Molecular strategies applied by bacteriophage T4 for efficient hijacking of *Escherichia coli*

Dissertation

“kumulativ“

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Chemie

der Philipps-Universität Marburg

vorgelegt von

Nadiia Pozhydaieva

Aus Dnipro, Ukraine

Marburg (Lahn), Deutschland, 2024

Die vorliegende Dissertation mit dem Titel „Molecular strategies applied by bacteriophage T4 for efficient hijacking of *Escherichia coli*“ wurde von 10.2020 bis 05.2024 am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg in der Arbeitsgruppe „Bacterial Epitranscriptomics“ unter der Leitung von Dr. Katharina Höfer angefertigt.

Vom Fachbereich Chemie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als

Dissertation angenommen am 26.03.2024

Erstgutachterin: Dr. Katharina Höfer
Zweitgutachter: Prof. Dr. Lennart Randau

Weitere Mitglieder der Prüfungskommission:
Prof. Dr. Peter Grauman
Prof. Dr. Julia Frunzke

Tag der Disputation: 24.05.2024

Erklärung

Ich erkläre, dass eine Promotion noch an keiner anderen Hochschule als der Philipps-Universität Marburg, Fachbereich Chemie, versucht wurde.

Hiermit versichere ich, dass ich die vorliegende Dissertation

“Molecular strategies applied by bacteriophage T4 for efficient hijacking of *Escherichia coli* “

selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

Ort/Datum	Nadiia Pozhydaieva

Erklärung zur kumulativen Dissertation

Ich versichere hiermit, dass die im kumulativen Teil der Dissertation mit dem Titel „Molecular strategies applied by bacteriophage T4 for efficient hijacking of *Escherichia coli* " aufgeführten Anteile der Autoren an den verfassten Publikationen und Manuskripten korrekt und vollständig dargelegt sind. Der kumulative Teil umfasst die Kapitel 2-6.

Kapitel 2 – Integrated omics reveal time-resolved insights into T4 phage infection of *E. coli* on proteome and transcriptome levels

Wolfram-Schauerte, M.*, **Pozhydaieva, N.***, Viering, M., Glatter, T., & Höfer, K. (2022). *Viruses*, 14(11), 2502. doi.org/10.3390/v14112502

In dieser Arbeit geleisteter Eigenbeitrag: Konzeption des Projekts, Planung und Durchführung von Experimenten, Analyse der Ergebnisse, Erstellung des Manuskriptes

Kapitel 3 – Shaping the bacterial epitranscriptome – 5′-terminal and internal RNA modifications

Schauerte, M.*, **Pozhydaieva, N.***, & Höfer, K. (2021). *Advanced Biology*, 5(8), 2100834

In dieser Arbeit geleisteter Eigenbeitrag: Erstellung und Beschreibung der Konzepte, Erstellung des Manuskriptes

Kapitel 4 – The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections

Pozhydaieva, N.*, Wolfram-Schauerte, M.*, Keuthen, H., & Höfer, K. (2024). *Current Opinion in Microbiology*, 77, 102417. https://doi.org/10.1016/j.mib.2023.102417

In dieser Arbeit geleisteter Eigenbeitrag: Erstellung und Beschreibung der Konzepte, Erstellung des Manuskriptes

Kapitel 5 – Temporal epigenome modulation enables efficient bacteriophage engineering and functional analysis of phage DNA modifications

Pozhydaieva, N., Billau, F. A., Wolfram-Schauerte, M., Ramírez Rojas, A. A., Paczia, N., Schindler, D., Höfer, K. (2024). bioRxiv. https://doi.org/10.1101/2024.01.28.577628

In dieser Arbeit geleisteter Eigenbeitrag: Konzeption des Projekts, Planung und Durchführung von Experimenten, Analyse der Ergebnisse, Erstellung des Manuskriptes

Kapitel 6 – A viral ADP-ribosyltransferase attaches RNA chains to host proteins

Wolfram-Schauerte, M., **Pozhydaieva, N.**, Grawenhoff, J., Welp, L. M., Silbern, I., Wulf, A., Billau, F. A., Glatter, T., Urlaub, H., Jäschke, H., Höfer, K. (2023). *Nature*, 620(7976), 1054-1062. https://doi.org/10.1038/s41586-023-06429-2

In dieser Arbeit geleisteter Eigenbeitrag: Konzeption des Projekts (Phagen Mutagenese und Charakterisierung der Mutante), Planung und Durchführung von Experimenten, Analyse der Ergebnisse

Nadiia Pozhydaieva	Dr. Katharina Höfer

„In the middle of difficulty lies opportunity.“

Albert Einstein

Table of Contents

Summary

Bacteriophages are bacterial predators that serve as excellent models to study host-pathogen interactions and hold significant potential for industrial and medical applications. These include the utilization of bacteriophages as alternatives to antibiotics in combating multi-resistant bacterial strains, controlling microbial communities, and engineering phages for specific diagnostic purposes. To fully exploit the potential of phages, a comprehensive understanding of phage infection mechanisms and the bacterial countermeasures is crucial. The T4 phage stands out as one of the model bacteriophages and its infection of *Escherichia coli* is one of the best-studied bacterium-bacteriophage interactions. Research on the molecular mechanisms of T4 phage infection has strongly shaped our understanding of the fundamental principles in molecular biology. Fundamental concepts like DNA being a blueprint of life, principles in molecular genetics, phage evolutionary mechanisms, and beyond were discovered upon research on T4 phage. Additionally, many T4 phage proteins became indispensable tools in molecular biology.

