RNA-Binding as a Potential Moonlighting Function of Metabolic Enzymes in *Escherichia coli*

DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von

Thomas Michael Klein

aus Mallersdorf-Pfaffenberg

Dezember 2022

Das Promotionsgesuch wurde eingereicht am: 12.12.2022

Die Arbeit wurde angeleitet von: PD Dr. Patrick Babinger

……..

(Thomas Klein)

Table of Contents

Abstract

Massive research efforts are being expended to comprehend all aspects of the complex interactome within living organisms. Much of the complexity of cellular regulation stems from the incredible broad spectrum of interactions between RNAs and proteins, two macromolecular species intertwined since the origin of life. While many RNA-binding protein domains have been characterized and classified, we might yet know relatively little about certain unconventional modes of RNA binding, like binding mediated by disordered protein regions. One particular field that is poorly researched relative to its potential extent is the hypothesized crosstalk between **R**NAs, metabolic **e**nzymes, and **m**etabolites, a generalized concept known as REM hypothesis. It is substantiated by only few isolated case examples like the dual-functional enzyme aconitase. Technological advancement has given rise to experiments producing extensive interactome data, and metabolic enzymes are frequently being found within RNA interactomes. This provoked request for more directive research investigating the scope of metabolic enzyme-RNA interactions.

Because prokaryotes feature particularly poor insights into the matter, this work aimed to rationally select bacterial enzyme candidates and examine their ability to bind RNAs. In this line, roughly twenty *Escherichia coli* enzymes were chosen from different metabolic pathways based on recent interactome data and literature references of any kind that hinted towards a physiologically relevant enzyme-RNA interaction. Experimental efforts were mainly focused on systematic evolution of ligands by exponential enrichment (SELEX), utilizing a confined sequence pool containing only RNAs that are transcripts of the *Escherichia coli* genome and thus potential biologically relevant targets. In addition, high-throughput sequencing of RNA fragments isolated by crosslinking immunoprecipitation (CLIP-Seq) was conducted for a subset of the enzymes, comprising a complementary method that captures and identifies RNAs in close proximity to the enzyme *in vivo*.

While the majority of enzyme candidates displayed no specific discrimination of RNAs when exposed to the genomic library, seven of them caused enrichment of specific RNA-sequences or -features: Pyruvate kinase enriched a particular set of RNAs, albeit with no apparent common feature. Phosphoglycerate kinase enriched A-rich loop sequences within hairpins. Glyceraldehyde-3-phosphate dehydrogenase enriched highly variable AU-rich sequence stretches. Aconitase enriched variable hairpins with a 3'-adjacent U-rich sequence. Thymidylate synthase enriched hairpins with a strictly conserved AGA-triloop. Glutamate-5-kinase enriched a small number of particular RNAs. Finally, quinone oxidoreductase enriched particular RNAs which share a moderately stringent sequence motif. The enrichments among these seven enzymes were of varying significance and interpretability. Not all of the potential interactions showed substantial specificity or were observable at all in subsequent electrophoretic mobility shift assay (EMSA) analysis, suggesting that genomic SELEX produced a 'best of the worst' selection in such instances. For pyruvate kinase and phosphoglycerate kinase, which constituted such cases, complementary CLIP-Seq data was obtained and revealed no specific RNA binding, challenging

physiological relevance of these enzyme's genomic SELEX results in conjunction with unremarkable EMSAs.

Out of the seven enzymes with notable genomic SELEX results, two were validated to have specific RNA-binding properties: Glutamate-5-kinase and quinone oxidoreductase both displayed unambiguous discriminatory RNA-binding in competition EMSAs probing enriched RNA fragments. A more detailed RNA-binding characterization was pursued for quinone oxidoreductase, which strongly enriched a transcript fragment of *yffO*. The respective RNA-enzyme complex could be disrupted by NADPH, indicating a participation of the Rossmann fold-domain of quinone oxidoreductase in binding. The K_D of quinone oxidoreductase binding to the *yffO* mRNA fragment was estimated to be moderate *in vitro* (7 µM), measured by both EMSA titration and microscale thermophoresis. This moderate K_D indicates that the interaction possibly relies on unidentified *in vivo* conditions if it was to be biologically significant. *YffO* is a grounded prophage gene, and it is unknown but conceivable that it has evolved towards a novel function which is beneficial for *Escherichia coli*.

Overall, specific RNA enrichment was absent or unverifiable for the majority of candidates. Nevertheless, the established RNA-binding specificities of glutamate-5-kinase and quinone oxidoreductase uphold the possibility that RNA-related moonlighting functions can be occassionally found among bacterial metabolic enzymes and certainly prompt further investigation on their biological relevance.

List of Figures

List of Tables

List of Abbreviations

(d)NTP	(desoxyribo)nucleoside triphosphate
∅	diameter
A, T, G, C, U, N, Y	adenosine, thymidine, guanosine, cytidine, uridine, any base, [cytidine or thymidine]
AAK	amino acid kinase
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
$A_{\text{x nm}}$	absorption at x nanometers
bp	base pairs (as unit length)
CLIP	crosslinking and immunoprecipitation
DNA	desoxyribonucleic acid; prefix '**s**': single stranded; prefix '**ds**': double stranded; prefix '**c**': copy
EMSA	electrophoretic mobility shift assay
F	fluorescence
IMAC	immobilized metal affinity chromatography
K_A	association constant
k_{cat}	catalytic turnover number
K_D	dissociation constant
K_M	Michaelis-Menten constant
LB	lysogeny broth
MEME	Multiple Em for Motif Elucidation, (Bailey et al. 2015)
MRW	molecular weight
MST	microscale thermophoresis
NAD(P)H	nicotineamide adenine dinucleotide (phosphate)
NC	complete genomic molecule ("GenBank entry")
NGS	next generation sequencing
NMR	nuclear magnetic resonance
nt	nucleotides (as unit length)
$OD_{\text{x nm}}$	optical density at x nanometers
OOPS	organic phase separation; (Queiroz et al. 2019)
P/C/I	phenol-chloroform-isoamyl alcohol
PAGE	polyacrylamide gelelectrophoresis
PCR	polymerase chain reaction
PTEX	phenol-toluol extraction; (Urdaneta et al. 2019)
PUA	pseudouridine synthase and archaeosine transglycosylase
RBP	RNA-binding protein
RBPome	entirety of RNA-binding proteins within an organism
REM	RNA-enzymes-metabolites
RNA	ribonucleic acid; prefix '**m**' - messenger; prefix '**as**' - antisense; prefix '**nc**' - non-coding; prefix '**t**': transfer; prefix '**r**': ribosomal
rpm	rotations per minute
RT	room temperature

SEC	size exclusion chromatography
SELEX	systematic evolution of ligands by exponential enrichment
SMI	size-matched input
TRAPP	total RNA-associated protein purification; (Shchepachev et al. 2019)
UTR	untranslated region
UV	ultraviolet
w/v	weight per volume

1 Introduction

1.1 The Origin of Life and Protein RNA Interactions

The origin of life is an everlastingly debated mystery of fundamental biology. The central dogma of molecular biology describes the interconnections of nucleic acids and proteins as fundamental prerequisite for living matter (Crick 1970). Research has come up with diverging theories about where to place these biomolecules in the timeline of life genesis, as well as about the development of complex machinery necessary for translation of genetic information into functional proteins. A well-renowned hypothesis is the primal emergence of an RNA world (Gilbert 1986). RNAs, or RNA-like polymers, are able to function as a catalyst for chemical reactions, but also as memories for genetic information (Bruce Alberts et al. 2002). In combination, these prerequisites could have paved the way for the evolution of self-replicating ribozymes, framing a self-persistent system upon which higher complexity could gradually develop. Spontaneous emergence of unaided, self-replicating ribozymes, even though supported by temperately successful attempts to develop self-replicating RNAs in the laboratory (Robertson and Joyce 2014; Tjhung et al. 2020), remains poorly understood (Bernhardt 2012). Considerations like the general hydrolysis tendency of RNA in water and the limited catalytic potential of RNA in isolation keep the RNA world hypothesis in further dispute (Oivanen et al. 1998; Bernhardt 2012).

An opposing scientific opinion is represented by the protein-first hypothesis, reasoning that peptide-based life initiation is evolutionary plausible (Andras and Andras 2005). Random polypeptide synthesis in a prebiotic world could have produced stable and catalytically active domains (Kurland 2010). Peptides and RNAs are also proposed to have evolved concomitantly, with primal peptides complementing the early functions of RNA (van der Gulik and Speijer 2015). The metabolism-first theory also passes over the protein-or-RNA-first dispute, stating that life might not have had its inception in a specific complex molecule type, but rather originated from a wide landscape of relatively simple organic molecules that engaged in constructive interactions and gradually developed key functions like autocatalysis and compositional replication (Hunding et al. 2006).

No matter which - if any - proposed hypothesis precisely captures the unbeknown reality, there is little doubt that the presence of all archetypes of biopolymers was a requirement for the gradual increase in complexity of life. The far backdating entanglement of RNAs and proteins is well illustrated by ribonucleoproteins engaging in translation. The conservation of translation across all domains of life is unique and gives hints about biology predating the last universal common ancestor (Roberts et al. 2008). Particularly, the evolution of the complex ribosomal apparatus, anecdotally referred to as the chicken-or-the-egg issue (Crick 1968), remains a poorly understood process up until today.

1.2 The Prevalence of Protein-RNA Interactions

The origin of life debate has always been extensive, and the above is merely intended to serve as a primer for this ongoing discussion. What can be deducted is a principal involvement of proteins, RNAs, and metabolites in primal biology - regardless of the order of emergence of macromolecules, the evolutionary pacing of an ancient metabolic network relative to the latter, and the time frame in which functional organization started to resemble the complexity of current life. This entanglement is well-reflected in the omnipresence of protein-RNA interactions that can be found in all extant species, and even more so in the higher degree of evolutionary conservation of RNA-binding proteins (RBPs) relative to non-RNA-binding proteins (Anantharaman et al. 2002).

Regulation in cells features an incredibly broad spectrum of protein-RNA interactions. RBPs process and determine the fate of mRNAs by acting on their biogenesis, directing function, and modulating stability (Glisovic et al. 2008). The landscape of RBP-interactions features a continuous spectrum from completely unspecific binding to high specificity based on RNA sequence, spatial properties, or a combination of the latter. The number of published RBP complex structures has been rising increasingly fast for many years (Berman et al. 2000), and many RBPs can be categorized based on domain classification. A single type of domain can be found in differemt proteins with diverging biological functions owing to alternative modes of RNA recognition, as can be seen from the abundant and well-characterized 'RNA recognition motif' (Cléry et al. 2008). Domains can act together to combinatorically increase specificity and expand the length of a recognition site (Maris et al. 2005). A lot of energy has been spent to feed any known protein-RNA interactions into comprehensive databases (Yi et al. 2017). Despite the large number of already recognized RBPs, it is being hypothesized that there might be a lot more RBPs to be discovered, many of which might feature unconventional interaction modes (Beckmann et al. 2016).

1.3 Transcriptome Complexity and Unconventional RNA-Binding

The maybe most surprising discovery by the human genome project, published in 2001, was that only a fraction of the human genome comprises protein-coding genes. By now, it has become apparent that much of the intergenic regions is transcribed, and non-coding RNAs pursue essential functions in cells from eukaryotes and prokaryotes alike - even though the genome of the latter is more compact and predominantly contains protein-coding genes (Rogozin et al. 2002). Discoveries like regulatory functions of long non-coding RNAs or RNA-interference contradict the opportunistic perception of non-coding RNA as 'junk'. While RNAs involved in protein assembly (rRNAs, tRNAs, mRNAs) comprise most of the total RNA mass found in cells (Palazzo and Lee 2015), the multitude of functions found for less abundant transcripts suggests incredible complexity across the large dynamic range of RNA species. This apparent complexity is stressed by unconventional functions found for non-coding RNAs. For example, long non-coding RNAs can act as a scaffold to guide inter-chromosomal interactions in eukaryotes

(Hacisuleyman et al. 2014) or act on architectural proteins to impact organization of chromatin (Creamer and Lawrence 2017). These are just particular case examples supporting the notion that not only proteins can act to guide RNA function, but also the other way around. Many RBPs that lack canonical RNA-binding domains feature intrinsically disordered regions. Intrinsic disorder can occur to different extents, and it becomes increasingly apparent that this attribute significantly contributes to protein-RNA interactions, facilitating specific and unspecific interactions alike (Järvelin et al. 2016), and engages in specialized functions like promoting phase separated ribonucleoprotein granules (Lin et al. 2015). Another considered 'unconventional' interaction is the potential interplay of metabolic enzymes and RNAs, as outlined in the following subsection.

1.4 The REM Hypothesis

Based on observations on metabolic enzymes being able to interact with specific RNAs, an opinion has emerged that secondary ('moonlighting'), RNA-related functions of enzymes might be a frequent occurrence in cells. While well-established examples are restricted to cases where the enzyme might post-transcriptionally regulate RNAs, different modes of interaction have been suggested. Such hypothesized functionalities include RNA-mediated modulation of enzyme activity through RNAs binding to allosteric or active sites of enzymes, as well as RNA-dependent complex formation, like multimer formation of single enzymes or RNA-scaffolding for metabolon formation (Castello et al. 2015).

The general idea of RNAs and proteins engaging in interdependent regulation, possibly in response to a certain metabolic cell state, is termed REM hypothesis (**R**NA-**E**nzymes-**M**etabolites), an acronym originally mentioned by Hentze and Preiss (2010). From an evolutionary standpoint, a regulatory REM network is plausible in the sense that enzymes, especially those involved in primal metabolism, had massive time to coevolve alongside RNAs and develop beneficial regulatory dependencies. From a different perspective, this long-lasting co-evolution could also be interpreted as counterargument that challenges the prevalence of a REM network, suggesting that many of the RNA-enzyme interactions found in interactome screenings or observed by *in vitro* studies merely constitute low-affinity fossils from early stages of life and lost purpose over time, getting replaced by more sophisticated regulation with increasing complexity of organisms (Cieśla 2006). In either case, the multitude of studies reporting RNA-binding enzymes asks for further endeavors investigating the nature and implications of such interactions.

The prime example and probably best-characterized enzyme with RNA-related moonlighting function is the eukaryotic aconitase IRP1, showcased in Figure 1. Cytosolic aconitase participates in the citric acid cycle by converting citrate to isocitrate. However, the enzymatic activity is dependent on an iron-sulfur cluster which gets depleted under iron-starvation conditions. In absence of the metallocofactor, the protein engages in posttranscriptional regulation and interacts with mRNAs which impact iron homeostasis (Hentze and Argos 1991; Constable et al. 1992).

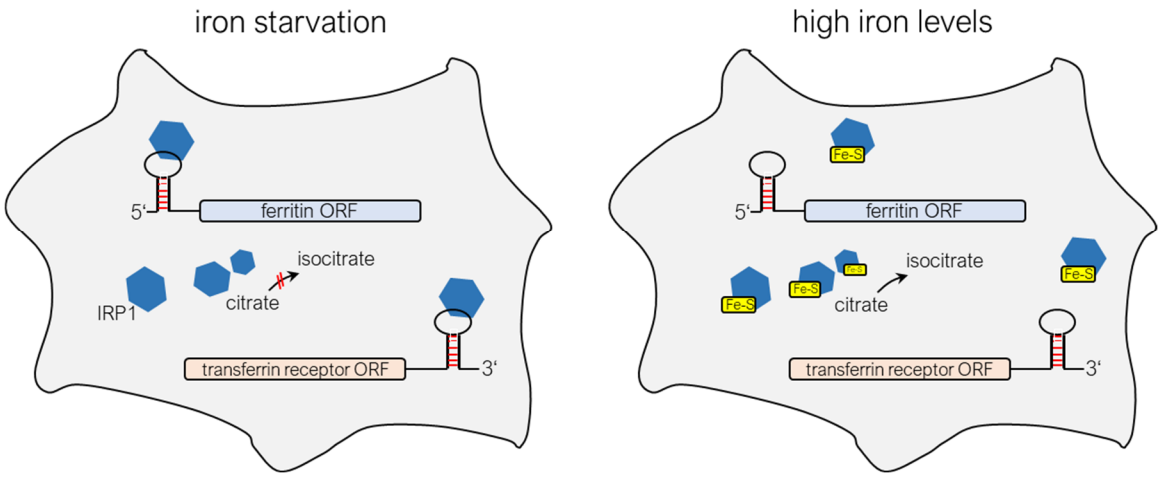


Figure 1: Moonlighting function of IRP1. left - The enzyme (blue hexagon) is unable to convert citrate under iron starvation due to missing iron-sulfur clusters. The lack of the metal cofactor enables IRP1 to exert post-transcriptional regulation of mRNAs involved in iron homeostasis through specific interactions with 5'- or 3'-untranslated regions of ferritin or transferrin receptor transcripts. right - When enough iron is present in the cell, IRP1 primarily engages as metabolic enzyme.

Binding within the 5'-UTR of ferritin mRNA inhibits ribosome recruitment, decreasing the level of iron storage protein. Additionally, binding within the 3'-UTR of transferrin receptor mRNA stabilizes the transcript by blocking endonulceolytic cleavage and thus promotes iron influx into the cell.

Another example for an enzyme frequently hypothesized to maintain an RNA-related moonlighting function is glyeraldehyde-3-phosphate dehydrogenase (GAPDH). The enzyme was initially found to bind tRNA in HeLa cells, discriminating between wildtype and mutant tRNAs (Singh and Green 1993). Subsequently, it was reported that GAPDH shows preferred binding towards AU-rich elements of specific 3'-mRNA regions in competition experiments (Nagy and Rigby 1995). Various studies reported further interactions between GAPDH and both 3'- and 5'-noncoding RNA regions, like those of GLUT1 mRNA (McGowan and Pekala 1996), hepatitis A virus RNA (Schultz et al. 1996), or parainfluenza virus type 3 RNA (De et al. 1996). Although these studies collectively list a diverse set of RNA species bound by GAPDH, they mutually suggested that binding might be conferred by the Rossmann fold-domain, based on the RNA-binding found to be in competition with cofactor binding. While the Rossmann fold-domain has long been suspected to be able to maintain a general, yet poorly characterized RNA-binding function (Hentze 1994), other regions of GAPDH are also discussed as potential interaction sites, like an RRM-like structural motif (White and Garcin 2016), the positively charged substrate-binding grove (Carmona et al. 1999), or the dimerization interface (Kramer et al. 2014). Given the apparent binding of cellular and viral RNAs alike and the RNA-targeted approach of *hitherto* identified interactions, it is conceivable that GAPDH might be able to bind a multitude of RNAs. While this appears intriguing with regards to the REM hypothesis, it also raises questions about the specificity and biological relevance of the spectrum of such interactions.

Besides GAPDH, more enzymes from glycometabolism are suspected to exhibit RNA-binding activity. Further examples include aldolase (Kiri and Goldspink 2002), phosphoglycerate kinase

(Shetty et al. 2004), glucose-6-phosphate dehydrogenase (McGowan and Pekala 1996), and lactate dehydrogenase (Pioli et al. 2002). Case examples of potential moonlighting functions are not limited to glycometabolism and extend to various pathways, as is outlined by a concise review of Cieśla (2006). Globally, reports of RNA-binding metabolic enzymes are more frequent in eukaryotes, though there are instances of simultaneous reports for respective eukaryotic and prokaryotic enzymes homologous to each other.

1.5 Advancements in RBP Identification through Interactome Capturing

The REM hypothesis has been brought up not least due to growing availability of interactome data. Owing to the profound linkage of RNA-protein interactions and gene regulation, method development for unbiased RBP identification has been pursued with great effort since decades, and many protocols are available for both RNA- and protein-centered approaches. For example, RBPs for a specific RNA of interest can be identified by pull-down assays using stable isotope labeling by amino acids in cell culture (SILAC), as described by Butter et al. (2009). As for protein-centered investigations, RNA immunoprecipitation (RIP-Seq) and cross-linking immunoprecipitation (CLIP-Seq) methods are broadly used to capture the as-is state of a protein of interest and its RNA-interactors *in vivo* (Ule et al. 2005). Several CLIP protocols able to determine the exact crosslinking site at nucleotide resolution have been published, like PAR-CLIP (Hafner et al. 2010) or iCLIP (König et al. 2010). The informative value of such experiments has increased considerately by the adoption of mass spectrometry and next-generation sequencing, offering quantitative evaluation of interacting species for RNAs and proteins of interest, respectively. The characterization of RBPs can also be performed *in vitro*, for example by using systematic evolution of ligands by exponential enrichment (SELEX), a method originally developed to probe small molecule-aptamer interactions (Ellington and Szostak 1990). During SELEX, RBP candidates are subjected to a library of target RNAs, and repeated selection and reamplification cycles reduce the contents of the library to its highest affinity sequences. Though the method is carried out *in vitro*, genomic libraries can be used to identify *in vivo* targets (Singer et al. 1997). The informative value of SELEX also benefits greatly from the integration of next generation sequencing.

Aforementioned and many more methods are available that focus on the characterization of specific candidate RBPs or RNAs. A more universal, untargeted approach for RBP identification is governed by the application of high throughput methods. Capture assays can utilize immobilized proteins as bait in screening applications. For example, a study by Scherrer et al. (2010) probed yeast protein microarrays with fluorescently labeled cellular RNA extracts to identify physiologically relevant interactions. Interestingly, the study reported RNA-binding for many proteins with well-established enzymatic activities. Several studies conducted interactome capturing through *in vivo* UV cross-linking, oligo(dT)-bead mediated precipitation of covalently linked complexes, and subsequent mass spectrometry analysis of eluted peptides (Baltz et al. 2012; Castello et al. 2012; Kwon et al. 2013) (see Figure 2, top). This approach is deemed complex due to the lack of PCR amplification implicating relatively low sensitivity (Shchepachev et al.

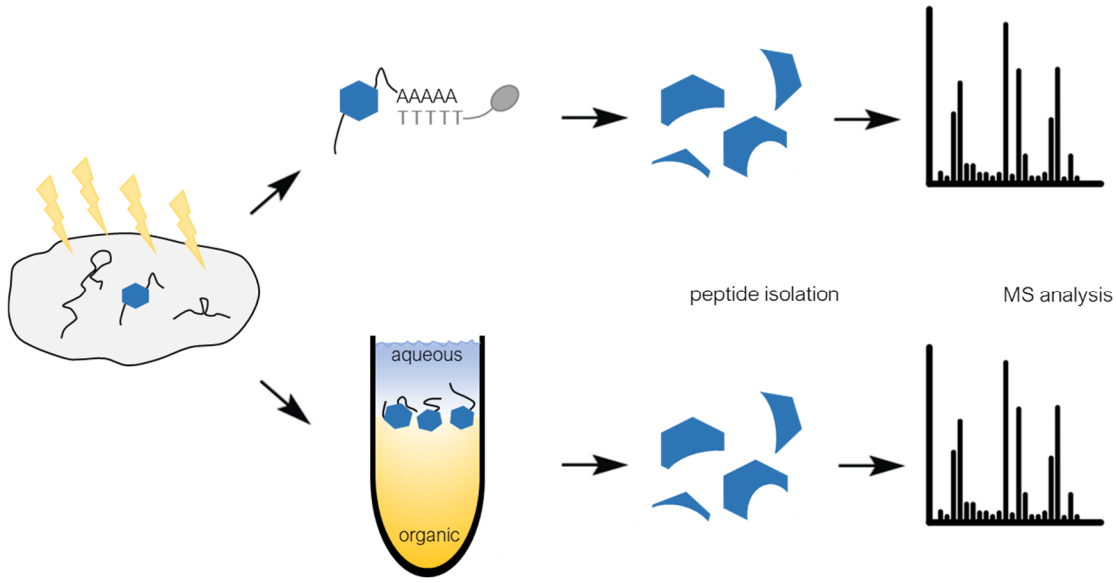


Figure 2: Schematic representation of RNA interactome capturing methods. Figure adopted and expanded based on Castello et al. (2015). The scheme illustrates the principal approach of interactome capturing via oligo(dT)-beads (top) or interphase-RBP-isolation (bottom) in grossly simplified manner. Experimental details of underlying published methods are not meant to be accurately depicted.

2019), and while it enables quick generation of large data sets, it could only ever capture mRNA binding proteins and was limited to eukaryotic systems. In prokaryotic systems, where polyadenylation is less abundant and carries only specialized functions (Slomovic et al. 2006), poly(A)-capturing is unsuited for global mRNA isolation. This limitation has led to slim data availability of prokaryotic interactomes. Computational methods like APRICOT (Sharan et al. 2017) have been developed to aid remedying this gap. Even though sequence-based computational pipelines can be proficient at identifying known RNA-binding domains, their ability to detect completely novel domains or interactions is by implication confined. Hence, the identification of RBPs from prokaryotes has long been limited.

Recently, adaptions of interactome capturing were published that circumvent the application of oligo(dT)-beads, greatly benefiting data set generation for prokaryotic systems. One recent approach, termed RICK or CARIC, extracts protein-RNA complexes after biotinylation of the RNA, using click chemistry nucleotides 4-thiouridine or 5-ethynyluridine (Bao et al. 2018; Huang et al. 2018). Shchepachev et al. (2019) developed a protocol for extraction of protein-RNA complexes through silica beads ('TRAPP') and used it to publish RBPome datasets for both yeast and *Escherichia coli*. Methods labeled OOPS, PTEX, and XRNAX utilize the dissenting chemical properties of proteins and nucleic acids to isolate crosslinked complexes from the interface between an aqueous and organic phase (Queiroz et al. 2019; Urdaneta et al. 2019; Trendel et al. 2019) (see Figure 2, bottom). These aforementioned interactome capturing experiments collectively unraveled an array of potential novel RBPs. For example, OOPS recovered 1838 proteins from three eukaryotic cell lines, and only 912 of them had an RBP GO-term assigned to them, or were previously identified as RBP in *hitherto* published capturing studies (Queiroz et al. 2019). PTEX recovered 3037 proteins from a HEK293 strain, far surpassing the number of well-established RBPs in that cell line (Urdaneta et al. 2019). Both OOPS and

PTEX were able to identify the majority (≤ 80 %) of proteins that were also recovered by RICK and CARIC, suggesting resemblance of different interactome capture methodologies. On top of that, OOPS and PTEX generated an interactome capture dataset for prokaryotic species *Escherichia coli* and *Salmonella typhimurium*, counting 364 and 172 proteins, respectively. For both sets, more than half of the hits were *hitherto* unrelated to RNA (Queiroz et al. 2019; Urdaneta et al. 2019).

The identification of candidate RBPs with no previous RNA-related annotation regularly includes a sizable number of metabolic enzymes, hinting at a potentially extensive REM network. Growing interest in protein-centered characterizations of such potential moonlighting enzymes was the motivation for work pursued within this thesis and framed its scope as summarized below.

1.6 Scope of this Thesis

Advancements in interactome screening have facilitated the rapid generation of RNA-protein interactome data. The regular occurrence of metabolic enzymes within those datasets, alongside isolated, established examples of enzymes pursuing physiologically relevant RNA-binding functions fueled the hypothesis that regulatory interactions between enzymes and RNA might be of disproportionally high prevalence relative to research efforts conducted on the matter. In this context, contributing to the further exploration of such RNA-enzyme interactions was the broadly defined goal of this work.

Provoked by a recent increase in interactome data availability for prokaryotes on the one hand, and the small number of reported bacterial REM interactions on the other hand, this work aimed to examine potential RNA-binding activities specifically in metabolic enzymes of *Escherichia coli*. As protein-centered studies are of particular sparse availability, the focus was set on such experiments, which was intended to complement and interpret published high-throughput interactome data. As this implicated that only a fraction of the whole enzymatic proteome could be covered, the first step was to rationally select candidate enzymes, maximizing the chance of finding novel interactions.

The experimental focus was mainly put on genomic SELEX. All enzyme candidates had to be properly purified to enable unbiased selection processes between the proteins of interest and the genomic RNA library. Treated libraries were to be analyzed by next generation sequencing. If RNAs were found to be significantly enriched, the respective RNA-enzyme interaction was to be verified by *in vitro* follow-up characterization.

As a parallel approach to genomic SELEX, CLIP-Seq experiments were opted for, providing complementary data for the same enzyme from an *in vivo* environment. CLIP-Seq was only pursued for a subset of the candidate pool due to time constraints.

2 Materials

2.1 Chemicals

If not stated otherwise, all chemicals employed in this work were obtained commercially from one of the vendors listed below.

AppliChem GMBH	Darmstadt, Germany
Bio-Rad Laboratories, Inc.	Hercules, USA
Biozym	Hessisch Oldendorf, Germany
Karl Roth GmbH & Co. KG	Karlsruhe, Germany
GE Healthcare	Little Chalfont, UK
Hartmann Analytic GmbH	Braunschweig, Germany
Illumina	San Diego, USA
Merck KGaA	Darmstadt, Germany
MP Biochemicals	Illkirch, France
New England Biolabs GmbH	Frankfurst, Germany
Roboklon GmbH	Berlin, Germany
Roche Diagnostics GmbH	Mannheim, Germany
Serva	Heidelberg, Germany
Sigma-Aldrich	St. Louis, USA
Thermo Fisher Scientific	Waltham, USA
VWR International	Radnor, USA

2.2 Kits

BM Chemiluminescence Western Blotting Substrate (POD)	Merck KGaA
GeneJET™ Gel Extraction Kit	Thermo Fisher Scientific
GeneJET™ Plasmid Miniprep Kit	Thermo Fisher Scientific
MagJET NGS Cleanup and Size Selection Kit	Thermo Fisher Scientific
Micellula DNA Emulsion & Purification Kit	Roboklon GmbH
MiSeq Reagent Kit v2 – 50 cycles	Illumina

MiSeq Reagent Kit v2 nano – 500 cycles	Illumina
NucAway™ Spin Columns	Thermo Fisher Scientific
NucleoSpin® Gel and PCR Clean-up Kit	Macherey-Nagel GmbH & Co. KG
Pure Link Genomic DNA Mini Kit	Thermo Fisher Scientific
SuperSignal™ West Femto Maximum Sensitivity Substrate	Thermo Fisher Scientific

2.3 Purchased Enzymes

DNA-polymerases:	
GoTaq polymerase	Promega
Pfu polymerase	Promega
Phusion polymerase	New England Biolabs
Q5 Polymerase	New England Biolabs
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific
L-Glutamic dehydrogenase from bovine liver	Sigma-Aldrich[1]
Hexokinase from *Saccharomyces cerevisiae*	Sigma-Aldrich
L-Lactic dehydrogenase from rabbit muscle	Sigma-Aldrich
Lysozyme from Egg White	Sigma-Aldrich
Peroxidase from horseradish	Sigma-Aldrich
Polynucleotide phosphorylase	Sigma-Aldrich
Pyruvate kinase from rabbit muscle	Sigma-Aldrich
Restriction enzymes	
BsaI-HF	New England Biolabs
*Dpn*I	New England Biolabs
*Not*I	New England Biolabs
RiboLock RNase-Inhibitor	Thermo Fisher Scientific
Superscript IV reverse transcriptase	Thermo Fisher Scientific
T4 DNA ligase	Thermo Fisher Scientific
T4 polynucleotide kinase	New England Biolabs
T7 polymerase	Thermo Fisher Scientific
Xanthine-oxidase	Sigma-Aldrich

[1] now a trademark of Merck

2.4 Bacterial Strains

Strain	Genotype	Reference
E. coli BL21-Gold (DE3)	B F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte	Agilent Technologies
E. coli BW25113	*lac*I^q *rrn*B$_{T14}$ Δ*lac*Z$_{WJ16}$ *hsd*R514 Δ*ara*BAD$_{AH33}$ Δ*rha*BAD$_{LD78}$	(Datsenko and Wanner 2000)
E. coli DY329	W3110 Δ*lac*U169 *nad*A::*Tn*10 *gal*490 λ*cI857* Δ(*cro-bio*A)	(Yu et al. 2000)
E. coli NEB Turbo	*F' pro*A$^+$B$^+$ *lac*I^q Δ*lac*ZM*15* / *fhu*A*2* Δ(*lac-pro*AB) *gln*V *gal*K*16* *gal*E*15* R(*zgb-210::Tn10*) Tet^S *end*A*1* *thi-1* Δ(*hsd*S-*mcr*B)5	New England Biolabs
E. coli TOP10	F-*mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*Lac*Z ΔM15 Δ*Lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Thermo Fisher Scientific

2.5 Plasmids

Vector	Resistance	Reference
pCA24N	Cat	(Kitagawa et al. 2005)
pCA24N_*pgk*	Cat	“
pCA24N_*pykF*	Cat	“
pCA24N_gapA	Cat	“
pCA24N_*eno*	Cat	“
pCA24N_*gpmA*	Cat	“
pCA24N_*talA*	Cat	“
pCA24N_*talB*	Cat	“
pCA24N_*rpiA*	Cat	“
pCA24N_*kdsA*	Cat	“
pCA24N_*upp*	Cat	“
pCA24N_*adk*	Cat	“
pCA24N_*thyA*	Cat	“
pCA24N_*ansB*	Cat	“
pCA24N_*proB*	Cat	“
pCA24N_*mdh*	Cat	“
pCA24N_*acnB*	Cat	“
pCA24N_*icd*	Cat	“

pCA24N_*sodA*	Cat	“
pCA24N_*ahpc*	Cat	“
pCA24N_*qorA*	Cat	“
pCA24N_*pdxH*	Cat	“
pCA24N_*pdxK*	Cat	“
pCA24N_*yhdh*	Cat	“
pCA24N_*ldhA*	Cat	“
pCA24N_*spoT*	Cat	“
pET21a (+)	Bla	Novagen - Merck
pET21a_*ansB*	Bla	(Funke 2019)
pET21a_*pgk*	Bla	this work
pET21a_*mdh*	Bla	this work
pET21a_*proB*	Bla	this work
pET21a_T7pol	Bla	AG Süß, TU Darmstadt
pQE60_*ybiB*_wt	Bla	(Schneider 2011)
pkD3	Cat	(Datsenko and Wanner 2000)
pkD4	Kan	(Datsenko and Wanner 2000)

2.6 Oligos

All oligos were purchased from biomers.net GmbH or metabion GmbH. All oligos listed in 5‘→3‘ direction.

2.6.1 DNA Oligos

name	sequence	usage
AcnB_middle	gcactgccaatcgaagtcg	introduction of genomic FLAG-Tag (AcnB)
AstC_colony_fwd	tatgcactttaaatgcatatg	introduction of genomic FLAG-Tag (AcnB)
AstC_colony_rev	gccacccgcgaagtcgatatac	introduction of genomic FLAG-Tag (AcnB)
Astc_Ntermflag_P1	Agcgcaaacattacttattattaacatataaataacgaattatttactgtgtgtaggctg gagctgcttc	introduction of genomic FLAG-Tag (AstC)
Astc_Ntermflag_P2	taaacaggtatcatccattcatcaaagttttcacgcgtaattggctgagacttgtcgtc atcgtctttgtagtccttgtcgtcatcgtctttgtagtccatcatatgaatatcctccttag	introduction of genomic FLAG-Tag (AstC)
ASTC_SEQ	gtgtgatcgttactgagaaacag	introduction of genomic FLAG-Tag (AstC)
bglG RNA motif as	tcaggttttgcctgcgaatgcagtaacaatcctatagtgagtcgtattaggatcc	transcription template (*bglG* RNA fragment)
bglG RNA motif s	ggatcctaatacgactcactataggattgttactgcattcgcaggcaaaacctga	transcription template (*bglG* RNA fragment)
CYPSTI	tcgccaagctagcttggattct	sequencing Primer
CYRI	tcacgaggccctttcgtctt	sequencing Primer
D701	caagcagaagacggcatacgagatcgagtaatgtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D702	caagcagaagacggcatacgagattctccggagtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing

D703	caagcagaagacggcatacgagataatgagcggtgactggagttcagacgtgtg ctcttccgatc	barcode Primer for Illumina sequencing
D704	caagcagaagacggcatacgagatggaatctcgtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D705	caagcagaagacggcatacgagatttctgaatgtgactggagttcagacgtgtgct cttccgatc	barcode Primer for Illumina sequencing
D706	caagcagaagacggcatacgagatacgaattcgtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D707	caagcagaagacggcatacgagatagcttcaggtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D708	caagcagaagacggcatacgagatgcgcattagtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D709	caagcagaagacggcatacgagatcatagccggtgactggagttcagacgtgtg ctcttccgatc	barcode Primer for Illumina sequencing
D710	caagcagaagacggcatacgagatttcgcggagtgactggagttcagacgtgtgc tcttccgatc	barcode Primer for Illumina sequencing
D711	caagcagaagacggcatacgagatgcgcgagagtgactggagttcagacgtgtg ctcttccgatc	barcode Primer for Illumina sequencing
D712	caagcagaagacggcatacgagatctatcgctgtgactggagttcagacgtgtgct cttccgatc	barcode Primer for Illumina sequencing
DinD random ctrl as	tattttctgtagcagagatgcattcttctgacttcttgtctgagaaaatgcataagaaaa aataggttaaaactgaggccctatagtgagtcgtattatac	transcription template (control RNA fragment)
Dind random ctrl s	gtataatacgactcactatagggcctcagttttaacctatttttttcttatgcattttctcag acaagaagtcagaagaatgcatctctgctacagaaaata	transcription template (control RNA fragment)
fecR FWD template	gtataatacgactcactatagggaagtgcttctcgccagtgccccgcgcaaaaac gcatcgtgaacgctggtgaaagcctgcagttcagcgcctctgagtttggcgcagtg aaa	transcription template (*fecR* RNA fragment, Pgk SELEX)
fecR REV template	tttcactgcgccaaactcagaggcgctgaactgcaggctttcaccagcgttcacgat gcgtttttgcgcgggggcactggcgagaagcacttccctatagtgagtcgtattatac	transcription template (*fecR* RNA fragment, Pgk SELEX)
GapA-Tev_bsaI_f	aaaaaaggtctcacatgactatcaaagtaggtatc	introduction of TEV cleavage site for GapA
GapA-Tev_bsaI_r	ttttttggtctctcttatttggagatgtgagcgat	introduction of TEV cleavage site for GapA
holB FWD template	gtataatacgactcactataggggaacagtacgccgtgacctggctttcacgcgaa gtgacaatgtcacaggatgcattacttgccgcattgcgcttaagcgcc	transcription template (*fecR* RNA fragment, PykF SELEX)
holB REV template	ggcgcttaagcgcaatgcggcaagtaatgcatcctgtgacattgtcacttcgcgtga aagccaggtcacggcgtactgttcccctatagtgagtcgtattatac	transcription template (*fecR* RNA fragment, PykF SELEX)
i5	acactctttccctacacgac	general reamplification primer (SELEX)
I5_adapter_restore	aatgatacggcgaccaccgagatctacactaagattaacactctttccctacacga c	preparation of SELEX libraries for sequencing
I7_adapter_restore	caagcagaagacggcatacgagatttctgaatgtgactggagttcaagacgtg	preparation of SELEX libraries for sequencing
i7_rev	gtgactggagttcagacgtg	general reamplification primer (SELEX)
Icd_Nterm_colonF	cgcgcatctttcatgacggcaaac	introduction of genomic FLAG-Tag (Icd)
Icd_Nterm_colonR	gcgtcgaccactttcagcatg	introduction of genomic FLAG-Tag (Icd)
Icd_Nterm_Seq	gagactagtagtagaactacc	introduction of genomic FLAG-Tag (Icd)
Icd_NtermFlag_p1	caaacgcatatgcaacgtggtggcagacgagcaaaccagtagcgctcgaagtgt aggctggagctgcttc	introduction of genomic FLAG-Tag (Icd)
Icd_NtermFlag_p2	ttttgcagggtgatcttcttgccttgtgccggaacaactactttactttccttgtcgtcatc gtctttgtagtccttgtcgtcatcgtctttgtagtccatcatcatatgaatatcctccttag	introduction of genomic FLAG-Tag (Icd)
lldd_TEV_BSAI_f	aaaaaaggtctcacatgattatttccgcagccagc	introduction of TEV cleavage site for LldD
lldd_TEV_BSAI_r	ttttttggtctctcttatgccgcattccctttcgc	introduction of TEV cleavage site for LldD
MDHatttactgacga as	aaaaaatcgtcagtaaataaaaaaccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
MDHatttactgacga s	gtataatacgactcactatagggttttttatttactgacgatttttt	transcription template (test RNA for Mdh; iCLIP validation)
MDH_BSAI_FWD	aaaaaaggtctcacatgaaagtcgcagtcctcggc	cloning of Mdh expression vector
MDH_BSAI_REV	ttttttggtctctcgagcttattaacgaactcttcg	cloning of Mdh expression vector
MDH_cagatcga_as	tccgatctgtccgatctgtccgatctgccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
MDH_cagatcga_s	gtataatacgactcactatagggcagatcggacagatcggacagatcgga	transcription template (test RNA for Mdh; iCLIP validation)

Mdh_ccagcgtcatcacttA	aaaaaagtgatgacgctggaaaaccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
Mdh_ccagcgtcatcacttS	gtataatacgactcactatagggttttccagcgtcatcacttttttt	transcription template (test RNA for Mdh; iCLIP validation)
Mdh_efeo_as	ttctccgatctggatggcagcattgacgcccgtgaagatgattacgagcaaaaagc cgccgacccaaaattcactggtttccaccgtctggaaaaagcattgtttggcgacaa caccctatagtgagtcgtattatac	transcription template (*efeO* RNA for Mdh; iCLIP validation)
Mdh_efeo_s	gtataatacgactcactatagggtgttgtcgccaaacaatgctttttccagacggtgg aaaccagtgaattttgggtcggcggctttttgctcgtaatcatcttcacgggcgtcaat gctgccatccagatcggagaa	transcription template (*efeO* RNA for Mdh; iCLIP validation)
MDHgctggcggaag as	aaaaaacttccgccagcaaaaaaccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
MDHgctggcggaag s	gtataatacgactcactatagggttttttgctggcggaagttttttt	transcription template (test RNA for Mdh; iCLIP validation)
Mdh_glnG seq as	cacattccgatctggatgctgccgtcagcgcctatcaacaaggggcgtttgattatct gcccaaaccgtttgatatcgacgaagcagtggcgctggttgagcgcgctatcagtc attccctatagtgagtcgtattatac	transcription template (*glnG* RNA for Mdh; iCLIP validation)
Mdh_glnG seq s	gtataatacgactcactatagggaatgactgatagcgcgctcaaccagcgccactg cttcgtcgatatcaaacggtttgggcagataatcaaacgccccttgttgataggcgct gacggcagcatccagatcggaatgtg	transcription template (*glnG* RNA for Mdh; iCLIP validation)
MDH_combMotif as	aaatccgatctgaaaaaacttccgccagcaaaaaaaaaaaaaaaaaaaaaaaa aagtgatgacgctggaccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
MDH_combMotif s	gtataatacgactcactatagggtccagcgtcatcacttttttttttttttttttttttttttgctg gcggaagtttttcagatcggattt	transcription template (test RNA for Mdh; iCLIP validation)
MDH_TEV_FWD	aaaaaaggtctcacatgaaagtcgcagtcctcggc	introduction of TEV cleavage site for Mdh
MDH_TEV_REV	ttttttggtctctcttacttattaacgaactcttcg	introduction of TEV cleavage site for Mdh
PGK_BSAI_FWD	aaaaaaggtctcacatgtctgtaattaagatgaccg	cloning of Pgk expression vector
PGK_BSAI_REV	ttttttggtctctcgagcttcttagcgcgctcttcg	cloning of Pgk expression vector
PGK_gadE seq as	gttgccgttctgccaacagttcctgccagcattcgggcaacgaagggttctactggt ggatacacataccaggggaataaccctatagtgagtcgtattatac	transcription template (*gadE* RNA for Pgk; iCLIP validation)
PGK_gadE seq s	gtataatacgactcactatagggttattcccctggtatgtgtatccaccagtagaacc cttcgttgcccgaatgctggcaggaactgttggcagaacggcaac	transcription template (*gadE* RNA for Pgk; iCLIP validation)
PGK_stem_term as	aaaaaatgaccccggctagaccggggtgccctatagtgagtcgtattatac	transcription template (test RNA for Pgk; iCLIP validation)
PGK_stem_term s	gtataatacgactcactatagggcaccccggtctagccggggtcatttttt	transcription template (test RNA for Pgk; iCLIP validation)
PGK_TEV_FWD	aaaaaaggtctcacatgtctgtaattaagatgaccg	introduction of TEV cleavage site for Pgk
PGK_TEV_REV	ttttttggtctctcttacttcttagcgcgctcttcg	introduction of TEV cleavage site for Pgk
PGK_ttgaagcattx cttcttc as	aaagaagaagaatgcttcaaaaaccctatagtgagtcgtattatac	transcription template (test RNA for Mdh; iCLIP validation)
PGK_ttgaagcattx cttcttc s	gtataatacgactcactatagggtttttgaagcattcttcttcttt	transcription template (test RNA for Mdh; iCLIP validation)
POLY T control AS	aaaaaaaaaaaaaaaaaaaaaaaccctatagtgagtcgtattatac	transcription template (general control RNA)
POLY T control s	gtataatacgactcactatagggttttttttttttttttttttttt	transcription template (general control RNA)
PROB BSAI FWD	aaaaaaggtctcacatgagtgacagccagacgctgg	cloning of ProB expression vector
PROB BSAI REV	ttttttggtctctcgagacgggtaatcatgtcatcac	cloning of ProB expression vector
PykF rrlX RRNA f	gtataatacgactcactatagggacagcctggccatcattacgccattcgtgcaggt cggaacttacccgacaaggaatttcgctaccttaggaccgttatagttacggccgc cgtttaccggggcttcgatcaa	transcription template (*rrlX* RNA fragment, PykF SELEX)
PykF rrlX RRNA r	ttgatcgaagccccggtaaacggcggccgtaactataacggtcctaaggtagcga aattccttgtcgggtaagttccgacctgcacgaatggcgtaatgatggccaggctgt ccctatagtgagtcgtattatac	transcription template (*rrlX* RNA fragment, PykF SELEX)
ran_control AS	gggtctggcctttgttgtttttgtcctgccctatagtgagtcgtattatac	transcription template (general control RNA)
ran_control sense	gtataatacgactcactatagggcaggacaaaaacaacaaaggccagaccc	transcription template (general control RNA)
rlmD fwd template	gtataatacgactcactatagggttatgatctgacgcgtcgtcgtgcgtcgttttgcag agtagaattgcgccattggcgagactt	transcription template (*rlmD* RNA fragment, ThyA SELEX)
rlmD rev template	aagtctcgccaatggcgcaattctactctgcaaaacgacgcacgacgacgcgtca gatcataaccctatagtgagtcgtattatac	transcription template (*rlmD* RNA fragment, ThyA SELEX)
rnd fwd template	gtataatacgactcactataggggtaaagcgtttaaagcgattaagtcgctgattact gacgtgagcgaaacgcataagatcagcgccgaattgctgg	transcription template (*rnd* RNA fragment, PykF SELEX)

rnd rev template	ccagcaattcggcgctgatcttatgcgtttcgctcacgtcagtaatcagcgacttaatcgctttaaacgctttacccctatagtgagtcgtattatac	transcription template (*rnd* RNA fragment, PykF SELEX)
scpC fwd template	gtataatacgactcactatagggcgattgcccgcagagctaacgaacagcatgaggccaaaaagccgtatcaaattcgccttctgacgggtgcgtcaatcagcgccgccgctg	transcription template (*scpC* RNA fragment, Pgk SELEX)
scpC rev template	cagcggcggcgctgattgacgcacccgtcagaaggcgaatttgatacggctttttggcctcatgctgttcgttagctctgcgggcaatcgccctatagtgagtcgtattatac	transcription template (*scpC* RNA fragment, Pgk SELEX)
SPEC_colony_F	cgtgctttgcgtggtacccg	introduction of genomic FLAG-Tag (SpeC)
SPEC_colony_R	tggttgattttcgctggttac	introduction of genomic FLAG-Tag (SpeC)
speC_ctermflag_P1	agcgaaaccgatgcggatggcgtgaaacggttgtacggttatgtgttgaaggactacaaagacgatgacgacaaggactacaaagacgatgacgacaagtaagtgtaggctggagctgcttc	introduction of genomic FLAG-Tag (SpeC)
speC_ctermflag_P2	gttagccactaattacgcaaagaaaaacgggtcgccagaaggtgacccgttcatatgaatatcctccttag	introduction of genomic FLAG-Tag (SpeC)
SPEC_SEQ	tcggagctggtgaccagtttgac	introduction of genomic FLAG-Tag (SpeC)
spoT_middle	tcgatcatggacatctacgc	cloning of SpoT expression vector
spoT_pet21BSAfwd	aaaaaaggtctcacatgtatctgtttgaaagcctg	cloning of SpoT expression vector
spoT_pet21BSArev	ttttttggtctctcgagatttcggtttcgggtgactttaatcacg	cloning of SpoT expression vector
T7_i5	gtataatacgactcactatagggacactctttccctacacgac	general reamplification primer (SELEX); T7 promoter for transcription
TruSeq Universal	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct	preparation of SELEX libraries for sequencing
ybiB fwd temp	gtataatacgactcactatagggtaacgcgcgcatgcagagaagccagaccagagcgatacagacgttctgcttctttcgggccagtcagcaccatgc	transcription template (*ybiB* RNA fragment, PykF SELEX)
ybiB rev temp	gcatggtgctgactggcccgaaagaagcagaacgtctgtatcgctctggtctggcttctctgcatgcgcgcgttaccctatagtgagtcgtattatac	transcription template (*ybiB* RNA fragment, PykF SELEX)
yjbH fwd temp	gtataatacgactcactatagggatcgttctgcacatactcgcgcacatgggtacgcacgcgcggtaagtgcgagtcctgcggaggattagtgtagttaaacttgt	transcription template (*ybiB* RNA fragment, PykF SELEX)
yjbH rev temp	acaagtttaactacactaatcctccgcaggactcgcacttaccgcgcgtgcgtaccatgtgcgcgagtatgtgcagaacgatccctatagtgagtcgtattatac	transcription template (*ybiB* RNA fragment, PykF SELEX)

2.6.2 RNA Oligos

All listed RNA oligos were employed in EMSA experiments intended to validate interactions between proteins and respective enriched RNAs identified in genomic SELEX or iCLIP. RNA oligos were purchased from biomers.net GmbH in HPLC-purified and lyophilized form. RNA oligos are listed in 5'→3' direction.

ACGCG_motif1 ftsH	cguggcaacaaacgcguugugucgaug
AcnB 3'UTR mRNA motif	aaaagucagcgcacgcgcugcgcauaa
AnsB target RNA	auuauuccuuuuuuauuuuguuc
arcZ stem	caccccggucuagccggggucauuuuu
Arich_control	aggacaaaaacaacaaaggccagaccc
AU-rich control	auuuauuuauuuauuua
copA asRNA fragment	aucuaccggcacgcgaucgccggucgu
fadE1 asRNA fragment	ccgcaucggaacccgcuuccgggcugg
fadE2 asRNA fragment	ggucagugcaaagcaggggaucuccug
fecR stemloop	cccgcgcaaaaacgcaucgu
kdpB stemloop	gccgcgcuaaacagcgcauu
KdsA target RNA	gggggggggggccgaagggcgg
lepB1 mRNA fragment	cgucuuuccgagcguaaagagacacug
LepB2 mRNA fragment	ggugaugugacgcaccgcauucugaca

motif CAGATCGGA	cagaucggacagaucggacagaucgga
narI mRNA fragment	uccgcucagcauauggacgguagcgag
QorA motif	ggagcauauggacuuc
rplA_intergenic	aggaguuucgcaagaaacuuaaucccc
rpsC asRNA fragment	augucagcacgcagagugugcagcggu
scpC MEME bulgeloop	cgucaaucagcgccgccgcugacgaug
scpC stemloop	augaggccaaaaagccguau
stemloop rpoc	agagcgauacagacguucu
stemloop tkta	agcgaugaacucuuucgcu
tktA asRNA fragment	uuuagcgaugaacucuuucgcuuuag
yffO	ggagcauauggacuucagggaagggau
yffO Alt1	agaacacauagauuucagggaagggau
yffO Alt2	ggagcauauggacuccaagggagagac
yffO Alt3	gaaacguguagacuucagggaagggau
yffO Alt4	ggagcauauggacuucaagaaggaggu
yffO Alt5	ggagcauaugaaucccggggaagggau
yffO mRNA fragment	ggagcauauggacuucagggaagggau
yffO_adj1	auaguuugcgcaagaucagcaugaugc
yffO_adj2	gagccaccgucagaugaugcgcuggca
yffO_motif_mutated	caacuggagagcgcuuaauucagggaa
yffO_shifted	caacuggagcauauggacuucagggaa
yffO_trunkL	ggacuucagggaagggau
yffO_trunkR	uggagcauauggacuuca
yhhA mRNA fragment	ccagcgcgcagcagaguugcugcgcug

2.7 Buffers and Solutions

Any culture medium was sterilized by autoclaving for 20 min at 121 °C. Thermosensitive compounds were added to the medium after autoclaving and cooling. Generally, all buffers and solutions were filter-sterilized and prepared with sterile, deionized water from a Milli-Q Advantage A10 Water Purification System. Solutions were stored at room temperature or 4 °C if not stated otherwise. While the following listing does not account for every single solution employed in this work, it provides an overview of routinely used recipes and solutions explicitly mentioned in the methods section.

2.7.1 Solutions for Molecular Biology Works

(d)NTP stock solution	100 mM (d)ATP 100 mM (d)CTP 100 mM (d)GTP 100 mM dTTP/UTP
x % (w/v) agarose	x g agarose was added to 100 ml 0.5x TBE, heated until dissolved, and stored at 60 °C

EtBr stock solution	10 mg/ml EtBr dissolved in H_2O
TBE (5x)	445 mM boric acid, 12.5 mM EDTA, 445 mM Tris-HCl, pH 8.15
P1-saline	145 mM NaCl, 50 mM Na_3citrate, autoclaved

2.7.2 Solutions for Microbiological Work

1000x ampicillin	150 mg/ml Na-ampicillin, filter sterilized
1000x chloramphenicol	30 mg/ml chloramphenicol in 70 % ethanol, filter sterilized
1000x kanamycin	75 mg/ml kanamycin, filter sterilized
1000x IPTG	0.5 M IPTG, sterile filtered, stored at -20 °C
LB medium	10 g/l Tryptone 5 g/l yeast extract 10 g/l NaCl For plates, 15 g/l agar was added
SOB medium	2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 10 mM NaCl *after autoclaving:* +2.5 mM KCl +10 mM $MgCl_2$ +10 mM $MgSO_4$
TFB I	50 mM MnCl 30 mM KAc 10 mM $CaCl_2$ 10 mM KCl 15 % glycerol
TFB II	75 mM $CaCl_2$ 10 mM Tris/HCl pH 7.0 10 mM KCl 15 % glycerol

2.7.3 Buffers and Solutions Related to Protein Biochemistry

KP buffer	1 M KH_2PO_4 / K_2HPO_4, desired pH adjusted by mixing ratio of the two compounds

Tris-HCl buffer	1 M Tris, pH adjusted by addition of 8 M HCl
HEPES buffer	1 M HEPES, pH adjusted by addition of 8 M NaOH
HisTrap equilibration buffer	50 mM of [Tris / KP / HEPES] buffer with desired pH, typically 7.5-8.0 150 mM NaCl 10 mM Imidazole
HisTrap elution buffer	50 mM of [Tris / KP / HEPES] buffer with desired pH, typically 7.5-8.0 150 mM NaCl 750 mM Imidazole
Gel filtration buffer	50 mM of [Tris / KP / HEPES] buffer with desired pH, typically 7.5-8.0 150 mM NaCl
APS for gel casting	10 % APS in water, stored at -20 °C
SDS separation gel buffer	10 % (w/v) SDS, 1.5 M Tris-HCl, pH 8.8
SDS stacking gel buffer	10 % (w/v) SDS, 1.5 M Tris-HCl, pH 6.8
SDS running buffer	0.1 % (w/v) SDS, 0.2 M Glycine, 25 mM Tris-HCl, pH 8.5
SDS sample buffer	5 % (w/v) SDS, 30 % (w/v) glycerin, 15 % (v/v) β-mercaptoethanol, 0.03 % (w/v) bromophenol blue,3.5 M Tris-HCl, pH 6.8
Coomassie staining solution	0.2 % (w/v) Coomassie Brilliant Blue R250, 50 % (v/v) ethanol, 10 % (v/v) acetic acid, stored in the dark
WB resuspension buffer	20 mM Tris/HCl pH 7.4 150 mM NaCl 1 mM EDTA
Blotting Buffer	50 mM Tris/HCl pH 7.4 40 mM Glycine 20 % (v/v) MeOH
1x PBS	137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4, pH 7.4

1x PBS-T	PBS with 0.5 % (v/v) Tween 20
5x TB	445 mM boric acid, 445 mM Tris-HCl pH 8.15
Blocking solution	PBS-T with 5 % (w/v) milk powder, prepared freshly.

2.7.4 Buffers and Solutions for SELEX

SELEX buffer	50 mM Tris/HCl or HEPES, pH 7.5 80 mM NaCl 10 mM KCl 0.8 mM MgCl2 0.5 mM DTT with RNase free Water, filter sterilized
3 M NaAc	pH 6.5 with RNase free Water, filter sterilized
7.5 M NH_4Ac	with RNase free Water, filter sterilized
0.5 M EDTA	pH 8.0 with RNase free Water, filter sterilized
250 mM MgAc	with RNase free Water, filter sterilized
6 % 8 M Urea PAGE	25.8 ml 40 % acrylamide/bisacrylamide solution 34.4 ml 5x TBE 82.6 g UREA *Ad* 173 ml with with RNase free Water polymerization with 0.1 % APS and 0.01 % (v/v) TEMED

2.8 Software

ÄKTA Unicorn	© GE Healthcare
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi
CentroidFold	http://rtools.cbrc.jp/centroidfold/
ChemDraw	© PerkinElmer
Citavi	© Swiss Academic Software GmbH
CLC Main Workbench 8	© Qiagen
Corel Draw 2017	© Corel Corporation
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
IGV	© Broad Institute and the Regents of the University of California

IPknots	https://bio.tools/ipknot
MEME Suite	https://meme-suite.org/meme/
MO.control	© NanoTemper
MS Office	© Microsoft
OptiQuant	© Packard Instrument Co
ProtParam Tool	https://web.expasy.org/protparam/
PureCLIP	https://github.com/skrakau/PureCLIP
Pymol	© Schrodinger LLC
RintD	http://rtools.cbrc.jp/
RintW	https://www.ncrna.org/rintw/
RNAfold	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi
Sigma Plot	© Systat Software Inc.
Spectra Manager	© Jasco Cooperation

3 Methods

3.1 Microbiological Methods

3.1.1 Preparation of Instrumentation and Solutions

Thermostable solutions and media were autoclaved for 20 min at 121 °C and 2 bar prior to use. Glassware and heat-stable expendable items like pipette tips were also autoclaved and subsequently dried at 50 °C. Glassware was sterilized for 4 h at 200 °C. Heat-labile solutions were prepared as stocks and filtered, either via a membrane filter (pore size 0.2 µm) and vacuum pump, or by usage of a syringe filter (pore size 0.2 µm). Solutions intended for use in chromatographic systems or oxygen-sensitive measurements were degassed for at least 30 min in a desiccator.

3.1.2 Transformation of Chemically Competent *E. coli* Cells

Chemically competent *E. coli* cells were prepared by inoculating 500 ml SOB-medium to an OD of 0.1. At an OD of 0.6 cells were cooled on ice for 10 min. Cells were centrifuged (10 min, 3200 g, 4 °C), resuspended in 50 ml TFB I, and centrifuged again. Pellets were then resuspended in 10 ml TFB II, and aliquots were stored at – 80 °C.

For transformation, chemically competent, shock-frozen cell suspensions were thawed on ice. After addition of approximately 100 ng plasmid DNA, cells were incubated on ice for 15 min, subjected to a 42 °C heat shock for 60 sec, and again incubated on ice for 5 min. Subsequently, 900 µl LB-medium was added and cells were incubated for 1 h at 37 °C under shaking (220 rpm) to develop corresponding antibiotic resistance. The suspension was split into varying volumes and plated on LB agar plates containing the appropriate antibiotic to ensure segregated growth of single clone colonies. Cells were grown overnight at 37 °C. For temporary storage, plates and suspensions were sealed with parafilm and stored at 4 °C.

3.1.3 Preparation and Transformation of Electrocompetent *E. coli* Cells

For preparation, 50 mL LB_0-medium was inoculated with approximately 100 µL of an overnight-culture. The cells were grown to an OD of 0.4 - 0.6 at 37 °C (30 °C in case of DY329 cells due to a temperature dependent repressor) and pelleted (8 min, 3200 g, 4 °C). Pellets were washed three times in ice-cold ddH_2O, resuspended in water, and finally stored in aliquots on ice until usage.

For transformation, an aliquot was supplied with at least 1 pg target DNA (at least 100 ng for homologous recombination) and thoroughly mixed. The cell suspension was then poured into pre-cooled electroporation cuvettes and subjected to an electric pulse (2500 V, 25 µF, 200 Ω) in an Electroporator 2510 (Eppendorf). Immediately after pulsing, 1 mL LB_0-medium was added, and cells were cured for 1 h at 37 °C (12 h at 25 °C for DY329 cells). Subsequently, the suspension was split into a dilution series (1:1, 1:10, 1:100, and 1:1000) and plated on antibiotic-containing agar to select for successful transformants.

3.1.4 Disposal of Microorganisms

All bacterial cultures and biologically contaminated equipment were autoclaved for 20 min before exposal.

3.2 Molecular Biology Methods

3.2.1 DNA Amplification by Polymerase Chain Reaction

Polymerase chain reaction (PCR) was used for amplification of specific in vitro DNA fragments (Mullis and Faloona 1987). Reactions were carried out on a Mastercycler personal (Eppendorf). DNA fragments were amplified using Polymerases Phusion, Pfu, or Taq, depending on requirements on elongation error rate and 3'→5' exonuclease proofreading. Standard reaction volumes were in the range of 20-100 µl containing variable amounts of template, 0.2 mM dNTP mix, 1 µM primers, and 1-2 U polymerase in respective reaction buffer. The exact PCR cycling conditions, especially annealing temperature with respect to employed primers and elongation time with respect to elongation length, were optimized for specific reaction. A general cycler program is shown in Table 1.

Table 1: Typical PCR cycling program

1	initial denaturation	95 °C	3 min
2	denaturation	95 °C	20 sec
3	annealing	55 °C (variable)	30 sec
4	elongation	72 °C	variable
	n repetitions of steps 2-4 depending on desired target DNA concentration		
5	final elongation	72 °C	5 min
6	storage	4 °C	∞

For PCR reactions requiring extensive optimization, a Mastercycler gradient (Eppendorf) was used to efficiently screen for best annealing temperatures of individual primer pairs within a temperature gradient.

3.2.2 Colony PCR

Colonies grown on selective agar were isolated with sterile plastic loops and resuspended in a small volume (~10 µL) of LB_0-medium. One microliter of this suspension was employed as template for standard PCR amplification (section 3.2.1) with primers adjacent to the respective target transformation.

3.2.3 Agarose Gel Electrophoresis

Agarose gels were employed for both composition analysis and purification of nucleic acids by size. Agarose gels with concentrations of 0.5-4 % (w/v) agarose were prepared by dissolving respective amounts of agarose in heated 0.5× TBE buffer and cooling until solidification. Visualization of nucleic acid migration by UV-light was enabled by incorporation of 2 µg/ml

ethidium bromide into gels. The gels were employed in 0.5x TBE running buffer. Samples were mixed with commercial loading dye for gel loading. GeneRuler™ 1 kb DNA Ladder Plus and GeneRuler™ Low Range DNA ladder (Thermo Fisher Scientific) served as standard ladders. Electrophoresis typically proceeded at 170 V for 20 min. For isolation of nucleic acids from agarose gels, respective loci were excised above an UV-light table and treated with GeneJet gel extraction kit (Thermo Fisher Scientific) to elute the nucleic acid from the gel matrix.

3.2.4 Isolation of Plasmid DNA from Bacterial Culture

Commercially available kits like the Promega PureYield Plasmid Miniprep system were used for purification of plasmid DNA from *E. coli* cultures. A volume of 1.5 ml of overnight cell culture was harvested by centrifugation (1 min, 16000 g, RT) and treated according to the supplied protocol. Plasmid DNA was eluted with 30 µl of water or provided elution buffer and stored at -20 °C.

3.2.5 Cleavage of DNA by restriction enzymes

Commercially available endonucleases (New England Biolabs) were used to cleave DNA at palindromic recognition sites. A typical restriction reaction contained 10 U of restriction enzyme per microgram of DNA, used the incubation buffer as recommended by the manufacturer, and proceeded for 1 hour at 37 °C. Fragments were routinely purified by separation and excision in agarose gels.

3.2.6 Ligation of DNA fragments

Ligation reactions were conducted by mixing vector and insert in a 1:3 molar ratio in the supplied ligation buffer and adding one unit of T4 DNA ligase (New England Biolabs). The ligation reaction proceeded at room temperature for at least 6 hours and could be used directly for transformation of competent *E. coli* cells.

3.2.7 DNA Sequencing

Standard Sanger sequencing of vectors and PCR products was outsourced to the service of Microsynth Seqlab GmbH. Next-generation sequencing is mentioned separately in section 3.4.12. For plasmid sequencing, up to 700 ng of plasmid DNA and 30 pmol of respective primer were mixed in a total volume of 15 µl. PCR products purified by agarose gel electrophoresis (section 3.2.3) have fairly low concentration. In this case, 12 µL of eluted DNA, supplemented with 30 pmol of respective primer, was used for sequencing without further dilution.

3.2.8 *In Vitro* Transcription and Radioactive Labeling of RNAs

Any RNA that was not obtained commercially by biomers.net GmbH was synthesized using T7 in vitro transcription. T7 RNA polymerase was either purchased from Thermo Fisher Scientific or expressed and purified according to protocols described in section 3.3, using the expression vector pET21a_T7pol (a kind donation of the Süß group of the synthetic RNA biology department in TU Darmstadt). Templates for transcription contained the promoter sequence of T7 RNA polymerase and were either prepared by annealing two short, complementary DNA oligos, or by linearization of a suitable vector. Typical reaction conditions are stated in Table 2,

transcription proceeded at least 6 hours at 37 °C. Reactions were scaled up by increasing the amount of T7 RNA polymerase if desired.

