verify protein-protein interactions *in vitro* (Louche et al., 2017). To analyse a hypothetical interaction, protein A (Strep-tagged) and B (His-tagged) are both overexpressed and purified in stable buffers. Then protein A is loaded onto its respective column matrix (Ni-NTA or Strep-Tactin). After washing several times until the wash fraction is free of protein, protein B is allowed to run over the column. Again, after washing until the flowthrough was free of protein content, protein A is now eluted from the column. Empty columns served as one negative control, whereas to guard against protein B interacting with the column on its own, a second control with only protein B is also performed. Pure protein samples and eluate fraction are analysed via SDS-PAGE.

Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse proteins based on their molecular weight (Laemmli, 1970). 8–15 % (v/v) Rotiphorese Gel 30 was used to prepare running gels depending on the expected protein mass. Denaturation of the protein samples was performed by boiling in SDS loading dye for 30 min at 95°C and loaded onto the prepared gel. PageRuler Plus Prestained Protein Ladder (Thermo

Fisher Scientific) was run on the same gel and served as a standard reference for estimation of protein size. The electrophoresis was carried out at 120-160 V in a tris-glycine buffer system until sufficient resolution of the protein ladder. Proteins were later visualized by fixation, Coomassie staining, silver staining or Western blotting depending on the project.

Coommassie Staining

Coomassie Brilliant Blue G-250 (Sigma-Aldrich Chemie) was used to visualize proteins separated using SDS-PAGE analysis (De St. Groth et al., 1963). The gels were incubated in colloidal or normal staining solution for 1 hr minimum at room temperature on a shaker at 50 rpm. Background staining of the gel was decreased by then incubating in de-staining solution until the desired contrast between the stained protein bands and background was reached. Stained gels were documented using the Molecular Imager Gel Doc XR+ system (Bio-Rad Laboratories).

Silver staining

Alternatively, when the concentrations of proteins were lower than 10 ng, or for fishing experiments (variation of a pulldown experiment using Protein A as bait and whole cell protein extract instead of Protein B), where the amount of protein that could bind to the bait (test protein) is unknown, silver staining was performed. However, it cannot be quantitatively interpreted (Winkler et al., 2007) as the level of staining strongly corresponds to the number of glutamate, aspartate and cysteine residues present in the proteins (they are reduced). The gels were stained as per the established protocol (Nesterenko et al., 1994).

Table 9: Silver staining protocol

Step	Solution	Time	Repeat
Fixing	Fixation solution	1 – 24 h	
Washing	EtOH (50%)	20 min	3x
Reduction	Thiosulfate solution	1.5 min	
Washing	dH_2O	20 s	3x
Staining	Impregnator	15 – 25 min	
Washing	dH_2O	20 s	3x
Development	Developer	Until sufficiently stained	
Washing	dH_2O	20 s	2x
Stop	Stop solution	5 min	

Western blotting

Protein detection in cell-free crude extracts, membrane and cytosolic protein fractions and purified protein elution fractions was performed by Western blotting (Towbin et al., 1979).

Protein concentration was first estimated, then the proteins fractions separated by SDS-PAGE, the stacking gel of the protein gels removed and lastly, they were shortly washed in transfer buffer. Immun-Blot polyvinylidene fluoride (PVDF) membrane (Bio-Rad Laboratories) were activated by incubating them in methanol for 5 min. A semi-dry blotting system (G&P Kunststofftechnik) was prepared by assembling three pieces of Whatman paper soaked in transfer buffer onto the anode of the blotting machine. The PVDF membrane was placed onto the Whatman paper, and the gel was placed on top of the membrane. Next, three additional pieces of Whatman paper soaked in transfer buffer were added on top of the gel and air bubbles were removed. The cathode was placed on top of the stack and the proteins were transferred from the gel to the membrane at 80 mA for 50 min. Unless otherwise stated, all following incubation steps were carried out at room temperature on a horizontal reciprocating shaker at 50 rpm. The membrane was washed in 50 ml Blotto 3 times for 30 min. After that, the Blotto was removed and the membrane was incubated at 4°C overnight in a diluted solution of the primary antibody in TBST.

The following day, the primary antibody was removed and the membrane was first washed in 50 ml of Blotto for 30 min and then twice in 50 ml TBST for 30 min. A 1:100,000 dilution of the secondary antibody in TBST was added to the membrane and it was incubated on a shaker for 30 min at RT. The secondary antibody was removed and the membrane was washed in 50 ml of TBST 3 times for 20 min. Then, the TBST was removed, the membrane was rinsed with dH2O and incubated in buffer III for 5 min to equilibrate the pH. Finally, 1 ml of 0.25 mM CDP-*Star* chemiluminescence substrate (Roche Diagnostics) was added on top of the membrane and it was put inside a flat plastic bag and imaged. Chemiluminescence emerging by substrate turnover was detected and documented using the ChemoCam Imager ECL (INTAS Science Imaging Instruments).

2.2.7 Whole cell crosslinking of Syn3A for protein-protein interactions

Cells were first grown as described in section 2.1. After ~40 hrs incubation cells were harvested from SP4 medium by centrifugation at 4,000 g for 15 min, followed by two washes with 10 ml HEPES buffer. Cells were pelleted once more, into pre-weighed tube, then wet cell mass determined via weighing. StageCL (patent pending) crosslinker was dissolved freshly in water-free Dimethyl sulfoxide (DMSO) at appropriate molar concentration. The crosslinker was quickly added to cells resuspended to a final concentration of 1 mg/ml cells in HEPES buffer (pH 7.8) at a final concentration of 0.25 – 200 mM (5% DMSO in final volume), depending on the experimental setup. The cells were incubated at 25°C for 30 min under gentle agitation, followed by quenching with a final concentration of 50 mM ammonium bicarbonate (pH 7.5) for 15 min. Cells were pelleted and snap-frozen with liquid nitrogen before storing at -80°C.

Further processing steps were mostly performed by Dr. Adam Belsom at the lab of Prof. Juri Rappsillber at the Technical University, Berlin. Further steps involved cell lysis and proteolysis for LC-MS analysis via an adapted SPEED protocol (Doellinger et al., 2020), peptide fractionation by strong cation exchange chromatography (SCX), size-exclusion chromatography (SEC), and hydrophilic strong anionic exchange chromatography (hSAX). LC-MS/MS acquisition of these fractionated crosslinked peptide species were then performed.

2.2.8 Whole cell UV crosslinking for RNA-protein interactions of Syn3A

Cells were grown as described in section 2.1. After ~40 hrs incubation cells were harvested from SP4 medium by centrifugation at 4,000 g for 15 min, followed by three washes with 10 ml HEPES buffer. The final wash pellet was resuspended in 11.6 ml HEPES buffer. 5.8 ml of this cell suspension was spread on a sterile plastic or glass petri plate (16 x 92 mm), to a depth of 1 mm. Cells were then crosslinked on metal blocks precooled to -20°C. Crosslinking was performed either using an in-house UV lamp as source about 3 cm above the cells for 10 minutes or, using CL-1000 UVP Crosslinker (Analytik Jena) set to function at a UV energy setting of 1500 x 10^2 µJ/cm^2. The crosslinked cells were then collected, pelleted, aliquoted at 2 mg/ml protein concentration (assumed protein content of cell is 10% of wet cell weight), and flash frozen in liquid nitrogen.

Further processing steps were performed by Luisa Welp or Dr. Aleksander Chernev, in the lab of Prof. Henning Urlaub, Max-Planck-Institute for Biophysical Chemistry. In short, cells were lysed using high urea concentration or sonication in Trizol reagent (Zymo Research®), proteins precipitated using acetone, and the RNA/DNA digested. This was followed by trypsin digestion, C18 desalting, TiO_2 cross-link enrichment and application to an LC-MS/MS acquisition pipeline.

3 RESULTS

3.1 Studies on the essentiality of NusA in *M. pneumoniae* due to its role in the 'Expressosome'

As in-cell CLMS had already indicated the presence of a linker between the RNAP and ribosome that could not be accounted for via NusG, attention shifted to NusA, another transcription factor (O'Reilly et al., 2020). Applying cryo-EM, a region of density between RNAP and the ribosome was determined that could only be explained by NusA. NusA was found to interact with RNAP via its N-terminal domain (NTD), and with the mRNA entry site of the ribosome via its C-terminal region. The C-terminal region in *Mycoplasmas* in comparison to NusA in *B. subtilis* has a disordered, low complexity region (Figure 8) that was hypothesized to serve as the anchor region between RNAP and the ribosome.

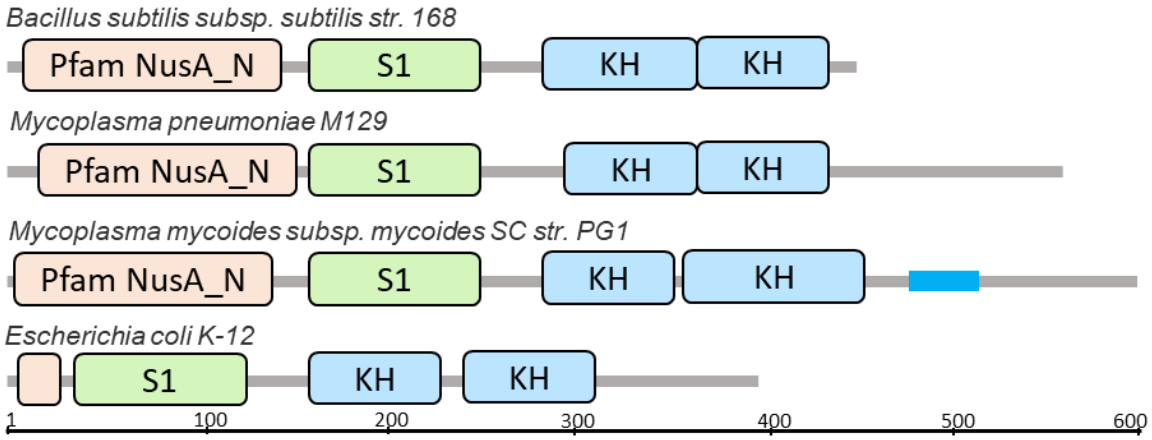


Figure 8: Domain annotation and arrangement for NusA accross *B. subtilis, M. pneumoniae, M. mycoides,* and *E. coli.* While each variant of NusA contains the same domains, the respective domains differ in size and *M. mycoides* contains an additional unknown domain.

To validate this hypothesis, the construction of deletion mutants lacking varying fragments of the C-terminal region of NusA was attempted. Based on the transposon insertion map for gene essentiality in *M. pneumoniae* (Lluch-Senar et al., 2015), two truncation variants were designed. One mutant lacked the C-terminal region after position I418 and the other was truncated after K440, as indicated with light blue spheres in Figure 9A.

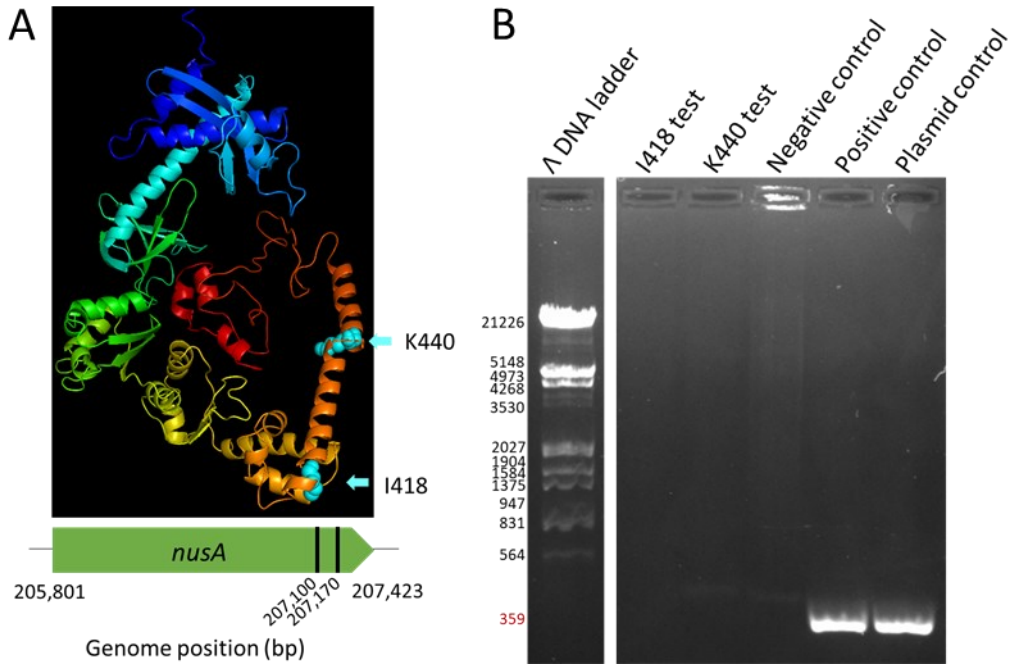


Figure 9: A. The structure of NusA. The structural model of NusA with the two points of termination in the experimental mutants highlight with light blue spheres. Corresponding locations in the genome are indicated**. B. Screening for C-terminal region mutants.** Picture of an electrophoresis gel to screen for PCR products from the chloramphenicol resistance gene. A product of 359 bp was synthesised from internal primers NS73 – NS74. From left to right: No PCR products were detected from passaged colonies of the desired mutants (I418, K440) indicating that the resistance marker is absent, and hence the truncation event has not taken place. Negative control- GP35 recombinase parent strain (GPM116) showed no PCR product. Both positive controls, a control deletion of *mpn668* (inessential gene) and a plasmid-based chloramphenicol resistance gene showed a product of expected size.

After transformation of fresh *M. pneumoniae* GP 35 with single-stranded DNA fragments, individual colonies were screened for the desired truncation mutants. As the number of transformant colonies generated was very low, this experiment was repeated several times initially. Later, mass testing of scrapings from all MP agar plates was undertaken to reduce the risk of any unpicked mutants.

Screening was performed by PCR testing for the product of the chloramphenicol resistance gene. Primers that bound internally to the *cat* gene were used to generate products of 359 bp. Two positive controls in the form of a deletion strain for MPN_668 or OsmC, an inessential gene, and a plasmid containing the *cat* gene were used. The *osmC* gene was freshly deleted during a transformation process parallel to the generation of these truncation mutants. As negative control, the original strain used for transformation, GPM 116, which is *M. pneumoniae* M129 harbouring a GP 35 recombinase that was constantly selected for by the antibiotic puromycin was used. The two truncation mutants yielded no check PCR product, and the negative control as well (Figure 9B). The positive control to form a deletion mutant of OsmC, yielded a product equal to that of resistance plasmid. Thus, it was not possible to isolate any truncated mutants of either the I418 or the K440 variant.

3.2 A whole cell protein-protein crosslinking study using a novel crosslinker

3.2.1 Establishing JCVI-*Syn3A* growth conditions

As previous user experience with crosslinkers suggested (the crosslinking of *M. pneumoniae* used commercially available crosslinkers DSS and DSSO), a new crosslinking experimental setup first necessitated the development of proper growth conditions of the bacteria and second, dosage concentrations of the crosslinker to be used. It was first tried to grow Syn3A in MP medium, as this was the medium previously in use for *M. pneumoniae*. However, growth in this medium was not satisfactory because the cells either lysed quickly upon handling or formed cellular floccules. MP medium contains PPLO broth, HEPES, a pH indicator, NaOH and heat inactivated horse serum. Correspondence with other labs revealed horse serum, a component of MP medium, likely to be responsible. Instead, SP4 medium was used with the inclusion of heat inactivated foetal bovine serum instead of Knockout™- a synthetic serum substitute due to easier availability and cost. SP4 additionally contains L-glutamine, sodium bicarbonate, yeastolate and CMRL 1066 which is normally used in cell culture.

Previously, growth of Syn3A was determined via a DNA quantification method that involved cell lysis, purification of DNA and hybridization of total DNA to a fluorescent probe (Dabrazhynetskaya et al., 2013). Crosslinking involves the harvesting of cells while still in log phase, to gain a snapshot of the cell in its most active phase as compared to one of stable dormancy. Further, the ability of the crosslinker to permeate cell membranes is reduced when stationary phase cultures are used, likely due to clumping of cell mass (Belsom & Rappsilber, 2021). Hence, a convenient means to accurately monitor cell growth was sought. Doubling times were studied first via manual and then plate reader growth curves, with a Syn3 strain expressing mCherry fluorescent dye (Figure 10). mCherry is excited at 587 nm and emits at 610 nm.

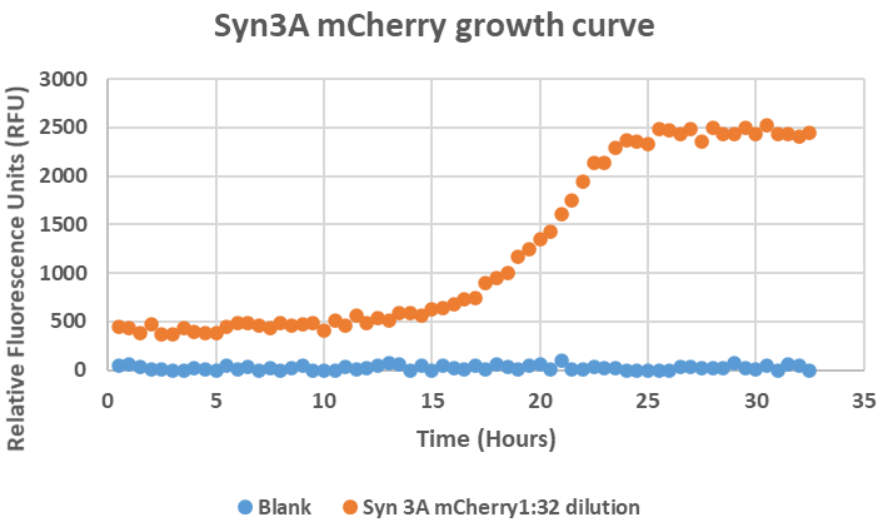


Figure 10: Fluorescent dye-based growth curve of Syn3A in SP4 medium. Doubling time was roughly 20 hours.

In strains that did not harbour mCherry, which would be used for crosslinking later, experiments revealed that growth monitoring at 495 nm was most ideal (Figure 11). The nominal 600 nm was not ideal because of the small size of the cells and due to the incorporation of a dye in SP4 medium.

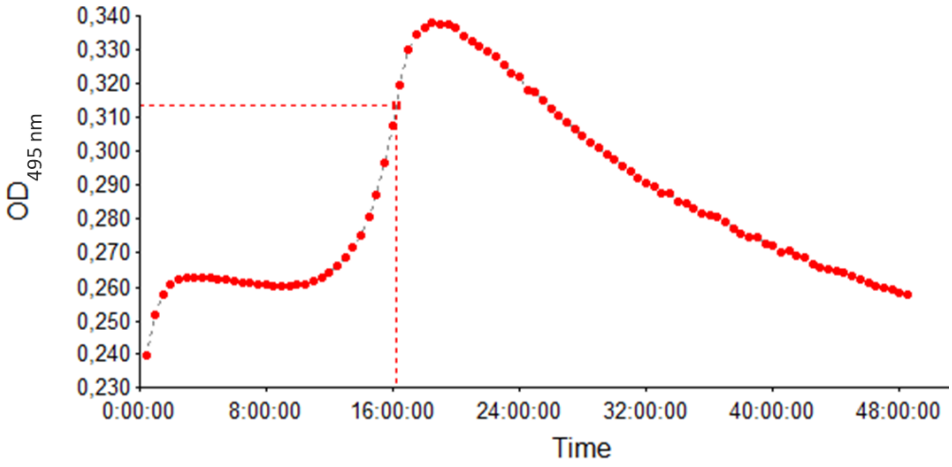


Figure 11: Syn3A growth curve measured at 495 nm. Rough doubling time 16 hours when grown at 37°C without constant shaking.

Second, harvesting of cells involved use of an appropriate buffer system. Previously, phosphate buffered saline was always used with *M. pneumoniae* cultures. Repeated experiments with Syn3 and Syn3A cells however revealed difficulties with handling of the cells. Cell pellets often became floccular and unstable even in the gentlest washing conditions. Further, after initial crosslinking these cells proved difficult to resuspend. After trial and error, PBS (pH=7.5) was replaced with 20 mM HEPES buffer (pH=7.8). The use of HEPES, was also

found to be compatible with the crosslinker used, and the issue of cellular resuspension in downstream solvents (Trizol, for RNA-binding proteins) was resolved.

3.2.2 Crosslinking of Syn3A

Syn3A was selected for crosslinking of the synthetic cell because it was the most stable strain to work with. Syn3 was particularly susceptible to lysis while handling and had a longer doubling time. To determine the most appropriate ratio of crosslinker to cells, initial titration experiments were performed. Fixed amounts of cells were crosslinked across a range of StageCL concentrations as depicted in Figure 12.

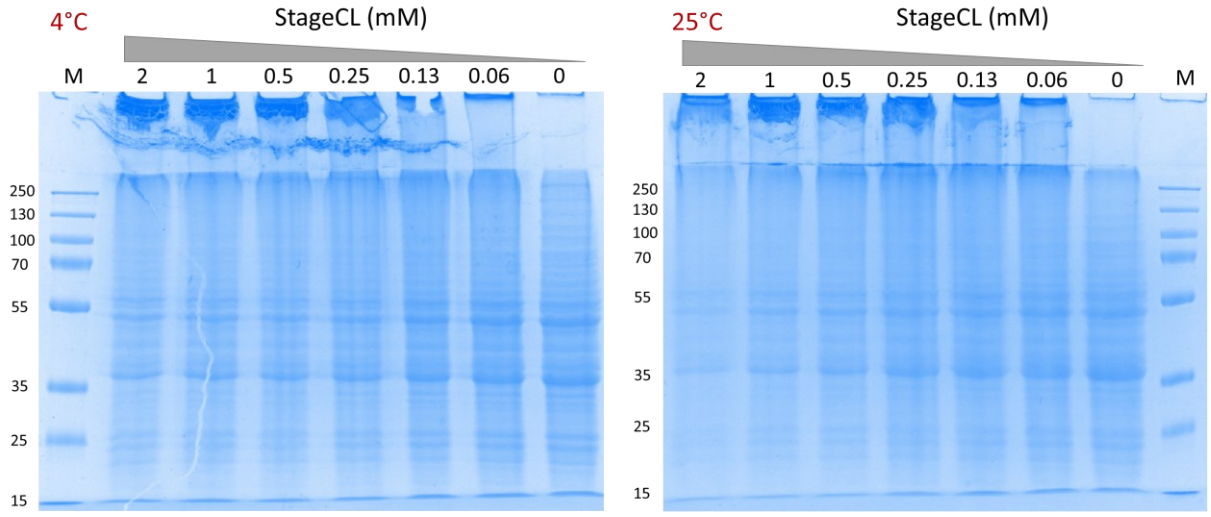


Figure 12: Titration of Cells vs Crosslinker. SDS-PAGE gels after Coomassie staining. From left to right, decreasing concentrations of StageCL is used to crosslink Syn3A cultures.

As the crosslinking reaction is favoured at higher temperatures but the preservation of the native cell state prefers cooled temperatures, the effect of temperature on the crosslinker performance and ability to cross the cell membrane had to be ascertained. Hence, two different temperature conditions, 4°C and 25°C were tested. The gels (Figure 12) indicate that crosslinking was so prevalent at higher concentrations of 2-0.5 mM that significant protein matter was still stuck in the wells of the gel. At lower concentrations distinct bands of increasing intensity are visible. Ultimately, a concentration of 0.25 mM StageCL at a reaction temperature of 25°C was selected for large-scale crosslinking.

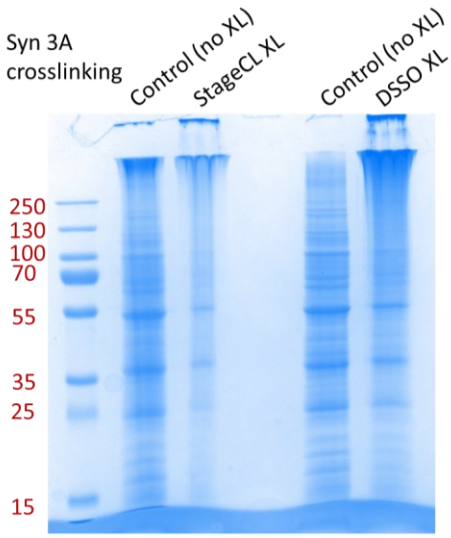


Figure 13: Test samples of large-scale crosslinking of Syn3A using StageCL and DSSO. Crosslinking of Syn3A was verified using a small test sample. Crosslinking with DSSO was applied as a control. Deeper penetration of the crosslinker into the cell and therefore more crosslinking is implied by fewer, more intense bands.

Assuming protein content to be 10% of wet cell mass, cells were adjusted to a concentration of 10 mg/ml and over a course of multiple experiments around 6 grams of cells were crosslinked. As a crosslinked control, the MS-cleavable crosslinker DSSO was also used on a smaller subset of cells. Figure 13 shows a non-crosslinked and crosslinked sample using StageCL, with much better crosslinking efficiency at the same concentration than DSSO. Processing of these samples would later be carried out by Dr. Adam Belsom.