However, despite the considerable knowledge gained from studying the T4 phage, numerous aspects of its infection remain unresolved. While approximately half of the T4 proteins have been associated with specific functions, the roles of the remaining 45% of T4 phage proteins are still unexplored. Therefore, this knowledge gap alone makes it evident that our molecular understanding of the T4 phage infection and the strategies the phage employs to execute an efficient infection is far from complete. This thesis aimed to enhance our knowledge of the T4 phage infection at the molecular level and to uncover previously unexplored mechanisms used by the T4 phage to carry out infection efficiently.

Chapter II describes a multi-omics study designed to provide a temporal resolution of an *E. coli* infection with T4 phage at the molecular level. The transcriptome and proteome of *E. coli* and T4 phage were analyzed throughout the infection. This enabled the identification of temporal gene expression patterns for T4 phage transcripts. Even more, a decoupling of transcription and translation processes was observed for certain T4 phage genes. The transcriptome and proteome analysis of *E. coli* revealed a general degradation of host transcripts and preservation of the host proteins. This study presents the molecular kinetics of T4 phage infection for the first time. The results strongly suggest the existence of additional, unexplored regulatory mechanisms that allow differential degradation of host and phage transcripts and decoupling of transcription and translation for specific phage genes.

A possible explanation for the differential RNA degradation upon infection and decoupling of transcription and translation observed for some T4 phage genes, could be the presence of RNA modifications. RNA modifications may provide the molecular basis for the discrimination between

bacterial and viral transcripts during T4 phage infection. **Chapter III** summarizes the current knowledge on the bacterial epitranscriptome, emphasizing mRNA modifications. The known writers, readers, and erasers that regulate RNA modifications, and techniques to identify and study specific RNA modifications are discussed. This work demonstrates a significant knowledge gap regarding RNA modifications on bacterial mRNA, their modulators, and their biological significance.

While some initial insights into the bacterial epitranscriptome already exist, the epitranscriptome of bacteriophages remains unexplored to date. In **Chapter IV**, the current knowledge of RNA modifications in bacteria is used to hypothesize how some bacterial and phage enzymes may shape and modulate the epitranscriptome of bacteriophages during infection.

To investigate the biological role of potential T4 phage-derived infection regulators, it is necessary to study how their absence or inactivity affects phage infection. However, this requires efficient tools for phage mutagenesis. CRISPR-Cas is a powerful tool for precise genome engineering, but its effectiveness for T4 phage mutagenesis is severely hampered by the highly abundant modifications of T4 DNA. **Chapter V** outlines an approach for temporal reduction of T4 phage DNA modifications. This enables efficient and scarless CRISPR-Cas-based mutagenesis of T4 phage DNA. This system not only facilitates T4 phage mutagenesis but also allows the study of the role of DNA modifications in phage infection and has the potential to be extended to other phages beyond T4 phage.

Chapter VI focuses on the T4 ADP-ribosyltransferase ModB and reveals that ModB not only accepts NAD as a substrate to perform ADP-ribosylation but also NAD-RNA to perform RNAylation – a novel post-translational modification. This modification was shown to be introduced by ModB *in vitro* and *in vivo*. Furthermore, it was demonstrated that ModB RNAylates several *E. coli* proteins, including ribosomal proteins S1 and L2. The biological role and molecular mechanisms of RNAylation were investigated. Additionally, a T4 phage mutant with a catalytically inactive ModB was generated to explore the impact of ModB activity on phage infection and phage phenotype.

This thesis contributes to an enhanced understanding of the T4 phage infection of *E. coli* by providing insights into the molecular organization of infection, showing the impact of phage DNA modifications on phage phenotype and mutagenesis efficiency, and through the discovery of RNAylation – a novel post-translational protein modification. Apart from deepening our understanding of the T4 phage infection and its regulation, the knowledge gained in this thesis also lays the groundwork for its translation into application. Particularly, understanding the molecular organization of the phage infection and its gene expression patterns is essential for

designing a synthetic phage or tailoring an existing phage for specific needs. To tailor the T4 phage, the mutagenesis strategy reported here can be efficiently applied. Furthermore, the discovery of RNAylation expands the arsenal of T4 phage-derived molecular tools, as RNAylation can find potential applications in synthetic biology for the development of novel artificial cellular RNA-protein constructs and opens up new possibilities for the design of next-generation RNA-based therapeutics.

Taken together, this study expands our understanding of the molecular mechanisms underlying efficient T4 phage infection and underscores that the discovery potential based on T4 phage research is far from being fully exploited.

Zusammenfassung

Bakteriophagen sind Viren von Bakterien, die nicht nur als hervorragende Modelle für die Untersuchung von Wirt-Pathogen-Interaktionen dienen, sondern auch ein erhebliches Potenzial für industrielle und medizinische Anwendungen haben. Dazu gehören der Einsatz von Bakteriophagen als Alternative zu Antibiotika bei der Bekämpfung multiresistenter Bakterienstämme, deren Nutzung für die Kontrolle mikrobieller Gemeinschaften und die Entwicklung von Phagen für spezifische Diagnosezwecke. Um das vollständige Potenzial der Phagen für medizinische und biotechnologische Anwendungen zu entfalten, ist ein umfassendes Verständnis der zugrundeliegenden Infektionsmechanismen von entscheidender Bedeutung. Der T4 Phage ist einer der Modell-Bakteriophagen und seine Infektion von *Escherichia coli* ist eine der am besten untersuchten Bakterium-Bakteriophagen-Interaktionen. Die Erforschung der molekularen Mechanismen der T4 Infektion hat maßgeblich zu unserem heutigen Verständnis der grundlegenden Prinzipien der Molekularbiologie beigetragen. Grundlegende Konzepte wie die DNA als Bauplan des Lebens, Prinzipien der Molekulargenetik, Evolutionsmechanismen von Phagen und vieles mehr wurden durch die Forschung an T4 Phagen entdeckt. Darüber hinaus sind viele Proteine der T4 Phagen zu unverzichtbaren Schlüsselwerkzeugen in molekularbiologischen Labors geworden.