Table 2: *In vitro* transcription reaction

template	1 µg
RNase-free water	67 µl
1 M $MgOAc_2$	2 µl
2 M Tris/HCl pH 8	10 µl
1 M DTT	2 µl
200 mM Spermidin	1 µl
100 mM NTP mix	16 µl
1 µl of 6000 Ci/mmol, 10 mCi/ml α-^{32}P-ATP (added if radioactive labeling of RNA was desired)	
200 U/µl T7 RNA polymerase	2 µl

Next to *in vitro* transcription, RNAs were also labeled with a terminal ^{32}P-Phosphate. End-labeling reactions were conducted with commercial T4 Polynucleotide kinase (New England Biolabs) according to Table 3. Incubation typically proceeded 2 hours at 37 °C.

Table 3: RNA end-labeling reaction

100 µM RNA	5 µl
100 µM ATP	2 µl
6000 Ci/mmol, 10 mCi/ml γ-^{32}P-ATP	3 µl
10x PNK buffer	2.5 µl
RNase-free water	11.5 µl
10U/µl T4 Polynucleotide kinase	1 µl

3.2.9 Purification of RNA

RNAs were purified either by phenol-chloroform-isoamyl alcohol (PCI) extraction and precipitation, and/or by commercial purification kits, including the Monarch® RNA cleanup Kit (50 µg) by New England Biolabs and NucAway™ spin columns by Invitrogen. For PCI extraction, RNA samples were supplemented with an equal volume of ROTI®Aqua-P/C/I pH 4.5 (ROTH). The low pH causes enrichment of not only proteins but also undesired DNAs in the interphase. After centrifugation (3 min, 3200 g) and isolation of the lower phase, ½ volume of 7.5 M ammonium acetate was added, followed by one volume of isopropanol. Precipitation occurred by incubating the mixture for 15 min at -20 °C, and RNA was pelleted by centrifugation (15 min, 16,000 g, 4 °C). For low amounts of RNA, 0.5 µl of GlycoBlue™ coprecipitant (Thermo Fisher Scientific) was added to the precipitation mix. RNAs were resolubilized in desired amount of RNase-free water.

3.2.10 Determination of Nucleic Acid Concentration

The light absorbance at 260 nm was used to determine DNA and RNA concentrations via spectroscopy. The linear dependency of absorbance A and absorbing molecules is defined by the Lambert-Beer Law and allows the calculation of plasmid DNA as follows, where ε is the

wavelength-dependent molar absorbance coefficient of DNA or RNA, and d the pathlength of the cuvette:

$$c_{\text{DNA}} = \frac{A_{260\text{ nm}}}{\varepsilon_{260\text{ nm}} \times d} \quad \left[\frac{\text{mg}}{\text{ml}}\right]$$

The purity of nucleic acid solutions was assessed by calculating the absorbance quotients $A_{260\text{ nm}}/A_{230\text{ nm}}$ and $A_{260\text{ nm}}/A_{280\text{ nm}}$, values which are indicative for impurities like proteins, solvents, and salts (Wilfinger et al. 1997).

3.2.11 Concentration Determination of RNA ^{32}P-Labeled by *In Vitro* Transcription

For α-^{32}P-ATP-labeled RNA, the yield of *in vitro* transcription reactions was determined by scintillation counting. Therefore, 2 µL of RNA sample were dissolved in 2 ml of ROTI scintillation cocktail, and the sample activity was measured on a scintillation counter device. The concentration c was calculated according to the below formula, taking calibration date of [α–^{32}P]ATP, adenosine frequency in the RNA strand, labeled-to-unlabeled ATP ratio, and scintillation counter efficiency (measured every 3 months, usually corresponding to a factor 0.7-0.8) into account:

$$c = \frac{\text{corrected scintillation counts}}{\text{label efficiency} \times n(\text{adenosines per strand})} \quad [\text{pmol/µL}]$$

$$\text{corrected scintillation counts} = \frac{\text{scintillation counts}}{2220 \times \text{counting efficiency}} \quad [\text{nCi/µL}]$$

$$\text{label efficiency} = LE_{t=0} \times \exp\left(-\frac{\text{days from calibration} \times \ln(2)}{14.3}\right) \quad [\text{µCi/nmol}]$$

$$LE_{t=0} = \frac{1000 \times c([\alpha\text{-}^{32}\text{P}]\text{-ATP}) \times V([\alpha\text{-}^{32}\text{P}]\text{-ATP})}{\text{total ATP amount}} \quad [\text{µCi/nmol}]$$

scintillation counts: counts per minute of employed sample volume [CPM/µL]

***c*([α-^{32}P]-ATP)**: molar concentration of labeled ATP molecules in stock solution [µCi/µL]

***V*([α-^{32}P]-ATP)**: amount of labeled ATP employed in initial transcription reaction [µL]

total ATP amount: sum of labeled and unlabeled ATP molecules in transcription reaction [nmol]

3.2.12 Modification of the *E. coli* Genome by Homologous Recombination

For iCLIP experiments described in section 3.4.10, it was necessary to modify the *E. coli* genome at selected positions by introducing double FLAG-Tags. To this end, the λ-red recombination system comprised of genes γ, β, and *exo* was utilized to perform homologous recombination (Datsenko and Wanner 2000; Yu et al. 2000). The genes are encoded in the genome of *E. coli* DY329 under the control of a temperature-dependent repressor. Gamma functions as an inhibitor of exonuclease V and prevents degradation of the linear, double-stranded DNA strand destined for modification. Beta and Exo engage in interactions with the dsDNA which initiate the recombination process. Exo acts as a dsDNA-specific exonuclease, creating 3' ssDNA overhangs.

Beta then binds to these overhangs and facilitates the annealing process between complementary DNA strands.

Figure 3 gives an overview over the construction of a suitable cassette and the subsequent integration into the *E. coli* genome. Next to FLAG-Tag and stop-codon, the cassette also had to carry a kanamycin resistance gene to facilitate isolation of successfully modified cells by colony growth. Homology regions H1 and H2 are genomic sequences adjacent to the desired tag location, at least 50 bp in size to prevent unwanted recombination events.

The cassette was created by PCR amplification of a kanamycin resistance gene located on a pKD4 vector with suitable primers carrying homology regions, FLAG-Tag, and stop-codon. Subsequent *Dpn*I digestion for 1 hour at 37 °C ensured removal of trace vector DNA. The success of the PCR reaction was verified by agarose gel electrophoresis (section 3.2.3), and the desired cassette was gel-purified. The purified cassette was inserted into electrocompetent *E. coli* DY329 cells carrying the λ-red recombination genes by transformation (section 3.1.3). Successfully modified cells were identified by colony PCR (section 3.2.2) with gene-specific primers.

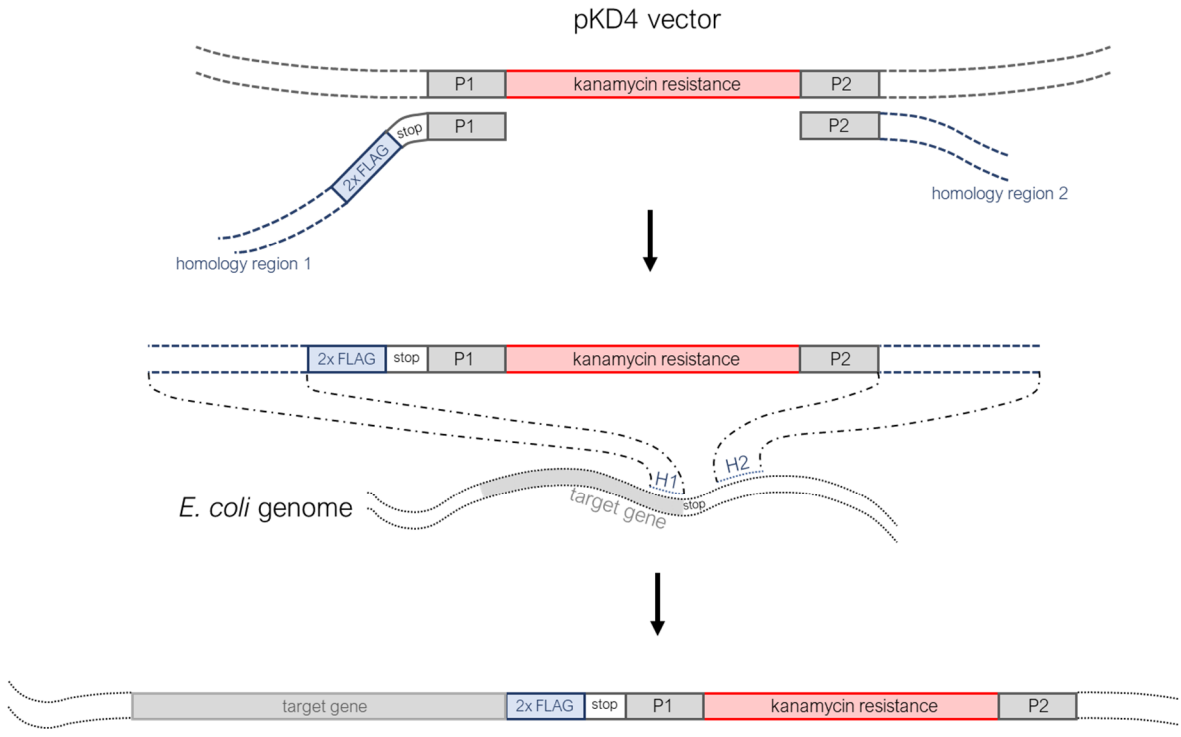


Figure 3 : Schematic representation of homologous recombination. The first step shows the construction of a cassette holding FLAG-Tag and kanamycin resistance by PCR amplification with suitable primers (priming sites are labeled P1 and P2). The resulting construct can then be integrated into the *E. coli* genome by λ-red mediated recombination via homology regions H1 and H2. The scheme shows the addition of a FLAG-Tag on the C-terminal end of the target gene. The method was analogously applied for addition of N-terminal FLAG-Tags by adapting the primers.

3.2.13 Gene Transfer by P1 Transduction

Section 3.2.12 described the modification of genomic DNA in *E. coli* DY329 cells, which carry a temperature-inducible promoter for recombination genes. This promoter renders the strain unsuited for follow-up experiments, since recombination is triggered at 30 °C and above. The genomic modification was therefore transferred into a *E. coli* BW25113 strain by P1 phage induced transduction. During the lytic cycle, the packaging process of new phages has a small chance to incorporate genomic DNA of the host instead of phage DNA, and this DNA can be incorporated into the genome of newly infected cells. Since phage particles can be transferred in air and cause unwanted contaminations in the lab, all experimental steps were conducted under a fume hood.

Cells from a modified *E. coli* DY329 were grown overnight and used to inoculate 50 mL LB-medium supplemented with 5 mM $CaCl_2$ and 0.2 % glucose. Cells were grown for 1 h at 37 °C. The suspension was split into 4 flasks, varying amounts of P1 lysate (0, 20, 100, 500 µL) were added, and the aliquots were incubated under shaking (220 rpm, RT) until some of the cultures became clear, indicating lysis of *E. coli* DY329 cells. The aliquot that both became clear and was supplemented with the lowest amount of P1 lysate was further used. To ensure quantitative destruction of cells, 100 µL of chloroform were added, and the suspension was shaken for 10 min at 37 °C. Subsequently, the cell debris was pelleted by centrifugation (10 min, 3200 g, 4 °C), and the isolated supernatant was stored at 4 °C until further use.

For P1 transduction, an overnight culture of the target strain *E. coli* BW25113 was grown and supplemented with 2.5 mM $CaCl_2$, which is required for efficient absorption of phage particles. Infection was initiated by adding either 400 µL of undiluted or 1:10 diluted P1 donor lysate to 1 mL of *E. coli* suspension. A third aliquot was supplemented with 400 µL LB_0-medium as negative control. The aliquots were incubated for 30 min at 37 °C, and the infection process was stopped by addition of 5 mL P1 saline, which contains citrate able to complex calcium ions. The cultures were incubated for 1 hour to develop antibiotic resistance (220 rpm, 37 °C), and cells were pelleted by centrifugation (10 min, 3200 g, 4 °C). The pellets were resuspended in 100 µL P1 saline and plated on selective agar, screening for cells carrying the kanamycin resistance cassette in their genome. After 1 day, colonies were picked, and the correct genome modification was verified by colony PCR (section 3.2.2) using gene- and cassette-specific primers. The PCR products were sequenced as described in section 3.2.7.

3.2.14 Libraries for SELEX

Selected RBP candidates were subjected to genomic SELEX (**S**ystematic **E**volution of **L**igands by **EX**ponential enrichment) in an attempt to identify high affinity target RNAs in *E. coli*. SELEX requires a library from which the target protein selects nucleic acids based on highest affinity interaction. The RNA library was transcribed from a DNA library whose construction was outsourced to Microsynth AG. In brief, genomic DNA was isolated from a OneShot™ TOP10 *E. coli* strain, and mechanically sheared into shorter DNA fragments. Subsequently, T4 DNA Polymerase was used for end-repair (creating blunt-ended fragments), 5' strands were phosphorylated by T4 polynucleotide kinase, and dA-tails were added using Taq Polymerase (Figure 4). TruSeq® adapters were ligated, and PCR-amplification (section 3.2.1) of the resulting construct with truncated version of primers 'TruSeq-universal' and 'D7*xx*' (called 'i5' and 'i7_rev', see oligo list section 2.6.1) resulted in the final library construct which can be subjected to experiments. For transcription into RNA, a primer with added T7 promoter was used ('T7_i5'). After employment of the library in experiments, different barcodes can be introduced by PCR with full-length D7*xx* primers to enable the parallel sequencing of multiple sample libraries.

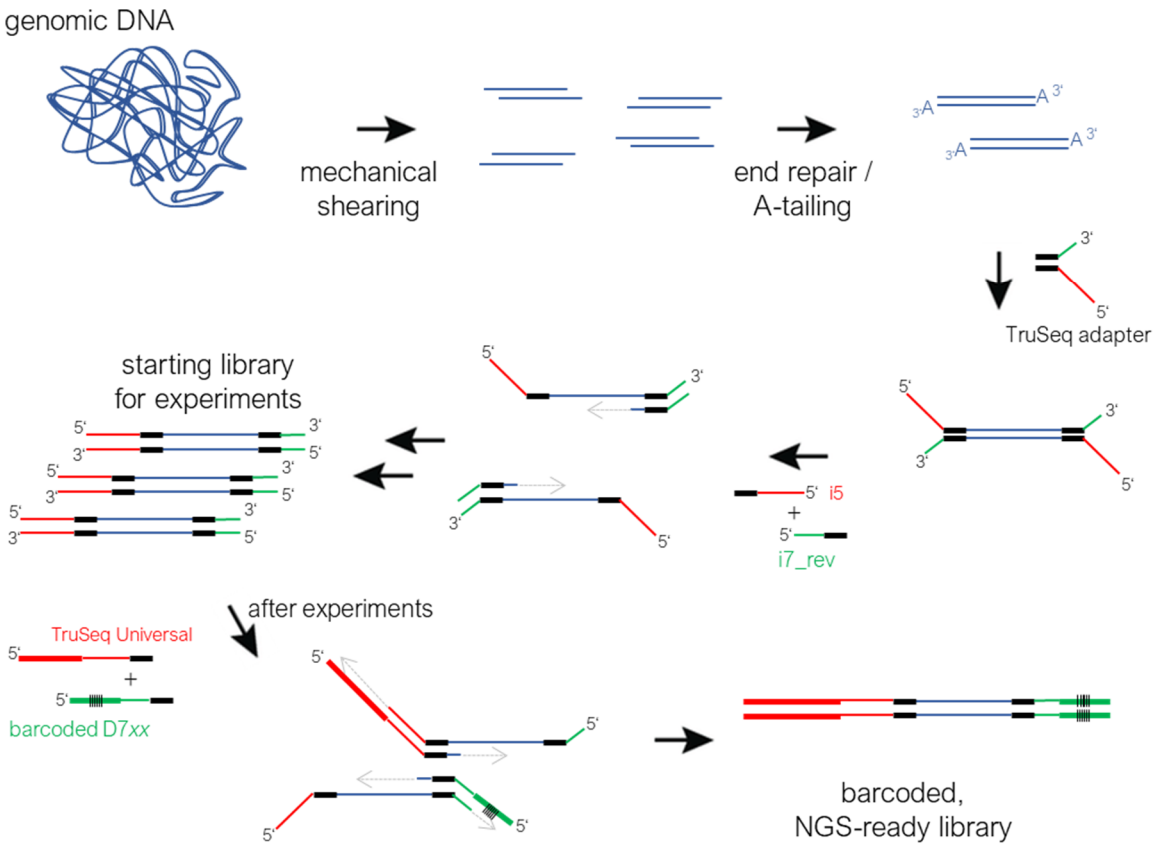


Figure 4: Schematic representation of genomic library creation and sequencing preparation.

The additional SELEX library used specifically for the lysozyme control experiment is comprised of two different pools that both have primer-flanked regions of fixed length. One pool features completely randomized 64 nucleotide length sequences, and the second a fixed, eight nucleotide stem loop structure flanked by randomized regions 26 nucleotides in length (Figure 5). The library was a kind donation from the department of synthetic RNA biology at TU Darmstadt, where the library was previously used for SELEX experiments (Lotz et al. 2019).

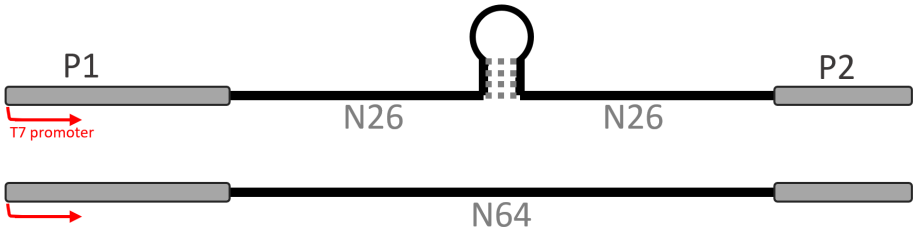


Figure 5: Random aptamer library used for the lysozyme control experiment. Figure adapted based on Lotz et al. (2019).

3.3 Protein Biochemistry

3.3.1 Analytical Scale Expression of Proteins

Expression of genes that needed optimization or initial expressions of genes were usually tested in small-scale expression batches, to allow for economic screening of multiple expression conditions. To this end, 5 ml of pre-culture was inoculated with a respective transformant and grown at 37 °C under shaking overnight. The pre-culture was used to inoculate 50 ml of selective LB-medium to an OD_{600} of 0.1. After the cultures were grown to an OD_{600} of 0.6 to 0.9 at 37 °C, expression was induced by induction with 0.5 mM IPTG. Expression proceeded overnight under desired testing conditions. Cultures were then harvested by centrifugation (15 min, 16000 g, 4 °C) and pellets were resuspended in 500 µl resuspension buffer. The cells were sonified (20 %, 1 min, alternating 2 s pulse and 2 s pause, Branson Digital Sonifier), and cell debris was removed by centrifugation (30 min, 16000 g, 4 °C). An aliquot of cell debris as well as the supernatant containing soluble protein was subjected to SDS-PAGE to examine the success of expression (section 3.4.3).

3.3.2 Preparative Scale expression of Proteins

For preparative scale expression, freshly prepared overnight 50 ml cultures of transformants were used to inoculate 2 to 4 l of selective LB-medium to an OD_{600} of 0.1. The new cultures were grown to an OD_{600} of 0.6 to 0.9 by shaking at 37 °C. Gene expression was then induced adding 0.5 mM IPTG, and expression proceeded overnight under desired temperature condition. Cells were harvested by centrifugation (20 min, 4000 rpm, 4 °C, Beckman Avanti J-26S XP, JLA-8.1 rotor) and pellets were resuspended in 20 ml resuspension buffer per 1 l of harvested cell culture. Cell disruption was achieved by sonication on ice (60 %, 3 min, alternating 2 s pulse and 2 s pause, Branson Digital Sonifier). Cell debris was removed by centrifugation (30 min, 16000 g, 4 °C) and recombinant proteins were purified from the supernatant as described below.

3.3.3 Immobilized Metal Ion Affinity Chromatography

The polyhistidine tag introduced to proteins of interest renders the proteins highly affine to Ni^{2+} ion-mediated complexation. The immobilized metal ion affinity chromatography is taking advantage of this property by capturing such His-tagged proteins on a Ni^{2+}-endowed sepharose resin. In this work, an Äkta micro device (GE HEALTHCARE) and HisTrap crude FF 5 ml columns (GE HEALTHCARE) were used for preparative purification of proteins of interest. Imidazole at high enough concentrations is able to displace the protein from Ni^{2+} binding sites, thus making it possible to elute proteins with an imidazole rich buffer.

The Äkta micro system was operated at a flowrate of 5 ml/min. Before application of a protein onto the column, pump system and column were equilibrated using HisTrap equilibration buffer. Subsequently, the disrupted crude extract (section 3.3.2) was filter sterilized (pore size 0.45 µm) and loaded onto the column. The column was then washed with 10 column volumes (CV) equilibration buffer, and the proteins were eluted by a gradient of elution buffer (0-100 % over 15 CV). Elution was monitored by absorbance levels at 260 nm and 280 nm. Identity and purity grade of collected elution fractions were examined using SDS-PAGE (section 3.4.3) and clean fractions containing the protein of interest were pooled and further purified by preparative gel filtration chromatography as described below.

3.3.4 Preparative Size Exclusion Chromatography

Preparative size exclusion chromatography (SEC), also commonly referred to as gel filtration, is a chromatographic technique utilizing the fractionation through a reverse molecular sieve. Larger molecules, being less prone to enter the porous matrix of the column material, will elute earlier, whereas smaller molecules are retained and will elute later on. Elution times are roughly proportional to the decadic logarithm of their molecular weight. An Äkta prime system (GE HEALTHCARE) and HiLoad 10/300GL Superdex 75 or 200 columns (GE HEALTHCARE) were used, depending on the molecular weight of the protein of interest. These columns contain cross-linked, porous agarose beads to achieve the molecular sieve effect. Runs were performed at a temperature of 4 °C and at a flowrate of 1.5 ml/min. After the column was equilibrated with at least 1.5 CV of gel filtration buffer, a maximum of 10 ml protein solution was applied. Following flow-through of about 1 CV of gel filtration buffer, complete elution of protein species collected in 4 ml fractions was assumed. Elution was monitored by continuous OD_{280} measurement. Fractions containing relevant amounts of protein were further analyzed by SDS-PAGE (3.4.3).

3.3.5 Dialysis of Protein Solutions

If proteins were desired to be stored in a specific buffer after purification or salts had to be removed, dialysis was applied. Therefore, proteins were transferred into pre-rinsed dialyzing tubes with a molecular weight cut-off of 10 kDa (Visking) and incubated in a large amount of target buffer (typically 1-3 l) at 4 °C for at least 12 h. For quantitative success of dialysis, the target buffer was exchanged one time during dialysis.

3.3.6 Concentrating Protein Solutions

The concentration of protein solutions was increased by employing Amicon centrifugal filter devices with a cut-off of 10 kDa (Sigma-Aldrich). Filter devices with applied protein solutions were centrifuged at 4,000 rpm and 4 °C until solution volume was decreased to desired extend.

3.4 Analytical Methods

3.4.1 Determination of Protein Concentration by UV-Absorption Spectroscopy

Determination of protein concentration in solution can be accomplished utilizing the characteristic absorbance bands of the amino acids tryptophan (Trp) and tyrosine (Tyr) as well as those of disulfide bonds between two cysteines (cystine) in the wavelength interval of 250 nm to 300 nm. The molar absorption coefficient of a protein at 280 nm can be derived from amino acid composition using the following equation, where n_i is the quantity in which the corresponding species is occurring in the amino acid sequence (Pace et al. 1995):

$$\varepsilon_{280\,\mathrm{nm}} = 5500 \times n_\mathrm{i}^{\mathrm{Trp}} \times 1490 \times n_\mathrm{i}^{\mathrm{Tyr}} \times 125 \times n_\mathrm{i}^{\mathrm{Cystine}} \quad \left[\mathrm{M^{-1}cm^{-1}}\right]$$

With the molar absorption coefficient, the protein concentration c can be calculated from the absorbance measured at 280 nm ($A_{280\,\mathrm{nm}}$) in a cuvette with pathlength d as follows (Swinehart 1962):

$$A_{280\,\mathrm{nm}} = \varepsilon_{280\,\mathrm{nm}} \times c \times \mathrm{d}$$

However, application of this equation leads to inaccurate protein concentration values, caused by light scattering contributing to absorbance, predominantly originating from aggregated protein. The following equation includes a correction term for improved accuracy:

$$c = \frac{A_{280\,\mathrm{nm}} - 2 \times A_{333\,\mathrm{nm}}}{\varepsilon_{280\,\mathrm{nm}} \times \mathrm{d}}$$

Absorption was measured on a Jasco V-650 spectrometer (JASCO GmbH) or on a NanoDrop One (Thermo Fisher Scientific).

3.4.2 Determination of Protein Concentration by Bradford Assay

The Bradford assay is a better suited method for protein concentration determination when dealing with proteins that display low absorbance at 280 nm. The method relies on a shift in absorbance maximum upon binding of Coomassie Brilliant Blue G-250 to certain cationic, nonpolar, and hydrophobic residues (Bradford 1976). The Bradford assay was routinely applied by mixing 200 µl of Bradford Reagent (Bio-Rad) with 800 µl of diluted protein solution. After incubation for 5 min, absorbance was measured and compared to a BSA calibration curve in order to obtain the concentration value. For blanking, 800 µl H_20 was employed instead of protein solution.

3.4.3 SDS-Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulfate (SDS) in high concentration is able to denature proteins. The negatively charged SDS binds to proteins with a ratio of roughly 1 molecule per 1.4 amino acid residues,

thus rendering the individual charge of the protein neglectable and unifying its mass to charge ratio. As a consequence, electrophoretic mobility depends only on the sieve effect of the gel, and the migration speed is inversely proportional to the logarithm of mass. Gels were routinely prepared according to Table 4:

Table 4: SDS-Polyacrylamide Gel Recipe

	resolving gel [ml]	stacking gel [ml]
H_20	31.6	16.0
resolving gel buffer	19.5	-
stacking gel buffer	-	7.4
40 % acrylamide/bisacrylamide 37.5:1	26.2	5.9
bromophenol blue	-	0.09
TEMED	0.09	0.03
10 % APS	0.2	0.2

Recipe for 13 gels poured in a multi-gel casting system (Hoefer)

Protein samples were mixed with commercial SDS sample buffer and incubated for 5 min at 95 °C. Electrophoresis was usually carried out at 300 V and 45 mA for approximately 45 min. Subsequently, gels were stained for 15 min by shaking them in a Coomassie Brilliant Blue R250 staining solution, providing a detection limit of 200 to 500 ng mm^{-2}. Excess dye was removed by repeatedly heating the gel in water using a microwave. The molecular weight of protein bands was determined by comparison to protein mass standard LMW (Thermo Fisher Scientific).

3.4.4 Western Blotting and Immunodetection of Proteins

Western blotting refers to the transfer of proteins separated by SDS-polyacrylamide gel electrophoresis to a polyvinylidene difluoride membrane (Towbin et al. 1979). Hydrophobic and dipole-mediated interactions lead to immobilization of proteins on the membrane. Proteins can then be detected by suitable interaction with an antibody. In this work, a primary-secondary antibody interaction was used to detect FLAG-tagged proteins. The primary *anti*-FLAG antibody can in turn be detected by a secondary antibody coupled to horseradish peroxidase, which can provide a chemiluminescence signal at 425 nm by oxidizing luminol to 3-aminophthalate.

For Western blotting, cell cultures were harvested by centrifugation (8 min, 3200 g, 4 °C) and resuspended in 300 µl WB resuspension buffer. Subsequently, cells were lysed by sonication on ice (20 % amplitude, 45 s, alternating 2 s pulse, 2 s pause) and centrifuged (5 min, 16000 g, 4 °C) to remove insoluble cell debris. 3 µl of the supernatant was subjected to SDS-polyacrylamide gel electrophoresis (section 3.4.3). The membrane was cut to the size of the resolving gel and activated by shaking in methanol for 2 min. After washing of the methanol, the membrane and six sheets of Whatman filter paper were soaked in WB blotting buffer for 15 min. The sandwich 3x Whatman paper – gel – membrane – 3x Whatman paper was assembled and locked into a Biometra Fastblot B34 apparatus (Analytik Jena AG). The blotting reaction proceeded at 100 mA for 1 h.

For immunodetection of FLAG-tagged proteins, the membrane was shaken in blocking solution for 40 min. Subsequently, the membrane was subjected to 2 µg/ml Rat Anti-DYKDDDDK

Antibody in blocking solution and incubated for 1 h at room temperature. After washing the membrane with PBS-T, 0.16 µg/ml Peroxidase AffiniPure Goat Anti-Rat IgG+IgM (H+L) antibody (Jackson ImmunoResearch Europe Ltd) in blocking solution was added, and the membrane was incubated for 1 h at room temperature again. Free antibody was washed off with PBS-T, and the SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) was used according to the manufacturer's instructions for detection. A Luminescent LAS-3000 Imager (Fujifilm) was used for visualization.

3.4.5 Electrophoretic Mobility Shift Assays

Native polyacrylamide gel electrophoresis was used to examine migration behavior of proteins in their native state and mainly for investigation of RNA-protein complexes in electrophoretic mobility shift assays (EMSAs). For high sensitivity, RNAs employed in EMSA experiments were routinely labeled with ^{32}P-radionuclide (section 3.2.8). Components for native gels were 40 % acrylamide/bisacrylamide 18:1, glycerol (final concentration 2.5 % (v/v)), H_2O, and suitable buffer. The buffer in the gel was also applied as respective running buffer. Polymerization was induced by addition of 5 µl TEMED and 50 µl 10% APS per 5 ml gel. For EMSA experiments conducted to validate interactions identified in genomic SELEX experiments, 1x TB buffer was employed as gel buffer and running buffer (see materials section 2.7.3). For examination of QorA migration behavior under different pH values, used buffer systems include TB (pH 8.15, 8.3, 8.6, 8.9), CHES (pH 9.5), Glycine-NaOH (pH 9.4, 9.7, 10.0), CAPS (pH 10.4). Pre-incubation of RBP candidates and RNAs identified in genomic SELEX was carried out in the same buffer conditions used for incubation of the respective RBP candidate and genomic RNA library during genomic SELEX (see also method section 3.4.11). These conditions were meant to resemble ions and pH of *E. coli* cytosol in a grossly simplified fashion:

Tris/HCl (or HEPES for PykF)	50 mM, pH 7.5
NaCl	80 mM
KCl	10 mM
$MgCl_2$	0.8 mM
DTT	0.5 mM

Samples to be subjected to native gel electrophoresis were supplemented with 10 % glycerol to facilitate sinking into gel pockets. Electrophoresis was typically carried out at 150 V and 20 mA for 0.5 – 1 hour. Bromophenol blue and xylene cyanol in a separate pocket was used to estimate migration speed of samples. For visualization of ^{32}P-labeled RNA, gels were wrapped into plastic foil and exposed to a Cyclone Storage phosphoscreen (Canberra Packard) for at least 2 hours. The phosphoscreen was subsequently visualized on a Cyclone Storage Phosphor System (Canberra Packard).

For quantification of band intensities from EMSA RNA-protein titration experiments, pixel densities were assessed using the OptiQuant v3.0 software. The bound RNA density was corrected by a factor corresponding to the pixel intensity at complex height in the lane without protein, and ratios of bound-to-total RNA densities were plotted against employed protein concentration. Data points were fitted using a hyperbolic equation:

$$y = \frac{100\,\% \times c^{\text{Protein}}}{K_{\text{D}} + c^{\text{Protein}}}$$

where y is the fraction of bound RNA. Fitting was conducted using SigmaPlot v14 software.

3.4.6 Urea Polyacrylamide Gel Electrophoresis

Urea polyacrylamide gels were used to validate integrity of single RNAs as well as RNA pools employed during SELEX (section 3.4.11). Gels were prepared and operated identical to native gels described in section 3.4.5, with the exception of the gel containing a final concentration of 8 M Urea. The Century$^{\text{TM}}$-Plus Marker template (Thermo Fisher Scientific) was used for *in vitro* transcription of a radiolabeled size-ladder if analyzed RNAs required size determination.

3.4.7 Analytical Size Exclusion Chromatography

Analytical size exclusion chromatography is based on the same principles described for preparative size exclusion (section 3.3.4). A Superdex S200 10/300GL column (GE HEALTHCARE) was employed in combination with an ÄKTA micro chromatography system (GE HEALTHCARE) to separate 20 µl samples of protein solution. Samples were separated at a flow rate of 0.3 ml/min. For exact size-determination, elution profiles of calibration proteins were generated alongside the elution profile of the protein of interest. For calibration, the commercial Cytiva Gel Filtration Calibration kit (GE HEALTHCARE) was employed.

3.4.8 Circular Dichroism Spectroscopy

Circular Dichroism spectroscopy was used as validation for protein folding after expression. Employed proteins were first centrifuged (30 min, 16000 g, 4 °C) to remove residual aggregate. Protein solutions were diluted (or concentrated beforehand) so that their minimal ellipticity signal between 190 and 260 nm was below -15 mdeg. Spectra were measured at a Jasco J-815 (JASCO GmbH) between 190 and 260 nm with 5 repetitions for smoothing. The value θ_{MRW}, which is a normalized form of the ellipticity θ generally used to compare different proteins, was calculated as follows:

$$\theta_{\text{MRW}} = \frac{\theta}{c \times \text{d} \times N_{\text{R}}}$$

θ: observed ellipticity [mdeg]

c: concentration [mol/l]

d: pathlength [cm]

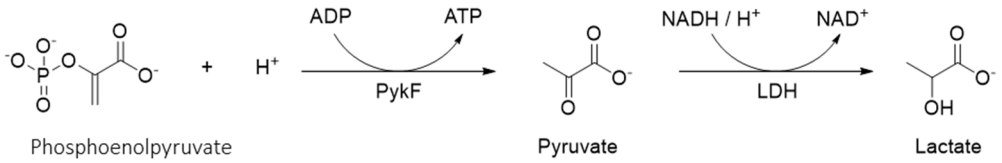
N_{R}: number of protein residues

3.4.9 Activity Assays for Metabolic Enzymes

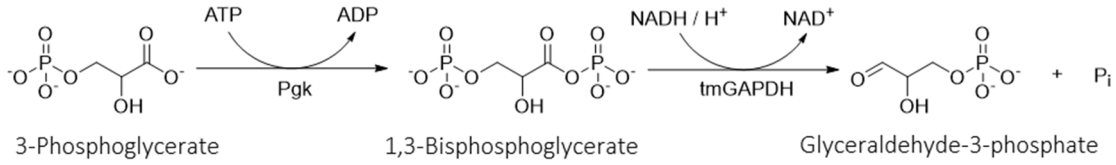
When possible based on enzymatic activity, photometric turnover assays were conducted for verification of functional integrity of enzymes that were to be subjected to RNA selection experiments. This includes any enzyme which either directly interconverts species with diverging molar absorption coefficient or for which a suchlike conversion was available through

straightforward follow-up reactions. All respective conversions are listed hereinafter. For conciseness, the specification of concentrations for all species is omitted, as activity assays were conducted in lose fashion to validate the principal presence of enzymatic activity rather than recording full kinetic profiles, and to this end it was not essential to optimize employed concentrations with respect to making the first step of coupled assays the rate-limiting one. Time-dependent absorption measurements were conducted in 200 µl quartz cuvettes on a Jasco V-650 spectrometer (JASCO GmbH). In all cases, enzymatic activity was indicated by one of the following conversions: NAD(P)H oxidation ($\Delta A_{340\,\mathrm{nm}}$), $NAD(P)^+$ reduction ($\Delta A_{340\,\mathrm{nm}}$), Oxidation of hypoxanthine to uric acid ($\Delta A_{293\,\mathrm{nm}}$), or oxidation of pyridoxine-5-phosphate to pyridoxal-5-phosphate ($\Delta A_{414\,\mathrm{nm}}$).

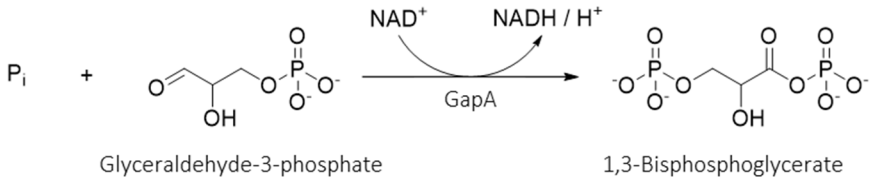
- PykF -



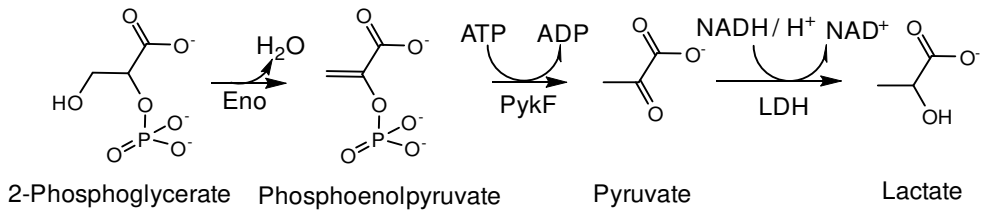
- Pgk –



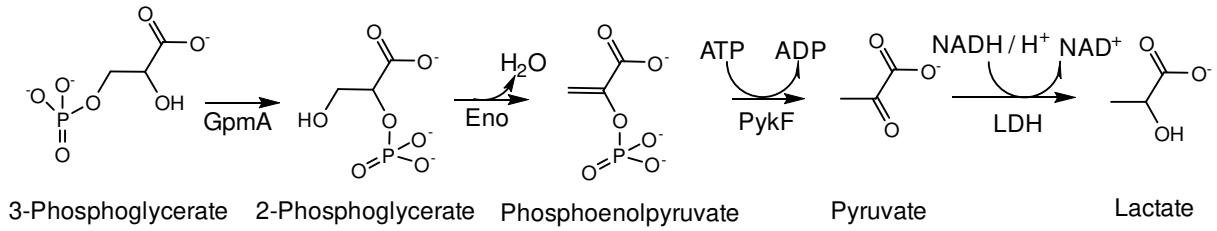
- GapA –



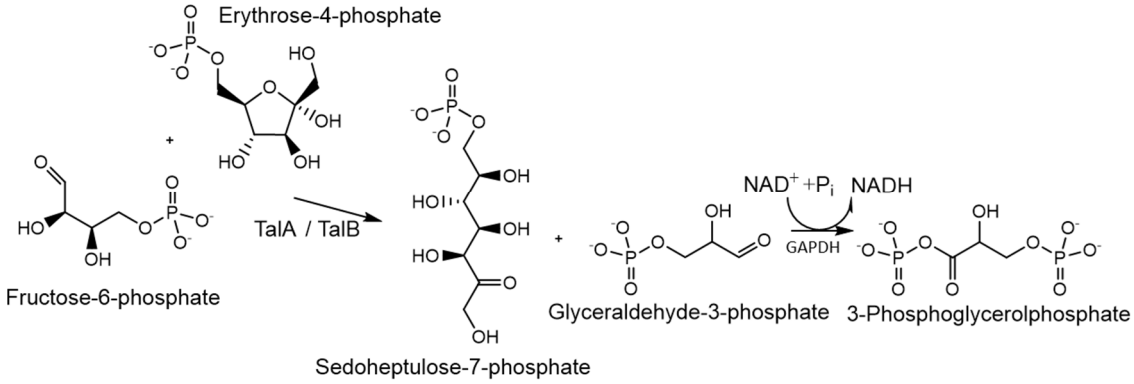
- Eno –



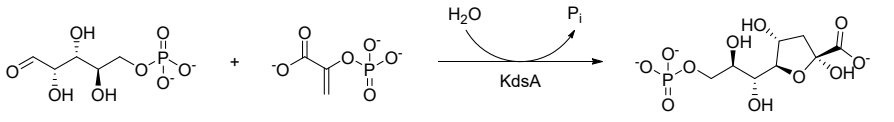
- GpmA –

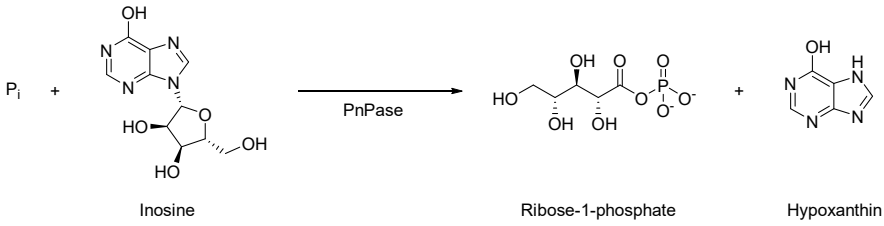


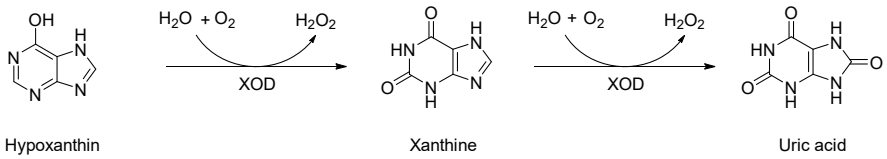
- TalA and TalB -



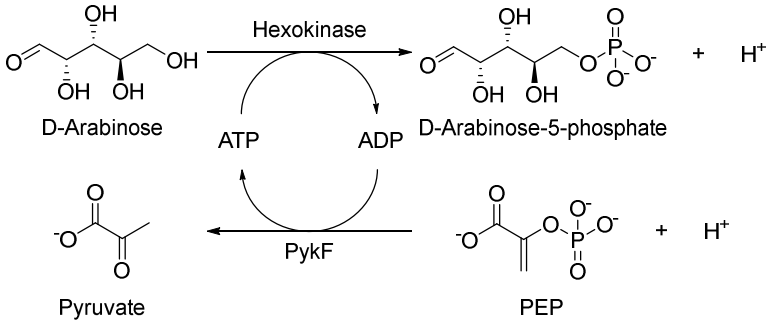
- KdsA –



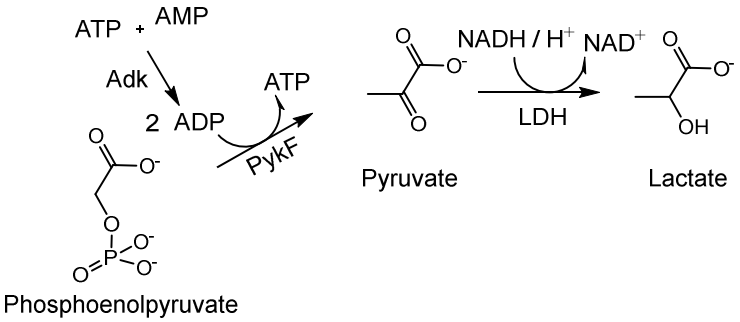




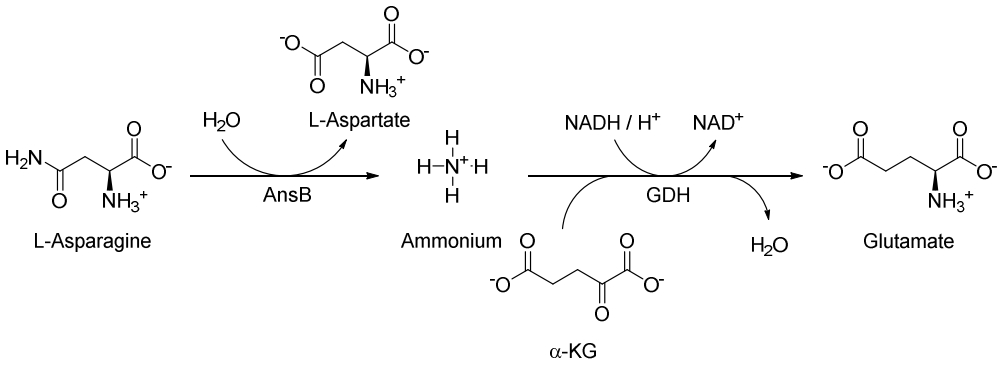
The substrate D-Arabinose-5-phosphate was synthesized enzymatically using Hexokinase as illustrated below. To avoid ADP accumulation, the reaction was coupled to pyruvate kinase. The reaction had low yield, but produced enough D-Arabinose-5-phosphate to assay KdsA activity.



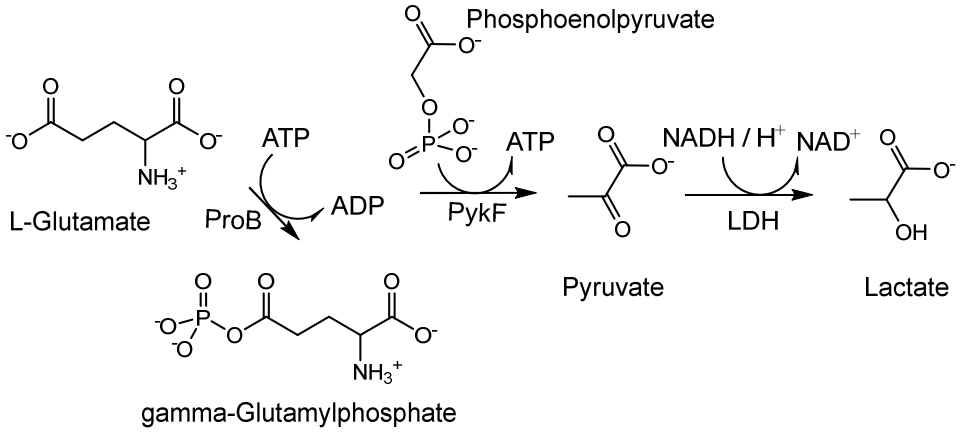
- Adk –



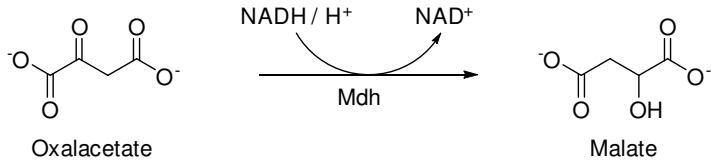
- AnsB –



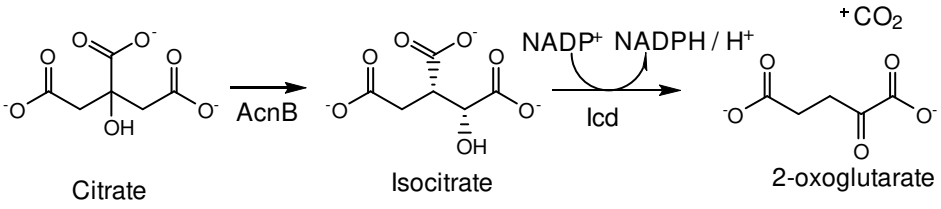
- ProB –



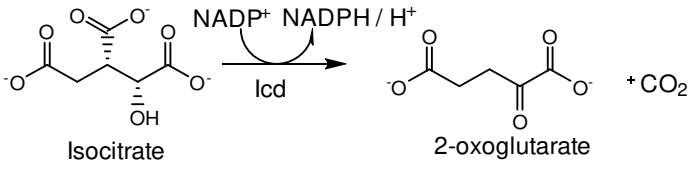
- Mdh –



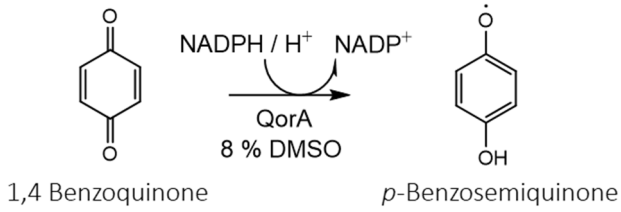
- AcnB –



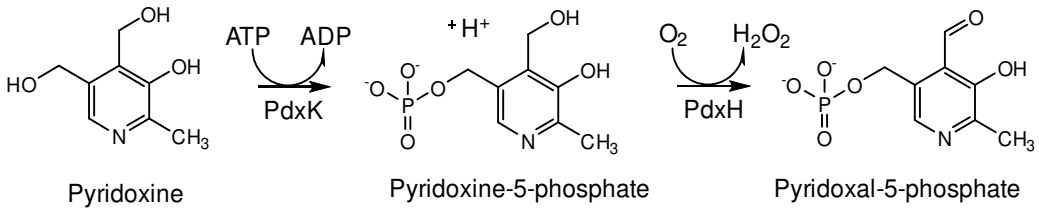
- Icd –



- QorA –



- pdxH –



3.4.10 iCLIP

Individual nucleotide resolution UV-crosslinking and immunoprecipitation (iCLIP) can provide a snapshot of all RNAs interacting with a target protein by freezing the interactions with UV-light, immunoprecipitating the complexes and subjecting isolated RNAs to next-generation sequencing (König et al. 2010).

In this work, iCLIP was used to examine proteins carrying a tandem FLAG-Tag. The introduction of FLAG-Tags by genomic DNA modification is described separately in section 3.2.12. A brief schematic of iCLIP is shown in Figure 6. All steps following UV-crosslinking and prior to next-generation sequencing were carried out by collaboration partner Dr. Oliver Rossbach, who routinely works with iCLIP experiments.

In preparation of UV-crosslinking, *E. coli* BW25113 strains carrying tandem FLAG-Tags at the respective gene of interest as genomic modification were grown in 5 ml selective LB-medium overnight. These pre-cultures were used to uniformly inoculate 1000 ml selective LB-medium, and cultures were shaken at 37 °C and 150 rpm until an OD of 1.6 was reached. Cells were harvested by centrifugation (10 min, 4000 g, 4 °C), washed, thoroughly resuspended in a fixed, small volume of PBS-T buffer, and poured into ∅ 15 cm petri dishes. The volume should match the petri dish, so that the liquid level is not too high, ensuring homogenous application of UV-

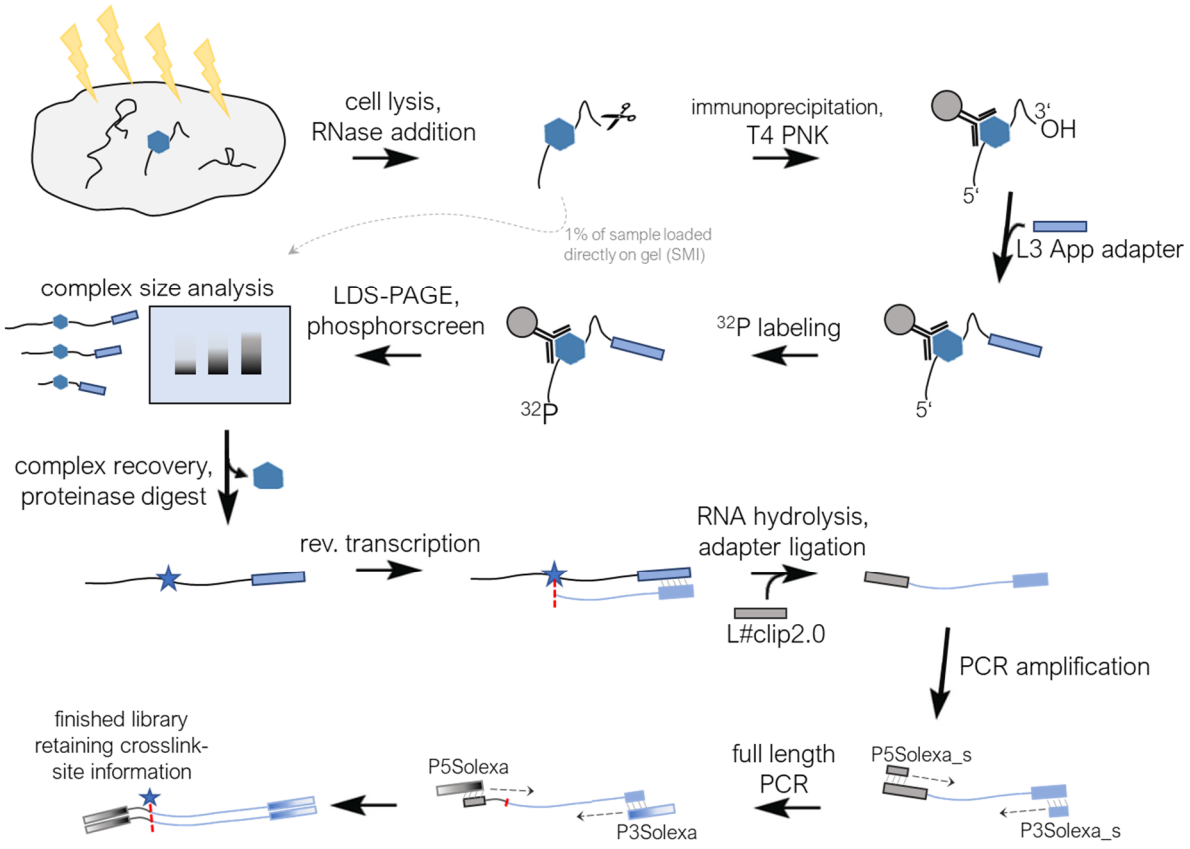


Figure 6: Schematic overview of iCLIP. The final library retains information about the exact crosslinking site through reverse transcription being forced to terminate by covalently linked amino acid leftovers.

light to all cells. From this point, cells were continuously kept on ice. Cells were exposed to 3 × 333 mJ/cm^2 UV-light (λ = 254 nm) with intermediary panning, harvested by centrifugation (10 min, 4000 g, 4 °C), shock-frozen in liquid nitrogen, and stored at -80 °C until use.

The principal procedure from cell lysis to library preparation as conducted by Dr. Oliver Rossbach can be looked-up in recent literature (Buchbender et al. 2020), with some aberrations due to the prokaryotic nature of proteins examined within this thesis. The experimental steps as described by Buchbender *et al.* shall be transcribed here in brief form to provide an overview.

For lysis, UV-light treated cell pellets were resuspended in buffer and subjected to sonication. Cell lysates were freed of cell debris and supplemented with RNase I for partial RNA digestion, initially to verify the presence of RNA molecules within the crosslinked complexes, and later on to shorten the RNA overhangs adjacent to crosslinking sites for preparative scale cell lysates. The target length of fragments for NGS sequencing is usually between 50 and 200 nt. The amount of RNase necessary varies from experiment to experiment and was optimized individually, depending on RNase batch, lysate concentration and cell type. The optimization of RNase concentration was eased by the adoption of sonification of cell lysis, which shears long RNA strands. After removal of an aliquot (1% of sample) used as size-matched input, digested cell lysates were subjected to immunoprecipitation by incubation in anti-FLAG® M2 magnetic beads (Sigma-Aldrich). Beads were washed in high-salt buffer and resuspended in PNK buffer. The 3' ends of RNA strands were dephosphorylated by addition of T4 polynucleotide kinase. After incubation, beads were washed again and 3' L3-App adapters were ligated with T4 RNA ligase. Subsequently, 5' RNA ends are end-labeled with T4 Polynucleotide kinase and ^{32}P-γ-ATP. Beads were washed again and then precipitated by placing the tube on a magnet. The supernatant as well as the size-matched input were loaded onto a 4-12 % Novex NuPAGE Bis-Tris gel (Thermo Fisher Scientific). The commercial gel system ensures a constant pH of approximately 7, preventing alkaline hydrolysis of RNA. The protein-RNA complexes were blotted from the gel onto a nitrocellulose using the Novex XCell II Blot Module (Life Technologies). The membrane was rinsed in cold PBS buffer, wrapped in plastic foil, and exposed to a phosphor screen to develop the autoradiogram. Development time depends on the amount of protein-RNA complex that could be immunoprecipitated. At this point, autoradiogram and Western blot are compared. If colocalized bands and the behavior of radioactive RNA-signals with respect to RNase digests substantiates the presence of sufficient amounts of protein-RNA complex, the RNA can be isolated and further processed for library preparation.

For RNA isolation, protein-RNA complexes were excised from the nitrocellulose membrane using the autoradiogram as a mask. The width of the excised band should be chosen generous but avoid unwanted protein-RNA complexes of wrong molecular weight, as can be judged from a high RNase experiment. The membrane was transferred into a tube and treated with proteinase K to degrade proteins. The crosslinking site is retained. The solution was then supplemented with urea and neutral pH phenol/chloroform (the neutral pH is necessary to purify DNA-RNA after adapter ligation). Phases were separated after thorough shaking and the aqueous layer was transferred into a new tube. RNA was precipitated at -20 °C with sodium acetate, ethanol, and

GlycoBlue. The precipitate was pelleted by centrifugation, dried, and resuspended in water. RNA was reverse-transcribed with SuperScript III reverse transcriptase (Thermo Fisher Scientific) and an RT-oligo pairing to the L3-App sequence. RNA templates were removed by NaOH treatment, and the cDNA was subjected to MyOne™ Dynabead cleanup from ThermoFisher. At the end of the cleanup-procedure, DNA was eluted from the beads with water, but the beads were kept in the sample. For second adapter ligation, L#clip2.0 oligos were added. These oligos carry unique molecular identifiers and a separate barcode for each sample to allow for multiplexed NGS sequencing. A ligation master-mix containing RNA ligase and ATP was added to the sample-adapter mix and the reaction proceeded overnight. Fresh dynabeads were added and the sample was subjected to MyOne™ cleanup again.

For first PCR amplification, cDNA was amplified with primers P5Solexa_s and P3Solexa_s and Phusion polymerase. After that, excess primer dimers were removed using the ProNex® size selection system by Promega, aiming to retain inserts < 20 nt. The short primer version facilitate the size selection step. For second PCR amplification, the number of cycles required was first optimized by checking the success of a PCR reaction containing an aliquot of cDNA, primers P5Solexa and P3Solexa, and Phusion polymerase by capillary gel electrophoresis on a TapeStation. After that, the PCR reaction was carried out in preparative scale with half of the cDNA sample. PCR success was verified once more on TapeStation, and at this point multiple barcoded samples were combined in equimolar amounts into one library. The library was subjected to ProNex® size selection again to remove residual primers. The sample-to-ProNex® bead ratio was adapted with respect to the longer primers in comparison to the first cDNA-amplification. The success of the primer removal can be checked on TapeStation. The library was stored at -20 °C until use.

3.4.11 SELEX Protocol

A schematic overview of the general SELEX selection procedure is illustrated in Figure 7. First, the library was routinely transcribed (section 3.2.8) with addition of [α–^{32}P]ATP. Incorporation of radioactively labeled nucleotides allowed for quantification of transcription yield and retained RNA during the subsequent selection process. The RNA transcript was processed by addition of ½ volume of 0.5 M EDTA pH 8.0 in order to complex magnesium ions. Subsequently, ½ volume of 7.5 M ammonium acetate and 1.5 volumes of isopropanol were added. The sample was incubated at -20 °C for at least 15 min to precipitate RNA, centrifuged (15 min, 16000 g, 4 °C), and the supernatant was discarded. The RNA pellet was washed at least twice by thoroughly swirling up the pellet with 70 % ethanol and re-isolating the pellet by centrifugation (3 min, 16,000 rpm). The purified pellet was resuspended in a suitable amount of H_20, up to 100 µl. The success of transcription was intermittently verified by Urea-PAGE in between selection rounds (section 3.4.6).

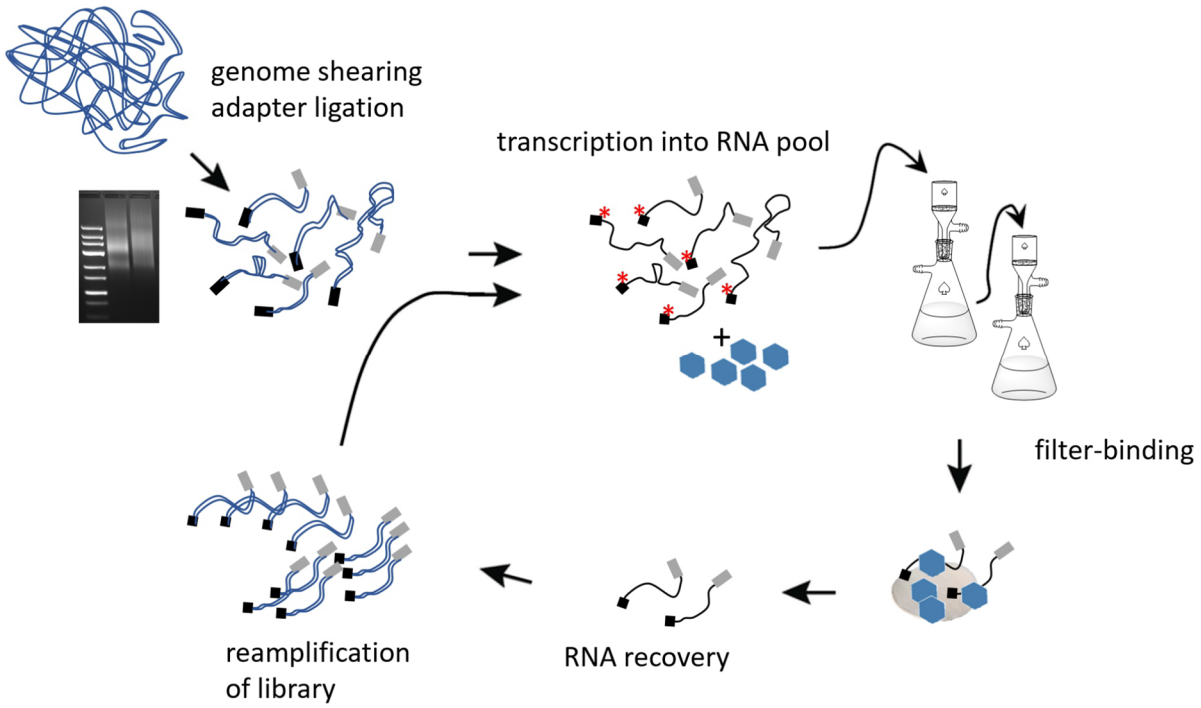


Figure 7: SELEX overview.

The concentration of RNA after resuspension was determined by scintillation counting, as UV-VIS spectrometric concentration determination for radioactive samples was not available within the isotope work area. 2 µL of RNA sample were dissolved in 2 ml of ROTI scintillation cocktail, and the sample activity was measured on an scintillation counter device. The concentration c was calculated by formulas listed in section 3.2.11. The counting efficiency was determined by calibration, employing a dilution series of a known amount of radioactivity. The calibration was conducted in regular intervals and usually corresponded to a factor of 0.7 - 0.8.

The RNA library was then employed in tenfold molar excess over protein amount and was incubated with the respective protein of interest in a final volume of 500 µL, according to the following scheme:

SELEX Buffer	50 mM Tris/HCl pH 7.5, 80 mM NaCl, 10 mM KCl, 0.8 mM $MgCl_2$*
DTT	0.5 mM
Protein of interest	1 µM
RNA library	10 µM

*HEPES instead of Tris for PykF, due to a Tris-inhibitory effect

Prior to protein addition, the RNA library was heated to 95 °C for 1 min and then immediately exposed to room temperature for 10 min, allowing for unfolding and redevelopment of RNA secondary structures in SELEX buffer. Incubation took place for 30 min at 37 °C. Lower temperature was deployed when deemed necessary based on thermal stability of the respective protein of interest. For initial rounds of SELEX, the sample was also subjected to a procedure referred to as negative selection (Ellington and Szostak 1992). After the refolding step, the RNA library is filtered using identical setup and filter-materials as used in the subsequent selection step (described below). This procedure helps eliminating unspecific target RNAs, which might get selected based on their increased affinity towards the filter material, instead of a heightened

affinity towards the target protein of interest (Figure 8). While this is most important for a randomized library which contains a large diversity of aptamers and likely contains unspecific binders, the negative selection step was also applied to the genomic RNA library to ensure maximum specificity.

A suction pump was used to filter the sample through a 0.2 µm pore sized M24 cellulose filter slice (Whatman) which was pre-wetted with SELEX buffer. The filter stops proteins and protein-RNA complexes but does not retain free RNAs (Rio 2012). A 1 µL aliquot of the filter binding assay input was saved and subjected to scintillation counting in order to quantify the RNA input amount by radioactivity. The filter was washed three times with 1 ml of SELEX buffer to promote flushing of free and low affinity RNAs. The stringency of this filtering step can be raised by increasing the number of washing steps if desired. The initial flowthrough as well as flowthroughs from washing steps were captured, and an aliquot of each was subjected to scintillation counting, in order to quantify the portion of eluted RNA.

Phenol-chloroform extraction was used in order to elute and isolate RNA complexed with proteins. The filter was put into a mixture of 400 µL 8 M urea and 500 µL ROTI®Aqua-P/C/I pH 4.5 (25:24:1) and shaken rigorously in a Eppendorf thermomixer comfort 5355 (10 min, 1400 rpm, RT). The filter was removed from the liquid, put into a new mixture of 200 µL 8 M urea and 200 µL ROTI®Aqua-P/C/I, and again subjected to rigorous shaking (10 min, 1400 rpm, RT). The filter was then removed and subjected to scintillation counting. The two eluate fractions were pooled, supplemented with 200 µL H_2O, and centrifuged (10 min, 16000 g, 4 °C) for phase separation. The upper phase was isolated and mixed with an equal volume of chloroform. The mixture was shaken thoroughly and centrifuged (3 min, 16000 g, 4 °C). The upper phase was isolated again, supplemented with 1/10 volume 3 M natrium acetate pH 6.5, 1 volume isopropanol, and 1 µL GlycoBlue™ coprecipitant (Thermo Fisher Scientific), and incubated for at least 15 min at -20 °C in order to precipitate RNA. The RNA was pelletized by centrifugation for (15 min, 16000 g, 4 °C), washed using 100 µL 70 % EtOH, and again pelletized by

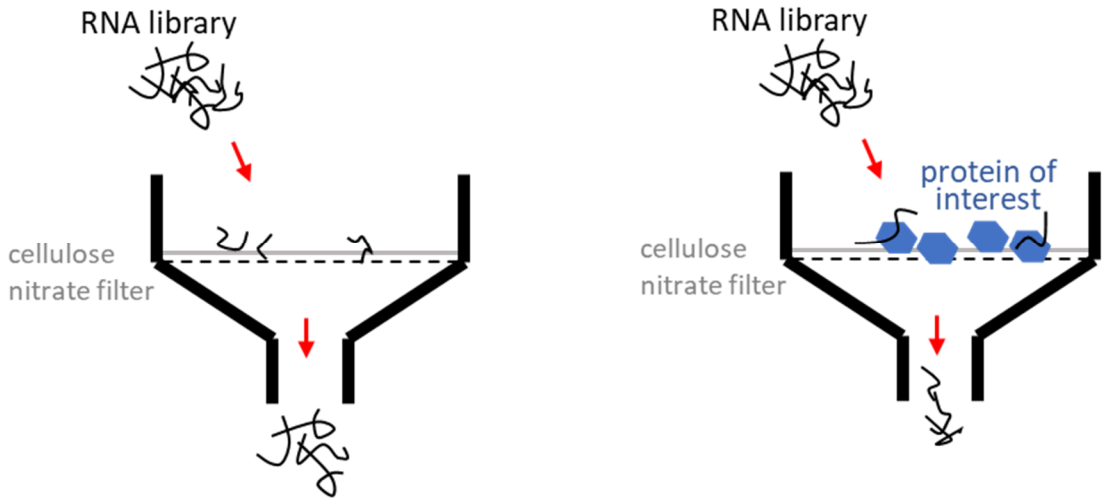


Figure 8: SELEX filtering steps. Left: negative SELEX – RNAs with high affinity to the filter are eliminated from the library. right – actual filter selection step.

centrifugation (5 min, 16000 g, 4 °C). The pellet was dissolved in 67 µL H_20, of which 4 µL were subjected to scintillation counting.