3.2.3 Crosslink-based protein interaction map of Syn3A

In order to characterize genes of unknown function in a synthetic minimal cell a whole cell crosslinking mass spectrometric approach was used to generate a map of the interacting proteome. Figure 14 shows the most recent iteration of this currently active project (Map 17048).

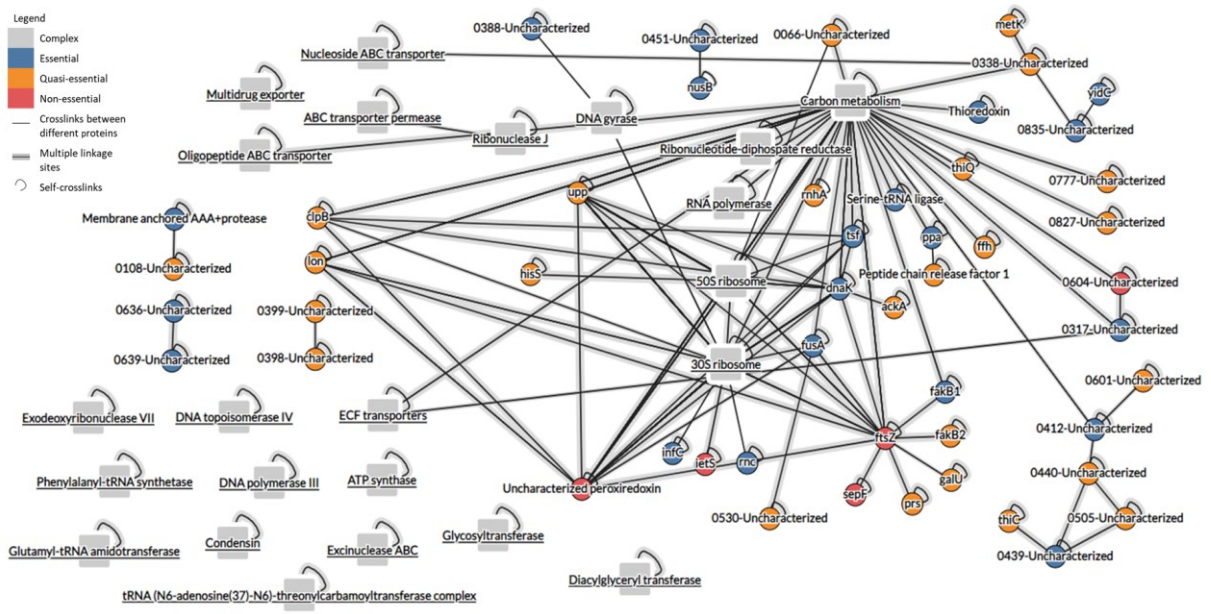


Figure 14: Protein-protein interactome map. Syn3A proteins arranged according to interacting complexes. Uncharacterized proteins that interact with any of the complexes or known proteins are indicated. All proteins are colour coded based on their essentiality status (consult legend). Ambiguous crosslinks between different proteins have been removed to allow for clearer visualization of the interacting proteins. GapA and TufA have also been excluded as they form numerous linkages to many proteins. Transporter related proteins are grouped in the upper left corner, complexes with crosslinks within themselves are collected in the lower left corner. Core cellular functions around which several known and unknown proteins seem to act are arranged in the centre. Uncharacterized proteins are arranged on the right.

We identified a total of 21,317 crosslinks of which 15,316 were cross-links involving the same protein (self-links). A further 6,001 were heteromeric crosslinks (involving two different proteins) with a 5% residue-pair false discovery rate (FDR). These represented 643 distinct protein-protein interactions (PPIs) at 1% PPI FDR. 311 of these identified crosslinks involved uncharacterized proteins and in all 28 uncharacterized proteins were indicated as interacting partners to either another uncharacterized protein or an annotated protein. A further 1652 self-links are spread amongst 80 uncharacterized proteins.

3.2.4 Validation of the interactome map

Several different known and expected complexes could already be recognized via the extensive crosslinking between their subunits. For example, as depicted in Figure 15, DNA gyrase A and B are linked to each other and two other interaction partners. RNA polymerase subunits α, β, β', and δ (A, B, C, and D) were also found in a complex.

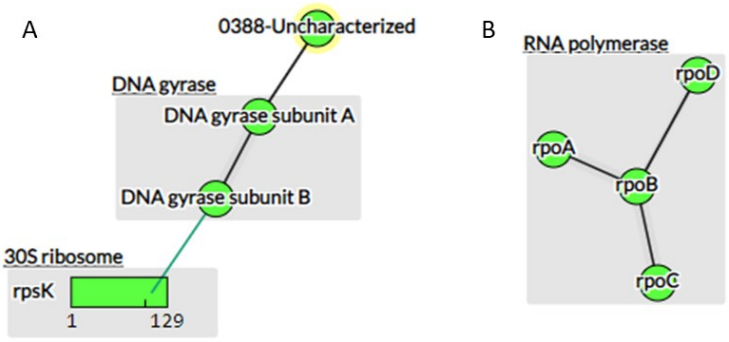


Figure 15: Known complexes from the PPI map. DNA gyrase subunits A and B are linked as well as RNA polymerase subunits α, β, β', and δ (A, B, C, and D). DNA gyrase also links to an essential, uncharacterized protein 0388, and a 30S ribosomal protein RpsK.

Focusing further on crosslinks between individual proteins that form known super-complexes, the 30S and 50S ribosomes are heavily crosslinked as depicted in Figure 16.

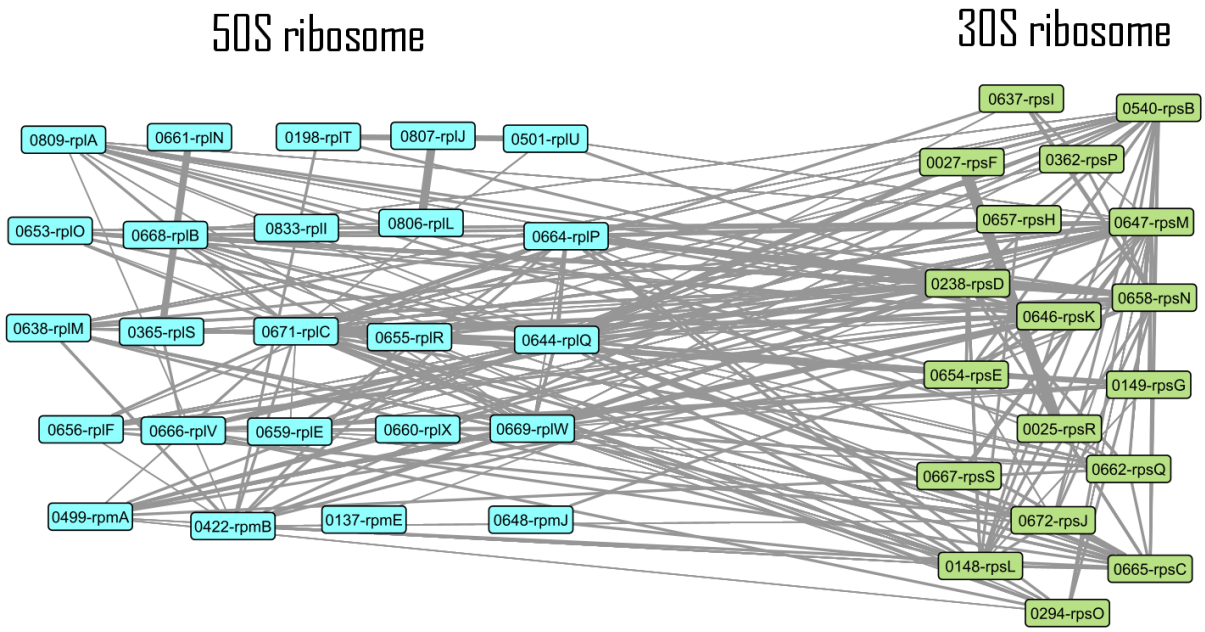


Figure 16: Crosslinking between 30S and 50S ribosomes. Extensive crosslinking between the ribosomal subunits covered most of the components.

In Syn3A, with regards to the 50S ribosome, 26 of the listed 30 proteins were found to be crosslinked and present in our map (Figure 16). RplD, RplJ, RpmC, RpmF, RpmH, and RpmI are missing from the crosslinked 50S ribosome. In the 30S ribosome 18 of the 20 ribosomal proteins are present in the crosslinked complex. RpsT and RpsU were not found to be crosslinked. Further, extensive crosslinking was observed between RplN - RplS, RplJ – RplI, RplP – RpsN, RplR – RpsE, and RpsF – RpsR. This provides quite good, almost complete

coverage of ribosomal components, which is one of the most intricate super-complexes in the cell.

As crosslinking data between different residues within a protein, and between different monomers of multimeric complexes was also available, reported crystal structures were used to verify the accuracy of our crosslinking data. As shown in Figure 17, the crosslinks could be fitted to the crystal structure of holo Glyceraldehyde-3-phosphate dehydrogenase 1 (GAPDH1) from methicillin resistant *Staphylococcus aureus* MRSA252. The purple lines show links between residues within one monomer, while the orange lines join residues from other monomers in the holoenzyme.

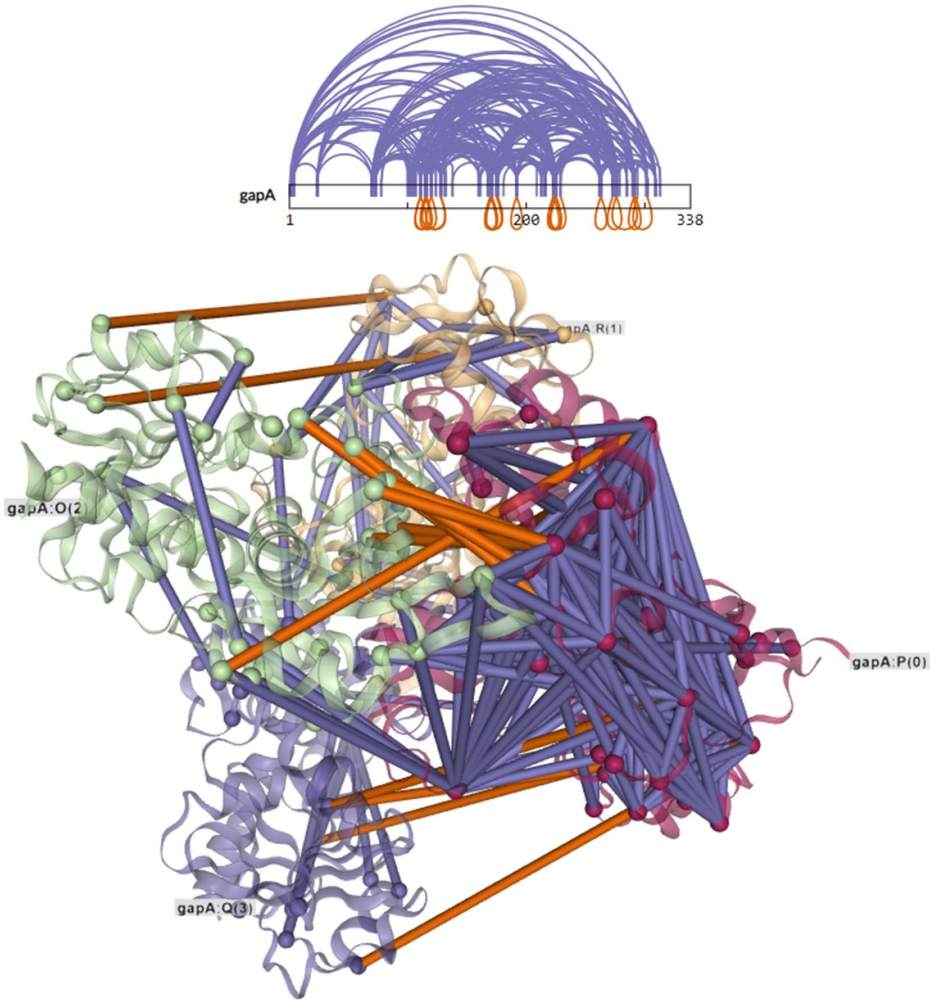


Figure 17: Crosslinks fitted to the structure of GapA. Crosslinks obtained for GapA are depicted in the upper representation. In purple are self-links, in orange are links between different monomers of the multimeric holoenzyme. In the lower image, these crosslinks were fitted to the reported crystal structure of holo Glyceraldehyde-3-phosphate dehydrogenase 1 (GAPDH1) from methicillin resistant *Staphylococcus aureus* MRSA252 (PDB ID: 3LVF). 273 crosslinks are depicted in full and 21 in part, of 294 filtered crosslinks for this protein. Again, in purple are self-links within one monomeric subunit, orange lines link residues from different subunits within the multimer.

Of the 80 uncharacterized proteins that had crosslinking data, 28 uncharacterized proteins that were found to interact with at least one other protein. Based on when the interactions became available, work on characterizing some of these proteins was started. The more promising candidates are described in the following sections.

3.2.5 Efforts to characterize Syn3A_0439, Syn3A_440 and Syn3A_505

Figure 18: A six protein complex with a central triad of uncharacterized proteins. Proteins Syn3A_0439, 440, and 505 form the central core of this complex. Thiamine substrate-binding protein ThiC is also linked to Syn3A_0439. Syn3A_0412 is linked to 440 and 0601.

Figure 18 shows a 6-protein complex that consists of three crosslinked proteins of unknown function and 3 neighbours that are less crosslinked. Uncharacterized proteins 0439, 0440 and 0505 are crosslinked to each other the most with crosslinked residues throughout 0439, in the C-terminal half of 0505 and in the first and last quarter of 0440. Syn3A_0439 is an essential, uncharacterized lipoprotein of 80 kDa size and 374 copies per cell. Syn3A_0505 (40 kDa, 252 copies) and Syn3A_0440 (108 kDa, 409 copies) are both quasi-essential uncharacterized lipoproteins with no discernible domain architecture outside of an N-terminal signal peptide region. Syn3A_0439 shows a similar lack of recognizable domains. Syn3A_0439 has no homologs outside of the Mollicutes, Syn3A_0440 is the same but it does show some similarity, but low identity, to LytF from *B. subtilis*. Syn3A_0505 on the other hand has more apparent homologs, among them PspA from *S. pneumoniae*, and YcgV from *E. coli.* However, self-crosslinks from Syn3A_0505 were unable to be matched to the predicted structure of either of these proteins. Further, 0505 has no reported interaction partners in the STRING database.

Syn3A_0412 (159 kDa, 216 copies, essential) is crosslinked once to 0440 and once to 0601 (at the C-terminus). 0412 has 11 transmembrane domains and thought to be similar to

the preprotein translocase ATPase subunit. SYN3A_0601 is a 36 kDa quasi-essential protein that has about 73 copies per cell. On the other side, ThiC (54 kdA, 211 copies) is a quasi-essential protein involved in thiamidine specific binding and ABC transport. It seems to exist only within the Mycoplasmas. Lastly, SYN3A_0439 is paralogous to SYN3A_0851 (75% overlap), SYN3A_440 is paralogous to SYN3A_0599 and SYN3A_0412 has a 50% overlap with SYN3A_0315. Of these six proteins, Syn3A_0439 was also later found to be an RNA-binding protein.

Cloning and overexpression of Syn3A_0439, 0440, 0505

Given the high degree of crosslinking between these three proteins and their uncharacterized status, the possibility of an uncharacterized protein complex was high. Hence, these three proteins were tried for purification as recombinant proteins, with binding and characterization assays in mind. Recombinant versions of these proteins were designed as depicted in Figure 19.

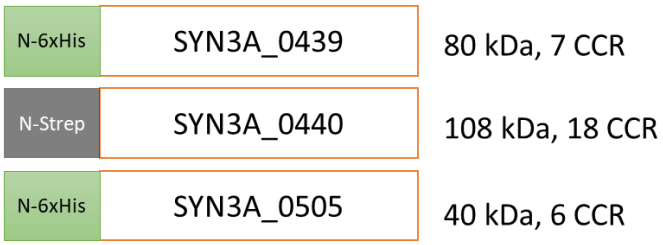


Figure 19: Scheme of recombinant constructs of SYN3A_0439, 0440, 0505. His- and strep-tagged variants of the proteins were designed for characterization studies.

To enable expression of these proteins in *E. coli*, multiple rounds of codon optimization were first completed. As stated in Figure 19 the number of CCRs implies the number of TGA to TGG codon switches that were required. TGA codes for Tryptophan in Mycoplasmas but, is a stop codon in *E. coli*. After cloning into their respective overexpression plasmids and transformation of over-expression strains BL21(DE3), C43(DE3) and Rosetta over-expression experiments were carried out several times.

The cloning construct for Strep-440 was the first to be completed as the gene was commercially synthesized. Test- expression in the three *E. coli* strains revealed a significant band only for Rosetta strain, but at the height of ~60 kDa instead of 80 kDa. Subsequent over-expressions failed to replicate this result. The mutagenesis reactions for the remaining two genes were not trivial. The process of creating a TGA-free version of Syn3A_0505 is complete. However, efforts to convert all TGA to TGG in Syn3A_439 have been hampered by a rearrangement of the N-terminal region of the gene during the cloning process. Whether or not

this is due to toxicity of the insert in *E. coli* is yet to be determined. Thus, the production of recombinant versions of these proteins has so far not been successful.

Structural prediction and validation of Syn3A_0439

In order to better understand the nature of the protein under consideration attempts at structural prediction were undertaken. Several protein prediction tools were considered Alphafold2 is a protein structural prediction algorithm that recently released protein structures for several more organisms (Jumper et al., 2021). 170 predicted structures were released for *Mycoplasma mycoides* subsp. mycoides SC (strain PG1), the parental strain of Syn3A. 169 structures were for proteins that were characterized while one (Uniprot-ID: P55802) was for an uncharacterized lipoprotein that was deleted during genome minimization and is not present in Syn3A. While this predicted structure was not reported by their team in any of their releases, another program, Colabfold, allows a more user friendly approach and faster access to the same algorithm with minimal trade-offs (Mirdita et al., 2021). After setting up Colabfold on the server, it was able to generate several models, the best of which is shown in the lower part of Figure 20.

Predicted models are often based of structural homology and sequence alignments. Hence, it is useful to have data to validate such a structure. The self-crosslinking data we had already obtained from the CLMS study could be used towards this validation step. In the upper image in Figure 20 we see the overall spread of self-links in the protein. There were no homomeric crosslinks reported i.e crosslinks between different monomeric units of a multimeric protein assembly. All 74 crosslinks matched to this protein could be fit to this structure. Further, as shown in Figure 21 the measure of the crosslink Cα to Cα distance was clustered around 10 Å (shown in pink). Shown in grey bars are a random distribution of crosslinks, which would have been the case if the predicted structure did not match the crosslinking data.

Figure 20: Predicted structure of SYN3A_0439 with fitted crosslinks from CLMS study. Model used: 439_cc0a9_unrelaxed_model_1.pdb, showing 74 in full of 74 filtered TT crosslinks.

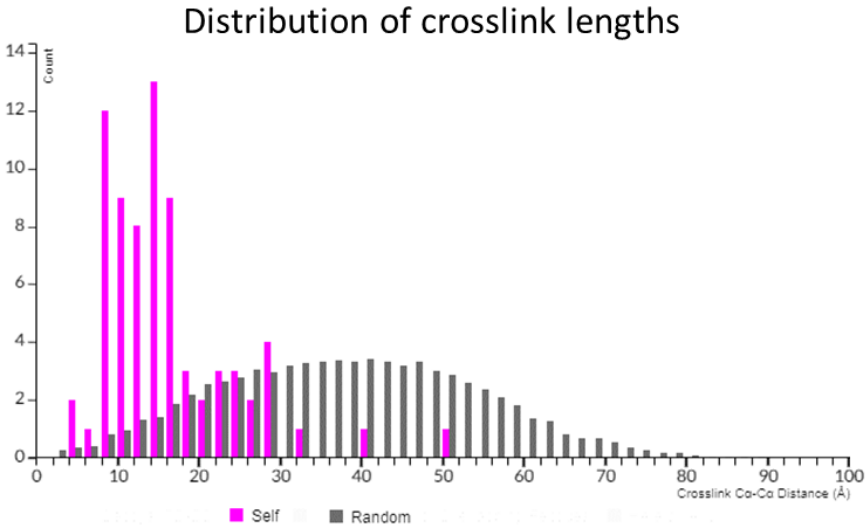


Figure 21: Distribution of the crosslinks as per their Cα - Cα distance. Self-crosslinks are displayed in pink, random spread of crosslinks in grey.

Thus, the predicted model for Syn3A could be verified with our crosslinking data. Efforts to replicate this success with the other proteins are hampered at the moment by the larger size of Syn3A_0440 (110 kDa) which requires the protein to most likely be broken into multiple parts and each modelled separately.

3.2.6 Other uncharacterized proteins of note

Three other complexes with proteins of unknown function are worth mentioning because they were all found to contain RNA-binding sites (experimental data presented in the next section) and efforts to characterize some of these candidates has already begun.

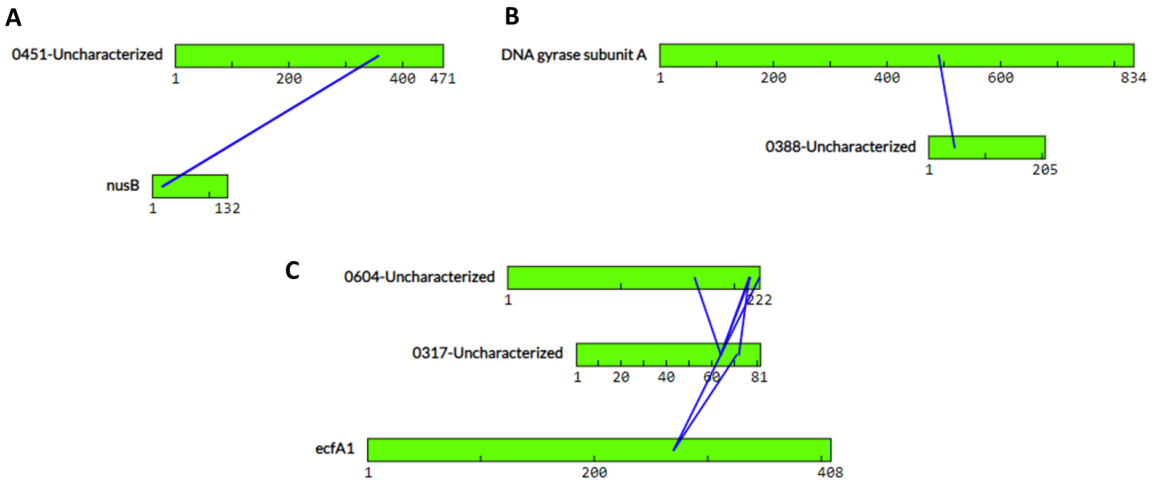


Figure 22: Complexes with one annotated protein. A. Syn3A_0451 shows 6 matches for a crosslink to NusB. **B.** Syn3A_0388 shows 5 matches for a link to GyrA. **C.** Syn3A_0317 has14 matches to four crosslinks with Syn3A_0604 and 4 matches to two crosslinks with ecfA1.

Figure 22 shows three different complexes wherein at least one interacting partner is annotated. In complex A, NusB, a known transcription antitermination factor of 15 kDa is shown to interact with JCVISYN3A_0451, a 51 kDa essential uncharacterized protein that is homologous to glyceraldehyde-3-phosphate dehydrogenase (GapN). NusB is thought to have 18 protein copies per cell, compared to 0451 or GapN' 248.

In complex B, DNA gyrase A (94 kDa, 299 copies per cell) interacts with JCVISYN3A_0388 (24 kDa, 87 copies), an essential protein with only one homolog in *Mesoplasma florum*, also uncharacterized.

Complex C consists of three proteins, JCVISYN3A_0604 (25 kDa, 346 copies), JCVISYN3A_0317 (8 kDa, 20 copies) and the ATP-binding A1 component of the energy-coupling factor (ECF) type transporter or ecfA1 (46 kDa). 0604 is non-essential while the other two are essential. 0604 has few known homologs, among them putative uncharacterized

members of the LemA family from *A. laidlawii* and *L. monocytogenes.* 3 crosslinks were found between the C-terminal halves of both 0604 and 0317. JCVISYN3A_0317, is essential like its uncharacterized homolog in *B. subtilis* YneF. Both 0604 and 0317 contain a single transmembrane domain found towards the N-terminal half of the protein. Two crosslinks link 0317 and ecfA1, with the crosslinked residues from 0317 being the same lysine residues (64K, 71K) that link to 0604. Efforts to validate these interactions via *in-vitro* interaction assays are underway (unpublished master thesis of Camilo Torres).

3.3 A study of the RNA-binding proteins in Syn3A

Given the conservation of ribonucleotide machinery within the cell, in comparison to parental wild type *M. mycoides*, it was hypothesized that this model organism must use RNA based gene or protein regulation to an increased extent. This is due to the fact that during the artificial genome reduction process several genes involved in gene regulation were likely marked as non-essential and hence deleted. Thus, this minimal cell could shed light on hitherto unexplored RNA-binding proteins as well as assign secondary functions for those that are already annotated.