Trotz des umfangreichen Wissens, das bei der Erforschung des T4 Phagen gewonnen wurde, sind zahlreiche Aspekte seines Infektionsprozesses nach wie vor ungelöst. Während etwa die Hälfte der T4 Proteine mit spezifischen Funktionen in Verbindung gebracht wurde, ist die Rolle der übrigen 45 % der T4 Proteine noch unerforscht. Allein diese Wissenslücke macht deutlich, dass unser molekulares Verständnis der T4 Phageninfektion und der eingesetzten Strategien, die dem Phagen eine effiziente Infektion ermöglichen, noch lange nicht vollständig ist. Daher wurden eine grundlegende Verbesserung unseres Verständnisses der T4 Phageninfektion auf molekularer Ebene und die Aufdeckung bisher unerforschte Infektionsmechanismen als Ziele dieser Dissertation gesetzt.

In **Kapitel II** wird eine Multi-Omics-Studie beschrieben, die eine zeitliche Auflösung einer Infektion von *E. coli* mit T4 Phagen auf molekularer Ebene bereitgestellt hat. Das Transkriptom und das Proteom von *E. coli* und T4 Phagen wurden während der gesamten Infektion analysiert. Dies ermöglichte die Identifizierung von zeitlichen Genexpressionsmustern für Transkripte des T4 Phagen. Bei bestimmten viralen Genen wurde eine Entkopplung von Transkriptions- und Translationsprozessen observiert. Die Transkriptom- und Proteomanalysen von *E. coli* ergaben einen allgemeinen Abbau der Wirts-Transkripte und die Erhaltung der Wirtsproteine. Zusätzlich

liefert diese Studie erstmals Einblicke in die molekulare Kinetik der T4 Phageninfektion. Die Ergebnisse deuten stark auf die Existenz zusätzlicher, unerforschter Infektionsregulationsmechanismen hin, die eine differenzielle Degradation von Wirts- und Phagentranskripten ermöglichen und überdies eine bisher unbekannte Entkopplung von Transkription und Translation für zahlreiche Phagengene begründen.

Eine mögliche Erklärung für die differenzielle RNA Degradation und damit die Unterscheidung zwischen bakteriellen und viralen Transkripten während der T4 Infektion könnte das Vorhandensein von RNA-Modifikationen sein. Zusätzlich würde dieses Phänomen eine fundierte Erklärung für die beobachtete Entkopplung von Transkription und Translation für einige T4 Phagengene liefern. **Kapitel III** fasst den aktuellen Wissensstand zum bakteriellen Epitranskriptom zusammen, wobei der Schwerpunkt auf mRNA-Modifikationen liegt. Die bekannten Proteine, die RNA Modifikationen einfügen, erkennen oder entfernen können, werden zusammen mit Techniken zur Identifizierung und Untersuchung spezifischer RNA Modifikationen diskutiert. Diese Arbeit zeigt eine erhebliche Wissenslücke in Bezug auf RNA Modifikationen in bakterieller mRNA, ihre Modulatoren und ihre biologische Bedeutung auf.

Während wenige Studien bereits Erkenntnisse über das bakterielle Epitranskriptom sammeln konnten, ist das Epitranskriptom von Bakteriophagen bisher hingegen unerforscht. In **Kapitel IV** wird das aktuelle Wissen über die RNA-Modifikation in Bakterien genutzt, um Hypothesen darüber aufzustellen, wie einige der bakteriellen und viralen Enzyme das Epitranskriptom von Bakteriophagen während der Infektion formen und modulieren könnten.

Um die biologische Rolle potenzieller Infektionsregulatoren aus dem Repertoire des T4 Phagen zu erforschen, muss untersucht werden, wie ihre Abwesenheit oder ihre Inaktivität die Phageninfektion beeinflusst. Dies erfordert jedoch effiziente Werkzeuge für die Phagenmutagenese. CRISPR-Cas ist in dieser Hinsicht ein vielversprechendes molekularbiologisches Werkzeug für präzises Genome Engineering, jedoch wird seine Effektivität für die Mutagenese von T4 Phagen durch abundante Modifikationen der T4 DNA stark beeinträchtigt. In **Kapitel V** wird ein Ansatz vorgestellt, der eine effiziente zeitliche Reduzierung der DNA-Modifikationen des T4 Phagen ermöglicht. Dadurch wird eine effiziente CRISPR-Cas-basierte Mutagenese von T4 Phagen ermöglicht. Darüber hinaus erlaubt dieses System auch die Untersuchung der Rolle von DNA-Modifikationen bei der Phageninfektion und hat das Potenzial auf andere Phagen ausgeweitet zu werden.

Kapitel VI befasst sich mit der T4 ADP-Ribosyltransferase ModB und zeigt, dass ModB nicht nur NAD als Substrat akzeptiert, sondern auch NAD-RNA. Auf diese Weise ist es ModB möglich eine neuartige post-translationale Modifikation, genannt RNAylierung, durchzuführen. Die Katalyse

der RNAylierung durch ModB wurde sowohl *in vitro* als auch *in vivo* nachgewiesen. Darüber hinaus wurde gezeigt, dass ModB mehrere *E. coli* Proteine, darunter die ribosomalen Proteine S1 und L2, RNAyliert. In diesem Kontext wurden die biologische Rolle und die molekularen Mechanismen der RNAylierung untersucht. Zusätzlich wurde eine T4 Phagenmutante mit einer katalytisch inaktiven Variante der ADP-Ribosyltransferase ModB erzeugt, um die Auswirkungen der katalytischen Aktivität auf die Phageninfektion und den Phänotypen zu untersuchen.