For reverse transcription, 10 mM dNTP mix and 2 µL of i7_rev primer was added to 63 µL of RNA. The sample was incubated at 65 °C for 5 min and cooled to 4 °C to enable primer annealing. After addition of 20 µL SS-IV buffer, 10 µL 0.1 M DTT, and 2 µL SuperScript IV reverse transcriptase, DNA synthesis proceeded for 10 min at 55 °C, followed by denaturation of the enzyme for 10 min at 80 °C. DNA was precipitated by incubation at -20 °C for at least 15 min after addition of 1/10 volume 3 M natrium acetate, 1 volume isopropanol, and 1 µL GlycoBlue™ coprecipitant. The precipitate was pelletized by centrifugation (15 min, 16000 g, 4 °C). Following removal of supernatant, the DNA was dissolved in 70 µL H_2O. Subsequently, PCR amplification (section 3.2.1) by Q5 Polymerase was used to obtain a DNA yield sufficient for subsequent transcription into RNA. The number of PCR cycles was adapted based on the RNA yield from filter binding assay. PCR amplification was verified by agarose gel electrophoresis (section 3.2.3). Additional PCR rounds were conducted when deemed necessary by visual inspection of the DNA library. After that, the next round of SELEX was started by transcription of the DNA library with [α–^{32}P]ATP.

The scintillation counter results for the input and elution fractions of the filter binding assays, the filter, and for the eluted RNA were used to track the progress of the selection procedure. The proportional distribution of RNA during the filter binding assay can be tracked through normalization to input. In case the protein preferably binds to specific sequences, it is to be expected that RNA concentration in filter assay eluates will decrease from one SELEX round to another, while the RNA amount isolated from the filter by phenol-chloroform extraction is expected to go up.

3.4.12 Next Generation-Sequencing and Read Processing

All next generation sequencing experiments were performed on an Illumina MiSeq Device located in the chair of Biochemistry I, University of Regensburg and generously made available by Prof. Dr. Gunter Meister. Read processing and genome mappings were conducted by collaboration partner Gerhard Lehmann.

iCLIP libraries were sequenced using a MiSeq Reagent Kit V2 MS-102-2001 50 cycles. The workflow in Linux `bash` and `R` as well as necessary external dependencies for iCLIP data analysis are listed in detail in recent literature (Busch et al. 2020). The protocol was adopted with minor adoptions related to the prokaryotic genome. The workflow started with a general quality control of sequencing data. Reads were then filtered based on barcode quality. To exclude contamination and carry-over from other experiments, it was checked that expected barcodes were the most abundant ones. Subsequently, demultiplexing and adapter trimming were performed, grouping the reads by biological sample. At this stage, unique molecular identifier (UMI) sequence information from L#clip2.0 oligos necessary for subsequent deduplication was kept and added to the ID of each read. After demultiplexing, reads were mapped to the complete genome of *E. coli* strain K12 substrain MG1655. After mapping, any number of reads with identical UMIs were reduced to one read. Mapped reads were transformed also into 1 nt length crosslinking events,

since the iCLIP sequencing reads start precisely at the position where cDNAs are truncated during reverse transcription. Peak calling was done both by manual inspection and by utilizing `PureCLIP` (Krakau et al. 2017).

SELEX libraries were sequenced using a Illumina MiSeq Reagent Nano Kit v2 MS-103-1003 500 cycles. Data was processed similarly as described above, conducting quality control, demultiplexing, and grouping of reads into biological samples. The mapping was done with `Bowtie`. Two datasets were created, one which includes PCR duplicates and one where duplicates are reduced to one read, as both were used for inspection of the experiments.

Mapping data (.bam files) and crosslinking-sites (.bed files) were examined in the most current version of the integrative genomics viewer IGV (Robinson et al. 2011).

3.4.13 Notes on Evaluation Principles of Genome Mapping Data

In genomic SELEX, the starting library features a more or less even distribution of reads. Principally, only nucleic acids retained by the protein of interest should become amplified, whereas unbound RNAs should become depleted with progressing selection rounds. Ideally, sequence enrichment at a certain genomic locus should only arise in one experiment and not across multiple libraries of experiments employing different proteins - the latter case would suggest selection pressure originating from the method itself rather than from the protein, like PCR amplification bias or filter-binding. Figure 9 shows scenarios of 'good' and 'bad' peaks within an evolved library, exemplarily extracted from SELEX experiments conducted in this work. A healthy peak should feature a surplus of reads relative to the average read density, no matter if PCR duplicates have been deleted or not during processing of sequencing data (Figure 9, top). The shape of coverage should ideally form a smooth curve, similar to a Gauss-distribution (Figure 9, left-bottom). This is because the original library covers the selection-defining sequence stretch with multiple nucleic acids of different lengths, and all of them should be retained and amplified equally. If single reads are overamplified (Figure 9 , right bottom), the impact of the protein on the final copy number is less clear, and disproportional PCR amplification might have occurred. However, scenarios like an overly high washing stringency thinning out the library in early iterations would raise the mapped-to-deduplicated read ratio of a potentially valid selection target, and peaks might still be valid despite lacking proper Gauss-distribution of reads. It is therefore critical to correctly contextualize peak shapes based on read numbers and read distribution for individual experiments, and constraints defining significant peaks were specified separately to the best of knowledge for each experiment. Owing to the relatively small size of the *E. coli* genome (~4,600,000 nt), mappings can also be surveyed manually in the genome browser within acceptable time expenditure. Thus, peaks were also manually assessed no matter if classified as significant or not, a potentially superior way to review the data and locate all enriched genomic loci. In the genomic SELEX library, the defining feature causing a particular RNA sequence to be enriched is expected to be located at a position where all reads within a cluster do overlap. When using a motif-finding algorithm to break down multiple clusters to a

common motif based on the underlying sequences of each cluster, this motif should be verified to colocalize with the apex of coverage within each cluster.

iCLIP captures any RNA species that has been cross-linked to the immunoprecipitated protein of interest. Depending on the respective protein, the pool of captured sequences has the potential to be highly variable. Theoretically, specific RNA-binding proteins should capture a narrower ensemble of RNA species, while the interactome composition of unspecific RNA-binding proteins might reflect transcript abundancy of the cell. However, transcript abundancy has the potential to impact any CLIP data set, since crosslinking captures any surrounding molecule based on spatial proximity, and even specific binders might 'see' abundant species like rRNAs a lot. To get a better picture of what is background and what comprises a specific RNA target, a size-

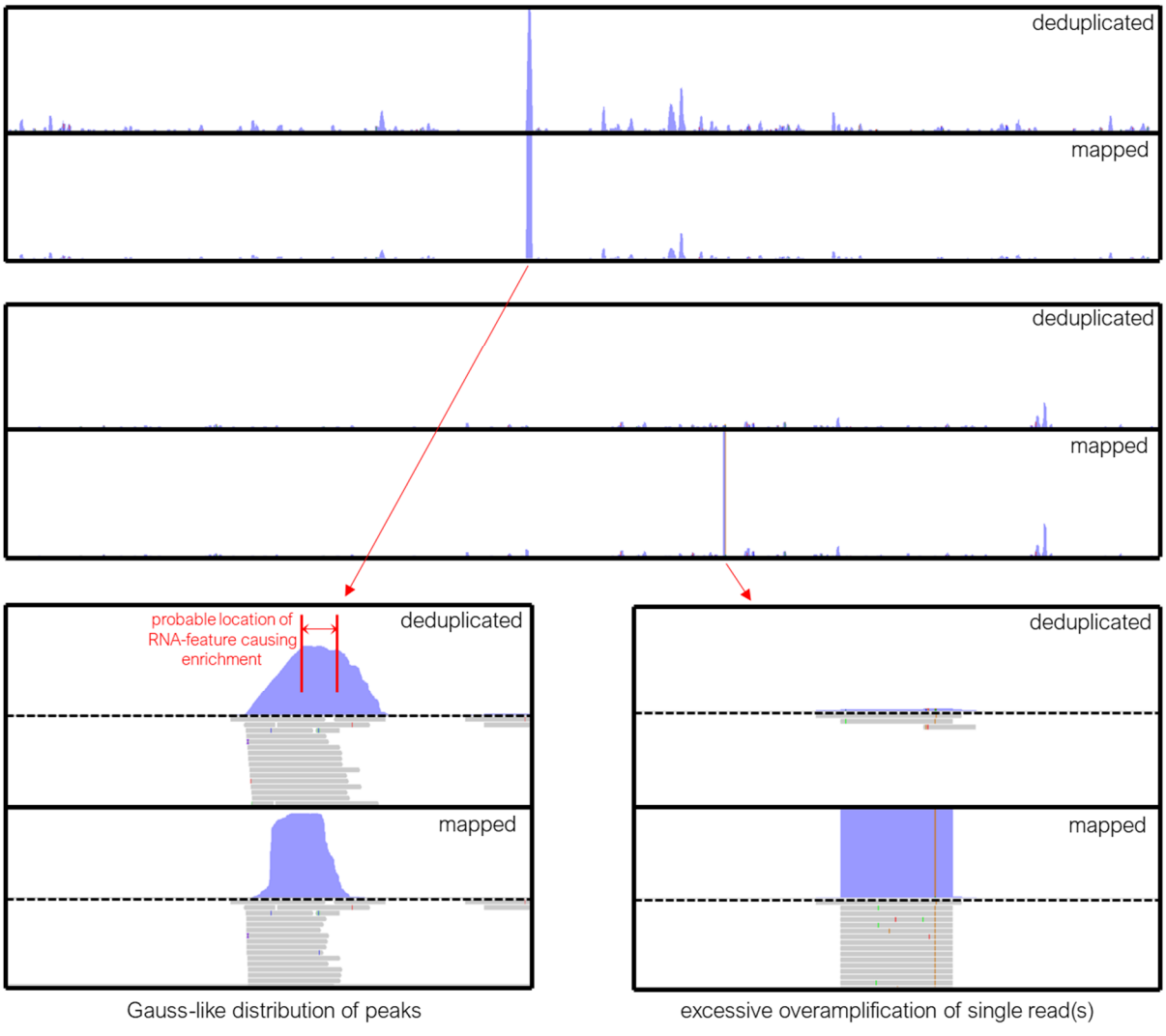


Figure 9: Illustratory examples of genome mappings after SELEX experiments. Windows at the top and in the middle show coverage traces of genome mapping excerpts. At the bottom, marked peaks are illustrated close up, and single reads (grey bars) are shown in addition to the coverage traces. Marked peaks highlight examples of a 'healthy' cluster with sufficient deduplicated reads (left-bottom) and a cluster mainly comprised of PCR duplicates, judging the high ratio mapped-to-deduplicated reads (right-bottom). In a healthy cluster, the RNA-feature responsible for enrichment likely coincides with the apex of coverage (marked by red double-sided arrow).

matched input control was adopted (Wheeler et al. 2018), and libraries of interest were compared to this background control. Aforementioned principals on healthy peak shapes still hold true for CLIP mappings, even though duplication should be less pronounced, with CLIP featuring less PCR amplification than SELEX. As for genomic SELEX libraries, mappings were manually surveyed in a genome browser. `PureCLIP` was deployed as a helper tool for initial peak calling, and genomic loci of uniquely called peaks were manually verified, since the algorithm can omit recognition of peak clusters based on technical reasons. The defining element within a sequence causing enrichment (putting enrichment based on transcript abundancy aside) is expected to be located at the very start of a read due to the abortion of cDNA synthesis at the cross-link site.

3.4.14 Microscale Thermophoresis

Microscale thermophoresis (MST) allows to measure ligand binding interactions by tracking the Soret effect with fluorescence. The Soret effect describes molecular movement in solutions bearing microscopic temperature gradients (Duhr and Braun 2006). The temperature fluctuation is induced by an infrared laser in MST, and the movement of biomolecules along the gradient (and therefore the progression of fluorescence signal) depends on their size, charge, and solvation entropy (Jerabek-Willemsen et al. 2014). A change in these properties based on a binding event leads to a quantifiable signal:

$$F_{\text{norm}} = (1 - x) \times F_{\text{norm}}^{\text{unbound}} + x \times F_{\text{norm}}^{\text{bound}}$$

F_{norm} is the normalized fluorescence signal which changes depending on the molecule being in its bound or unbound state. This difference in fluorescence allows the determination of the bound fraction and thus the K_D. MST measurements were conducted on a Monolith NT.115Pico device from NanoTemper, generously made available by Prof. Dr. Gernot Längst from the chair of Biochemistry III. The Pico RED detector (excitation wavelength 600-650 nm) was used to measure fluorescence-labeled Cy5-RNA (obtained commercially by Metabion). Prior to titration experiments, base intensities of free RNA and RNA-protein complex solutions were measured to ensure homogeneity of the fluorescence signal. Solutions were supplemented with 0.05 % Tween 20 in addition to experimental conditions to avoid adsorption of RNA to reaction tubes.

4 Results and Discussion

4.1 Selection of Enzyme Candidates

The general goal of this thesis was contributing to the exploration of the hypothesized REM network within *E. coli* by conducting suitable protein-centered experiments. With well-established, RNA-related moonlighting functions being *hitherto* restricted to isolated case examples (e.g. aconitase), the frequency in which REM interactions might be found is principally an unknown variable, and a scarce occurrence could not be principally ruled out. In this context, an approach was taken that aimed at generating streamlined data (mainly from genomic SELEX, but also from CLIP-Seq) for a sizable number of enzymes. This way, enough throughput should be generated to hopefully identify promising target interactions, or to be able to make a valid judgement in case that experiments turn out to be uneventful. Potentially RNA-interacting enzymes were chosen primarily based on two criteria described in the following subchapter and grossly summarized in Table 5.[2]

4.1.1 Selection Based on Specific Interactome Screenings in Prokaryotes

The first criterium for RBP candidate selection was the recovery of enzymes in one or more recently conducted high throughput screenings in prokaryotic systems, namely PTEX (Urdaneta et al. 2019), OOPS (Queiroz et al. 2019) and TRAPP (Shchepachev et al. 2019), also mentioned in introductory notes (see chapter 1.5). Based on interactome capturing experiments, these studies hypothesized RNA-binding properties for many proteins with no previous RNA-related annotation, among them a sizeable number of enzymes. Data from these experiments in particular was gauged promising for this work due to the isolation of cross-linked protein-RNA complexes *via* liquid phase separation (PTEX, OOPS) or silica beads (TRAPP) instead of poly(A) capturing or molecular tagging. Since polyadenylation of mRNA is rare in bacteria (Slomovic et al. 2006) and chemical modification of RNA can be undesired with respect to generation of unbiased datasets, it was contemplated that these studies constitute meaningful additions to the scarce interactome data available for prokaryotes and might in fact shed light on *hitherto* unidentifiable RNA-binding proteins. Within PTEX, TRAPP, and OOPS experiments, the recovered protein fraction was analyzed *via* mass spectrometry, assigning detected peptide fragments with appropriate software.[3] The detection was quantified and reported as FC-value (a measure for protein enrichment in crosslinked over non-crosslinked sample). The respective FC-performances of proteins ultimately chosen to be RBP candidates for this work are illustrated in Table 5 (middle column). Some of the candidates were selected mainly or exclusively based on this metric, namely GpmA, TalA, TalB, RpiA, KdsA, Upp, SodA, and AhpC. While the interactome screenings were considered a valuable resource, it is apparent that individual FC-values are not consistently coherent in between experiments. For example, comparing FC-values of GpmA

[2] Table 5 lists short names of proteins based on genes for compactness. Full enzyme names are listed in Table 15.

[3] PTEX and TRAPP used MaxQuant proteomics software; OOPS used python scripts deposited under https://github.com/TomSmithCGAT/CamProt

antithetically indicates slight depletion for PTEX and significant enrichment for TRAPP and OOPS. Gerovac et al. (2020) pointed out that contradictions within these screening results also extent to well-established RNA-binding proteins like CsrA (not listed here), which is enriched in TRAPP, but not in OOPS or PTEX. While interactome performances are listed for all candidates in Table 5, the remainder of the candidates were chosen based on additional criteria described below, in order to limit exposure to these particular screenings.

Table 5: Summary of enzymes selected as RBP candidates.

		High throughput screening results			
metabolism	enzyme	PTEX[a] (FC-value)	TRAPP[b] (FC-value)	OOPS[c] (replicates)	references suggesting RNA-binding activity[d]
glycolysis	PykF[e]	-	2.354	5/5	-
	Pgk[e]	0.575	2.156	5/5	binds coding mRNA regions (euk)
	GapA	-1.287	2.977	5/5	binds AU-rich elements (euk)
	Eno	-0.535	2.726	5/5	part of RNA degradasome (prok)
	GpmA	-0.040	2.965	5/5	-
pentose phosphate pathway	TalA	-	2.431	5/5	-
	TalB	-0.067	2.402	5/5	-
	RpiA[e]	1.023	3.230	5/5	-
lipopoly-saccharides	KdsA[e]	0.557	3.088	4/5	-
nucleotide metabolism	Upp[e]	1.229	2.604	5/5	-
	Adk	-2.368	3.037	5/5	general RNA co-purification (euk)
	ThyA	-	1.747	-	binds *thyA* mRNA (prok)
amino acid metabolism	AnsB	0.356	-	5/5	-
	ProB	-	3.038	-	Features a PUA domain (prok)
citric acid cycle	Mdh[e]	0.411	1.844	5/5	binds 3' UTRs of mRNA (euk)
	AcnB[e]	-2.300	1.024	5/5	binds *acnB* mRNA (prok)
	Icd[e]	-1.814	2.861	5/5	binds 5' UTR of mitochondrial RNA (euk)
oxidoreductases	SodA	1.017	4.233	-	-
	AhpC	1.585	3.500	5/5	-
	QorA	-	1.773	2/5	binds pH-responsive elements (euk)
	PdxH	-	2.050	-	-
uncharacterized proteins	YbiB	-	4.037	-	unspecific RNA binding (prok)
	YggX	-	2.434	-	-
positive control	MS2	-	-	-	binds phage-specific stem loops (prok)

[a, [b, [c Performance of each protein in high throughput screenings PTEX, TRAPP, and OOPS. FC-values are the $\log_2$-fold change in protein intensity between crosslinked and non-crosslinked samples, as detected by mass spectrometry. For OOPS, the table lists the number of replicates out of 5 that showed statistically significant FC-values >0, staying in line with original result presentation. TRAPP trialed varying UV dosages, data from highest irradiation is listed. All values are extracted from Urdaneta et al. (2019), Shchepachev et al. (2019), and Queiroz et al. (2019).

[d Data from studies with eukaryotes (euk) or prokaryotes (prok). Respective references are mentioned and cited in chapter 4.1.3.

[e Selection was pursued in the Master's Thesis of Funke (2019) based on criteria described in chapters 4.1.1 and 4.1.2. Experiments involving these enzymes were continued in this work.

4.1.2 Selection Based on Interactome Screenings in Eukaryotes

Interactome data for eukaryotes is more abundant in literature and was therefore considered as well, despite this work being limited to enzymes from *E. coli*. It was assumed that an evolutionary conservation of metabolic function in between homologous eukaryotic and prokaryotic enzymes might extend to potential moonlighting functions – an assumption supported by the dual functioning aconitase, for which RNA-binding activity has been reported in both domains of life (Hentze and Argos 1991; Tang and Guest 1999). An array of methodically varied RBP screenings in eukaryotes, concisely referenced in a recent review dedicated to undiscovered RBPs (Hentze et al. 2018)[4], was assessed already in the scope of a Master's Thesis preceding this work (Funke 2019). Among those screenings, four enzymes had been found to feature a high number of annotations as RNA-binding proteins, and respective prokaryotic homologues PykF, Pgk, Mdh, and Icd had been selected. Interestingly, none of the prokaryotic pendants carried confirmed annotations as RNA-binding proteins. The exploration of these enzymes was carried on in this work, based on purification and initial characterization attempts of these proteins within the Master's Thesis.

4.1.3 Specific RNA-Binding References for Selected RBP Candidates

In order to take data originating from individual, protein-centered studies into account as well, literature was consulted for references of RNA-binding enzymes outside of high throughput screenings, considering data from both eukaryotes and prokaryotes. Identified references of proteins ultimately considered as RBP candidates are described in the following paragraphs, sorted by metabolic pathway. These references are also grossly summarized in Table 5 (right column).

- Glycometabolism -

***ec*GapA -** As already mentioned in introductory notes, eukaryotic GAPDHs have been reported to interact with an array of different RNAs. Thereby, AU-rich elements adjacent to coding regions of mRNAs are frequently mentioned (Nagy and Rigby 1995; McGowan and Pekala 1996; Schultz et al. 1996; De et al. 1996). In contrast, essentially no coverage is existent for prokaryotic GAPDH, making it an ideal candidate RBP for this project. One study has postulated general RNase activity across both eukaryotic and prokaryotic GAPDHs (Evguenieva-Hackenberg et al. 2002). While noteworthy, this also asks for a verification of applied purification procedures to rule out RNase copurification, since the report lacks reaffirmation by further studies. Recently, after completion of the practical part of this work, GapA was mentioned in a study conducting PTEX for the *Staphylococcus aureus* interactome, where the enzyme was recovered alongside many FAD/NAD-binding proteins (Chu et al. 2022).

[4] These HT screenings are comprised of seven screens in *Homo sapiens* (Baltz et al. 2012; Beckmann et al. 2015; Castello et al. 2012; Castello et al. 2016; Conrad et al. 2016; Kramer et al. 2014), six screens in *Mus musculus* (Boucas et al. 2015; He et al. 2016; Kwon et al. 2013; Liao et al. 2016; Liepelt et al. 2016), and eight screens in *Saccharomyces cerevisiae* (Beckmann et al. 2015; Kramer et al. 2014; Matia-González et al. 2015; Mitchell et al. 2013; Scherrer et al. 2010; Tsvetanova et al. 2010).

***ec*Eno -** The eukaryotic representative of enolase was individually characterized by Shchepachev et al. (2019) following good performance in TRAPP experiments. A potential tRNA binding specificity could be demonstrated based on RNA-Seq analysis of *in vivo* crosslinked enolase-RNA complexes. Originally, a potential RNA binding function of enolase was recognized in a study describing involvement of enolase in tRNA transport to mitochondria (Entelis et al. 2006). At the same time, another study postulated the participation of *ec*Eno in the RNA degradosome of *E. coli* (Chandran and Luisi 2006). Only recently, specific RNAs were identified that inhibit human enolase activity *in vitro* and impair glycolysis in cultured HeLa cells (Huppertz et al. 2022).

***ec*Pgk -** Human phosphoglycerate kinase was isolated from bronchial epithelial cells as urokinase-type plasminogen activator receptor (uPAR) mRNA binding protein (Shetty et al. 2004), and the interaction was also demonstrated in subsequent immunoprecipitation and mobility shift assays.

***ec*PykF** – The unspecific RNA-binding activity of *E. coli* pyruvate kinase was assessed during the master's thesis of Franziska Funke (2019). It was found that *ec*PykF binds to random RNA fragments with relatively high affinity ($K_D < 1$ µM, see Figure S 27). Mammalian pyruvate kinase was found to show preferential binding to rRNAs and mRNAs related to the endoplasmic reticulum (Simsek et al. 2017). Recently, *ec*PykF was shown to interact with specific sites of the 70S ribosome, tackling the topic of ribosome-amplified metabolism (Yu et al. 2021). Also recently, a human pyruvate kinase isoform PKM2 was found to interact with long noncoding RNAs associated with cancer pathogenesis (Pu et al. 2022).

- Nucleotide metabolism -

***ec*Adk -** Adenylate kinase was reported to be co-purified with RNA from plant tissue, based on ribonuclease treatment restoring the ability of an adenylate kinase-containing fraction to bind MonoQ-resin (Deppert et al. 1992; Schlattner et al. 1995), and a non-enzymatic, regulatory role was suggested.

***ec*ThyA -** Thymidylate synthase is one of the better characterized examples of allegedly moonlighting enzymes. Human thymidylate synthase was originally reported to bind its own mRNA (Chu et al. 1993). Since then, an array of studies were published further characterizing this autoregulatory interaction, summarized in a review by Tai et al. (2004). It is hypothesized that the apo-form of thymidylate synthase inhibits its own expression, while substrate binding disrupts the complex with the own mRNA and promotes translation. The interaction was also observed for bacterial thymidylate synthase (Voeller et al. 1995). Even though the protein has been extensively studied, ThyA was selected as a candidate to complement existing results with *hitherto* unadopted methods.

- Amino acid biosynthesis -

***ec*ProB -** Glutamte-5-kinase from *E. coli* is comprised of an amino acid kinase (AAK) domain and a pseudouridine synthase and archaeosine transglycosylase (PUA) domain (Pérez-Arellano

et al. 2005), the latter of which is principally able to mediate RNA-binding (Aravind and Koonin 1999). While it can be found in many RNA-modifying enzymes, the role of the PUA domain in glutamate kinases remain elusive (Pérez-Arellano et al. 2007). Even though deletion of the PUA domain causes reduced activity and altered allosteric regulation in the AAK domain of ProB (Pérez-Arellano et al. 2005), the latter remains enzymatically active. Additionally, the interface which typically confers RNA-binding in the PUA domain is exposed, suggesting that it might be functional (Marco-Marín et al. 2007). Since putative RNA targets are still undiscovered, ProB was deemed ideal for selection experiments to gain hints on potential interaction partners.

- Citric acid cycle -

***ec*AcnB -** As outlined in introductory notes (see chapter 1.4 and Figure 1), aconitase is the prime example for RNA-moonlighting, regulating iron homeostasis in eukaryotes (Constable et al. 1992), and reportedly acting on its own mRNA in prokaryotes (Benjamin and Massé 2014). While partially considered as positive control, AcnB was also selected to potentially complement existing knowledge, especially since prokaryotic representatives are significantly less studied relative to eukaryotic aconitases.

***ec*Mdh -** Human malate dehydrogenase isoform MDH2 was reported to bind specifically to a conserved element in the 3' UTR of sodium voltage-gated channel alpha subunit 1 (SCN1A) mRNA, based on pulldown-assays, mobility shifts, and upregulation of respective gene expression in a MDH2-knockdown (Chen et al. 2017).

***ec*Icd** – For an eukaryotic homologue of the *E. coli* isocitrate dehydrogenase, the NAD^+-dependent mitochondrial IDH-1 from yeast, a potentially allosterically controlled interaction with conserved 5' UTR regions of mitochondrial RNAs was postulated by Anderson et al. (2002), conjecturing an *in vivo* regulatory interplay between IDH-1, mRNAs, isocitrate, and AMP.

- Oxidoreductases -

The GO term oxidoreductase shows up frequently in results of certain RNA interactome screenings, like for example in microarray screens in yeast (Scherrer et al. 2010). While 'oxidoreductase' often coincides with the sub-term 'dehydrogenase', which is already represented sufficiently amongst aforementioned candidates (GapA, Mdh, Icd), a more holistic representation of the enzyme class was desired. Hence, quinone oxidoreductase (***ec*QorA**), pyridoxamine 5'-phosphate oxidase (***ec*PdxH**), alkyl hydroperoxide reductase (***ec*AhpC**) and superoxide dismutase (***ec*SodA**) were included as candidates, all of which were significantly recovered in the TRAPP experiment (Shchepachev et al. 2019). *ec*QorA is structurally homologous to ζ-crystallin, an enzyme suspected to bind and stabilize mRNAs (Tang and Curthoys 2001).

- Uncharacterized proteins -

***ec*YbiB -** This structural homologue of anthranilate phosphoribosyltransferase, which binds to nucleic acids without detectable sequence specifity (Schneider et al. 2015), was added to the selection due to being a housekeeping protein with yet undiscovered function that has been examined previously in our working group (Schneider et al. 2015).

***ec*YggX -** a possible Fe^{2+}-trafficking protein without confirmed function, might act on aconitase regulation (Pomposiello et al. 2003). Even though a linkage to the moonlighting function of aconitase might be far-fetched and the protein might possibly be no metabolic enzyme, it was chosen in an attempt to unravel an unknown protein function using genomic SELEX.

4.2 Purification of Enzymes

All proteins mentioned in section 4.1 were to be expressed and purified for SELEX experiments and potential *in vitro* follow-up characterizations. Respective genes were derived from the ASKA library (Kitagawa et al. 2005), which provides a complete set of N-terminal histidine-tagged ORF clones for *E. coli* K12. The gene for the bacteriophage MS2 coat protein was ordered from GeneArt. Since nickel-affinity purification, employed for purification of all proteins, is reported to potentially cause trace Hfq impurification (Milojevic et al. 2013), which could majorly affect RNA selection experiments, all proteins were subjected to follow-up size exclusion chromatography irrespective of the purity grade observed after nickel-affinity chromatography. The purification of each protein is documented in supplements (section 7, Figure S 2 - Figure S 23),[5] including chromatography elution profiles, SDS-PAGE analysis of chromatography fractions, protein yields, and where applicable substrate turnover verified by suitable photometric activity assays (assays were conducted for all metabolic enzymes except RpiA, Upp, ThyA, and SodA). Detecting enzymatic activity can be considered a straightforward approach to diagnose an enzyme's native fold prior to subjecting them to further experiments. The reaction schemes of all conducted assays are schematized in methods section 3.4.9.

With the exception of AhpC, which was not found in significant amounts in the soluble fractions, all proteins of interest listed in Table 5 were successfully purified, with the following noteworthy remarks:

- MS2 coat protein was purified with C-terminal His-Tag as a double mutant V75E A81G. This mutant simplifies experiments involving the RNA-binding function of MS2, as it prevents capsid-multimer formation, while retaining affinity to the MS2 operator stemloop motif (LeCuyer et al. 1995).
- AnsB was purified with C-terminal His-Tag, as it is a periplasmic protein with an N-terminal signal peptide, which might be cleaved off during purification (Bonthron 1990).
- Repositioning of the His-Tag of ProB from N- to C-terminus was empirically found to significantly improve yield. This might be due to the N-terminal end being involved in the formation of the native tetramer (Marco-Marín et al. 2007).
- By default, Tris-HCl was used as buffer system for all proteins. The buffer system was changed to HEPES for PykF due to an inhibitory effect of Tris on pyruvate kinase activity (see Figure S 1e). Furthermore, KP-buffer was used for ProB, as it was empirically found to increase yield.

[5] Purification work was partially conducted as part of a Master's Thesis (Funke 2019). Respective instances are labeled in the supplements section.

4.3 iCLIP Experiments

Cross-linking and immunoprecipitation (CLIP) methods constitute a widely used, protein-centered approach for characterization of protein-RNA interactions (Hafner et al. 2021). The method utilizes UV irradiation in living cells to covalently capture RNAs in proximity of a protein of interest, which can then be isolated *via* immunoprecipitation. RNase digestion can be used to reduce the bound RNA to a trimmed down fragment that is covalently linked to the protein, narrowing down the site of protein-RNA interaction. In general, the implementation of next generation sequencing considerably enhanced the informative value CLIP approaches. Several experimental variants have been published. The variant iCLIP, employed in this work, takes advantage of reverse transcription aborting exactly at the first cross-link position, yielding aborted cDNAs with nucleotide-resolution information about the exact binding site (Huppertz et al. 2014). By immunoprecipitating potential RNA-enzyme complexes, the envisioned goal was to directly 'catch' the enzymes while exerting a moonlighting function *in vivo*. As opposed to *in vitro* methods like SELEX (also pursued in this work and subsequently described in section 4.4), the concern that specific conditions like presence or absence of substrates, cofactors, or heterologous protein oligomerization are required for observing a dual-enzyme function was deemed unproblematic for iCLIP experiments. Even if a moonlighting function of a protein would be sparsely populated relative to an enzymatic function, the sensitivity of PCR-amplification and NGS-sequencing would help identifying the former. The feasibility of catching enzymes exerting their moonlighting function *via* cross-linking and immunoprecipitation was demonstrated previously, for example in the case of aconitase (Cho et al. 2021).

The iCLIP method was employed in cooperation with Dr. Oliver Rossbach (University of Gießen), who owns superior expertise and therefore conducted experimental steps from lysis of UV-irradiated cells up until library sequencing. The immunoprecipitation experiments are described hereinafter with kind permission of Dr. Oliver Rossbach. It shall also be noted that only a subset of all selected RBP candidates was subjected to iCLIP, owing to stepwise addition of new enzyme candidates later on in this work, as well as cooperation work requiring a wider time frame. Candidates selected later on were only characterized by genomic SELEX.

For enabling immunoprecipitation experiments, tandem FLAG-Tags had to be introduced to the genes of interest. Genomic modifications were created by utilizing λ-red mediated homologous recombination (section 3.2.12), followed by P1 Phage transduction (section 3.2.13) to transfer the modifications into target strain *E. coli* BW25113. The position of the FLAG-Tags (C- or N-terminus, respectively) was adapted if Western blots indicated poor expression levels.[6] Western blots for all genes that were to be subjected to iCLIP studies are depicted in Figure S 25, testifying successful integration of FLAG-Tags. FLAG-Tag modifications were partially conducted as part of a Master's Thesis (Funke 2019).

[6] C-terminal FLAG: *ansB, astC, hfq, kdsA, mdh, pgk, rho, rpiA, speC, yggS*
N-terminal FLAG: *icd, pykF, upp*

4.3.1 Isolation of RNA-Protein Coprecipitates and Library Generation

Initial iCLIP experiments were conducted for RBP candidates Mdh, PykF, RpiA, KdsA, Upp, Pgk, and AnsB, as well as known RNA binders Rho (Roberts et al. 2008) and Hfq (Muffler et al. 1996) serving as positive controls, and supposed negative control YggS. The purpose of first experiments was to achieve a principal co-isolation of RNAs upon immunoprecipitation of RBP candidates. To this end, immunoprecipitates were subjected to gradient PAGE and subsequently visualized by Western blotting and autoradiography. Colocalization of bands in nitrocellulose membrane (protein) and phosphor screen (^{32}P-endlabeled RNA) served as indicator for the presence of cross-linked protein-RNA complexes. Respective results are shown in Figure 10. Blue arrows indicate the presence of protein bands in the overlying Western Blot-nitrocellulose membrane.

Both positive controls showed clear formation of protein-RNA complexes in Figure 10a. Rho (~49 kDa with tag) forms a distinct complex above 55 kDa. Hfq (~13 kDa with tag) showed multiple colocalized bands at a range of 15 to 100 kDa. Hfq is known to form hexamers, depending on RNA binding, Hfq concentration, and microenvironment (Panja and Woodson 2012). No quantitative disruption of oligomerization seemed to occur in the LDS-PAGE, possibly due to the Hfq subunits being tied together *via* cross-linked RNA. The 'no FLAG' control lysate expectedly did not produce any radioactive signal. Both Rho and Hfq samples lacked protein-

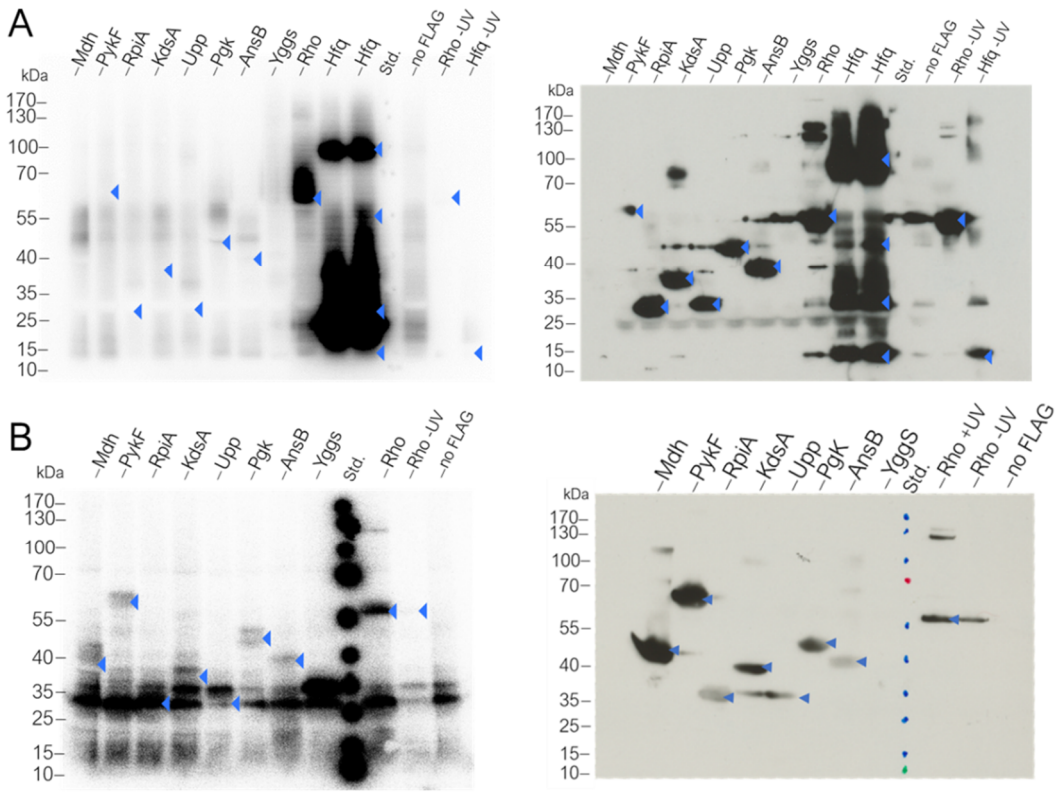


Figure 10: Initial attempts of protein-RNA complex immunoprecipitation. Autoradiograms (left) and Western blots (right) of initial iCLIP experiments with 10 µl (A) or 100 µl (B) lysate input. Blue arrows indicate the position of proteins as derived from Western blots. These particular results were previously published in the Master's thesis of Franziska Funke (2019).

RNA complexes when omitting the UV-irradiation step ('Rho -UV' and 'Hfq -UV'), confirming sufficiently stringent washing steps during immunoprecipitation. The weak radioactive signal around 55 kDa visible in all samples is attributed to unspecific interactions of the heavy antibody chain. No relevant radioactive signal was observed at heights of respective RBP candidate protein bands (blue arrows for Mdh, PykF, RpiA, KdsA, Upp, Pgk, and AnsB in Figure 10a), indicating that no detectable amount of RNA was cross-linked to those proteins. Supposed negative control YggS proved to be unsuited due to the lack of a protein band in the Western blot. Overall, it was demonstrated that none of the enzyme candidates was able to coprecipitate RNAs to an extent comparable to that of RNA-binding proteins Rho and Hfq, considering that the latter did not show superior band intensity (Figure S 25), and their RNA coprecipitation levels can thus not be attributed to higher expression levels in the cell.

Since it was already established in previous work by Franziska Funke (2019) that a subset of the RBP candidates hold unspecific RNA binding activity (namely PykF, Pgk, and KdsA, see Figure S 27), the experiment was repeated despite initially negative results, considering that CLIP experiments for bacteria are fairly uncharted and input conditions might not have been chosen appropriately for *E. coli* cells. When subjecting 100 µl instead of 10 µl of cell lysates to immunoprecipitation, weak radioactive signals were observed for Mdh, PykF, KdsA, Pgk, and AnsB, colocalizing with protein bands of the Western blot (see blue arrows within Figure 10b). No assessment could be made for RpiA and Upp, since protein height coincided with unspecific signal near 30 kDa present in all lanes that originated from the light antibody chain. Unspecific signal arising from the light chain as opposed to the heavy chain in Figure 10a is presumably explained by a more stringent RNase digest of lysates, producing smaller RNA fragments that preferably bind to the light chain (personal communication Oliver Rossbach). Positive control Rho expectedly showed clear coprecipitation of RNA.

While the detection of bands in the autoradiogram upon employment of increased lysate input levels seemed intriguing, the fact that these bands were detected for all analyzable proteins (i.e., for all loci not covered by the light antibody chain) simultaneously questioned the specific nature of RNA coprecipitates. The crowded composition of prokaryotic cytoplasm (Zimmerman and Trach 1991) could lead to low levels of random capturing of abundant RNAs upon UV-irradiation, which becomes increasingly visible when employing high enough amounts of cell lysate. It is also possible that radioactive signal is visible irrespective of RNA binding. For example, a coprecipitated kinase could utilize the ^{32}P-ATP to label protein residues.

To validate the presence of coprecipitated RNA, the cell lysate of KdsA strain was exemplarily treated with a series of RNase concentrations instead of one fixed concentration, testing whether band height would change with respect to RNase amount (Figure 11a). It could be observed that the alleged KdsA-RNA complex became slightly smaller at the highest RNA concentration, verifying the presence of RNA. Figure 11a also shows new negative controls ornithine decarboxylase (SpeC) and succinylornithine transaminase (AstC), employed after YggS proved to be a dysfunctional control. Both enzymes have no known relation to RNAs, do not use nucleotide cofactors and are not found to be RBP-candidates according to the OOPS-screen

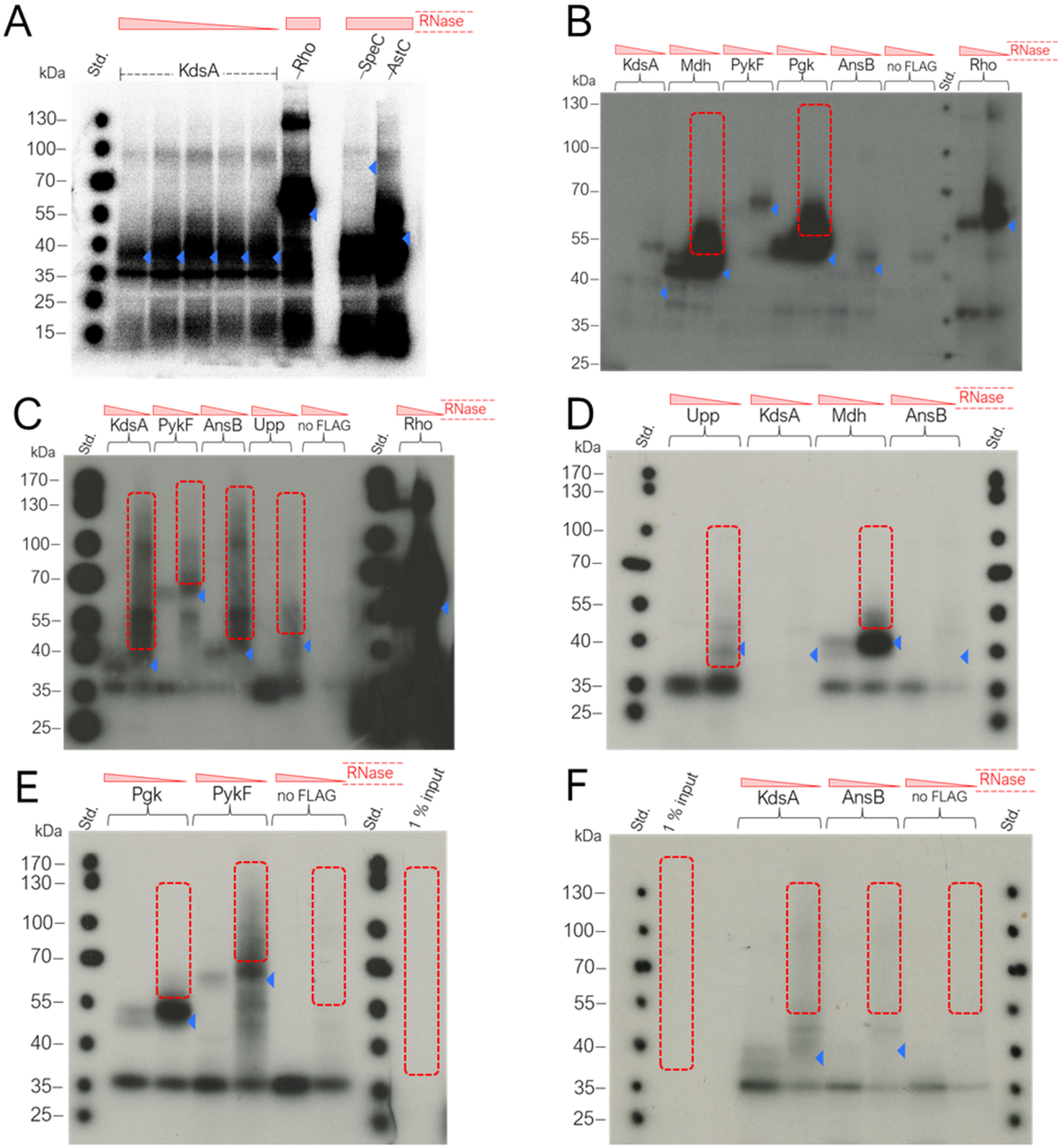


Figure 11: Autoradiograms of iCLIP cell lysates separated by 4-12 % gradient LDS-PAGE. Red bars indicate RNase digests of lysates prior to PAGE separation. Blue arrows indicate colocalized bands on Western blots of respective LDS-gels (Western blots not shown separately for conciseness). **A:** RNase test digestion for KdsA lysate, and new negative controls SpeC and AstC on the right. **B-F:** Preparative scale cell lysates for first (**B,C**) and second (**D,E,F**) CLIP library replicate. For all strains, RNA-containing gel fragments were excised as indicated by red rectangles.

(Queiroz et al. 2019). Considering the simultaneous occurrence of RNA bands for all proteins of interest, new negative controls SpeC and AstC should help examine whether 'any protein' could produce autoradiogram bands under employed conditions. Unfortunately, results turned out ambiguous (Figure 11a, right). AstC (~46 kDa) produced a strong radioactive signal. SpeC (~82 kDa) principally did not produce a radioactive signal at respective protein band height but showed a signal at lower than expected height, indicating potential coprecipitate of unknown nature. Overall, negative controls failed their purpose to produce empty lanes in the

autoradiogram, and even though RNA coprecipitation could be confirmed by serial RNase digest, negative controls suggest the amount of input that yields detectable RNA signals for RBP candidates seems to be at a level already prone to unspecific effects.

Despite these ambiguous results and considering the principally successful verification of protein-RNA complexes in aforementioned immunoprecipitations, the full iCLIP protocol was subsequently carried out for AnsB, KdsA, Mdh, Pgk, PykF, and Upp strains in preparative scale to create a library suited for next generation sequencing and assess the nature of RNA coprecipitates. To that end, RNA-containing gel fragments were excised according to red rectangles depicted in Figure 11b-c. Later on, the iCLIP experiment was repeated as replicate (respective excision patterns depicted in Figure 11d-f). RNA isolation *via* gel excision was split among multiple gels because immunoprecipitation had to be repeated for some strains using even higher amounts of cell lysate, in order to obtain detectable amounts of RNA (for example, see low signal of KdsA compared to Mdh in Figure 11b). The diverging radioactive signal in between samples probably indicated differences in expression levels rather than variable RNA-binding affinities – for example Mdh and Pgk, featured the strongest RNA bands and concordantly showed the highest expression (see Figure 11b and Western Blots in Figure S 25). After their isolation, the RNAs were reverse-transcribed and transformed into a multiplexed library according to procedure described in section 3.4.10.

4.3.2 Sequencing Results of iCLIP Samples

Due to retention of the crosslinking site information directly adjacent to the P5Solexa adapter, a sequencing kit with 50 cycles (implicating a maximum read length of 50 nucleotides) was employed for sequencing of the iCLIP libraries. A total number of 7,932,167 reads for the first and 8,277,069 reads for the second replicate library could be mapped to the genome of *E. coli* K12 substr. MG1655, respectively. The resulting genome mappings were to be screened for clusters of closely aligned read termini, indicating recurring crosslinking events between the protein of interest and the transcript of the respective genomic locus. To this end, mappings of samples were compared to mappings of two replicate background controls referred to as size-matched input (SMI). SMI background controls were prepared by directly separating the cell lysate on a gel, omitting the immunoprecipitation step (see method scheme in Figure 6, '1% input') and should help distinguishing sample-specific RNA enrichment from enrichment that occurs unconditionally based on abundancy of certain RNAs *in vivo*. Autoradiogram and gel excision of the SMI lysates is depicted in Figure 11e-f ('1 % input').

Figure 12 shows the number of clusters in SMI samples identified by analysis with `PureCLIP`, a hidden-Markov model based approach to crosslink site detection (Krakau et al. 2017). *E. coli* possesses a total of 86 transfer RNA genes and 22 ribosomal RNA genes (Berlyn 1998), and read clustering was identified for essentially all of the tRNA and rRNA transcripts present in *E. coli*, but only for a fraction of thousands of protein-coding and non-coding RNAs. Instances of tRNAs or rRNAs that were called as cluster in only one of the two SMI replicates were manually

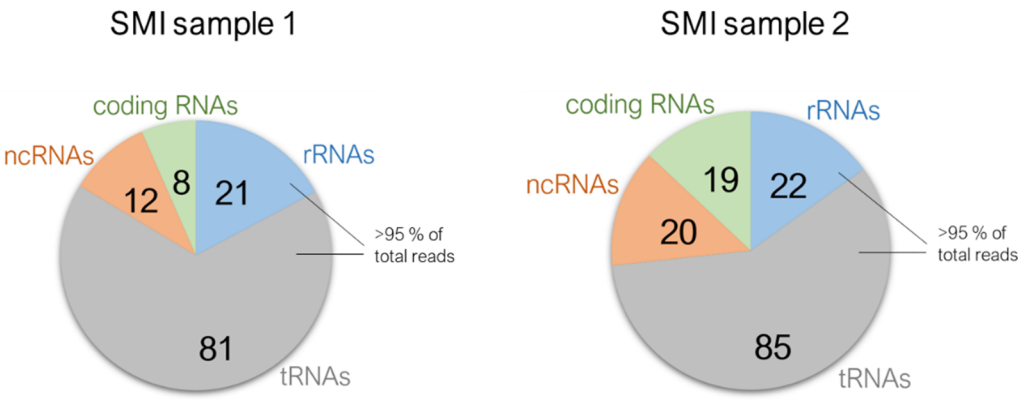


Figure 12: Read clusters identified by PureCLIP for iCLIP controls SMI 1 and SMI 2. Clusters are sorted by RNA type: non-coding, protein-coding, ribosomal, and transfer RNAs.

inspected in the mapping, and respective loci featured read numbers of similar scale, indicating that `PureCLIP` omitted them in one replicate due to technical limitations. Transfer and ribosomal RNA clusters consistently showed absolute read numbers one to two orders of magnitude above those of protein-coding and non-coding RNAs, and the former RNA types accounted for >95 % of total mapped reads. tRNAs and rRNAs make up the majority of RNA molecules *in vivo* (Li and Deutscher 2008), and it is the expected result for the SMI control samples that these abundant species are comprising the majority of reads. The remainder of the clusters found at protein-coding or non-coding regions partially shared loci in between the two replicates, coinciding with abundant RNAs like the 4.5S signal recognition particle ffs (Griffin 1975), 4.5S RNA (Wassarman and Storz 2000) or lipoprotein mRNA lpp, coding for the most abundant protein in *E. coli* (Li et al. 2014). Taken together, the distribution of reads across tRNAs, rRNAs, coding RNAs and non-coding RNAs was consistent in between SMI replicates, and it seemed to resemble a brief footprint of general transcript abundancy.

Inspection of mapped libraries of RBP candidates revealed a trend similar to that of SMI samples. Transfer and ribosomal RNA clusters were proportionally overrepresented, as seen from the number of clusters identified by `PureCLIP` sorted by RNA type (Table 6). For almost all RBP samples, a significant proportion out of the existing 86 tRNAs and 22 rRNAs in *E. coli* had clusters designated to them. Moreover, analysis of absolute read number distribution across RNA types revealed that just like for the SMI samples, tRNAs and rRNAs make up >95 % of total mapped reads, while coding RNAs and ncRNAs account only for a maximum of a few percent. The identified clusters among coding and ncRNAs were coinciding with abundant transcripts like for example the lipoprotein mRNA lpp, and almost entirely co-occurred in at least two libraries from different proteins, indicating that enrichment was based on general RNA abundancy rather than specific protein-selection. The few peaks only called in one experiment were manually assessed by comparing respective gene loci across experiments, and it was apparent that these loci were occupied by reads in multiple experiments as well, meaning that they failed recognition by `PureCLIP` merely due to technical limitations.

Since functional targets could potentially be found among tRNAs and rRNAs as well, respective enrichments were compared across RBP candidate experiments. However, read numbers of tRNA

and rRNA clusters showed no variation above noise level, implying that under employed experimental conditions, these abundant molecules have a similar crosslink rate for all RBP candidates. The majority of reads being attributed to tRNAs and rRNAs is not unexpected since CLIP read distribution generally depends not only on the specificity of a binding site, but also on the abundance of the underlying transcript (Busch et al. 2020). However, examples of strong RNA binders from both eukaryotes and prokaryotes breaking this pattern are available in literature. For example, CLIP studies of NF90 in human HEK-293 cells yielded 85 % of signals located at certain mRNAs (Lodde et al. 2022). A CLIP experiment with suspected 3'UTR-binding protein MSI2 in K562 cells revealed an indorsing read cluster enrichment at 3' UTR regions, comprising 50 % of total clusters (Karmakar et al. 2022). For prokaryotes, CLIP experiments have been conducted for well characterized RNA-binding proteins like Hfq, CsrA, or ProQ. Hfq CLIP experiments in *Salmonella typhimurium* showed that the majority of peaks mapped to sRNAs and UTRs of mRNAs, the two RNA classes known to be targets of Hfq, and yielded defined sequence motifs derived from these peaks (Holmqvist et al. 2016). CsrA of *Escherichia coli* and *Salmonella typhimurium* was shown to consistently enrich mRNAs with a defined sequence motif, which comprised the majority of hits in respective peak calling evaluation (Holmqvist et al. 2016; Potts et al. 2017). ProQ CLIP experiments in *Salmonella enterica* yielded a significantly high peak occurrence within 3' UTR regions of mRNAs (Holmqvist et al. 2018). Putting peak generation of such well-established RNA-binding proteins into perspective to own results, the `PureCLIP` analysis pointed to the conclusion that unambiguous, RNA subtype-dependent affinity discrepancies are not to be found within tested RBP candidates.

Table 6: Number of clusters identified by PureCLIP, sorted by RNA type.

RBP candidate	replicate	rRNA	tRNA	protein-coding RNA	ncRNA	
Mdh	1	9	29	9	5	
	2	16	78	6	5	
Pgk	1	11	38	22	17	
	2	15	51	1	0	*
PykF	1	12	36	4	7	
	2	22	77	17	11	
Upp	1	11	31	1	4	
	2	22	78	5	5	
AnsB	1	12	40	6	8	
	2	0	67	14	6	
KdsA	1	12	31	10	9	
	2	0	63	27	8	
		>95% of total mapped reads in all samples		<5 % of total mapped reads in all samples		

*Pgk replicate 2 sequencing was accompanied by technical difficulties, producing low read count and avg. size ~20 nt reads instead of target ~45 nt

While `PureCLIP` was useful for a time-efficient, initial evaluation of the large data set, manual inspection of the entirety of the mapping could also be conducted with reasonable thoroughness owing to the small size of *E. coli* genome (~4,600,000 nt). Hence, to exhaustively interpret the iCLIP data set and look for unique clusters with small read numbers that might have been omitted by `PureCLIP`, the data was reviewed in the Integrative Genomics Viewer, comparing mapping traces of samples, and looking for unique clusters that did not occur unitary across samples.

For replicate 1 of the Mdh sample, around 30 unique clusters could be identified, comprised of <40 reads each. Even though these read numbers were low compared to the total number of reads in that replicate (approximately one million), it was attempted to verify these interactions with respective transcript RNAs and also with a combinatory sequence motif by EMSA analysis. Unfortunately, Mdh did not display specific binding (data not shown), implying that this range of read numbers can already be considered as noise level. Unique clusters of samples other than Mdh featured even smaller numbers of reads, and no further experiments were conducted.

Beyond aforementioned candidates, no further enzymes were subjected to CLIP-Seq, and characterization of subsequently selected enzymes was focused exclusively on genomic SELEX experiments described in following chapters. While the general reflection of universal transcript abundancy and lack of unique read clustering within tested libraries was disappointing, it was not interpreted as ineptness of the method in probing these enzymes. Rather, tested candidates might not feature specific affinity towards any genomic RNAs, and could have simply delivered the appropriate result. This conclusion is *a fortiori* plausible as the examined enzymes produced similar libraries despite featuring diverging K_D-values when it comes to unspecific RNA binding (as is judged from titrations with random RNA fragments; $K_D^{PykF} = 0.6$ µM, $K_D^{Pgk} = 42$ µM, $K_D^{KdsA} = 0.7$ µM, see Figure S 27). The subsequent focus on genomic SELEX was mainly owed to a quicker experimental pipeline and result attainment. However, it cannot be completely ruled out that CLIP-Seq required further optimization with respect to cell growth and crosslinking conditions. Sequencing of a positive control like Hfq or Rho should be conducted if CLIP experiments are to be continued.

4.4 Validation of SELEX Functionality

SELEX, a method originally developed for characterization of small molecule-aptamer interactions (Ellington and Szostak 1990), was employed in this work for screening of RBP candidates against an RNA library comprised of *E. coli* genomic transcripts. SELEX protocols were to be established in the home department, and necessary knowledge was generously shared by the Süß group during a research visit at the Technical University of Darmstadt. Before subjecting proteins of interest to genomic SELEX, it was necessary to verify the successful establishment of the method setup by conducting suitable control experiments.

In this context, egg white lysozyme was employed as first control. The protein should be able to enrich specific aptamers out of a randomized RNA aptamer library (Cox and Ellington 2001).

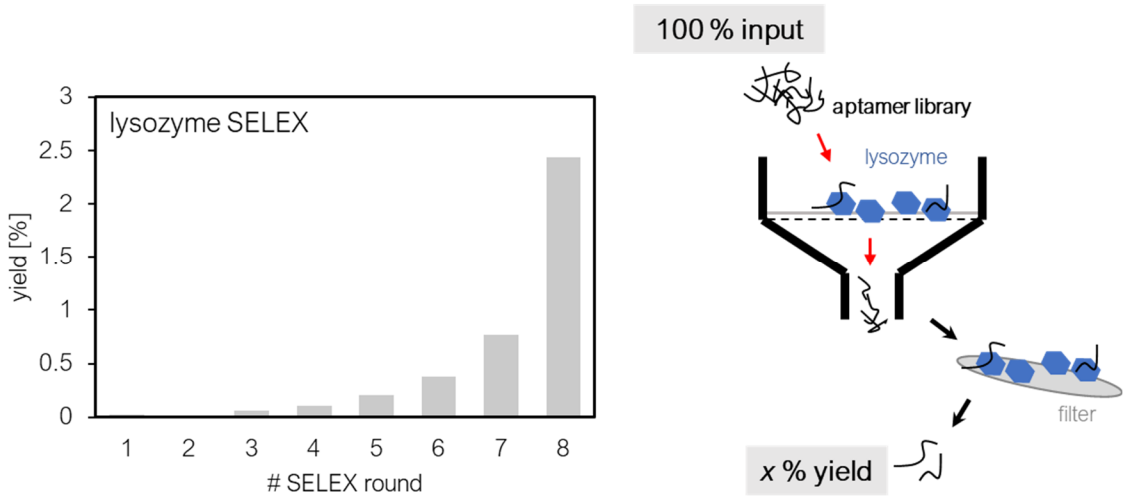


Figure 13: RNA retention during lysozyme SELEX. The yield of RNA isolated from the filter is plotted for each SELEX round. Graphical abstract of the filtering step on the right.

The library employed for lysozyme SELEX is schematically shown in Figure 5 (methods section). It consists of a primer-flanked, fixed length region that either contains a completely randomized 64 nt sequence or a central 8 nt stem loop structure flanked by two 26 nt randomized sequences. The stem loop is 'mixed' into the random sequence as it can function as a structural anchor for recognition loops and is reportedly able to increase the success rate of aptamer selection (Sassanfar and Szostak 1993; Davis and Szostak 2002). Lysozyme was subjected to the aptamer library in eight SELEX rounds, and the proportional amount of RNA retained by the lysozyme during each filtering step was tracked through radioactivity counting (see method section 3.4.11). Retained RNA expectedly showed exponential growth, with a more than 100-fold increase between the first and eight round of selection (Figure 13). Since the protein was exposed to the aptamer library under identical conditions within each round, the growth in RNA retention implicates a rapid evolution of the library towards affine aptamers. This result was in agreement with literature, where aptamers have been successfully selected and amplified by hen egg lysozyme from a randomized 30 nt RNA pool (Cox and Ellington 2001), and confirmed the correct application of the SELEX setup, paving the way for the exploration of novel potential RNA binding proteins. Follow-up sequencing of enriched aptamers was omitted, as lysozyme was no protein of interest and subsequent SELEX efforts utilized a genomic library rather than an aptamer library.

4.5 Adoption of Genomic SELEX

Due to this work being aimed at the exploration of functional partnerships between enzymes and RNA species which occur in the natural *in vivo* environment of *E. coli*, the randomized RNA library utilized for the lysozyme control experiment was only of limited suitability. While a randomized pool typically features around 10^{15} sequences (Gold 1995) and can contain coincidental genome fragments that occur *in vivo*, it cannot provide complete coverage of the genome. Furthermore, aptamers cover a huge structural landscape and can be viewed as

nucleotide analogues of antibodies, suggesting that there is a reasonable chance to find a high affinity aptamer for any given protein (Conrad et al. 1996; Lakhin et al. 2013). Hence, a randomized library would be prone to contain random high affinity aptamers with no biological relevance, competing with an actual, physiological RNA target. The initial low copy number of the biological relevant target in the randomized library would further impede the success chances of the selection process. Hence, genomic SELEX, which utilizes a transcribed library derived from the actual genome of interest (Singer et al. 1997), was deemed a more suitable approach to unravel biological relevant interactions with proteins.

A library derived from the *E. coli* genome was used for all subsequent SELEX experiments. The library was obtained commercially from Microsynth AG, who prepared it by genome shearing and adapter ligation (described in method section 3.2.14). Due to the relatively small size of the *E. coli* genome (4.64 million nucleotides), the library easily covers the sequence space of the genome. PCR amplification, an intermediary step iteratively applied during SELEX for replenishment of the library copy number, was found to cause significant adapter dimer formation. These dimers are a result of self-ligation of adapters without a library insert sequence and constitute a common by-product of NGS library preparations (Head et al. 2014), which can potentially disrupt the selection process and impede quality and quantity of reads during the final next-generation sequencing step. To contain the proportion of adapter dimer throughout experiments, size-based gel excision and commercial clean-up solutions (like the GeneJet NGS Cleanup Kit from Thermo Fisher Scientific) were used in regular intervals when conducting SELEX rounds (Figure 14).