To begin, a collaborative study with the research group of Professor Urlaub was undertaken to use whole cell UV crosslinking to crosslink nucleic acid species to their interacting proteins, and process these cells using an established mass spectrometry-based pipeline (Kramer et al., 2014). In the first approach, a set of experiments were setup to crosslink nucleic acids and proteins in live cells at 254 nm. A summary of the findings is depicted in Table 10.

Table 10: An overview of the RNA-binding proteins in SYN3A

Proteins in Syn3A	**452**
Annotated proteins that bind RNA	**122**
Peptides matched to annotated proteins	516
Annotated proteins that are not classified as RNA-binding protein	**19**
Peptides matched to annotated unclassified RNA-binding proteins	26
Uncharacterized proteins	**20**
Peptides matched to uncharacterized proteins	38

Two different types of experiments were planned which would yield complementary data. In the first type of experiment, termed ‘RNA-binding protein localization study’ log phase cells were harvested, crosslinked at UV-254 nm and then the samples processed as depicted in Figure 6. This yielded information on the proteins that bound RNA and identified a sequence of 5-20 amino acids where the interaction took place. In most cases, the exact amino acid

residue that interacts with RNA was also determined (based on the quality of the spectra and manual curation). Processing involved cell lysis, digestion of peptides and DNA, enrichment with TiO_2 or SiO_2, C18 chromatography and then analysis of the resultant RNA-peptide complexes via MS. The search engine used to analyze these results was RNP^{xl}, a dedicated data analysis workflow based on the OpenMS platform (Kohlbacher et al., 2007; Sturm et al., 2008). These results were calculated with a 1% false discovery rate (FDR).

The results from this proteome wide screening allowed the identification of a total of 161 RNA-binding protein interactions. These 161 proteins were searched against a database (RBP2GO) of reported RNA-binding proteins in *E. coli.* 122 proteins were found to be known RNA-binding proteins. Their interaction and functions can be described as well as their RNA-binding motifs predicted, if not identified. These proteins are grouped according to their cellular function and listed in Table 11.

Table 11: List of annotated proteins already known to be RNA-binding

Carbon core metabolism	**50S ribosome**	**30S Ribosome**	**Translation**	**Unassigned**
0131-fbaA	0137-rpmE	0025-rpsR	0061-serS	0065-trxA
0213-eno	0198-rplT	0027-rpsF	0064-lysS	0095-secA
0221-pyk	0199-rpmI	0082-rpsT	0150-fusA	0115-galU
0227-pdhC	0365-rplS	0148-rpsL	0151-tufA	0168-oppF
0234-crr	0422-rpmB	0149-rpsG	0200-infC	0300-nusA
0262-rpe	0499-rpmA	0238-rpsD	0202-rsmD	0360-ffh
0475-ldh	0501-rplU	0294-rpsO	0240-thiI	0378-nadE
0606-pgk	0638-rplM	0362-rpsP	0263-cpgA	0381-mtnN
0607-gapA	0644-rplQ	0482-rpsU	0287-aspS	0407-rpoD
0729-pgm	0648-rpmJ	0540-rpsB	0289-rbfA	0427-pstB
0779-ptsG	0653-rplO	0637-rpsI	0308-trpRS	0522-ftsZ
	0655-rplR	0646-rpsK	0329-rluB	0523-ftsA
DNA related	0656-rplF	0647-rpsM	0348-engA	0542-dnaK
0001-dnaA	0659-rplE	0654-rpsE	0361-rlmH	0543-grpE
0006-gyrB	0660-rplX	0657-rpsH	0390-fmt	0645-rpoA
0007-gyrA	0661-rplN	0658-rpsN	0434-rlmFO	0793-atpH
0097-exoR	0663-rpmC	0662-rpsQ	0519-ileS	0804-rpoB
0254-uvrC	0664-rplP	0665-rpsC	0528-pheT	
0406-dnaG	0666-rplV	0667-rpsS	0535-argS	
0452-parE	0668-rplB	0672-rpsJ	0539-tsf	
0453-parC	0669-rplW		0548-cspR	
0609-dnaB	0670-rplD		0640-truA	
0690-ligA	0671-rplC		0650-map	
	0806-rplL		0004-ksgA	
Nucleotide metabolism	0807-rplJ			
0045-tmk	0809-rplA	**RNA synthesis and degradation**		
0203-gmk	0810-rplK	0003-rnmV		
0330-dgk	0833-rplI	0257-rnjB		

0549-rgdB	0910-rpmH	0359-rny
0798-upp		0600-rnjA
0831-prs		

In each of these interactions the relevant peptide and in most cases the amino acid residue that directly interacts with RNA (mostly Uridine) was identified. In some of the proteins, the same peptide was found multiple times with the same or different XL positions while specifically for some proteins, like ribosomal proteins, multiple RNA localization sites were found throughout the protein. These detailed results can be found in the Supplements, Table 15.

This study revealed that 19 proteins that had not been described as RNA-binding proteins in *E. coli* were found to interact at atleast one location with RNA. More importantly, a further 20 uncharacterized proteins were found to bind RNA. These proteins are briefly listed in Table 12.

Table 12: New RNA-binding proteins.

Unclassified previously annotated proteins		**Uncharacterized proteins**	
0009-rnsC	0512-plsC	0030	0444
0169-oppA	0617-fakB2	0034	0451
0195-potC	0643-ecfA1	0060	0493
0264-prkC	0706-thiB	0138	0546
0303-polC	0787-mgtA	0317	0602
0305-papA	0817-whiA	0338	0636
0327-scpA	0822-ecfS3	0346	0827
0371-ywjA1	0887-cdr	0352	0835
0420-fakA	0908-yidC	0388	0852
0430-ylxM		0439	0878

These results provided us with a subset of RNA-binding proteins in the organism Syn3A. The next step was to experimentally verify this RNA-protein interaction. Rather than immediately perform a protein-by-protein RNA pulldown followed by RNA-sequencing analysis, a whole-cell approach was selected. In the second experiment, efforts were made to confirm the previous results using RNA-binding protein enrichment studies, first in two pilot studies with only technical triplicates and then in a scaled-up series of experiments, with biological and technical triplicates.

Log phase cells were harvested, split into two halves, one crosslinked at UV-254 nm and the other processed identically without UV crosslinking. The resulting peptides were analyzed on two different MS modes ion trap (IT) and Orbitrap (OT). The peptides in the form of MS data were processed via the RNP^{xl} pipeline and then crosslinked candidates were selected for spectrum curation. Results were validated by Dr. Chernev (Proteomics Service Facility, University of Goettingen). A setting of 1% FDR was set for this analysis.

The results of this set of enrichment experiments are listed in Table 13. Compared to the 122 known RNA-binding proteins from the previous study, 85 proteins were enriched and are indicated in blue.

Table 13: Enriched known RNA-binding proteins. 85 proteins could be experimentally confirmed as RNA-binding proteins and are indicated in blue.

Carbon core metabolism	**50S ribosome**	**30S Ribosome**	**Translation**	**Unassigned**
0131-fbaA	0137-rpmE	0025-rpsR	0061-serS	0065-trxA
0213-eno	0198-rplT	0027-rpsF	0064-lysS	0095-secA
0221-pyk	0199-rpmI	0082-rpsT	0150-fusA	0115-galU
0227-pdhC	0365-rplS	0148-rpsL	0151-tufA	0168-oppF
0234-crr	0422-rpmB	0149-rpsG	0200-infC	0300-nusA
0262-rpe	0499-rpmA	0238-rpsD	0202-rsmD	0360-ffh
0475-ldh	0501-rplU	0294-rpsO	0240-thiI	0378-nadE
0606-pgk	0638-rplM	0362-rpsP	0263-cpgA	0381-mtnN
0607-gapA	0644-rplQ	0482-rpsU	0287-aspS	0407-rpoD
0729-pgm	0648-rpmJ	0540-rpsB	0289-rbfA	0427-pstB
0779-ptsG	0653-rplO	0637-rpsI	0308-trpRS	0522-ftsZ
	0655-rplR	0646-rpsK	0329-rluB	0523-ftsA
DNA related	0656-rplF	0647-rpsM	0348-engA	0542-dnaK
0001-dnaA	0659-rplE	0654-rpsE	0361-rlmH	0543-grpE
0006-gyrB	0660-rplX	0657-rpsH	0390-fmt	0645-rpoA
0007-gyrA	0661-rplN	0658-rpsN	0434-rlmFO	0793-atpH
0097-exoR	0663-rpmC	0662-rpsQ	0519-ileS	0804-rpoB
0254-uvrC	0664-rplP	0665-rpsC	0528-pheT	0823-folC
0406-dnaG	0666-rplV	0667-rpsS	0535-argS	
0452-parE	0668-rplB	0672-rpsJ	0539-tsf	
0453-parC	0669-rplW		0548-cspR	
0609-dnaB	0670-rplD		0640-truA	
0690-ligA	0671-rplC		0650-map	
	0806-rplL		0004-ksgA	
Nucleotide metabolism	0807-rplJ			
0045-tmk	0809-rplA	**RNA synthesis and degradation**		
0203-gmk	0810-rplK	0003-rnmV		
0330-dgk	0833-rplI	0257-rnjB		
0549-rgdB	0910-rpmH	0359-rny		
0798-upp		0600-rnjA		
0831-prs				

Of the 19 unclassified RNA-binding proteins, 4 were enriched while for the uncharacterized proteins, 6 of 20 proteins were enriched (Table 14).

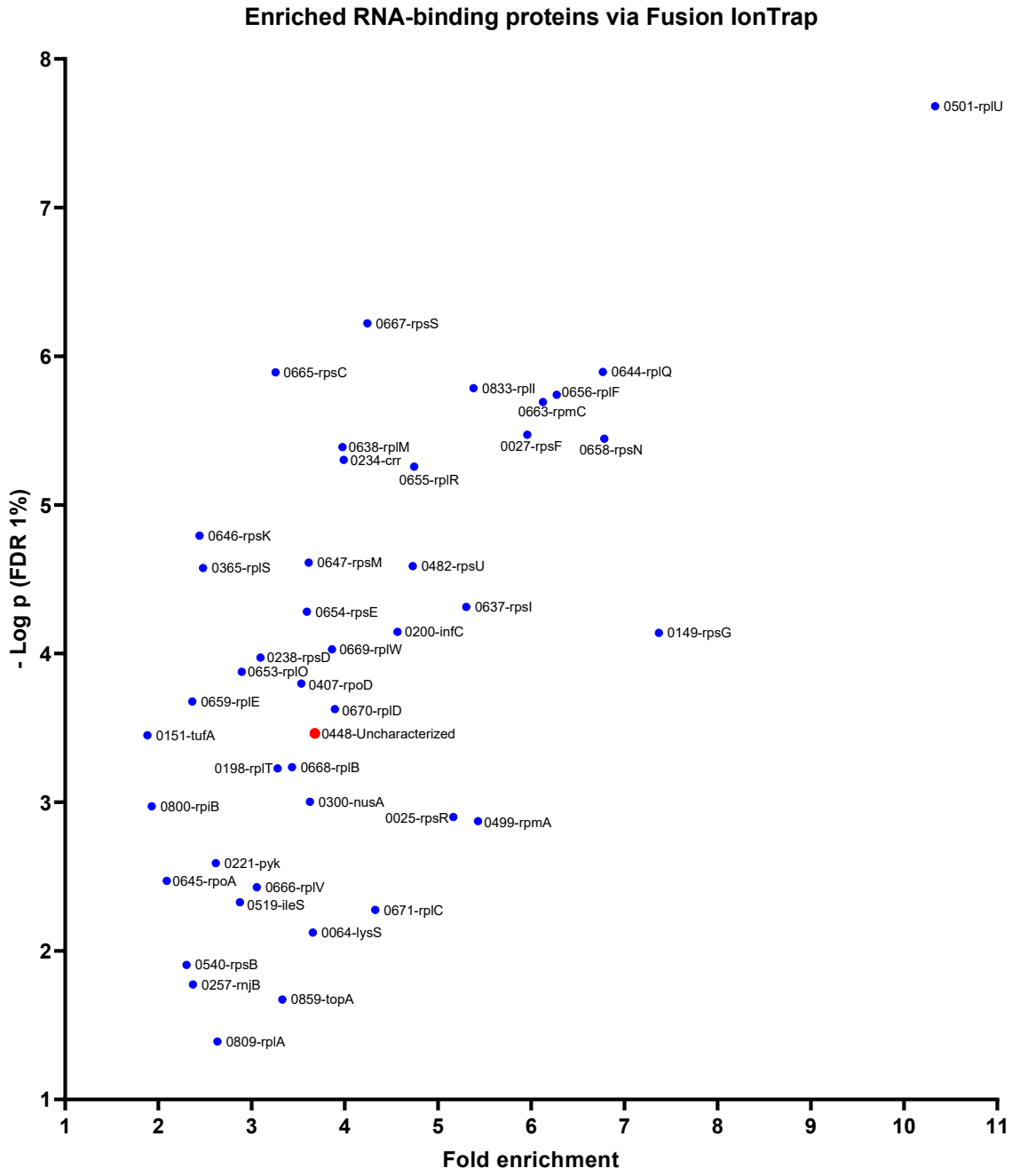
Table 14: Enriched new RNA-binding proteins.

Unclassified previously annotated proteins		Uncharacterized proteins	
0009-rnsC	0512-plsC	0030	0444
0169-oppA	0617-fakB2	0034	0451
0195-potC	0643-ecfA1	0060	0493
0264-prkC	0706-thiB	0138	0546
0303-polC	0787-mgtA	0317	0602
0305-papA	0817-whiA	0338	0636
0327-scpA	0822-ecfS3	0346	0827
0371-ywjA1	0887-cdr	0352	0835
0420-fakA	0908-yidC	0388	0852
0430-ylxM		0439	0878

Besides confirming the RNA-binding ability of these proteins, this experimental approach also indicated the abundance of these RNA-protein interactions in the cell via fold-increase levels under crosslinked and non-crosslinked conditions.

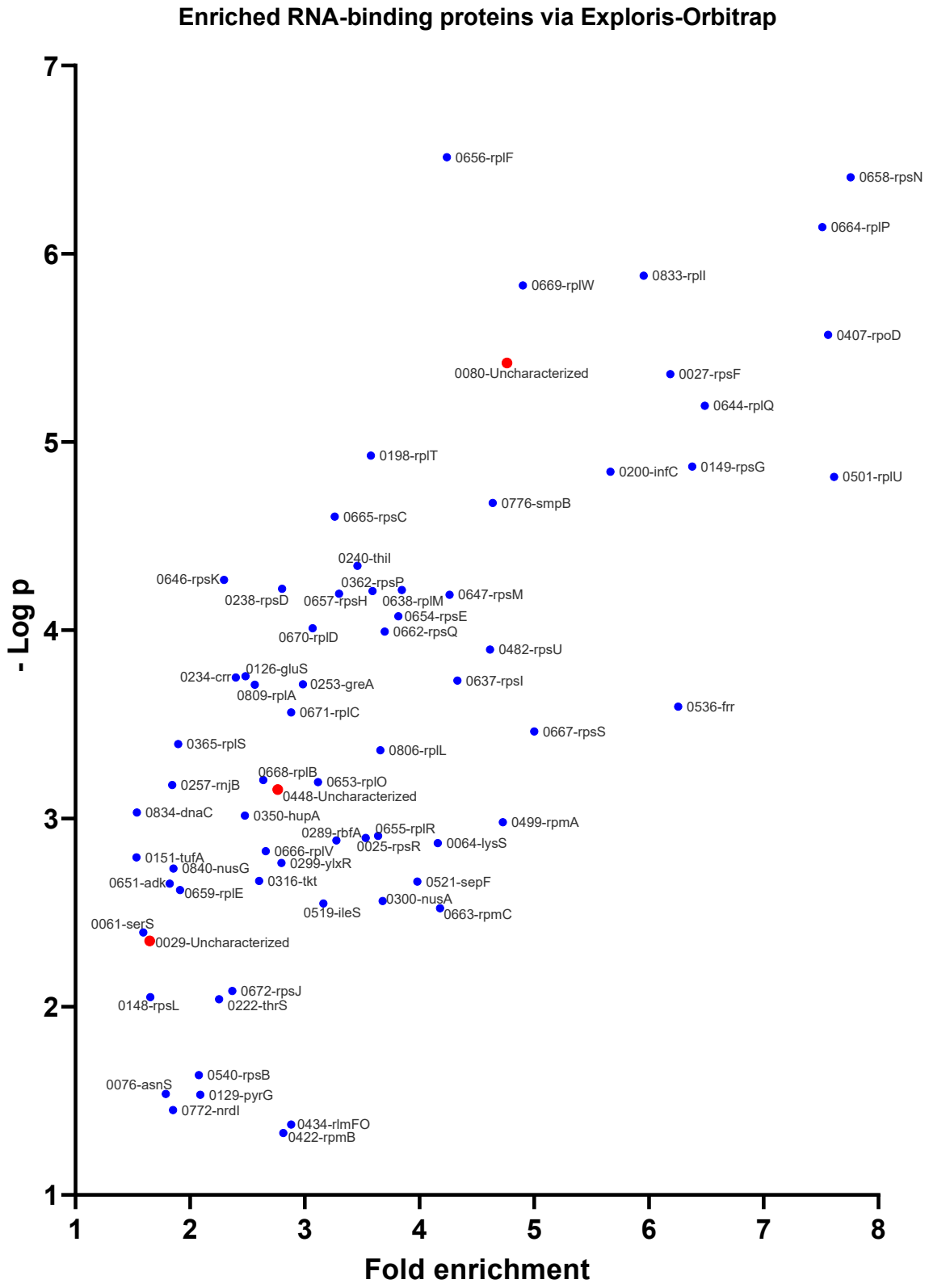
In the four Plots below, the proteins that were enriched in the crosslinked sample have been plotted below against their -log P values. Uncharacterized proteins that were enriched in the crosslinked samples are marked with red dots. Where possible the protein-locus IDs are displayed. The fold enrichment is indicative of the abundance of the RNA-protein heterocomplex and therefore the abundance of these proteins in the cell. Hence, there could be a tendency for RNA-binding proteins that are more abundant in the cell to be confirmed.

Plot 1: Enriched proteins from the large-scale study, MS mode: Fusion IonTrap



In Plot 1 proteins that were found to be enriched using the Fusion-Ion Trap mode on the MS are depicted. As expected, a wide variety of ribosomal proteins were captured. Syn3A_0448 is the only uncharacterized protein to be enriched more than 3-fold. In total 44 proteins were found enriched using Fu-IT mode.

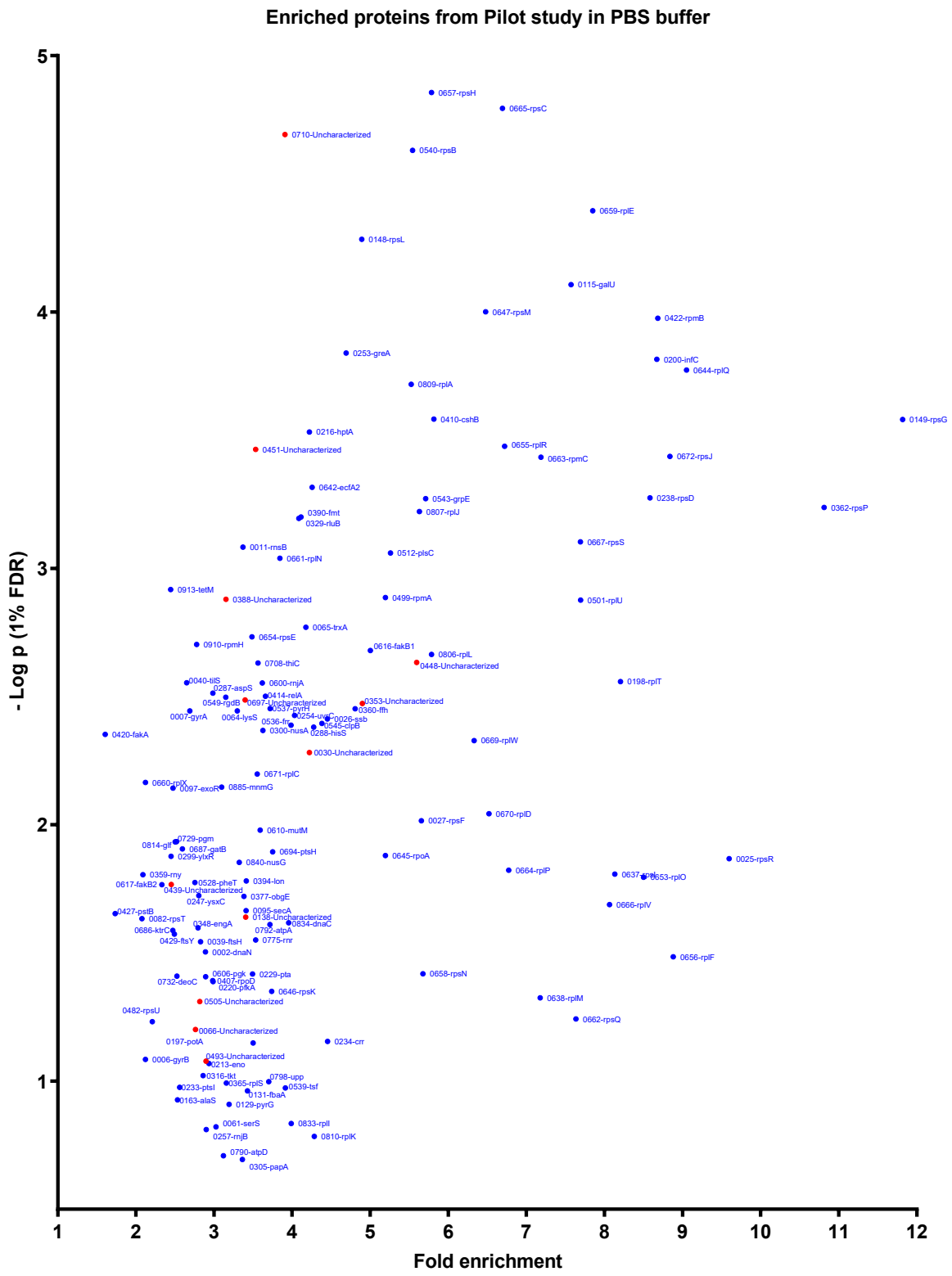
Plot 2: Enriched proteins from the large-scale study, MS mode: Orbitrap



In Plot 2 Syn3A_0448 is again enriched, but also, SYN3A_0029 and SYN3A_0080. 69 proteins were found enriched using OrbiTrap mode. NusA was found to be enriched in both these modes. Further, in the pilot studies different buffers, phosphate buffered saline and killing

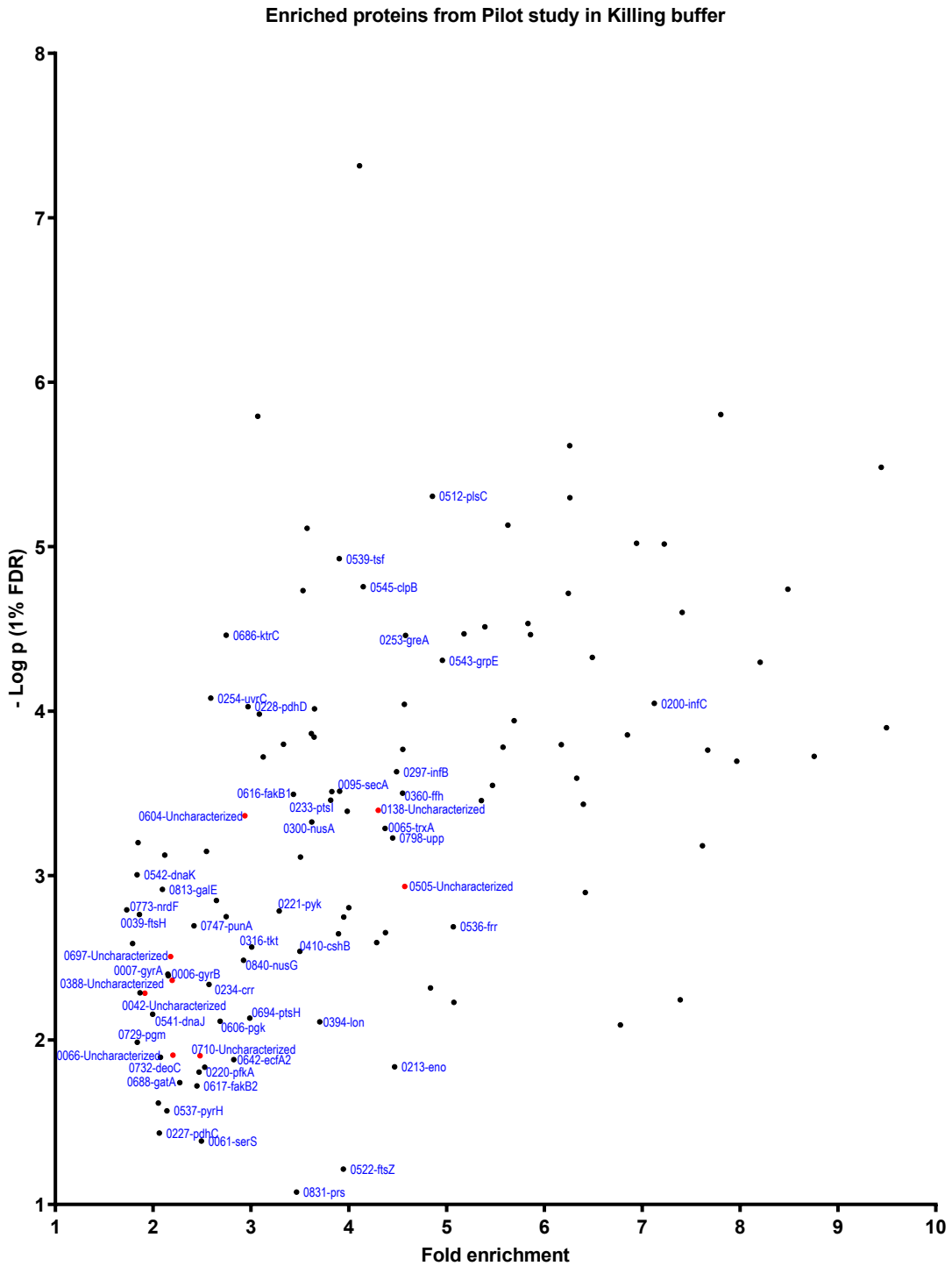
buffer, had also been tested which lent another dimension to which proteomic data could be accessed. The results for these pilot studies are also reported below.