Zusammenfassend kann festgestellt werden, dass diese Arbeit zum besseren Verständnis der Infektion von *E. coli* durch T4 Phagen beiträgt, indem sie Einblicke in die zeitliche Organisation molekularer Abläufe der Infektion, die Auswirkungen von viralen DNA-Modifikationen auf den Phänotypen der Phagen und die Entdeckung der RNAylierung – eine neuartige posttranslationale Proteinmodifikation – liefert. Die in dieser Arbeit gewonnenen Erkenntnisse legen auch den Grundstein für den Transfer in die Anwendung. Insbesondere das Verständnis der zeitlichen Organisation der Phagengenexpression ist für die Erzeugung synthetischer Phagen oder die Anpassung der bereits existierenden Phagen für spezifische Anwendungen von wesentlicher Bedeutung. Weiterhin kann die in dieser Arbeit entwickelte und vorgestellte Mutagenesestrategie effizient eingesetzt werden, um solche Phagen mit verbesserten Funktionen zielgerichtet zu erzeugen.

Außerdem erweitert die Entdeckung der RNAylierung das Arsenal der von T4 Phagen abgeleiteten molekularen Werkzeuge. Die RNAylierung könnte in der synthetischen Biologie für die Entwicklung neuartiger künstlicher zellulärer RNA-Protein-Konstrukte eingesetzt werden und eröffnet somit neue Möglichkeiten für die Zukunft von RNA-basierten Therapeutika.

Insgesamt erweitert diese Studie unser Verständnis der molekularen Mechanismen, die einer effizienten T4 Phageninfektion zugrunde liegen, und unterstreicht, dass das Entdeckungspotenzial der T4 Phagenforschung noch lange nicht ausgeschöpft ist.

Chapter I

1. Introduction

1.1. Bacteriophages – viruses targeting bacteria

Bacteriophages, also referred to as phages, are viruses of bacteria and the most abundant biological entities in nature[1,2]. Their presence in various environments, from the deep ocean to gut microbiome, evidenced by unexplored phage DNA in metagenomic sequencing data ("phage dark matter")[3], suggests that we just began to uncover phage variety[4]. Moreover, the observed variations in morphology, size, and genomic organization among already isolated phages underscore their extensive diversity[5-7]. Phage genetic variation is ever-increasing, driven by the rapid evolution of phages in their arms race with their hosts[8,9] and through gene exchange via horizontal gene transfer[10-12]. In this way, phages can acquire genes from other phages, even those originating from phylogenetically distant hosts, resulting in phages representing the greatest genetic diversity on Earth[13].

Nevertheless, despite their diversity, all phages uniformly possess the ability to infect and replicate within bacterial cells. Phage infections are highly specific, characterized by a particular phage targeting a particular bacterial species (host) or even one specific strain within that species. The infection process begins with the injection of phage genetic material, which can be either DNA or RNA, into the host cell. Subsequently, phages can follow distinct, genetically predetermined lifestyles, such as lytic, lysogenic, or chronic. In the lytic life cycle, phages initiate replication at the expense of their host directly after injection of their genetic material into the bacterial cell. Phage replication leads to the assembly and release of new phage progeny, culminating in the lysis of the infected cell. In contrast, the lysogenic cycle involves the integration of the phage genome into the bacterial host genome (prophage) or its maintenance as an episomal element, with the potential to revert to the lytic cycle under specific triggers[14]. Differently from lytic and lysogenic infections, chronic infections lead to continuous replication and release of new viral particles while not causing the immediate lysis of the host cell[1,15].

Regardless of their life cycle, all phages offer a valuable platform for fundamental research. Their scientific value is strongly reflected in the number of groundbreaking discoveries that were made through phage research and have profoundly shaped our understanding of essential cellular processes. Such investigations led to the identification of DNA as the genetic blueprint of life[16], elucidation of principles in molecular genetics[17], and provided insights into key concepts of phage infection[18] and phage evolutionary mechanisms[19]. Additionally, phage-derived proteins, such as

polymerases and ligases, have become standard tools in molecular biology and were essential for the development of modern cloning and sequencing techniques[20,21].

While phage-derived proteins offer vast potential for biotechnological applications, focusing only on phages would disregard an important perspective: the interaction of phages with their hosts and the hosts' metabolic adaptations to resist phage infections. For instance, CRISPR-Cas (**c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats and **C**RISPR-**a**ssociated **p**roteins), which was discovered in the past years, revolutionized our idea about bacterial immunity and its adaptivity[22]. CRISPR-Cas was transformed into a powerful genetic tool with enormous application potential across all domains of life[23]. Another example is prokaryotic restriction enzymes, which contribute to innate bacterial immunity by degrading the nucleic acids of invading phages and are also widely utilized in molecular biology[24,25]. Furthermore, the rapidly advancing field of anti-phage defense systems[26] reflects the complex yet little-explored interplay between the virus and its host. Hence, phages remain an excellent model for exploring the molecular principles of the ongoing arms race between host and phage, their co-evolution, and the yet-to-be-discovered mechanisms driving efficient phage infection.

Today, phages, but in particular lytic phages, are of increasing interest for medical and biotechnological applications due to their ability to lyse host cells in a highly specific and rapid manner. For example, the outstanding ability of phages to control bacterial populations is exploited in industrial processes to prevent bacterial contaminations[27]. Furthermore, phages are applied for environmental monitoring, where they are used to determine the presence of specific bacteria in water and soil samples[28]. In diagnostics, phages enable the accurate detection of bacterial pathogens[29]. Apart from this, the potential to utilize phages to cure bacterial diseases, known as phage therapy, was already recognized in the early 20th century[30]. However, the discovery of antibiotics and their broader effectiveness overshadowed the application of phages for therapeutic purposes and consequently slowed down the research on phages. The current threat of widespread antimicrobial resistance among clinically relevant pathogens caused the renaissance of bacteriophage research. Nonetheless, effective phage application requires a comprehensive molecular understanding of infection mechanisms, which is essential for selecting or tailoring the ideal phage for specific applications.