Even though PCR amplification was conducted with conservative cycle numbers to produce just enough DNA copies enabling the transcription reaction of the subsequent SELEX round, PCR is inherently introducing selection bias proportional to the number of cycles conducted (Dabney and Meyer 2012). Therefore, PCR impact on SELEX was examined by performing another control experiment, completely omitting protein selection and filtering, and instead exposing the library directly to repeated rounds of re-amplification, with total cycle number comparable to that of a 'real experiment'. Next generation sequencing of the resulting library yielded 91,984 reads, of which 90,988 reads (98.9 %) could be aligned to the genome. The GC distribution of

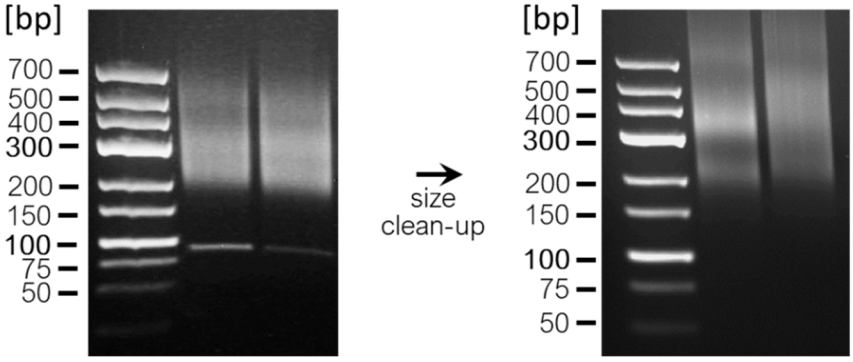


Figure 14: Genomic library before and after size clean-up on agarose gel. The library with genomic insert is located at 200-700 bp, faulty adapter dimers visible at 90 bp.

mapped reads (Figure 15a) was found to be closely centered to the actual GC content of *E. coli* (50.8 %), ruling out critical levels of amplification GC bias. As opposed to an RNA-seq library where copy number reflects transcript abundance and identical reads are to be expected, the fragments in the genomic library, originating from random genome shearing, are expected to be nearly 100 % unique. Figure 15b shows that more than 93 % of reads (blue line) are aligned exactly one time to the genome. The remaining fraction is aligned multiple times due to either PCR duplicates or genomic sections like the ribosomal RNA operon which occur multiple times in the *E. coli* genome. Strikingly, reads with three or more copy numbers are neglectable sparse, and the elimination of PCR duplicates resulted in a loss of merely 2.84 % of total aligned reads (red line). A representative excerpt of the resulting genome mapping (two million nt) is depicted in Figure 15c, showing that the genome is still covered with reasonably uniform density after PCR treatment. Slight variation in coverage is observable, which might be attributed to PCR, but also to inconsistencies during the original DNA shearing process when generating the library. The variation should be outweighed by the strong selection pressure exerted when an employed protein of interest indeed was to have specific RNA binding properties, as can be seen from literature data where more than half of single clone isolates featured a consensus sequence after genomic SELEX (Shtatland et al. 2000). Employed next generation sequencing kits allowed for a maximum of roughly 100,000 reads per sample (a limitation imposed by pricing). Assuming an average read length of 150 nt and uniform distribution of reads, the *E. coli* genome would be covered by an average factor of 3.3. This was deemed sufficient for mapping analysis considering

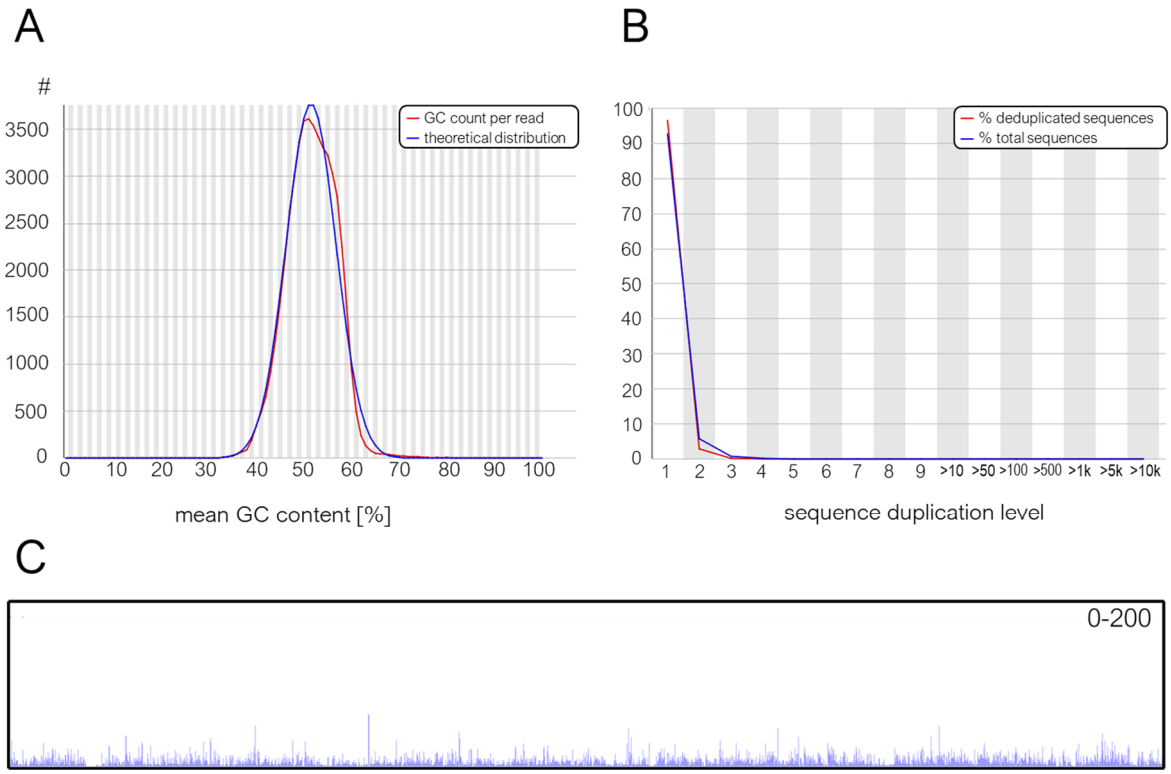


Figure 15: Genomic library composition after repeated rounds of PCR re-amplification. A: FastQC analysis of GC-distribution. **B:** FastQC analysis of read duplication – sequence duplication level on the X-axis refers to the number of positions in the genome leading to a correct alignment for a particular sequence. **C:** Mapping excerpt (two million nt) visualizing read distribution. Y-axis represents read numbers from 0 to 200 in linear scale. NC_000913:1000000-3000000.

that libraries repeatedly exposed to specific RNA-binders would feature sequences enriched by orders of magnitude. It shall however be noted that the capped read number is obviously too low to mirror the actual coverage as resulted from genome shearing.

4.6 Validation of Genomic SELEX with Bacteriophage Coat Protein MS2

The coat protein of bacteriophage *Emesvirus zinderi* (MS2) is a well characterized model RBP for RNA recognition. The bacteriophage utilizes a specific stem loop within its own messenger RNA for two key functions (Bernardi and Spahr 1972). First, the protein MS2 acts as a translational repressor at high enough concentrations, blocking the replicase initiation codon located at the stem loop site. Secondly, the interaction is necessary to accomplish encapsulation. While the primary target for encapsulation is the viral genome, the *E. coli* genome also carries sequence stretches that resemble the consensus MS2 recognition stem loop, and a study could confirm the integration of heterologous RNA into virus particles (Pickett and Peabody 1993). While the biological significance of heterologous interactions between MS2 and host RNAs remains unclear, they have been tackled by a study conducting genomic SELEX experiments with MS2 and a *E. coli* genomic library (Shtatland et al. 2000). These experiments allowed to isolate specific sequences that carry the stem loop consensus sequence, among them mRNA fragments from *rffG* (outer membrane lipopolysaccharide synthesis) and *ebgR* (lactose utilization repressor).

On behalf of identifying moonlighting protein-RNA interactions for metabolic enzymes, it was necessary to first ascertain the principal ability of the SELEX method to isolate acknowledged targets from the *E. coli* library. To that end, coat protein MS2 was selected as a robust positive control due to the readily available literature data. Next to the purpose of method validation, key differences to the genomic SELEX experiment previously conducted by Shtatland et al. (2000) rendered the reproduction of MS2 SELEX interesting. These differences include utilization of next generation sequencing as opposed to single clone sequencing and diverging methods of library creation (genome shearing versus random priming), potentially facilitating more elaborate data sets. MS2 was subjected to six selection cycles, and sequencing of the evolved library yielded a total of 75,455 reads that could be mapped to the *E. coli* genome. Comparison of the resulting mapping to that of the PCR-exposed control library described in section 4.5 shows that the general read density in the MS2-evolved library is reduced in favor of conglomerated clusters with high read numbers, e.g. at the *dacB* gene locus (Figure 16). The read cluster at the *dacB* locus in particular comprised the highest coverage in the mapping, featuring 29,418 reads over a sequence span of 350 nucleotides (with a total of 75,455 sequenced reads, a 350 nt region is expected to contain no more than 2-3 reads when assuming a theoretical, uniform read distribution).

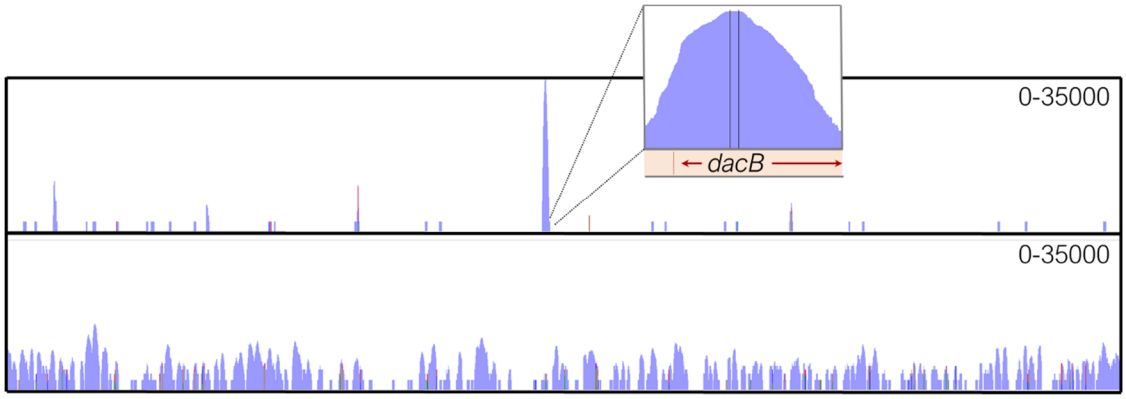


Figure 16: Genome mappings of MS2-exposed and control library. Top – Genomic library evolved by exposure to MS2 coat protein. Inlet: close-up view of the *dacB* locus. Bottom – control library evolved only by repeated PCR without any protein-binding selection. Y-axis represents read numbers from 0 to 35,000 in log-scale.

In order to survey the allocation of all loci that featured significant read conglomeration for the MS2-treated library, the genome mapping was screened for clusters defined as ≤ 400 nt regions with more than 20 reads. Table 7 lists genes located at thusly identified genome regions, as well as respective *pro rata* read numbers. Respective proportional numbers of mapped and deduplicated reads of each locus showed loose correlation (average aberration factor 5.6), indicating no critical levels of PCR-induced bias. Identified sequences coincided with five of the eight sequences found by Shtatland et al. (2000) through single clone sequencing, suggesting strong correlation between own and reported data. These five matches are either *sense* or *antisense* transcripts of gene ORFs *rffG, elfG, mreD, ygjI,* and *ebgR*. Remarkably, the sequence at the *rffG* locus, comprising 14.5 % of mapped reads, was reported by Shtatland et al. to be identified in 98 out of a total of 173 sequenced isolates. This loose correlation of coverage affirmed consistency between experiments. Based on the functions of aforementioned genes[7] Shtatland et al. further hypothesized that MS2 selects heterologous RNAs related to cell surface function. Since next generation sequencing provided superior data depth, more RNAs related to cell surface function could be identified, comprising a highly significant proportion of hits listed in Table 7. This includes all top hits - *dacB* (peptidoglycan synthesis), *tesA* (fatty acid metabolism), *sslE* (biofilm maturation), *fliO* (flagellar protein), *alx* (putative membrane-bound redox modulator), and *ecpC* (pilus formation, biofilm).

[7] *rffG* – outer membrane biogenesis; *elfG* – putative adhesion-related function; *mreD* – putative cell shape-related function; *ygjI* – inner membrane transporter

Table 7: Read clusters identified in the MS2-evolved genomic library.

gene locus	sense/antisense[a]	mapped reads[b]	deduplicated reads[b]
dacB	as	38.99 %	3.40 %
rffG	s	14.45 %	2.76 %
tesA	s	9.49 %	2.71 %
sslE	as	8.28 %	2.69 %
fliO	as	6.65 %	2.31 %
alx	as	4.52 %	2.41 %
ecpC	s	4.45 %	0.71 %
elfG	as	0.83 %	1.40 %
mnmA	as	0.74 %	0.67 %
yhiN	s	0.70 %	1.87 %
mreD	as	0.64 %	1.90 %
yeJL	as	0.41 %	1.36 %
damX	as	0.41 %	0.97 %
entF	as	0.36 %	0.67 %
ygjI	as	0.34 %	1.29 %
mdtC	as	0.31 %	0.82 %
bcsF	as	0.31 %	0.41 %
pppA	as	0.26 %	1.36 %
rsmF	as	0.26 %	0.52 %
gshA	as	0.25 %	0.54 %
oppA	s	0.24 %	1.14 %
yjgR	as	0.23 %	0.45 %
murF	s	0.22 %	1.10 %
glnE	as	0.20 %	1.21 %
torS	as	0.14 %	0.97 %
yqeB	as	0.13 %	0.73 %
xanP	as	0.12 %	0.73 %
torZ	as	0.12 %	0.37 %
ybdH	as	0.11 %	0.71 %
ebgR	s	0.11 %	0.99 %
yeeJ	s	0.10 %	0.86 %
proA	as	0.09 %	0.41 %
ybjT	as	0.09 %	0.60 %
panF	as	0.08 %	0.37 %
tynA	as	0.07 %	0.39 %
emrD	as	0.07 %	0.62 %
nuoK	as	0.06 %	0.54 %
fryC	as	0.06 %	0.34 %
hrpB	as	0.06 %	0.45 %
nuoN	as	0.05 %	0.39 %
bglX	as	0.05 %	0.41 %
trpD	s	0.05 %	0.52 %
yagG	s	0.04 %	0.43 %
yhfG/fic	s	0.04 %	0.45 %
ribD	as	0.03 %	0.32 %
topA	as	0.03 %	0.32 %
yqeC	s	0.03 %	0.30 %

[a] This indicator refers to the RNA orientation with respect to the gene, irrespective of the gene being located at + or – strand.
[b] Read numbers listed as percentage of global number of mapped or deduplicated reads, respectively.

Underlying genomic sequences at positions of mapped read clusters listed in Table 7 were used as input for a MEME motif search (Bailey and Elkan 1994). The algorithm calculated a motif with an E-value of 2.1×10^{-49} (Figure 17a).[8] The E-value could be improved to 2.1×10^{-69} when including additional gene loci that did not fulfil quality criteria due to insufficient mapped or deduplicated reads but were subjectively assessed as significant read clusters (listed in Table S 1). The identified motif matches the established minimal consensus motif for MS2 recognition, constituting a stem with a pronged adenosine and a four base loop ANYA (Romaniuk et al. 1987). The underlying sequences of genomic loci occupied by the top eight read clusters, jointly accounting for 87 % of mapped reads, are aligned in Figure 17b. Interestingly, the loop consistently displays a cytosine at the -5 position. This position was found to be critical for MS2 affinity ($K_A \sim 3 \times 10^8$ M^{-1} as determined by Romaniuk et al.). While the native bacteriophage sequence maintains an uracil at -5 (see Figure 17c), replacement with a purine results in a 10- to 100-fold decrease in K_A (Uhlenbeck et al. 1983), and replacement with cytosine induces a 50-fold increase in affinity *in vitro* (Lowary and Uhlenbeck 1987). The MS2 SELEX experiment stringently selected the highest affinity loop sequence from the genomic pool. It is unclear why the bacteriophage genome conserves the uracil over the higher affinity inducing cytosine. One hypothesis suggests that overly high affinity can cause super-repression, preventing necessary levels of replicase translation (Johansson et al. 1997).

To conclude the MS2 control experiment, the interaction between MS2 and a *rrfG* mRNA fragment carrying the recognition motif was tested in EMSA analysis (Figure 17d). The observed interaction was not compromised by a 100-fold excess of unlabeled competitor RNA. Overall, the control experiment confirmed that the established SELEX method can successfully isolate a both sequence specific and structural motif when employing an RNA-binding protein. MS2 selectively enriching stem-loops with a cytosine but not an uracil at the -5 position highlights that the unconditional isolation of the highest affinity aptamer has the potential to obscure, but also detect novel biological functions. While the native operator hairpin sequence was not enriched in the library, factors like high a copy number of phage mRNAs in the cell might alleviate the difference in affinity. It is not unreasonable that the unraveled high affinity interactions between MS2 and host mRNAs or asRNAs carry biological functions, and the unambiguous overrepresentation of cell surface-related genes in the MS2-evolved library, a trend already observed by Shtatland et al., further reinforces this assumption. The bacteriophage might benefit from altered cell adhesion properties, which are pivotal for viral propagation.

[8] The E-value is an estimate for the expected number of motifs with equal width, side count, and at least equal log-likelihood ratio that would be found in a similary sized set of completely randomized sequences (Bailey and Elkan 1994).

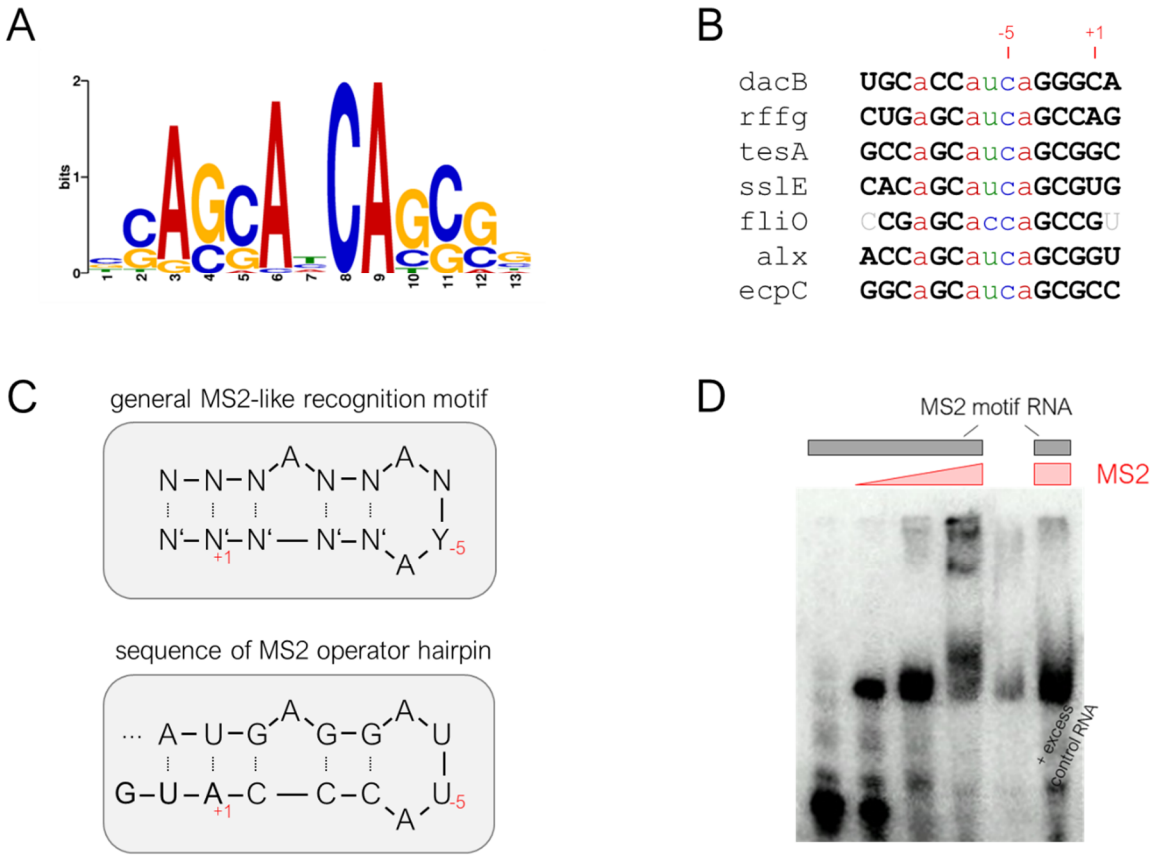


Figure 17: Motif characterization of the specific MS2-RNA interaction. A: Sequence motif generated by MEME based on clustered reads. B: Sequences constituting the stem-loop motif found at the eight gene loci that jointly account for 87 % of mapped reads. +1 indicates the position of start-codon in the similar, native bacteriophage hairpin. C: Generalized consensus motif for MS2 recognition (Romaniuk et al. 1987) and sequence of the native bacteriophage hairpin. D: EMSA showing the interaction of MS2 (0.34, 3.4, and 34 µM) with a ^{32}P-labeled *rffG* mRNA fragment (700 nM). In the last lane, competition with 70 µM unlabeled control RNA. Sequence of *rffG* mRNA fragment: NC_000913:3972719-3972754 (GenBank entry). Control RNA: GGGUU CUAGA GAGGU GAGCU UGGCA ACCUC UGAUG UAGGU.

4.7 Results of Genomic SELEX Applied to RBP Candidates

After successfully establishing genomic SELEX and confirming its principal ability to enrich biological relevant sequence targets from a genomic *E. coli* transcript library, all selected enzyme candidates (section 4.1) were employed in the former in search of specific RNA interactions. Hereinafter, SELEX results of all enzymes with notable RNA are described in separate subchapters. For the principal guidelines applied to evaluate mapping data the reader is referred to section 3.4.13 of methods.

4.7.1 Genomic SELEX with Pyruvate Kinase (PykF)

E. coli pyruvate kinase was considered a feasible candidate not only due to its occurrence in the datasets from OOPS and TRAPP interactome screenings but also due to detection of substantial unspecific RNA binding activity in previous work (K_D = 0.6 µM, see Figure S 27). Regarding an unspecific base affinity as an indicator for potential biological relevant interactions is supported by known examples of initially considered nonspecific binders found to display affinity variations

by orders of magnitude, like the *E. coli* C5 protein (Guenther et al. 2013). Three independent SELEX experiments were conducted for PykF, each consisting of multiple selection rounds starting from the same genomic library. PykF was the only enzyme for which multiple SELEX replicates had been created in the course of this work, a result of the enzyme being preliminary deployed for the purpose of method acquirement in the Technical University of Darmstadt.

Seven, four, and eight selection rounds were performed, respectively. Sequencing and data processing of evolved libraries yielded 30,649, 145,221, and 34,904 mapped reads and 1,121, 136,151, and 933 deduplicated reads[9]. Since read numbers in between replicates were unequal, clusters featuring significant read conglomeration were defined in relative fashion and should contain a minimum of both 0.1 % of total mapped reads as well as 0.5 % of total deduplicated reads in a ≤ 400 nt region. The mappings of all replicates were screened and resulting hits are summarized in Table 8. A highly significant proportion of the clusters were co-occurring in between two replicates (see color-coding in Table 8), considering the total number of 4,401 genes in *E.coli*. Clusters that did not fulfil the stringent read number constraint but were subjectively assessed as relevant are listed as grey entries.

Despite co-occurring clusters in between replicates, a motif query in MEME using the entirety of underlying genomic sequences as input yielded no statistically significant sequence motif for PykF. Selecting specific subsets of reads as input for MEME (e.g. considering only the clusters with highest read numbers, or exclusively clusters that occurred in two experiments simultaneously) did not substantially improve the output. Secondary structure prediction (RNAfold, Gruber et al. 2008) was subsequently used to investigate whether structural features drove the enrichment of identified clusters. While a stringent structural motif was not discernable, central underlying sequences of most clusters seemed to comprise long stem-loops, albeit with no close conservation (Figure 18a). The stems are of varying lengths and feature one or more bulge nucleotides. The loops are also heterogeneous in both length and sequence. For a functional characterization of enriched gene loci, a GO-term analysis of all respective gene products was conducted with PANTHER. While this analysis should be surveyed with care considering the small input query of 41 genes, it revealed no overrepresentation functional classes and resembled the distribution of the entirety of *E. coli* genes (see Figure S 28). The isolation of *antisense* transcripts of ribosomal genes *rrl*[*A-H*] seems at least noteworthy considering the alleged ability of ribosomes to slightly enhance the activity of PykF (Yu et al. 2021).[10] However, a clear connection between these findings is elusive considering that SELEX enriched only the *antisense* sequence of the 23S rRNA.

[9] The deduplicated read number was highly variant. While not ideal, this number is hard to control throughout SELEX experiments. Factors in play are, *inter alia,* 'thinning out' of the library during filtering steps, number of SELEX rounds, number of PCR cycles necessary to maintain the experiment (tied to extra cycles necessary after adapter dimer removal steps), and final steps to prepare the NGS-ready library.

[10] In this context, it is worth mentioning that PykF was assayed for altered catalytic activity in presence of RNA with a random sequence. No significant impact on k_{cat} or K_M was observable. Respective results are described in the Bachelor's Thesis of Hannah Osterholz (2019).

Table 8: Unique read clusters identified in the PykF-exposed genomic libraries.[a]

replicate 1				replicate 2				replicate 3			
gene	s/as	map [%]	ded [%]	gene	s/as	map [%]	ded [%]	gene	s/as	map [%]	ded [%]
holB	s	33.21	5.80	*rrlA-H*	as	2.01	0.17	*fadE*	as	85.26	0.54
sdhE	s	11.27	0.18	*clpB*	s	0.36	0.01	*sslE*	as	4.35	0.96
fadE	as	10.40	0.09	*yjgR*	s	0.10	0.03	*livF*	s	1.22	0.11
fadB	s	8.27	1.43	*napG*	s	0.08	0.03	*ytfF*	s	0.97	3.00
surA	as	6.07	2.59	*ubiH*	as	0.05	0.01	*surA*	as	0.72	0.11
yedK	as	5.03	2.50	*yjbH*	as	0.03	0.02	*yhcB*	as	0.59	0.21
ydjN	s	3.94	1.07	*yhcB*	as	0.02	0.01	*rrlA-H*	as	0.46	6.54
yjbH	as	2.50	6.69	*ygjH*	s	0.02	0.02	*sad*	s	0.42	1.07
citF	as	1.99	0.18					*cheB*	as	0.38	1.18
sad	s	1.74	2.41					*ydjN*	s	0.31	0.64
rnd	s	1.58	5.17					*ubiH*	as	0.18	0.11
mtr	s	1.19	0.98					*mtr*	s	0.12	0.54
fdoG	as	1.19	2.14					*murA*	as	0.09	1.07
yjgR	s	0.74	3.48					*nrfD*	s	0.09	1.82
rpoN	s	0.69	3.66					*ynfA*	s	0.08	2.14
umuD	s	0.69	2.50					*fdoG*	as	0.08	0.64
ygjH	s	0.66	1.61					*rhsE*	s	0.07	1.82
degQ	s	0.61	2.94					*fhuB*	s	0.07	1.82
kdsD	s	0.35	2.85					*rhsD*	s	0.06	1.18
iap	s	0.16	1.25					*clpB*	s	0.06	0.21
fadR	s	0.15	0.80					*yrbG*	s	0.04	0.11
aceE	as	0.14	1.52					*phnF*	as	0.04	1.18
dgt	s	0.12	1.25								
opgH	as	0.10	1.52								
yrbG	s	0.07	2.85								
sslE	as	0.06	0.27								

[a] For each iteration the gene locus of the clustered mapped reads, their orientation towards the respective gene (s/as), and fractions of mapped (map [%]) and deduplicated reads (ded [%]) relative to the total number is listed. Concordant color-coding indicates occurrence of clusters in two replicates. Grey letters mark clusters with either low mapped or deduplicated leads that were still included in the list.

Since no functional pattern or definite sequence motif was observable and structural resemblance of enriched RNAs seemed only ill-definable, biological relevance of the interaction between PykF and enriched sequences could not be assumed. This conclusion would be in line with CLIP-Seq not yielding target RNAs either. Nevertheless, representative RNA transcripts from gene loci *holB, yjbH,* and *rnd* were analyzed *in vitro* in order to rule out preferential binding to PykF, relative to a control RNA. The EMSA displayed in Figure 18b shows that 2 µM PykF was not enough to produce visible band shifts for either *yjbH* asRNA or the control RNA. Increasing the PykF concentration to 25 µM (Figure 18c) consistently produced band shifts for *yjbH* asRNA, *rnd* mRNA, *holB* mRNA, and control RNA with only marginal variance in bound-to-unbound ratio. While only comprising a brief screening, these results are indicating that the affinity of these enriched RNA fragments does not significantly deviate from the general, unspecific RNA affinity towards PykF. The stem-loop aptamers have been selected during SELEX for unknown reasons, and feature too heterogenous structures to draw conclusions. Overall, results provoked no further efforts to identify an RNA-related moonlighting function in *E. coli* pyruvate kinase.

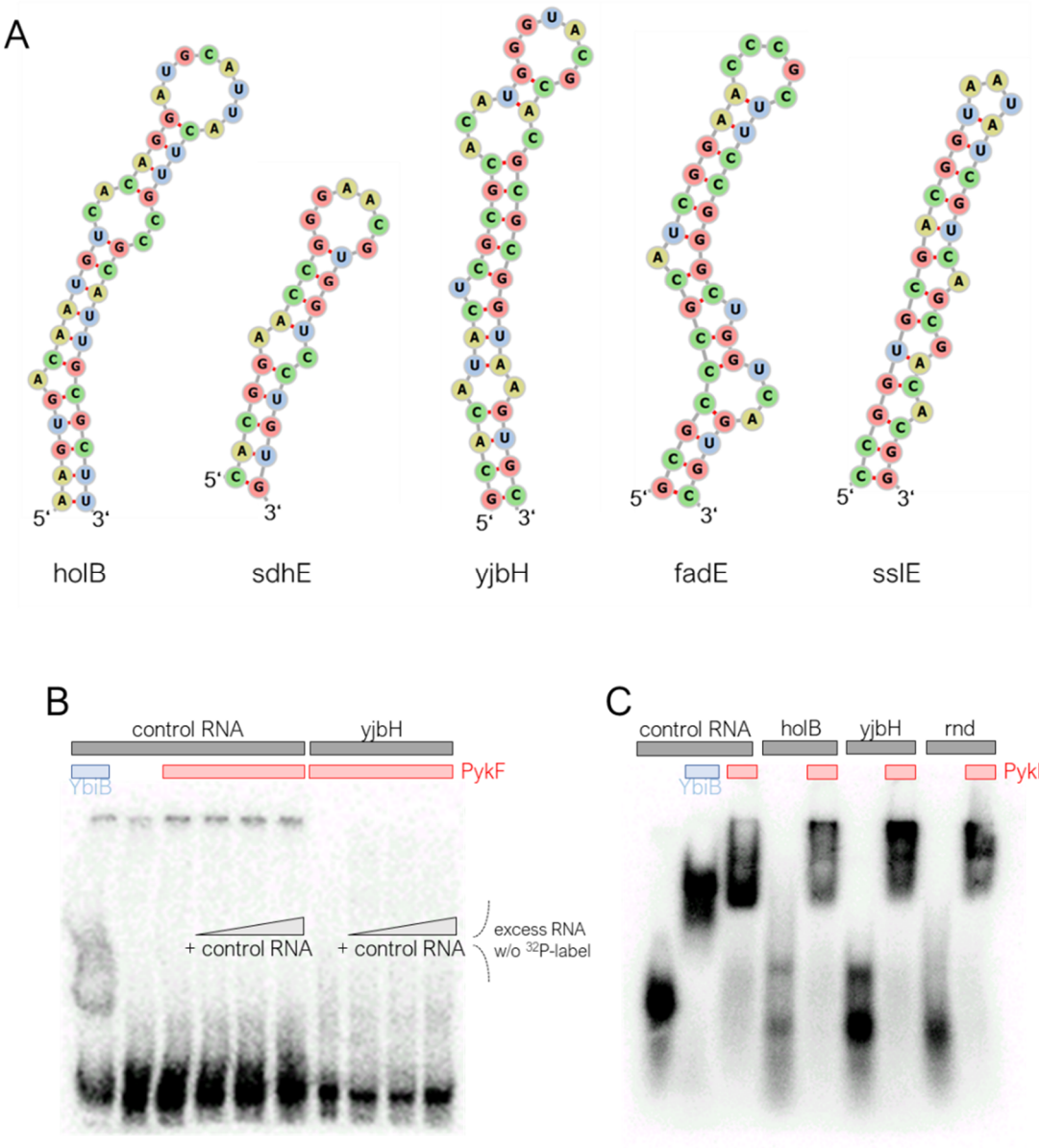


Figure 18: Representation and *in vitro* analysis of RNA fragments enriched by PykF. A: Secondary structure prediction for RNAs enriched during the PykF SELEX experiment. **B:** EMSA analysis employing 2 µM PykF and 400 nM of ^{32}P-labeled *yjBH* asRNA or control RNA. Competition with [1,2,7] µM unlabeled control RNA. **C:** EMSA analysis employing 25 µM PykF and 5 µM of ^{32}P-labeled *holB* mRNA, *yjbH* asRNA, or *rnd* mRNA. Unspecific RNA-binder YbiB was employed solely to verify the functionality of the assay.
RNA-sequences (stated as GenBank entries NC_000913**):** *holB* mRNA: 1156266-1156346, *yjbH* asRNA: 4239057-4239139, *rnd* mRNA: 1887007-1887081, control RNA: 3818686-3818763.

4.7.2 Genomic SELEX with Phosphoglycerate Kinase (Pgk)

Sequencing and data processing of the Pgk-exposed genomic library yielded 127,689 mapped reads and deletion of duplicates yielded 123,283 remaining reads. Table 9 lists all gene loci harboring read clusters with >0.2 % of total deduplicated reads and >~0.2 % of total mapped reads. A MEME query with respective sequences yielded a motif rated with an E-value of 1.7×10^{-6}. The alleged motif, occurring in 15 out of the 19 input sequences, is depicted in Figure 19a. While the E-value can be smaller for motifs of highly specific RNA-binding proteins (for example

Table 9: Unique read clusters identified in the Pgk-exposed genomic library.

gene locus	sense/antisense	mapped reads	deduplicated reads
scpC	s	12.01 %	3.82 %
fecR	s	8.24 %	3.83 %
kdpB	as	2.34 %	0.96 %
5‘ UTR *ypfG*	as	1.17 %	1.24 %
yehX	s	0.89 %	1.13 %
mlaE	as	0.84 %	0.59 %
yiaC	as	0.79 %	0.73 %
dnaG	s	0.76 %	0.54 %
yjjV	as	0.69 %	0.59 %
mutL	s	0.58 %	0.61 %
yjbb	as	0.55 %	0.26 %
putp	s	0.47 %	0.55 %
priA	as	0.42 %	0.53 %
hycC	as	0.40 %	0.36 %
lysP	as	0.30 %	0.22 %
ppsa	s	0.28 %	0.35 %
bglX	s	0.25 %	0.34 %
psts	as	0.23 %	0.27 %
3‘ UTR *yebF*	s	0.19 %	0.24 %

the MS2 coat protein, see section 4.6), it was significant enough to initiate further investigation. The motif suggested an A-rich central sequence with adjacent GC-pairings. Secondary structure predictions (RNAfold) of all enriched RNAs were inspected in order to identify common elements. Within each cluster, a similar stem-loop, presumably responsible for the calculated motif, was detected in the region of maximum read density. For the top eight clusters (with respect to mapped read number), the respective stem loops are displayed in Figure 19b. Base-pairing stems (not depicted fully) are of diverse length and composition and do not display a comprehensible pattern. Loops are enriched predominantly in adenosines and subordinately in cytosines. Both loop length and sequence of nucleotides was found to be heterogenous, contradicting the discovery of a stringent loop recognition motif. The functions of underlying genes were reviewed but did not show notable biological relationship (data not shown).

Pgk is comprised of two Rossmann-fold domains responsible for ATP binding and substrate binding, respectively (Young et al. 2007). While the Rossmann-fold is reported to confer specific RNA-binding in certain proteins (Nagy and Rigby 1995), it is also comprehensible that Pgk selected the aptamers based on an coincidental interaction involving the nucleotide binding pocket and the AC-rich loop, considering the adenine nucleotide binding function of Pgk. EMSA analysis of the interaction was conducted to estimate the affinity of enriched aptamers, and a representative EMSA is depicted in Figure 19c. Only minor complex bands could be detected for respective fragments of *kdpB* asRNA, *fecR* mRNA, and *scpC* mRNA (representing clusters with the most reads) when employing 30 µM Pgk, indicating that the interaction shows no higher affinity than estimated for the unspecific RNA-binding activity of Pgk determined in previous work (K_D^{Pgk} = 42 µM, see Figure S 27). More strikingly, a short control RNA fragment (AGGACAAAAACAA) showed stronger band formation than the RNAs enriched in the SELEX experiment, reinforcing the alleged adenine-based aptamer selection. While it could not be

completely ruled out that the aptamer selection is hinting towards a physiological protein-RNA interaction whose affinity is obscured by lack of particular *in vivo* conditions, results on hand did not suffice to further investigate the interaction. It should also be noted that similar structures could occur many times in the *E. coli* genome. Though, a number is hard to estimate owing to the heterogeneity of identified loops.

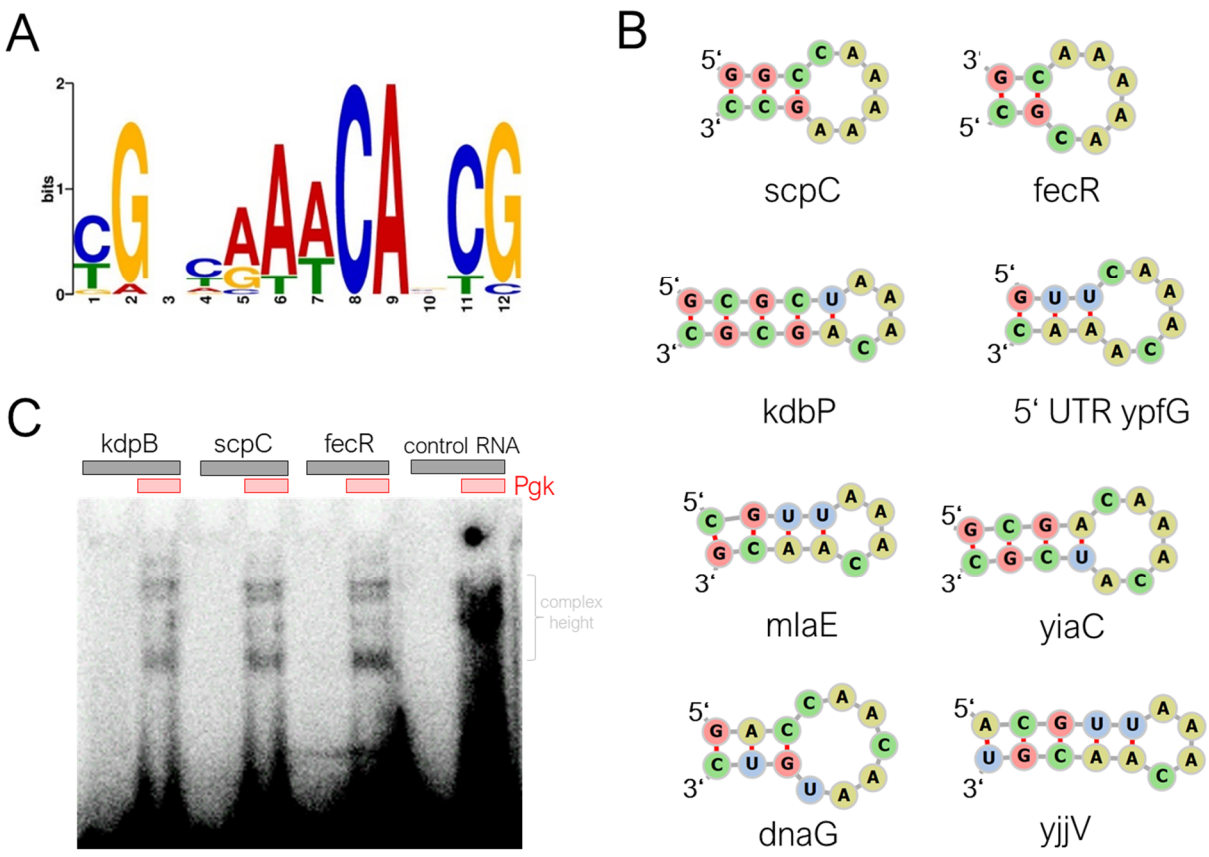


Figure 19: Representation and *in vitro* analysis of RNA fragments enriched by Pgk. A: Sequence motif calculated by MEME based on enriched RNAs. **B:** Secondary structures predicted for the eight most abundant RNAs in the Pgk-exposed library. **C:** EMSA analysis employing 30 µM Pgk and 7 µM of respective ^{32}P-labeled RNA. Complex band height is indicated.
RNA-sequences (stated as GenBank entries NC_000913**):** *kdpB* asRNA: 726837-726856, *scpC* mRNA: 3064947-3064966, *fecR* mRNA: 4517071-4517090. Control RNA: AGGACAAAAACAA.

4.7.3 Genomic SELEX with Thymidylate Synthase (ThyA)

Sequencing and data processing of the ThyA-exposed genomic library yielded 72,556 mapped reads and deletion of duplicates yielded 1,093 remaining reads. Even though the deduplicated read number was low, unambiguously clustered reads could be identified in the mapping. Table 10 lists all respective gene loci harboring unique clusters with >0.5 % of total deduplicated reads and >0.05 % of mapped reads. Gene locus *rpsC* with mappings of ThyA-evolved library, PykF-evolved library (processed in parallel to ThyA), and PCR-treated control library is representatively shown in Figure 20a, demonstrating that despite the low number of deduplicated reads, unambiguously enriched clusters still were clearly identifiable for ThyA. Strikingly, the top three clusters accounted for ~85 % of mapped reads and also comprised most deduplicated reads, suggesting strong favorable selection of these RNAs.

To find hints on commonalities between clusters listed in Table 10, respective genomic sequences were used as input for a MEME motif search. MEME calculated a motif with an E-value of 1.2×10^{-19} (Figure 19b) which shows strong conservation of nucleotides and strikingly appears in every input sequence. Exemplary close-up views of reads at gene loci *rpsC, yhdY, rlmD,* and *ugpQ* are depicted in Figure 19c. Positions of sequences comprising the motif are marked, showing that the position of the motif consistently colocalized with the apex of coverage. An alignment of sequences from all hits with respect to the motif is displayed in Figure 20d. Since every last hit showed close resemblance to the motif, the MEME search was repeated including clusters initially omitted owing to low read numbers (Table 10, grey entries), and it produced the same

Table 10: Unique read clusters identified in the ThyA-exposed genomic library.[a]

Gene locus	sense/antisense	mapped reads	deduplicated reads
rpsC	as	37.60 %	6.68 %
yhdY	s	29.60 %	7.14 %
rlmD	as	17.56 %	4.76 %
ugpQ	s	9.34 %	1.37 %
yqhD	as	0.99 %	2.93 %
yhcB	s	0.73 %	0.82 %
yjeI	s	0.54 %	1.01 %
fecB	s	0.53 %	2.10 %
aceA	as	0.37 %	2.84 %
yjhP	as	0.17 %	2.29 %
hcp	s	0.15 %	1.10 %
nagE	as	0.14 %	2.84 %
hypF	s	0.12 %	1.65 %
ybgL	s	0.09 %	2.65 %
yjgR	s	0.07 %	0.91 %
pppA	s	0.07 %	2.10 %
putA	s	0.03 %	0.64 %
metE	as	0.03 %	0.82 %
pflC	s	0.02 %	0.37 %
fliH	s	0.02 %	0.46 %
uvrB	s	0.02 %	0.64 %
gpp	as	0.02 %	0.82 %
yjcO	as	0.01 %	0.55 %

[a] Gene loci listed in grey were excluded from the original MEME query due to low read number occupation.

motif with an improved E-value of 8.4×10^{-33}. Respective aligned sequences are shown in Figure Figure 20d, bottom.

The calculated motif clearly represents a stem loop structure with the consensus form schematized in Figure 20e. The third base of the loop seems to be least conserved, albeit mostly occupied by an adenosine or cytosine. The two base-pairs closest to the loop are highly conserved, and notably, the second base-pair shows a strong preference towards the Wobble-base interaction G-U. The read distribution suggests that a longer stem might have been beneficial in the selection process, as top four hits *rpsC*, *yhdY*, *rlmD*, and *ugpQ*, accounting for a combined share of 94 % of mapped reads, are all predicted by RNAfold to form a stem-structure 5 nucleotides in length (see Figure 20f). Intriguingly, the structure strongly resembles an iron-responsive element found in human amyloid-beta-precursor protein mRNA, which features a stem-loop GC(AGA)GC with the same AGA tri-loop (Cho et al. 2010). No read enrichment was observed for *thyA* mRNA, which reportedly binds to its own gene product ThyA in *E. coli* (Voeller et al. 1995). Looking for a connection of own results and this reported feedback interaction, the *thyA* sequence was screened for the enriched motif, but no such stem-loop structure was identified.

The interaction between ThyA and the identified stem loop structure was subsequently probed by EMSAs. Unexpectedly, no complex band was observable in any experiment. Figure 20g shows a representative gel probing *rpsC* asRNA and *ugpQ* mRNA for an interaction with 40 µM of ThyA. The contradiction of promising SELEX results and negative binding assays could not be resolved within this work. While an artificial enrichment of identified aptamers caused by filter-binding or PCR-bias cannot be ruled out, this explanation is not line with the stem-loop motif not being observable in any other SELEX experiment conducted in the course of this work. SELEX might be able to discriminate fine differences in affinity even if found in a very low affinity regime. If the aptamer selection was indeed caused by ThyA, it remains unclear whether the interaction carries biological significance in the cell, and whether observation of higher affinity in EMSA analysis was obscured by lack of necessary *in vivo* conditions.

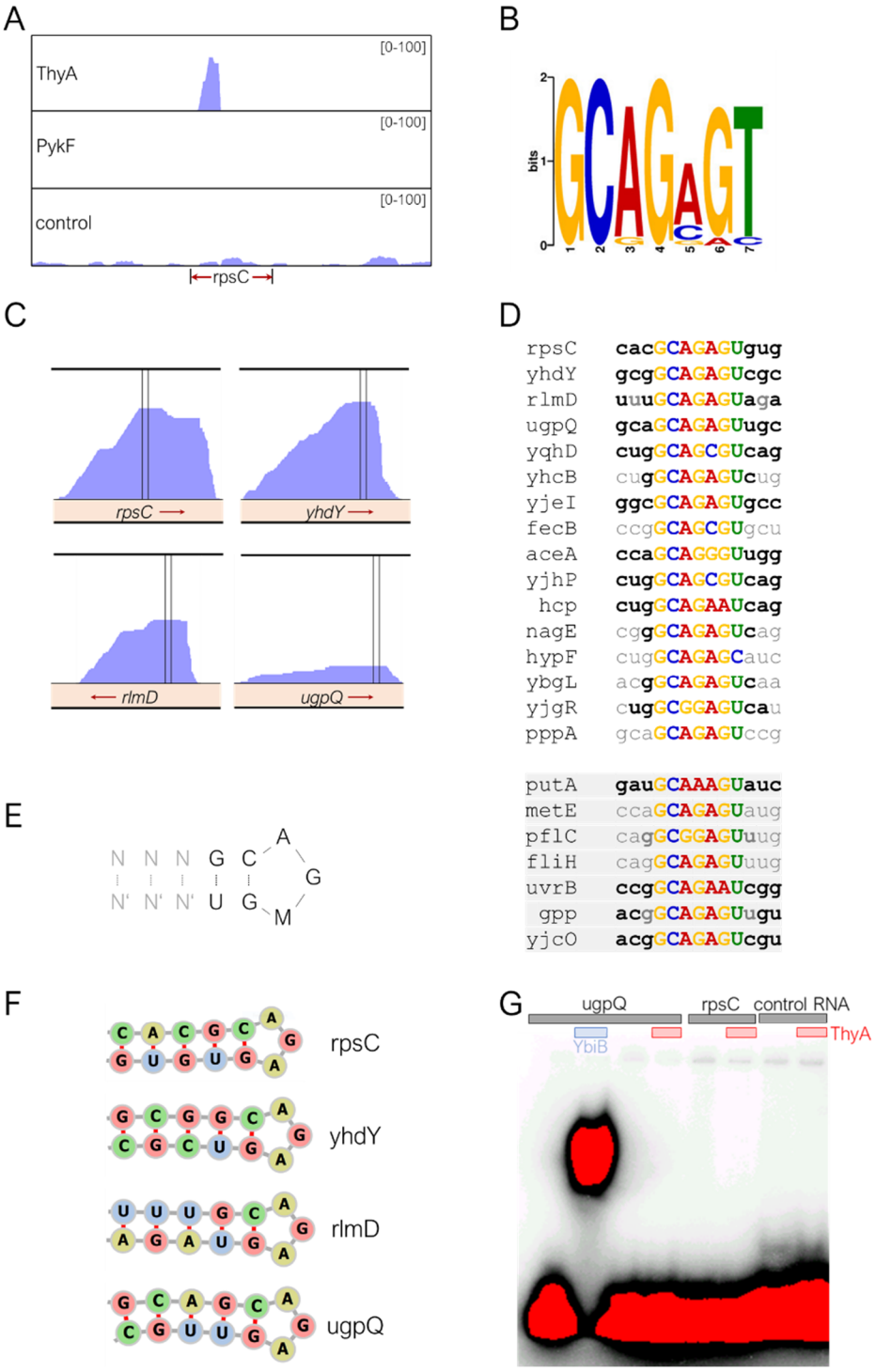


Figure 20: Representation and *in vitro* analysis of RNA fragments enriched by ThyA. **A:** Traces showing mapped reads at NC_000913:3449000-3452200 for ThyA-exposed, PykF-exposed, and PCR-treated control library. The *rpsC* locus is indicated. **B:** Sequence motif calculated by MEME based on enriched RNAs. **C:** Close-up view of representative read clusters. The motif sequence is indicated by vertical bars. Arrows indicate orientation of underlying genes. **D:** Relevant sequences of identified RNAs aligned to the stem-loop motif. Nucleotides in black indicate continuation of the probable stem. **E:** Generalized consensus motif of identified RNAs. **F:** Secondary structures calculated by RNAfold for *rpsC* asRNA, *yhdY* mRNA, *rlmD* asRNA, and *ugpQ* mRNA. **G:** EMSA deploying 40 µM ThyA and 10 µM of respective ^{32}P-labeled RNA. **RNA-sequences** (stated as GenBank entries NC_000913)**:** *rpsC* asRNA: 3449339-3449381, *ugpQ* mRNA: 3587434-3587460, control RNA: 2562311-2562337. Unspecific RNA-binder YbiB was employed to verify EMSA functionality. Red signalizes oversaturation.

4.7.4 Genomic SELEX with Glutamate Kinase (ProB)

E. coli glutamate kinase was a promising candidate for moonlighting enzyme function due to its PUA domain. The domain is speculated to carry an RNA-binding function, albeit unconfirmed (Marco-Marín et al. 2007). Sequencing and data processing of the ProB-exposed genomic library yielded 40,336 mapped reads, and deletion of duplicates yielded 907 remaining reads. The low number of remaining deduplicated reads were spread out fairly evenly across the genome. This can indicate that no specific binding is to be found, as the library was thinned out early during the SELEX experiments. However, three read clusters were identifiable within the mapping, summarized in Table 11. The significance of these clusters was assessed as low, owing to unhealthy peak shape (high mapped-to-deduplicated read ratio, see Figure 21a). For comparison only, a 'healthy' cluster comprised of many reads found in the ThyA experiment is shown, forming a brief bell distribution. In contrary to that, *tktA* is only occupied by a single read, duplicated many times (22,247-fold). The significance of the three identified clusters was therefore ambiguous.

Comparing the clusters, a communal sequence or structural motif could be identified neither by MEME motif search nor by manual comparison. To investigate whether the high rate of duplication was merely a result of experimental bias or if underlying deduplicated reads were actually enriched by the enzyme, binding of respective RNA to ProB was examined in EMSA analysis. In fact, initial EMSA experiments showed preferential binding of a *tktA* asRNA fragment to ProB, relative to two different control RNAs (Figure 21b). Following up on this result, the interaction was challenged with a 50-fold excess of unlabeled RNAs (Figure 21c). Adding unlabeled *tktA* asRNA to the *tktA* asRNA-ProB complex expectedly impeded band intensity, but two different control RNAs had seemingly no impact on complex formation (last lane vs. third and fourth lane).

While *in* vitro analysis of the *tktA* asRNA fragment demonstrated that clusters in Table 11 were indeed found based on an preferential binding interaction with ProB, the significance of the interaction still has to be assessed critically. The aforementioned high ratio of mapped-to-deduplicated reads indicates a quick depletion of the genomic library, challenging the presence of any highly affine species with respect to ProB in the RNA pool. In line with that, the affinity of the interaction was estimated to be low with respect to requirements for biologically relevant interaction, with an estimated $K_D >> 8$ µM judging from band intensity ratios (Figure 21c) and difficulties obtaining a measurable isothermal calorimetry signal for the interaction (data not shown). However, it is uncertain whether ideal conditions for ProB and its interaction had been established in this work. While functionality of the enzyme was ascertained by assaying substrate turnover (Figure S 14), the protein displayed poor long-term thermal stability, and concentration

Table 11: Unique read clusters identified in the ProB-exposed genomic library.

Gene locus	sense/antisense	mapped reads	deduplicated reads
tktA	as	55.15 %	0.11 %
lepB	s	30.45 %	1.32 %
yfjR	s	5.75 %	0.88 %

measurements might have been inaccurate, falsifying the affinity estimation within EMSA experiments. Further investigations building on herein presented results were outside of the work's time frame and are required for definite conclusions about a potential specific binding interaction of ProB. A key factor of ProB stability might be a suspected complex formation with glutamate-5-phosphate reductase (ProA), supported by structural data (Marco-Marín et al. 2007), by instability of the ProB enzymatic product L-glutamyl 5-phosphate (Seddon et al. 1989), and by bifunctionality (dual kinase-reductase activity) of the analogous eukaryotic enzyme (Hu et al. 1992). At the very least, the observed specificity directs further attention to a possible RNA-binding functions of the PUA-domain in metabolic enzymes.

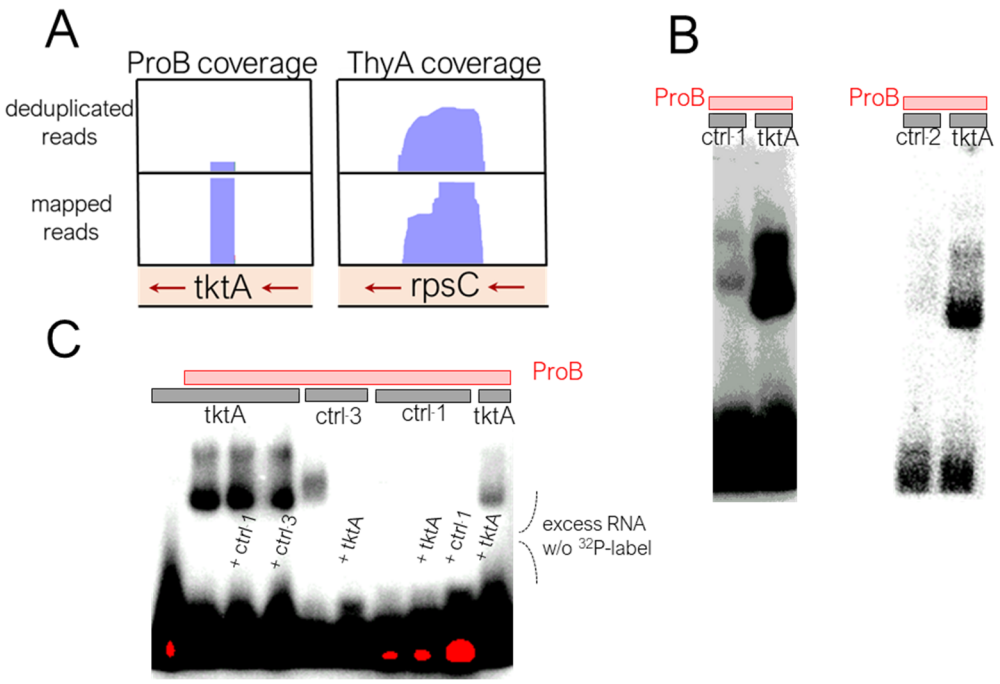


Figure 21: Read cluster representation and in vitro analysis of RNAs enriched by ProB. A: Mapped reads at the *tktA* gene locus. On the right, reads of a cluster in the ThyA-evolved library (section 4.7.3) are shown as reference for a healthy read distribution. **B:** Initial EMSA experiments pointing to preferred binding of ^{32}P-labeled *tktA* asRNA to ProB, relative to control RNAs; left – c_{ProB} = 8 µM, c_{RNAs} = 2 µM; right - c_{ProB} = 8 µM; c_{RNA} = 1 µM. **C:** EMSA verifying preferential binding of ^{32}P-labeled *tktA* asRNA in competition experiments. ProB (8 µM) and ^{32}P-RNAs (2 µM) were supplemented with excess (100 µM) unlabeled RNA where indicated. Red signalizes oversaturation.
RNA-sequences (stated as GenBank entries NC_000913**):** *tktA* asRNA: 3080595-3080620; ctrl-1: 510306-510332. ctrl-2: 2562281-2562307; ctrl-3: 2562244-2562270.

4.7.5 Genomic SELEX with Aconitase (AcnB)

AcnB is the *E. coli* pendant to the prime example for moonlighting enzymes, aconitase. Sequencing and data processing of the AcnB-exposed library yielded 180,535 mapped reads and 132,821 deduplicated reads. Initially, it was observed that no reads were found at the 3' UTR locus of *acnB*. Published data indicates that inactivated AcnB, i.e. AcnB not occupied by an iron-sulfur cluster, self-regulates its own mRNA stability by binding to 3' UTR secondary structure elements, albeit differing from iron-responsive elements (IREs) identified as aconitase interactors in eukaryotes (Tang and Guest 1999). Even though the 3' UTR *acnB* locus showed no enrichment, five other genomic loci were identified that harbored a significant number of reads (Table 12). Reads were not perfectly unique to the AcnB-exposed library but enriched only in the latter one. Figure 22a representatively shows the mapping around gene locus *rplJ*. For AcnB, the marked read cluster shows an increased number in both deduplicated and predominantly mapped reads. The PdxH-evolved library (treated in parallel, shown for comparison only) also harbors the cluster, but not above background level. It was concluded that peaks most likely originate from AcnB-based selection, and respective RNAs potentially contaminated parallel samples despite stringent washing of filtering equipment, making them appear but not be enriched in the contaminated experiment.

In order to gain a lead as to why these specific RNA fragments were enriched in the AcnB experiment, sequences and corresponding secondary structures (predicted by RNAfold) were examined and checked against available literature on AcnB-RNA interactions in prokaryotes. While MEME queries and manual inspection did not lead to identification of a stringent sequence motif within the five RNA fragments, comparison of secondary structures revealed a potential, albeit diffuse resemblance of the 3' UTR of *acnB* in *E. coli* (Tang and Guest 1999; Benjamin and Massé 2014). These studies suggested that the 3' UTR contains stem-loops with adjacent AU-rich stretches and conducted mutational analysis to narrow down the binding site. While the exact pinpointing of a secondary structure element responsible for binding is inherently tied to the yet challenging effort to accurately predict base-pairings (the reader is referred to different structure representations between forecited publications), an involvement of the 3' UTR region of *acnB* in translation rate is elaborately established. An own comparative analysis of the 3' UTR region of *acnB* with different prediction suites (RNAfold, CentroidFold, IPknots, RintD, RintW) coherently predicted a stem-loop structure at positions nt +34 to nt +63 (counting from the *acnB* stop codon) with a 3'-adjacent AU-rich stretch (see Figure 22b, left). Strikingly, enriched RNA fragments likewise contain a probable, albeit heterogenous stem-loop with a 5'-adjacent AU-rich (predominantly U-rich) segment, according to calculations with the same prediction

Table 12: Significant read clusters identified in the AcnB-exposed library.

gene locus	sense/antisense	mapped reads	deduplicated reads
5' UTR *rplJ*	s	12.01 %	3.82 %
copA	as	8.24 %	3.83 %
yejM	as	2.34 %	0.96 %
3' UTR *ymgC*	as	1.17 %	1.24 %
sbcC	as	0.89 %	1.13 %

tools (Figure 22b). The stem-loop positioning coincides with the coverage apex of all clusters and was likely responsible for the RNA enrichment. However, it is unclear whether these identified interactions carry biological significance, considering their high sequence heterogeneity.

The stem loop of the RNA with highest enrichment is located on a transcript of an intergenic region ~250 nt downstream of nearest upstream gene *rplA* and ~140 nt upstream of nearest downstream gene *rplJ*. Enriched transcripts of gene loci *copA*, *yejM*, *sbcC,* and *ymgC* are all oriented antisense (with the latter transcript lying 20 nt downstream of the stop-codon). A biological function would only be conceivable in the way that AcnB stabilizes or destabilizes respective non-coding RNAs and asRNAs, which in turn have regulatory function. In the case of *copA*, a regulator of copper homeostasis (Meydan et al. 2017), a functional relationship could be easily deduced, since copper, disrupting iron-sulfur cluster assembly, is known to be toxic for dehydratases like AcnB (Macomber and Imlay 2009), and the enrichment of antisense *copA*

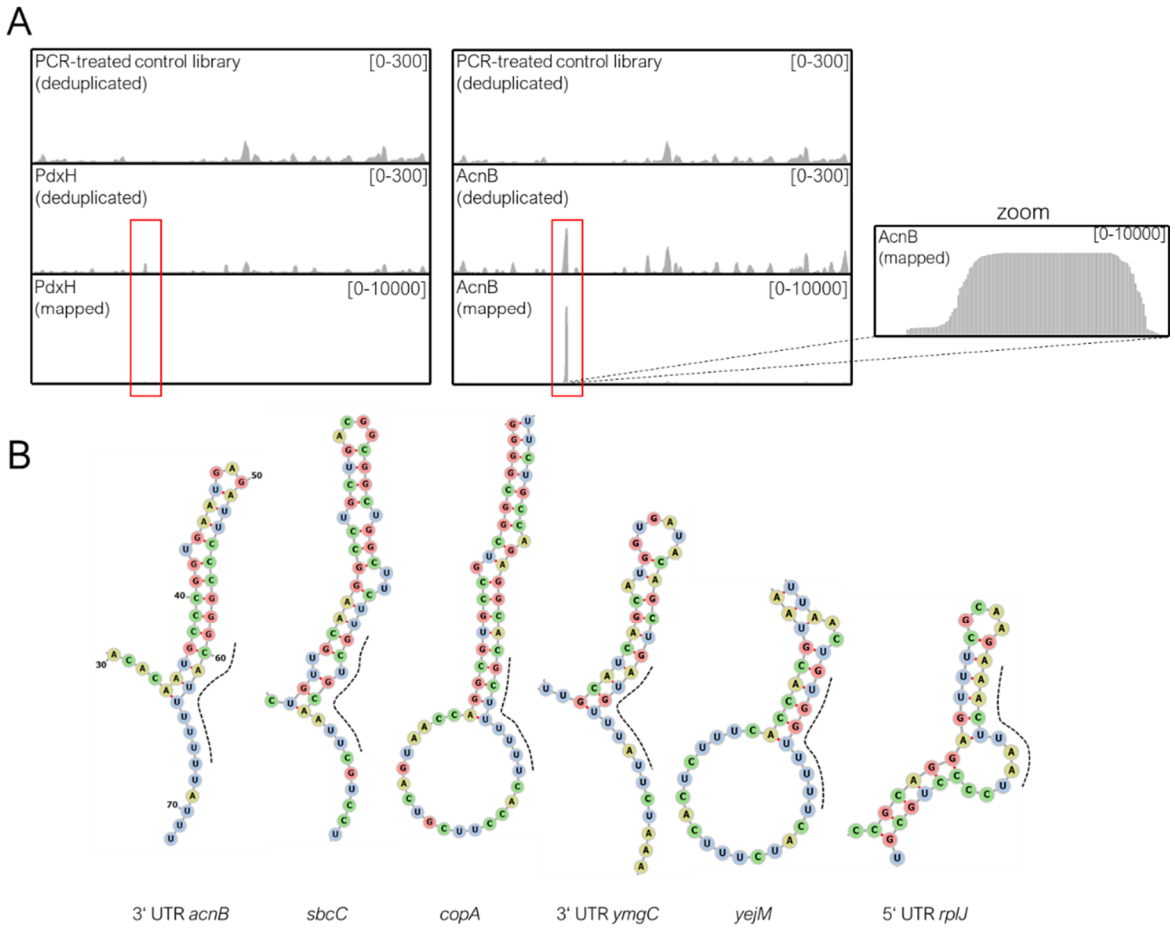


Figure 22: Representative read cluster and secondary structure prediction of enriched RNAs. A: Genome mapping NC_000913:4175950-4189670. Red rectangle highlights the position of read-enrichment in the AcnB experiment (zoom-inlet in on the right). Traces of the control library and the PdxH-evolved library (employed in parallel to AcnB) are aligned for comparison. **B:** secondary structure visualization of the 3' UTR region of *acnB* (left, numbers indicate nucleotides downstream of *acnB* stop-codon) and of RNAs enriched in the AcnB SELEX experiment. Black dashed lines indicate the communal feature of RNAs, an U-rich or AU-rich stretch 3'-adjacent to a stem.

transcripts in the SELEX experiment might not be coincidental. However, it is unclear why the alleged binding of AcnB to its own mRNA was not reproducible, impeding the significance of presented results. It can be speculated that respective sequences were coincidentally poorly resembled in the original library.

4.7.6 Genomic SELEX with Glyceraldehyde-3-Phosphate Dehydrogenase (GapA)

Eukaryotic glyceraldehyde-3-phosphate dehydrogenase is frequently mentioned in literature as non-canonical, AU-rich RNA binding protein (Garcin 2019). The SELEX experiment with *E. coli* GapA yielded 16,377 mapped reads and 4,333 deduplicated reads. Table 13 lists unique read clusters with >0.1 % mapped reads and 0.3 % deduplicated reads. Initially, neither a MEME analysis nor examination of secondary structure predictions provided hints on a common selection criterium within RNAs. However, a close comparison of sequences revealed that most of the sequences carry a linear, AU-rich nucleotide stretch (Table 13, right). These AU-rich stretches are consistently at loci of maximum read coverage, making them presumably responsible for enrichment of respective RNAs. It shall be noted however that the combined mapped read-number of RNAs sharing this pattern comprises only 6.25 % of total reads, and other clusters, co-occurring in parallelly treated libraries (thus being safely attributed to PCR bias), showed higher mapped read numbers, contesting the selection pressure of the AU-rich stretches.

The biological background of genes underlying the AU-rich RNAs was examined and showed no commonality, raising the question if all of the potentially identified interactions could have a

Table 13: Read clusters identified in the AcnB experiment.

gene	sense/ as	mapped reads	dedup. reads	Transcript of genomic sequence colocalizing with apex of read cluster coverage.
cmoB	as	0.88 %	1.85 %	acuggcacag**aaa**c**a**g**uu**gcguagggucgaucccccaccgcgaggugc
yihA	as	0.88 %	1.78 %	accguuaaacgccaguacagcuucacgcacc**auauu**c**aauu**gcgcuu
ydiQ	as	0.68 %	1.22 %	accguuggcggc**u**c**auu**g**uu**gcagaacucgaaagugcgcaaagacgugcu
entE	s	0.58 %	1.22 %	cgcacucauaacgacuacuacuacagcgugcguc
citX	s	0.49 %	0.99 %	gaugccugcaacgucaac**uaaaa**c**ca**c**aaa**gcuugcg
trua	as	0.46 %	0.99 %	cacaaccagccggagagcuggauagcagaguugcuggc
malX	as	0.31 %	0.60 %	aauaccgaacaccacgaagucg**au**g**auauu**gccgucgguau
cobT	as	0.31 %	0.53 %	gucc**aaaa**gc**aa**c**uuuu**cagccugacggcgacucauugccggagcuga
umpH	s	0.21 %	0.72 %	aucgcagagg**uauaaaa**c**a**cgcugucagg
ddpA	as	0.20 %	0.42 %	gguagaccaggaaauggcccgacgcagauccgcc
yfcU	s	0.18 %	0.30 %	ucauugcggacuacagcaucacugcgcaaacacgacacg**aa**g**aaaau**ggcggcga
yagE	s	0.18 %	0.44 %	uucuuccugggcagcgguggcgaguucucccagcucggcgccgaagagcguaaag
yebZ	s	0.16 %	0.39 %	ccugccucaugcgcggu**acaa**g**aa**c**auaa**cgauuagccagcg
dxr	s	0.14 %	0.35 %	agugcg**aaa**c**uu**c**uuaaaa**cgaugcuac
pauD	s	0.13 %	0.46 %	cc**u**c**auuuau**gcccgucuuauccguuuccgcuuugcccuucacc
hofP	as	0.13 %	0.44 %	uugcugcacccgucgcc**auuu**c**uuuu**gccc
yhes	s	0.12 %	0.44 %	aucuuuacgugccugggcgcuguucgc**auu**c**u**c**uuuu**ggcgcuucgucg
yafN	as	0.12 %	0.42 %	uugcugaacaagagg**a**g**aaaaa**gcccauaaaggcgcg
arfB	s	0.11 %	0.42 %	uucc**uu**c**uuuuuauu**cccgaccgcugcgc

On the right, respective RNA sequences of each cluster are shown, constricted to the area of maximum coverage. Where present, AU-rich stretched are marked in bold & underline.

physiological function. However, the ability of eukaryotic GAPDHs to bind AU-rich elements is frequently mentioned in literature (Nagy and Rigby 1995; McGowan and Pekala 1996; Schultz et al. 1996). Principally, the enrichment of the AU-rich elements for *E. coli* GapA can therefore be considered an interesting observation. A possible interpretation of the diversity of identified RNAs in sequence and function is either a limited suitability of SELEX for unambiguous isolation of highest affinity biological targets of GapA (if existent), or the existence of a coincidental, unspecific affinity towards AU-rich stretches. In either case, transferring results from eukaryotic GAPDHs to GapA is vague, even when considering the structural homology and the mutual presence of a Rossmann-fold domain. Owing to the diversity of identified RNAs, no further *in vitro* studies were conducted for GapA.