Plot 3: Enriched proteins from pilot study in PBS buffer, MS mode: IonTrap



In Plot 3, 132 proteins were classified as enriched. This included 12 uncharacterized proteins namely, 0030, 0066, 0138, 0353, 0388, 0439, 0451, 0493, 0505, 0697, and 0710. Also, five previously annotated proteins were also found to be enriched after crosslinking. They are 0300-NusA, 0305-papA, 0420-fakA, 0512- plsC, 0617-fakB2. In Plot 4, a different cutoff for the enrichment was selected based on the total RNA yield from the samples submitted for processing.

Plot 4: Enriched proteins from pilot study in Killing buffer, MS mode: IonTrap



In Plot 4, all data points except ribosomal proteins have been labelled. A total of 118 proteins were found to be enriched. Of those 11 were uncharacterized, namely 0030, 0042, 0066, 0138, 0388, 0439, 0451, 0505, 0604, 0697, and 0710.

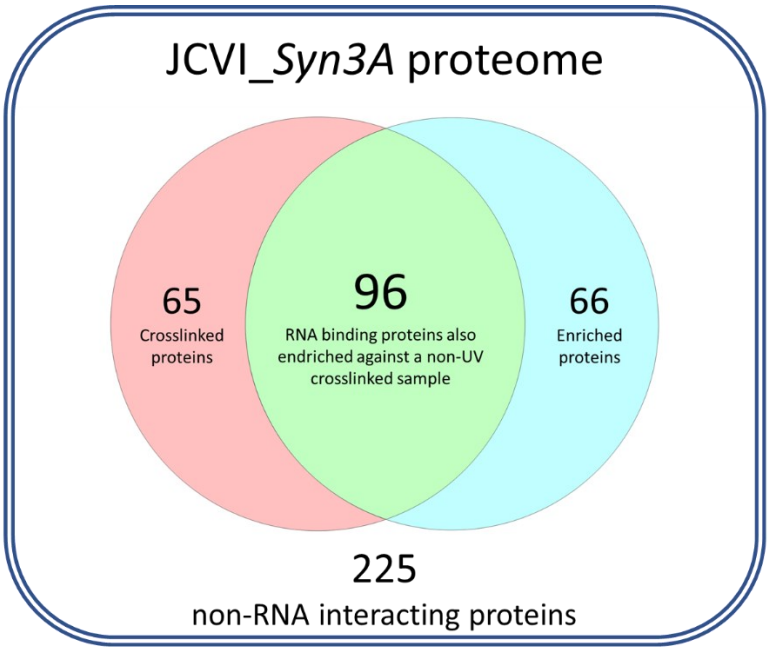


Figure 23: RNA-binding proteins in Syn3A. 161 proteins were found to contain a localized RNA-binding site and 96 were confirmed for RNA-binding ability. A further 66 proteins were only enriched in the UV-crosslinked sample versus the non-crosslinked sample.

Cumulatively, as depicted in Figure 23, a total of 161 proteins were crosslinked which is the definitive measure for whether a protein interacts with RNA or not. A further 162 proteins were enriched in a UV-crosslinked sample as compared to a non-crosslinked control. 227 proteins were found to interact with RNA in this study. 225 proteins out of the complete proteome of 452 proteins in Syn3A were found to not interact with RNA.

The scientific value of this study has thus been to provide an often elusive, RNA-interaction dimension to the characterization of the unknown proteins while confirming the annotations of known RNA-binding proteins.

4 DISCUSSION

The cellular activity of proteins, the complexes they arrange themselves into to perform specific functions and the role of yet uncharacterized proteins is an extremely fascinating layer of knowledge. In the pursuit of designing synthetic genomes and artificial cells (Chi et al., 2019) the complete genome must be annotated. As demonstrated by recent efforts to create a minimal cell from *M. mycoides* (Hutchison et al., 2016), almost one third of the genome remained poorly or completely uncharacterized despite being essential in nature. Our experiments through cutting edge CLMS protein interactome maps developed in *M. pneumoniae* and already in their next phase of evolution with Syn3A are a collaborative effort towards defining the essential functions and genes required for a minimal natural and synthetic cell. In this thesis, results from *M. pneumoniae* were confirmed, strengthening the evidence for coupled transcription-translation in Mycoplasma. The first interactome map was created from CLMS experiments in Syn3A, and several protein complexes of unknown function are under study. A whole cell RNA-interacting protein study was also conducted revealing unexplored RNA-binding abilities for several known proteins and providing useful context for some uncharacterized proteins.

4.1 NusA

The experiments to determine the essentiality of the disordered C-terminal region showed that this region is essential. While NusA can tolerate an excessive amount of transposon insertions in the first passage (Lluch-Senar et al., 2015), as indicated initially by the presence of only one transposon insertion event in this region by passage ten- this region does appear to be essential for cell viability.

In-cell structural studies of *M. pneumoniae* showed via whole cell cross-linking that the RNA polymerase core consisting of the conserved subunits α, β and β' was interacting with the proteins GreA, NusG, NusA, SpxA, SigA, and RpoE (O'Reilly et al., 2020). Of these, NusA, appeared to link RNAP via its N-terminal domain and the mRNA entry site of the ribosome via its C-terminal region.

Confirmation of this linkage was sought first via cryo-Electron tomography data of *M. pneumoniae* cells. 53.3% of the sub-tomograms were sorted as 70S ribosomes and resulted in a 5.6 Å ribosome density. This data enabled the fitting of a ribosome homology model on PDB 3J9W to which most *M. pneumoniae* ribosomal proteins could be mapped. Focusing on the mRNA entry site allowed the identification of a subset of ribosomes in complex with RNAP. This complex was refined into a 9.2 Å map that showed additional density at the interface between the RNAP and the ribosome (O'Reilly et al., 2020). This additional density could not

be explained by NusG or any of the other proteins that had been found to interact with the complex in the CLMS PPI map except for NusA without its disordered C-terminal region. However, this was the exact region of NusA that was found to have multiple links to the 30S ribosome.

The NusA N-terminal region was shown to bind RNAP similarly to a complex reported in *E.coli* (Guo et al., 2018) near the RNAP mRNA exit tunnel. The model predicted that the two KH domains in the middle of NusA were positioned near the ribosomal mRNA entry site. Thus, the only missing information that had not been explained either by electron density maps or the integrative model was the role of NusA C-Terminal domain (CTD). The experiments that failed to generate any NusA truncation mutants despite targeting two different truncation points and multiple mutation approaches (Section 3.1) provided further confirmation that this disordered region is essential for viable *M. pneumoniae* cells. This lent support to the hypothesis that NusA CTD does indeed bridge the RNAP and ribosome in the expressome.

Transcription and translation was already reported to be coupled in bacteria (Miller et al., 1970). In experiments on the leader sequence of tryptophan operon in *E. coli*, paused transcription could be resumed by the activity of its coupled ribosome (Landick et al., 1985). NusG had also been shown via NMR to form a complex with the 30S ribosomal protein S10 in *E. coli* (Burmann et al., 2010). At the time, NusG was proposed to bridge the elongating RNAP to the ribosome via its bond to S10, thereby controlling the rate of transcription based on the capacity of the cell to translate new proteins (Proshkin et al., 2010). Furthermore, this coupling was found to be essential to prevent RNAP back-tracking, and in addition to the previously found NusG-S10 bridge, a NusG homolog RfaH could also interact with S10 (McGary & Nudler, 2013).

The cryo-EM structure of the entire expressome complex consisting of the necessary elongating RNAP and the ribosome, was however found to not include NusG or RfaH as the linker arm of either protein is too short (Kohler et al., 2017). The postulated NusG binding sites were also found to be on opposite ends of a cryo-EM *E. coli* RNAP core bound to the small ribosomal 30S subunit (Demo et al., 2017). Our reported crosslinking data coupled with the cryo-ET structures all pointed to NusA being this link between transcribing and translating complexes. That the deletion of the disordered NusA CTD yielded no viable mutants serves to confirm that this disordered C-terminal region is essential for maintaining integrity of the expressome.

However, this field of work is still expanding. Recently, in *E. coli* it was shown that two types of expressomes could exist, one where the transcription-translation complex (TTC) has a short mRNA spacer between the RNAP and the ribosomal active-centre P site and does not seem to use NusA or NusG as a bridge. When this mRNA spacer is longer however, then NusG appears to bridge the TTC while NusA stabilizes it (C. Wang et al., 2020). This finding

was partially confirmed by the study of a series of cryo-EM acquired coupled, uncoupled, and collided expressomes. It was found that when a long mRNA sequence is present NusG bridges the transcription and translation complex. However, upon shortening of the mRNA spacer, significant re-arrangement takes place which orients the ribosome entrance channel to the RNAP exit channel, removing the NusG bridge (Webster et al., 2020). Perhaps as this picture develops further and one keeps in mind that while most of this work has been done in *E. coli* and also shown in Archaea (French et al., 2007), variety is often a feature of life. Indeed, it has also been shown that initial contact between the RNAP and ribosome is actually direct and via *in vivo* reporter assays that the recruitment of NusG occurs late in transcription and is translation dependant (Washburn et al., 2020). Lastly, in terms of bacteria other than *E. coli*, uncoupled RNAPs in *B. subtilis* have been shown to outpace pioneering ribosomes and genomic signatures of uncoupled transcription-translation have been reported to exist in several bacterial phyla (Johnson et al., 2020).

In order to add further proof to NusA linking the RNAP to the ribosome, perhaps, a CRISPR/Cas interference based method could be applied to confirm that the disordered C-terminal region is essential (Zhang et al., 2021).

Emboldened by our work with *M. pneumoniae* but also aware of the limitations imposed by a slow growing, S2 organism, we also began exploring similar work with a new minimal cell that was gaining popularity and had potential to be a more accessible minimal model.

4.2 A Protein-Protein Interaction map of Syn3A

Our previous work involving the crosslinking of *M. pneumoniae* (O'Reilly et al., 2020) had demonstrated already the ability of whole cell CLMS to shed light on the working of proteomic machines like the transcription coupled to translation assemblies. While *M. pneumoniae* had an expressed proteome of 689 genes, Syn3A has only 472 protein coding genes. This lower value is somewhat closer to the upper limit of total proteins that can be handled by mass spectrometric analysis with good coverage of the proteome depth (Chavez & Bruce, 2019).

The conditions for crosslinking were set up with a new experimental affinity tagged crosslinker that was to be used with this new minimal cell. Both parameters had therefore been optimized from the previous study with *M. pneumoniae*, where we used CLMS on a larger genome (therefore more complexity in starting sample) and two different crosslinkers, DSS and DSSO (O'Reilly et al., 2020). The PPI maps soon revealed that several reported protein complexes like DNA gyrase, and RNA polymerase could be visualized (Figure 15). Furthermore, associations between complexes of proteins could be established as depicted in Figure 16 showing the 50S and 30S ribosomes and their subunits with interconnecting links. As extremely few structures are recorded for *M. mycoides* and none for Syn3A (as these

proteins are deleted), structural homologs showing the most similarity could be fitted with the crosslinking data as depicted in Figure 17. Using the Alphafold2 algorithm for predicting structures, a similar effort was made for Syn3A_0439 (Figure 20), one of the uncharacterized proteins of a larger uncharacterized membrane protein complex (Figure 18).

A long sought-after goal of the structural biology-crosslinking research community is the crosslinking of whole, intact cells in their native growth environments. This would mean the analysis of a highly complex, intricate network of proteins spanning the entire proteome and physiological environment of the cell. While this has been partially successfully applied to soluble cell fractions in cell lysates (B. Yang et al., 2012), the data generated are sparse due to most crosslinks occurring between abundant, well-characterized protein complexes such as the proteasome and ribosome. In the *M. pneumoniae* study 10,552 self-links and 1579 heteromeric crosslinks representing 577 distinct protein-protein interactions were recovered from close to 700 MS runs, half with DSS and half from DSSO, an MS-cleavable crosslinker. Comparatively, the still in-progress, Syn3A PPI project has 15,316 self-links and 6,001 heteromeric crosslinks representing 643 distinct protein-protein interactions from around 50 MS runs. The Syn3A study also uses a lower FDR of 1% (for PPI) compared to the 5% used for *M. pneumoniae*. These numbers demonstrate the advantage of the experimental affinity tag on the enrichable crosslinker, StageCL. With 311 of the identified crosslinks involving at least one of the 28 uncharacterized proteins found so far, it is also delivering on the less abundant, unknown proteins. Four unidentified complexes comprising uncharacterized proteins have also been indicated (Figures 18 and 22). Another advantage of this study is that even though more cellular material was required; the cells lysates have not been fractionated prior to trypsin digestion into peptides. This means the entire fractionation into membrane and soluble proteins is skipped and true whole cell crosslinking data can be acquired in one step.

The importance of the crosslinking data generated so far cannot be understated. The uncharacterized proteins now under study along with their interaction partners have been identified as functioning together for the very first time. As even one crosslink at an FDR lower than 5% is enough to indicate interaction (O'Reilly & Rappsilber, 2018) it is possible to imagine all the 79 uncharacterized proteins appearing at least once with more MS runs and as more fractionation of the samples takes place. Complete fractionation using anionic, cationic affinity-based chromatography, followed by size exclusion chromatography and then C18 phase separation prior to MS analysis is planned. This will also help towards the as-yet unrealized secondary aim of this project which is to generate sufficient self-link data to enable structural prediction of all complexes when coupled to computational model prediction.

When completed this resource should enable visualization of the complete proteomic interactome of the cell, sufficient crosslinking to identify topology of protein complexes and a self-link density that will enable tertiary structure predictions of uncharacterized proteins. The

first iteration of this structural prediction was already demonstrated via using an SDA crosslinker on human serum albumin and broadly agreed with the known crystal structure (Belsom et al., 2016). Such 'high-density' information from CLMS had already been deputed in CASP11 (Schneider et al., 2016) and while still at a modest stage, already has potential for flexible structures whose multiple conformations have eluded X-ray crystallography and EM. More importantly, such a CLMS pipeline with an enrichable-crosslinker that outperforms older crosslinkers, provides a quickly implementable process for the characterization of unknown, new organisms and delivers whole cell data and specific protein structures from the same starting material. Especially in minimal cells where only the core genome comprising of essential genes remains, this data is invaluable as traditional methods of deletion mutants are impossible.

In the complex containing Syn3A_0439, Syn3A_0440 and Syn3A_0505 all bound together by multiple linkages, based on cell-free protein expression data that was presented at the Synthetic Minimal Cell workshop, 2021, where RNase activity of a membrane bound nature was found in cell extracts, it is proposed that these 3 proteins form a membrane bound extra cellular nuclease-nucleotide uptake system (Sharma et al., 2015). Syn3A_0439 additionally has RNA binding ability with one peptide matched, and Syn3A_0505 was enriched in the second RNA-binding protein-UV crosslinking enrichment experiment. Cloning of these heavily trans-membraned proteins has been partially completed and over-expression experiments are planned. Perhaps in the future, research on this complex can build on earlier work done to characterize ribonucleotide ABC transporters (Webb & Hosie, 2006), explore old reports of deoxy mononucleotide uptake (Neale et al., 1984; Youil & Finch, 1988) and confirm nucleotide uptake as this activity is reported for *M. mycoides* but no genes have been assigned (Wise et al., 2019). Neighbouring interaction partners Syn3A_0412 and 0601, and ThiC cannot be commented upon at the current time. Finally, it should be noted that while Alphafold2 database currently does not contain any structures of the unknown proteins, Colabfold can be applied to generate structural predictions which can be in turn validated via the self-link data, when available. This pipeline has been successfully applied for Syn3A_0439 (Result section 3.2.5).

From the protein complexes where at least one partner is characterized, perhaps their characterization might be easier. In complex A, (Figure 22) NusB, a transcription anti-termination factor is linked to Syn3A_0451 which is hypothesized to be a member of the Aldehdye dehydrogenase family. Furthermore, Syn3A_0451 was also found to be an RNA-binding protein with one peptide matched. Of the known 19 known Aldehyde dehydrogenase enzymes, 10 have structures, though not all are of bacterial origin. The structure does include a Rossman fold and NADH binding cleft (Shortall et al., 2021), both known to also bind RNA. NusB binding to an aldehyde dehydrogenase enzyme has not been reported as per STRING. Perhaps given the moonlighting functions of glycolytic enzymes (Curtis & Jeffery, 2021), one

can hypothesize a similar role for 0451, especially since there are reports of ALDH enzymes involved in several cellular processes like detoxification, antioxidation and regulation (Chang et al., 2013; Shortall et al., 2021).

The complex containing DNA gyrase subunit A and 0388 (Figure 22) is also fascinating because Syn3A_0388 was identified as an RNA-binding protein with 6 different peptides matching to this small 24 kDa protein. A Pfam search confirms it lacks any recognizable domains. Given the extremely important role GyrA plays in multiple cellular processes such as replication, transcription and cell division (Vanden Broeck et al., 2019), perhaps one can postulate that 0388 forms part of a regulatory complex towards its functioning in these various pathways, one that also incorporates elements of RNA sensing. Interactions between and possibly also containing DNA gyrase subunits have been shown to be heterologous and optimized in a species specific manner (Weidlich & Klostermeier, 2020).

In one of the complexes that in-vitro experiments have already begun on, Syn3A_0317 shows crosslinks to both Syn3A_0604 and ECF A1. Subsequently we learnt that four peptides were matched to ECF A1, marking it as an RNA-binding protein, along with Syn3A_0317 which mapped 3 peptides. Thus, we have a complex where a small 8 kDa, uncharacterized protein (Syn3A_0317) of relatively low copy number (20), binds RNA, and links Syn3A_0604, a 25 kDa protein of 346 copies and the ATP binding A1 component of ECF transporter, gene Syn3A_0643, which also binds RNA, with all the links localizing around the C-terminal region of 0317. That the links come from the same lysine residues could imply a binding site near that region. Syn3A_0317 contains a predicted transmembrane (residues: 12-38 of 81aas), and from residues 19-79, a domain belonging to uncharacterized protein family 0154, a family of short bacterial proteins of unknown function also called DUF1043. Like its homolog in *B. subtilis*, YneF, 0317 is essential. Syn3A_0604 contains a single LemA family domain (residues 73-220, total length: 222 aa). This LemA domain was described structurally from *L. monocytogenes* (Lenz et al., 1996) and consists of a four-helical forming transmembrane structure where its N-terminus is extracellular. This structure would fit with 0604, leaving a predicted lengthier disordered N-terminus region. The crosslinks likely happen within the trans-membrane region. While Ecf A1 has never been reported to bind RNA, Syn3A does contain a Group II ECF transport system (Slotboom, 2014) whose architecture has been accounted for (Wise et al., 2019).

Further afield, the similarities in predicted domains for 0639 and 0399 (Figure 14) indicate that they may both be membrane pores or transport permeases especially as they contain a Pfam domain P57382 linking them to FtsX-like permease family, predicted to transport lipids. Given that Syn3A has no biosynthetic potential and hence, all biobricks must be scavenged-the question is which substrates to these gated permeases take up? With 0636 and 0398 both having lipoprotein architecture which is not widely conserved outside of Mycoplasma, parallels

may be drawn to the variety of *Opp* solute-transport systems (Levdikov et al., 2005) found in pathogenic bacteria like *Borrelia burgdoferi* (X. G. Wang et al., 2002, 2004) which has 13 such solute-binding protein genes and *Streptococcus pneumoniae* (Kerr et al., 2004) which has 3 *Opp* paralogues required for colonization of the nasopharynx. Lastly, Syn3A_0636 was also found to be an RNA-binding protein with one matched peptide. Given the Pfam evidence, and domain structure, perhaps a model wherein RNA-binding protein 0636 is used to control amino acid levels may be proposed, similar to that of RNA-binding protein YBX3 which regulates solute carrier amino acid transporter mRNA abundance (Cooke et al., 2019).

The crosslinking of cells was performed on a large volume of log phase culture. Thus, protein interactions that only occur during stationary phase may have been missed. Further, it is possible, that as the cells were washed twice before crosslinking, that they had begun to express some sort of stress response. A comparative study to crosslink in medium could perhaps address this, however whether this is feasible in SP4 which is undefined and protein rich itself must be considered first. The use of a crosslinker with a smaller spacer arm than DSS and DSSO allows for a map of higher resolution. This coupled to its affinity enrichment tag has already increased the data yield with fewer MS runs. However, using a new crosslinker comes with unknown risks as well. There is significantly less experimental and established data on this crosslinker compared to those that have been in use for thirty years. Therefore, user knowledge is still being created, and questions of stability during storage and transport of the crosslinker do exist. Utmost care was taken to use the crosslinker as prescribed by our collaborators.

In addition, using an experimental crosslinker on a synthetic-somewhat-experimental minimal cell represents an additional level of consideration. It is hypothesized that the genes deemed unessential and hence removed, often functioned towards long term cell homeostasis, or were involved in protein stability and degradation of misfolded proteins as well as cell signalling of a non-essential but higher-secondary function. While the genome of *M. mycoides* was thought to already be highly adapted to its ecological niche in the host, and therefore minimized to a sustainable equilibrium and still be pathogenic, Syn3A lacking all pathogenic ability may have several loose ends (Danchin & Fang, 2016; Wise et al., 2019). One issue that the crosslinking experiments have already revealed is that when focusing only on self-links, internal ribosomal architecture does not agree with that reported for crystal structures. While this does not seem to be an issue for PPI maps, that crosslinking data matches well with established crystal structures is used as a means of quality checking in the CLMS community. This is not an exact quality benchmark though, as while CLMS data is assumed to be static, it may capture protein machinery in a variety of confirmations while *at work inside the cell*. Thus, CLMS data while sometimes not fitting established structures due to over-length crosslinks

may represent structures present *in situ* but uncapturable in crystal structures (Ding et al., 2017).

As this is a secondary aim of our crosslinking endeavour in Syn3A with StageCl, several steps are underway to shed more light on this situation. As we have excellent crosslinking data for the ribosome from *M. pneumoniae* (O'Reilly et al., 2020), we have crosslinked *M. pneumoniae* with StageCL, with the aim of confirming the results we obtained using DSS and DSSO crosslinkers. This would validate the functioning of the crosslinker in a known cell, with native ribosomal architecture. Further, the possibility exists that Syn3A ribosomes are incompletely folded or unstable due to the lack of chaperones or ribosomal assembly proteins. Hence, we have crosslinked Syn1.0 cells which while containing a synthesized genome are essentially wild type *M. mycoides*. The addition of these control studies will enable us to better understand the ribosomal crosslinking results obtained, thus far. Lastly, since the *M. pneumoniae* samples were separated into membrane and soluble fractions, it could be possible that similar variable ribosomal architectures were present there as well but weren't caught as only the soluble fraction was processed. This once again, highlights the advantages of this crosslinking study namely SPEED processing of the sample and using an enrichable crosslinker.

In the future, the remaining fractions (SCX, SEX, HCX) need to be processed and the sample penetration enhanced, thereby covering the entire proteome. Currently some 5-10 percent of the proteins have not yet appeared in the maps. Protein pull downs, can be used to verify each of the reported interactions while also generating recombinant tagged versions of the uncharacterized proteins and their partners. These would enable ITC studies, down the line, or protein characterization assays, once more information is available. Of considerable importance is the creation of deletion mutants, which is a challenge when so many of these proteins are essential. Hence, gene silencing strategies may be employed perhaps based on CRISPRi (Zhang et al., 2021). Especially in the case of proteins that may interact with lipids or metabolites, MS analysis of cell lysate after a fishing experiment using the uncharacterized protein as 'bait' should be undertaken. Overall, the prospects of Syn3A interactome studies are linked to overlaying several -omics datasets on this interactome map, in an approach originally highlighted for *M. pneumoniae* (Catrein & Herrmann, 2011), but which should be considerably easier in Syn3A, due to the lack of complexity derived from considering host-pathogen interacting proteins and influences- Thus, Syn3A, a truly minimal cell has tremendous uncovered potential in the cell modelling arena.