1.2. T4 phage – a prime model for studying virus-host-interactions

The selection of an appropriate phage-bacterium pair is crucial for a sophisticated and representative investigation of the fundamental mechanisms of phage infection. Early research on phage molecular biology focused on those phages that are now recognized as model phages,

such as T4, T7, Lambda, M13, and MS2. Many of these model phages infect *Escherichia coli* – a model Gram-negative bacterium from the family of gammaproteobacteria. *E. coli* is a well-studied, genetically tractable, and easily culturable bacterium under laboratory conditions, which significantly streamlines research efforts to study phage-host interactions.

Among all model phages infecting *E. coli*, bacteriophage T4, shortly T4 phage, stands out as a lytic phage that has contributed significantly to our understanding of molecular biology and virology over the years. The advantage of T4 phage as a model phage is based on the total inhibition of host gene expression by the virus, which allows for differentiation between host and phage molecular synthesis[31]. The genome of T4 phage consists of 169 kbp of double-stranded DNA, encoding 288 genes[31]. The expression of T4 phage genes is highly efficient and strictly regulated. To ensure tight regulation during infection, the T4 phage, lacking its own RNA Polymerase (RNAP), hijacks and reprograms the *E. coli* RNAP (*Ec*RNAP)[32,33]. This reprogramming enables the phage to orchestrate a temporally regulated gene expression throughout the infection cycle, subdivided into early, middle, and late infection phases (Figure 1). During the early phase of the infection, those genes that are required to take control over the host are expressed. In the middle phase, the phage DNA replication is taking place. In the late phase, the synthesis of structural proteins and the assembly of structural components occur, which is crucial for generating new infectious phage progeny and facilitating their release. To run such an efficient and coordinated infection, the infection stages are strictly regulated at different molecular levels, including complex regulation within the phage genome, at transcriptional and translational levels, and through the modification of proteins.

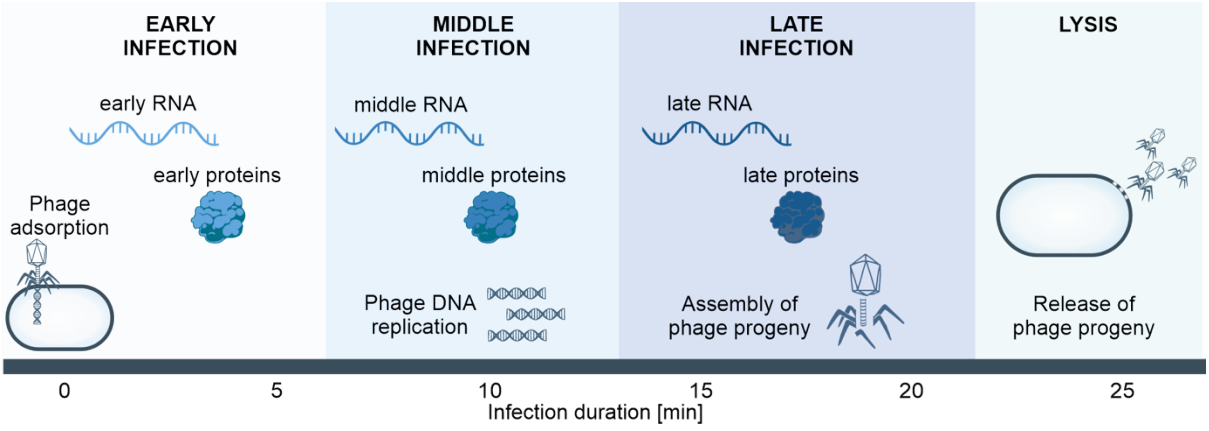


Figure 1: A schematic illustration of the process of T4 phage infection of *E. coli*. The infection of *E. coli* by T4 phage is a highly regulated process with tight regulation of the expression of early, middle, and late transcripts and the biosynthesis of corresponding proteins.

1.3. The blueprint of T4 phage – genome architecture and epigenetics

The infection of *E. coli* by T4 phage starts with attachment of the phage to its specific receptors at the *E. coli* surface, namely OmpC or lipopolysaccharide (LPS)[34,35], followed by injection of the phage DNA into the target cell. Within the 169 kbp T4 genome comprising a total of 288 encoded genes, 278 genes are encoding proteins, eight – tRNAs, and two – non-coding small RNAs[31]. Based on this correlation, the gene density within the T4 phage genome is very high and twice that of its host *E. coli*[31].

The high gene density of the T4 phage genome is also reflected in the highly complex gene arrangement within the DNA. For example, some genes have internal starts within a sequence of other, longer genes (e.g., *gene products (gp) 17* and *19*). Thus, both gene products, long and short, originating from the same phage genome segment, are distinct from each other in function[36,37]. The T4 phage genome also features overlapping genes, including the *gp30.3'*, *repEA*, and *repEB*[38,39]. Introns that are later spliced out post-transcriptionally, e.g., in the genes *td*, *nrdB*, and *nrdD*, have also been observed in T4 phage[40,41]. Conversely, the T4 phage gene *gp60* is flanked by regulatory elements that facilitate translation bypass. In this process, ribosomes skip a non-coding mRNA region of 50 nucleotides to translate Gp60[42]. Collectively, all these aspects highlight the complex coding arrangements within the T4 phage DNA that enable it to encode 288 genes within its compact 169 kbp genome.

The phage DNA is one of the targets for bacterial anti-phage defense systems[43]. Bacteria attempt to neutralize invaders early by nucleolytically degrading phage DNA to prevent phage gene expression and replication. However, phages have evolved strategies to avoid the degradation of their genetic material by extensively modifying their DNA and thereby protecting it from host nucleases[44]. This adaptation justifies the enormous diversity of phage DNA modifications, which entity for a specific phage is defined as its epigenome[45]. Maintaining extensive genome modifications to protect phage DNA from host nucleases is also a strategy encountered in the T4 model phage.