4.7.7 Genomic SELEX with Quinone Oxidoreductase (QorA)

Sequencing and data processing of the QorA-exposed library yielded 128,201 reads, and 31,412 reads remained after deletion of duplicates. Most of the deduplicated reads were evenly distributed across the genome, similar to the PCR-treated control library. However, a small number of unambiguous peak clusters were identified based on an overproportionate accumulation of mapped reads, summarized in Table 14.

Strikingly, one single cluster at the gene locus *yffO* contained 88,507 reads, accounting for 69 % of all total reads and more than 91 % of duplicated reads. While this suggests that the overrepresentation is at least partially owed to PCR-amplification, the enrichment of the *yffO* transcript fragment displayed a healthy deduplicated read pattern (Figure 23a). A MEME motif calculation based on genomic sequences underneath the nine identified clusters as input yielded a motif (see Figure 23b) with a moderate E-value of 4.9×10^{-4}. The motif is consistently located at the point of highest read coverage for all clusters (Figure 23d), making it plausible that it was responsible for respective RNA enrichments. However, a sequence logo based on only nine input sequences should only be seen as hint for potential specificity features.

The alignment of the enriched RNAs from each cluster to the motif sequence is ambiguous when looking at structural features (Figure 23c). While the *yffO* mRNA fragment forms a stem loop with a loop sequence CAUAUGGA according to RNAfold, an identical stem loop shape is adopted only by one other RNA, *aslB* mRNA (predicted stem-loop formation is indicated by underscoring in Figure 23c). Remaining sequences form alternative stem loops or even little to

Table 14: Unique read clusters identified in the QorA-exposed genomic library.

genlocus	sense/antisense	mapped reads	deduplicated reads
yffO	s	69.04 %	0.35 %
hrpB	as	0.99 %	0.01 %
narI	s	0.87 %	0.25 %
htpX	as	0.77 %	0.04 %
yeaX	as	0.27 %	0.01 %
paaH	s	0.25 %	0.03 %
aslB	s	0.18 %	0.04 %
gudD	s	0.16 %	0.05 %
yncD	s	0.06 %	0.01 %

no probability for base-pairing. Some of the divergently predicted secondary structures are exemplarily illustrated in Figure 23e. Considering the high surplus of mapped reads of *yffO* mRNA compared to remaining RNAs, as well as the moderate sequence conservation among

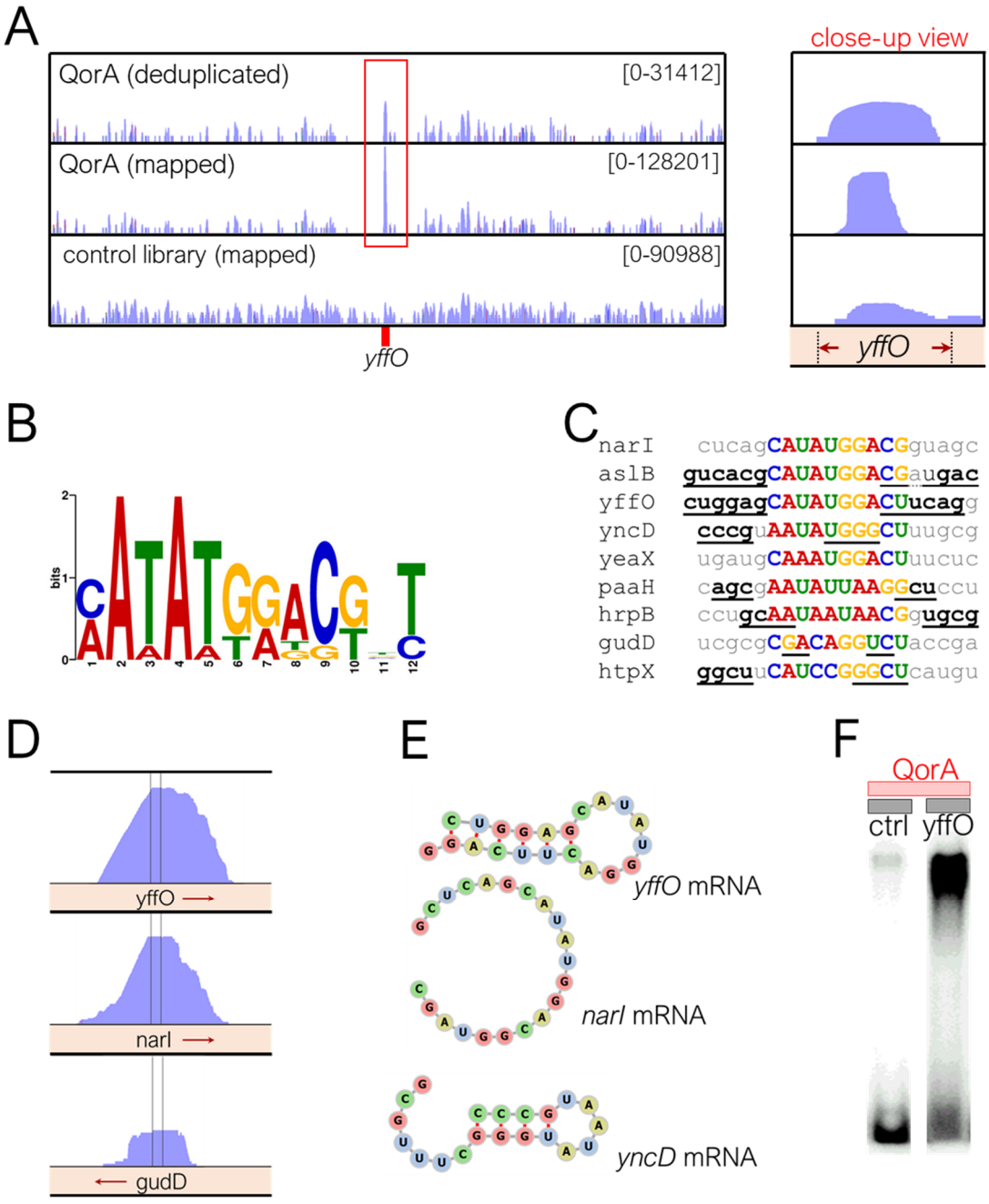


Figure 23: Analysis of SELEX results for QorA. **A:** Genome mapping NC_000913:2534120-2592550. Y-axis are log scaled, the region of enriched *yffo* mRNA is highlighted in red and shown enlarged on the right to display peak shapes. **B:** Sequence motif calculated by MEME based on enriched RNAs. **C:** Sequences of enriched RNAs aligned to the motif. Nucleotides likely involved in base-pairing according to RNAfold are indicated by black underscoring. **D:** Close-up view of representative read clusters. The localization of the motif is indicated by vertical bars. Arrows indicate orientation of genes. **E:** Representation of varying secondary structures adopted by *yffO, narI,* and *yncD* mRNA fragments based on RNAfold prediction. **F:** EMSA displaying the specific interaction between QorA (14 µM) and ^{32}P-labeled *yffO* mRNA (12 µM).

RNA-sequences (stated as GenBank entries NC_000913)**:** *yffO* mRNA: 2562311-2562337, control RNA: 3587434-3587460.

identified hits, a sequence-specific and/or structure-specific binding interaction directed towards only one or few RNAs could not be ruled out. For initial verification of the interaction, an EMSA analysis showed that QorA is able to strongly distinguish between an unrelated control RNA and the *yffO* mRNA fragment (Figure 23f). Due to this promising result, further analysis about the specificity of the interaction was conducted and is described separately in section 4.8.

4.7.8 Summary of SELEX Experiments of Remaining Candidates

SELEX experiments of candidates which were not aforementioned in dedicated sections showed only insignificant read enrichment or clusters that are not unique to the experiment but occurred in parallelly treated libraries based on filter binding or PCR bias. Genomic libraries exposed to any of KdsA, Upp, Mdh, Eno, GpmA, YggX, Adk, TalA, TalB, PdxH, and SodA fell into this category, and resulting mappings are not described in further detail.

Unspecific nucleic acid binding protein YbiB with unelucidated function showed strong enrichment for a limited set of RNA fragments. Most notable species were an *rpoC* asRNA fragment (23.4 % of reads), *rhaT* mRNA fragment (5.6 % of reads), *recJ* mRNA fragment (0.8 % of reads), and *iaaA* mRNA fragment (0.7 % of reads). The interaction between a respective *rpoC* asRNA fragment and YbiB was probed in EMSA analysis (data not shown). Binding of the transcript was observable but expected, considering the unspecific RNA property of YbiB (Schneider et al. 2015). A common denominator of the four identified fragments was found to be a nucleotide array GAAGCC central within each respective read cluster. Examination of this sequence is recommended as starting point for further studies of the YbiB RNA-binding function. The RNA enrichment was compared to results of SELEX experiments with a random RNA and ssDNA aptamer library conducted in the PhD thesis of Daniel Schneider (2011), but showed no resemblance to motifs identified from single clone sequencing at that time. The potential interaction was not further explored, as elucidation of YbiB, having no confirmed enzymatic function, was not a main focus with respect to the scope of this work.

In the AnsB experiment, reads with long uracil stretches were heavily duplicated. However, the amplification of these reads was attributed to effects other than protein selection and results are not further elaborated (data not shown).

While negative results for the majority of RBP candidates were disappointing, respective libraries served as important control to generally differentiate protein-based enrichments from artificial bias, especially in between parallelly treated experiments. It could also be deduced that a relatively strong unspecific RNA-binding activity does not necessarily entail RNA enrichment in SELEX experiment, as can be seen from KdsA (K_D = 0.7 µM, Figure S 27).

4.8 Characterization of QorA as RNA-Binding Protein

SELEX experiments with QorA led to strong enrichment of a transcript fragment of *yffO*, and initial EMSA experiments indicated increased affinity compared to a control RNA. A summary of the current biological background knowledge about QorA and *yffO* is provided in sections 4.8.5 and 4.8.6, respectively. Based on the SELEX result and under consideration of the alleged RNA-binding activity of homologous protein ζ -crystallin (Tang and Curthoys 2001), more elaborate *in vitro* binding studies were conducted to examine the specificity of the interaction.

4.8.1 Competition EMSAs Verifying Sequence-Dependent RNA-Binding by QorA

The *yffO* transcript fragment bearing the predicted motif was initially employed as 27 nt long RNA (NC_000913:2562311-2562337) when it was tested for interaction specificity (Figure 23f). These nucleotides roughly comprise the complete apex of coverage in the genome mapping. While the apex of coverage *per se* indicates the localization of the affinity determinant within the read cluster (see section 3.4.13 and Figure 9 of methods), increased affinity of adjacent sequences was precautionary ruled out by EMSAs (data not shown for conciseness). Especially considering that the predicted motif is shorter than 27 nt, it cannot be ruled out that the affinity determinant is only a subset of all nucleotides comprising the apex of coverage. Attempts to further narrow down the RNA length were conducted only later on (see section 4.8.4) and thus the 27 nt RNA was employed for all experiments in this subsection. The fragment is referred to as RNA^{yffO} in the following main text.

The ability of QorA to bind RNA^{yffO} was probed against equal-length 27 nt RNAs that are randomly selected transcripts of unrelated genomic regions. It was observable that equal concentrations of QorA induced almost quantitative complex formation for RNA^{yffO} but failed to significantly shift bands of unrelated transcript fragments RNA^{rpsC} and RNA^{yhhA} (Figure 24a, lanes 4,6,8). Addition of excess unlabeled, HPLC-purified RNA^{yffO} efficiently erased the complex band, ensuring that the latter was comprised of protein-RNA complex rather than a protein-nucleotide complex that might have emerged as a result of trace ^{32}P-ATP impurifications from the RNA labeling reaction (lane 5). The faintly visible RNA^{rpsC}-QorA complex which is quantitatively replaced by unlabeled RNA^{yffO} indicates that QorA bears a general, unspecific RNA-binding ability with reduced affinity (lanes 6,7). RNA^{yhhA} showed no significant complex formation (lane 8). Figure 24b demonstrates that excess unlabeled RNA^{rpsC} fails to displace the RNA^{yffO}-QorA complex, whereas equimolar excess of unlabeled RNA^{yffO} efficiently disrupts it (lanes 3,4). Furthermore, unspecific complex RNA^{rpsC}-QorA is displaced more efficiently by RNA^{yffO} than by RNA^{rpsC} (lanes 6,7).

While molecular weights of QorA and YbiB are nearly identical (35.2 and 35.1 kDa, respectively) and both proteins natively form dimers (Thorn et al. 1995; Schneider et al. 2015), the RNA^{yffO}-QorA complex migrated significantly slower in polyacrylamide gels in comparison to YbiB (see Figure 24a, lane 2 vs lane 4). Slow migration can generally be caused by too high polyacrylamide concentration or an unsuited pH-value with respect to the pI of the protein of interest. However, QorA migration was empirically found to be largely independent of such variables in extensive

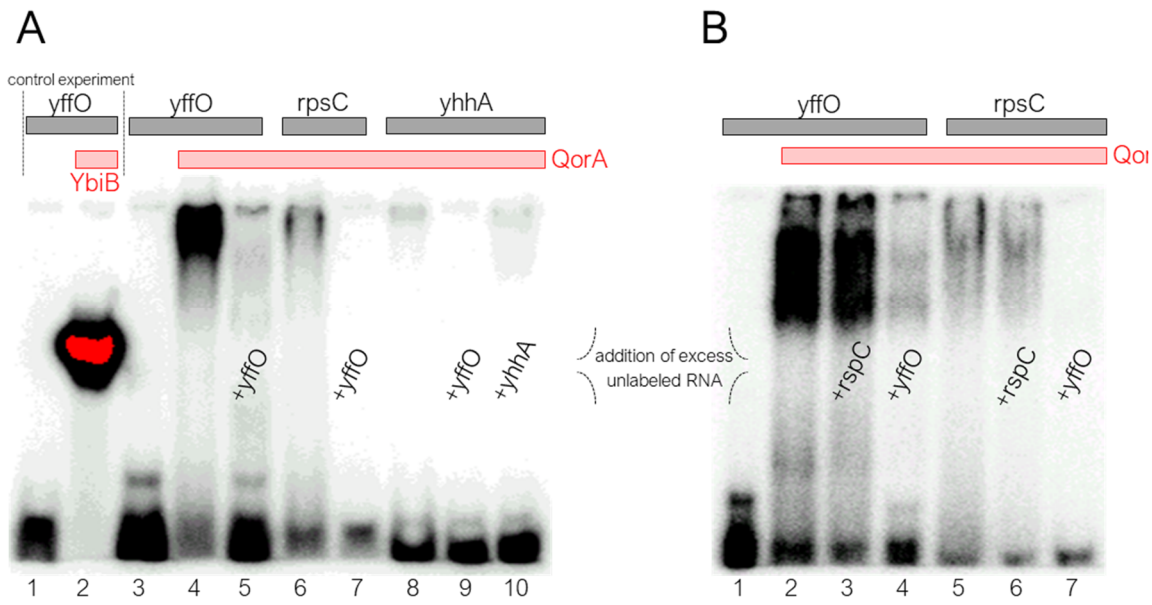


Figure 24: Preferential binding of QorA to a transcript fragment of *yffO*. Red and grey bars indicate addition of proteins and ^{32}P-labeled RNA, respectively. Addition of 11-fold molar excess of unlabeled competitor RNA is indicated within the gel image. **A)** EMSA comparing QorA affinity towards RNA^{yffO} and control transcripts of gene loci *rpsC* and *yhhA*. Unspecific RNA-binder YbiB was employed to ensure EMSA functionality. **B)** EMSA examining complex displacement behavior of RNA^{yffO} and the lower affinity *rpsC* control transcript. Concentrations: QorA (14 µM), YbiB (90 µM), labeled RNA^{yffO} (12 µM), labeled RNA^{rpsC} (12 µM), labeled RNA^{yhhA} (12 µM).
RNA-sequences (stated as GenBank entries NC_000913)**:** mRNA^{yffO}: 2562311-2562337, asRNA^{rpsC}: 3449339-3449365, mRNA^{yhhA}: 3587434-3587460.

trialing, no matter if the pH of the gel system was set below or above of QorA's pI-value (data not shown). Hence, to ascertain that the intensity of shifted RNA^{yffO}-QorA bands was not caused by unspecific effects like RNA-induced aggregation, free and complexed QorA was visualized by Coomassie-staining after native gel electrophoresis (Figure S 26), confirming slow migration irrespective of the presence of interacting RNAs. Taken together, presented results strongly suggest a specific interaction between QorA and a sequence within the transcript of *yffO*.

4.8.2 Affinity Quantification of the Specific RNA-Binding Activity of QorA

To obtain a reliable estimate for the *in vitro* affinity of RNA^{yffO} towards QorA, titrations were conducted using two separate methods, band shift titration and microscale thermophoresis (MST). Figure 25 shows the results of EMSA titration of radioactively labeled RNA. Incubation of RNA^{yffO} with up to 44 µM QorA (dimeric concentration) induced essentially quantitative complex formation, whereas the control RNA displayed a fairly weak band shift. Readout of band intensities and hyperbolic fitting of respective data points yielded a K_D of 7.4 µM for the RNA^{yffO}-QorA complex. Fitting of the control RNA dataset was not meaningful due to the low proportion of complex, and the K_D is estimated to be in a range of ≥100 µM based on extrapolation. The hyperbole-like progression of data points far from 100 % bound fraction in Figure 25a is an artificial effect stemming from technical limitations in band intensity readout accuracy and should not be confused with actual saturation, which is per definition only reached at 100 % fraction bound.

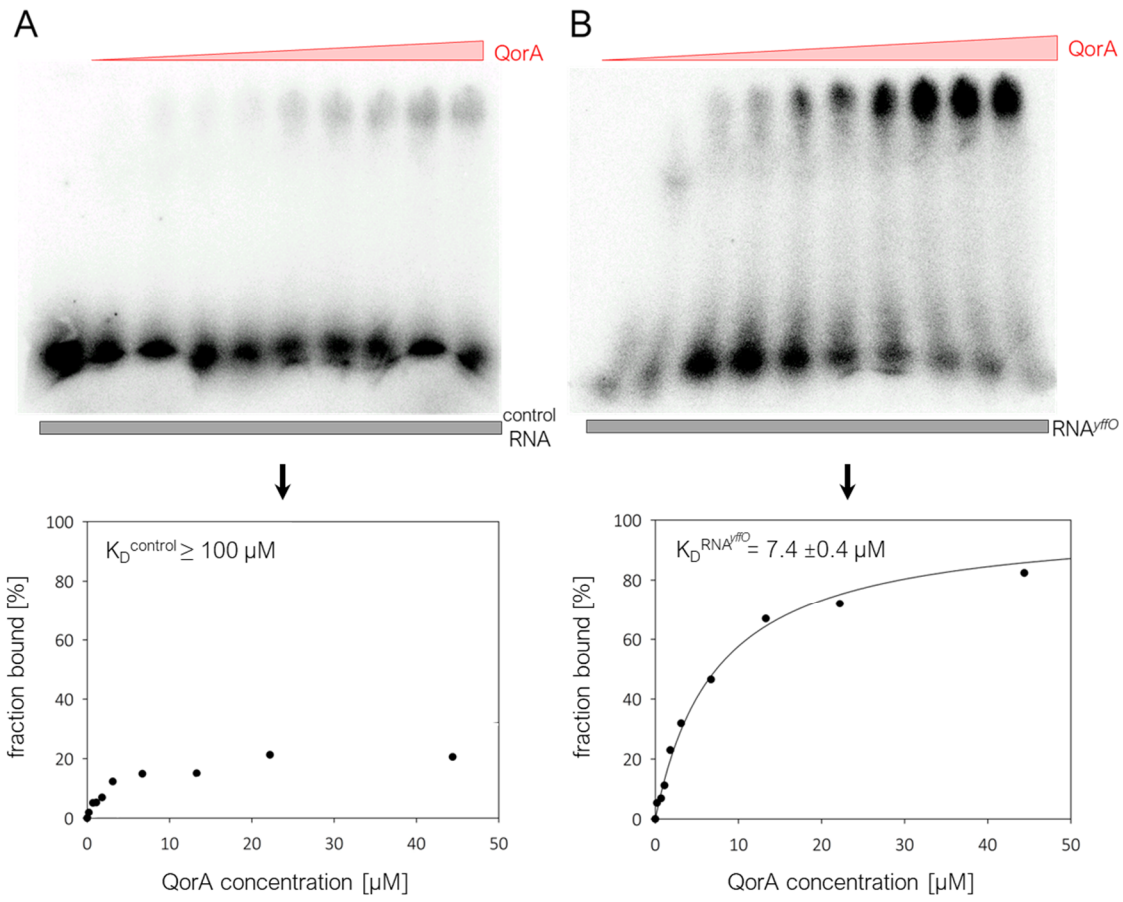


Figure 25: K_D determination for the RNA^{yffO}-QorA complex by mobility shift titration. Red and grey bars indicate addition of QorA and ^{32}P-labeled RNA, respectively. EMSA QorA titration of 350 nM control RNA (**A**) or 350 nM RNA^{yffO} (**B**). Dimeric QorA concentrations are stated in plots. Data points were fitted with a hyperbolic function using SigmaPlot v14 (see method section 3.4.5). The fit of data points in (**A**) was not constructive due to poor intensity ratio of complex band and background (a side effect of low binding) and is thus not displayed. K_D in (**A**) estimated based on extrapolation.
RNA-sequences (stated as GenBank entries NC_000913**):** RNA^{yffO}: 2562311-2562337, control RNA: 3080595-3080620.

MST experiments were conducted in order to cross-validate the K_D-estimation from EMSA titration. Cy5-labeled RNA was used as fluorescence reporter for complex formation. Figure 26 shows that titration of Cy5-RNA^{yffO} invokes larger changes in relative fluorescence when compared to the Cy5-labeled control RNA, which shows only noise signal up until ≥100 µM. Titration data points of RNA^{yffO} were fitted according to the law of mass action, deriving a K_D of 7.2 µM (dimer). For the control RNA, the lack of amplitude prevented a confident estimate for the K_D and indicates reduced affinity. The stoichiometry of the interaction could not be resolved, as this would require generation of a saturation curve employing Cy5-RNA^{yffO} concentrations >> K_D, and not enough labeled RNA was available. Taken together, affinity quantification by EMSA titration and MST titration both indicates a K_D in the low micromolar range for RNA^{yffO}, as well as significantly worse affinity for the control RNA.

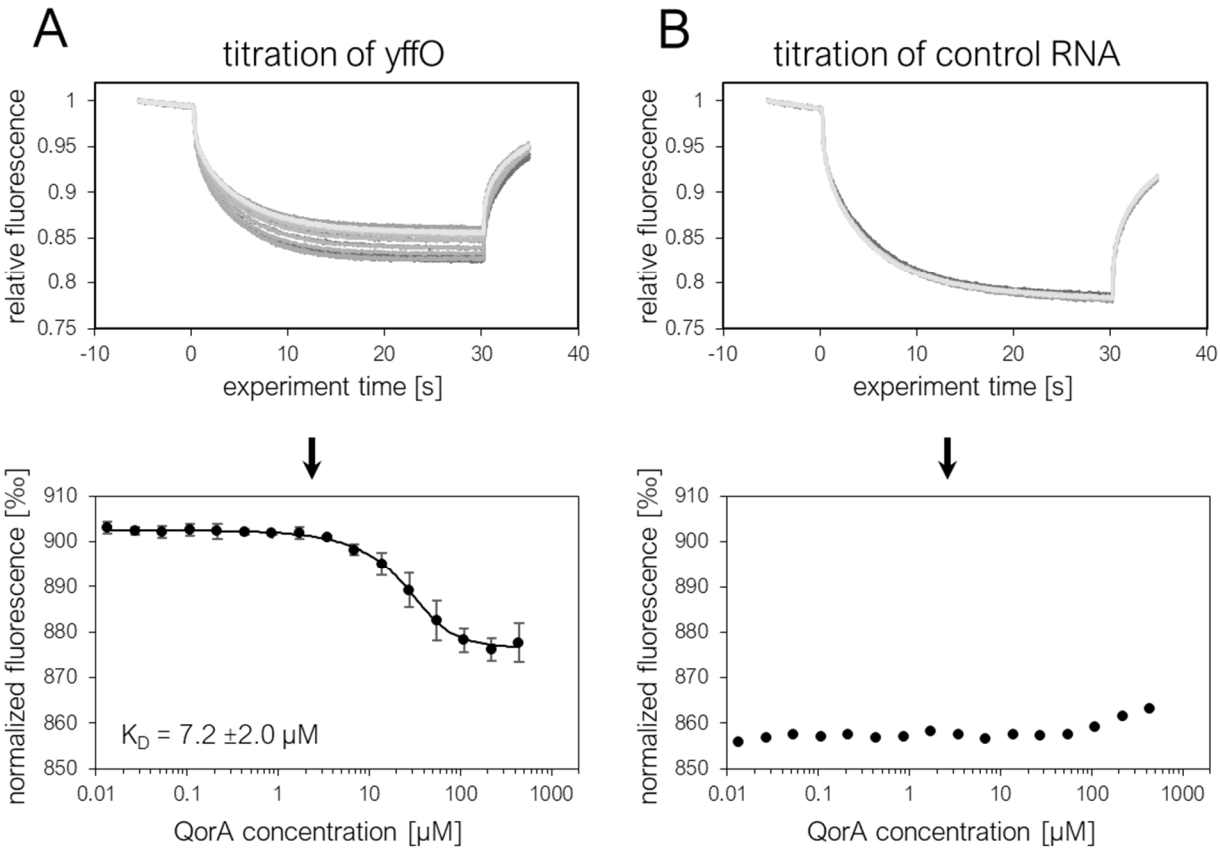


Figure 26: MST titrations for RNA^{yffO} and an unrelated control RNA. A) Relative fluorescence (670 nm) upon titration of RNA^{yffO} using QorA. **B)** Relative fluorescence (670 nm) upon titration of control RNA using QorA. For both measurements, 10 nM of Cy5-labeled RNA was titrated with QorA up to a final concentration of 440 µM (dimeric concentration). Data was fitted according to the law of mass action using NanoTemper proprietary software.
RNA-sequences (**stated as GenBank entries NC_000913**): RNA^{yffO}: 2562311-2562337, control RNA: 3080595-3080620.

4.8.3 Disruption of the RNA-QorA Complex by NADPH Binding

As elaborated in more detail in the biological background chapter (section 4.8.5), the Rossmann-fold domain of QorA is known to hold a NADPH-binding function (Thorn et al. 1995). Since the Rossmann-fold is suspected to confer RNA binding activity (Nagy and Rigby 1995), NADPH competition experiments were conducted. Binding disruption by NADPH was previously reported for the structurally homologous ζ-crystallin in complex with AU-rich RNA strands (Fernández et al. 2007).

Comparative MST titration showed that the fluorescence amplitude significantly decreases in presence of 10 mM NADPH (Figure 27a), not surpassing MST noise levels even at maximum QorA concentration (440 µM). This suggested a severely decreased affinity and rendered a fit for K_D estimation inappropriate. Mobility shift assays were in line with MST titration and confirmed that NADPH efficiently disrupts the RNA^{yffO}-QorA complex (Figure 27b). While the stoichiometry and uniformity of the interaction is not characterized, an involvement of the

cofactor-binding domain in RNA binding seems likely. Figure 27c shows an electrostatic surface map of the homodimeric QorA with bound NADPH based on the crystal structure 1QOR (Thorn et al. 1995), to illustrate where the NADPH disruption takes place. The adenine moiety of NADPH (green stick model, highlighted by green arrows) is visible at the cofactor binding site, which features positive surface charge. The NADPH molecule is deeply buried in the enzyme, and its nicotinamide group protrudes into the alleged active site, where it catalyzes a reduction reaction (active sites highlighted by black arrows and alternative view in Figure 27c). The positioning of NADPH is shown more precisely in Figure 27d. The Rossmann fold-domain (grey cartoon), featuring the typical array of six parallel β-sheets, is mainly responsible for NADPH

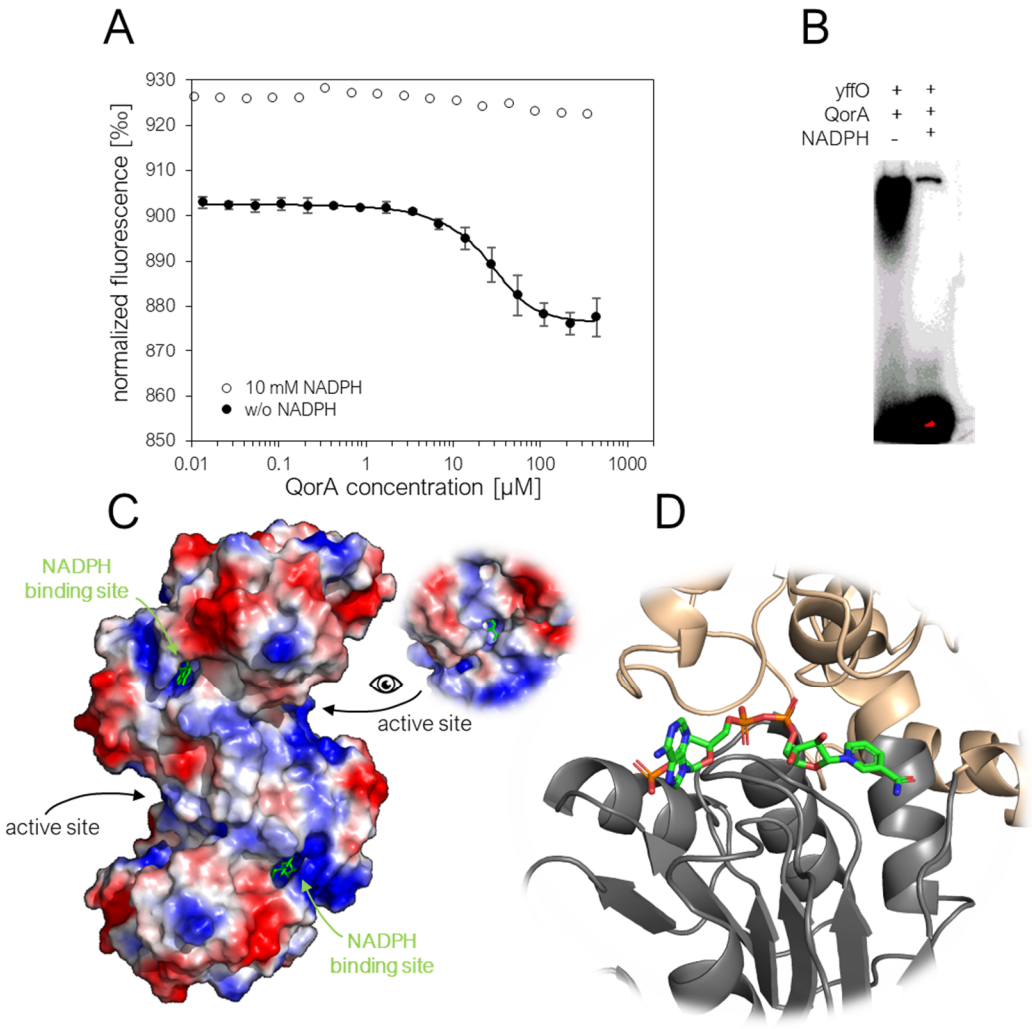


Figure 27: Impact of NADPH on RNA binding of QorA, and NADPH positioning in QorA. A: Comparative titration of 10 nM Cy5-labeled RNA^{yffO} with QorA (final concentration 440 µM) in presence and absence of 10 mM NADPH. Relative fluorescence measured at 670 nm. Data was fitted according to the law of mass action using NanoTemper proprietary software. The decreased signal upon addition of NADPH prevented fitting of respective data points. **B:** Mobility shifts surveilling RNA^{yffO}-QorA complex formation in presence and absence of 20 mM NADPH. **C:** Vizualization of dimeric QorA crystal structure (1QOR, Thorn et al. 1995). Red and blue indicates negative and positive electrostatic potential, respectively. NADPH is represented as green stick molecules, and their adenine moieties are visible within the NADPH binding pockets. A view into the substrate groove is provided (black arrow), and the nicotinamide moiety of NADPH protruding into the active site is faintly visible. **D:** Cartoon representation illustrating burial of NADPH (stick model) bound to QorA. Grey and beige color indicate Rossmann-fold and N-terminal domain of the protomer.

binding, even though the latter also requires some interactions to the N-terminal domain (beige). An RNA base mimicking the nucleobase moiety of NADPH is a reasonable assumption, but it is unclear whether it can account for the observed sequence specificity on its own. To the best of knowledge, resolved RNA-protein complexes involving similarly structured Rossmann fold domains, which could be used to get a picture of how the RNA might possibly contact the cofactor domain in QorA, are not available in literature. While aminoacyl-tRNA synthetases feature a Rossmann-fold, they contain an extra domain (CP1) attached to it that is involved in RNA-binding (Liu et al. 2012). For eukaryotic lactate dehydrogenases, it could be shown that DNA aptamers can bind with organism-tier specificity, and respective crystal structures revealed an interaction network which involves multiple direct contacts, water molecule-mediated interactions, and, most importantly, base flipping of hairpin or bulge loop-nucleobases into the cofactor groove (Cheung et al. 2013; Choi and Ban 2016). Since the latter is comprised by a Rossmann fold similar to that of QorA, it can be assumed that the fold is principally able to comprise a binding site for specific RNAs as well.

4.8.4 Motif Characterization for the Specific RNA-QorA Interaction

In order to investigate whether the high number of mapped reads found for RNA^{yffO} dictates that this particular RNA has the highest affinity among all fragments enriched during the QorA SELEX experiment, a comparative experiment with RNA^{narI} was conducted, which still showed significant enrichment but featured an 80-fold lower mapped read number compared to RNA^{yffO} (Table 14). While fragments RNA^{yffO} and RNA^{narI} are highly similar in the sequence predicted as motif by MEME (identical sequence stretch CAUAUGGAC, see Figure Figure 23c), the quality of the sequence motif and the low number of input sequences used to calculate it does not allow to view it as ascertained and exclusive affinity determinant. Notably, the motif is part of a stem-loop in RNA^{yffO} according to secondary structure prediction, while RNA^{narI} is showing only small probability for base pairing and is likely linear (see Figure 23c and Figure 23e).

Figure 28 shows that the complex band for RNA^{narI} is considerably weaker when compared to RNA^{yffO} (lane 3 vs lane 1). Furthermore, it exhibited identical displacement behavior after addition of either excess unlabeled control RNA or RNA^{narI} (lanes 4,5), challenging the existence of significant affinity discrepancies between the latter two. The displacement of a control RNA-QorA complex through addition of excess unlabeled competitor RNA showed a quantitative effect when adding excess RNA^{yffO} (lane 7), but not when adding either excess RNA^{narI} or excess control RNA (lanes 8,9). Taken together, there seems to be significant affinity discrepancy between RNA^{yffO} and RNA^{narI} despite the similar sequence motif, suggesting that the affinity of RNA^{yffO} might not be based solely on sequence specificity. Two conclusions are possible – the stem loop predicted for RNA^{yffO} but not for RNA^{narI} is a structural requirement for the interaction, or further sequence stretches within RNA^{yffO} outside of the predicted motif locus are also affinity determinants.

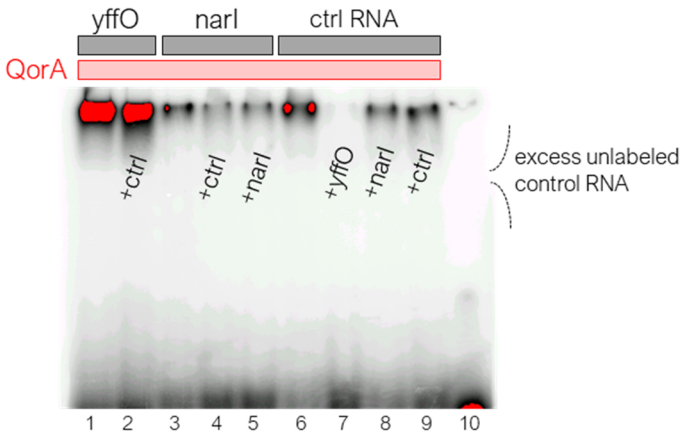


Figure 28: Comparison of QorA affinity towards RNA^{yffO} and RNA^{narI}. Red and grey bars indicate addition of proteins and ^{32}P-labeled RNA, respectively. Addition of 10-fold molar excess of unlabeled competitor RNA is indicated within the gel image. Concentrations: QorA (17 µM), labeled RNA^{yffO} (7.5 µM), labeled RNA^{narI} (7.5 µM), labeled control RNA (7.5 µM). Unbound RNA partially migrated out of gel. Red signalizes oversaturation.
RNA-sequences (**stated as GenBank entries NC_000913**): RNA^{yffO}: 2562311-2562337, RNA^{narI}: 1286287-1286313, control RNA: 3080595-3080620.

To further explore these conclusions conclusion, altered RNA fragments based on RNA^{yffO} were tested (Figure 29): For the purpose of examining the relevance of the very last nucleotides 3'-adjacent to the stem-loop, the sequence was shifted five nucleotides in 5' direction (yffO-s). The motif sequence located on the loop was completely altered to probe if the interaction is based purely on loop-sequence recognition (yffo-sm). Additionally, the original stem-loop was completely disrupted by cutting of nucleotides at the 5' side (yffo-trunc5', the stem-loop predicted for this sequence sits at the 3' end and is not to be confused with the original stem-loop). Finally, the original stem-loop was reduced to its bare minimum by cutting of any nucleotides at the 3' side (yffo-trunc3').

Figure 29b shows that RNA$^{yffO\text{-}s}$ and RNA$^{yffO\text{-}sm}$ principally bind to QorA (lanes 3,6). However, while excess unlabeled RNA^{yffO} is able to erase the complex band of RNA$^{yffO\text{-}s}$ (lane 5), excess RNA$^{yffo\text{-}s}$ shows little effect on RNA^{yffO} complex formation (lane 4). Furthermore, displacement of RNA$^{yffO\text{-}sm}$ by excess RNA$^{yffO\text{-}s}$ (lane 7) is more efficient than the *vice versa* displacement of RNA$^{yffO\text{-}s}$ by excess RNA$^{yffO\text{-}sm}$ (lane 8). This suggests an affinity breakdown RNA^{yffO} > RNA$^{yffO\text{-}s}$ > RNA$^{yffO\text{-}sm}$, implicating that both the loop sequence and the sequence 3'-adjacent to the stem-loop are influencing QorA affinity. The affinity assessment could be confirmed in a separate competition experiment (Figure 29c), showing that an excess of both control RNA and RNA$^{yffO\text{-}sm}$ induces stronger displacement in a QorA-RNA$^{yffO\text{-}sm}$ (lanes 6,8) than in a QorA- RNA$^{yffO\text{-}s}$ complex (lanes 2,4).

In line with the conclusion that both features are affinity determinants, RNA$^{yffo\text{-}trunc3'}$, featuring a completely alternative stem-loop, and RNA$^{\text{yffO-trunk5'}}$, which is truncated even more than RNA$^{yffO\text{-}s}$ at the 3' side, showed greatly reduced complex formation. (Figure 29b, lanes 9,10). Separate competition experiments (Figure 29d) did also confirm this trend, with RNA$^{yffo\text{-}trunc5'}$ (lanes 7,8,9) notably being more impaired than RNA$^{yffo\text{-}trunc3'}$ (lanes 4,5,6).

To summarize, the experiments with altered RNA fragments seem to indicate that both the stem-loop, but also the 3'-adjacent linear nucleotides are defining the affinity of the 27 nt RNA^{yffO} fragment towards QorA. While the complexity of the interaction, allegedly relying on both structural and sequence context, perfectly explains read number and affinity discrepancies observed between RNA^{yffO} and RNA^{narI}, it would suggest a highly specialized interaction between QorA and *yffO* mRNA. Further EMSA experiments testing sequences with sparingly introduced mutations at the stem-loop or the 3'-adjacent site were conducted but turned out ambiguous thus far (data not shown), suggesting that an entirely exact sequence breakdown of the potential binding motif is non-trivial. In summary, it could be deduced that both the predicted stem-loop and the 3'-adjacent sequence are potential mediators of the specific interaction, with the stem-loop being an unconditional requirement.

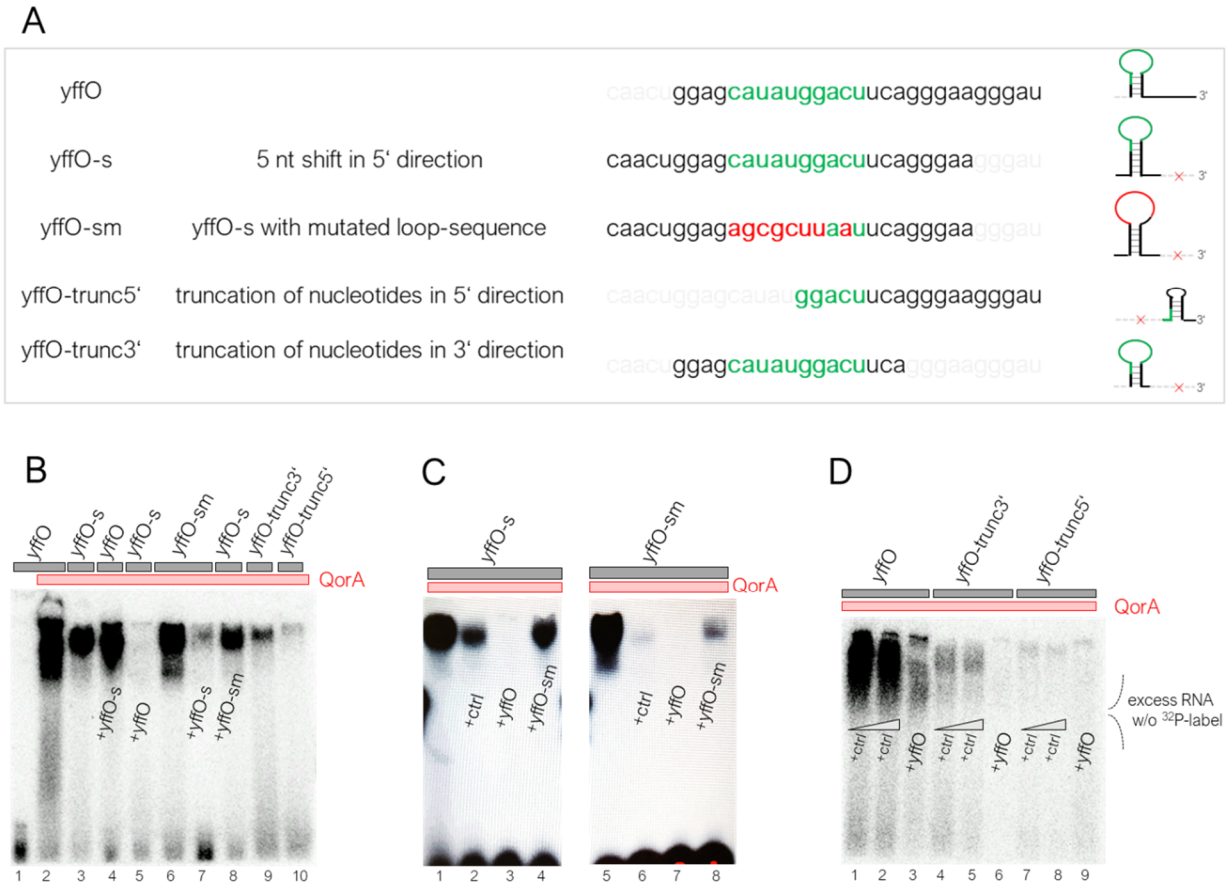


Figure 29: Experiments conducted to break down the RNA^{yffO}-QorA binding motif. A) Overview of altered RNA fragments (yffO-s, yffO-sm, yffO-trunc5', yffO-trunc3') based on the original transcript fragment found to bind preferably to QorA (yffO). **B)** 8 µM of radiolabeled RNA fragments (grey bars) were subjected to gel electrophoresis in presence of 15 µM QorA (red bar) and 220 µM unlabeled RNA (where indicated in the gel). **C)** 2 µM of radiolabeled RNA fragments (grey bars), 11 µM QorA (red bars), 100 µM unlabeled RNA (where indicated in the gel). Ctrl RNA-sequence (stated as GenBank entry): NC_000913:3080595-3080620. **D)** 2 µM of radiolabeled RNA fragments (grey bars), 10 µM QorA (red bars), 44-110 µM unlabeled RNA (where indicated in the gel; triangle indicates 44 and 110 µM in adjacent lanes, no triangle: 110 µM). Ctrl RNA-sequence (stated as GenBank entry): NC_000913:3080595-3080620. Unbound RNA partially migrated out of gel.

4.8.5 Biological Background Information on QorA

Due to the identification of a specific RNA-binding activity for QorA, a more elaborate description of the enzyme is provided in this subsection.

QorA belongs to the medium-chain dehydrogenase/reductase (MDR) superfamily, an ancient superfamily present in the last universal common ancestor (Riveros-Rosas et al. 2003). The MDR superfamily occurs in all domains of life and contains a very diverse set of enzymatic activities, including alcohol dehydrogenases, quinone oxidoreductases, enoyl reductases, sorbitol dehydrogenases, cinnamyl reductases, threonine dehydrogenases, and numerous others. MDRs typically consist of two domains, a Rossmann fold domain conferring cofactor-binding and a substrate binding domain comprised of a core of antiparallel β-strands and surface-positioned α-helices (Hedlund et al. 2010). All representatives utilize either NADH or NADPH as cofactor, and several but not all enzymes feature a catalytic zinc atom at the active site (Nordling et al. 2002), such as classical alcohol dehydrogenases in animals and plants. Some MDRs feature a second zinc atom, binding to a structural site to stabilize an external loop, possibly affecting oligomerization state (Persson et al. 1994). Most MDRs form either dimers or tetramers. It has been attempted to categorize all MDRs into families and subfamilies, and while exact categorizations keeps changing with newly identified MDR representatives (Nordling et al. 2002; Riveros-Rosas et al. 2003; Knoll and Pleiss 2008; Hedlund et al. 2010), they all agree on a principal separation of zinc-binding and non-zinc-binding MDRs.

ζ-Crystallin is an eukaryotic representative of a non-zinc-binding, NADPH-binding MDR with quinone oxidoreductase activity. The name refers to the protein's occurrence in the eye lens of mammals (Huang et al. 1987), and while the relevance of its enzymatic activity in the lens is unclear, truncated ζ-crystallin has been shown to cause cataract (Rodriguez et al. 1992). The abundance of the protein is highly divergent, ranging from 10 % of crystallin in guinea pig and camel to only catalytic amounts in humans, owing to an additional, tissue-specific promoter that causes high expression in the lens and is not present in all organisms (Huang et al. 1987; Garland et al. 1991; Gonzalez et al. 1994b). ζ-Crystallin is referred to as moonlighting player, allegedly administering different cell functions. At high levels in the eye lens, the protein might function mainly as structural protein despite its catalytic properties, or generally preserve a reducing environment based on its NADPH binding activity (Bazzi 2001). The protein occurs in catalytic amounts in tissues other than eye lens in various organisms (Rao and Zigler 1992; Gonzalez et al. 1993; Gonzalez et al. 1994a), and has been shown to reduce *ortho*-quinones by catalyzing a one-electron transfer, generating a semiquinone radical in the process (Rao et al. 1992). Generally, the full substrate spectrum across homologous quinone oxidoreductases in different species is unknown, considering the existence of numerous quinone derivates. ζ-Crystallin also carries alleged RNA-binding functions. The RNA-binding activity was first proposed after isolation of the protein from cytosolic extracts of rat renal cortex using a biotinylated oligo with AU-rich sequence (Tang and Curthoys 2001). Further studies observed binding of ζ-crystallin to the ARE-consensus sequence $A(UUUA)_5$ as well as $poly(U)_{14}$, and proposed an involvement of the Rossmann fold domain in RNA binding (Fernández et al. 2007; Porté et al. 2009). ζ-Crystallin

supposedly stabilizes glutaminase mRNA and *bcl*-2 mRNA by binding to AREs, affecting expression levels of respective gene products (Ibrahim et al. 2008; Lapucci et al. 2010).

Proteins homologous to ζ-crystallin are found in bacteria as well. Based on the alleged RNA-binding function of ζ-crystallin, the *E. coli* pendant QorA was chosen as RBP candidate in this work. QorA shares 31 % sequence identity to ζ-crystallin and shows significant residue conservation at the NADPH binding site as well as the N-terminal substrate binding domain (Thorn et al. 1995). While enzymatic activity of QorA is not confirmed, examination of the crystal structure suggests that the substrate binding pocket would fit large substrates like naphtoquinones or phenanthrenequinones (Thorn et al. 1995; Sulzenbacher et al. 2004), species that are known to be metabolized by the homologous eukaryotic ζ-crystallin (Rao et al. 1992). Besides a crystal structure reported by Thorn et al. (1995), no research on QorA has been published so far. The proteome of *E. coli* maintains a second soluble quinone oxidoreductase with 16 % sequence identity to QorA, named YhdH, for which a crystal structure has been reported as well (Sulzenbacher et al. 2004). YhdH also binds NADPH and features a narrower substrate binding crevice, suggesting that it might metabolize smaller species like benzoquinones.

A *Staphylococcus aureus* quinone oxidoreductase (*sa*QorA) which is homologous to QorA and ζ-Crystallin has been characterized (Maruyama et al. 2003). The study observed turnover of 9,10-phenanthrenequinone to the corresponding semiquinone radical in a one-electron reduction process. Since O_2-induced, spontaneous reoxidation of the semiquinone radical leads to the formation of hazardous superoxide anions (Cadenas 1995), the role of one-electron reducing quinone oxidoreductases like *sa*QorA or ζ-crystallin in stress response is not well understood. However, a study reported decreased growth of yeast cells with deleted ζ-crystallin when supplementing menadione or hydrogen peroxide (Fernández et al. 2007), attesting a definite beneficial effect. While subject to speculation, it is possible that these enzymes provide a physiological effect by a property other than its quinone reductase activity, and the RNA-binding activity is hinting towards posttranscriptional regulation. Due to time limitations, a comprehensive examination of putative catalytic activities of *ec*QorA was not performed within this work.

4.8.6 Biological Background of *yffO*, a Grounded Prophage Gene

The RNA preferentially bound by QorA as described in chapter 4.8 is a transcription fragment from *yffO*, one of nine genes constituting the grounded prophage operon CPZ-55 in *E. coli*. Grounded prophages (as opposed to active prophages) are DNA stretches permanently integrated into the host genome due to mutation-induced phage entrapment in the lysogenic cycle. Prophages contribute to horizontal gene transfer in bacteria and are interesting from an evolutionary perspective. They seemingly invoke disadvantages based on the additional metabolic burden of integrated DNA and ultimately by the eventual lysis of the host. However, the prophage can also carry functions that increase the fitness of the host, like immunity to secondary phage infection or physiological benefits (Canchaya et al. 2003). Such beneficial traits are

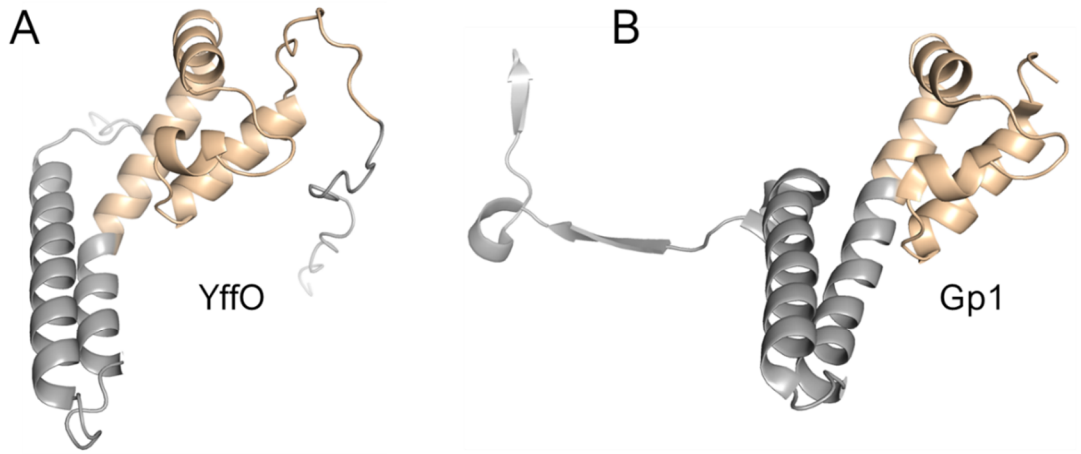
plausible due to the interests of prophage and host being at least partially aligned during the lysogenic phase (Touchon et al. 2016), and they provide an explanation for the permanent retention of grounded prophage genes. The frequency of phage-DNA in bacterial genomes is highly variable. A study pursued by Touchon et al. (2016) showed that 46 % of 2110 analyzed bacterial genomes carried at least one prophage, and the latter account for an average of 3.1 % of the bacterial genomes, up to 18 %. Furthermore, the analysis of the different strains concluded that no particular pattern of serial prophage acquisition is to be observed in *E. coli* and *Shigella* strains, suggesting horizontal propagation in between them.

The occurrence of the grounded prophage CPZ-55 is limited to certain species of the family of *Enterobacteriaceae.* CPZ-55 presumably originates from a P22-like phage type (Casjens 2003). In *E. coli* K-12*,* CPZ-55 has a total size of 6,782 nucleotides and is located between ethanolamine catabolism genes *eutA* and *eutB*, presumably due to a former transposon-like integration event (Touchon et al. 2016). CPZ-55 carries nine genes, the putative phage integrase-coding *intZ* as well as eight hypothetical, virtually uncharacterized genes *yff[L-S].* While the operon is not essential for cell growth, it could be shown that a ΔCPZ55 *E. coli* strain confers slightly confined growth rates in LB-medium and M9C minimal medium alike (Wang et al. 2010). Moreover, the study identified *via* screening that stressors like nalidixic acid, H_2O_2, or low pH can severely impede the cell survival of *E. coli* cells carrying deletions of certain prophage operons relative to wildtype cells. Even if the effects of such stressors were marginal on the ΔCPZ55 strain in particular, these results suggest the existence of biological functions for grounded prophages in bacteria, in line with the notion that 'junk' prophages would lose the evolutionary arms race due to the metabolic burden of extra DNA.

The function of the *yffO*-encoded hypothetical 15.7 kDa protein is unknown. Due to the gene originating from a temperate phage, it is self-evident that it was once involved in phage physiology. A amino acid sequence analysis using BLAST yielded almost exclusively hypothetical proteins, but also the small terminase subunit gp1 from *Shigella* phage Sf6 as a related protein (24.9 % identity, 33.7 % similarity according to a pairwise sequence alignment generated by EMBOSS Needle). Figure 30a-b illustrates an AlphaFold prediction of the YffO structure as well as a published crystal structure of Gp1, suggesting structural homology between the proteins. The terminase of bacteriophage Sf6 packs newly synthesized concatemeric phage DNA into capsid precursors during virus morphogenesis (Zhao et al. 2012). The small terminase subunit Gp1 is responsible for the specific recognition of the concatemeric DNA *via* the helix-turn-helix motif (Zhao et al. 2010), and the large subunit cleaves the concatemeric DNA and uses ATP hydrolysis to direct it into the procapsid.

The structural homology of helix-turn-helix motifs of YffO and Gp1, as well as an N-terminal AT-hook motif G-R-P (residues 24-26, Figure 30c) provide strong indication that YffO has retained its DNA-binding properties. Notably, the surrounding of the AT-hook motif is more enriched in basic residues for YffO than for Gp1 (PKKR**GRP**AK vs. EPKA**GRP**SD). While YffO might have historically served as small terminase subunit for virus morphogenesis, the protein has by implication lost its function during the transformation of CPZ-55 bacteriophage

into a permanently grounded prophage. As functionless genes are considered metabolic burden that are evolutionary disfavored, the retention of the whole CPZ-55 operon might imply that associated genes, among them *yffO*, carry out functions beneficial to the bacterial host. This interpretation is supported by the gene being found to be upregulated by the transcriptional master regulator $FlhD_2C_2$, which is vital for the regulation of genes involved in flagellar biosynthesis and chemotaxis (Stafford et al. 2005). Reportedly, the region upstream of *yffO* which resembles the $FlhD_2C_2$ consensus recognition sequence is bound by $FlhD_2C_2$ in the low nanomolar range according to *in vitro* assays and is able to significantly increase the expression levels of a fused *lacZ* reporter gene. Run-off micro-array experiments have shown that the corresponding promoter region affected by $FlhD_2C_2$-binding has higher affinity towards RNA polymerases carrying the σ^{38}-subunit as opposed to the σ^{70}-subunit (Maciag et al. 2011). The consensus motif of both subunits differs slightly, and σ^{38}, which is mostly active during slow growth, is associated specifically with the transcription of genes involved in responses to oxidative stress or nutrient starvation (Tanaka et al. 1993).



C

```
YffO      1 MEGELIENNGLDIYDTSETPKKRGRPAKYNEKIATQIVLLVSEGYSLRKI   50
                     .....|..|.|||:.|..::|..|..|:|.|.||.|:
Gp1       1 -------------GSHMATEPKAGRPSDYMPEVADDICSLLSSGESLLKV   37

YffO     51 SMMPGMPSHRQMMRWQLEHMDFREGIAWMSWLWCAEAGRRAVEIIDEVDI  100
            ...||||....:.||..:|.|||:     .:....||  ||..|.:|:
Gp1      38 CKRPGMPDKSTVFRWLAKHEDFRD-----KYAKATEA--RADSIFEEI--   78

YffO    101 NAEDGPKQLRKAEAKAKALL-------AAAKLNSLKHSPFGDDKQ-----  138
             .|.....:..|...|||.|        |.|::|..|   :||...
Gp1      79 -FEIADNAIPDAAEVAKARLRVDTRKWALARMNPRK---YGDKVTNELVG  124

YffO    139 ------------------    138

Gp1     125 KDGGAIQIETSPMSTLFGK    143
```

Figure 30: Structures and sequence alignment of hypothetical protein YffO and Gp1. A: AlphaFold prediction of the structure of YffO. Suspected DNA-binding domain colored in beige. **B:** Crystal structure of Gp1 (PDB entry: 1HEF, Zhao et al. 2010). DNA-binding domain colored in beige. **C:** Pairwise sequence alignment between YffO and Gp1. Green bar indicates the AT-hook motif followed by the helix-turn-helix motif for both sequences.

5 Perspectives and Outlook

5.1 Contextualization of Results with Respect to the REM Hypothesis

Regulatory interplay of RNAs and metabolic enzymes is frequently hypothesized based on either high-throughput screenings that capture interactomes of RNA, or based on observations of physiological, RNA-related effects induced by enzymes, like impact on specific mRNA levels. Taking glycolysis and TCA cycle as example, essentially all eukaryotic enzymes from those pathways have been successfully identified in one or more poly(A) RNA interactome studies over time (Albihlal and Gerber 2018). Especially for glycolytic enzymes, the list of possible RNA-related moonlighting functions is ever-growing and becomes hard to summarize - the reader is referred to a recent review of Wegener and Dietz (2022). Putting that data availability into perspective, precise biochemical research characterizing the exact nature of hypothesized RNA-enzyme interactions is relatively scarce and not found at all for many enzymes – do interactions occur at defined binding sites? How high is the affinity, and to what extent do the enzymes discriminate between RNA sequences? Can interactions occur on a standalone basis or are they part of larger complexes ('metabolon')? Are enzymes affecting RNA-fate or affected by RNA? It can be argued that the only thoroughly understood metabolic enzyme with RNA-related moonlighting function is aconitase, featuring ample biochemical data including structural data of the protein-RNA complex (Walden et al. 2006). For example, the binding site of eukaryotic GAPDH (often referred to as model RNA-interacting metabolic enzyme and featuring many proposed moonlighting functions[11]) is still elusive and discussed to coincide with cofactor binding site, substrate groove, dimer interface, or a combination of the latter (Garcin 2019). Even though it has been concisely established that the enzyme can bind RNA with nanomolar K_D-level (White et al. 2015), overall, biochemical understanding on how GAPDH possibly effects a multitude of RNAs in specific fashion is shallow. For less popular case examples of allegedly RNA-interacting enzymes, protein-centered characterization is often not available. A recent study conducted by Vaishali et al. (2021) opportunistically touched upon this issue by characterizing the RNA-binding properties of six enzymes suggested to be RBPs based on reported mRNA interactome capture. It used NMR-spectroscopy and biochemical assays to conclude that binding was either absent or nonspecific, and advised caution when embarking RNA binding studies involving 'unconventional, novel RBPs'. This case report highlights the need of more protein-centered investigations.

Experimental results in this work contribute to the particulary scant protein-centered data available for prokaryotic metabolic enzymes and might also help answering if RNA-binding is conserved in homologous enzymes across domains of life. Genomic SELEX probed whether the candidate enzymes are able to discriminate sequences in the RNA pool of *Escherichia coli*, and

[11] Example references of proposed functions: Zhou et al. (2008), Rodríguez-Pascual et al. (2008), Backlund et al. (2009), Kondo et al. (2011), Ikeda et al. (2012), Chang et al. (2013), Millet et al. (2016), Xu et al. (2016), Lin et al. (2017).

selected follow-up validation verified associated affinity and specificity. This targeted approach opposes interactome data generation which identifies lists of proteins interacting with unknown parts of a whole transcriptome. By choosing candidate enzymes that occurred in one or more interactome datasets, one side goal was to cross-validate existing interactome data. Considering that SELEX employs candidate proteins in their isolated state *in vitro*, it should be noted that resulting data might not hint on complex moonlighting functions that require specific *in vivo* environment. Nevertheless, the genomic SELEX data gives useful information on which of the enzymes might engage in physiologically relevant, direct protein-RNA contacts, even if a higher, native affinity might be obscured by improper conditions. To complement genomic SELEX and extract RNA-binding information *in vivo*, iCLIP was conducted. Due to timely constraints, iCLIP could only be conducted for a subset of all enzymes employed in genomic SELEX, and completion of the iCLIP dataset would be a valuable continuation of this work.

5.2 Condensed Review of Results

A total of 23 proteins were characterized by genomic SELEX, among them 20 metabolic enzymes from Escherichia coli. iCLIP was conducted for 8 of those metabolic enzymes, and for 6 of them iCLIP experiments were carried out up until evaluation of sequencing libraries. Table 15 provides a simplified, condensed summary of genomic SELEX results of proteins that showed significant RNA enrichment.

Notably, this included three enzymes from glycometabolism, namely PykF, Pgk, and GapA. RNAs enriched by PykF were among the most ambiguous to interpret, featuring no communal sequence or structural feature nor a communal biological context, and displaying no coinciding binding of these RNAs in the PykF iCLIP experiment. As neither iCLIP nor EMSA experiments could provide additional evidence for preferential binding, it cannot be excluded that the heterogeneous set of RNAs was enriched in SELEX either randomly or based on insignificant affinity fluctuations. Pgk selected A-rich sequences embedded into stem-loops, albeit featuring high variation in loop size and composition. To the best of knowledge, this is the first report of a potential specific RNA interaction for prokaryotic phosphoglycerate kinase. Strikingly, preferred binding to poly(A) RNA was reported previously for a chloroplast phosphoglycerate kinase (Lin et al. 2007), which is structurally homologous to *ec*Pgk and also adopts a Rossmann fold. This might point towards an evolutionary conserved, possibly ancient relation of phosphoglycerate kinase and poly(A) RNA species, considering the bacterial descent of chloroplasts (Miyagishima 2005). However, the interaction between Pgk and enriched stem-loop structures should definitly be interpreted carefully, considering the lack of specificity and general affinity observed in EMSA analysis, and also the lack of affirmative iCLIP results for Pgk. A more opportunistic explanation of the affinity towards an adenosine-rich loop is comprised by the obvious chemical resemblance of cofactor ATP, though such analogous correlation was not found in SELEX results for Adk and adenosines, or for Upp and uraciles. SELEX employing GapA, which also adopts a Rossmann fold domain, enriched RNAs containing AU-rich sections. Owing to the strong heterogeneity of the sequences, no further characterization was pursued.