4.3 RNA-interacting proteins, a whole cell approach

The creation of Syn3A involved the deletion of a total of 432 genes from Syn 1.0 which is almost wild type *M. mycoides* capri LC GM12. Of those deleted, only one gene was involved in ribosomal biogenesis- YqeH or MMSYN1_0488. Four redundant methyltransferases, and one tRNA-lys was also deleted. This indicates that Syn3A has only six genes short of the complete complement of ribosomal/RNA-interacting genes found in the wild-type bacterium (Hutchison et al., 2016). However, the removal of 426 other protein coding genes implies that several proteins involved in regulation of various cellular processes, genetic events and homeostasis maintenance are likely to be missing, if deemed non-essential. Given that essentiality was determined using transposon insertion-based survivability rates, those genes that could tolerate insertions over four passages were deemed non-essential (Hutchison et al., 2016). While this approach was elucidative, it does not comment on the quality or robustness of the cell in terms of 'ability to adapt to its environment'. Within the established framework of control and regulation of cellular processes, with the number of proteins severely reduced, it seems plausible to hypothesize that several other proteins may have taken on additional secondary roles, and that different RNA transcripts or DNA topologies may be the sole remaining regulatory elements in the cell. In the absence of non-essential proteins, this cell provides the possibility to characterize new RNA-interacting proteins in the core proteome.

Our two-stage study was an attempt to do so.

In the first stage, UV crosslinking of cells from a log phase culture was performed in a pilot study. For this pilot study, cells were crosslinked in medium which yielded much less insightful data as SP4 medium is a very rich, undefined medium. However, as that was one of the established protocols for *E. coli*, it was performed. Cells were also washed in PBS twice and then crosslinked, yielding far better results. Following on the success of the pilot experiment, a larger scale experiment in quadruplicate was performed, and various parameters such as Silica or TiO_2 based enrichment, use of a standard or digestion mix, and search and analysis via NuXL or RNPxl (different search engines) was experimented with to obtain the optimum results.

The second stage of this approach was to investigate whether these same proteins would also be enriched after crosslinking cells against a non-crosslinked control. As this approach would select for peptide species containing an RNA nucleotide (A, G, U, C), only potential RNA-binding proteins should be enriched. Such a setup allows a side-by-side comparison of the expected vs. observed MS/MS spectrums for the intensities of control vs UV-irradiated sample. Again, first a pilot stage with only technical triplicates was performed to determine which buffer was most suitable for RNA stability, and then a much larger scale biological and technological study was carried out in quadruplicate. Two different MS acquisition modes were also experimented with, OrbiTrap and IonTrap, which differ in their handling of the ions. OrbiTrap was found to provide greater depth or gleam more information from the samples, but

IonTrap did provide some hits that were different. As shown in Plots 1- 4, mostly ribosomal proteins were captured in each of these approaches, but also a significant number of annotated proteins. The overall discovery of uncharacterized proteins as RNA-binding proteins was much lower using this method.

Taken together, the first approach provides data on whether a protein interacts with RNA, and if possible, localizes the residue of interaction in the amino acid sequence. The probability of this being a random event due to a RNA species or protein occupying the same cellular environment, for example, is miniscule. Further, the number of peptides capturing this interaction can also be used to ascertain how many times the MS measured it. Hence, the data from the first UV-crosslinking project alone is strongly indicative of an RNA-binding protein. The second project then provides confirming data as to whether, in a large enough sample of cells, these RNA-binding proteins were also enriched when UV-crosslinked vs a non-UV-crosslinked control.

While there is no definitive list on the complete number of RNA-binding proteins in Syn3A, those processes most likely to involve RNA-protein interactions can be considered to estimate effectiveness of our experimental setup.

Of the known RNA-binding proteins, naturally most come from the ribosomes and transcription. All 20 of the 30S ribosomal proteins have their RNA-protein binding region localized and were enriched, thus confirming their RNA-protein interactions. Similarly, 29 of the 30, 50S ribosomal proteins were localized and of these 27 were enriched. Nine proteins are known to play a role in transcription in Syn3A. Four of these, three subunits of the RNA polymerase (α, β, δ) and NusA were listed and confirmed to be RNA-binding proteins. RNA-binding proteins are also widespread in translation and related processes like tRNA synthases, translation factors, ribosome assembly, and rRNA modification and maturation. Compared to the number of proteins reported to take part in these processes as per Synwiki datasets (Pedreira et al., 2022), a smaller subset was found and confirmed (Table 11, translation and unassigned group).

Consulting Synwiki (Pedreira et al., 2022), of the known endoribonucleases, RnmV was listed and Rny were confirmed. RnjA and RnjB, exoribonucleases for RNA metabolism and turnover were confirmed as well. Of the proteins involved in nucleotide metabolism, Upp, Prs, and RgdB were confirmed while three more were listed. That these expected proteins were caught in the RNA-binding protein screens lends confidence to the results but also implies that there is room for improvement as not all expected RNA-binding proteins are localized and confirmed.

Considering less logical RNA-binding proteins, that have been previously reported in the literature, the number of metabolic enzymes confirmed to bind RNA was also quite significant. Specifically, enzymes of core carbon metabolism: Crr, Eno, FbaA, GapA, Ldh, PdhC, Pgk,

Pgm, PtsG, and Rpe were reported and six of these were confirmed. FtsA and FtsZ of cell envelope and division machinery were also noted. While the multiple roles of metabolic enzymes have been recently reported (Curtis & Jeffery, 2021), this is confirmation that Syn3A also uses some form of complex RNA-moonlighting enzyme interactions. These interactions are likely to be wider than that of enzyme control, and possibly are conditional to the different intracellular and environmental conditions (Buccitelli & Selbach, 2020; Cieśla, 2006).

Surprisingly while GalU was the only known RNA-binding protein involved in lipid metabolism, several other members of lipid metabolism namely FakA, FakB2 and PlsC were also noted and confirmed to be new RNA-binding proteins (Table 14). Recently, an mRNA dependant motility control mechanism involving GalU was also described (Zhao et al., 2021). In a cell where lipid acquisition is the only means of cell membrane synthesis it is of significant interest that these proteins also interact with RNA.

Ten DNA related proteins with known RNA interacting functions were also detected and four of these were confirmed (GyrA, GyrB, ExoR and UvrC). GyrA and GyrB are helicases whose ability to bind RNA has been reported (Corley et al., 2020) while the other two are nucleases. In addition, one of the new RNA-binding proteins, WhiA, also has an uncharacterized DNA-binding function. While DNase treatment was performed on the samples during processing, there is a slight chance that this protein is a false positive. In order to resolve this, the spectra would have to be manually analysed.

Nineteen previously annotated proteins were identified as RNA-interacting proteins. None of these proteins were known to have any record of RNA interaction based on searching Synwiki (Pedreira et al., 2022), RBP2GO (Caudron-Herger et al., 2021) (selection: *E. coli*), and STRING databases. The only proteins that were confirmed from these 19 were PapA, a peptidase for proline containing peptides and the three lipid proteins mentioned above. The remaining proteins are all annotated with a transport related function (Table 14). Often, they are ATP-binding components of their transporter complexes. The ability of ATP-binding domains to bind RNA with varying specificities has been reported (Sassanfar & Szostak, 1993; Wei & Richardson, 2001). This could be one mechanism of action for this subset of new RNA-binding proteins.

Another twenty proteins, all uncharacterized, were found to be RNA-binding proteins, 7 are essential, 1 possibly essential, 8 quasi-essential and 4 are non-essential which were included for fitness purposes (Table 14). Of the uncharacterized proteins that were caught in this RNA-binding protein study, several have homologs and preliminary annotations. Syn3A_0030 is essential, and a likely ABC transporter, ATP-binding protein. It shows some similarity to MacB from *E. coli* and BceA from *B. subtilis* and could probably be annotated as so. Similarly, Syn3A_0451 resembles aldehyde dehydrogenase- AldY (*B. subtilis*), GabD (E. coli). Syn3A_0493 is similar to YtjP and DapE in *B. subtilis* and *E. coli* respectively and is

annotated as a putative N-acetylglucosamine-6-phosphate deacetylase. Lastly, Syn3A_0878 is annotated as an amino acid permease. While it does not appear to be conserved across several species, its homologs in *B. subtilis*, SteT and *E. coli*, FrlA along with the knowledge that *M. mycoides* capri LC Y can take up all amino acids in their free form (Rodwell, 1969) suggests that this protein too can be annotated.

The second RNA-binding protein enrichment project aimed to confirm results of the first. UV-crosslinking of the test sample creates peptide-nucleic acid conjugates which should be absent from the non-crosslinked control. An algorithm calculates the modified masses of all possible peptides, when crosslinked to a nucleic acid residue. Then the proteomes of the sample and control are processed. RNA-binding proteins from the test sample should thus be 'enriched' in the proteome, their level of enrichment reflecting that proteins abundance as an RNA-binding protein (K. Sharma et al., 2015). Twenty uncharacterized RNA-binding proteins were found in the first project. Of these, only Syn3A_0030, 0138, 0388, 0439, 0451, and 0493 were confirmed in the second enrichment project. In addition, Syn3A_0505, 0042, 0066, 0080, 0353, 0697 and 0710 were also found to be enriched in the crosslinked sample. Thus, while direct binding evidence does not exist for this latter group of proteins, they may also be considered as tentative RNA-binding protein.

The first localization project allowed us to determine the RNA binding peptide as well as the exact 1 or 2 amino acid residues where the interaction takes place. It is yet to be determined how many of these interactions take place within known RNA binding sites and how many are in disordered regions or yet to be characterized RNA-binding domains. A sub-project to determine how many RNA-binding proteins could be predicted from known domains and short amino acid sequences using the TriPepSVM pipeline (Bressin et al., 2019) was undertaken in the still underway-masters thesis of Camilo Torres. This study will provide an insight onto hitherto unknown RNA-binding motifs. Meanwhile, several moonlighting metabolic proteins have been found to contain only Rossman folds which were characterized to bind redox cofactors: NAD+, NADPH, and FAD (Hentze, 1994). Disordered protein regions are not commonly seen in glycolytic enzymes but often in transcription factors like NusA and uncharacterized proteins. Metabolic enzymes do however often contain loops or flexible regions that could play the same role (Castello et al., 2012; Fuller et al., 2020).

While the best efforts have been made to make this a robust study of RNA-binding proteins in Syn3A, the RNA yields of individual samples still varied in the different replicates. Especially for the pilot enrichment studies, this could be a source of error as some proteins may appear to be enriched when not. We have tried to compensate for this by using a higher enrichment fold cut-off when this happened (Plot 3). Besides this, some proteins may exhibit an RNA-binding function only under certain cellular or environmental conditions. Lastly, while data has

been generated from a protein centric perspective, it remains to be determined which RNA species bind to which proteins.

Thus, in addition to re-assessing the samples from the point of view of growth phases, other protein-centric studies should also be conducted in the future. As the most abundant proteins that were picked up in this screen were ribosomal or other annotated metabolic enzymes, perhaps specific uncharacterized proteins could now be targeted using a crosslinking immunoprecipitation (CLIP) or CLIP-seq approach to sequence the bound RNA (Licatalosi et al., 2008). Alternatively, RNA-centric methods that again utilize crosslinking via UV (PAIR, (Zeng et al., 2006)) or formaldehyde (CHART, (Simon et al., 2011)) and then streptavidin capture of the RNA oligomer-protein complex may be applied. Alternatively, their overexpression could be undertaken from a plasmid and cell lysate used to pull down RNAs followed by RNA sequencing. Electrophoretic mobility shift assays (EMSA) could also be used to initially confirm these hits. Lastly, a whole cell transcriptomic experiment is in the publishing step. The coupling of this data to transcriptome abundancies would yield insightful data on when these proteins and their RNA interactions most likely come into play.

4.4 Outlook

NusA was found to be an integral part of the expressome and was proposed to bridge the RNA polymerase and the ribosome (O'Reilly et al., 2020). It has been shown here that the disordered NusA C-terminal domain region is essential for cell growth. Whether this disordered region can bind 30S ribosomal proteins S3 and S2 individually, as predicted by the crosslinking data, could be also tested. NusA may act as a transcription factor that senses the change from an inactive to an active expressome and strengthens the coupling of this dynamic super-complex. If stalled ribosomes lead to dissolution of the expressome (O'Reilly et al., 2020), then further research into what triggers NusA recruitment is required. Further, as transcription and translation seem to be uncoupled in *B. subtilis* (Johnson et al., 2020), and the linker-protein is still unconfirmed in *E. coli*, (Demo et al., 2017; C. Wang et al., 2020), both the prevalence of this linking mechanism and the coupling of these two processes throughout the wider bacterial domain needs to be explored.

The protein-protein interaction maps have confirmed the existence of several conserved complexes and protein interactions in Syn3A. While it has captured hetero crosslinks for 28 uncharacterized proteins, the entire proteome of Syn3A has not yet been visualized in the interactome map. To capture the missing annotated and uncharacterized proteins, experiments using higher crosslinker concentration need to be undertaken. A pilot experiment towards this endeavour has already been performed to test whether the current digestion protocol and sample processing is suitable for higher crosslinker usage. Once deeper

coverage of the proteome is achieved, further investigation on the issue of abnormal ribosomal architecture needs to be addressed. The control experiments involving crosslinking both in *Syn1.0* and *M. pneumoniae* will prove whether the genome rearrangement process was responsible for the over-length crosslinks seen in the ribosome or whether ribosomal architecture is different in the *Mycoplasmas* itself.

Prediction of protein structures should be undertaken using Colabfold followed by the self-links being used to validate predicted models. Structure-based functional characterization can then be used to characterize those proteins without interacting partners. Regarding those proteins that are predicted to have a disordered structure but a high copy number (Wright & Dyson, 1999) crosslinks to other proteins would be invaluable starting points for further investigation. Finally, drawing on results from the RNA-binding protein study, the complexes containing Syn3A_0317, 0388, 0439 and 0451 should be studied further as they all have protein interacting partners, self-links, and are confirmed RNA-binding proteins. The RNA-binding study localized RNA-interacting sites and provided confirmation for many of these proteins. These results should be validated again by *in vitro* RNA-pulldowns using purified target RNA-binding proteins and cell lysate. RNA-sequencing should then be used to determine the RNA transcript. The study of each RNA-localized peptide in its respective protein should also be undertaken to investigate the possibility of new RNA-binding motifs. Aside from the three proteins already mentioned, the components of lipid metabolism FakA, FakB2 and PlsC have never been connected with RNA-binding functions (Frank et al., 2020). Hence, they should be studied immediately, from the point of view of RNA-based regulation. Altogether, these two whole cell protein-interaction studies have provided several potential projects which will ultimately help yield an annotation for every gene in the minimal cell Syn3A.

5 REFERENCES

Antczak, M., Michaelis, M., Wass, M. N. 2019. Environmental conditions shape the nature of a minimal bacterial genome. *Nat. Commun.* **10**: 1–13.

Belsom, A., Rappsilber, J. 2021. Anatomy of a crosslinker. *Curr. Opin. Chem. Biol.* **60**: 39–46.

Belsom, A., Schneider, M., Fischer, L., Brock, O., Rappsilber, J. 2016. Serum albumin domain structures in human blood serum by mass spectrometry and computational biology. *Mol. Cell. Proteomics* **15**: 1105–1116.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248–254.

Bressin, A., Schulte-Sasse, R., Figini, D., Urdaneta, E. C., Beckmann, B. M., Marsico, A. 2019. TriPepSVM: *de novo* prediction of RNA-binding proteins based on short amino acid motifs. *Nucleic Acids Res.* **47**: 4406.

Brimacombe, R., Stiege, W., Kyriatsoulis, A., Maly, P. 1988. Intra-RNA and RNA-protein cross-linking techniques in *Escherichia coli* ribosomes. *Methods Enzymol.* **164**: 287–309.

Buccitelli, C., Selbach, M. 2020. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**: 630–644.

Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B., et al. 2010. A NusE:NusG complex links transcription and translation. *Science.* **328**: 501–504.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.

Catrein, I., Herrmann, R. 2011. The proteome of *Mycoplasma pneumoniae*, a supposedly “simple” cell. *Proteomics* **11**: 3614–3632.

Caudron-Herger, M., Jansen, R. E., Wassmer, E., Diederichs, S. 2021. RBP2GO: A comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic Acids Res.* **49**: D425–D436.

Chang, C.-H., Curtis, J. D., Maggi, L. B. J., Faubert, B., Villarino, A. V, et al. 2013. Post-transcriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**: 1239–1251.

Chavez, J. D., Bruce, J. E. 2019. Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Curr. Opin. Chem. Biol.* **48**: 8–18.

Chen, W. H., Van Noort, V., Lluch-Senar, M., Hennrich, M. L., Wodke, J. A. H., et al. 2016. Integration of multi-omics data of a genome-reduced bacterium: Prevalence of post-transcriptional regulation and its correlation with protein abundances. *Nucleic Acids Res.*

44: 1192–1202.

Chi, H., Wang, X., Shao, Y., Qin, Y., Deng, Z., et al. 2019. Engineering and modification of microbial chassis for systems and synthetic biology. *Synth. Syst. Biotechnol.* **4**: 25–33.

Cieśla, J. 2006. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim. Pol.* **53**: 11–32.

Commichau, F. M., Blötz, C., and Stülke, J. 2015. Methods in molecular biology of bacteria. *Dep. Gen. Microbiol. Georg. Göttingen*.

Commichau, F. M., Stülke, J. 2008. Trigger enzymes: Bifunctional proteins active in metabolism and in controlling gene expression. *Mol. Microbiol.* **67**: 692–702.

Conn, A. B., Diggs, S., Tam, T. K., Blaha, G. M. 2019. Two old dogs, one new trick: A review of RNA polymerase and ribosome interactions during transcription-translation coupling. *Int. J. Mol. Sci.* **20**: 1–14.

Cooke, A., Schwarzl, T., Huppertz, I., Kramer, G., Mantas, P., et al. 2019. The RNA-binding protein YBX3 controls amino acid levels by regulating SLC mRNA abundance. *Cell Rep.* **27**: 3097-3106.e5.

Corley, M., Burns, M. C., Yeo, G. W. 2020. How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* **78**: 9–29.

Curtis, N. J., Jeffery, C. J. 2021. The expanding world of metabolic enzymes moonlighting as RNA binding proteins. *Biochem. Soc. Trans.* **49**: 1099–1108.

Dabrazhynetskaya, A., Volokhov, D. V, Lin, T.-L., Beck, B., Gupta, R. K., et al. 2013. Collaborative study report: Evaluation of the ATCC experimental mycoplasma reference strains panel prepared for comparison of NAT-based and conventional mycoplasma detection methods. *Biologicals* **41**: 377–383.

Dallo, S. F., Kannan, T. R., Blaylock, M. W., Baseman, J. B. 2002. Elongation factor Tu and E1 β subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. *Mol. Microbiol.* **46**: 1041–1051.

Danchin, A., Fang, G. 2016. Unknown unknowns: essential genes in quest for function. *Microb. Biotechnol.* **9**: 530–540.

De St. Groth, S. F., Webster, R. G., Datyner, A. 1963. Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. *Biochim. Biophys. Acta* **71**: 377–391.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A. B., et al. 2017. Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* **6**: 1–17.

Ding, Y. H., Gong, Z., Dong, X., Liu, K., Liu, Z., et al. 2017. Modeling protein excited-state structures from “over-length” chemical cross-links. *J. Biol. Chem.* **292**: 1187–1196.

Doellinger, J., Schneider, A., Hoeller, M., Lasch, P. 2020. Sample preparation by easy extraction and digestion (SPEED) - a universal, rapid, and detergent-free protocol for

proteomics based on acid extraction. *Mol. Cell. Proteomics* **19**: 209–222.

Eaton, M. D., Meiklejohn, G., Van Herick, W. 1944. Studies on the etiology of primary atypical pneumonia: A filterable agent transmissible to cotton rats, hamsters, and chick embryos. *J. Exp. Med.* **79**: 649–668.

Frank, M. W., Yao, J., Batte, J. L., Gullett, J. M., Subramanian, et al. 2020. Host fatty acid utilization by *Staphylococcus aureus* at the infection site. *MBio* **11**: e00920-20.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science.* **270**: 397–404.

French, S. L., Santangelo, T. J., Beyer, A. L., Reeve, J. N. 2007. Transcription and translation are coupled in Archaea. *Mol. Biol. Evol.* **24**: 893–895.

Fuller, G. G., Han, T., Freeberg, M. A., Moresco, J. J., Niaki, A. G., et al. 2020. RNA promotes phase separation of glycolysis enzymes into yeast g bodies in hypoxia. *Elife* **9**: e48480.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., et al. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science.* **329**: 52–56.

Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., et al. 2014. Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. *PLoS Genet.* **10**: e1004363.

Güell, M., Van Noort, V., Yus, E., Chen, W. H., Leigh-Bell, J., et al. 2009. Transcriptome complexity in a genome-reduced bacterium. *Science.* **326**: 1268–1271.

Guo, X., Myasnikov, A. G., Chen, J., Crucifix, C., Papai, G., et al. 2018. Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* **69**: 816-827.e4.

Hames, C., Halbedel, S., Schilling, O., Stülke, J. 2005. Multiple-mutation reaction: A method for simultaneous introduction of multiple mutations into the *glpK* gene of *Mycoplasma pneumoniae*. *Appl. Environ. Microbiol.* **71**: 4097–4100.

Hentze, M. W. 1994. Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? *Trends Biochem. Sci.* **19**: 101–103.

Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., et al. 2016. Design and synthesis of a minimal bacterial genome. *Science.* **351**: aad6253.

Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., et al. 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**: 2165–2169.

Iacobucci, C., Piotrowski, C., Aebersold, R., Amaral, B. C., Andrews, P., et al. 2019. First community-wide, comparative cross-linking mass spectrometry study. *Anal. Chem.* **91**: 6953–6961.

Inamine, J. M., Ho, K. C., Loechel, S., Hu, P. C. 1990. Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae, Mycoplasma genitalium,* and *Mycoplasma gallisepticum*. *J. Bacteriol.* **172**: 504–506.

Johnson, G. E., Lalanne, J. B., Peters, M. L., Li, G. W. 2020. Functionally uncoupled transcription–translation in *Bacillus subtilis*. *Nature* **585**: 124–128.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589.

Kerr, A. R., Adrian, P. V, Estevão, S., de Groot, R., Alloing, G., et al. 2004. The Ami-AliA/AliB permease of *Streptococcus pneumoniae* is involved in nasopharyngeal colonization but not in invasive disease. *Infect. Immun.* **72**: 3902–3906.

Kim, S. H., Shin, D. H., Liu, J., Oganesyan, V., Chen, S., et al. 2005. Structural genomics of minimal organisms and protein fold space. *J. Struct. Funct. Genomics* **6**: 63–70.

Kim, Y., Li, H., Binkowski, T. A., Holzle, D., Joachimiak, A. 2009. Crystal structure of fatty acid/phospholipid synthesis protein PlsX from *Enterococcus faecalis*. *J. Struct. Funct. Genomics* **10**: 157–163.

Kohlbacher, O., Reinert, K., Gröpl, C., Lange, E., Pfeifer, N., et al. 2007. TOPP—the OpenMS proteomics pipeline. *Bioinformatics* **23**: e191–e197.

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., Cramer, P. 2017. Architecture of a transcribing-translating expressome. *Science.* **356**: 194–197.

Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, et al. 2014. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**: 1064–1070.

Krause, D. C., Balish, M. F. 2004. Cellular engineering in a minimal microbe: Structure and assembly of the terminal organelle of *Mycoplasma pneumoniae*. *Mol. Microbiol.* **51**: 917–924.

Kühner, S., Van Noort, V., Betts, M. J., Leo-Madas, A., Batisse, C., et al. 2009. Proteome organization in a genome-reduced bacterium. *Science.* **326**: 1235–1240.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Landick, R., Carey, J., Yanofsky, C. 1985. Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc. Natl. Acad. Sci. U. S. A.* **82**: 4663–4667.

Lenz, L. L., Dere, B., Bevan, M. J. 1996. Identification of an H2-M3-restricted *Listeria* epitope: implications for antigen presentation by M3. *Immunity* **5**: 63–72.

Levdikov, V. M., Blagova, E. V., Brannigan, J. A., Wright, L., Vagin, A. A., Wilkinson, A. J. 2005. The structure of the oligopeptide-binding protein, AppA, from *Bacillus subtilis* in complex with a nonapeptide. *J. Mol. Biol.* **345**: 879–892.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., et al. 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464–469.

Lluch-Senar, M., Delgado, J., Chen, W.-H., Lloréns-Rico, V., O'Reilly, et al. 2015. Defining

a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol. Syst. Biol.* **11**: 780.

Louche, A., Salcedo, S. P., Bigot, S. 2017. Protein–protein interactions: Pull-down assays. *Methods Mol. Biol.* **1615**: 247–255.