The epigenomic regulations used by T4 phages to protect their DNA from host defense systems comprise hypermodifications of cytidines (Figure 2). The cytidines within the T4 genome are entirely modified to 5-hydroxymethyl-2'-deoxycytidine (5hmdC)[46] and further glycosylated to 5-α-/β-glycosylhydroxymethyl-2'-deoxycytidines. The initial 5hmdC modification occurs at the single nucleotide level, where deoxycytidine monophosphate (dCMP) is hydroxymethylated to 5hmdCMP by the dCMP hydroxymethyltransferase Gp42. 5hmdCMP is further converted to 5hmdCTP (5-hydroxymethyl-2'-deoxycytidine triphosphate) by Gp1, a deoxynucleotide monophosphate kinase (dNMP kinase). In the next step, 5hmdCTP is incorporated into the phage

DNA polynucleotide chain by the phage DNA polymerase. Once part of the DNA, 5hmdC is exclusively glycosylated to 5ghmdC by DNA glycosyltransferases. These bulky DNA glycosylations protect phage DNA from the host-derived nucleases.

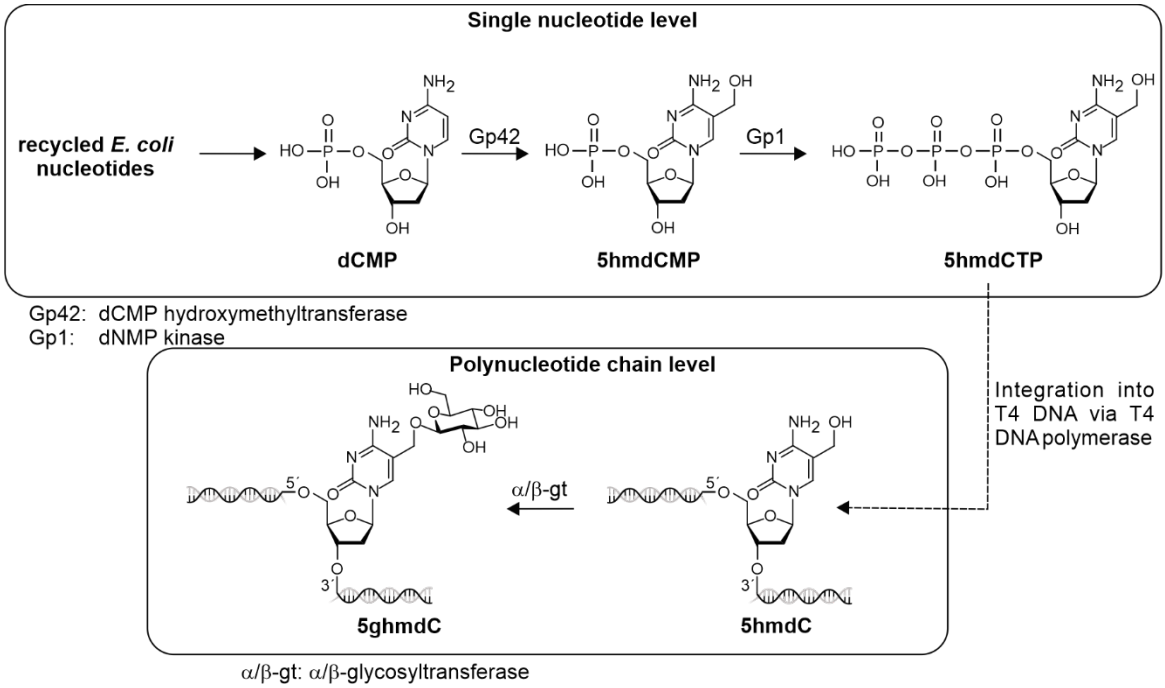


Figure 2: Biosynthesis and incorporation of cytidine hypermodifications into the T4 phage genome.

Apart from protection from host nucleases, these cytidine modifications are crucial for the phage to distinguish between self and host DNA, enabling the selective degradation of the latter by phage-encoded DNA nucleases[47-49]. The degradation of bacterial DNA facilitates the takeover of the host cell's metabolism and secures a nucleoside pool for phage DNA replication.

Taken together, the T4 epigenome plays a central role in establishing an efficient expression of phage genes in a host cell, selective degradation of host DNA, and preserving the integrity of the highly complex T4 genome to serve the primary infection goal: phage replication and formation of the viral progeny.

1.4. The T4 phage gene expression is a highly organized process

The transcription of T4 phage genes is a highly efficient process, which phages maintain by utilizing the transcription machinery of its host *E. coli*, as T4 phage does not possess its own transcriptional machinery. The takeover is realized by reprogramming the *Ec*RNAP to prioritize phage gene expression over the hosts directly after the phage DNA injection into the bacterial cell. To achieve efficient reprogramming, T4 phage employs specialized proteins known as host

acquisition factors (HAFs)[31]. The HAFs either repurpose, inhibit, or modulate essential bacterial processes, ensuring the successful phage infection. Among the known T4 phage HAFs facilitating the transcriptional reprogramming are adenosine diphosphate-ribosyltransferases (ADP-ribosyltransferases) and transcriptional regulators that alter *Ec*RNAP to recognize and bind specific phage promoters[31,50]. The T4 phage promoters and the genes under their control can be categorized into distinct temporal groups: early, middle, and late. This results in sequential expression of specific phage gene sets at distinct infection stages. Specifically, following the onset of initiation, ~40 early phage promoters are recognized by the unmodified sigma factor σ70 *Ec*RNAP holoenzyme[51]. Next, the ADP-ribosyltransferase Alt, which is a HAF incorporated within the viral capsid and delivered into the bacterium alongside phage DNA, introduces a post-translational modification to *Ec*RNAP, named ADP-ribosylation. The modification of *Ec*RNAP enhances its specificity toward distinct early phage promoters[52,53] and increases twofold the transcriptional activity for associated genes[54]. Even though the ADP-ribosylation impact on promoter specificity and transcriptional activity has been proven, the precise molecular consequences initiated by this post-translational modification have yet to be determined. At the same time, another T4 HAF, Alc, suppresses the transcription of host DNA containing unmodified cytidines by acting as a site-specific termination factor[55-57].