Still, the result lines up with eukaryotic GAPDHs generally being characterized as AU-binding protein based on the high number of affirming reports (Wegener and Dietz 2022). The result also lines up with a recent study by Chu et al. (2022) reporting recovery of *Staphylococcus aureus*

Table 15: Summary of noteworthy results from SELEX[a]

protein	enriched RNAs in genomic SELEX	specific binding in *in vitro* analysis[b]	description of enriched RNA features
Pyruvate kinase (PykF)[c]	✓	✗	various mRNAs / asRNAs with no apparent communal attribute
Phosphoglycerate kinase (Pgk)[c]	✓	✗	poorly defined hairpins, showing A-rich loops of variable length as common feature
Glyceraldehyde-3-phosphate dehydrogenase (GapA)	✓	no data	poorly defined, AU-rich sequence stretches
Enolase (Eno)	✗		-
Phosphoglycerate mutase (GpmA)	✗		-
Transaldolase A (TalA)	✗		-
Transaldolase B (TalB)	✗		-
Ribose-5-phosphate isomerase (RpiA)	✗		-
2-dehydro-3-deoxyphospho-octonate aldolase (KdsA)[c]	✗		-
Uracil phosphoribosyl-transferase (Upp)[c]	✗		-
Adenylate kinase (Adk)	✗		-
Thymidylate synthase (ThyA)	✓	✗	hairpin with strictly conserved sequence GC(AGA)GU (brackets indicating triloop)
L-asparaginase (AnsB)[c]	✗		-
Glutamate-5-kinase (ProB)	✓	✓	small number of specific fragments (two mRNAs and one asRNA)
Malate dehydrogenase (Mdh)[c]	✗		-
Aconitase (AcnB)	✓	no data	poorly defined stem-loops with 3'-adjacent U-rich stretches
Isocitrate dehydrogenase (Icd)	✗		-
Superoxide dismutase (SodA)	✗		-
Quinone oxidoreductase (QorA)	✓	✓	few mRNAs / asRNAs with a potential motif (see Figure 23b). The motif being part of a stem-loop and further sequence-features 3'-adjacent to this stem-loop seem to be affinity determinants based on *in vitro* mutational studies of the *yffO* mRNA fragment.
Pyridoxine/pyridoxamine-5-phosphate oxidase (PdxH)	✗		-
YbiB[d]	✓	no data	three mRNAs and one asRNA with a potential motif GAAGCC
YggX[d]	✗		-
positive control MS2	✓	✓	bacteriophage operator hairpin-like motif NANN(ANCA)N'N'N' (N and N' indicate complementary bases, brackets indicate loop)

[a] The color code refers to the metabolic pathway the enzymes belong to, see Table 5.
[b] ✓ - successful verification of specific binding in EMSA; ✗ - no binding or specific binding visible in EMSA; 'no data' – no EMSA experiments conducted due to the enriched RNA features being overly heterogenous across clusters.
[c] iCLIP was conducted for these enzymes, and resulting libraries showed no specific RNA enrichment.
[d] unknown function – possibly no metabolic enzymes, listed for completeness.

GapA in a phenol-toluol extraction interactome capturing experiment (PTEX), mutually suggesting that prokaryotic GAPDHs might exert RNA-related moonlighting functions as well. However, the nature of such potential functions remains unidentified.

The remaining two enzymes from glycometabolism employed in genomic SELEX, Eno and GpmA, showed no specific RNA enrichment. Eno is reported to participate in formation of the RNA degradosome (Chandran and Luisi 2006). The crystal structure of dimeric Eno in complex with the minimal binding segment of RNase E (PDB accession 2FYM) revealed that the small RNase E fragment extending into the cleft between N-terminal domains of the individual Eno protomers is distant from the catalytic RNase domain. Eno was also shown to have no activity impairment when engaging in the RNA degradosome (Py et al. 1996). This makes it plausible that Eno attends a moonlighting function in the degradosome that does not involve direct interaction with RNA (like metabolic sensoring), and the lack of RNA enrichment during SELEX is in good agreement to that hypothesis. However, it opposes the alleged interactions of eukaryotic enolase with RNAs (Shchepachev et al. 2019; Huppertz et al. 2022). Reviewing examined glycolytic enzymes with respect to their performance in interactome screenings listed in Table 5, no strict correlation was apparent. Eno and GpmA with no notable SELEX results were recovered by TRAPP and OOPS with similar significance as PykF, Pgk, and GapA. PTEX observed an increased recovery in crosslinked samples over non-crosslinked samples only for Pgk.

Mdh, AcnB, and Icd were enzyme candidates selected from citric acid cycle. Out of the three, only AcnB SELEX showed RNA enrichment deviating from the control experiment. Notably, the experiment did not validate the reported feedback interaction between AcnB and the 3' UTR of *acnB* mRNA (Tang and Guest 1999). However, enriched RNAs were identified carrying structure-sequence features similar to those in the 3' UTR of *acnB* mRNA, namely stems with 3'-adjacent U-rich sequences. This finding indirectly agrees with the postulated feedback interaction and might be a first hint that AcnB is able to not only interact with its own mRNA, but also with additional target RNAs *in vivo*.

For ThyA, SELEX did not validate the alleged feedback interaction with *thyA* mRNA (Voeller et al. 1995), but yielded unambiguous enrichment of a stem-loop motif accounting for a significant proportions of total reads. The motif, comprising a AGA-triloop and two adjacent base pairs, resembles the hairpin ending of an iron-responsive element (IRE) in human amyloid-beta-precursor mRNA (Cho et al. 2010), albeit not extending to the conserved stem typical to an IRE. The relevance of this resemblance cannot be judged and is uncertain, as IREs are considered to play functional roles exclusively in metazoans (Piccinelli and Samuelsson 2007). While efforts to validate this interaction *in vitro* failed within this work for unknown reasons, the IRE-like RNA enrichment might possibly hint towards a general suitability of such stem-loop elements to confer specific interactions, considering its apparent widespread occurrence across domains of life – though this is subject to speculation.

For ProB, RNAs enriched in SELEX were successfully tested for binding *in vitro*, showing greatly improved affinity compared to control RNA in EMSA analysis. However, the observed affinity was estimated to be way too low for biological relevance. The result *per se* is intriguing, as the

PUA domain of ProB is known to confer RNA-binding in many different proteins, like e.g. MCT-1-like proteins or Nip7-like proteins (Reinert et al. 2006; Coltri et al. 2007). The domain occurs in RNA modifying enzymes and ribonucleoproteins (Pérez-Arellano et al. 2007). A survey on the PUA domain examining 20 species scattered across domains of life revealed that glutamate-5-kinase, while present in all examined organisms, contained a PUA domain only sporadically, distributed across eukaryotes and prokaryotes alike (Cerrudo et al. 2014). While the PUA domain was shown to impact glutamate kinase activity, it is not a requirement for the latter, possibly indicating an unidentified moonlighting function, and sequences identified through SELEX might provide a first hint. *Bacillus subtilis proB* was shown to exert control over the expression of other genes (Ogura et al. 1994; Ogura and Tanaka 1996), supporting the possibility of a physiological RNA-binding function of the PUA domain. Further investigation is recommended.

QorA SELEX strongly enriched RNAs carrying a transcript fragment of prophage gene *yffO*. The interaction was quantified by EMSA titration and MST titration, which agreed on a $K_D \approx 7$ µM. Based on a large surplus in reads containing the *yffO* mRNA fragment and affinity comparisons with another enriched RNA with fewer reads, the current working hypothesis is that the interaction might be highly specific towards the *yffO* mRNA fragment, and the remainder of identified RNAs were enriched to lower extent based on similarity. However, not all enriched RNAs were probed by *in vitro* experiments. While the K_D is moderate with respect to physiologically relevant interactions, which can display K_D-values in the low nanomolar range, it is not unusual that higher affinity is only achieved through specific circumstances like the cooperative effect of multiple interaction partners, while K_D-values of isolated domains are in the micromolar or even millimolar range (Bertini et al. 2012). To the best of knowledge, the identified RNA-binding specificity of QorA is the first report of RNA-binding for prokaryotic quinone oxidoreductases. Homologous quinone oxidoreductases from eukaryotes (ζ-crystallin) feature multiple reports of RNA-binding, including AU-rich or U-rich binding observed *in vitro* (Fernández et al. 2007; Porté et al. 2009), and stabilization of specific mRNAs (glutaminase mRNA, *bcl*-2 mRNA) by ARE-binding leading to post-transcriptional control (Ibrahim et al. 2008; Lapucci et al. 2010). While the sequence discovered for QorA does not match reported ARE-binding, it still indicates that the RNA-binding property of the quinone oxidoreductase subfamily of MDR-proteins extends to prokaryotes and might thus be widely conserved. Intriguingly, the finding that NADPH disrupted RNA-binding yet again links the Rossmann fold domain to RNA-binding activity, as postulated for other enzymes like glyceraldehyde-3-phosphate dehydrogenase and lactate dehydrogenase (Garcin 2019).

5.3 Notes on Literature Data Consulted for Candidate Enzyme Selection

The selection of RBP candidates was partially based on specific screenings that created interactome data sets for prokaryotes based on UV-crosslinking and subsequent isolation of RNA-protein complexes by phase separation, namely PTEX (Urdaneta et al. 2019), TRAPP (Shchepachev et al. 2019), and OOPS (Queiroz et al. 2019). In this context, it is an interesting matter whether proteins that displayed specific RNA-enrichment in genomic SELEX were also hits within these screenings ('hits' referring to preferential recovery of these proteins, i.e. positive

FC-values, see Table 5). Among the metabolic enzymes that showed noteworthy RNA enrichment in genomic SELEX (Table 15) only Pgk was a hit in all three screens. PykF, GapA, and AcnB were hits in TRAPP and OOPS. ThyA and ProB were hits only in TRAPP. QorA was a hit in TRAPP and was ambiguously identified as hit in two out of five replicates in OOPS. Taken together, TRAPP recovery is best in line with SELEX enrichment, and PTEX is worst. However, TRAPP also features the highest number of total recovered proteins (no matter if positive or negative FC-values), and PTEX features the smallest (TRAPP: >1000, OOPS: ~650, PTEX: ~250). Due to this general variance in mass spectrometry sensitivity, benchmarking the screenings against each other might principally be not constructive. This is stressed by the fact that a fair number of enzymes showing no enrichment in genomic SELEX was likewise preferably recovered by TRAPP and OOPS, but not by PTEX (examples: Eno, GpmA, TalA, TalB, Adk, Icd). None of the candidate enzymes that were chosen within this work solely based on occurrence in the interactome data sets showed specific RNA-enrichment in genomic SELEX or iCLIP, among them enzymes that were preferably recovered in all three screenings: RpiA, KdsA, Upp, Mdh. Taken together, a correlation between genomic SELEX or iCLIP results in this work and interactome data of PTEX, TRAPP, OOPS cannot be hypothesized. However, this work was mainly focused on *in vitro* selection, while the screenings isolated RNA-protein complexes from the cell. It is thus complex to judge whether a strict correlation of data was to be expected.

Reviewing genomic SELEX results from a broad perspective, it could be demonstrated that within a considerate selection of enzyme candidates that is based on literature RNA-binding references for the enzymes itself or homologues, a fair fraction produces results that hint towards possible affinity discrepancies within the RNA pool of *Escherichia coli*. It is reasonable to assume that observed instances of RNA enrichment were in fact protein-based and not a product of artificial selection bias. This is based on the control experiment displaying unambiguous enrichment of the well-characterized phage operator hairpin sequence by MS2 coat protein, as well as the multiple established correlations to literature, including ARE-binding by GapA, AcnB enriching motifs resembling a previously reported structure that mediates the feedback interaction of *acnB* mRNA with its own gene product, RNA-binding of the PUA domain-containing ProB, and RNA-binding of the prokaryotic ζ-crystallin pendant QorA. With that said, only a subset of these enriched RNAs were validated as specific binders of respective proteins. It is to be assumed that SELEX will enrich whatever sequence exerts the highest affinity towards the employed protein, irrespective of the actual range of probed affinities being large or minuscule. Hence, it should also be noted that obtained results cannot be interpreted as unconditional indicators for *in vivo*-occurring RNA-interactions of high affinity and biological significance. For example, the enrichment of A-rich loops observed for Pgk could opportunistically be explained by a coincidental interaction with the ATP-grove with no further biological implications, factoring in the general low affinity of Pgk towards RNAs (K_D = 42 µM) and lack of preferential binding observed in EMSAs.

While biological significance cannot be ruled out, it is advisable that the interactions which were observed in the form of RNA enrichment during genomic SELEX but were undetectable in follow-up analyis should be judged very carefully. However, the unambiguous discriminatory RNA-

binding of ProB and QorA intruigingly prompts for further investigations that establish physiological significance of unraveled interactions in irrevocable manner, and also suggests that RNA-related moonlighting functions, while not necessarily being utterly frequent, are yet to be unraveled amongst bacterial metabolic enzymes.

5.4 Outlook on Specific RNA-Interactions Observed for ProB and QorA

For ProB specifically, it would be advised to first continue the *in vitro* characterization of the interaction with *tktA* asRNA, as the band shift intensity observed thus far indicated extremly low affinity despite preferential binding. ProB displayed poor long-term stability especially when employed in higher concentrations, which prevented a K_D-determination. Co-purification of ProA, which stabilizes and interacts with ProB (Marco-Marín et al. 2007), might alleviate solubility issues observed during concentration of the protein for RNA-binding experiments. The heterocomplex formation might not only be useful for ProB stability but also be worth examining for altered affinity towards the specific interaction with the *tktA* asRNA fragment.

For QorA, the specific interaction with the *yffO* mRNA fragment was comprehensively demonstrated by *in vitro* experiments. Hence, verifying the interaction or physiological effects of the interaction *in vivo* should be the next step when further investigating the system, as the moderate *in vitro* K_D of 7 µM still suggests caution. A straightforward approach to gain insights on biological relevance would be comprised by follow-up CLIP-Seq studies, which could not be conducted in the timeline of this work. However, after having successfully identified a specific target RNA for the enzyme, switching to RNA-centric experiments might comprise a good way to validate the findings. A suitable strategy would be small molecule-labeling of target RNA (e.g. biotinylation), exposure to cell lysate and subsequent search for the target protein via mass spectrometry analysis. Another recommended method is aptamer modification. Aptamer-ligand affinities comparable to those of antibody-antigen-interactions are available (Gemmill et al. 2020), and can be used for RBP detection when incorporated during *in vivo* transcription. In the context of such experiments, it might be useful to first quantify the expression level of QorA, as the discovery of the interaction *in vivo* is ultimately dependent on the concentration of QorA in the cell.

If the interaction could be verified *in vivo*, the unknown gene function of the suspected former terminase and now grounded gene *yffO* would be of interest. Terminases are repeatedly reported to form higher order multimers (like octamers or decamers), including those of phages T4 (Lin et al. 1997), T7 (White and Richardson 1987), SPP1 (Chai et al. 1995), and P22 (Nemecek et al. 2007). In that context, a straightforward first step would be to simply purify the protein and characterize its oligomerization behavior, examining whether it still assembles as higher order multimer.

Another point of interest is the effect of RNA on NADPH binding. Since NADPH was shown to disrupt the RNA-QorA complex, activity might be affected *vice versa* by RNA. Thus far, assaying quinone oxidoreductase activity has not been very productive, as the native substrate for *ec*QorA is not known, and attempts with 1,4-benzoquinone showed possible residual, but uncertain

activity (Figure S 19). A quinone oxidoreductase from *Staphylococcus aureus*, which closely resembles *ec*QorA, is reported to catalyze the reduction of 9,10-phenanthrenequinone (Maruyama et al. 2003). The size of the substrate also fits the substrate grove of *ec*QorA based on crystal structure analysis (Thorn et al. 1995; Sulzenbacher et al. 2004). Thus, it is recommended to test this compound first when attempting to unravel the substrate spectrum of *ec*QorA. It might also be interesting to test the RNA-binding ability of YhdH, another *E. coli* quinone oxidase that is structurally homologous to QorA and also binds NADPH *via* a Rossmann fold domain (Sulzenbacher et al. 2004).

While further *in vitro* characterization of QorA's RNA-binding activity is of subordinate importance compared to validating its biological relevance, some suggestions for it are provided in the following. Crystallization of the *yffO* mRNA fragment-QorA complex would be of ultimate use but might be difficult to achieve owing to the moderate *in vitro* K_D. Since the exact relevance of each nucleotide position within the fragment could not be determined, additional mutational analysis could provide further inside on binding specificity, but is also expected to be difficult to interpret, based on *hitherto* pursued efforts. Another yet unresolved matter which should be explored is the stoichiometry of the interaction. Also, testing more RNA fragments from genomic loci that also showed enrichment during QorA SELEX will elucidate whether QorA's binding activity is fully wired to the *yffO* mRNA. Within this work, the *narI* mRNA fragment, which is predicted to maintain a deviating secondary structure, was shown to have lower affinity. In this context, some of the RNAs enriched in QorA SELEX are predicted to form stem-loops similar to the *yffO* mRNA fragment, like for example the *aslB* mRNA fragment (Figure 23c), and should be prioritized.

5.5 Outlook on SELEX Results for MS2

Due to the absence of enzymatic function and its origin outside of the *E. coli* cell, MS2 was of no central interest in this work, and merely served for validation of the genomic SELEX method. However, the MS2 SELEX experiment yielded noteworthy results. The evolved library showed a strict enrichment of sequences resembling the bacteriophage operator hairpin stem-loop, but with the slightly differing loop sequence ANCA. The cytosine within this loop sequence replaces an uracil found in the original bacteriophage stem-loop and increases the affinity 50-fold (Lowary and Uhlenbeck 1987), explaining the ANCA loop enrichment over the original sequence. The underlying gene functions of respective RNAs overwhelmingly relate to cell surface function. The same observation was made previously in genomic SELEX experiments (Shtatland et al. 2000). Due to NGS providing higher sequencing power, RNAs identified in this work greatly extended the pre-existing list of possible MS2 RNA targets and partially confirmed targets predicted by Shtatland et al. based on bioinformatical genome screening, like for example *bglX* (β-glucosidase), *nuoN* and *nuoK* (membrane-bound NADH quinone oxidoreductase subunits), or *fic* (probable adenylyltransferase, involved in cell filamentation (Kawamukai et al. 1989)), and also comprise genes of unknown functions like *yjgR, yhiN*, *yqeB,* or *ybjT*. By implication, the enrichment of corresponding mRNAs / asRNAs might give a hint that some of these genes of unknown function are involved in cell surface physiology in one way or another. While the phage mechanism of

host-translational manipulation to promote self-proliferation is well-understood (Altuvia et al. 2018), implications of a possible specific posttranscriptional control of host cell surface genes are not described in literature and might be an interesting matter to explore.

6 Publication Bibliography

Albihlal, Waleed S.; Gerber, André P. (2018): Unconventional RNA-binding proteins: an uncharted zone in RNA biology. In *FEBS letters* 592 (17), pp. 2917–2931. DOI: 10.1002/1873-3468.13161.

Altuvia, Shoshy; Storz, Gisela; Papenfort, Kai (2018): Cross-Regulation between Bacteria and Phages at a Posttranscriptional Level. In *Microbiology spectrum* 6 (4). DOI: 10.1128/microbiolspec.RWR-0027-2018.

Anantharaman, Vivek; Koonin, Eugene V.; Aravind, L. (2002): Comparative genomics and evolution of proteins involved in RNA metabolism. In *Nucleic Acids Research* 30 (7), pp. 1427–1464. DOI: 10.1093/nar/30.7.1427.

Anderson, Sondra L.; Schirf, Virgil; McAlister-Henn, L. (2002): Effect of AMP on mRNA binding by yeast NAD+-specific isocitrate dehydrogenase. In *Biochemistry* 41 (22), pp. 7065–7073. DOI: 10.1021/bi0200662.

Andras, Peter; Andras, Csaba (2005): The origins of life -- the 'protein interaction world' hypothesis: protein interactions were the first form of self-reproducing life and nucleic acids evolved later as memory molecules. In *Medical hypotheses* 64 (4), pp. 678–688. DOI: 10.1016/j.mehy.2004.11.029.

Aravind, L.; Koonin, E. V. (1999): Novel predicted RNA-binding domains associated with the translation machinery. In *Journal of molecular evolution* 48 (3), pp. 291–302. DOI: 10.1007/pl00006472.

Backlund, Michael; Paukku, Kirsi; Daviet, Laurent; Boer, Rudolf A. de; Valo, Erkka; Hautaniemi, Sampsa et al. (2009): Posttranscriptional regulation of angiotensin II type 1 receptor expression by glyceraldehyde 3-phosphate dehydrogenase. In *Nucleic Acids Research* 37 (7), pp. 2346–2358. DOI: 10.1093/nar/gkp098.

Bailey, T. L.; Elkan, C. (1994): Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings. International Conference on Intelligent Systems for Molecular Biology* 2, pp. 28–36.

Bailey, Timothy L.; Johnson, James; Grant, Charles E.; Noble, William S. (2015): The MEME Suite. In *Nucleic Acids Research* 43 (W1), W39-49. DOI: 10.1093/nar/gkv416.

Baltz, Alexander G.; Munschauer, Mathias; Schwanhäusser, Björn; Vasile, Alexandra; Murakawa, Yasuhiro; Schueler, Markus et al. (2012): The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. In *Molecular cell* 46 (5), pp. 674–690. DOI: 10.1016/j.molcel.2012.05.021.

Bao, Xichen; Guo, Xiangpeng; Yin, Menghui; Tariq, Muqddas; Lai, Yiwei; Kanwal, Shahzina et al. (2018): Capturing the interactome of newly transcribed RNA. In *Nature methods* 15 (3), pp. 213–220. DOI: 10.1038/nmeth.4595.

Bazzi, M. D. (2001): Interaction of camel lens zeta-crystallin with quinones: portrait of a substrate by fluorescence spectroscopy. In *Archives of Biochemistry and Biophysics* 395 (2), pp. 185–190. DOI: 10.1006/abbi.2001.2538.

Beckmann, Benedikt M.; Castello, Alfredo; Medenbach, Jan (2016): The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. In *Pflugers Archiv : European journal of physiology* 468 (6), pp. 1029–1040. DOI: 10.1007/s00424-016-1819-4.

Beckmann, Benedikt M.; Horos, Rastislav; Fischer, Bernd; Castello, Alfredo; Eichelbaum, Katrin; Alleaume, Anne-Marie et al. (2015): The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. In *Nature communications* 6, p. 10127. DOI: 10.1038/ncomms10127.

Benjamin, Julie-Anna M.; Massé, Eric (2014): The iron-sensing aconitase B binds its own mRNA to prevent sRNA-induced mRNA cleavage. In *Nucleic Acids Research* 42 (15), pp. 10023–10036. DOI: 10.1093/nar/gku649.

Berlyn, M. K. (1998): Linkage map of Escherichia coli K-12, edition 10: the traditional map. In *Microbiology and molecular biology reviews : MMBR* 62 (3), pp. 814–984. DOI: 10.1128/MMBR.62.3.814-984.1998.

Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H. et al. (2000): The Protein Data Bank. In *Nucleic Acids Research* 28 (1), pp. 235–242. DOI: 10.1093/nar/28.1.235.

Bernardi, A.; Spahr, P. F. (1972): Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. In *Proceedings of the National Academy of Sciences of the United States of America* 69 (10), pp. 3033–3037. DOI: 10.1073/pnas.69.10.3033.

Bernhardt, Harold S. (2012): The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). In *Biology direct* 7, p. 23. DOI: 10.1186/1745-6150-7-23.

Bertini, Ivano; McGreevy, Kathleen S.; Parigi, Giacomo (Eds.) (2012): NMR of biomolecules. Towards mechanistic systems biology. Weinheim, Germany: Wiley-VCH. Available online at https://onlinelibrary.wiley.com/doi/book/10.1002/9783527644506.

Bonthron, D. T. (1990): L-asparaginase II of Escherichia coli K-12: cloning, mapping and sequencing of the ansB gene. In *Gene* 91 (1), pp. 101–105. DOI: 10.1016/0378-1119(90)90168-q.

Boucas, Jorge; Fritz, Christian; Schmitt, Anna; Riabinska, Arina; Thelen, Lisa; Peifer, Martin et al. (2015): Label-Free Protein-RNA Interactome Analysis Identifies Khsrp Signaling Downstream of the p38/Mk2 Kinase Complex as a Critical Modulator of Cell Cycle Progression. In *PloS one* 10 (5), e0125745. DOI: 10.1371/journal.pone.0125745.

Bradford, M. M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. In *Analytical biochemistry* 72, pp. 248–254. DOI: 10.1006/abio.1976.9999.

Bruce Alberts; Alexander Johnson; Julian Lewis; Martin Raff; Keith Roberts; Peter Walter (2002): The RNA World and the Origins of Life. In Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter (Eds.): Molecular Biology of the Cell. 4th edition: Garland Science. Available online at https://www.ncbi.nlm.nih.gov/books/NBK26876/.

Buchbender, Andreas; Mutter, Holger; Sutandy, F. X. Reymond; Körtel, Nadine; Hänel, Heike; Busch, Anke et al. (2020): Improved library preparation with the new iCLIP2 protocol. In *Methods (San Diego, Calif.)* 178, pp. 33–48. DOI: 10.1016/j.ymeth.2019.10.003.

Busch, Anke; Brüggemann, Mirko; Ebersberger, Stefanie; Zarnack, Kathi (2020): iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. In *Methods (San Diego, Calif.)* 178, pp. 49–62. DOI: 10.1016/j.ymeth.2019.11.008.

Butter, Falk; Scheibe, Marion; Mörl, Mario; Mann, Matthias (2009): Unbiased RNA-protein interaction screen by quantitative proteomics. In *Proceedings of the National Academy of Sciences of the United States of America* 106 (26), pp. 10626–10631. DOI: 10.1073/pnas.0812099106.

Cadenas, Enrique (1995): Antioxidant and prooxidant functions of DT-diaphorase in quinone metabolism. In *Biochemical Pharmacology* 49 (2), pp. 127–140. DOI: 10.1016/S0006-2952(94)00333-5.

Canchaya, Carlos; Proux, Caroline; Fournous, Ghislain; Bruttin, Anne; Brüssow, Harald (2003): Prophage genomics. In *Microbiology and molecular biology reviews : MMBR* 67 (2), 238-76, table of contents. DOI: 10.1128/MMBR.67.2.238-276.2003.

Carmona, Pedro; Rodríguez-Casado, Arantxa; Molina, Marina (1999): Conformational structure and binding mode of glyceraldehyde-3-phosphate dehydrogenase to tRNA studied by Raman and CD spectroscopy. In *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1432 (2), pp. 222–233. DOI: 10.1016/s0167-4838(99)00113-2.

Casjens, Sherwood (2003): Prophages and bacterial genomics: what have we learned so far? In *Molecular microbiology* 49 (2), pp. 277–300. DOI: 10.1046/j.1365-2958.2003.03580.x.

Castello, Alfredo; Fischer, Bernd; Eichelbaum, Katrin; Horos, Rastislav; Beckmann, Benedikt M.; Strein, Claudia et al. (2012): Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. In *Cell* 149 (6), pp. 1393–1406. DOI: 10.1016/j.cell.2012.04.031.

Castello, Alfredo; Fischer, Bernd; Frese, Christian K.; Horos, Rastislav; Alleaume, Anne-Marie; Foehr, Sophia et al. (2016): Comprehensive Identification of RNA-Binding Domains in Human Cells. In *Molecular cell* 63 (4), pp. 696–710. DOI: 10.1016/j.molcel.2016.06.029.

Castello, Alfredo; Hentze, Matthias W.; Preiss, Thomas (2015): Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. In *Trends in endocrinology and metabolism: TEM* 26 (12), pp. 746–757. DOI: 10.1016/j.tem.2015.09.012.

Cerrudo, Carolina S.; Ghiringhelli, Pablo D.; Gomez, Daniel E. (2014): Protein universe containing a PUA RNA-binding domain. In *The FEBS journal* 281 (1), pp. 74–87. DOI: 10.1111/febs.12602.

Chai, S.; Lurz, R.; Alonso, J. C. (1995): The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. In *Journal of Molecular Biology* 252 (4), pp. 386–398. DOI: 10.1006/jmbi.1995.0505.

Chandran, Vidya; Luisi, Ben F. (2006): Recognition of enolase in the Escherichia coli RNA degradosome. In *Journal of Molecular Biology* 358 (1), pp. 8–15. DOI: 10.1016/j.jmb.2006.02.012.

Chang, Chih-Hao; Curtis, Jonathan D.; Maggi, Leonard B.; Faubert, Brandon; Villarino, Alejandro V.; O'Sullivan, David et al. (2013): Posttranscriptional control of T cell effector function by aerobic glycolysis. In *Cell* 153 (6), pp. 1239–1251. DOI: 10.1016/j.cell.2013.05.016.

Chen, Yong-Hong; Liu, Shu-Jing; Gao, Mei-Mei; Zeng, Tao; Lin, Guo-Wang; Tan, Na-Na et al. (2017): MDH2 is an RNA binding protein involved in downregulation of sodium channel Scn1a expression under seizure condition. In *Biochimica et biophysica acta. Molecular basis of disease* 1863 (6), pp. 1492–1499. DOI: 10.1016/j.bbadis.2017.04.018.

Cheung, Yee-Wai; Kwok, Jane; Law, Alan W. L.; Watt, Rory M.; Kotaka, Masayo; Tanner, Julian A. (2013): Structural basis for discriminatory recognition of Plasmodium lactate dehydrogenase by a DNA aptamer. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (40), pp. 15967–15972. DOI: 10.1073/pnas.1309538110.

Cho, Hyun-Hee; Cahill, Catherine M.; Vanderburg, Charles R.; Scherzer, Clemens R.; Wang, Bin; Huang, Xudong; Rogers, Jack T. (2010): Selective translational control of the Alzheimer amyloid precursor protein transcript by iron regulatory protein-1. In *The Journal of biological chemistry* 285 (41), pp. 31217–31232. DOI: 10.1074/jbc.M110.149161.

Cho, Soo-Yeon; Jung, Soo-Jin; Kim, Kyoung-Dong; Roe, Jung-Hye (2021): Non-mitochondrial aconitase regulates the expression of iron-uptake genes by controlling the RNA turnover process in fission yeast. In *Journal of microbiology (Seoul, Korea)* 59 (12), pp. 1075–1082. DOI: 10.1007/s12275-021-1438-4.

Choi, Sung-Jin; Ban, Changill (2016): Crystal structure of a DNA aptamer bound to PvLDH elucidates novel single-stranded DNA structural elements for folding and recognition. In *Scientific Reports* 6, p. 34998. DOI: 10.1038/srep34998.

Chu, E.; Voeller, D.; Koeller, D. M.; Drake, J. C.; Takimoto, C. H.; Maley, G. F. et al. (1993): Identification of an RNA binding site for human thymidylate synthase. In *Proceedings of the National Academy of Sciences of the United States of America* 90 (2), pp. 517–521. DOI: 10.1073/pnas.90.2.517.

Chu, Liang-Cui; Arede, Pedro; Li, Wei; Urdaneta, Erika C.; Ivanova, Ivayla; McKellar, Stuart W. et al. (2022): The RNA-bound proteome of MRSA reveals post-transcriptional roles for helix-turn-helix DNA-binding and Rossmann-fold proteins. In *Nature communications* 13 (1), p. 2883. DOI: 10.1038/s41467-022-30553-8.

Cieśla, Joanna (2006): Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? In *Acta biochimica Polonica* 53 (1), pp. 11–32.

Cléry, Antoine; Blatter, Markus; Allain, Frédéric H-T (2008): RNA recognition motifs: boring? Not quite. In *Current opinion in structural biology* 18 (3), pp. 290–298. DOI: 10.1016/j.sbi.2008.04.002.

Coltri, Patrícia P.; Guimarães, Beatriz G.; Granato, Daniela C.; Luz, Juliana S.; Teixeira, Elaine C.; Oliveira, Carla C.; Zanchin, Nilson I. T. (2007): Structural insights into the interaction of the Nip7 PUA domain with polyuridine RNA. In *Biochemistry* 46 (49), pp. 14177–14187. DOI: 10.1021/bi7015876.

Conrad, Richard C.; Giver, Lori; Tian, Yu; Ellington, Andrew D. (1996): [20] In vitro selection of nucleic acid aptamers that bind proteins. In John N. Abelson (Ed.): Combinatorial chemistry, vol. 267. San Diego, Calif. [u.a.]: Acad. Press (Methods in Enzymology, 267), pp. 336–367.

Conrad, Thomas; Albrecht, Anne-Susann; Melo Costa, Veronica Rodrigues de; Sauer, Sascha; Meierhofer, David; Ørom, Ulf Andersson (2016): Serial interactome capture of the human cell nucleus. In *Nature communications* 7, p. 11212. DOI: 10.1038/ncomms11212.

Constable, A.; Quick, S.; Gray, N. K.; Hentze, M. W. (1992): Modulation of the RNA-binding activity of a regulatory protein by iron in vitro: switching between enzymatic and genetic function? In *Proceedings of the National Academy of Sciences of the United States of America* 89 (10), pp. 4554–4558. DOI: 10.1073/pnas.89.10.4554.

Cox, J.Colin; Ellington, Andrew D. (2001): Automated selection of anti-Protein aptamers. In *Bioorganic & medicinal chemistry* 9 (10), pp. 2525–2531. DOI: 10.1016/s0968-0896(01)00028-1.

Creamer, K. M.; Lawrence, J. B. (2017): XIST RNA: a window into the broader role of RNA in nuclear chromosome architecture. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372 (1733). DOI: 10.1098/rstb.2016.0360.

Crick, F. (1970): Central dogma of molecular biology. In *Nature* 227 (5258), pp. 561–563. DOI: 10.1038/227561a0.

Crick, F.H.C. (1968): The origin of the genetic code. In *Journal of Molecular Biology* 38 (3), pp. 367–379. DOI: 10.1016/0022-2836(68)90392-6.

Dabney, Jesse; Meyer, Matthias (2012): Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. In *BioTechniques* 52 (2), pp. 87–94. DOI: 10.2144/000113809.

Datsenko, K. A.; Wanner, B. L. (2000): One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. In *Proceedings of the National Academy of Sciences of the United States of America* 97 (12), pp. 6640–6645. DOI: 10.1073/pnas.120163297.

Davis, Jonathan H.; Szostak, Jack W. (2002): Isolation of high-affinity GTP aptamers from partially structured RNA libraries. In *Proceedings of the National Academy of Sciences of the United States of America* 99 (18), pp. 11616–11621. DOI: 10.1073/pnas.182095699.

De, B. P.; Gupta, S.; Zhao, H.; Drazba, J. A.; Banerjee, A. K. (1996): Specific interaction in vitro and in vivo of glyceraldehyde-3-phosphate dehydrogenase and LA protein with cis-acting

RNAs of human parainfluenza virus type 3. In *The Journal of biological chemistry* 271 (40), pp. 24728–24735. DOI: 10.1074/jbc.271.40.24728.

Deppert, Wolfgang R.; Normann, Johannes; Wagner, Edgar (1992): Adenylate kinase from plant tissues. In *Journal of Chromatography A* 625 (1), pp. 13–19. DOI: 10.1016/0021-9673(92)87216-U.

Duhr, Stefan; Braun, Dieter (2006): Why molecules move along a temperature gradient. In *Proceedings of the National Academy of Sciences of the United States of America* 103 (52), pp. 19678–19682. DOI: 10.1073/pnas.0603873103.

Ellington, A. D.; Szostak, J. W. (1990): In vitro selection of RNA molecules that bind specific ligands. In *Nature* 346 (6287), pp. 818–822. DOI: 10.1038/346818a0.

Ellington, A. D.; Szostak, J. W. (1992): Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. In *Nature* 355 (6363), pp. 850–852. DOI: 10.1038/355850a0.

Entelis, Nina; Brandina, Irina; Kamenski, Piotr; Krasheninnikov, Igor A.; Martin, Robert P.; Tarassov, Ivan (2006): A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in Saccharomyces cerevisiae. In *Genes & development* 20 (12), pp. 1609–1620. DOI: 10.1101/gad.385706.

Evguenieva-Hackenberg, Elena; Schiltz, Emile; Klug, Gabriele (2002): Dehydrogenases from all three domains of life cleave RNA. In *Journal of Biological Chemistry* 277 (48), pp. 46145–46150. DOI: 10.1074/jbc.M208717200.

Fernández, M. R.; Porté, S.; Crosas, E.; Barberà, N.; Farrés, J.; Biosca, J. A.; Parés, X. (2007): Human and yeast zeta-crystallins bind AU-rich elements in RNA. In *Cellular and molecular life sciences : CMLS* 64 (11), pp. 1419–1427. DOI: 10.1007/s00018-007-7091-1.

Funke, Franziska (2019): In Search of RNA-Binding Metabolic Enzymes in E. Coli. Master's Thesis. University of Regensburg, Regensburg. Institute for Biophysics and Physical Biochemistry.

Garcin, Elsa D. (2019): GAPDH as a model non-canonical AU-rich RNA binding protein. In *Seminars in cell & developmental biology* 86, pp. 162–173. DOI: 10.1016/j.semcdb.2018.03.013.

Garland, Donita; Rao, P.Vasantha; Del Corso, Antonella; Mura, Umberto; Zigler, J.Samuel (1991): ζ -Crystallin is a major protein in the lens of Camelus dromedarius. In *Archives of Biochemistry and Biophysics* 285 (1), pp. 134–136. DOI: 10.1016/0003-9861(91)90339-K.

Gemmill, Darren; D'souza, Simmone; Meier-Stephenson, Vanessa; Patel, Trushar R. (2020): Current approaches for RNA-labelling to identify RNA-binding proteins. In *Biochemistry and cell biology = Biochimie et biologie cellulaire* 98 (1), pp. 31–41. DOI: 10.1139/bcb-2019-0041.

Gerovac, Milan; El Mouali, Youssef; Kuper, Jochen; Kisker, Caroline; Barquist, Lars; Vogel, Jörg (2020): Global discovery of bacterial RNA-binding proteins by RNase-sensitive gradient profiles

reports a new FinO domain protein. In *RNA (New York, N.Y.)* 26 (10), pp. 1448–1463. DOI: 10.1261/rna.076992.120.

Gilbert, Walter (1986): Origin of life: The RNA world. In *Nature* 319 (6055), p. 618. DOI: 10.1038/319618a0.

Glisovic, Tina; Bachorik, Jennifer L.; Yong, Jeongsik; Dreyfuss, Gideon (2008): RNA-binding proteins and post-transcriptional gene regulation. In *FEBS letters* 582 (14), pp. 1977–1986. DOI: 10.1016/j.febslet.2008.03.004.

Gold, L. (1995): Oligonucleotides as research, diagnostic, and therapeutic agents. In *The Journal of biological chemistry* 270 (23), pp. 13581–13584. DOI: 10.1074/jbc.270.23.13581.

Gonzalez, P.; Hernández-Calzadilla, C.; Rao, P. V.; Rodriguez, I. R.; Zigler, J. S.; Borrás, T. (1994a): Comparative analysis of the zeta-crystallin/quinone reductase gene in guinea pig and mouse. In *Molecular biology and evolution* 11 (2), pp. 305–315. DOI: 10.1093/oxfordjournals.molbev.a040111.

Gonzalez, P.; Rao, P. V.; Zigler, J. S. (1993): Molecular cloning and sequencing of zeta-crystallin/quinone reductase cDNA from human liver. In *Biochemical and biophysical research communications* 191 (3), pp. 902–907. DOI: 10.1006/bbrc.1993.1302.

Gonzalez, P.; Rao, P. V.; Zigler, J. S. (1994b): Organization of the human zeta-crystallin/quinone reductase gene (CRYZ). In *Genomics* 21 (2), pp. 317–324. DOI: 10.1006/geno.1994.1272.

Griffin, B. E. (1975): Studies and sequences of Escherichia coli 4.5 S RNA. In *Journal of Biological Chemistry* 250 (14), pp. 5426–5437. DOI: 10.1016/S0021-9258(19)41199-X.

Gruber, Andreas R.; Lorenz, Ronny; Bernhart, Stephan H.; Neuböck, Richard; Hofacker, Ivo L. (2008): The Vienna RNA websuite. In *Nucleic Acids Research* 36 (Web Server issue), W70-4. DOI: 10.1093/nar/gkn188.

Guenther, Ulf-Peter; Yandek, Lindsay E.; Niland, Courtney N.; Campbell, Frank E.; Anderson, David; Anderson, Vernon E. et al. (2013): Hidden specificity in an apparently nonspecific RNA-binding protein. In *Nature* 502 (7471), pp. 385–388. DOI: 10.1038/nature12543.

Hacisuleyman, Ezgi; Goff, Loyal A.; Trapnell, Cole; Williams, Adam; Henao-Mejia, Jorge; Sun, Lei et al. (2014): Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. In *Nature structural & molecular biology* 21 (2), pp. 198–206. DOI: 10.1038/nsmb.2764.

Hafner, Markus; Katsantoni, Maria; Köster, Tino; Marks, James; Mukherjee, Joyita; Staiger, Dorothee et al. (2021): CLIP and complementary methods. In *Nat Rev Methods Primers* 1 (1), pp. 1–23. DOI: 10.1038/s43586-021-00018-1.

Hafner, Markus; Landthaler, Markus; Burger, Lukas; Khorshid, Mohsen; Hausser, Jean; Berninger, Philipp et al. (2010): Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. In *Cell* 141 (1), pp. 129–141. DOI: 10.1016/j.cell.2010.03.009.

He, Chongsheng; Sidoli, Simone; Warneford-Thomson, Robert; Tatomer, Deirdre C.; Wilusz, Jeremy E.; Garcia, Benjamin A.; Bonasio, Roberto (2016): High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. In *Molecular cell* 64 (2), pp. 416–430. DOI: 10.1016/j.molcel.2016.09.034.

Head, Steven R.; Komori, H. Kiyomi; LaMere, Sarah A.; Whisenant, Thomas; van Nieuwerburgh, Filip; Salomon, Daniel R.; Ordoukhanian, Phillip (2014): Library construction for next-generation sequencing: overviews and challenges. In *BioTechniques* 56 (2), 61-4, 66, 68, passim. DOI: 10.2144/000114133.

Hedlund, Joel; Jörnvall, Hans; Persson, Bengt (2010): Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement. In *BMC bioinformatics* 11, p. 534. DOI: 10.1186/1471-2105-11-534.

Hentze, M. W.; Argos, P. (1991): Homology between IRE-BP, a regulatory RNA-binding protein, aconitase, and isopropylmalate isomerase. In *Nucleic Acids Research* 19 (8), pp. 1739–1740. DOI: 10.1093/nar/19.8.1739.

Hentze, Matthias W. (1994): Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? In *Trends in Biochemical Sciences* 19 (3), pp. 101–103. DOI: 10.1016/0968-0004(94)90198-8.

Hentze, Matthias W.; Castello, Alfredo; Schwarzl, Thomas; Preiss, Thomas (2018): A brave new world of RNA-binding proteins. In *Nature reviews. Molecular cell biology* 19 (5), pp. 327–341. DOI: 10.1038/nrm.2017.130.

Hentze, Matthias W.; Preiss, Thomas (2010): The REM phase of gene regulation. In *Trends in Biochemical Sciences* 35 (8), pp. 423–426. DOI: 10.1016/j.tibs.2010.05.009.

Holmqvist, Erik; Li, Lei; Bischler, Thorsten; Barquist, Lars; Vogel, Jörg (2018): Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. In *Molecular cell* 70 (5), 971-982.e6. DOI: 10.1016/j.molcel.2018.04.017.

Holmqvist, Erik; Wright, Patrick R.; Li, Lei; Bischler, Thorsten; Barquist, Lars; Reinhardt, Richard et al. (2016): Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. In *The EMBO journal* 35 (9), pp. 991–1011. DOI: 10.15252/embj.201593360.

Hu, C. A.; Delauney, A. J.; Verma, D. P. (1992): A bifunctional enzyme (delta 1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. In *Proceedings of the National Academy of Sciences of the United States of America* 89 (19), pp. 9354–9358. DOI: 10.1073/pnas.89.19.9354.

Huang, Q. L.; Russell, P.; Stone, S. H.; Zigler, J. S. (1987): Zeta-crystallin, a novel lens protein from the guinea pig. In *Current eye research* 6 (5), pp. 725–732. DOI: 10.3109/02713688709034836.

Huang, Rongbing; Han, Mengting; Meng, Liying; Chen, Xing (2018): Capture and Identification of RNA-binding Proteins by Using Click Chemistry-assisted RNA-interactome Capture (CARIC) Strategy. In *Journal of Visualized Experiments : JoVE* (140). DOI: 10.3791/58580.

Hunding, Axel; Kepes, Francois; Lancet, Doron; Minsky, Abraham; Norris, Vic; Raine, Derek et al. (2006): Compositional complementarity and prebiotic ecology in the origin of life. In *BioEssays : news and reviews in molecular, cellular and developmental biology* 28 (4), pp. 399–412. DOI: 10.1002/bies.20389.

Huppertz, Ina; Attig, Jan; D'Ambrogio, Andrea; Easton, Laura E.; Sibley, Christopher R.; Sugimoto, Yoichiro et al. (2014): iCLIP: protein-RNA interactions at nucleotide resolution. In *Methods (San Diego, Calif.)* 65 (3), pp. 274–287. DOI: 10.1016/j.ymeth.2013.10.011.

Huppertz, Ina; Perez-Perri, Joel I.; Mantas, Panagiotis; Sekaran, Thileepan; Schwarzl, Thomas; Russo, Francesco et al. (2022): Riboregulation of Enolase 1 activity controls glycolysis and embryonic stem cell differentiation. In *Molecular cell*. DOI: 10.1016/j.molcel.2022.05.019.

Ibrahim, H.; Lee, Y. J.; Curthoys, N. P. (2008): Renal response to metabolic acidosis: role of mRNA stabilization. In *Kidney international* 73 (1), pp. 11–18. DOI: 10.1038/sj.ki.5002581.

Ikeda, Yuki; Yamaji, Ryoichi; Irie, Kazuhiro; Kioka, Noriyuki; Murakami, Akira (2012): Glyceraldehyde-3-phosphate dehydrogenase regulates cyclooxygenase-2 expression by targeting mRNA stability. In *Archives of Biochemistry and Biophysics* 528 (2), pp. 141–147. DOI: 10.1016/j.abb.2012.09.004.

Järvelin, Aino I.; Noerenberg, Marko; Davis, Ilan; Castello, Alfredo (2016): The new (dis)order in RNA regulation. In *Cell communication and signaling : CCS* 14, p. 9. DOI: 10.1186/s12964-016-0132-3.

Jerabek-Willemsen, Moran; André, Timon; Wanner, Randy; Roth, Heide Marie; Duhr, Stefan; Baaske, Philipp; Breitsprecher, Dennis (2014): MicroScale Thermophoresis: Interaction analysis and beyond. In *Journal of Molecular Structure* 1077, pp. 101–113. DOI: 10.1016/j.molstruc.2014.03.009.

Johansson, Hans E.; Liljas, Lars; Uhlenbeck, Olke C. (1997): RNA Recognition by the MS2 Phage Coat Protein. In *Seminars in Virology* 8 (3), pp. 176–185. DOI: 10.1006/smvy.1997.0120.

Karmakar, Shilpita; Ramirez, Oscar; Paul, Kiran V.; Gupta, Abhishek K.; Kumari, Vandana; Botti, Valentina et al. (2022): Integrative genome-wide analysis reveals EIF3A as a key downstream regulator of translational repressor protein Musashi 2 (MSI2). In *NAR cancer* 4 (2), zcac015. DOI: 10.1093/narcan/zcac015.

Kawamukai, M.; Matsuda, H.; Fujii, W.; Utsumi, R.; Komano, T. (1989): Nucleotide sequences of fic and fic-1 genes involved in cell filamentation induced by cyclic AMP in Escherichia coli. In *Journal of bacteriology* 171 (8), pp. 4525–4529. DOI: 10.1128/jb.171.8.4525-4529.1989.

Kiri, Arpna; Goldspink, Geoffrey (2002): RNA-protein interactions of the 3' untranslated regions of myosin heavy chain transcripts. In *Journal of muscle research and cell motility* 23 (2), pp. 119–129. DOI: 10.1023/a:1020211729728.

Kitagawa, Masanari; Ara, Takeshi; Arifuzzaman, Mohammad; Ioka-Nakamichi, Tomoko; Inamoto, Eiji; Toyonaga, Hiromi; Mori, Hirotada (2005): Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. In *DNA research : an international journal for rapid publication of reports on genes and genomes* 12 (5), pp. 291–299. DOI: 10.1093/dnares/dsi012.

Knoll, Michael; Pleiss, Jürgen (2008): The Medium-Chain Dehydrogenase/reductase Engineering Database: a systematic analysis of a diverse protein family to understand sequence-structure-function relationship. In *Protein science : a publication of the Protein Society* 17 (10), pp. 1689–1697. DOI: 10.1110/ps.035428.108.

Kondo, Seiji; Kubota, Satoshi; Mukudai, Yoshiki; Nishida, Takashi; Yoshihama, Yasuto; Shirota, Tatsuo et al. (2011): Binding of glyceraldehyde-3-phosphate dehydrogenase to the cis-acting element of structure-anchored repression in ccn2 mRNA. In *Biochemical and biophysical research communications* 405 (3), pp. 382–387. DOI: 10.1016/j.bbrc.2011.01.034.

König, Julian; Zarnack, Kathi; Rot, Gregor; Curk, Tomaz; Kayikci, Melis; Zupan, Blaz et al. (2010): iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. In *Nature structural & molecular biology* 17 (7), pp. 909–915. DOI: 10.1038/nsmb.1838.

Krakau, Sabrina; Richard, Hugues; Marsico, Annalisa (2017): PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. In *Genome Biol* 18 (1), p. 240. DOI: 10.1186/s13059-017-1364-2.

Kramer, Katharina; Sachsenberg, Timo; Beckmann, Benedikt M.; Qamar, Saadia; Boon, Kum-Loong; Hentze, Matthias W. et al. (2014): Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. In *Nature methods* 11 (10), pp. 1064–1070. DOI: 10.1038/nmeth.3092.

Kurland, Charles G. (2010): The RNA dreamtime: modern cells feature proteins that might have supported a prebiotic polypeptide world but nothing indicates that RNA world ever was. In *BioEssays : news and reviews in molecular, cellular and developmental biology* 32 (10), pp. 866–871. DOI: 10.1002/bies.201000058.

Kwon, S. Chul; Yi, Hyerim; Eichelbaum, Katrin; Föhr, Sophia; Fischer, Bernd; You, Kwon Tae et al. (2013): The RNA-binding protein repertoire of embryonic stem cells. In *Nature structural & molecular biology* 20 (9), pp. 1122–1130. DOI: 10.1038/nsmb.2638.

Lakhin, A. V.; Tarantul, V. Z.; Gening, L. V. (2013): Aptamers: problems, solutions and prospects. In *Acta Naturae* 5 (4), pp. 34–43.

Lapucci, Andrea; Lulli, Matteo; Amedei, Amedeo; Papucci, Laura; Witort, Ewa; Di Gesualdo, Federico et al. (2010): zeta-Crystallin is a bcl-2 mRNA binding protein involved in bcl-2 overexpression in T-cell acute lymphocytic leukemia. In *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 24 (6), pp. 1852–1865. DOI: 10.1096/fj.09-140459.

LeCuyer, K. A.; Behlen, L. S.; Uhlenbeck, O. C. (1995): Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. In *Biochemistry* 34 (33), pp. 10600–10606. DOI: 10.1021/bi00033a035.

Li, Gene-Wei; Burkhardt, David; Gross, Carol; Weissman, Jonathan S. (2014): Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. In *Cell* 157 (3), pp. 624–635. DOI: 10.1016/j.cell.2014.02.033.

Li, Zhongwei; Deutscher, Murray P. (2008): Chapter 2 Analyzing the Decay of Stable RNAs in E. coli. In Lynne E. Maquat (Ed.): RNA turnover in bacteria, archaea and organelles, vol. 447. Amsterdam [u.a.]: Elsevier, Acad. Press (Methods in Enzymology, 447), pp. 31–45.

Liao, Yalin; Castello, Alfredo; Fischer, Bernd; Leicht, Stefan; Föehr, Sophia; Frese, Christian K. et al. (2016): The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. In *Cell reports* 16 (5), pp. 1456–1469. DOI: 10.1016/j.celrep.2016.06.084.

Liepelt, Anke; Naarmann-de Vries, Isabel S.; Simons, Nadine; Eichelbaum, Katrin; Föhr, Sophia; Archer, Stuart K. et al. (2016): Identification of RNA-binding Proteins in Macrophages by Interactome Capture. In *Molecular & cellular proteomics : MCP* 15 (8), pp. 2699–2714. DOI: 10.1074/mcp.M115.056564.

Lin, Guo-Wang; Lu, Ping; Zeng, Tao; Tang, Hui-Ling; Chen, Yong-Hong; Liu, Shu-Jing et al. (2017): GAPDH-mediated posttranscriptional regulations of sodium channel Scn1a and Scn3a genes under seizure and ketogenic diet conditions. In *Neuropharmacology* 113 (Pt A), pp. 480–489. DOI: 10.1016/j.neuropharm.2016.11.002.

Lin, H.; Simon, M. N.; Black, L. W. (1997): Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. In *Journal of Biological Chemistry* 272 (6), pp. 3495–3501. DOI: 10.1074/jbc.272.6.3495.

Lin, Jen-Wen; Ding, Min-Pey; Hsu, Yau-Heiu; Tsai, Ching-Hsiu (2007): Chloroplast phosphoglycerate kinase, a gluconeogenetic enzyme, is required for efficient accumulation of Bamboo mosaic virus. In *Nucleic Acids Research* 35 (2), pp. 424–432. DOI: 10.1093/nar/gkl1061.

Lin, Yuan; Protter, David S. W.; Rosen, Michael K.; Parker, Roy (2015): Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. In *Molecular cell* 60 (2), pp. 208–219. DOI: 10.1016/j.molcel.2015.08.018.

Liu, Cuiping; Sanders, Jeffrey M.; Pascal, John M.; Hou, Ya-Ming (2012): Adaptation to tRNA acceptor stem structure by flexible adjustment in the catalytic domain of class I tRNA synthetases. In *RNA (New York, N.Y.)* 18 (2), pp. 213–221. DOI: 10.1261/rna.029983.111.

Lodde, Valeria; Floris, Matteo; Munk, Rachel; Martindale, Jennifer L.; Piredda, Davide; Napodano, Catello Mario Panu et al. (2022): Systematic identification of NF90 target RNAs by iCLIP analysis. In *Scientific Reports* 12 (1), p. 364. DOI: 10.1038/s41598-021-04101-1.

Lotz, Thea S.; Halbritter, Thomas; Kaiser, Christoph; Rudolph, Martin M.; Kraus, Leon; Groher, Florian et al. (2019): A light-responsive RNA aptamer for an azobenzene derivative. In *Nucleic Acids Research* 47 (4), pp. 2029–2040. DOI: 10.1093/nar/gky1225.

Lowary, P. T.; Uhlenbeck, O. C. (1987): An RNA mutation that increases the affinity of an RNA-protein interaction. In *Nucleic Acids Research* 15 (24), pp. 10483–10493. DOI: 10.1093/nar/15.24.10483.

Maciag, Anna; Peano, Clelia; Pietrelli, Alessandro; Egli, Thomas; Bellis, Gianluca de; Landini, Paolo (2011): In vitro transcription profiling of the σ S subunit of bacterial RNA polymerase: re-definition of the σ S regulon and identification of σ S-specific promoter sequence elements. In *Nucleic Acids Research* 39 (13), pp. 5338–5355. DOI: 10.1093/nar/gkr129.

Macomber, Lee; Imlay, James A. (2009): The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. In *Proceedings of the National Academy of Sciences of the United States of America* 106 (20), pp. 8344–8349. DOI: 10.1073/pnas.0812808106.

Marco-Marín, Clara; Gil-Ortiz, Fernando; Pérez-Arellano, Isabel; Cervera, Javier; Fita, Ignacio; Rubio, Vicente (2007): A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of Escherichia coli glutamate 5-kinase. In *Journal of Molecular Biology* 367 (5), pp. 1431–1446. DOI: 10.1016/j.jmb.2007.01.073.

Maris, Christophe; Dominguez, Cyril; Allain, Frédéric H-T (2005): The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. In *The FEBS journal* 272 (9), pp. 2118–2131. DOI: 10.1111/j.1742-4658.2005.04653.x.

Maruyama, Atsushi; Kumagai, Yoshito; Morikawa, Kazuya; Taguchi, Keiko; Hayashi, Hideo; Ohta, Toshiko (2003): Oxidative-stress-inducible qorA encodes an NADPH-dependent quinone oxidoreductase catalysing a one-electron reduction in Staphylococcus aureus. In *Microbiology (Reading, England)* 149 (Pt 2), pp. 389–398. DOI: 10.1099/mic.0.25796-0.

Matia-González, Ana M.; Laing, Emma E.; Gerber, André P. (2015): Conserved mRNA-binding proteomes in eukaryotic organisms. In *Nature structural & molecular biology* 22 (12), pp. 1027–1033. DOI: 10.1038/nsmb.3128.

McGowan, K.; Pekala, P. H. (1996): Dehydrogenase binding to the 3'-untranslated region of GLUT1 mRNA. In *Biochemical and biophysical research communications* 221 (1), pp. 42–45. DOI: 10.1006/bbrc.1996.0541.

Meydan, Sezen; Klepacki, Dorota; Karthikeyan, Subbulakshmi; Margus, Tõnu; Thomas, Paul; Jones, John E. et al. (2017): Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene. In *Molecular cell* 65 (2), pp. 207–219. DOI: 10.1016/j.molcel.2016.12.008.

Millet, Patrick; Vachharajani, Vidula; McPhail, Linda; Yoza, Barbara; McCall, Charles E. (2016): GAPDH Binding to TNF-α mRNA Contributes to Posttranscriptional Repression in Monocytes: A Novel Mechanism of Communication between Inflammation and Metabolism. In *Journal of immunology (Baltimore, Md. : 1950)* 196 (6), pp. 2541–2551. DOI: 10.4049/jimmunol.1501345.

Milojevic, Tetyana; Sonnleitner, Elisabeth; Romeo, Alessandra; Djinović-Carugo, Kristina; Bläsi, Udo (2013): False positive RNA binding activities after Ni-affinity purification from Escherichia coli. In *RNA biology* 10 (6), pp. 1066–1069. DOI: 10.4161/rna.25195.

Mitchell, Sarah F.; Jain, Saumya; She, Meipei; Parker, Roy (2013): Global analysis of yeast mRNPs. In *Nature structural & molecular biology* 20 (1), pp. 127–133. DOI: 10.1038/nsmb.2468.

Miyagishima, Shin-Ya (2005): Origin and evolution of the chloroplast division machinery. In *Journal of plant research* 118 (5), pp. 295–306. DOI: 10.1007/s10265-005-0226-2.

Muffler, A.; Fischer, D.; Hengge-Aronis, R. (1996): The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for rpoS translation in Escherichia coli. In *Genes & development* 10 (9), pp. 1143–1151. DOI: 10.1101/gad.10.9.1143.

Mullis, Kary B.; Faloona, Fred A. (1987): [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Methods in enzymology* 155, pp. 335–350. DOI: 10.1016/0076-6879(87)55023-6.

Nagy, E.; Rigby, W. F. (1995): Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). In *The Journal of biological chemistry* 270 (6), pp. 2755–2763. DOI: 10.1074/jbc.270.6.2755.

Nemecek, Daniel; Gilcrease, Eddie B.; Kang, Sebyung; Prevelige, Peter E.; Casjens, Sherwood; Thomas, George J. (2007): Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. In *Journal of Molecular Biology* 374 (3), pp. 817–836. DOI: 10.1016/j.jmb.2007.08.070.

Nordling, Erik; Jörnvall, Hans; Persson, Bengt (2002): Medium-chain dehydrogenases/reductases (MDR). Family characterizations including genome comparisons and active site modeling. In *European journal of biochemistry* 269 (17), pp. 4267–4276. DOI: 10.1046/j.1432-1033.2002.03114.x.

Ogura, M.; Kawata-Mukai, M.; Itaya, M.; Takio, K.; Tanaka, T. (1994): Multiple copies of the proB gene enhance degS-dependent extracellular protease production in Bacillus subtilis. In *Journal of bacteriology* 176 (18), pp. 5673–5680. DOI: 10.1128/jb.176.18.5673-5680.1994.

Ogura, M.; Tanaka, T. (1996): Transcription of Bacillus subtilis degR is sigma D dependent and suppressed by multicopy proB through sigma D. In *Journal of bacteriology* 178 (1), pp. 216–222. DOI: 10.1128/jb.178.1.216-222.1996.

Oivanen, Mikko; Kuusela, Satu; Lönnberg, Harri (1998): Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Brønsted Acids and Bases. In *Chemical reviews* 98 (3), pp. 961–990. DOI: 10.1021/cr960425x.

Osterholz, Hannah (2019): Kinetic Characterization of Metabolic Enzymes with Potential Moonlighting Function as RNA-Binders. Bacherlor's Thesis. University of Regensburg, Regensburg. Institute for Biophysics and Physical Biochemistry.

Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. (1995): How to measure and predict the molar absorption coefficient of a protein. In *Protein science : a publication of the Protein Society* 4 (11), pp. 2411–2423. DOI: 10.1002/pro.5560041120.

Palazzo, Alexander F.; Lee, Eliza S. (2015): Non-coding RNA: what is functional and what is junk? In *Frontiers in genetics* 6, p. 2. DOI: 10.3389/fgene.2015.00002.

Panja, Subrata; Woodson, Sarah A. (2012): Hexamer to monomer equilibrium of E. coli Hfq in solution and its impact on RNA annealing. In *Journal of Molecular Biology* 417 (5), pp. 406–412. DOI: 10.1016/j.jmb.2012.02.009.

Pérez-Arellano, Isabel; Gallego, José; Cervera, Javier (2007): The PUA domain - a structural and functional overview. In *The FEBS journal* 274 (19), pp. 4972–4984. DOI: 10.1111/j.1742-4658.2007.06031.x.

Pérez-Arellano, Isabel; Rubio, Vicente; Cervera, Javier (2005): Dissection of Escherichia coli glutamate 5-kinase: functional impact of the deletion of the PUA domain. In *FEBS letters* 579 (30), pp. 6903–6908. DOI: 10.1016/j.febslet.2005.11.037.

Persson, Bengt; Zigler, J. Samuel; Jörnvall, Hans (1994): A Super-Family of Medium-Chain Dehydrogenases/Reductases (MDR). In *European journal of biochemistry* 226 (1), pp. 15–22. DOI: 10.1111/j.1432-1033.1994.00t15.x.

Piccinelli, Paul; Samuelsson, Tore (2007): Evolution of the iron-responsive element. In *RNA (New York, N.Y.)* 13 (7), pp. 952–966. DOI: 10.1261/rna.464807.

Pickett, G. G.; Peabody, D. S. (1993): Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein. In *Nucleic Acids Research* 21 (19), pp. 4621–4626. DOI: 10.1093/nar/21.19.4621.

Pioli, Patricia A.; Hamilton, B. JoNell; Connolly, John E.; Brewer, Gary; Rigby, William F. C. (2002): Lactate dehydrogenase is an AU-rich element-binding protein that directly interacts with AUF1. In *The Journal of biological chemistry* 277 (38), pp. 35738–35745. DOI: 10.1074/jbc.M204002200.

Pomposiello, Pablo J.; Koutsolioutsou, Anastasia; Carrasco, Daniel; Demple, Bruce (2003): SoxRS-regulated expression and genetic analysis of the yggX gene of Escherichia coli. In *Journal of bacteriology* 185 (22), pp. 6624–6632. DOI: 10.1128/JB.185.22.6624-6632.2003.

Porté, Sergio; Crosas, Eva; Yakovtseva, Evgenia; Biosca, Josep A.; Farrés, Jaume; Fernández, M. Rosario; Parés, Xavier (2009): MDR quinone oxidoreductases: the human and yeast zeta-crystallins. In *Chemico-biological interactions* 178 (1-3), pp. 288–294. DOI: 10.1016/j.cbi.2008.10.018.

Potts, Anastasia H.; Vakulskas, Christopher A.; Pannuri, Archana; Yakhnin, Helen; Babitzke, Paul; Romeo, Tony (2017): Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. In *Nature communications* 8 (1), p. 1596. DOI: 10.1038/s41467-017-01613-1.

Pu, Feifei; Liu, Jianxiang; Jing, Doudou; Chen, Fengxia; Huang, Xin; Shi, Deyao et al. (2022): LncCCAT1 interaction protein PKM2 upregulates SREBP2 phosphorylation to promote osteosarcoma tumorigenesis by enhancing the Warburg effect and lipogenesis. In *International journal of oncology* 60 (4). DOI: 10.3892/ijo.2022.5334.

Py, B.; Higgins, C. F.; Krisch, H. M.; Carpousis, A. J. (1996): A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. In *Nature* 381 (6578), pp. 169–172. DOI: 10.1038/381169a0.

Queiroz, Rayner M. L.; Smith, Tom; Villanueva, Eneko; Marti-Solano, Maria; Monti, Mie; Pizzinga, Mariavittoria et al. (2019): Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). In *Nature biotechnology* 37 (2), pp. 169–178. DOI: 10.1038/s41587-018-0001-2.

Rao, P. V.; Krishna, C. M.; Zigler, J. S. (1992): Identification and characterization of the enzymatic activity of zeta-crystallin from guinea pig lens. A novel NADPH:quinone oxidoreductase. In *Journal of Biological Chemistry* 267 (1), pp. 96–102. DOI: 10.1016/S0021-9258(18)48464-5.

Rao, P. V.; Zigler, J. S. (1992): Purification and characterization of zeta-crystallin/quinone reductase from guinea pig liver. In *Biochimica et biophysica acta* 1117 (3), pp. 315–320.

Reinert, Line S.; Shi, Bo; Nandi, Suvobroto; Mazan-Mamczarz, Krystyna; Vitolo, Michele; Bachman, Kurtis E. et al. (2006): MCT-1 protein interacts with the cap complex and modulates messenger RNA translational profiles. In *Cancer research* 66 (18), pp. 8994–9001. DOI: 10.1158/0008-5472.CAN-06-1999.

Rio, Donald C. (2012): Filter-binding assay for analysis of RNA-protein interactions. In *Cold Spring Harbor protocols* 2012 (10), pp. 1078–1081. DOI: 10.1101/pdb.prot071449.

Riveros-Rosas, Héctor; Julián-Sánchez, Adriana; Villalobos-Molina, Rafael; Pardo, Juan Pablo; Piña, Enrique (2003): Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. In *European journal of biochemistry* 270 (16), pp. 3309–3334. DOI: 10.1046/j.1432-1033.2003.03704.x.

Roberts, Elijah; Sethi, Anurag; Montoya, Jonathan; Woese, Carl R.; Luthey-Schulten, Zaida (2008): Molecular signatures of ribosomal evolution. In *Proceedings of the National Academy of Sciences of the United States of America* 105 (37), pp. 13953–13958. DOI: 10.1073/pnas.0804861105.