Mahamid, J., Baumeister, W. 2012. Cryo-electron tomography: the realization of a vision. *Microsc. Anal.* **26**: 45–48.

McGary, K., Nudler, E. 2013. RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* **16**: 112–117.

Merzbacher, M., Detsch, C., Hillen, W., Stülke, J. 2004. *Mycoplasma pneumoniae* HPr kinase/phosphorylase: Assigning functional roles to the P-loop and the HPr kinase/phosphorylase signature sequence motif. *Eur. J. Biochem.* **271**: 367–374.

Michalik, S., Reder, A., Richts, B., Faßhauer, P., Mäder, U., et al. 2021. The *Bacillus subtilis* minimal genome compendium. *ACS Synth. Biol.* **10**: 2767–2771.

Miles, R. J. 1992. Catabolism in mollicutes. *J. Gen. Microbiol.* **138**: 1773–1783.

Miller, O. L., Hamkalo, B. A., Thomas, C. A. 1970. Visualization of bacterial genes in action. *Science.* **169**: 392–395.

Mirdita, M., Ovchinnikov, S., Steinegger, M. 2021. ColabFold - making protein folding accessible to all. *BioRxiv* 2021.08.15.456425.

Neale, G. A., Mitchell, A., Finch, L. R. 1984. Uptake and utilization of deoxynucleoside 5'-monophosphates by *Mycoplasma mycoides* subsp. *mycoides*. *J. Bacteriol.* **158**: 943–947.

Nesterenko, M. V., Tilley, M., Upton, S. J. 1994. A simple modification of Blum's silver stain method allows for 30 minite detection of proteins in polyacrylamide gels. *J. Biochem. Biophys. Methods* **28**: 239–242.

O'Reilly, F. J., Rappsilber, J. 2018. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat. Struct. Mol. Biol.* **25**: 1000–1008.

O'Reilly, F. J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., et al. 2020. In-cell architecture of an actively transcribing-translating expressome. *Science.* **369**: 554–557.

Pashev, I. G., Dimitrov, S. I., Angelov, D. 1991. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem. Sci.* **16**: 323–326.

Pedreira, T., Elfmann, C., Singh, N., Stülke, J. 2022. SynWiki: Functional annotation of the first artificial organism *Mycoplasma mycoides* JCVI-*syn3A*. *Protein Sci.* **31**: 54–62.

Piñero-Lambea, C., Garcia-Ramallo, E., Martinez, S., Delgado, J., Serrano, L., Lluch-Senar, M. 2020. *Mycoplasma pneumoniae* genome editing based on oligo recombineering and Cas9-mediated counterselection. *ACS Synth. Biol.* **9**: 1693–1704.

Pollack, J. D., Williams, M. V., McElhaney, R. N. 1997. The comparative metabolism of the

mollicutes (*Mycoplasmas*): The utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit. Rev. Microbiol.* **23**: 269–354

Proshkin, S., Rachid Rahmouni, A., Mironov, A., Nudler, E. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science.* **328**: 504–508.

Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., Grierson, C. 2020. Designing minimal genomes using whole-cell models. *Nat. Commun.* **11**: 836.

Rodwell, A. W. 1969. A defined medium for *Mycoplasma* strain *Y*. *J. Gen. Microbiol.* **58**: 39–47.

Sanger, F., Nicklen, S., Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**: 5463–5467.

Sassanfar, M., Szostak, J. W. 1993. An RNA motif that binds ATP. *Nature* **364**: 550–553.

Schneider, M., Belsom, A., Rappsilber, J., Brock, O. 2016. Blind testing of cross-linking/mass spectrometry hybrid methods in CASP11. *Proteins Struct. Funct. Bioinforma.* **84**: 152–163.

Sharma, K., Hrle, A., Kramer, K., Sachsenberg, T., Staals, R. H. J., Randau, et al. 2015. Analysis of protein-RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. *Methods* **89**: 138–148.

Sharma, S., Tivendale, K. A., Markham, P. F., Browning, G. F. 2015. Disruption of the membrane nuclease gene (*MBOVPG45_0215*) of *Mycoplasma bovis* greatly reduces cellular nuclease activity. *J. Bacteriol.* **197**: 1549.

Shortall, K., Djeghader, A., Magner, E., Soulimane, T. 2021. Insights into Aldehyde dehydrogenase enzymes: A structural perspective. *Front. Mol. Biosci.* **8**: 1–15.

Simon, M. D., Wang, C. I., Kharchenko, P. V, West, J. A., Chapman, B. A., et al. 2011. The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108**: 20497–20502.

Slotboom, D. J. 2014. Structural and mechanistic insights into prokaryotic energy-coupling factor transporters. *Nat. Rev. Microbiol.* **12**: 79–87.

Stieger, C. E., Doppler, P., Mechtler, K. 2019. Optimized fragmentation improves the identification of peptides cross-linked by MS-cleavable reagents. *J. Proteome Res.* **18**: 1363–1370.

Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, et al. 2008. OpenMS - An open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**: 1–11.

Taylor, R. G., Walker, D. C., McInnes, R. R. 1993. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* **21**: 1677–1678.

Towbin, H., Staehelin, T., Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**: 4350–4354.

Vanden Broeck, A., Lotz, C., Ortiz, J., Lamour, V. 2019. Cryo-EM structure of the complete *E. coli* DNA gyrase nucleoprotein complex. *Nat. Commun.* **10**: 1–12.

Wach, A. 1996. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12**: 259–265.

Waites, K. B., Talkington, D. F. 2004. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev.* **17**: 697–728.

Wang, C., Molodtsov, V., Firlar, E., Kaelber, J. T., Blaha, G., et al. 2020. Structural basis of transcription-translation coupling. *Science.* **1365**: 1359–1365.

Wang, X. G., Kidder, J. M., Scagliotti, J. P., Klempner, M. S., Noring, R., Hu, L. T. 2004. Analysis of differences in the functional properties of the substrate binding proteins of the *Borrelia burgdorferi* oligopeptide permease (*opp*) operon. *J. Bacteriol.* **186**: 51–60.

Wang, X. G., Lin, B., Kidder, J. M., Telford, S., Hu, L. T. 2002. Effects of environmental changes on expression of the oligopeptide permease (*opp*) genes of *Borrelia burgdorferi*. *J. Bacteriol.* **184**: 6198–6206.

Warburg, O., Christian, W. 1942. Isolation and crystallization of enolase. *Biochem* **310**: 384–421.

Washburn, R. S., Zuber, P. K., Sun, M., Hashem, Y., Shen, B., et al. 2020. *Escherichia coli* NusG links the lead ribosome with the transcription elongation complex. *IScience* **23**: 101352.

Webb, A. J., Hosie, A. H. F. 2006. A member of the second carbohydrate uptake subfamily of ATP-binding cassette transporters is responsible for ribonucleoside uptake in *Streptococcus mutans. J. Bacteriol.* **188**: 8005–8012.

Webster, M. W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Weixlbaumer, A. 2020. Structural basis of transcription-translation coupling and collision in bacteria. *Science.* **3**: 1355–1359.

Wei, R. R., Richardson, J. P. 2001. Identification of an RNA-binding Site in the ATP Binding domain of *Escherichia coli* Rho by H2O2/Fe-EDTA cleavage protection studies. *J. Biol. Chem.* **276**: 28380–28387.

Weidlich, D., Klostermeier, D. 2020. Functional interactions between gyrase subunits are optimized in a species-specific manner. *J. Biol. Chem.* **295**: jbc.RA119.010245.

Winkler, C., Denker, K., Wortelkamp, S., Sickmann, A. 2007. Silver-and Coomassie-staining protocols: Detection limits and compatibility with ESI-MS. *Electrophoresis* **28**: 2095–2099.

Wise, K. S., Gonzalez, D. J., Sun, L., Breuer, M., Smith, H. O et al. 2019. Essential metabolism for a minimal cell. *Elife* **8**: 8–11.

Wodke, J. A. H., Puchałka, J., Lluch-Senar, M., Marcos, J., Yus, E., et al. 2013. Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol. Syst. Biol.* **9**: 653.

Wright, P. E., Dyson, H. J. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**: 321–331.

Yang, B., Wu, Y.J., Zhu, M., Fan, S.B., Lin, J., et al. 2012. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**: 904–906.

Yang, Z., Tsui, S. K. W. 2018. Functional annotation of proteins encoded by the minimal bacterial genome based on secondary structure element alignment. *J. Proteome Res.* **17**: 2511–2520.

Youil, R., Finch, L. R. 1988. Isolation and characterization of *Mycoplasma mycoides* subsp. mycoides mutants deficient in nucleoside monophosphate transport. *J. Bacteriol.* **170**: 5922–5924.

Yus, E., Lloréns-Rico, V., Martínez, S., Gallo, C., Eilers, H., et al. 2019. Determination of the gene regulatory network of a genome-reduced bacterium highlights alternative regulation independent of transcription factors. *Cell Syst.* **9**: 143-158.e13.

Yus, E., Maier, T., Michalodimitrakis, K., Van Noort, V., Yamada, T., et al. 2009. Impact of genome reduction on bacterial metabolism and its regulation. *Science.* **326**: 1263–1268.

Zeng, F., Peritz, T., Kannanayakal, T. J., Kilk, K., Eiríksdóttir, et al. 2006. A protocol for PAIR: PNA-assisted identification of RNA binding proteins in living cells. *Nat. Protoc.* **1**: 920–927.

Zhang, R., Xu, W., Shao, S., Wang, Q. 2021. Gene silencing through CRISPR interference in bacteria: current advances and future prospects. *Front. Microbiol.* **12**: 567.

Zhao, X., Yang, F., Wang, Y., Zhang, Y. 2021. *hns* mRNA downregulates the expression of GalU and attenuates the motility of *Salmonella enterica* serovar *Typhi*. *Int. J. Med. Microbiol.* **311**: 151525.

6 APPENDIX

6.1 Supplementary data

6.1.1 Known RNA-binding proteins

Table 15: Annotated proteins that were previously known to be RNA-binding proteins

Protein locus number	Sequence (XL position in lower case)	Score (Q value)	AA start	AA end	Annotated function
Carbon core metabolism					
0475-ldh	IIEWDLTkEEQEKFDK	0.000577	285	300	Metabolic process
	DADVVViTAGRPQKPGETR	0.000427	73	91	
	SCeTLHK	0.000451	301	307	
	AGTYEDCKDADVVVITAGRPQKPGETR	0.00046			
	VTGFDPHKVIGSGTTLDSAR	0.000632			
	EEQEKFDKSCETLHK	0.001609			
	ATFYGIGACLTK	0.00444			
	CGCLTLEDLDQIQK	0.002041			
0213-eno	KAANsTLIK	0.007373	351	359	Glucose transport & catabolism
	GVIKAVK	0.000389	62	68	
	VQIVGDDLFTTNPK	0.001992			
0221-pyk	lNFSHGSYEEHGYR	0.000589	41	54	Glucose transport & catabolism
	EILNKANANHIQIISK	0.00444			
0234-crr	KGDvvAILK	0.000707758	145	153	Glucose transport & catabolism
0606-pgk	KSLApVAK	0.000453	74	81	Glucose transport & catabolism
	VSAIGFLVEKELKM(Oxidation)LSQAVNEPK	0.001317			
	IVVPVDTACAK	0.002588			
0607-gapA	VDLVIEsTGFYTDKEK	0.000723	89	104	Glucose transport & catabolism
	LLDlNHKDVR	0.000516	185	194	
	yETRPIVSSDIIGSK	0.000613	276	290	
	QLVKVCSWyDNESSyVSQLVR	0.000466	309	329	
	DASNLPWSELKVDLVIESTGFYTDKEK	0.000632			
0779-ptsG	gNNIQIVYGGEQEAIKPR	0.000657	585	602	Glucose transport & catabolism
	LYTkADFk	0.000403	500	507	
	DGNAEAkLYTK	0.000499	493	503	
	VVDeMKVeKPK	0.000373	676	686	
	ASkGLNVDGSK	0.000531	508	518	
	MQkLLEQQR	0.000444	603	611	
	kADIDGIK	0.000607	566	573	
	SLGGTTGAIVK	0.00074	574	584	
	MLTQTNkSSFK	0.000349	0	10	
0131-fbaA	TTFLELTNWFGCQGK	0.000460193			Glucose transport & catabolism
0729-pgm	DLLELIKQNNNIGVISISGR	0.00204082			glycolysis
0227-pdhC	ISeLANKAK	0.000248	336	344	Metabolic process

	VNSEIPAPVAGK	0.001992			
0262-rpe	NTKNYK	0.00398804			Pentose phosphate pathway
DNA related					
0453-parC	SIAFDQPALANNFNDQDTNDQIEAKPK	0.00199203			DNA condensation/segregation, DNA topoisomerase IV subunit A
0452-parE	GIPTGMHKTGKSTPEVIFSVLHAGGK	0.00755985			DNA condensation/segregation, DNA topoisomerase IV subunit B
0097-exoR	ITYyGLIR	0.000404204	268	275	DNA metabolism, magnesium-dependent 5'->3' exonuclease
0254-uvrC	yFIVDNEK	0.000547537	62	69	DNA repair
0001-dnaA	KIeTMLK	0.000935156	424	430	DNA replication
0406-dnaG	IIVLEGFM(Oxidation)DVISLSK	0.00762631			DNA replication
0609-dnaB	SINYNYeTMFDSNDFNLQTeAIFeFSeeeLKCLV	6.88815E-05	360	393	DNA replication
0690-ligA	ALLRINQLKEQLNLWSK	0.00131694			DNA replication
0006-gyrB	INPNIFYVDSK	0.00535			DNA topology
	KFQAILPLR	0.006715			
0007-gyrA	LTdNQEINAITLEYRK	0.000486	807	822	DNA topology
	ITLEMTNIK	0.005941			
Nucleotide metabolism					
0203-gmk	GNNVILEIEVDGATQVLNK	0.00782779			nucleotide salvage
0330-dgk	tQIELIMNK	0.000505063	188	196	Nucleotide salvage
0549-rgdB	LIQyIDNVK	0.000639	191	199	Nucleotide salvage
	KAYM(Oxidation)TTAIALYDAINK	0.001992			
0798-upp	gYIVPgLgDAgDR	0.000298638	189	201	Nucleotide salvage
0831-prs	ESQEHHSLTEVyNK	0.000449	313	326	Nucleotide salvage
	VHKVDSYTANM(Oxidation)TNGIAVIGKR	0.002588			
	VHKVDSYTANMTNGIAVIGK	0.001992			
0045-tmk	VIKPAIEK	0.00042736	77	84	Nucleotide salvage
50S ribosome					
0198-rplT	GyFGTK	0.000442	22	27	
	SMAyAFIGR	0.000684	41	49	
	SLwIVR	0.000461	58	63	
	LAkGYFGTK	0.000595	19	27	
	kAHEQVIR	0.000612	33	40	
	YGkVTR	0.006922	5	10	
	WIkLAK	0.000424	16	21	
0365-rplS	IyyIR	0.000597	103	107	
	IQSFEGlVIK	0.000371	46	55	
	TQGsGITYsVVVR	0.000406	56	68	
	NLSGkAAR	0.000319	108	115	
	EIMPTkQAK	0.000256	118	126	
	IKEIMPTk	0.000455	116	123	
	mSNGVFVER	0.000526	70	78	
	KMSNGvFvER	0.000743	69	78	
	IKEImPTKQAK	0.000412	116	126	
0501-rplU	VEPGQEIfIEK	0.000525124	13	23	

	VLgTIIK	0.000389	54	60
	VLGTIIkQGK	0.000629	54	63
	mFAIIK	0.000273	0	5
	MFAIIkTGGK	0.00049	0	9
	VKiDEiSAK	0.000545	91	99
	vIRYHPK	0.000618	69	75
	NVNkIYGHR	0.000416	77	85
	QVKVEPGQEIFIEK	0.00046		
	TGGKQVKVEPGQEIFIEK	0.00046		
	VEPGQEIFIEKIKGEVNDK	0.000632		
0638-rplM	ESNKFYYHHSMHPGGLK	0.00069539	69	85
	KESNKfYYHHSMHPGGLK	0.000249	68	85
	FYYHHSMHPGGLK	0.000296	73	85
	GSQHPFaAQKPEVLEISTK	0.000758	127	145
	MKQTTMIsAK	0.000338	0	9
	lMLPK	0.000437	107	111
	NVQGSNQyR	0.000638	112	120
	SvEvQRELDATK	0.000402	88	99
	wYIVDAENK	0.000478	15	23
	QTTmISAK	0.000825	2	9
	ELDATkILER	0.000238	94	103
	QTTMISAkDINKK	0.000404	2	14
	AIFsGKK	0.005913	62	68
	KESNKFYYHHSM(Oxidation)HPGGLK	0.00046		
	ESNKFYYHHSMHPGGLKK	0.00046		
	GSQHPFAAQKPEVLEISTKKGDVK	0.00046		
	ESNKFYYHHSM(Oxidation)HPGGLK	0.002041		
0644-rplQ	NGGYTsILK	0.000377525	93	101
	DIDaDKKETaLQK	0.000406	68	80
	LDNRKGDNAPmVIIK	0.000643	102	116
	KGdNAPMVIIK	0.000373	106	116
	GDNAPmVIIK	0.000508	107	116
	QAASwLR	0.000925	61	67
	LFNkLAk	0.000449	81	87
	HFdHMITLAK	0.000341	43	52
0653-rplO	ILANGtLtK	0.000443384	109	117
	DAIEkLGGK	0.000304	132	140
	VGFtSLNQK	0.000354	63	71
	SGGGVRPGFeGGQTPLFR	0.000701	41	58
	MKLNELK	0.000492	0	6
	ILANGTLTkK	0.000385	109	118
	IlkNNASLIK	0.000275	99	108
	tKATIVGR	0.000445	13	20
	lNElKYTPGSK	0.000407	2	12
	VGFTSLNQKQYTILNLSDLETLGLEK	0.00046		
0655-rplR	VAeeIAKK	0.000446971	72	79

	FTkAEAR	0.000475	2	8
	IKaFaETaR	0.000583	101	109
0656-rplF	KPEPYK	0.000345569	153	158
	yKNETIIR	0.000686	164	171
	IEVEENkLITK	0.000691	43	53
	KELQITgVgYK	0.000397	85	95
	ELQITGVGyK	0.000905	86	95
	AYRKPEPYKGK	0.000891		
	IAENNLITITGSKGTLSKQFSPLIK	0.001992		
0659-rplE	SIMQVPkIQk	0.000460021	24	33
	kSLAVFK	0.000573	72	78
	LSgQKPIVTKAK	0.000616	60	71
	MYDFLDK	0.000277	96	102
	KMyDFLDK	0.000691	95	102
	kLDAAISELEK	0.000476	49	59
	SImQVPK	0.00059	24	30
	LSGQKPiVTK	0.000354	60	69
	SLAvFK	0.000501	73	78
0660-rplX	GDVVkVIAGSHK	0.000506101	8	19
	GQIGPITSITkDK	0.00034	20	32
	KHVKPTNEDSEggIK	0.000382	44	58
	VGfEIIDGK	0.000408	83	91
	iLKGDVVK	0.000338	5	12
	VGFEIIDGKk	0.000757	83	92
0661-rplN	IfGPIAR	0.000740092	97	103
	MIQTLSkLK	0.000453	0	8
	GQVVkAVIVR	0.000395	54	63
	EIkEAGFAk	0.000533	104	112
	LKvADNSGAK	0.000597	7	16
	mIQTLSK	0.000291	0	6
	KGQVVkAVIVR	0.000408	53	63
	IAsLAPEVL	0.000291	113	121
	GTRIFGPIAR	0.000632		
0664-rplP	wVAVVK	0.00066129	92	97
	VSyEGKAK	0.0004	14	21
	iFPHMAMTK	0.000327	67	75
	EINFGEFGLM(Oxidation)ALDGAWIDNHQIEAAR	0.000632		
	YRKPHRVSYEGK	0.006715		
	KPHRVSYEGK	0.009524		
0666-rplV	AyEIFKR	0.000907112	92	98
	TIFVNeGPTLKR	0.000466	72	83
	AYEiFK	0.000375	92	97
	TIFVNeGPTLK	0.000502	72	82
	TSHvvIvvSDEK	0.000324	99	110
	NLnKDAAEPILK	0.000481	37	48
	lVADTIR	0.00032	18	24
	LSmIR	0.000365	6	10
	AKlSMIR	0.000402	4	10

0668-rplB	GSVMNPNDhPhGGGEGR	0.000243659	225	241
	gIRPTVR	0.000521	218	224
	KSPVTPWGK	0.000504	247	255
	NMTTIDySAVLTTK	0.000346	14	27
	IIdFKR	0.00035	65	70
	yILFAK	0.000585	104	109
	YrIIDFK	0.00059	63	69
	kALGVK	0.000528	256	261
	NNPeKSLVVSK	0.000384	28	38
	TSEkLIVR	0.00051	268	275
	SLVVSkNSK	0.000414	33	41
	ALGVkTR	0.007495	257	263
	KTSEkLIVR	0.000621	267	275
	NNRGIITTR	0.000416	45	53
	SPVTPwGK	0.000282	248	255
	VLAECYATIGEVGNEEYNLVNWGK	0.00046		
	GSVMNPNDHPHGGGEGRAPIGR	0.001155		
	KSPVTPWGKK	0.001609		
0669-rplW	KPVLTEK	0.000434057	8	14
	kAIITLK	0.000506	76	82
	ANkVQIK	0.000432	32	38
	AIITLkEGQK	0.000434	77	86
	LGkYVGK	0.000475	64	70
	mHITEVLK	0.000549	0	7
	YVgKKPSYK	0.000395	67	75
	kTFEEIFEVK	0.000482	39	48
	YVGKKPSYKK	0.00046		
	LGKYVGKKPSYK	0.00046		
0670-rplD	GGGVTFGPTPDINyK	0.000673202	83	97
	MKLQVLDIK	0.000356	0	8
	GGGVTfGPTPDINYKK	0.000956	83	98
	SVlSIKVK	0.00042	111	118
	TLIVTkEK	0.000367	152	159
	IDDqKTLIVTK	0.000465	147	157
	FEFAKpSTK	0.000954	128	136
	TLIVTKEKEELVVK	0.001609		
	GTGRARQGSIR	0.009524		
0671-rplC	VNLvNKPELGHFK	0.000574781	56	68
	VNLVNkPELGHFkk	0.000289	56	69
	MKGILGR	0.000305	0	6
	gIgSMgAIINR	0.000677	139	149
	GFaGGIK	0.000471	115	121
	GKGFAGGIK	0.000457	113	121
	gFAggIKR	0.000592	115	122
	IFkSK	0.006299	150	154
0806-rplL	VVkELTGVGLMDAK	0.00055087	71	84
	VAVIkVVK	0.00058	66	73
	ELTgVgLMDAK	0.000461	74	84

	ELTGVGLMDAkK	0.000368	74	85
0807-rplJ	SAQGVAIAEyK	0.000887989	23	33
	KAeIVAeIVSK	0.000343	10	20
	KAEIVAEIVSKIK	0.00046		
0809-rplA	NVTLSTTmGPGIK	0.000699411	210	222
	VLVLTNTkTK	0.009987	74	83
	FkEALSK	0.000563	7	13
	FKEALSkVEK	0.000366	7	16
0810-rplK	TTpTAYMLK	0.000496646	67	75
0833-rplI	LQTNDDKVFGsIssQD IVNQLK	0.000812632	80	101
0137-rpmE	VEqFK	0.000337353	59	63
	yFDQAK	0.000332	9	14
0199-rpmI	RVTVkSNGTLK	0.000476625	12	22
	VTVkSNGTLKR	0.000282	13	23
	QLsKATIISK	0.00044	42	51
	VDmKNLK	0.000443	52	58
	ATIISkVDMK	0.000446	46	55
	VTVKSNGtLK	0.000601	13	22
0422-rpmB	wNLNLQK	0.000700032	28	34
	SHALNATkR	0.000305	18	26
	VMdENGSVFNIK	0.000344	37	48
	SHaLNaTK	0.000361	18	25
	DALTgKSALSgQSR	0.000525	4	17
0499-rpmA	GTkIHPGLNVGR	0.000527876	50	61
	YEKFGKNR	0.007185		
	KSDGQFTNAGSIIYR	0.000632		
0648-rpmJ	VMIIcVTPK	0.000413557	22	30
0663-rpmC	SKmLDLR	0.000299351	3	9
	AMQAAEQETVEPDT KGETK	0.002588		
	KEIARIELALSEK	0.000632		
	KFLEEQYGQQSQTE LNEADIQK	0.002041		
	IELALSEK	0.001992		
0910-rpmH	TWqPSK	0.000690982	3	8
	VHgFR	0.000446	14	18
30S Ribosome				
0025-rpsR	kVNFFQK	0.000624	10	16
	DQRQlAlAIK	0.000449	52	61
	QLALAIkR	0.000421	55	62
	ITGTSpKDQR	0.000475	45	54
	DIELLKk	0.007016	26	32
	VNFFQkNNIK	0.000287	11	20
	NNIkYIDYk	0.000208	17	25
	QMALLPyVIE	0.000405	65	74
	QLaLaIK	0.000339	55	61
	FISPNGQILpR	0.000639	33	43
0027-rpsF	KPfIK	0.000664981	122	126
	IDkNVVR	0.000453	78	84