After just a few minutes of infection, the transcription shifts toward T4 middle genes[58]. This transition is partially mediated by ModA, another ADP-ribosyltransferase, which modifies the *Ec*RNAP α-subunits to shift transcription to middle phage genes[58]. However, the detailed mechanism by which ADP-ribosylation introduced by ModA influences this activity shift remains unclear. In addition to post-translational *Ec*RNAP modifications by ADP-ribosyltransferases, activation of middle promoters requires two T4-encoded early proteins, MotA and AsiA[58-60]. Both proteins interact with the *E. coli* sigma factor σ70 to alter the specificity toward middle phage genes, a process known as σ appropriation[53,61].

To initiate the expression of genes controlled by late phage promoters, a T4-encoded sigma factor, the gene product 55 (Gp55), is required for late promoter recognition[62]. The Gp55, a distant homolog of σ70[63], replaces *E. coli* σ70 in *Ec*RNAP, effectively abrogating *E. coli* σ70-dependent transcription[64]. However, an efficient recognition of late promoters and transcription of late genes also demands the presence of Gp33 and Gp45. The inclusion of Gp45 in the Gp33-Gp55-*Ec*RNAP complex enhances the transcription of T4 late genes under the control of late promoters by more than 1000-fold[65,66]. However, the mechanistic details of this process remain to be elucidated.

In summary, the redirection of *Ec*RNAP for phage gene expression involves three key steps: (I) introduction of post-translational modifications to *Ec*RNAP, (II) appropriation of *E. coli* σ70, and (III) replacing it with proteins encoded by the T4 phage. While some T4 proteins have been identified as being relevant to one of these steps, the molecular mechanisms by which they contribute to a particular process are unknown for most of them. This knowledge gap highlights that our current understanding of the molecular strategies used by T4 phages for host redirection is incomplete and requires further investigation.

1.5. Bacterial and viral nucleases shape the dynamic transcriptome during infection

The transcriptional shift toward phage gene expression during T4 phage infection is not only achieved by reprogramming of *Ec*RNAP but is also contributed by selective RNA degradation[67]. During infection, specific host and phage transcripts are actively degraded. On the one hand, the degradation of host RNA facilitates the switch to T4 metabolism by abolishing host gene expression, creating a pool of free ribonucleotides and ribosomes, and stimulating transcription and translation of T4 genes[68]. On the other hand, the degradation of phage transcripts in a stage-dependent manner allows the functional resolution of the infection[69-72]. Modulation of the transcriptome, an entity of the RNA, during T4 phage infection, is achieved by host nucleases, whose activity is influenced by phage-derived proteins and phage-derived nucleases.

To date, only a few proteins originating from T4 phage and *E. coli* have been associated with RNA turnover during T4 phage infection[72]. One of these is RNase E, an essential *E. coli* endoribonuclease[73] that was the first identified endonuclease involved in T4 mRNA processing[74,75]. RNase E was shown to mature gp32 mRNA[74] and prevent premature expression of the late gene *soc*[76]. Furthermore, RNase E fulfills another role during the infection – degradation of host mRNA[67]. This degradation is stimulated by the early T4 phage protein Srd, which exemplifies another phage mechanism to reprogram the cellular machinery of the bacterial host to meet the needs of the virus[68]. Both regulation mechanisms, namely RNase E-mediated maturation of the phage transcripts and the involvement in transcript degradation, show the importance of RNase E in transcript turnover upon infection[71].

Another *E. coli* RNase, RNase LS, targets specific groups of T4 phage transcripts, the middle and late transcripts, to prevent phage progeny formation and release via phage-mediated cell lysis[77,78]. However, T4 counteracts the degradation with its early gene product Dmd, which inhibits RNase LS's action on phage transcripts, allowing the infection cycle to be completed[69,77].

Beyond the role of nucleases in the selective degradation of the phage or host transcripts, certain nucleases are orchestrating the time-specific presence of T4 mRNA. For instance, the T4-

encoded sequence-specific endoribonuclease RegB targets early phage mRNAs, potentially facilitating a shift of phage gene expression from early to middle phase[70,79].

The provided examples highlight the few known nuclease-mediated regulatory mechanisms that affect the transcriptomes of *E. coli* and T4 phage during infection. To improve our understanding of transcriptional regulation during T4 phage infection, it is necessary to elucidate the dynamics of the transcriptome, identify other potential transcriptome modulators, and explore their mechanisms.

1.6. The regulation of the dual-transcriptome of a T4 phage infection of *E. coli*

The preceding overview of dual-transcriptome (transcriptome of phage and host) regulation during a T4 phage infection of *E. coli* (Chapter 1.5) is based on numerous valuable studies focused on specific transcripts and proteins. Nonetheless, such investigations are limited in assessing the role of specific processes within the broader cellular context. Recent advances in RNA detection and sequencing technologies enable comprehensive transcriptome studies, facilitating the investigation of specific biological processes in a global cellular context and providing a holistic understanding of transcriptional regulation and dynamics[80].

The first comprehensive study aiming to elucidate the temporal expression patterns of T4 phage genes was conducted using a microarray technique[81]. This study verified the temporally controlled transcription of T4 phage genes, and the relative abundance of specific phage transcripts was determined. The study offered the first temporally-resolved analysis of the T4 phage transcriptome during infection. However, it focused solely on phage transcripts, leaving the dynamics of the host transcriptome upon phage infection disregarded.