Robertson, Michael P.; Joyce, Gerald F. (2014): Highly efficient self-replicating RNA enzymes. In *Chemistry & biology* 21 (2), pp. 238–245. DOI: 10.1016/j.chembiol.2013.12.004.

Robinson, James T.; Thorvaldsdóttir, Helga; Winckler, Wendy; Guttman, Mitchell; Lander, Eric S.; Getz, Gad; Mesirov, Jill P. (2011): Integrative genomics viewer. In *Nat Biotechnol* 29 (1), pp. 24–26. DOI: 10.1038/nbt.1754.

Rodriguez, Ignacio R.; Gonzales, Pedro; Samuel Zigler, J.; Borrás, Teresa (1992): A guinea-pig hereditary cataract contains a splice-site deletion in a crystallin gene. In *Biochimica et Biophysica*

Acta (BBA) - Molecular Basis of Disease 1180 (1), pp. 44–52. DOI: 10.1016/0925-4439(92)90025-I.

Rodríguez-Pascual, Fernando; Redondo-Horcajo, Mariano; Magán-Marchal, Noemi; Lagares, David; Martínez-Ruiz, Antonio; Kleinert, Hartmut; Lamas, Santiago (2008): Glyceraldehyde-3-phosphate dehydrogenase regulates endothelin-1 expression by a novel, redox-sensitive mechanism involving mRNA stability. In *Molecular and Cellular Biology* 28 (23), pp. 7139–7155. DOI: 10.1128/MCB.01145-08.

Rogozin, Igor B.; Makarova, Kira S.; Natale, Darren A.; Spiridonov, Alexey N.; Tatusov, Roman L.; Wolf, Yuri I. et al. (2002): Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. In *Nucleic Acids Research* 30 (19), pp. 4264–4271. DOI: 10.1093/nar/gkf549.

Romaniuk, P. J.; Lowary, P.; Wu, H. N.; Stormo, G.; Uhlenbeck, O. C. (1987): RNA binding site of R17 coat protein. In *Biochemistry* 26 (6), pp. 1563–1568. DOI: 10.1021/bi00380a011.

Sassanfar, M.; Szostak, J. W. (1993): An RNA motif that binds ATP. In *Nature* 364 (6437), pp. 550–553. DOI: 10.1038/364550a0.

Scherrer, Tanja; Mittal, Nitish; Janga, Sarath Chandra; Gerber, André P. (2010): A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. In *PloS one* 5 (11), e15499. DOI: 10.1371/journal.pone.0015499.

Schlattner, U.; Wagner, E.; Greppin, H.; Bonzon, M. (1995): Binding of adenylate kinase to RNA. In *Biochemical and biophysical research communications* 217 (2), pp. 509–514. DOI: 10.1006/bbrc.1995.2805.

Schneider, Daniel (2011): Identifikation und funktionelle Charakterisierung von TrpD2, einer neuen Klasse nukleinsäurebindender Proteine. PhD thesis. University of Regensburg, Regensburg.

Schneider, Daniel; Kaiser, Wolfgang; Stutz, Cian; Holinski, Alexandra; Mayans, Olga; Babinger, Patrick (2015): YbiB from Escherichia coli, the Defining Member of the Novel TrpD2 Family of Prokaryotic DNA-binding Proteins. In *The Journal of biological chemistry* 290 (32), pp. 19527–19539. DOI: 10.1074/jbc.M114.620575.

Schultz, D. E.; Hardin, C. C.; Lemon, S. M. (1996): Specific interaction of glyceraldehyde 3-phosphate dehydrogenase with the 5'-nontranslated RNA of hepatitis A virus. In *The Journal of biological chemistry* 271 (24), pp. 14134–14142. DOI: 10.1074/jbc.271.24.14134.

Seddon, A. P.; Zhao, K. Y.; Meister, A. (1989): Activation of glutamate by gamma-glutamate kinase: formation of gamma-cis-cycloglutamyl phosphate, an analog of gamma-glutamyl phosphate. In *Journal of Biological Chemistry* 264 (19), pp. 11326–11335.

Sharan, Malvika; Förstner, Konrad U.; Eulalio, Ana; Vogel, Jörg (2017): APRICOT: an integrated computational pipeline for the sequence-based identification and characterization of RNA-binding proteins. In *Nucleic Acids Research* 45 (11), e96. DOI: 10.1093/nar/gkx137.

Shchepachev, Vadim; Bresson, Stefan; Spanos, Christos; Petfalski, Elisabeth; Fischer, Lutz; Rappsilber, Juri; Tollervey, David (2019): Defining the RNA interactome by total RNA-associated protein purification. In *Molecular systems biology* 15 (4), e8689. DOI: 10.15252/msb.20188689.

Shetty, Sreerama; Muniyappa, Harish; Halady, Prathap K. S.; Idell, Steven (2004): Regulation of urokinase receptor expression by phosphoglycerate kinase. In *American journal of respiratory cell and molecular biology* 31 (1), pp. 100–106. DOI: 10.1165/rcmb.2003-0104OC.

Shtatland, T.; Gill, S. C.; Javornik, B. E.; Johansson, H. E.; Singer, B. S.; Uhlenbeck, O. C. et al. (2000): Interactions of Escherichia coli RNA with bacteriophage MS2 coat protein: genomic SELEX. In *Nucleic Acids Research* 28 (21), E93. DOI: 10.1093/nar/28.21.e93.

Simsek, Deniz; Tiu, Gerald C.; Flynn, Ryan A.; Byeon, Gun W.; Leppek, Kathrin; Xu, Adele F. et al. (2017): The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. In *Cell* 169 (6), 1051-1065.e18. DOI: 10.1016/j.cell.2017.05.022.

Singer, B. S.; Shtatland, T.; Brown, D.; Gold, L. (1997): Libraries for genomic SELEX. In *Nucleic Acids Research* 25 (4), pp. 781–786. DOI: 10.1093/nar/25.4.781.

Singh, R.; Green, M. R. (1993): Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. In *Science (New York, N.Y.)* 259 (5093), pp. 365–368. DOI: 10.1126/science.8420004.

Slomovic, Shimyn; Portnoy, Victoria; Liveanu, Varda; Schuster, Gadi (2006): RNA Polyadenylation in Prokaryotes and Organelles; Different Tails Tell Different Tales. In *Critical Reviews in Plant Sciences* 25 (1), pp. 65–77. DOI: 10.1080/07352680500391337.

Stafford, Graham P.; Ogi, Tomoo; Hughes, Colin (2005): Binding and transcriptional activation of non-flagellar genes by the Escherichia coli flagellar master regulator FlhD2C2. In *Microbiology (Reading, England)* 151 (Pt 6), pp. 1779–1788. DOI: 10.1099/mic.0.27879-0.

Sulzenbacher, Gerlind; Roig-Zamboni, Véronique; Pagot, Fabienne; Grisel, Sacha; Salomoni, Aurelia; Valencia, Christel et al. (2004): Structure of Escherichia coli YhdH, a putative quinone oxidoreductase. In *Acta crystallographica. Section D, Biological crystallography* 60 (Pt 10), pp. 1855–1862. DOI: 10.1107/S0907444904020220.

Swinehart, D. F. (1962): The Beer-Lambert Law. In *J. Chem. Educ.* 39 (7), p. 333. DOI: 10.1021/ed039p333.

Tai, Ningwen; Schmitz, John C.; Liu, Jun; Lin, Xiukun; Bailly, Michelle; Chen, Tian-min; Chu, Edward (2004): Translational autoregulation of thymidylate synthase and dihydrofolate reductase. In *Frontiers in bioscience : a journal and virtual library* 9, pp. 2521–2526. DOI: 10.2741/1413.

Tanaka, K.; Takayanagi, Y.; Fujita, N.; Ishihama, A.; Takahashi, H. (1993): Heterogeneity of the principal sigma factor in Escherichia coli: the rpoS gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase Escherichia coli. In *Proceedings*

of the National Academy of Sciences of the United States of America 90 (8), pp. 3511–3515. DOI: 10.1073/pnas.90.8.3511.

Tang, A.; Curthoys, N. P. (2001): Identification of zeta-crystallin/NADPH:quinone reductase as a renal glutaminase mRNA pH response element-binding protein. In *Journal of Biological Chemistry* 276 (24), pp. 21375–21380. DOI: 10.1074/jbc.M101941200.

Tang, Yue; Guest, John R. (1999): Direct evidence for mRNA binding and post-transcriptional regulation by Escherichia coli aconitases. In *Microbiology (Reading, England)* 145 (Pt 11), pp. 3069–3079. DOI: 10.1099/00221287-145-11-3069.

Thorn, J. M.; Barton, J. D.; Dixon, N. E.; Ollis, D. L.; Edwards, K. J. (1995): Crystal structure of Escherichia coli QOR quinone oxidoreductase complexed with NADPH. In *Journal of Molecular Biology* 249 (4), pp. 785–799. DOI: 10.1006/jmbi.1995.0337.

Tjhung, Katrina F.; Shokhirev, Maxim N.; Horning, David P.; Joyce, Gerald F. (2020): An RNA polymerase ribozyme that synthesizes its own ancestor. In *Proceedings of the National Academy of Sciences of the United States of America* 117 (6), pp. 2906–2913. DOI: 10.1073/pnas.1914282117.

Touchon, Marie; Bernheim, Aude; Rocha, Eduardo Pc (2016): Genetic and life-history traits associated with the distribution of prophages in bacteria. In *The ISME Journal* 10 (11), pp. 2744–2754. DOI: 10.1038/ismej.2016.47.

Towbin, H.; Staehelin, T.; Gordon, J. (1979): Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. In *Proceedings of the National Academy of Sciences of the United States of America* 76 (9), pp. 4350–4354. DOI: 10.1073/pnas.76.9.4350.

Trendel, Jakob; Schwarzl, Thomas; Horos, Rastislav; Prakash, Ananth; Bateman, Alex; Hentze, Matthias W.; Krijgsveld, Jeroen (2019): The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. In *Cell* 176 (1-2), 391-403.e19. DOI: 10.1016/j.cell.2018.11.004.

Tsvetanova, Nikoleta G.; Klass, Daniel M.; Salzman, Julia; Brown, Patrick O. (2010): Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. In *PloS one* 5 (9). DOI: 10.1371/journal.pone.0012671.

Uhlenbeck, O. C.; Carey, J.; Romaniuk, P. J.; Lowary, P. T.; Beckett, D. (1983): Interaction of R17 coat protein with its RNA binding site for translational repression. In *Journal of biomolecular structure & dynamics* 1 (2), pp. 539–552. DOI: 10.1080/07391102.1983.10507460.

Ule, Jernej; Jensen, Kirk; Mele, Aldo; Darnell, Robert B. (2005): CLIP: a method for identifying protein-RNA interaction sites in living cells. In *Methods (San Diego, Calif.)* 37 (4), pp. 376–386. DOI: 10.1016/j.ymeth.2005.07.018.

Urdaneta, Erika C.; Vieira-Vieira, Carlos H.; Hick, Timon; Wessels, Hans-Herrmann; Figini, Davide; Moschall, Rebecca et al. (2019): Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. In *Nature communications* 10 (1), p. 990. DOI: 10.1038/s41467-019-08942-3.

Vaishali; Dimitrova-Paternoga, Lyudmila; Haubrich, Kevin; Sun, Mai; Ephrussi, Anne; Hennig, Janosch (2021): Validation and classification of RNA binding proteins identified by mRNA interactome capture. In *RNA (New York, N.Y.)* 27 (10), pp. 1173–1185. DOI: 10.1261/rna.078700.121.

van der Gulik, Peter T. S.; Speijer, Dave (2015): How amino acids and peptides shaped the RNA world. In *Life (Basel, Switzerland)* 5 (1), pp. 230–246. DOI: 10.3390/life5010230.

Voeller, D. M.; Changchien, L. M.; Maley, G. F.; Maley, F.; Takechi, T.; Turner, R. E. et al. (1995): Characterization of a specific interaction between Escherichia coli thymidylate synthase and Escherichia coli thymidylate synthase mRNA. In *Nucleic Acids Research* 23 (5), pp. 869–875. DOI: 10.1093/nar/23.5.869.

Walden, William E.; Selezneva, Anna I.; Dupuy, Jérôme; Volbeda, Anne; Fontecilla-Camps, Juan C.; Theil, Elizabeth C.; Volz, Karl (2006): Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. In *Science (New York, N.Y.)* 314 (5807), pp. 1903–1908. DOI: 10.1126/science.1133116.

Wang, Xiaoxue; Kim, Younghoon; Ma, Qun; Hong, Seok Hoon; Pokusaeva, Karina; Sturino, Joseph M.; Wood, Thomas K. (2010): Cryptic prophages help bacteria cope with adverse environments. In *Nature communications* 1, p. 147. DOI: 10.1038/ncomms1146.

Wassarman, Karen Montzka; Storz, Gisela (2000): 6S RNA Regulates E. coli RNA Polymerase Activity. In *Cell* 101 (6), pp. 613–623. DOI: 10.1016/S0092-8674(00)80873-9.

Wegener, Melanie; Dietz, Karl-Josef (2022): The mutual interaction of glycolytic enzymes and RNA in post-transcriptional regulation. In *RNA (New York, N.Y.)* 28 (11), pp. 1446–1468. DOI: 10.1261/rna.079210.122.

Wheeler, Emily C.; van Nostrand, Eric L.; Yeo, Gene W. (2018): Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. In *Wiley interdisciplinary reviews. RNA* 9 (1). DOI: 10.1002/wrna.1436.

White, J. H.; Richardson, C. C. (1987): Gene 18 protein of bacteriophage T7. Overproduction, purification, and characterization. In *Journal of Biological Chemistry* 262 (18), pp. 8845–8850. DOI: 10.1016/S0021-9258(18)47492-3.

White, Michael R.; Garcin, Elsa D. (2016): The sweet side of RNA regulation: glyceraldehyde-3-phosphate dehydrogenase as a noncanonical RNA-binding protein. In *Wiley interdisciplinary reviews. RNA* 7 (1), pp. 53–70. DOI: 10.1002/wrna.1315.

White, Michael R.; Khan, Mohd M.; Deredge, Daniel; Ross, Christina R.; Quintyn, Royston; Zucconi, Beth E. et al. (2015): A dimer interface mutation in glyceraldehyde-3-phosphate dehydrogenase regulates its binding to AU-rich RNA. In *The Journal of biological chemistry* 290 (3), pp. 1770–1785. DOI: 10.1074/jbc.M114.618165.

Wilfinger, W. W.; Mackey, K.; Chomczynski, P. (1997): Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. In *BioTechniques* 22 (3), 474-6, 478-81. DOI: 10.2144/97223st01.

Xu, Yang; Chaudhury, Arindam; Zhang, Ming; Savoldo, Barbara; Metelitsa, Leonid S.; Rodgers, John et al. (2016): Glycolysis determines dichotomous regulation of T cell subsets in hypoxia. In *The Journal of Clinical Investigation* 126 (7), pp. 2678–2688. DOI: 10.1172/JCI85834.

Yi, Ying; Zhao, Yue; Huang, Yan; Wang, Dong (2017): A Brief Review of RNA-Protein Interaction Database Resources. In *Non-coding RNA* 3 (1). DOI: 10.3390/ncrna3010006.

Young, Tracy A.; Skordalakes, Emmanuel; Marqusee, Susan (2007): Comparison of proteolytic susceptibility in phosphoglycerate kinases from yeast and E. coli: modulation of conformational ensembles without altering structure or stability. In *Journal of Molecular Biology* 368 (5), pp. 1438–1447. DOI: 10.1016/j.jmb.2007.02.077.

Yu, D.; Ellis, H. M.; Lee, E. C.; Jenkins, N. A.; Copeland, N. G.; Court, D. L. (2000): An efficient recombination system for chromosome engineering in Escherichia coli. In *Proceedings of the National Academy of Sciences of the United States of America* 97 (11), pp. 5978–5983. DOI: 10.1073/pnas.100127597.

Yu, JianChao; Ramirez, Lisa M.; Premo, Aaron; Busch, Devin B.; Lin, Qishan; Burz, David S.; Shekhtman, Alexander (2021): Ribosome-Amplified Metabolism, RAMBO, Measured by NMR Spectroscopy. In *Biochemistry* 60 (24), pp. 1885–1895. DOI: 10.1021/acs.biochem.1c00074.

Zhao, Haiyan; Finch, Casey J.; Sequeira, Reuben D.; Johnson, Brian A.; Johnson, John E.; Casjens, Sherwood R.; Tang, Liang (2010): Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. In *Proceedings of the National Academy of Sciences of the United States of America* 107 (5), pp. 1971–1976. DOI: 10.1073/pnas.0908569107.

Zhao, Haiyan; Kamau, Yvonne N.; Christensen, Theodore E.; Tang, Liang (2012): Structural and functional studies of the phage Sf6 terminase small subunit reveal a DNA-spooling device facilitated by structural plasticity. In *Journal of Molecular Biology* 423 (3), pp. 413–426. DOI: 10.1016/j.jmb.2012.07.016.

Zhou, Yi; Yi, Xiaofang; Stoffer, Jha'nae B.; Bonafe, Nathalie; Gilmore-Hebert, Maureen; McAlpine, Jessica; Chambers, Setsuko K. (2008): The multifunctional protein glyceraldehyde-3-phosphate dehydrogenase is both regulated and controls colony-stimulating factor-1 messenger RNA stability in ovarian cancer. In *Molecular cancer research : MCR* 6 (8), pp. 1375–1384. DOI: 10.1158/1541-7786.MCR-07-2170.

Zimmerman, Steven B.; Trach, Stefan O. (1991): Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. In *Journal of Molecular Biology* 222 (3), pp. 599–620. DOI: 10.1016/0022-2836(91)90499-V.

7 Supplement

7.1 Summary of Protein Purifications

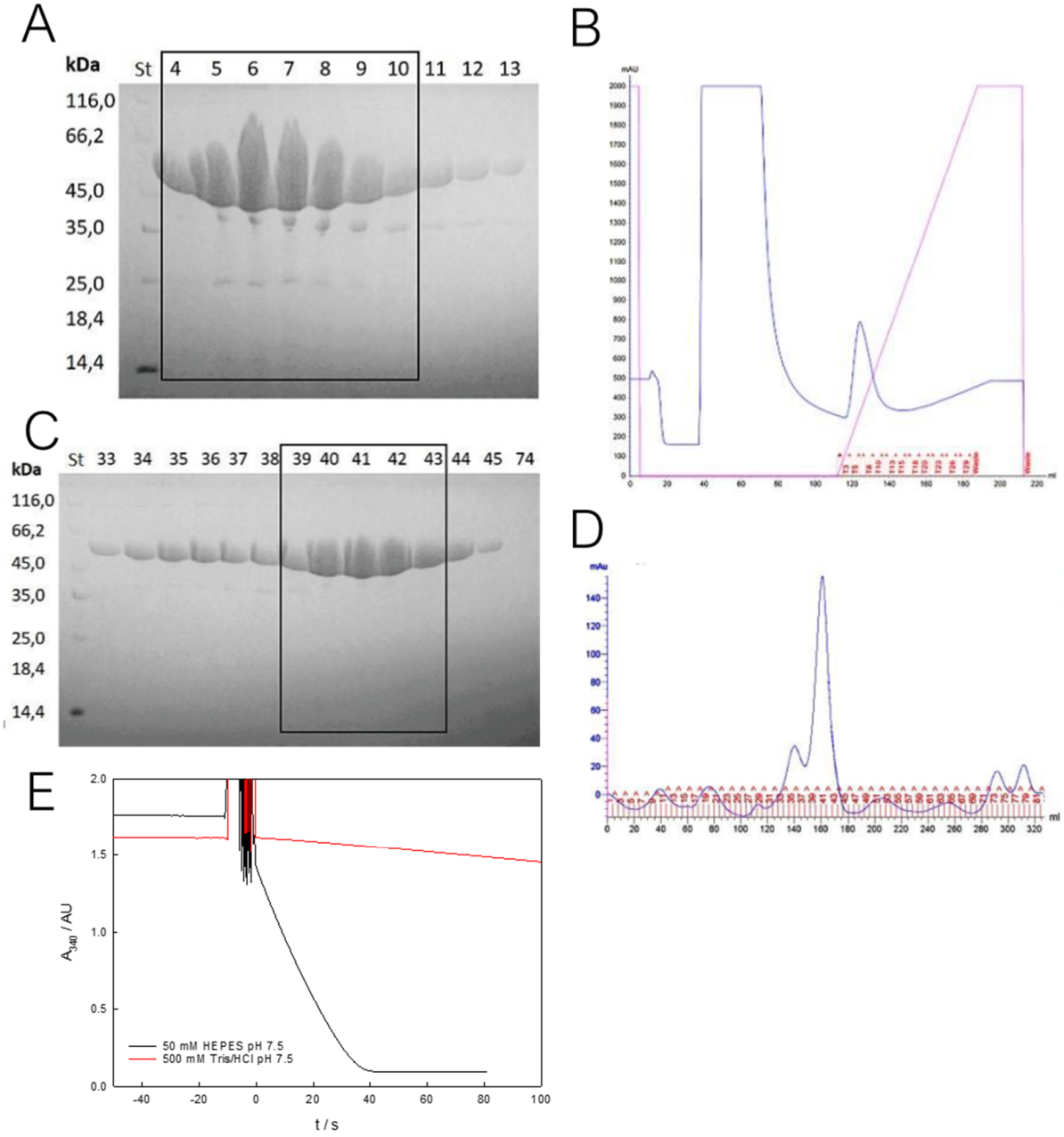


Figure S 1: Purification of Pykf. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Pooled fractions indicated by black rectangle. PykF has a molecular weight of 51 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (pink). **C:** S200 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S200 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of PykF. Change in $A_{340\ nm}$ indicates NADH oxidation. Comparison of Tris- and HEPES-buffer to demonstrate inhibitory effect of Tris on PykF. **Yield: 17.2 mg/l expression culture.** Purification conducted within the Master's Thesis of Franziska Funke (2019).

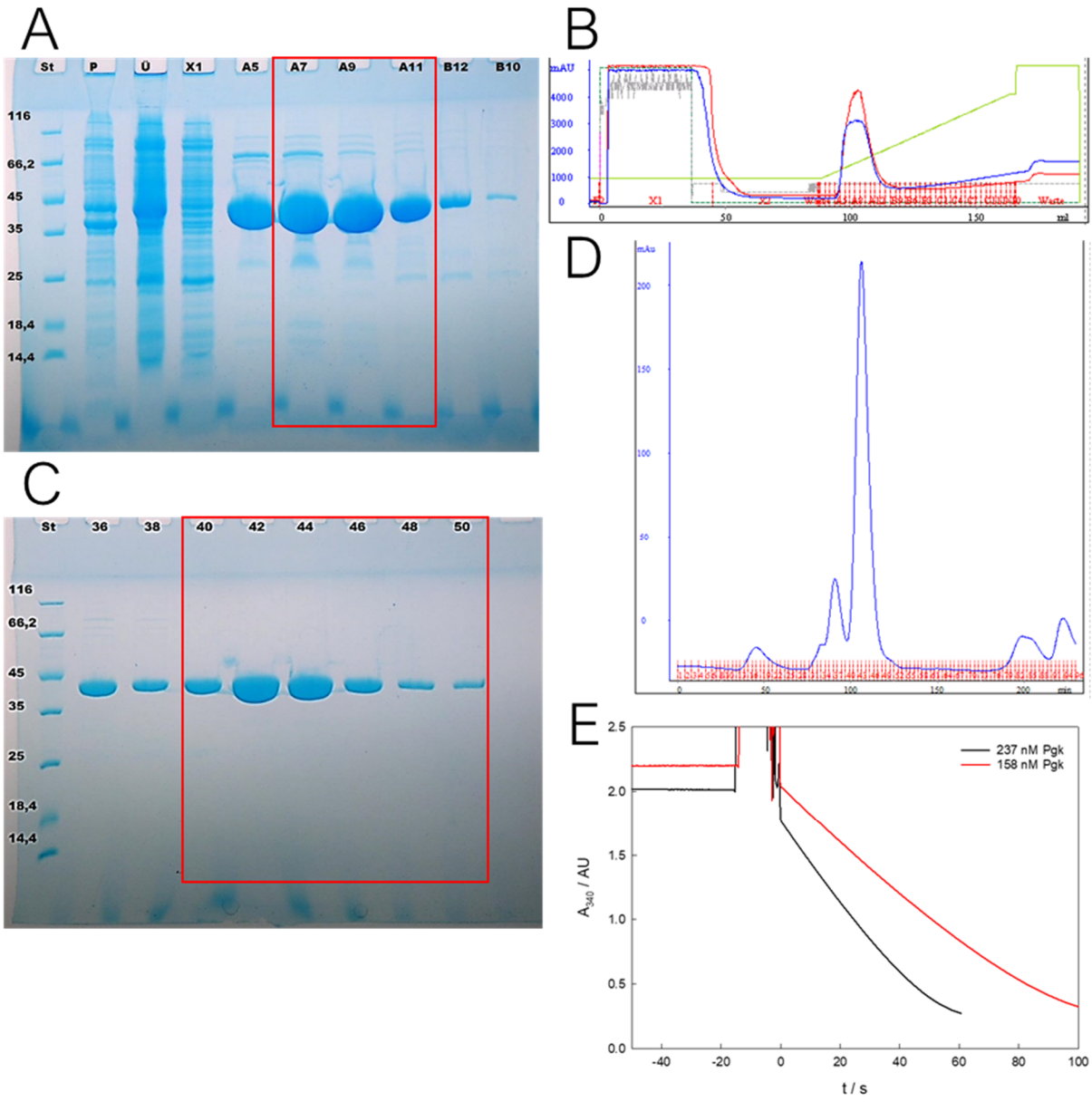


Figure S 2: Purification of Pgk. A: IMAC elution fractions on SDS-PAGE. ST – protein mass standard, P – pellet, Ü – crude extract, X1– flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. Pgk has a molecular weight of 41 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), $A_{260\ nm}$ (red), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of Pgk. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 41.3 mg/l expression culture.

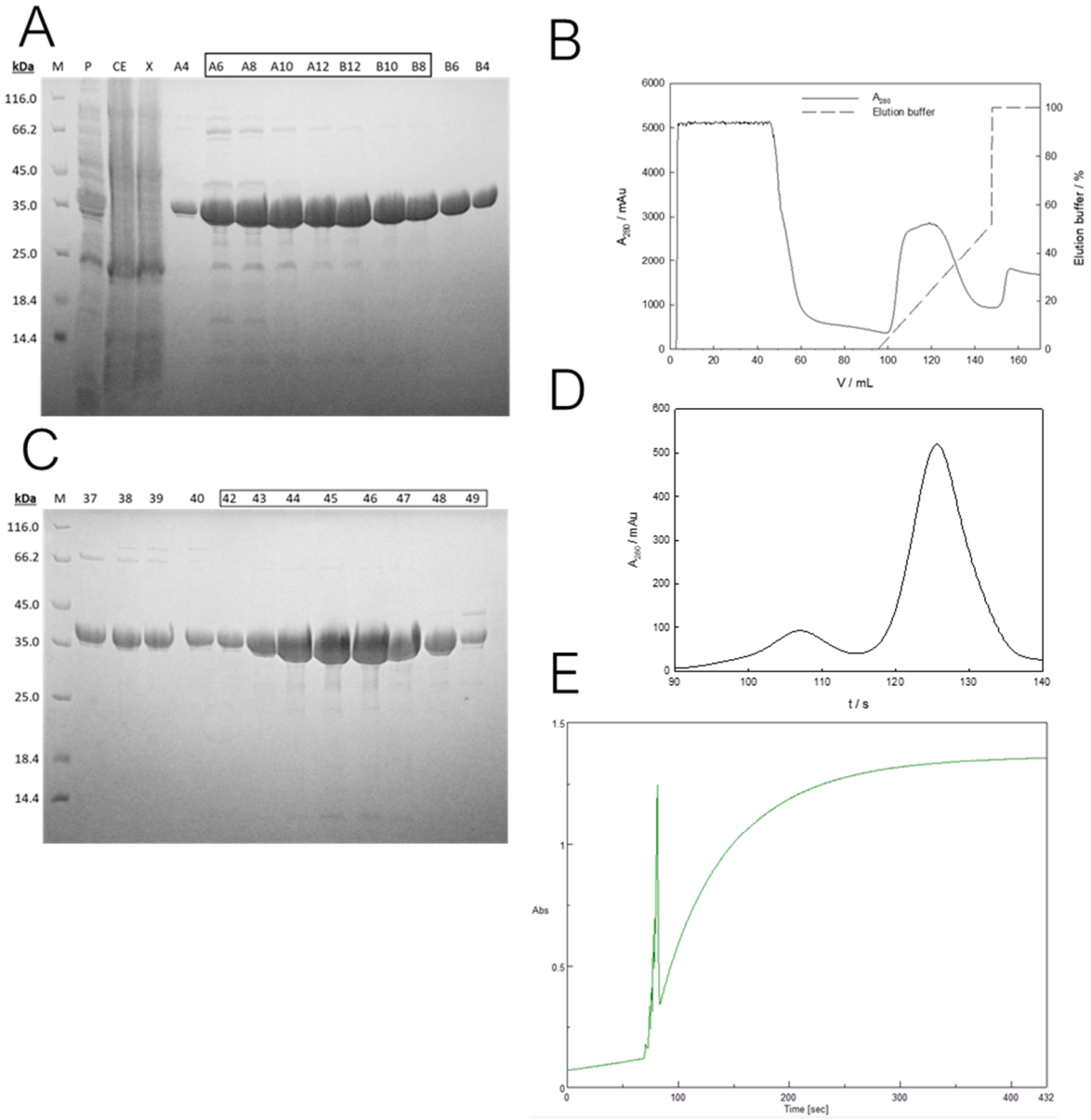


Figure S 3: Purification of GapA. A: IMAC elution fractions on SDS-PAGE. M – protein mass standard, P – pellet, CE – crude extract, X – flowthrough fraction, and elution fractions. Pooled fractions indicated by black rectangle. GapA has a molecular weight of 36 kDa. **B:** IMAC profile. $A_{280\,nm}$ (solid line), gradient of 750 mM imidazole (dashed line). **C:** S75 gel filtration elution fractions on SDS-PAGE. M – protein mass standard, and elution fractions. Fraction marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\,nm}$ (black line). **E:** Photometric assay indicating enzymatic activity of GapA. Change in $A_{340\,nm}$ indicates NAD^+ reduction.
Yield: 9.7 mg/l expression culture. Purification also presented in the Bachelor's Thesis of Hannah Osterholz (2019).

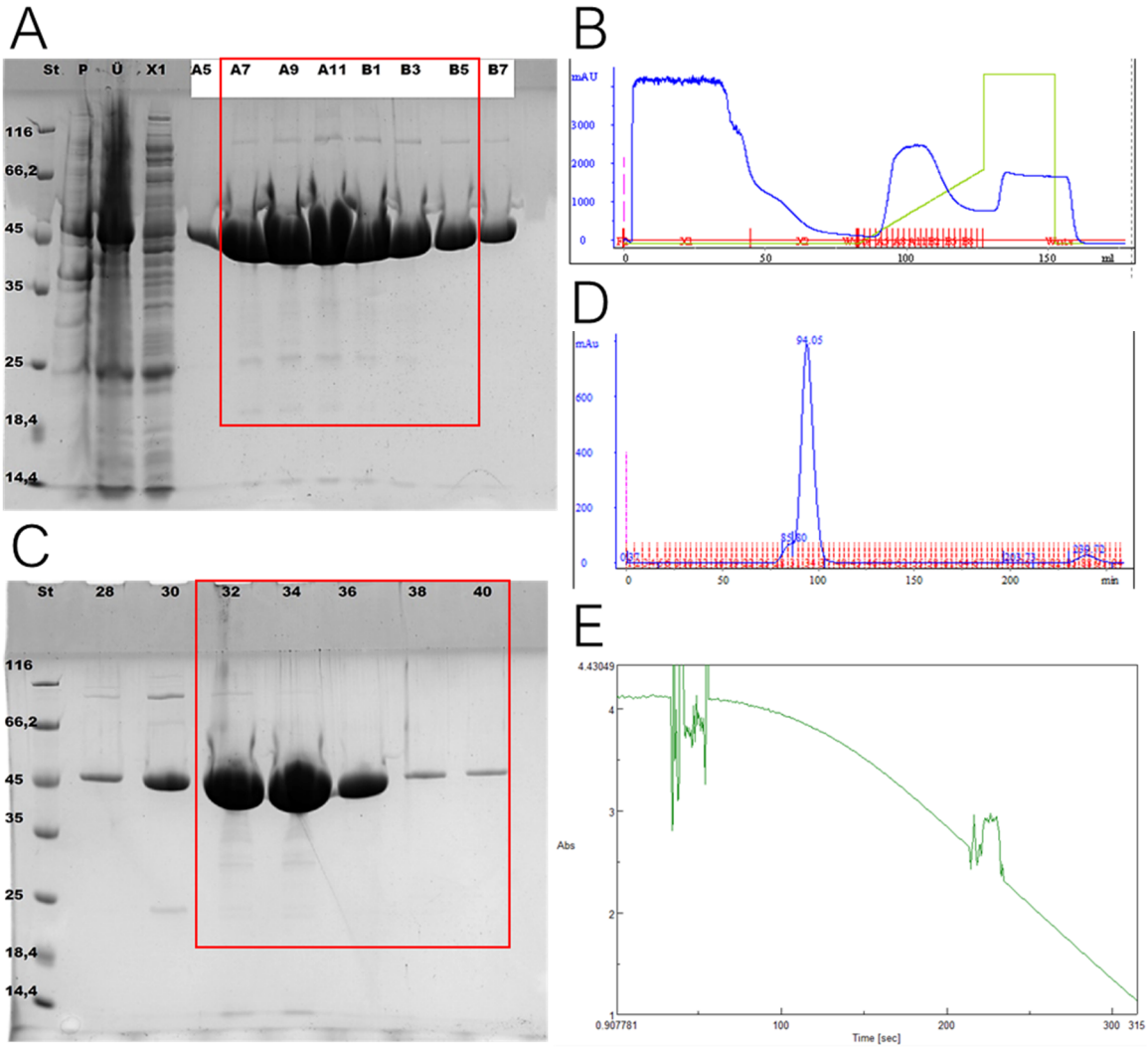


Figure S 4: Purification of Eno. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. Eno has a molecular weight of 47 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of Eno. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 39.6 mg/l expression culture.

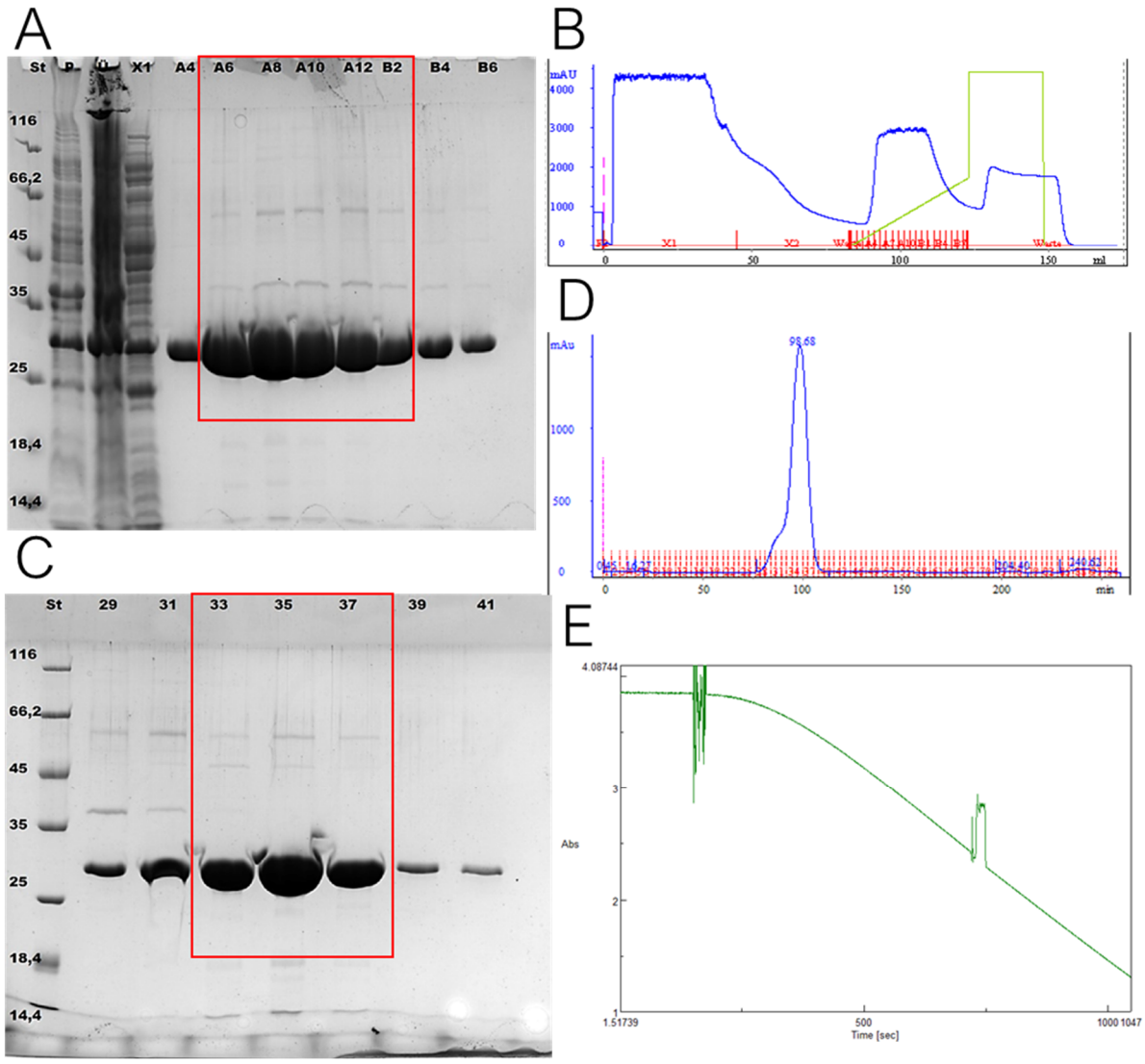


Figure S 5: Purification of GpmA. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. TalA has a molecular weight of 30 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of GpmA. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 31.2 mg/l expression culture.

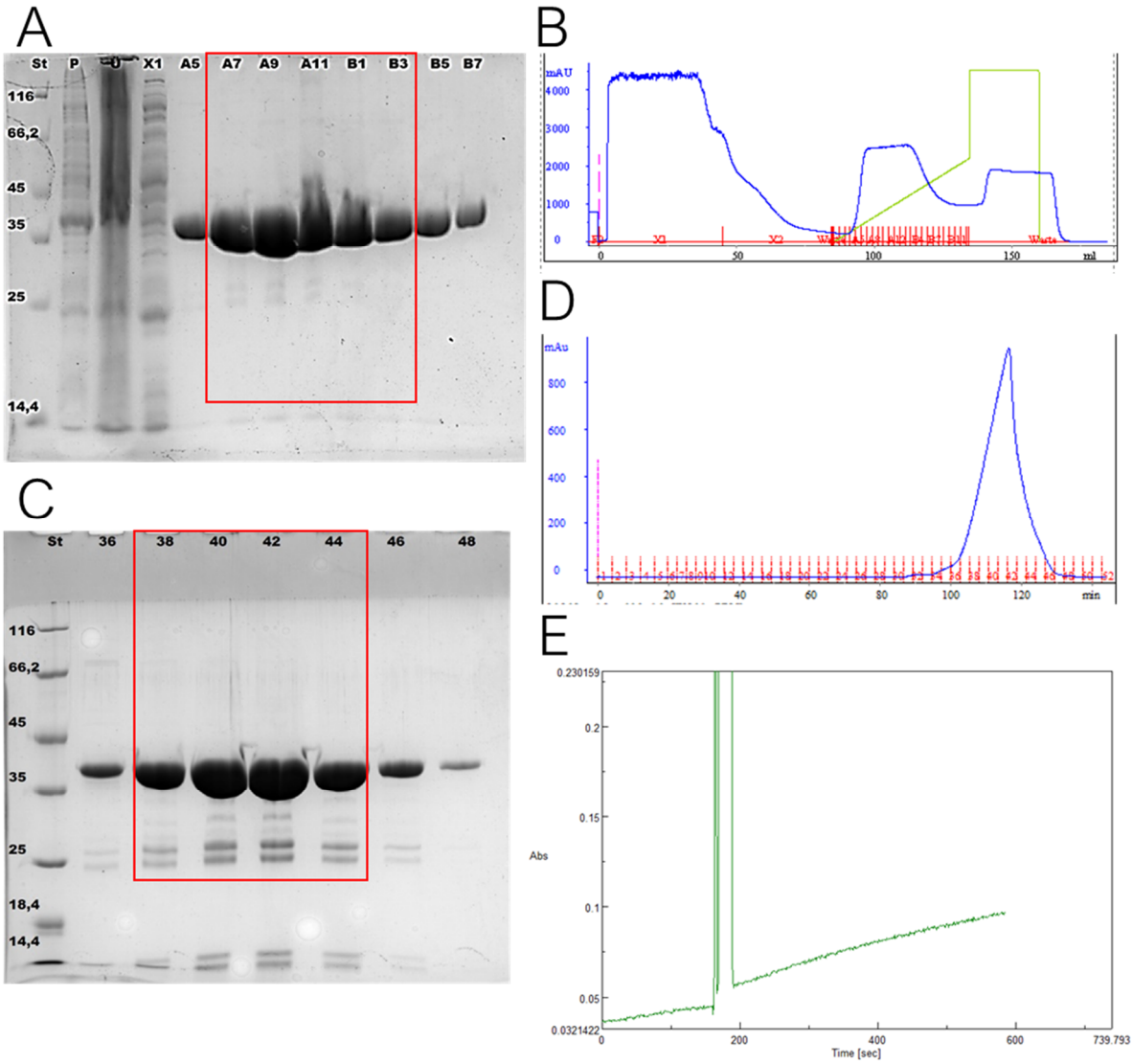


Figure S 6: Purification of TalA. A: IMAC elution fractions on SDS-PAGE. ST – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. TalA has a molecular weight of 37 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of TalA. Change in $A_{340\ nm}$ indicates NAD^+ reduction.
Yield: 39.6 mg/l expression culture.

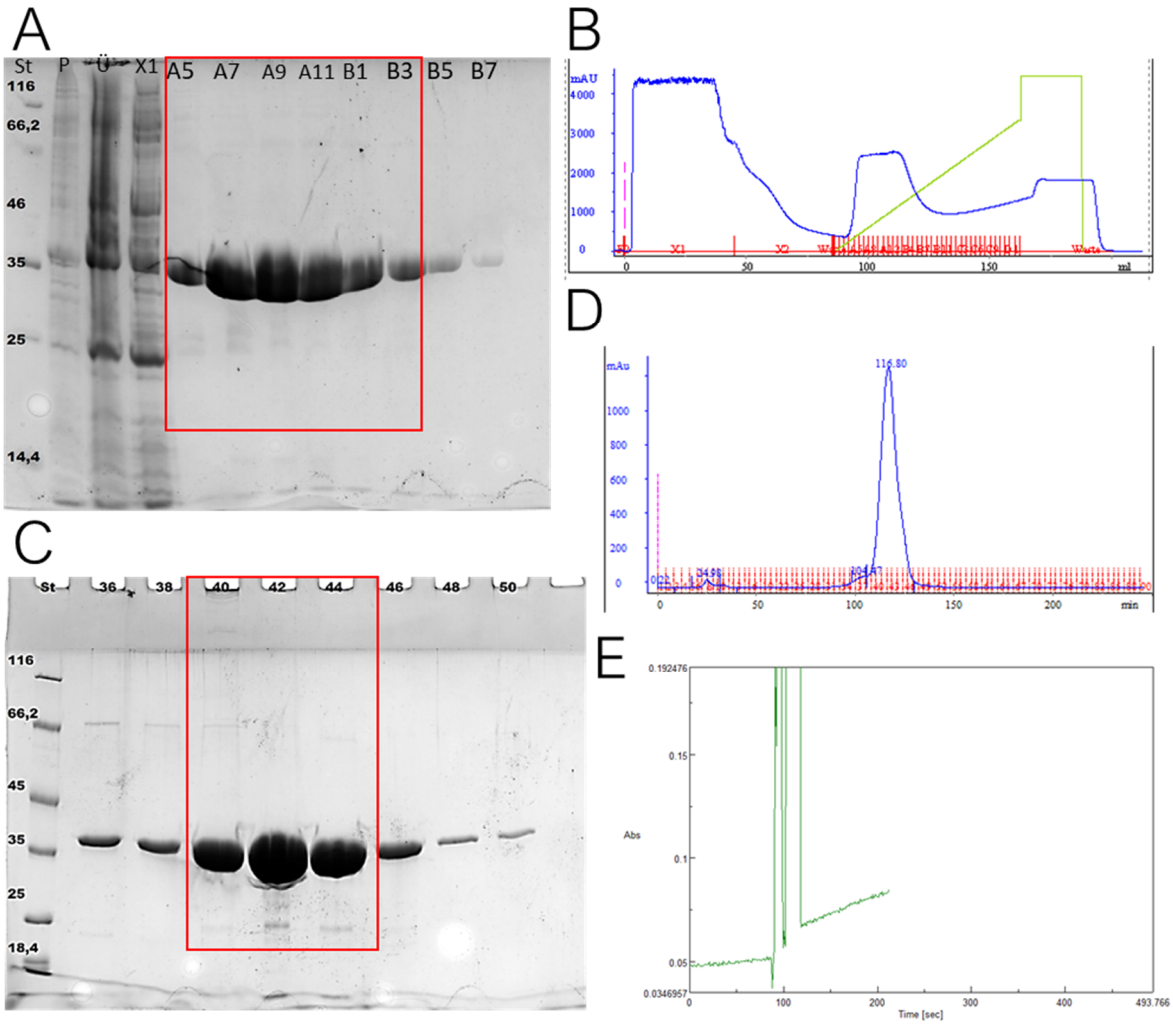


Figure S 7: Purification of TalB. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. TalB has a molecular weight of 37 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of TalB. Change in $A_{340\ nm}$ indicates NAD^+ reduction.
Yield: 44.1 mg/l expression culture.

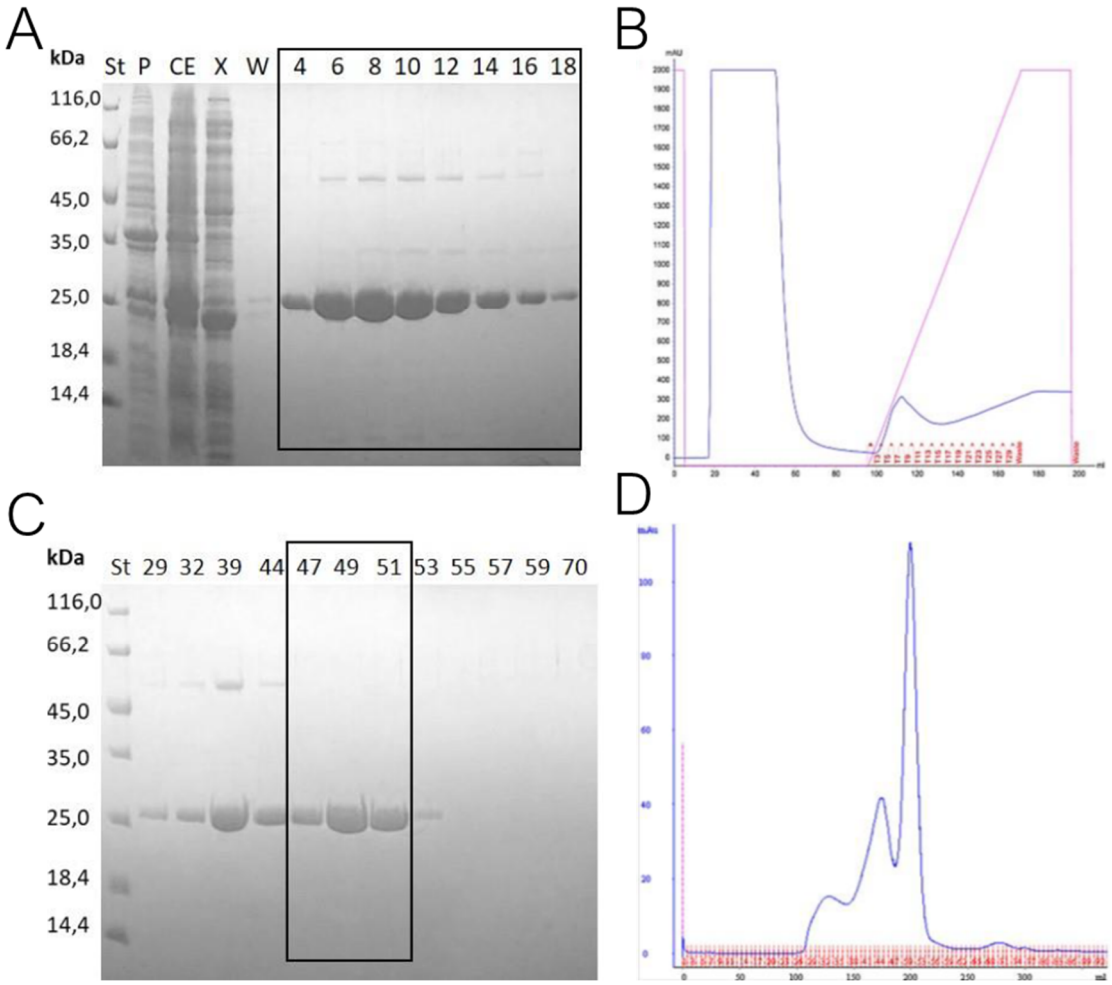


Figure S 8: Purification of RpiA. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, CE – crude extract, X – flowthrough fraction, W – waste, and elution fractions. Pooled fractions indicated by black rectangle. RpiA has a molecular weight of 23 kDa. **B:** IMAC profile. $A_{280\,nm}$ (blue), gradient of 750 mM imidazole (pink). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\,nm}$ (blue).
Yield: 6.1 mg/l expression culture. Purification conducted within the Master's Thesis of Franziska Funke (2019).

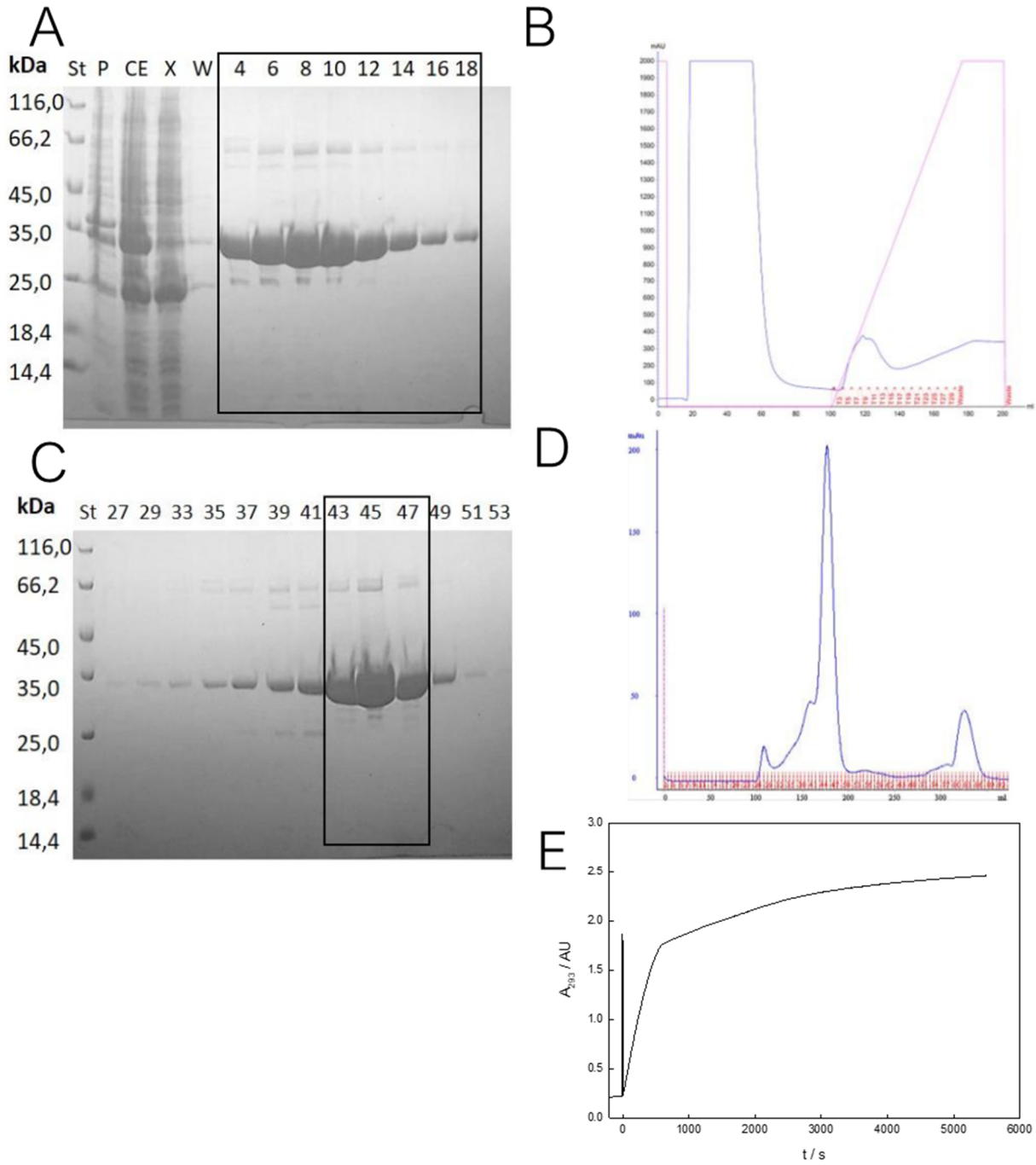


Figure S 9: Purification of KdsA. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, CE – crude extract, X – flowthrough fraction, W – waste, and elution fractions. Pooled fractions indicated by black rectangle. KdsA has a molecular weight of 31 kDa. **B:** IMAC profile. $A_{280\,nm}$ (blue), gradient of 750 mM imidazole (pink). **C:** S200 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S200 gel filtration profile. $A_{280\,nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of KdsA. Change in $A_{293\,nm}$ indicates xanthine-oxidation to uric acid. **Yield: 16.0 mg/l expression culture.** Purification conducted within the Master's Thesis of Franziska Funke (2019).

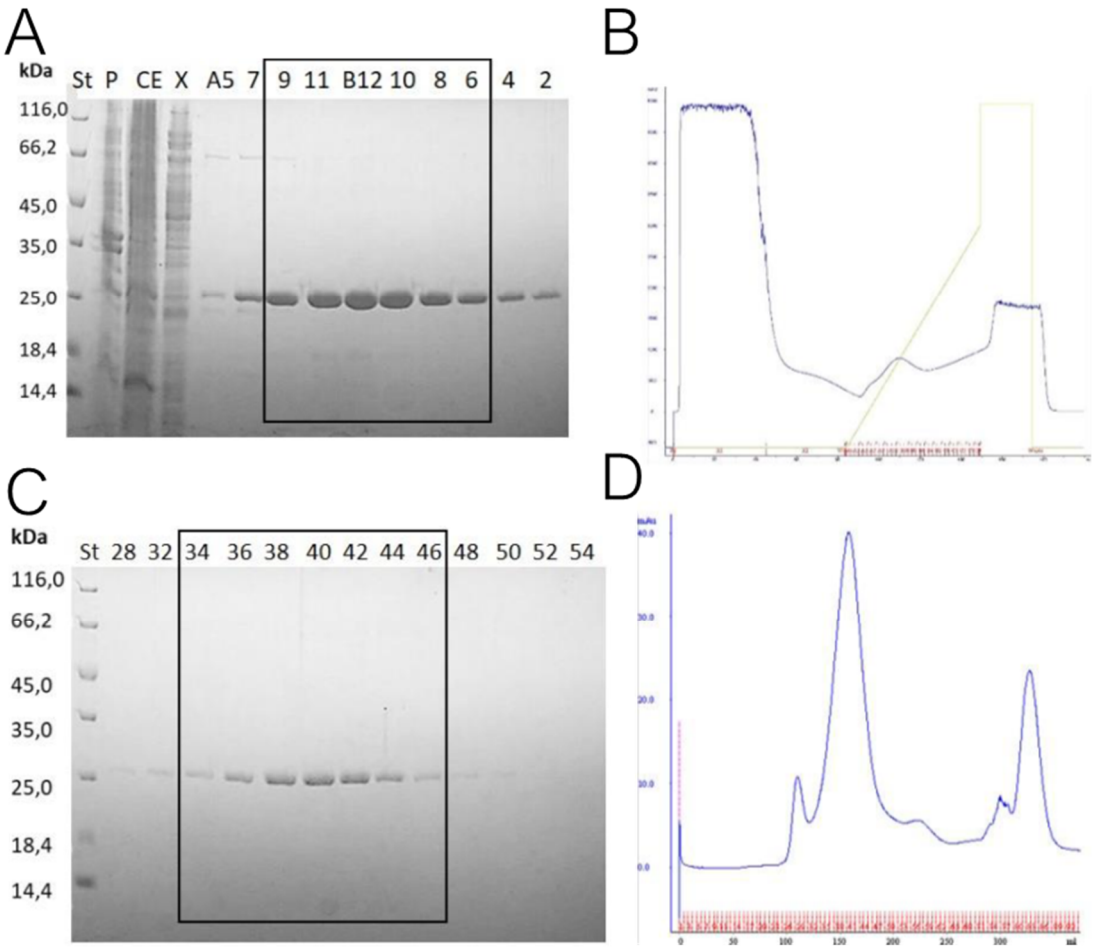


Figure S 10: Purification of Upp. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, CE – crude extract, X – flowthrough fraction, and elution fractions. Pooled fractions indicated by black rectangle. Upp has a molecular weight of 23 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue).
Yield: 12.2 mg/l expression culture. Purification conducted within the Master's Thesis of Franziska Funke (2019).

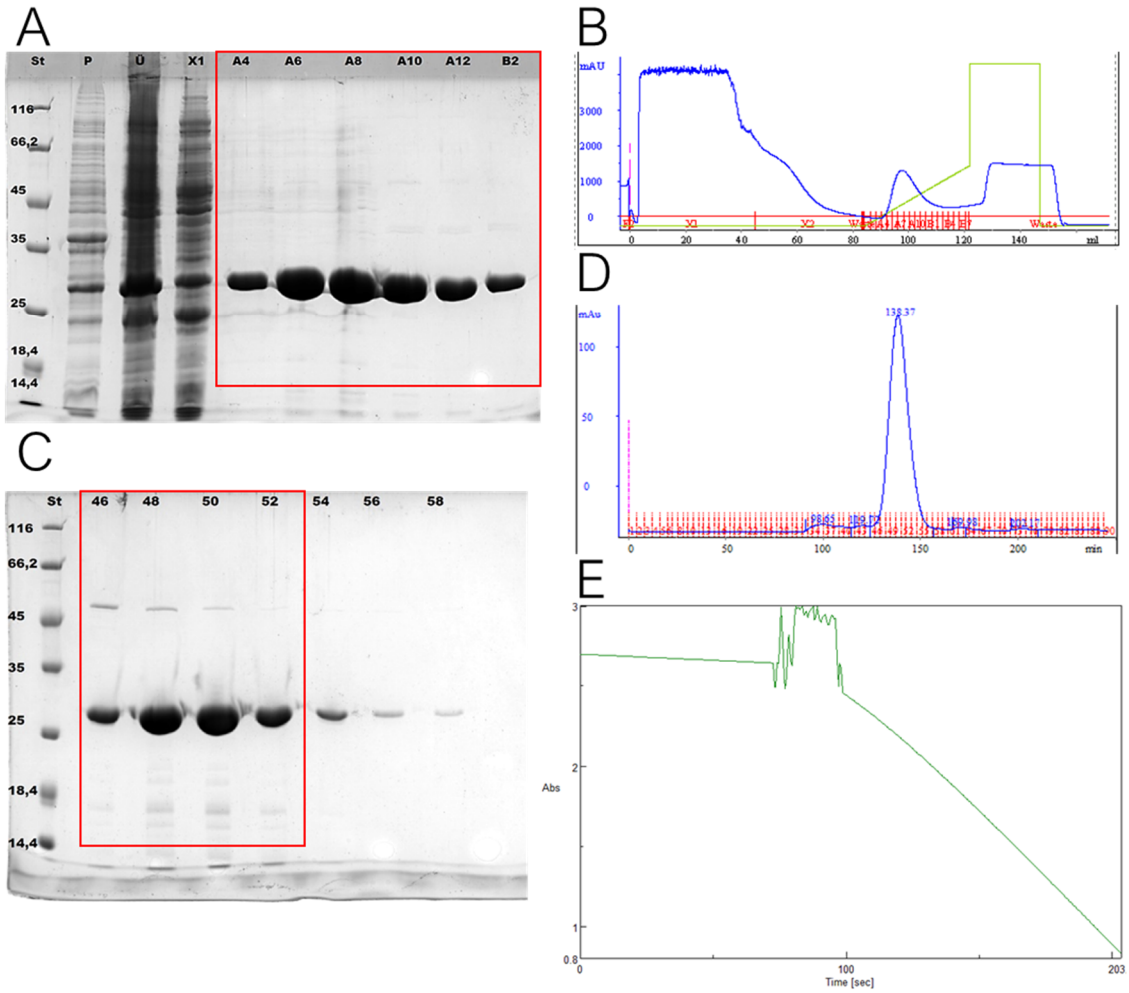


Figure S 11: Purification of Adk. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. Adk has a molecular weight of 25 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of Adk. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 16.2 mg/l expression culture.

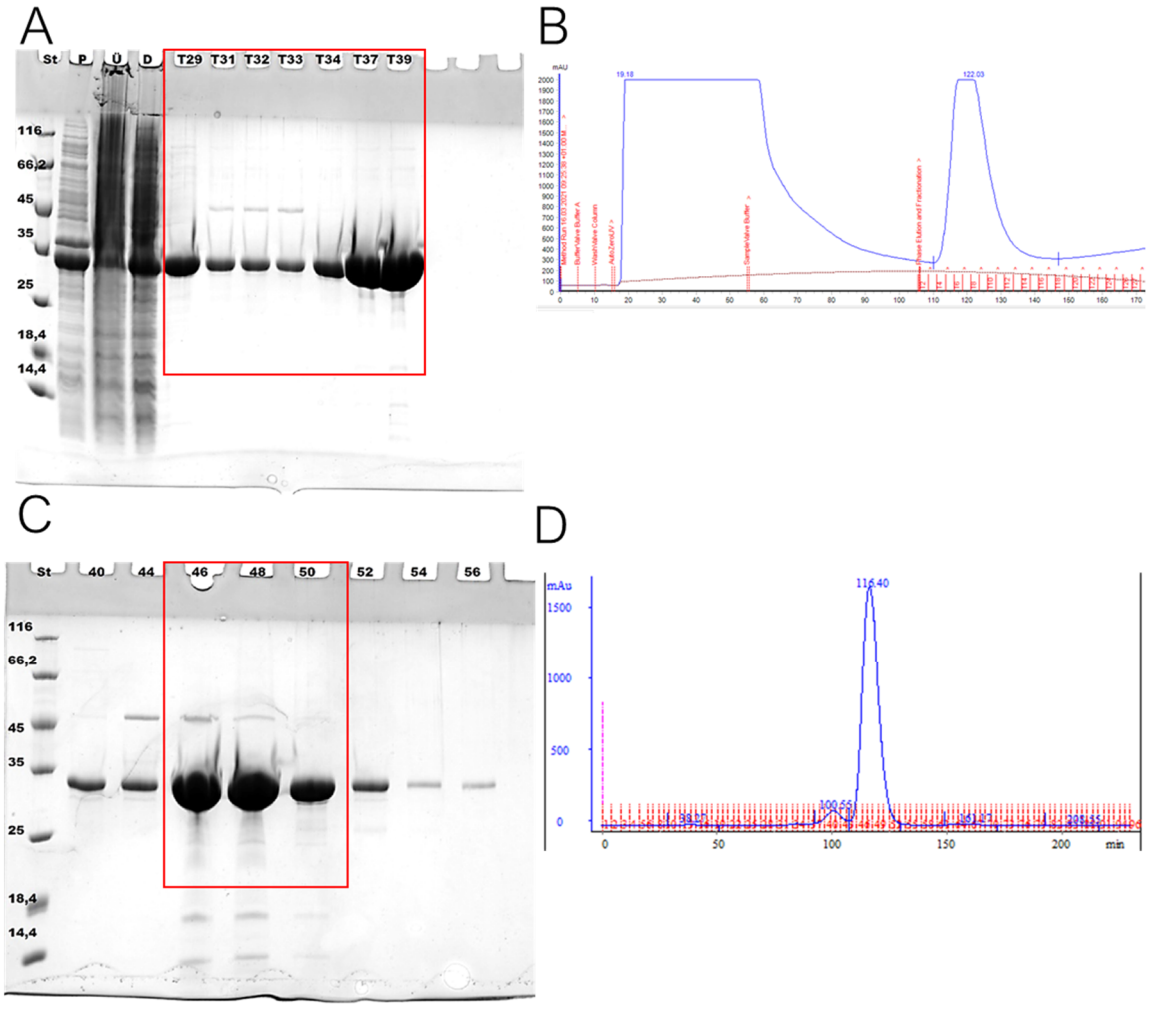


Figure S 12: Purification of ThyA. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, D – flowthrough fractions, and elution fractions. Pooled fractions indicated by red rectangle. ThyA has a molecular weight of 31 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue).
Yield: 37.4 mg/l expression culture.