	KNFRKPFIK	0.00046		
	NFRKPFIKR	0.000632		
	NFRKPFIK	0.003074		
0082-rpsT	LANkAFK	0.00054131	18	24
	VLTNEKSR	0.000361	10	17
	SEIkTAIk	0.000537	25	32
	AFkSEIK	0.00059	22	28
	SEIkTAIkk	0.000485	25	33
	kGIFK	0.000367	62	66
0148-rpsL	VTkIASPQKR	0.000486642	34	43
	KVTkIASPQK	0.000583	33	42
	VATMTpKKpNSALR	0.000287	49	62
	gINTLIK	0.000402	26	32
	SLyGVK	0.000651	127	132
	GINtLIKK	0.000676	26	33
	GTIDTQGVAK	0.000402	112	121
	yHVIR	0.000878	107	111
	kPNSALR	0.00037	56	62
	TKApALNR	0.000341	18	25
	GTLDTQGVAkR	0.000488	112	122
	VKdLPGVR	0.000415	99	106
	VTkIASPQK	0.000526	34	42
	SLyGVKR	0.000365	127	133
0149-rpsG	IGGANyQVPVEVSAER	0.00071193	78	93
	IMLDGK	0.00033	28	33
	VMtIK	0.000552	113	117
	NEKVmTIK	0.000719	110	117
	DVLADPIYNSK	0.006852	9	19
	WLINyAR	0.000767	101	107
	LRNEKVmTIK	0.000659	108	117
	AFAHyR	0.000691	148	153
	IMlDGKR	0.000339	28	34
	NEKVMTIKLANEIIDASNNIGGSVK	0.00046		
	VMTIKLANEIIDASNNIGGSVK	0.000632		
	RIGGANYQVPVEVSAER	0.001609		
	MAEPNKAFAHYRW	0.00353		
	HYAHVDCPGHADYVK	0.000632		
	SLPHVNIGTIGHVDHGK	0.000632		
0238-rpsD	fIGSTFK	0.000541576	3	9
	FMYGLsER	0.000517	62	69
	FGFSILETGk	0.000417	14	23
	FIgSTFKK	0.000569	3	10
	VSEyGQQLQEK	0.000367	47	57
	NTFaKaK	0.000453	73	79
	ITTpGQHGK	0.000532	32	40
	ITTPGQHGkER	0.009942	32	42
	VSEYGQQLQEkQK	0.000463	47	59
	LDNIVyR	0.000655	97	103

	VKVSEYGqqLqEK	0.000696	45	57
	VnKnEVKGEFVR	0.000443	170	181
0294-rpsO	YkEVIEK	0.000428283	76	82
	KdIPTR	0.00687	46	51
0362-rpsP	IDkELTLK	0.000488738	47	54
	ALHEAkFSK	0.00038	78	86
	IVaaDSRINR	0.000473	18	27
	kQAPFYR	0.000499	11	17
0482-rpsU	EYHLSKk	0.000415667	33	39
	FQkVASSNK	0.009981	18	26
0540-rpsB	EELSAAGVQyGHQTK	0.000380594	7	21
	ALGGIKQMHK	0.000286	147	156
	ILFVGTkR	0.000421	67	74
	LLETLGSkQQK	0.000356	56	66
	aKLEKALGGIK	0.00032	142	152
	SylyGVK	0.000395	29	35
	wLGGTLTNLK	0.000505	96	105
	SGNFyINNR	0.000717	87	95
	NKNhIIDLEK	0.000752	36	45
	EQIkILK	0.000563	133	139
	NAVkEAALR	0.000551	78	86
	MKsYIYGVK	0.000453	27	35
	LEkALGGIK	0.000588	144	152
	GDEIAVkEAK	0.000649	168	177
	QM(Oxidation)HKLPAAIVVVDPK	0.001317		
	EELSAAGVQYGHQTKR	0.003988		
0637-rpsI	VIgggFTgQAgATR	0.000766171	68	81
	DQGllTRDAR	0.000475	101	110
	kYGLR	0.000438	116	120
	RAPQYSkR	0.000418	124	131
0646-rpsK	GIAHIHSTFnNTIVTVSDEK	0.000791608	18	37
	STPyAAQLISEAAAK	0.000342	57	71
	kNIPK	0.000367	13	17
	TVSVEVkGPGPGR	0.000373	80	92
0647-rpsM	ISGvEIPNNKR	0.000474896	3	13
	VKdLTEEQIK	0.000649	44	53
	YKTEGELRR	0.00031	62	70
	NISmEISK	0.000498	54	61
	LMEIGSyR	0.000674	79	86
	RLMEIGSYR	0.00042	78	86
	kGLPVR	0.007444	92	97
	EVSLnIKR	0.00058	71	78
	KGLPVrGQSSK	0.000474	92	102
	VVVSLTYIYGIGLPTAQSVLK	0.00046		
0654-rpsE	ATFEGLSSMQTLKR	0.000613813	175	188
	AVIELAGISDVyAK	0.0009	149	162
	iLIKPAK	0.000398	130	136
	sLGsNNAINMIR	0.000374	163	174

	VGTGIIAGGPAR	0.000765	137	148
	NTTVPHeVLGTFGAGK	0.000327	114	129
	ANeVPeAIKK	0.00055	89	98
	VKPVEQK	0.00024	203	209
	kNLVSVTLR	0.000507	105	113
	VAEVkSVEK	0.000562	210	218
	kGLVGMGTGK	0.000586	79	88
	KQPkQVVK	0.0004	219	226
	GLVGmGTGK	0.000414	80	88
	VKPVEQKVAEvK	0.00029	203	214
	tFDtQKVKPVEQK	0.000601	197	209
	ATFEGLSSMQTlK	0.00079	175	187
	sKQVKDEFEEK	0.000705	40	50
	QVKDEfEEK	0.00039	42	50
	YGKTFDTQK	0.00046		
	ATFEGLSSM(Oxidation)QTLKR	0.00046		
0657-rpsH	KiSKPGLR	0.000259107	75	82
	VIQGlKK	0.000381	69	75
	YLkTVSVPSSK	0.000452	19	29
	mTTDVIADMLTR	0.000338	0	11
	vKLEIAR	0.000385	30	36
	ISkPGLR	0.000406	76	82
	TVSVpSSKVK	0.005792	22	31
	VYAQANEIPQVLNGLGISIVSTSQGIMTGK	0.00046		
0658-rpsN	HQkFNVR	0.000340335	12	18
	FAYeGQIPGIK	0.000668	46	56
	FGIcR	0.000307	36	40
	HQKFNVRNYTR	0.001609		
0662-rpsQ	LVrVIEK	0.000505325	75	81
	VLIGkVVSDK	0.000276	7	16
	RVLIGK	0.000564	6	11
	NHPiYK	0.000474	30	35
	TITVLVETyK	0.001107	20	29
0665-rpsC	VWINhGEVFKK	0.000562606	202	212
	LGGVEmAR	0.000345	159	166
	IVlAVR	0.000339	86	91
	TVQkLAIK	0.000489	134	141
	kLIVNVR	0.000334	98	104
	wLDQDIK	0.000532	31	37
	DaaVSKIDIER	0.000549	48	58
	VWiNHGEVFK	0.000663	202	211
	TALfKLLK	0.000335	40	47
	lGIVR	0.000381	11	15
	TTKdLTLFIK	0.009912	59	68
	MNNSQImAKPR	0.00032	215	225
	NIEkIVLAVR	0.000472	82	91
	LLkDAAVSK	0.000637	45	53
	VSpNVLR	0.000376	4	10

	GIKtAVSGR	0.000545	150	158	
	PAIVLGQEGkNIEK	0.000642	72	85	
	TEGyLEGSVPLSTLR	0.000482	167	181	
	DLTLfIK	0.000531	62	68	
	wIGEQISNR	0.000636	121	129	
0667-rpsS	KGPFVDESIFK	0.000639169	6	16	
	DGEVIkTWSR	0.00033	23	32	
	TFGVyNGK	0.000498	44	51	
	GPfVDESLFK	0.000604	7	16	
0672-rpsJ	GYDHAiVDQSiVK	0.000604055	11	23	
	IIQAAeGTGAKVR	0.000388	24	36	
	LLEILNPTAATMDILK R	0.00046			
	LLEILNPTAATM(Oxid ation)DILK	0.00046			
	LLEILNPTAATMDILK	0.00046			
RNA synthesis and degradation					
0257-rnjB	KYDLNeIK	0.005809057	519	526	RNAses, Exoribonucleases
0359-rny	LDAgILDLEK	0.000737085	107	116	RNAses, Endoribonucleases
	KQVLANGyK	0.001087	51	59	
0600-rnjA	GSFyVK	0.000704091	542	547	RNAses, Exoribonucleases
0003-rnmV	KVGIAEASDDAIK	0.00762631			RNAses, Endoribonucleases
0004-ksgA	VVVIeIDK	0.000443176	53	60	rRNA modifications and maturation
Translation					
0061-serS	KeILVK	0.000466035	40	45	Aminoacyl-tRNA synthase
0064-lysS	ALRPLPDKHAGIQDIE EK	0.000632111			Aminoacyl-tRNA synthase
0150-fusA	GVINADFFPVLAGSAf KNK	0.000624279	246	264	Translation factors, Translation elongation factor G
0151-tufA	HKPFfNK	0.000366328	320	326	Translation factors
	gQVLAKPGTIKPHTVL K	0.000554	290	306	
0200-infC	IVNfGKLK	0.000367352	69	76	Translation factors, Translation initiation factor IF-3
	LKyEQQK	0.000344	75	81	
	vKvSLK	0.000416	126	131	
	ITVNIgKHDLETK	0.00035	100	112	
0202-rsmD	MHiiSGK	0.000290911	0	6	rRNA modifications and maturation
0240-thiI	YGELTLkGNNR	0.005751736	7	17	tRNA modifications and maturation
	FIDTYETSILPFDDVC SMFVPKDPIIKPK	0.003074			
0263-cpgA	SFLENKISVFTGQTG AGK	0.00797314			Ribosome assembly
0287-aspS	YVAAHHPFTSPK	0.00199203			Aminoacyl-tRNA synthase
0289-rbfA	kESLLLR	0.000501765	9	15	Translation factors, ribosome biogenesis
	IFYQFIPIPENLTIQSIE EELENK	0.009524			
0308-trpRS	QIMVSGITPSGTMTL GNYLGVVK	0.00635863			Aminoacyl-tRNA synthase
	QVEQWLELGANK	0.001992			
0329-rluB	LVLTTMyDPK	0.000697488	74	83	rRNA modifications and maturation
0348-engA	WDLIK	0.007054881	294	298	Ribosome assembly, Ribosome biogenesis GTPase
0361-rlmH	IKELNSDLLLNK	0.00948767			rRNA modifications and maturation
0390-fmt	SSKVIDLDNQNKNVM PGTIIDINK	0.00199203			tRNA modification and maturation

0434-rlmFO	IYSLKPMK	0.000277127	387	394	tRNA modification and maturation
	ILNFKLQVM(Oxidation)RK	0.005789			
0519-ileS	KPVIyR	0.000958521	406	411	Aminoacyl-tRNA synthase
0528-pheT	IYNVFDHNLDLKKYIGSDVK	0.00671463			Aminoacyl-tRNA synthase
	LVNIIGLDTQNEFNVDNNSK	0.002041			
	LEFINHDNKELLCLTNNNK	0.001992			
0535-argS	IFHSYyAEIK	0.000571301	505	514	Aminoacyl-tRNA synthase
	KITPLIK	0.00046			
0539-tsf	LIkELR	0.000396845	6	11	Translation factors
	FLSLVDEIANALLNSNASSLEEGLQVK	0.001992			
0548-cspR	INIVLYQPEIAQNVGAIM(Oxidation)R	0.00988701			tRNA modification and maturation
0640-truA	TIDDIKVLRINNK	0.00952381			tRNA modification and maturation
0650-map	TICISINDQLIHGIPK	0.00808754			Translation factors; Chaperones, protein folding and maturation
Ungrouped					
0168-oppF	DLLIeFGNGR	0.000481045	13	22	Amino acid acquisition
	NEAILKVRDLLIEFGNGR	0.003868			
0793-atpH	KVEIVNK	0.000432219	139	145	ATP synthesis
0522-ftsZ	TVSLNDAYDAVGVISQAVNNK	0.00204082			Cell envelope and cell division, pentose phosphate pathway
	VVEGADLIFIAAGM(Oxidation)GGGTGTGAAPVIAK	0.002041			
0523-ftsA	TSSTTNLEyIVK	0.000977144	103	114	Cell envelope and cell division
0542-dnaK	TiTiSNSGNLSEAEVER	0.000304085	469	485	Chaperones, protein folding and maturation
0543-grpE	kNQYLNLK	0.000608824	32	39	Chaperones, protein folding and maturation
	NQYLNLkTK	0.000319	33	41	
	TkAEFQK	0.000524	21	27	
	NQyLNLK	0.000776	33	39	
	QLNLAEISNLTK	0.00353			
0378-nadE	KAAQILNVPEIIINR	0.000460193			Cofactor acquisition
0381-mtnN	VISDVM(Oxidation)FVSDSNM(Oxidation)LQFDQFINK	0.00535019			Cofactor acquisition
0823-folC	VLEYLNIFQTNIFQLQPPLGR	0.00398804			Cofactor acquisition
0115-galU	vCSVVEKPDEQNAPSNVAIAGR	0.000877299	182	203	Lipid metabolism
	FLYNK	0.007187			
0427-pstB	DKKILDQICEESLVK	0.00444006			Phosphate metabolism, Phosphate ABC transporter ATP-binding protein
0095-secA	FGkIADK	0.000390692	10	16	Protein secretion
0360-ffh	VLkAISR	0.000327732	389	395	Protein secretion, Signal recognition particle protein
0065-trxA	IMSIPTLITFENGNQVNK	0.00204082			Redox homeostasis
0407-rpoD	IGQAPILTK	0.00920502			RNAP
0645-rpoA	VAySVDSAK	0.00043601	181	189	RNAP
	SLkEIK	0.000376	298	303	
	SLNcLK	0.00041	265	270	
	SLNcLKR	0.000526	265	271	
0804-rpoB	VIAEDIVDANNNVLVAK	0.00204082			RNAP

0300-nusA	SKVAVITHNNNVEPI GAIIGVGGNR	0.002041		Transcription elongation factor of RNA polymerase

6.1.2 New RNA-binding proteins

Table 16: Unclassified previously annotated RNA-Binding proteins

Protein locus number-gene	Sequence (XL position in lower case)	Score (Q value)	AA start	AA end	Current annotation
0009-rnsC	TvLGFK	0.000532	229	234	Ribose transporter in *E. coli*, nucleotide permease in *B. subtilis*
	kAALVLK	0.000485	670	676	
0169-oppA	IVTKEQFK	0.007185			Oligopeptide ABC transporter substrate-binding protein
0195-potC	AGnSInFIYGR PnK	0.000225	896	909	Spermidine/putrescine ABC transporter permease
0264-prkC	KDLGINK	0.009488			Serine/threonine protein kinase C
0303-polC	QIITNIINIAKQ ENK	0.007828			PolC-type DNA polymerase III
0305-papA	QDLIDILNQN NVK	0.007828			Degradation of proline-containing peptides, Xaa-Pro amino-peptidase
0327-scpA	KDLVFENPD PLIDLNDLDL DK	0.001992			Participates in chromosomal partition during cell division. May act via the formation of a condensin-like complex containing Smc and ScpB that pull DNA away from mid-cell into both cell halves.
0371-ywjA1	LStIR	0.000325	575	579	Similar to multidrug ABC transporter, Lipid A export ATP-binding/permease protein MsbA in *E. coli*
0420-fakA	KLeLLK	0.000402	2	7	Fatty acid kinase
	IyVIQTK	0.000909	405	411	
0430-ylxM	QDlIISlINK	0.000291	78	87	Uncharacterized DNA-binding protein, Might take part in the signal recognition particle (SRP) pathway. This is inferred from the conservation of its genetic proximity to *ftsY*. May be a regulatory protein
0512-plsC	MNQVNSMQ EPiSQNNK	0.000256	0	15	acyl-ACP:1-acylglycerolphosphate acyltransferase
	LIDcIPLNR	0.000435	144	152	
0617-fakB2	LIkNLSNK	0.000338	215	222	Binds long-chain fatty acids, such as palmitate, and plays a role in lipid transport or fatty acid metabolism
0643-ecfA1	LFkQNLK	0.000455	77	83	ATP-binding A1 component of ECF transporters
	LNDLNNQyIN LIR	0.000546	33	45	
	ISAttKLFK	0.000346	71	79	
	LAEsLYKK	0.00045	84	91	
0706-thiB	KPFLVK	0.009887			Thiamine ABC transporter permease
0787-mgtA	KfSPK	0.000333	7	11	Magnesium-translocating P-type ATPase
	KADPTLLK	0.007626			
0817-whiA	MSfALEVK	0.006167	0	7	Uncharacterized DNA-binding protein
0822-ecfS3	TfTIK	0.000264	36	40	Folate ECF transporter S component
0887-cdr	KVAVIGSGFI GLECCEMLE HFNK	0.005957			Redox homeostasis, Coenzyme A Disulfide Reductase
0908-yidC	VMSyLNASK	0.000552	7	15	Sec-independent membrane protein translocase

Table 17: Uncharacterized proteins that bind RNA.

Protein locus number	Sequence (XL position in lower case)	Score (Q value)	AA start	AA end	Essentiality
0030	LDKSDFIVILGPSGSGK	0.002588			Essential
	FEPGIkGLK	0.000346	65	73	
	YSkDILNK	0.000472	86	93	
	syITGDLETPVLK	0.000458	119	131	

	TIkSEPVLLEQM(Oxidation)DQR	0.000418	40	54	
0034	NNDSNNAEILTPYK	0.000542	936	949	Quasi-essential
0060	YFcNLK	0.000433	3	8	Quasi-essential
0138	HLSEKELLQLFETIK	0.005789			Quasi-essential
0317	kpSEADIKK	0.000389	63	71	Essential
	TM(Oxidation)NAIkR	0.000256	72	78	
	DNPPITEkqiR	0.000317	44	54	
0338	EIIkSAFVK	0.00042	60	68	Quasi-essential
	TLVkVTK	0.000502	139	145	
	SAFVKQNGLNDPK	0.000421	64	76	
	TDGkVLVK	0.000472	180	187	
	SIQEEKIDLNK	0.000332	34	44	
	DLGFVSKNEK	0.000633	50	59	
0346	NISTNDLVDLKNYSDFIVLDEQk	0.000573	94	116	Quasi-essential
0352	EVEIfLKSTNK	0.006134	159	169	possibly Essential
0388	kLNISNK	0.000458	188	194	Essential
	kFVPSR	0.000341	195	200	
	LKEVEKIQAK	0.000485	148	157	
	TIkLTK	0.000464	137	142	
	IAIGEkTIK	0.000498	131	139	
	GSLDIyqeR	0.00052	179	187	
0439	NDNKSASNkK	0.007027	317	326	Essential
0444	LKNNTWLTSQTINK	0.003447			Non-essential
	KQANNLLDDPNLINFAK	0.002041			
	KDVENMVSHMINIYK	0.001992			
0451	LTIGLPKDNCDITPLIDQK	0.002041			Essential
0493	KIDNSLYVHQESDLIKK	0.00353			Quasi-essential
0546	KIEQIIK	0.001155			Non-essential
0602	KPQQIDK	0.002041			Non-essential
0636	KIAMSLR	0.001992			Essential
0827	LSkQEQLK	0.000511	139	146	Quasi-essential
0835	LTKKDDSYINLSNNGNNNNQFVYNINQK	0.002041			Essential
0852	KEQNLQVYK	0.001992			Non-essential
0878	LIGGLcLK	0.000361	480	487	Quasi-essential

6.2 Bacterial strains

Bacterial strains used in this work.

Name	Genotype/Description	Reference
E. coli		
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac[F´proAB lacI qZΔM15 Tn10 (Tetr)]*	Agilent Technologies
DH5α	F− *endA1 hsdR17 (rK−, mK+) supE44 thi-1 λ− recA1 gyrA96 relA1 deoR Δ(lacZYA-argF)U169 φ80ΔlacZΔM15*	Taylor et al., 1993
BL21 (DE3)	*B F− lon ompT hsdS(rB−, mB−) gal dcm lacI lacUV5-T7 gene1 ind1 sam7 nin5 [mal+]K-12(λS)*	Miroux, B., & Walker, J. E. (1996). Over-production of proteins in Escherichia coli
	F⁻ *ompT gal dcm lon*? *hsdS*$_B$($r_B^-m_B^-$)λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	
Rosetta	pLysSRARE[*T7p20 ileX argU thrU tyrU glyT thrT argW met T leuW proL ori*$_{p15A}$](Cm^R)/ for protein over-expression	Novagen

C43 (DE3)	*F – ompT hsdSB (rB- mB-) gal dcm (DE3)*	Lucigen
M. pneumoniae		
M129 (ATCC 29342)	Wild type	Somerson et al., 1963
GPM116	harboring GP35 recombinase (*mpn560*)-*arcA*::GP35-Puro^R	Piñero-Lambea et al., 2020
Synthetic cell		
M. mycoides J CVI-Syn1.0	originating from the sMmYCp235 synthetic genome, considered wild type *Mycoplasma mycoides* subspecies *capri* strain GM12	Gibson et al., 2010
Syn3	Unstable strain, with 149 uncharacterized genes	Hutchison et al., 2016
Syn3A	Working strain, same as *Syn3* but with 18 genes added back, including *sepF* and *ftsZ*	JCVI, Wise et al., 2019
Syn3A_mcherry	mCherry gene introduced into *Syn3A* via a Tn-mediated insertion	JCVI, Wise et al., 2019

6.3 Oligonucleotides

Oligonucleotides were purchased from Sigma Aldrich (Munich, Germany). Underlined are restriction sites.

Name	Sequence (5'-3')	Gene/Purpose
NS46	*GAATTTAAGGTTAAGGGAACCCCA*	Fwd, locationchk
NS47	AAACTCGAG *TAACACCGGTTTTGACCTGG*	Fwd, *nusA*_MPN154_I418 truncation mutant_HRupstrm_*XhoI*
NS48	GTATAATGTATGCTATACGAACGGTA *TCGTTAATCTAACCATTTCACACTATG*	Rev, *nusA* _MPN154_I418 truncation mutant_HRupstrm
NS49	GTATAGCATACATTATACGAACGGTACTA *GATCTCATCATCAACCACATTGG*	Fwd, *nusA* _MPN154_I418 truncation mutant_HRdownstrm
NS50	TTTGCGGCCGC *AGGACAAGTTAAAGTTCTTGGTCAA*	Rev, *nusA* _MPN154_I418 truncation mutant_HRdownstrm_*NotI*
NS51	*GTGATTTCCAAAGACCCAGTAGT*	Rev, insertion location chk
NS52	TACCGTTCGTATAGCATACATTATACGAAGTTATCTGCAGTAGTATTTAGAATTAATAAAGT *ATGAACTTTAATAAAATTGATTTAGACAATTGGAA*	Fwd, lox71_*PstI*_P438_*Cat*
NS53	ACTTTATTAATTCTAAATACTAAAGCTT *CCAGCGTGGACCGGC*	Rev, P438_complimentary_*HindIII_Cat*
NS54	AAGCTTTAGTATTTAGAATTAATAAAGT *ATGAAAATTATTAATATTGGAGTTTTAGCTCAT*	Fwd, *HindIII*_P438_*tetM*
NS55	TACCGTTCGTATAATGTATGCTATACGAAGTTATGGATCC *AGTTATTTTATTGAACATATATCGTACTTTATCT*	Rev, Lox66-complimentary_*BamHI_tetM*
NS56	TACCGTTCGTATAATGTATGCTATACGAAGTTATGGATCC *CCAGCGTGGACCGGC*	Rev, Lox66_complimentary_*BamHI_Cat*
NS57	*TAACACCGGTTTTGACCTGG*	Fwd, protected
NS58	*AGGACAAGTTAAAGTTCTTGGTCAA*	Rev, biotinylated
NS59	GTATAGCATACATTATACGAACGGTACTA *GTTTAGTTGCTTCCTGCTTGATTTG*	Fwd, *nusA* _MPN154_K440 truncation mutant_HRdownstrm
NS60	AAAGGTACCAATGAAAAAGTTGTCAGTTAATCAGATCC	JCVI-Syn3A, MMSYN3A_0353 region_fwd_KpnI
NS61	TTTGGATCCCTATCTTGATTTACTTATTAATAATTTATAAATTTCATC	JCVI-Syn3A, MMSYN3A_0353 region_rev_BamHI
NS62	GAACTAGATCAATATGTTAAAAAGAACTGGAAACTAGTAAAAG	MMSYN3A_0353_CCR primer A240G
NS63	AAAGAGCTCATGAAAAAGTTGTCAGTTAATCAGATCC	JCVI-Syn3A, MMSYN3A_0353 region_fwd_SacI
NS64	TTTGGATCCTCTTGATTTACTTATTAATAATTTATAAATTTCATCAAT	JCVI-Syn3A, MMSYN3A_0353 region_rev_no stop_BamHI
NS65	AAAGGTACCAGTGAGTACAATTGATGAATTTGTTGTTC	JCVI-Syn3A, MMSYN3A_0421 region_fwd_KpnI
NS66	TTTGGATCCTTAGATTAAATCTTCAACAATTACATCAACTTTA	JCVI-Syn3A, MMSYN3A_0421 region_rev_BamHI
NS67	AAAGGTACCAATGAAAGAAATTAATTTAGAAAATACAAAAGAAATTATTG	JCVI-Syn3A, MMSYN3A_0503 region_fwd_KpnI

NS68	TTTGGATCCTTAATTAAAAATACTAATGTTTTGATCAAAA TCATTATAAAT	JCVI-Syn3A, MMSYN3A_0503 region_rev_BamHI
NS69	AAAGGTACCAATGAAACTAAATGATAAGTTAAAGAATTTT TTTAATAAT	JCVI-Syn3A, MMSYN3A_0511 region_fwd_KpnI
NS70	TTTGGATCCCTATTAATTTGAAATGATTTTGTCTAAATCT TTGATTTT	JCVI-Syn3A, MMSYN3A_0511 region_rev_BamHI
NS71	AAAGAGCTCGTGAGTACAATTGATGAATTTGTTGTTC	JCVI-Syn3A, MMSYN3A_0421 region_fwd_SacI
NS72	TTTGGATCCGATTAAATCTTCAACAATTACATCAACTTTA TG	JCVI-Syn3A, MMSYN3A_0421 region_rev_no stop_BamHI
NS73	*CCTGCATTTATTTTCTTAGTGACAAG*	Fwd, CmR check
NS74	*CCTGCTGTAATAATGGGTAGAAG*	Rev, CmR check
NS75	AAAGGATCCATGATCACATACAAAGAGAAAAAAGATAAT	JCVI-Syn3A_0877 gene_fwd_Bam HI
NS76	TTTCTGCAGAGTTTATGATCTAATTATTTTTCTAATCAATA CTTTTAC	JCVI-Syn3A_0877 gene_rev_stop codon_PstI
NS77	5'-P-GAAATTAATTATTGGGTTTCAACTGTAGTGTTAACTGC	JCVI-Syn3A_0877 gene_ A381G
NS78	5'P-TTTTGGAGCTTAATTTTTACAGTTGCTTCAGTTTGGATG	JCVI-Syn3A_0877 gene PCR_A420G_A450G
NS79	AAAGGATCCATGAAAAATAAAGGAAAATTACTAGAATTTT TAAC	JCVI-Syn3A_0876 gene_fwd_Bam HI
NS81	AAAGGATCCATGAAAAACAAGTCTAGATTTTTTGAATTTC	JCVI-Syn3A_0878 gene_fwd_Bam HI
NS82	TTTGAATTCTTATCTTTTTTCTCTTTGTATGTGATCATT	JCVI-Syn3A_0878 gene_rev_Eco R1
NS83	TTTCCCGGGCTAAACTGTAGGTTTTGCTATGATTT	JCVI-Syn3A_0876 gene_rev_XmaI
NS128	AAAGGTACCATGATTTTAAAAATGTTAGAAAAAGGAATC ATTTC	fwd_SYN3A_0352_KpnI
NS129	TTTGGATCCTTAAACATTTTGATCATCTAACCAATCAAAT TG	rev_SYN3A_0352_BamHI_A537G
NS130	AAAGGTACCATGTACAAATTCAAAGCACTTTTAGATG	fwd_SYN3A_0451_kpnI
NS131	TTTGGATCCTTAATTATTAATTACTATTCCTTTATATCTAG TTG	rev_SYN3A_0451_BamHI
NS132	GCAGCAAAATCTAGTCAAAAAGCTTGGGAAAGTACTGAT TTAGAAAAAAGAATTTCTATTTTAGATAAATGGAAACAAT TAATTGATC	2xCCR_SYN3A0451_A379G_A424G
NS133	GTGATATGAGATTAGCTTGGGAAGAACCATTTGG	CCR_SYN3A0451_A1330G
NS134	AAAGGTACCATGAAAAAACTTCTATCGTTATTAGCTTG	SYN3A0505_fwd_KpnI
NS135	TTTCTCGAGTTATTTTGCAGGAGTTGGAGCTG	SYN3A0505_rev_XhoI
NS136	GTAATTTAAGCTGGAATAGTACAAATAATCACACAAAAG	SYN3A0505_CCR_A509G
NS137	GTAAAACTAAAAAAGAAAGATGGATAATTGATCAAAAGA CTTC	SYN3A0505_CCR_A593G
NS138	GATTCAGATGATGAAATATGGACTAAAGGTTTAACTTC	SYN3A0505_CCR_A698G
NS139	GAATATCTTGTTAAAAACTGGGCTAAAGTATTTTATGG	SYN3A0505_CCR_A749G
NS140	GAAAAAATACTTATGATCAATGGATAAGCTATTTAAAACA AGC	SYN3A0505_CCR_A1004G
NS141	GATTACTCCTAAAACTGAATGGAAATACCAAGG	SYN3A0505_CCR_A1094G
NS142	GGAACAACTTGGCTAAGATGGCAAACAAC	SYN3A_0439_CCR_A582G_A591G
NS143	GGCAATGAATCTGTAGAATGGTACAAAAATTC	SYN3A_0439_CCR_A801G
NS144	GTTAAAATGGTTTGGGCCTTTTCTCTAGATAA	SYN3A_0439_CCR_A1074G
NS145	GAATTAAACGTAATGCAAAATACATGGGTAAATTTAAAT G	SYN3A_0439_CCR_A1560G
NS146	GTAGCTAATAATAATGATAAATGGTATGTATCATTAAAAA ATGG	SYN3A_0439_CCR_A1677G
NS147	CAAAATGATGCTTATATTTGGAACAATGATCCTA	SYN3A_0439_CCR_A1932G
NS148	AAAGGATCCATGAAGAAACTATTAACAATATTAGGTTCT AT	fwd_SYN3A_0439_BamHI
NS149	TTTCTGCAGTTAGTCTAATGTAGGTTTTTCACTTTCT	rev_SYN3A_0439_PstI
NS150	GAATTCAAACAGGGGATAACGCAGGAAA	fwd_OriC_Syn3A_plox plasmid_ EcoR1
NS151	TTTTCTAGACTGCTGCCAGTGGCGATA	rev_OriC_Syn3A_plox plasmid_ Xba1
NS152	CTGATTATACATCAAAAAGAGCATCTG	SYN3A_0439_CCR check prime_200 bp dwnstr, NS142
NS153	AAAGAGCTCGTGAAAAAGCTTTTAACTTGGCTTAG	fwd_synthesized SYN3A_0440_SacI
NS154	TTTGGTACCTTAGCTATCACCTCTTCTTTCAC	rev_synthesized SYN3A_0440_KpnI
NS156	AAAGGATCCAATGAAAAAATTAGAATTATTAAAAAACATG ATCACAA	fwd_Syn3A_FakA_primer dimer removed

NS157	AAAGGATCCATGAAAAAACTTCTATCGTTATTAGCTTG	fwd_his-Syn3A_0505_BamHI
NS158	TTTAAGCTTTTATTTTGCAGGAGTTGGAGCTG	rev_his-Syn3A_0505_HindIII
NS159	GTTGTTTGCCATCTTAGCC	rev_complimentary to NS142_for LFH and reinsertion of first 600 bp of SYN§A_0439
NS162	AAAGGATCCATGACTAAAAATGATAAATTTGATAAACCATCTATAAC	fwd_SYN3A_0439 gene lipoprotein signal omitted for pWH844_BamHI
NS163	AAAGAGCTCATGACTGAACAAGTAAAAAATGAAAATTCACTTTT	fwd_SYN3A_0440 gene lipoprotein signal omitted for pGP172
NS164	AAAGGATCCATGACAACAGATAAACAATTTCAAGAATTTGAAAA	fwd_SYN3A_0505 gene lipoprotein signal omitted for pWH844_BamHI
NS165	GATGTAAAAAATCATATCACAACTA	fwd_SYN3A_0440 gene_all 18 CCRs_internal sequencing primer
NS166	ATAGAGCTTAAAGAAGAATTTTTGT	rev_SYN3A_0440 gene_all 18 CCRs_internal sequencing primer
NS167	GGCTGAAAGTAACTGGAATGATG	fwd_SYN3A_440_pGP3702_internal sequencing primer
NS168	CTCCATTTTAGCTTCCTTAGCTC	rev_sequencing primer_between Qe30-rev n SH71
NS169	GTTCTTACTCTCAAGATTGTAAGG	SYN3A_0439 free genes_internal rev seq primer

6.4 Plasmids

Plasmids used in this study are listed below.

Name	**Vector**	**Insert**
pGP3376	pBSKII(-)	deletion cassette for I418 NusA truncation mutant, HR_Upstream-CmR-HR_downstream (3 segments)_NS47-50_*XhoI- NotI*
pGP3377	pBSKII(-)	deletion cassette for I418 NusA truncation mutant, HR_Upstream-CmR-TetR-HR_downstream (4 segments)_NS47-50_*XhoI- NotI*
pGP3378	pBSKII(-)	deletion cassette for K440 NusA truncation mutant, HR_Upstream-CmR-HR_downstream (3 segments)_NS47-50_*XhoI- NotI*
pGP2727	pBSK II (-)	PCR product CB85-CB86 (lox71-P438-cat-lox66) --> pBSK II (-) *BamH1 PstI*
pGP2777	pET3c	PCR product *tetM* CB61-62 *EcoRI BamHI* in pET3c
pGP3379	pBSKII(-)	MMSYN3A_0439_TGA still present
pGP3380	pGP172 (n-term STREP tag)	MMSYN3A_0421
pGP3381	pBSKII(-)	MMSYN3A_0421
pGP3382	pGP172 (n-term STREP tag)	MMSYN3A_0503
pGP3383	pGP172 (n-term STREP tag)	MMSYN3A_0511
pGP3384	pGP172 (n-term STREP tag)	MMSYN3A_0353
pGP3400	pBSKII(-)	MMSYN3A_0505_TGA present
pGP3701	pUC57-BsaI-Free	MMSYN3A_0440_TGA free
pGP3702	pGP172 (n-term STREP tag)	MMSYN3A_0440_TGA free
pGP3704	pBSKII(-)	MMSYN3A_0505_TGA free
pGP3705	pWH844	MMSYN3A_0505_TGA free
pGP172	For expression of proteins carrying a Strep-tag at their N-terminal in *E. coli*	(Merzbacher et al., 2004)me
pWH844	For expression of proteins carrying His-tag at their N-terminal in *E. coli*	Schirmer *et al.,* 1997
pBSKII(-)	For cloning purposes	

6.5 Materials

6.5.1 Chemicals

Chemicals were at minimum 95% purity. If not listed here, other chemicals were purchased from Merck, Serva, Sigma or Roth.

Name	Supplier
2-Mercaptoethanol	Carl Roth, Karlsruhe
Acetic acid (glacial)	Th. Geyer, Renningen
Acetonitrile	Carl Roth, Karlsruhe
Acrylamide	Carl Roth, Karlsruhe
Agar-Agar	Carl Roth, Karlsruhe
$AgNO_3$	Carl Roth, Karlsruhe
Albumin Fraction V, pH 7.0 (BSA)	AppliChem, Darmstadt
Ammonium peroxydisulfate (APS)	Carl Roth, Karlsruhe
Ampicillin sodium salt	Carl Roth, Karlsruhe
Bacto Agar	Becton, Dickinson & Co., FR
Brilliant Blue G-250	Sigma-Aldrich Chemie, Steinheim
Bromophenol blue sodium salt	Sigma-Aldrich Chemie, Steinheim
$CaCl_2 \cdot 2\ H2O$	Carl Roth, Karlsruhe
CDP-Star Chemiluminescence substrate (25 mm)	Roche Diagnostics, Mannheim
CHAPS	Roche Diagnostics, Mannheim
Chloramphenicol	Carl Roth, Karlsruhe
d(+)-Glucose monohydrate	Carl Roth, Karlsruhe
d-Desthiobiotin IBA Lifesciences, Göttingen	IBA Lifesciences, Göttingen
DMSO	Carl Roth, Karlsruhe
dNTP mix	Roche Diagnostics, Mannheim
DSS	Thermo Fisher Scientific
DSSO	Thermo Fisher Scientific
EDTA disodium salt hydrate	Carl Roth, Karlsruhe
Erythromycin	Sigma-Aldrich Chemie, Steinheim
Ethanol absolute	VWR International, Rue Carnot, FR
Fetal bovine serum	Invitrogen, Karlsruhe
Formaldehyde solution (37 %)	Carl Roth, Karlsruhe
Glycerol	Carl Roth, Karlsruhe
Glycine	Carl Roth, Karlsruhe
HCl	VWR International, Rue Carnot, FR
HEPES	Carl Roth, Karlsruhe
Horse serum	Invitrogen, Karlsruhe
Imidazole	Carl Roth, Karlsruhe
IPTG	Carl Roth, Karlsruhe
K_2HPO_4	Carl Roth, Karlsruhe
$K_2HPO_4 \cdot 2\ H2O$	Sigma-Aldrich Chemie, Steinheim
Kanamycin sulfate	AppliChem, Darmstadt
KCl	Carl Roth, Karlsruhe

KOH	Carl Roth, Karlsruhe
Methanol	VWR International, Rue Carnot, FR
$MgCl_2 \cdot 6\ H2O$	Carl Roth, Karlsruhe
$MgSO_4 \cdot 7\ H2O$	Carl Roth, Karlsruhe
$MnCl_2 \cdot 4\ H2O$	Carl Roth, Karlsruhe
Na_2CO_3	Carl Roth, Karlsruhe
NaCl	Carl Roth, Karlsruhe
NaOH	Carl Roth, Karlsruhe
Ni-NTA Sepharose (50 % suspension)	IBA Lifesciences, Göttingen
Nutrient broth	Carl Roth, Karlsruhe
Orthophosphoric acid (85%)	Merck, Darmstadt
PageRuler Plus Prestained Protein Ladder	Thermo Fisher Scientific
Paraformaldehyde	Carl Roth, Karlsruhe
peqGOLD Universal Agarose	VWR International, Erlangen
Phenol red	Carl Roth, Karlsruhe
Powdered milk	Carl Roth, Karlsruhe
PPLO broth	Becton, Dickinson and Company, France
RotiphoreseR Gel 30 (37.5:1)	Carl Roth, Karlsruhe
RotiR-Quant (5×)	Carl Roth, Karlsruhe
SDS Pellets	Carl Roth, Karlsruhe
StageCL	Synthesized by Dr. Ana Perez, TU Berlin
Strep-Tactin Sepharose (50 % suspension)	IBA Lifesciences, Göttingen
TEMED	Carl Roth, Karlsruhe
Tetracycline hydrochloride	Sigma-Aldrich Chemie, Steinheim
Tris	Carl Roth, Karlsruhe
Tryptone/Peptone ex casein	Carl Roth, Karlsruhe
TWEEN 20/80	Sigma-Aldrich Chemie, Steinheim
X-Gal	Thermo Fisher Scientific
Yeast extract	Carl Roth, Karlsruhe
λ DNA (0.3 µg/µl)	Thermo Fisher Scientific

6.5.2 Enzymes and Antibodies

Name	**Supplier**
Ampligase thermostable DNA ligase	Epicentre, Madison, USA
Anti-His-tag (ABIN1573880)	Antibodies-online GmbH
Anti-Rabbit IgG (Fc), AP Conjugate [1:100,000]	Promega, Madison, USA
Anti-Strep-tag (StrepMAB-Classic)	IBA Lifesciences, Göttingen
DNase I (from bovine pancreas, grade II)	Roche Diagnostics, Mannheim
DreamTaq DNA Polymerase (5 U/µl)	Thermo Fisher Scientific
FastAP Thermosensitive Alkaline Phosphatase (1 U/µl)	Thermo Fisher Scientific
FastDigest restriction endonucleases	Thermo Fisher Scientific
Phusion High-Fidelity DNA Polymerase (2 U/µl)	Thermo Fisher Scientific
SUMO protease	In-house purified
T4 DNA Ligase (5 U/µl)	Thermo Fisher Scientific

6.5.3 Commercial systems

Name	Supplier
Bradford reagent Carl	Carl Roth, Karlsruhe
NucleoSpin Plasmid Kit	Macherey-Nagel, Düren
PefablocR SC protease inhibitor	Carl Roth, Karlsruhe
peqGOLD Bacterial DNA Kit	PEQLAB Biotechnologie, Erlangen
QIAquick PCR purification Kit	Qiagen, Hilden

6.5.4 Consumables

Name	Supplier
Beakers (50–5000 ml, DURAN)	DURAN Group, Wertheim/Main
Blotting paper sheets (BF 2, 195 g/m2)	Sartorius, Göttingen
Cell scrapers (24 cm, 30 cm)	TPP, Switzerland
Centrifuge bottles Nalgene (500/1000 ml, PPCO)	Thermo Fisher Scientific, Bonn
Concentrator Vivaspin Turbo 15 (5000 MWCO)	Sartorius, Göttingen
Cryo boxes (9×9, PC)	Sarstedt, Nürmbrecht
Culture Tubes (1 ml, borosilicate glass, disposable)	Kimble Chase, Vineland, USA
Cuvettes (semi-micro, PS)	Sarstedt, Nürmbrecht
Dewer carrying flask 26 B	KGW Isotherm, Karlsruhe
Dialysis tubing MEMBRA-CEL (MWCO 3500)	Serva, Heidelberg
Disposal bags	Sarstedt, Nürmbrecht
Dynabeads, Streptavidin	Thermo Scientific, Bonn
Erlenmeyer flasks (100–2000 ml, DURAN, un/-baffled)	DURAN Group, Wertheim/Main
Falcon centrifuge tubes (15/50 ml, PP, sterile)	Sarstedt, Nürmbrecht
Filtration unit Filtropur S/BT50 (0.2 µm)	Sarstedt, Nürmbrecht
Glas pipettes (1, 5, 10, 25 ml graduated)	Brand, Wertheim
Hollow needles	B.Braun Melsungen AG
Immun-Blot PVDF membrane	Bio-Rad Laboratories, München
Inoculation loops (1 µl, sterile)	Sarstedt, Nürmbrecht
Laboratory film Parafilm M	Bemis, Neenah, USA
Laboratory flasks (100–5000 ml, DURAN)	DURAN Group, Wertheim/Main
Magnetic stirring bars (PTFE)	Brand, Wertheim
Measuring cylinders (100–2000 ml, DURAN)	DURAN Group, Wertheim/Main
Membrane filters S-Pak (0.45 µm, sterile)	Millipore, Molsheim, FR
Microcentrifuge tubes (1.5 ml, PP)	Beckman Coulter, Brea, USA
Microliter pipette Research (8-channel, 100 µl)	Eppendorf, Hamburg
Microliter pipettes Research (2.5, 20, 200, 1000, 5000 µl)	Eppendorf, Hamburg
Microtest plate 96 well, F	Sarstedt, Nürmbrecht
Petri dishes with cams (60×15 /150×20 mm, PS)	Sarstedt, Nürmbrecht
Pipette tips (10, 200, 1000, 5000 µl)	Sarstedt, Nürmbrecht
Poly-Prep chromatography columns	Bio-Rad Laboratories, München
Protein LoBind tubes (1.5 ml)	Eppendorf, Hamburg
Reaction tubes (1.5/2 ml, PP)	Sarstedt, Nürmbrecht

Reaction tubes GeneAmp (0.5 ml, PP)	PerkinElmer, Weiterstadt
Scalpel Cutfix (sterile)	B. Braun, Tuttlingen
Single use syringes	Becton, Dickinson & Company, France
Syringes (1–50 ml, sterile)	Terumo Europe, Leuven, BE
Test tubes (DURAN)	DURAN Group, Wertheim/Main
Test tubes with white cap (5 ml)	Malvern Panalytical, Westborough, USA
Tissue flasks (25 cm^2, 75 cm^2)	TPP, Switzerland
Ultracentrifuge tubes (26 ml, PC)	Beckman Coulter, Brea, USA

6.5.5 Instruments

Name	Supplier
ÄKTAprime plus chromatography system	GE Healthcare, Uppsala, Sweden
Analytical balance MSE224S-100-DU Cubis	Sartorius, Göttingen
Autoclave LTA 2x3x4	Zirbus Technology, Bad Grund
Centrifuge Heraeus Biofuge primo R	Thermo Fisher Scientific, Bonn
Centrifuge Heraeus Megafuge 16R	Thermo Fisher Scientific, Bonn
Centrifuge Sorvall RC6SEP	Thermo Scientific, Bonn
Centrifuge Sorvall WX Ultraseries	Thermo Scientific, Bonn
ChemoCam Imager ECL	INTAS Science Imaging Instruments, Göttingen
CO_2 Incubator	Labotect, Göttingen
Degassing Station	ThermoVac MicroCal, Northhampton, USA
French pressure cell FA-003 (20 k Mini)	Thermo Fisher Scientific
French pressure cell FA-032 (40 k Standard)	Thermo Fisher Scientific
French pressure cell press FA078E1	SLM Aminco, Lorch
French pressure cell press HTU-DIGI-Press	G. Heinemann, Schwäbisch Gmünd
Gel electrophoresis device	Waasetec, Göttingen
Hamilton GASTIGHT syringe 1002 (2.5 ml)	Altmann Analytik, München
Heating block	Waasetec, Göttingen
Horizontal reciprocating shaker 3006	GFL, Burgwedel
Ice flaker MF36	Scotsman, Suffolk, UK
Incubator shaker Innova 44R	New Brunswick Scientific, Edison, USA
Liquid nitrogen container	Apollo Messer Cryotherm, Kirchen (Sieg)
Magnetic stirrer REO basic C	IKA-Werke, Staufen im Breisgau
Microbial incubator	Heraeus B12 Thermo Fisher Scientific, Bonn
Microcentrifuge Heraeu Fresco 21	Thermo Fisher Scientific, Bonn
Microplate reader Synergy Mx	BioTek Instruments, Bad Friedrichshall
Microwave Privileg 8020	Quelle, Fürth
Mini centrifuge NG002R	Nippon Genetics Europe, Düren
Mini-PROTEAN Tetra System	Bio-Rad Laboratories, München
Molecular Imager Gel Doc XR+ system	Bio-Rad Laboratories, München
pH-Meter 766	Calimatic Knick, Berlin
Sterile bench HERA safe KS12	Thermo Scientific, Bonn
Thermocycler labcycler	SensoQuest, Göttingen
ThermoStat Plus Vortex	Schütt Labortechnik, Göttingen
UV-Vis spectrophotometer NanoDrop ND-1000	PEQLAB Biotechnologie, Erlangen

UV-Vis spectrophotometer Ultrospec 2100 pro	Amersham Biosciences, Freiburg im Breisgau
Water filtration plant	Millipore, Schwalbach

6.5.6 Software and websites

Name	Provider/Reference	Purpose
ChemoStar Imager 0.2.37	INTAS Science Imaging Instruments, Göttingen	Western blot imaging and documentation
Gen5 2.09.2	BioTek Instruments, Bad Friedrichshall	Microplate reading, analysis and documentation
Geneious 2019-2022	Biomatters, Auckland, NZ	Oligo design and calculation, in silico cloning, analysis of sequencing results
http://www.ncbi.nlm.nih.gov/pubmed	National Institutes of Health, Bethesda, USA	Literature
http://www.rcsb.org	Research Collaboratory for Structural Bioinformatics	Protein structures
Image Lab (Beta 2) 3.0.1	Bio-Rad Laboratories, München	Gel imaging and documentation
Interpro	Elixir Core Data Resource, EMBL-EBI	Resource for functional analysis of proteins
Mendeley Desktop	Elsevier	Literature administration
Microsoft Office 365	Microsoft, Redmond, USA	Data processing
Mycowiki, http://mycowiki.uni-goettingen.de/	University of Göttingen, Bingyao, Pedreira, unpublished work	*M. pneumoniae* database
ND-1000 3.8.1	Thermo Fisher Scientific, Bonn	Nucleic acid concentration/purity
Pfam 35.0	European Bioinformatics Institute	Database of protein families
PrimeView 5.31	GE Healthcare Bio-Sciences, Uppsala, Sweden	ÄKTA chromatography analysis and documentation
Protparam	Expeasy Server, Swiss Institute of Bioinformatics	Estimating protein extinction coefficients
RBP2GO	Diederichs Lab/Projects - DKFZ Heidelberg	Database of RNA-binding proteins
Synwiki, http://synwiki.uni-goettingen.de/	University of Göttingen, doi.org/10.1002/pro.4179	*JCVI Syn3A* database
UniProt, http://www.uniprot.org	Uniprot Consortium EMBL- EBI, SIB, PIR	Information about protein sequences
Xinet	Rappsliber research group	Syn3A Protein-Protein Interaction map visualization