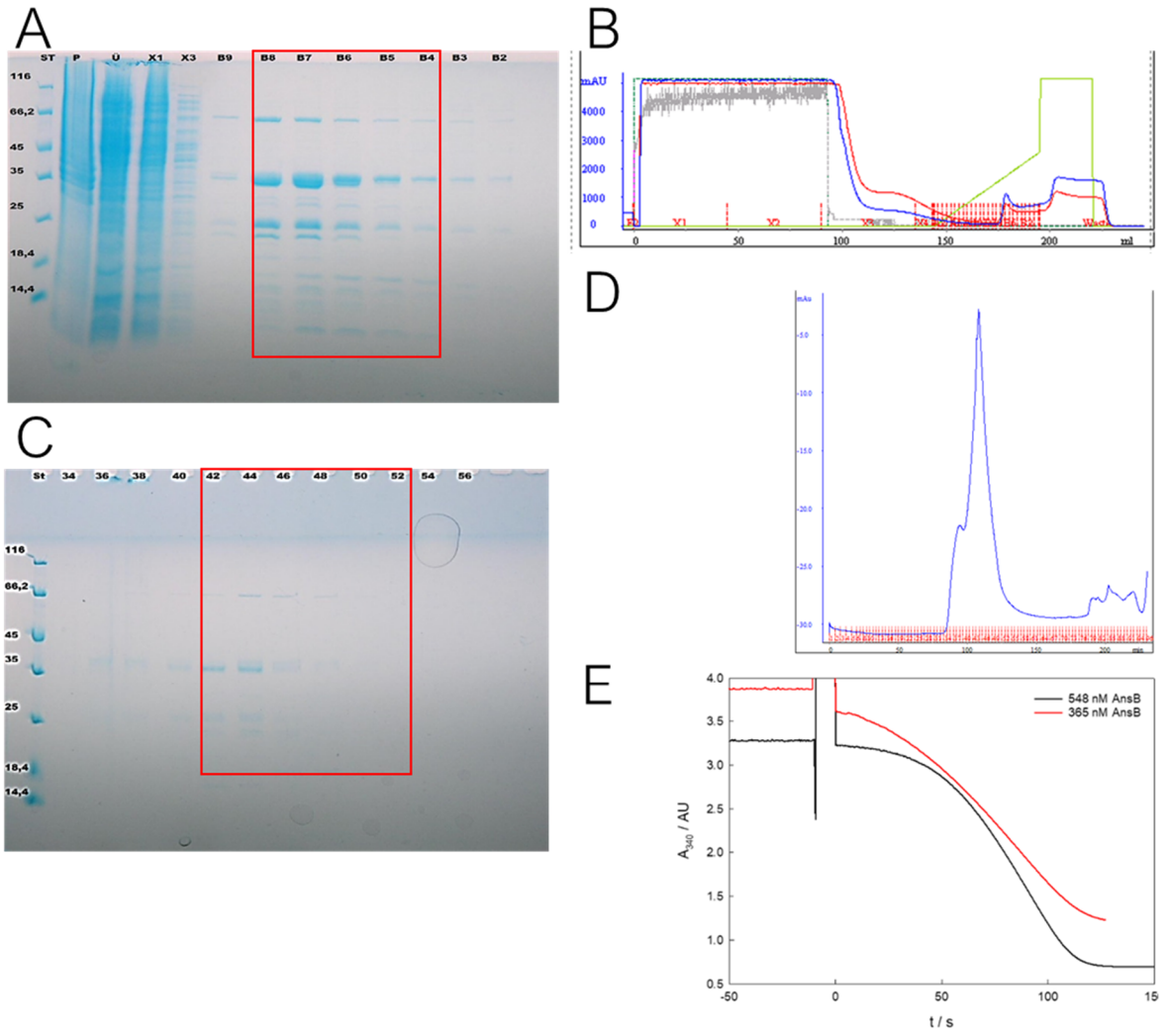


Figure S 13: Purification of AnsB. A: IMAC elution fractions on SDS-PAGE. ST – protein mass standard, P – pellet, Ü – crude extract, X1, X3 – flowthrough fractions, and elution fractions. Pooled fractions indicated by red rectangle. AnsB has a molecular weight of 37 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), $A_{260\ nm}$ (red), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of AnsB. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 0.53 mg/l expression culture.

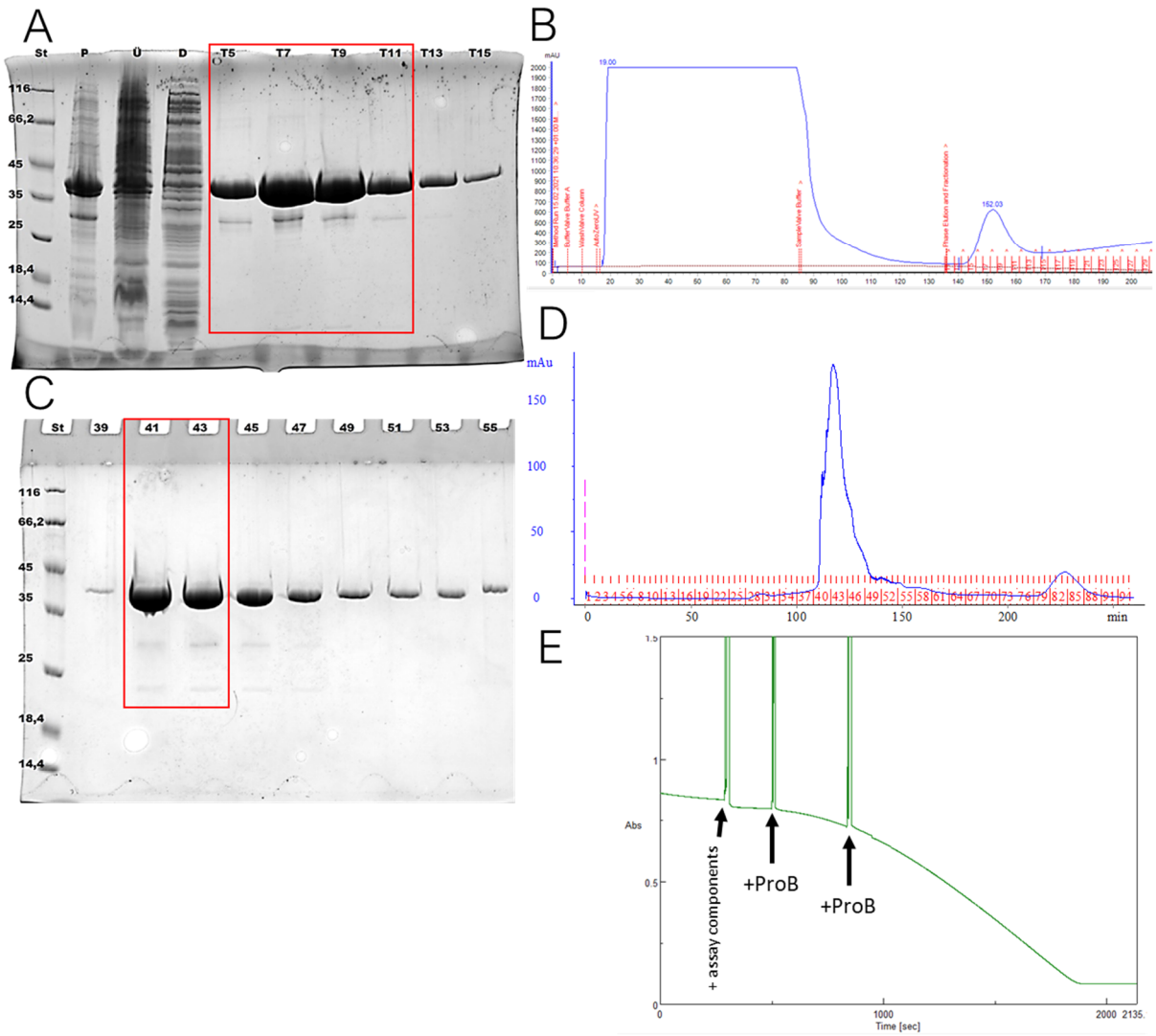


Figure S 14: Purification of ProB. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, D – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. ProB has a molecular weight of 40 kDa. **B:** IMAC profile. $A_{280\,nm}$ (blue). **C:** S200 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S200 gel filtration profile. $A_{280\,nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of ProB. Change in $A_{340\,nm}$ indicates NADH oxidation.
Yield: 8.6 mg/l expression culture.

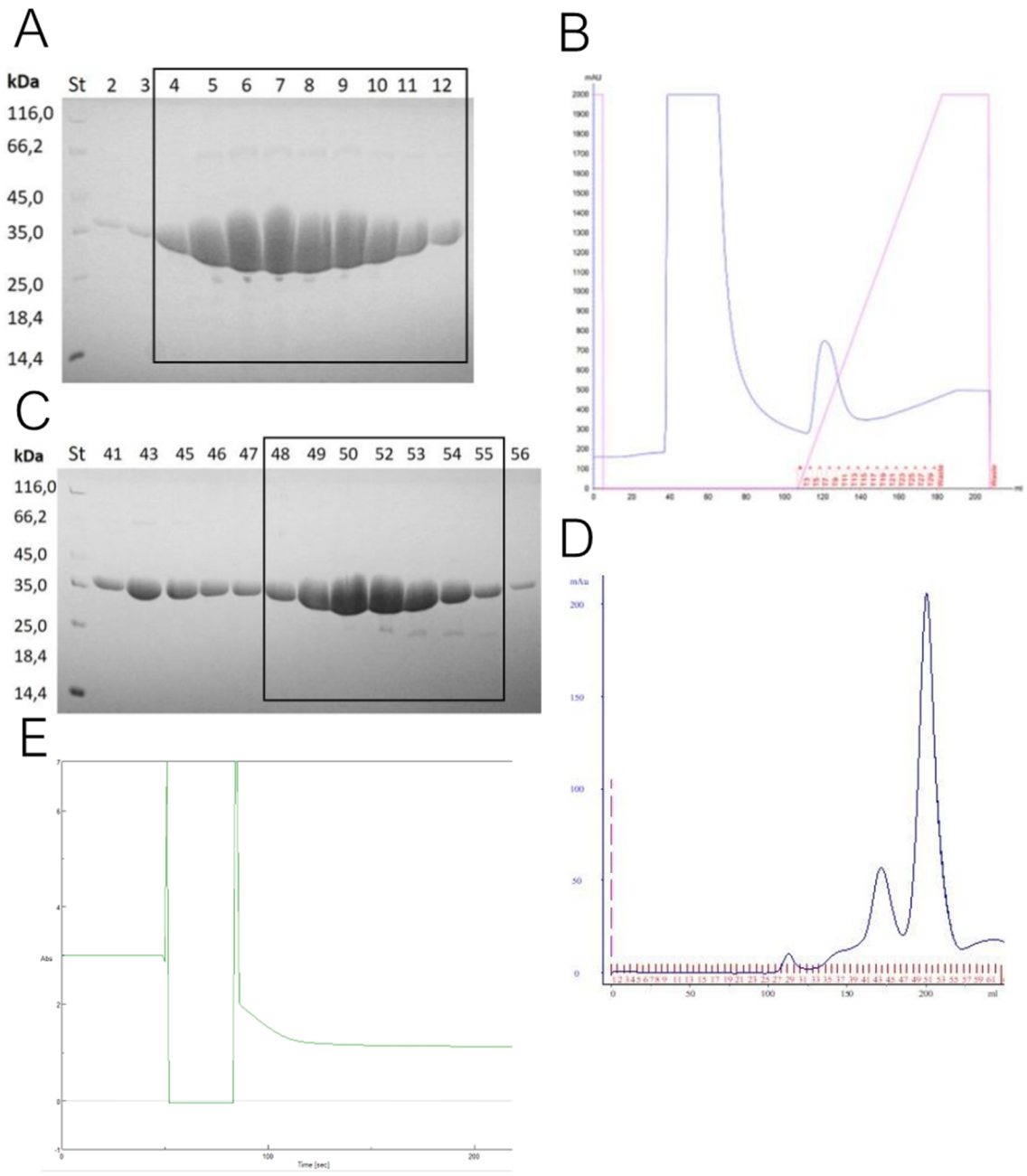


Figure S 15: Purification of Mdh. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Pooled fractions indicated by black rectangle. Mdh has a molecular weight of 32 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (pink). **C:** S200 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S200 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Photometric assay indicating enzymatic activity of KdsA. Change in $A_{340\ nm}$ indicates NADH oxidation.
Yield: 20.3 mg/l expression culture. Purification conducted within the Master's Thesis of Franziska Funke (2019).

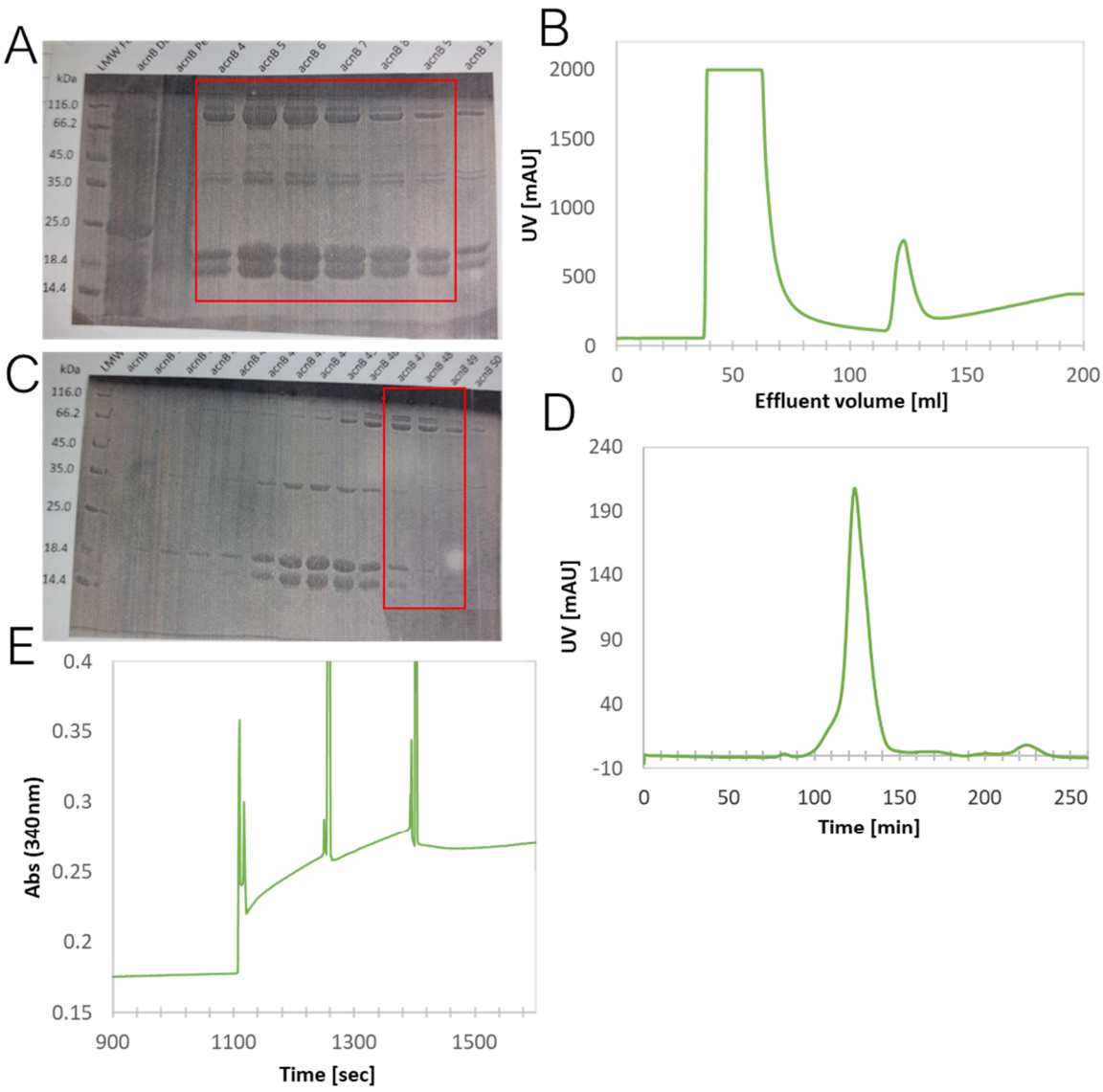


Figure S 16: Purification of AcnB. A: IMAC elution fractions on SDS-PAGE. From left to right: protein mass standard, flow-through, pellet, and elution fractions. Pooled fractions indicated by red rectangle. AcnB has a molecular weight of 94 kDa **B:** IMAC profile. $A_{280\,nm}$ (green). **C:** S200 gel filtration elution fractions on SDS-PAGE. LMW – protein mass standard, and elution fractions. Fractions marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. AcnB used in experiments was filtered with a molecular weight cut-off of 45 kDa prior to use, as light peptides eluted in fractions close to the target protein for unknown reasons and might have comprised a trace contaminant. **D:** S200 gel filtration profile. $A_{280\,nm}$ (green). **E:** Coupled photometric assay indicating enzymatic activity of AcnB. Change in $A_{340\,nm}$ indicates $NADP^+$ reduction.
Yield: 8.3 mg/l expression culture.

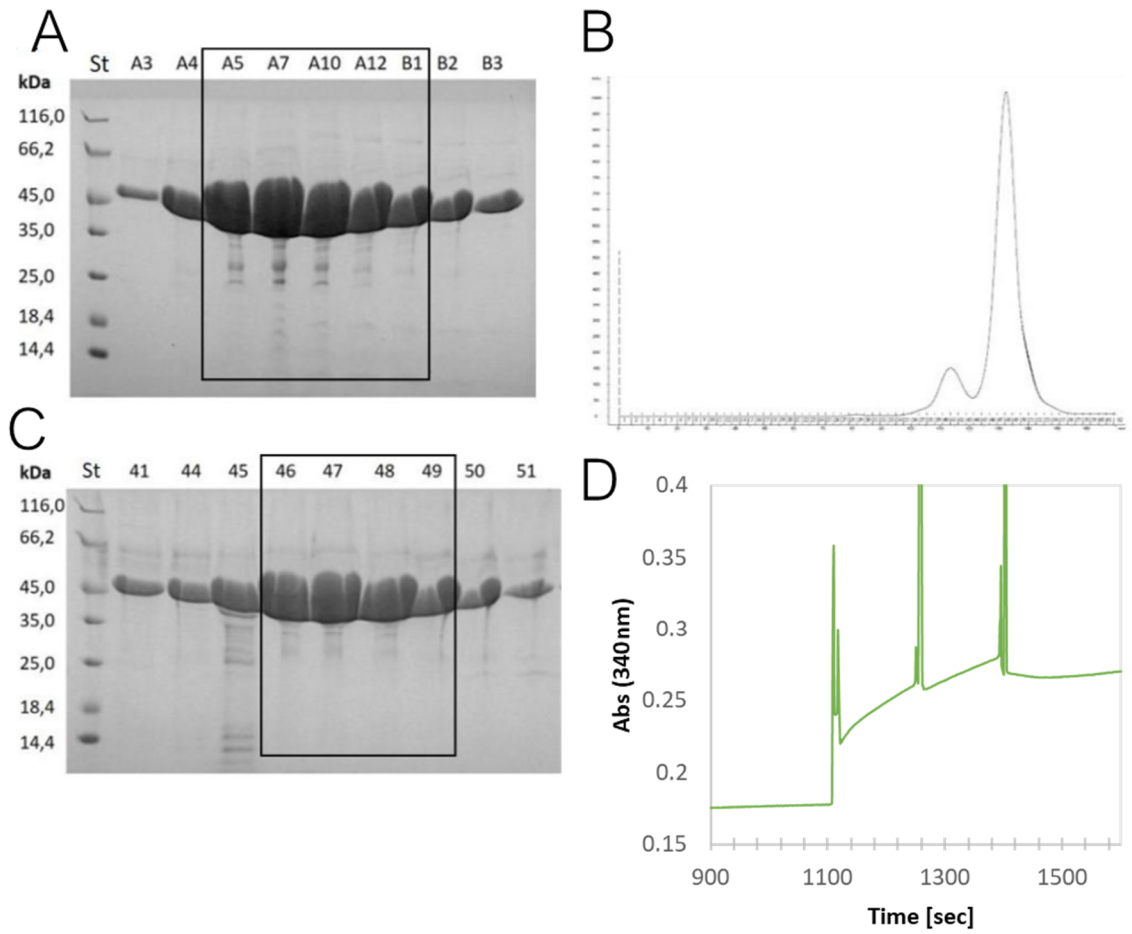


Figure S 17: Purification of Icd. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Pooled fractions indicated by black rectangle. Icd has a molecular weight of 46 kDa. **B:** S200 gel filtration profile. $A_{280\ nm}$ (black). **C:** S200 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fractions marked by black rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** Photometric assay indicating enzymatic activity of Icd. Change in $A_{340\ nm}$ indicates NADH oxidation (identical to assay in **Figure S 16**, as A and Icd share a coupled assay).
Yield: 13.1 mg/l expression culture. Purification conducted within the Master's Thesis of Franziska Funke (2019).

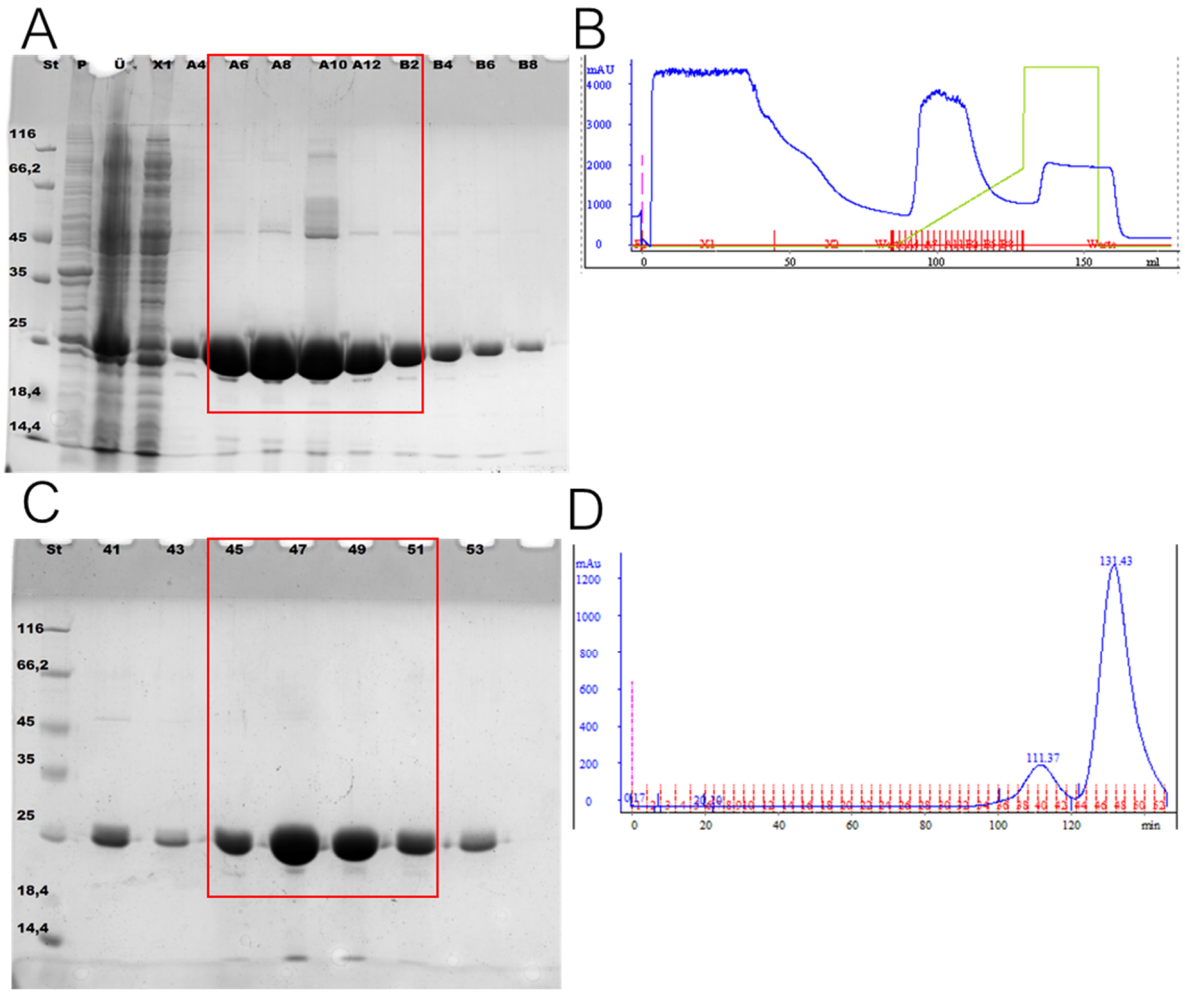


Figure S 18: Purification of SodA. A: IMAC elution fractions on SDS-PAGE. ST – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. SodA has a molecular weight of 24.5 kDa. **B:** IMAC profile. $A_{280\,nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\,nm}$ (blue).
Yield: 27.1 mg/l expression culture.

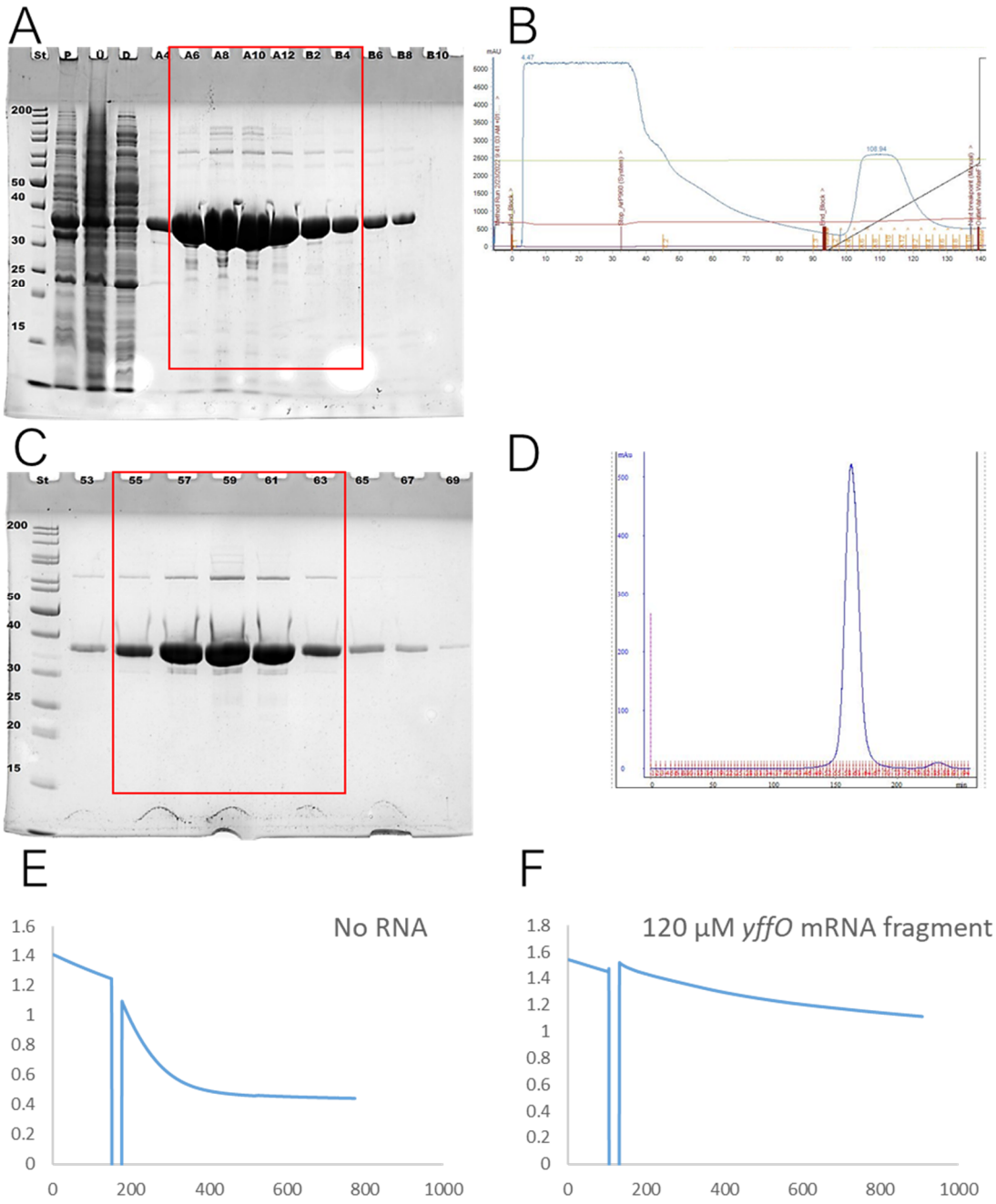


Figure S 19: Purification of QorA. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, D – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. QorA has a molecular weight of 32 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (black) **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Photometric assay which might indicate enzymatic activity of QorA towards 1,4-benzoquinone (assay reproducibility poor). Change in $A_{340\ nm}$ indicates NADH oxidation. **F:** same as E, but in presence of 120 µM *yffO* mRNA fragment.
Yield: 14.6 mg/l expression culture.

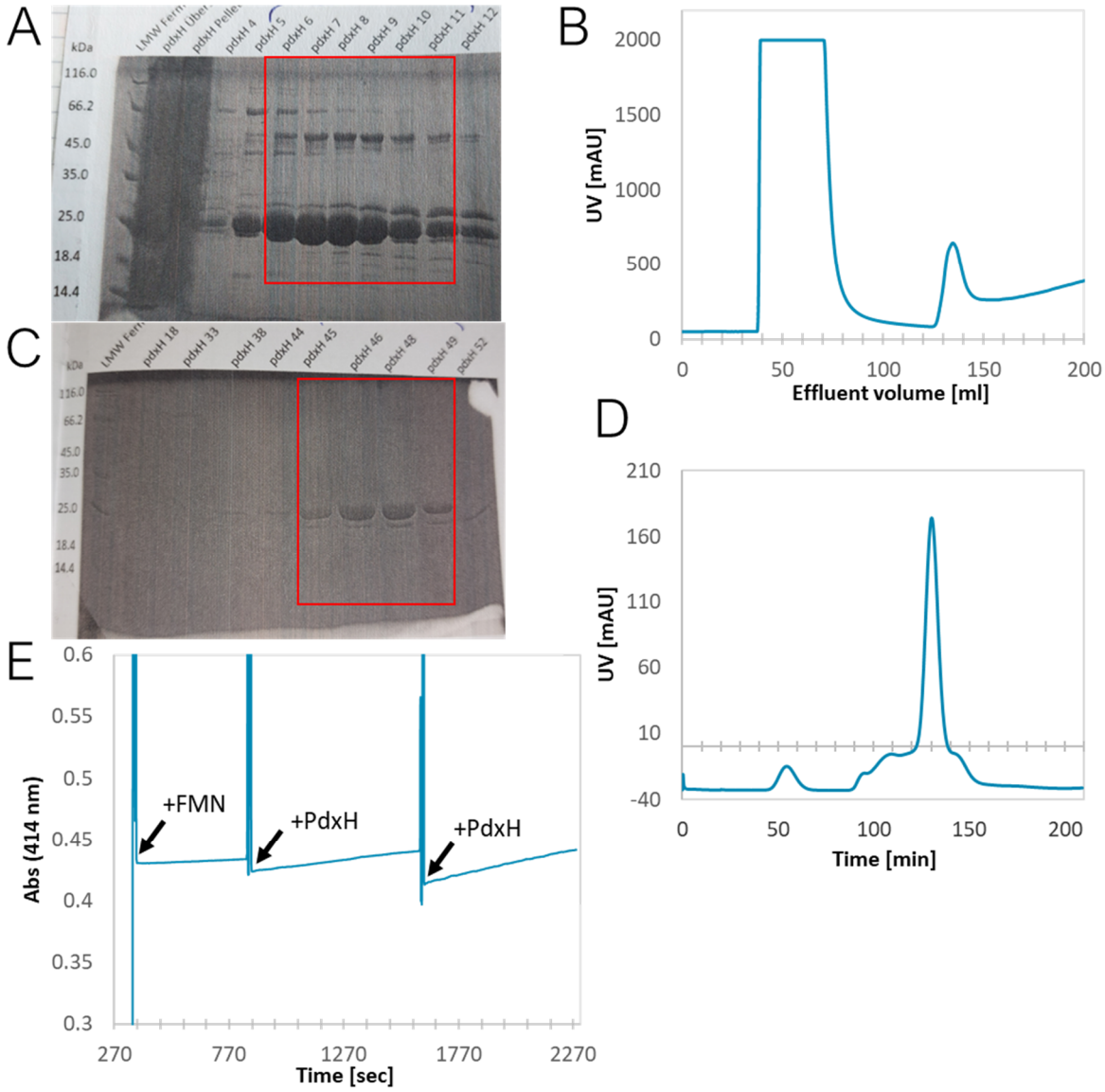


Figure S 20: Purification of PdxH. A: IMAC elution fractions on SDS-PAGE. From left to right: protein mass standard, crude extract, pellet, and elution fractions. Pooled fractions indicated by red rectangle. PdxH has a molecular weight of 26 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue). **C:** S75 gel filtration elution fractions on SDS-PAGE. LMW – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue). **E:** Coupled photometric assay indicating enzymatic activity of PdxH. Change in $A_{414\ nm}$ indicates pyridoxine-5-phosphate oxidation.
Yield: 3.2 mg/l expression culture.

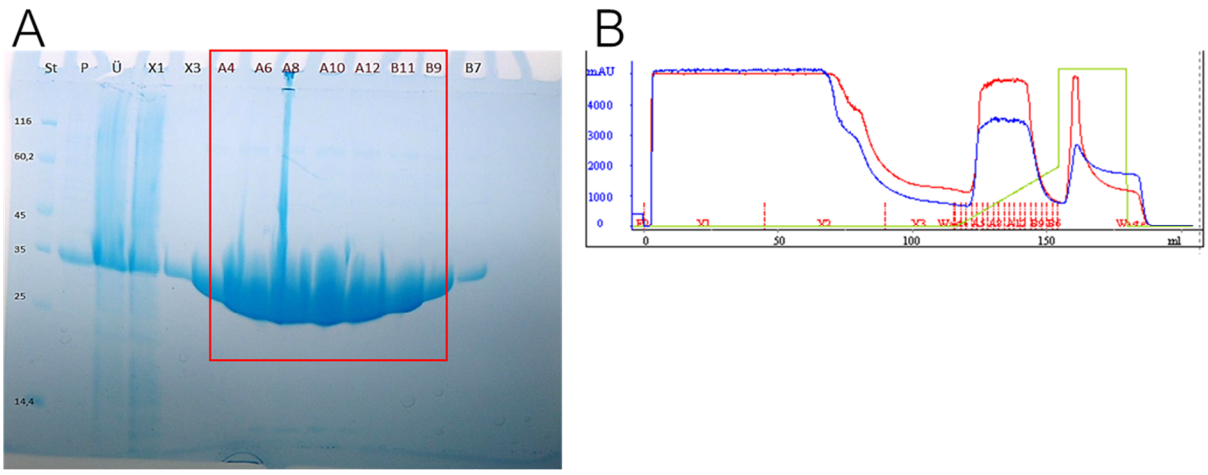


Figure S 21: Purification of YbiB. A: IMAC elution fractions on SDS-PAGE. St – protein mass standard, P – pellet, Ü – crude extract, X1,X3 – flowthrough fractions, and elution fractions. Pooled fractions indicated by red rectangle. YbiB has a molecular weight of 35 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), $A_{260\ nm}$ (red), gradient of 750 mM imidazole (green).
Yield: 24.3 mg/l expression culture.

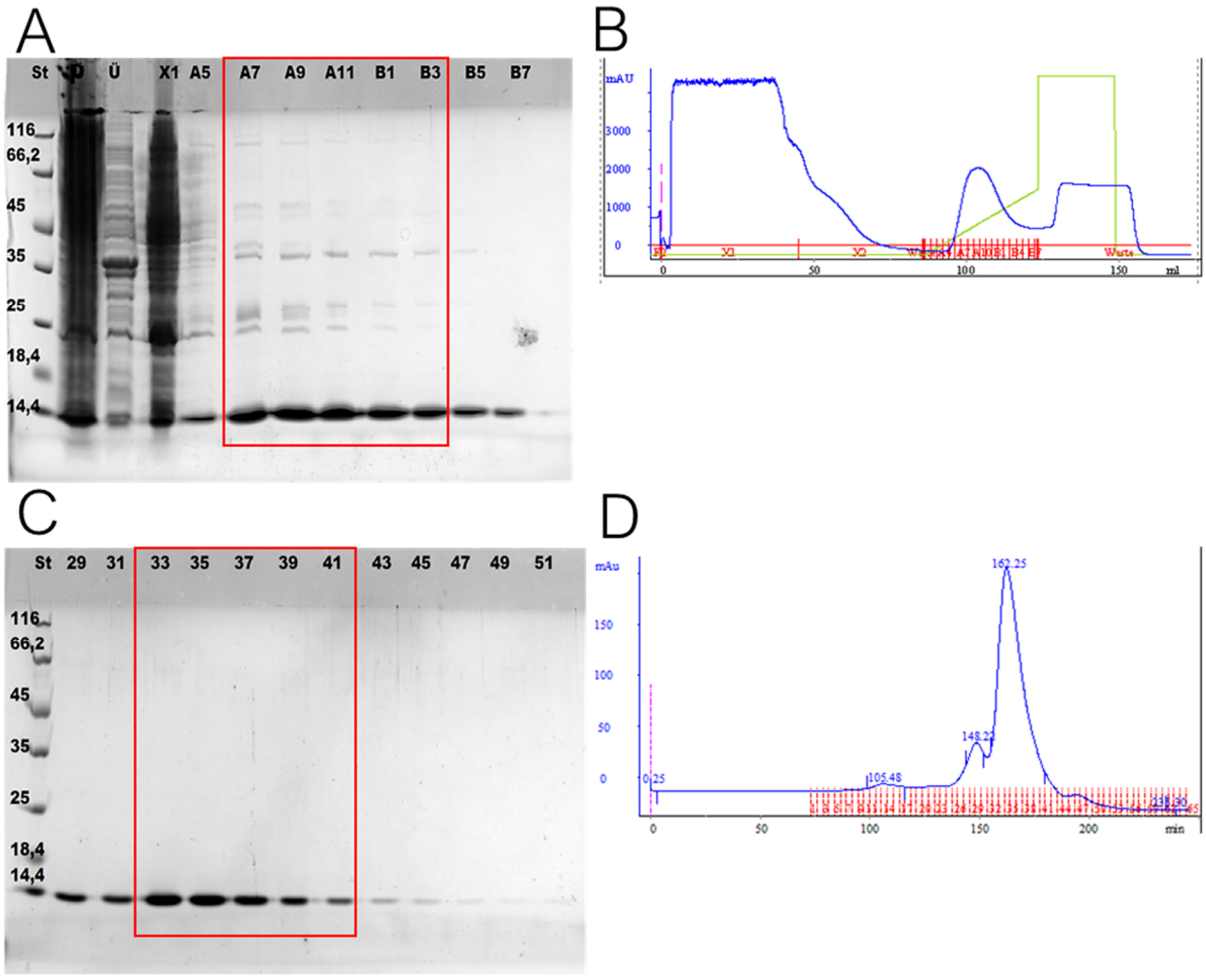


Figure S 22: Purification of YggX. A: IMAC elution fractions on SDS-PAGE. ST – protein mass standard, P – pellet, Ü – crude extract, X1 – flowthrough fraction, and elution fractions. Pooled fractions indicated by red rectangle. YggX has a molecular weight of 12 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), gradient of 750 mM imidazole (green). **C:** S75 gel filtration elution fractions on SDS-PAGE. St – protein mass standard, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\ nm}$ (blue).
Yield: 18.8 mg/l expression culture.

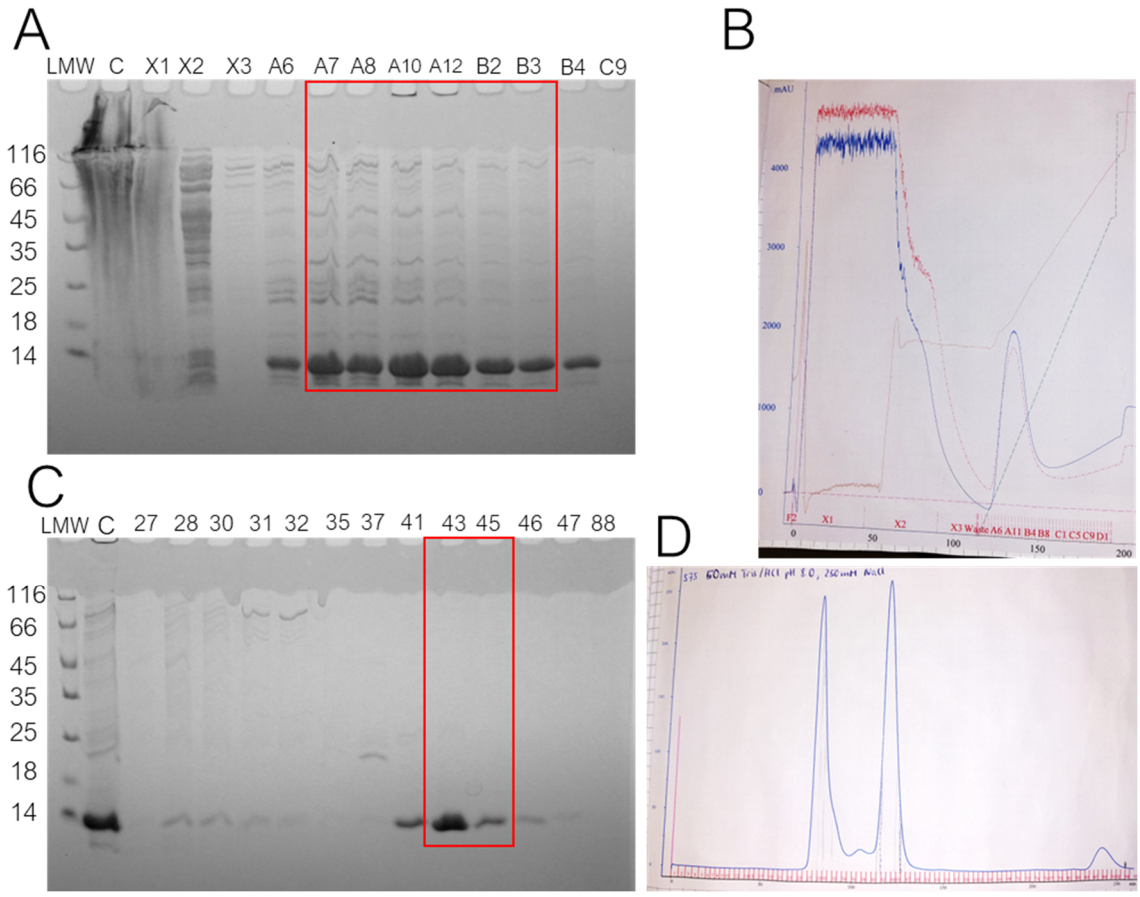


Figure S 23: Purification of MS2 coat protein. A: IMAC elution fractions on SDS-PAGE. LMW – protein mass standard, C – crude extract, X1-X3 – flowthrough fractions, and elution fractions. Pooled fractions indicated by red rectangle. MS2 has a molecular weight of 14 kDa. **B:** IMAC profile. $A_{280\,nm}$ (blue), $A_{260\,nm}$ (red), gradient of 750 mM imidazole (dashed green). **C:** S75 gel filtration elution fractions on SDS-PAGE. LMW – protein mass standard, C – crude extract, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S75 gel filtration profile. $A_{280\,nm}$ (blue). **Yield: 3.3 mg/l expression culture.**

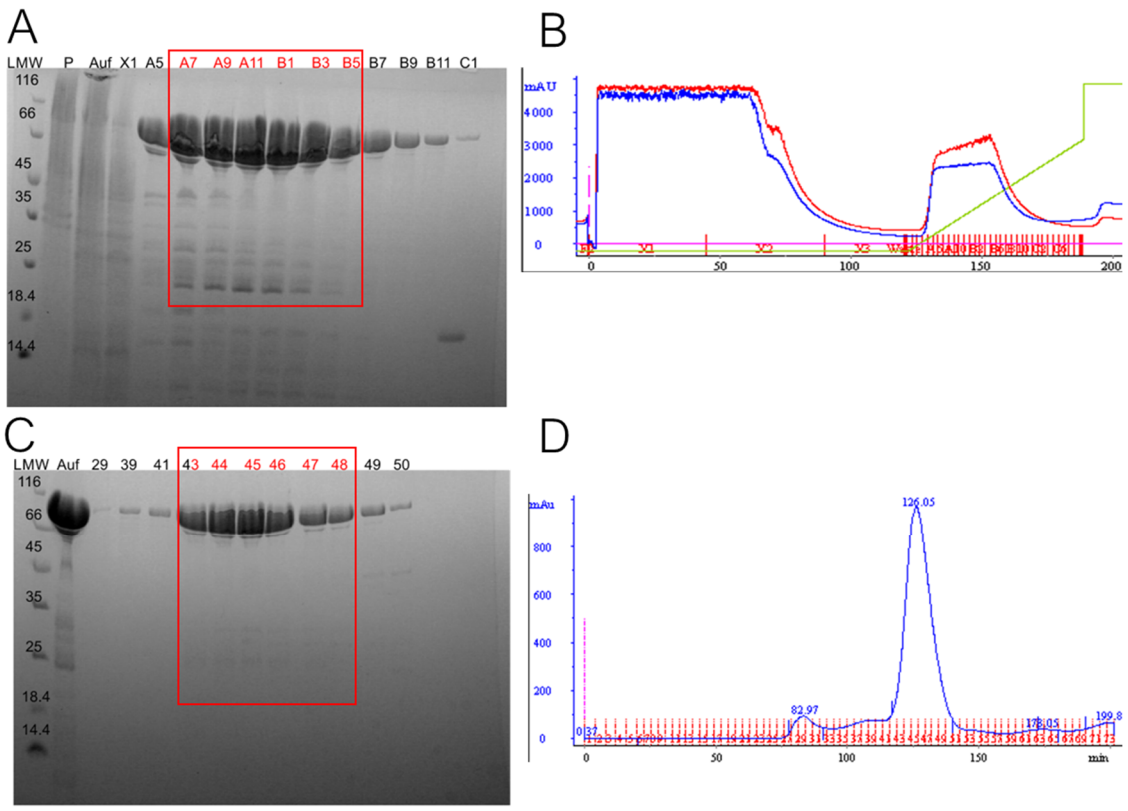


Figure S 24: Purification of T7-Polymerase. A: IMAC elution fractions on SDS-PAGE. LMW – protein mass standard, P – pellet, Auf – lysate, X1 – flowthrough, and elution fractions. Pooled fractions indicated by red rectangle. T7 Polymerase has a molecular weight of 99 kDa. **B:** IMAC profile. $A_{280\ nm}$ (blue), $A_{260\ nm}$ (red), gradient of 750 mM imidazole (green). **C:** S200 gel filtration elution fractions on SDS-PAGE. LMW – protein mass standard, Auf – lysate, and elution fractions. Fraction marked by red rectangle were pooled, concentrated, and frozen in liquid nitrogen for long time storage. **D:** S200 gel filtration profile. $A_{280\ nm}$ (blue). **Yield: 15.9 mg/l expression culture.**

7.2 Further Supplements

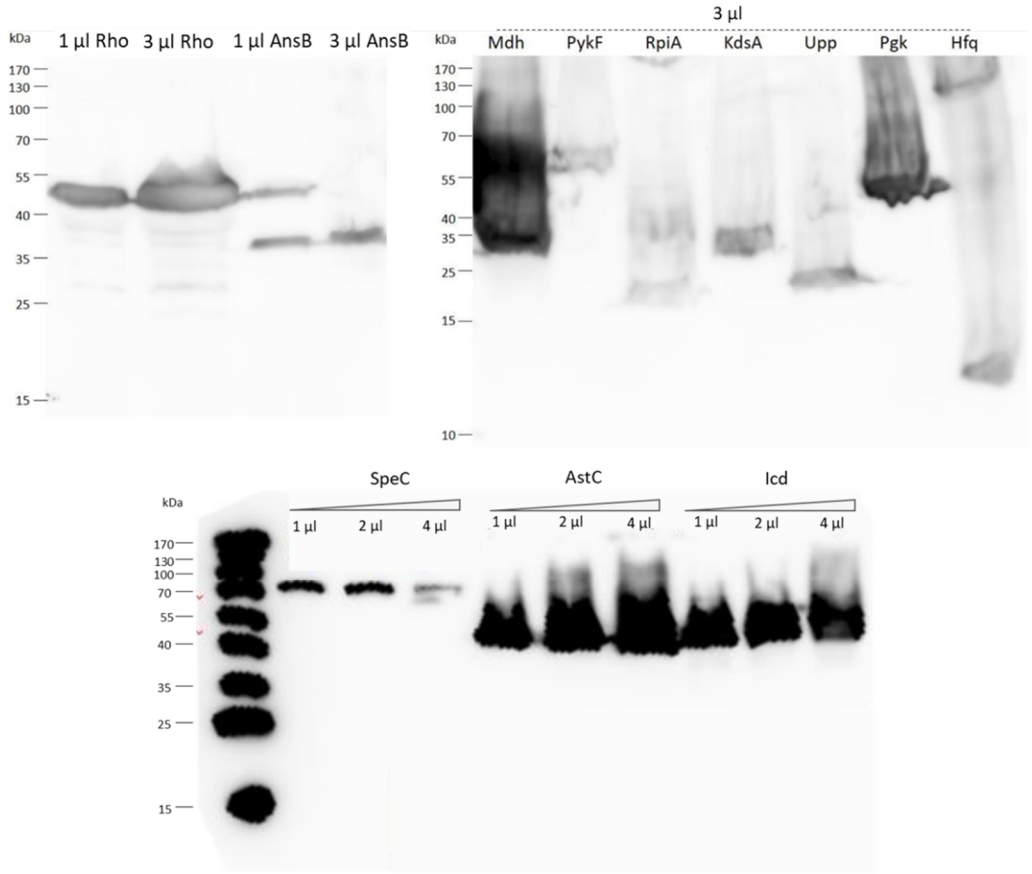


Figure S 25: Western-blots of SDS-PAGEs loaded with cell lysates of strains carrying genomic FLAG-modifications at respective genes. Indicated amounts of cell lysate were loaded onto each lane. Visualization by horseradish peroxidase coupled to a secondary anti-rat antibody detecting the primary rat anti-FLAG antibody. Expected molecular weights: 47 kDa (Rho), 35 kDa (AnsB), 32 kDa (Mdh), 51 kDa (PykF), 23 kDa (RpiA), 31 kDa (KdsA), 23 kDa (Upp), 41 kDa (Pgk), 11 kDa (Hfq), 80 kDa (SpeC), 44 kDa (AstC), 46 kDa (Icd) . Upper gel images were originally published in the Master's thesis of Franziska Funke (2019).

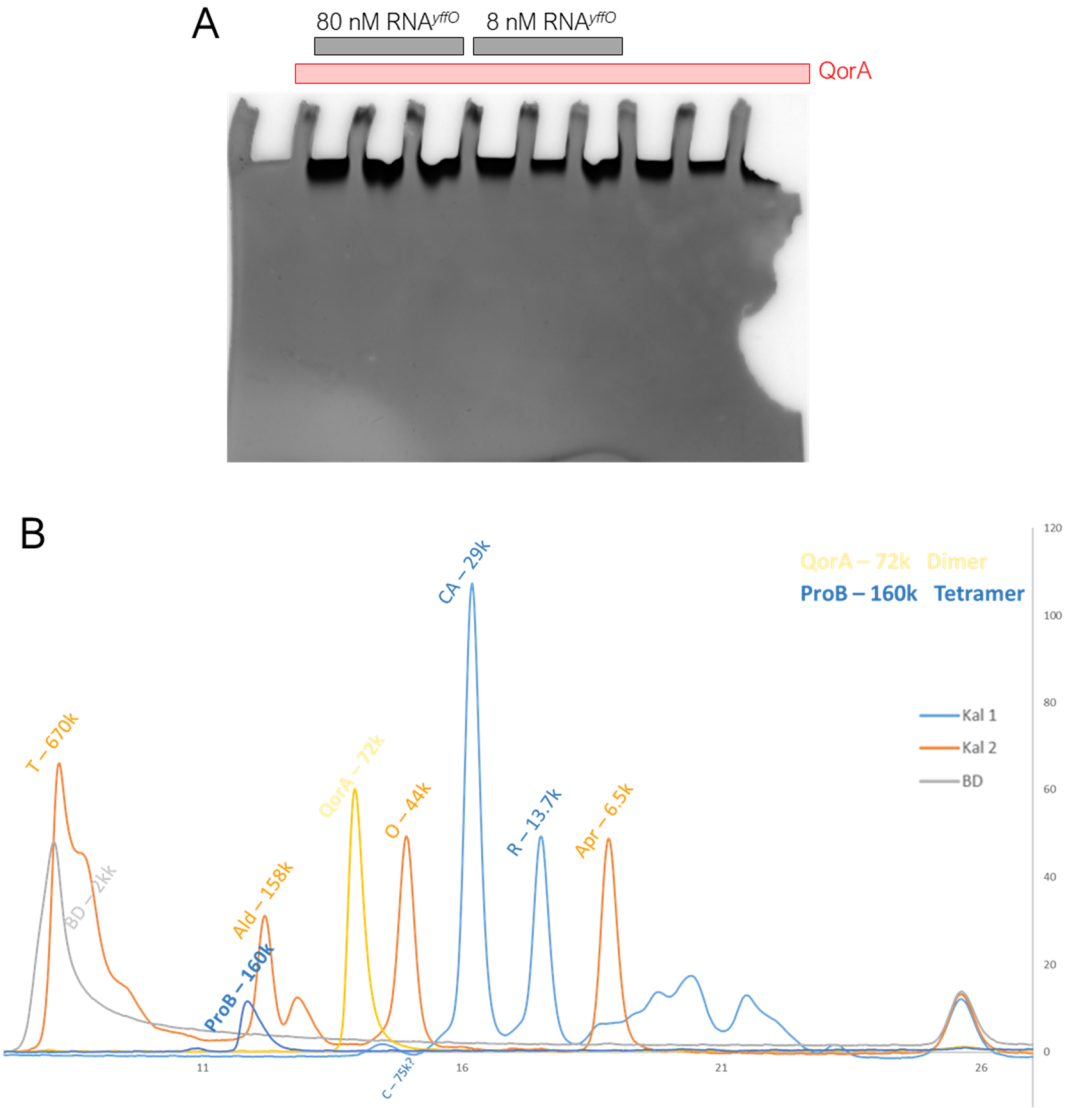


Figure S 26: QorA oligomerization behavior. A) Coomassie-stained native PAGE loaded with free QorA (12 µM) or yffO-QorA complex as indicated. Band visible right below pocket in all lanes. Each condition is loaded three times, probing potential impact of preincubation pH values of 7.5, 8.4, and 9.3 on complex. Gel running buffer in use was TBE pH 8.15. **B)** Analytical gel filtration verifying dimeric oligomerization state of QorA (36 kDa monomer) and tetrameric oligomerization state of ProB (40 kDa monomer). Experiment calibrated with. Calibration peaks: Apr – Aprotinin (6500 Da), R – Ribonuclease (13700 Da), CA – Carbonic anhydrase (29000 Da), O – Ovalbumin (44000 Da), Aldolase (158000 Da), T – Thyroglobulin (670000 Da).

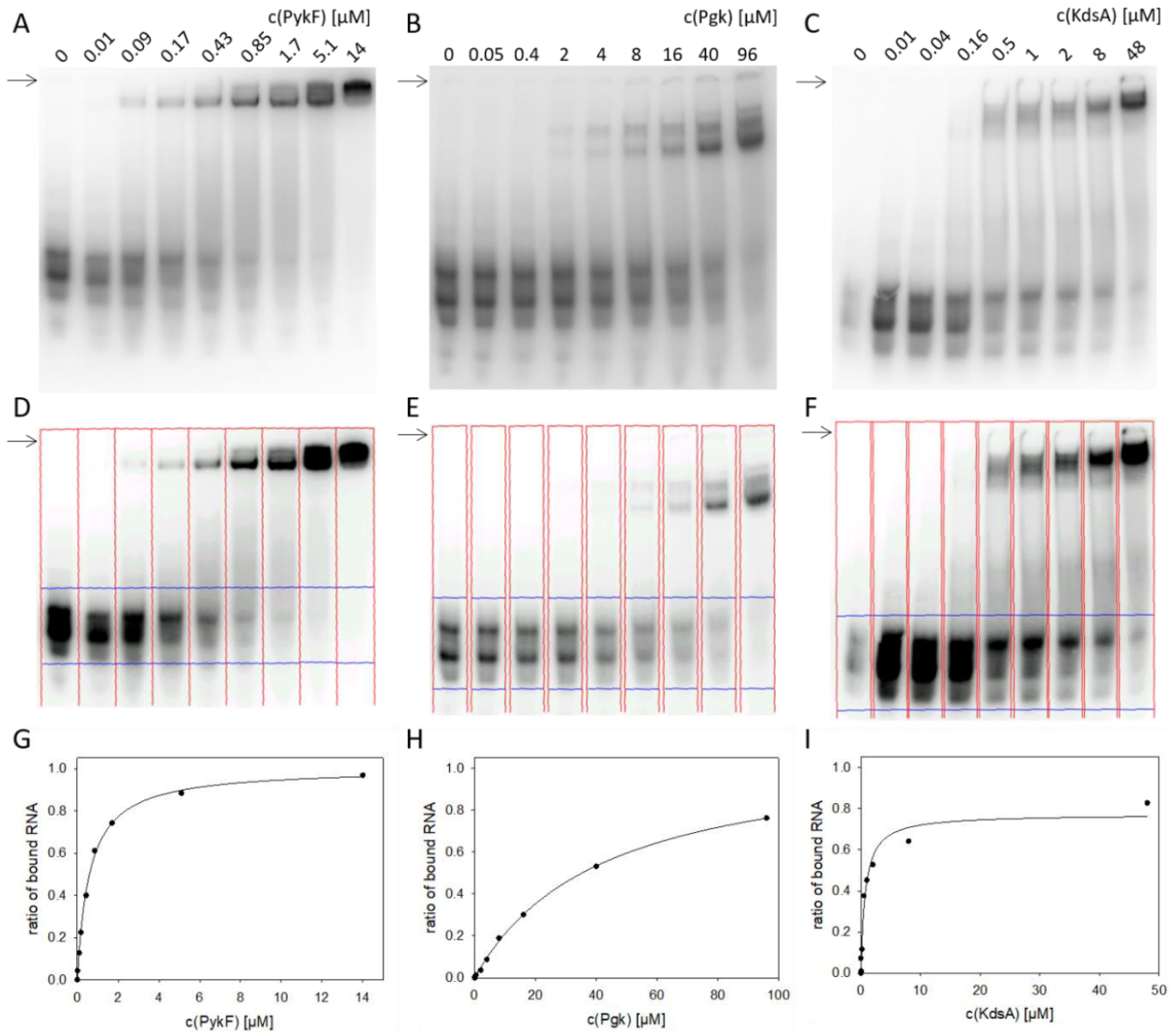


Figure S 27: EMSAs attesting unspecific RNA binding activity to PykF, Pgk, and KdsA. A-C: Gel images displaying EMSA titrations against a 40 nt control RNA fragment (GGGUU CUAGA GAGGU GAGCU UGGCA ACCUC UGAUG UAGGU). **D-F:** Segmentation for calculation of titration curves from band intensities. **G-I:** Fitting of titration data points, delivering K_D-estimates of K_D^{PykF} = 0.6 µM, K_D^{Pgk} = 42 µM, and K_D^{KdsA} = 0.7 µM. Data from Master's Thesis Franziska Funke (2019).

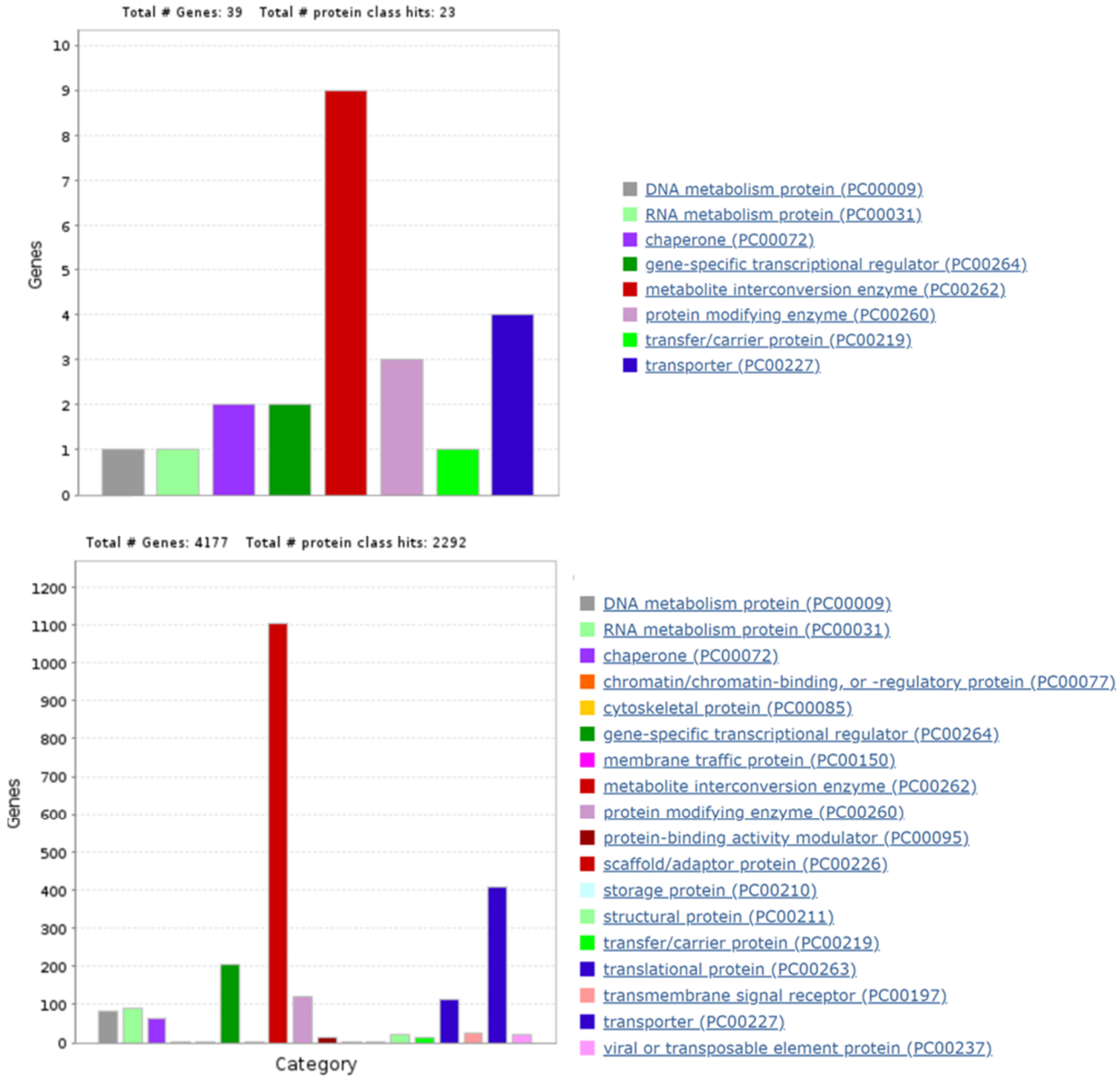


Figure S 28: GO-term analysis of protein classes for the PykF SELEX experiment. Top – GO-term distribution of read clusters found in the PykF-exposed genomic libraries. Bottom – GO-term distribution of the entire *E. coli* genome for reference. Plots created using PANTHER v17.0.

Table S 1: Additional gene loci with read clusters in the MS2 SELEX experiment.

gene locus	sense/antisense	mapped reads	deduplicated reads
gfcB	s	0.930%	0.065%
prlC	as	0.082%	0.366%
fsr	s	0.061%	0.108%
UTR near yneL	-	0.046%	0.022%
yhbW	s	0.041%	0.323%
dbpA	as	0.038%	0.172%
fadI	as	0.037%	0.172%
argG	as	0.030%	0.259%
ynfF	s	0.028%	0.215%
eamA	as	0.027%	0.129%
yigZ	as	0.021%	0.215%
hyfB	as	0.019%	0.172%
hcaT	as	0.016%	0.194%
hslu	as	0.016%	0.215%
yaaX	s	0.013%	0.215%
yhhJ	as	0.013%	0.215%
nepI	as	0.013%	0.215%
cusA	as	0.012%	0.194%
pykF	as	0.011%	0.172%

8 Acknowledgments

First, I want to sincerely thank PD Dr. Patrick Babinger for mentoring me during my career as PhD student. He was always friendly, very trustful, and aided me in manifold ways, providing advice and covering my back so I could focus on my research as much as possible.

I also want to thank Prof. Dr. Reinhard Sterner for the same steadfast kindness and trust, as well as for good advice in many situations. He suggested and initiated my research visit in TU Darmstadt, which turned out to be vital for this work.

Both PD Dr. Patrick Babinger and Prof. Dr. Reinhard Sterner were incredible supportive and provided staggering encouragement during periods of fruitless research and especially towards the final phase of the thesis. This is something that I will keep as fond memory.

A special thanks to Prof. Dr. Beatrix Süß, who welcomed me in her department in TU Darmstadt and generously shared knowledge on SELEX methods. Thank you also Dr. Michael Vockenhuber, who spent a lot of time providing guidance during this research visit. Everyone I met in Darmstadt was so friendly and it was a great time!

Thank you to Dr. Jan Medenbach, who took time whenever I approached him for fruitful discussions about my topic and owns an infectious enthusiasm for science.

My gratitude goes to Prof. Dr. Gunter Meister, who generously provided access to the NGS device in his lab. On this note, I also want to cordially thank Norbert Eichner for guiding me on library sequencing and Gerhard Lehmann for processing and genome mapping of sequencing data.

A special thanks to Dr. Oliver Rossbach, who shared his CLIP-Seq expertise and spared neither trouble nor expenses, personally conducting time-consuming immunoprecipitations and library preparation.

Thank you to Franziska Funke. The work you conducted was immensely helpful for this project and is deeply appreciated.

Thank you to all students I had the pleasure supervising and working with in the lab – I have only good memories!

I want to thank our technicians, Sabine Laberer, Chrisitane Endres, Jeanette Ueckert, and Sonja Fuchs, for lab assistance that was indispensable for this work. On top of providing help, you all have been incredible friendly throughout the years and are the best colleagues one could wish for.

The same applies to virtually all current and former PhD students, post-docs, and master students I happened to run across in the office or lab. The atmosphere in the department has always been incredible and might be one of a kind, making the PhD student life so much better. Thank you to everyone!

Lastly, I want to sincerely thank my family for their everlasting, heartfelt support.