Simultaneous investigations on phage and host transcriptome became first possible with the advent of next-generation sequencing techniques. Next-generation sequencing was already applied to elucidate the functional role of an *E. coli* toxin-antitoxin system during T4 phage infection[82]. This work greatly exemplifies the utility of RNA-sequencing (RNA-Seq) in studying the host's response to phage infection on the transcriptome level. Nevertheless, the study is focused on the functionality of the toxin-antitoxin system, therefore not addressing the dynamics of the dual-transcriptome of T4 phage-infected *E. coli*.

Although RNA-Seq has not been applied to explore the transcriptome of *E. coli* with T4 phage, it has been used to investigate the infection of marine bacteria by various phages on transcriptional level[83,84]. The transcriptome was studied at different time points of infection, revealing metabolic reprogramming following infection, specific changes in resource acquisition, and shifts in central carbon and energy metabolism. Thus, extending RNA-Seq-based dual-transcriptome studies to

the model T4 phage-*E. coli* pair could significantly improve the understanding of infection-regulatory mechanisms and uncover complex interactions that cannot be achieved by focusing on individual transcripts or mechanisms. Furthermore, time-resolved studies of T4 phage infection could elucidate the transcriptional kinetics of T4 phage and shed light on the transcriptional response of *E. coli* to infection, providing valuable insights into the interplay between phage and host during the infection process.

1.7. RNA modifications as a putative layer of infection regulation

During the T4 phage infection of *E. coli*, both phage and bacterial proteins influence the dual-transcriptome by differentially processing and degrading the transcripts. The mechanisms underlying the distinction between host and phage transcripts remain unexplored. A possible explanation for the distinction of transcripts upon infection could be the presence of RNA modifications. As the T4 phage modifies its DNA to evade bacterial nucleases and to differentiate between viral T4 DNA and the bacterial host DNA, modifications on phage RNA could similarly regulate the stability and the fate of T4 phage transcripts during the infection.

However, the presence of RNA modifications and their potential impact on the phage infection cycle and the interaction between T4 phage and *E. coli* has not been studied yet. Also, the knowledge about the *E. coli* epitranscriptome, defined as the entity of RNA modifications, remains limited. To date, only a few modifications have been identified in *E. coli* mRNA, such as inosine (I, deamination product of adenosine)[85,86], N^6-methyladenosine (m^6A)[87], and certain 5'-terminal RNA-caps.

Inosine and m^6A are internal RNA modifications affecting adenine bases. Their impact and their biological role in the context of *E. coli* mRNA is largely unexplored. Although the m^6A modification is prevalent within bacterial mRNA, its functional role is yet not known[87]. Analogously to m^6A, the biological function of inosine in *E. coli* mRNA is not thoroughly explored. The deamination of adenosine to inosine (A-to-I editing) is either spontaneous or an enzyme-catalyzed process[88]. In *E. coli*, the TadA enzyme catalyzes adenosine deamination[85]. The TadA-mediated A-to-I editing within protein-coding sequences was shown to contribute to the re-coding of tyrosine codons to cysteine codons. Specifically, among the re-coded transcripts is the hokB transcript, which encodes the toxin HokB involved in the self-destruction of *E. coli*. The A-to-I editing in this transcript results in a more toxic and stable variant of the HokB protein[85].

The terminal RNA modifications are modifications at the 5'- or 3'-terminus of RNA. The 5'-terminus of the RNA determines its origin. Typically, transcription is initiated with nucleoside triphosphates, resulting in a 5'-triphosphorylated primary RNA (5'-PPP-RNA) transcript. The 5'-PPP-RNA can be

nucleolytically processed, yielding a monophosphorylated 5'-terminus (5'-P-RNA)[89]. Interestingly, *Ec*RNAP can use non-canonical nucleotides for transcription initiation instead of the conventional nucleoside triphosphates (adenosine, cytidine, guanosine, and uridine triphosphates), leading to the creation of 5'-capped transcripts. Such non-canonical caps, including nicotinamide adenine dinucleotide (NAD, used in this work to refer to oxidized NAD^+), flavin adenine dinucleotide (FAD), uridine diphosphate glucose (UDP-Glc), and uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), have been detected on *E. coli* RNA[90]. Some of these modifications, such as the NAD-cap, are hypothesized to protect RNA from degradation by nucleases recognizing the 5'-terminus of RNA[91,92].

While RNA modifications could affect the stability of host transcripts, their presence in phage transcripts and their potential function in phage infection remains unexplored. Given the abundance of RNA modifications observed in eukaryotic viruses[93-96], it is plausible that phage RNA might also be modified. These modifications could be a molecular basis for discrimination between phage and host transcripts by nucleases. Investigations of RNA modifications in both phage and host during viral infection could unveil infection regulatory mechanisms that are yet unknown.

1.8. The dual-proteome of a T4 phage infection of *E. coli*

The T4 phage proteome, the entity of the proteins encoded by the T4 phage, comprises 278 proteins. The functionality of approximately 55% of T4 phage proteins was revealed in numerous studies, focusing mainly on specific proteins. Nevertheless, despite the T4 phage being one of the most extensively studied phages, the biological functions of the other 45% of its proteins remain to be elucidated. The rapid evolution of the T4 phage genome and the predominantly small size of its uncharacterized proteins define the challenge of revealing their specific role and function based on their sequence[97]. The phylogenetic diversity of these proteins also complicates the sequence-based determination of function: Some T4 proteins show orthologous relationships across all domains of life (eukaryotes, archaea, or bacteria) and may even share sequence homology with eukaryotic viruses, underscoring the complexity and evolutionary dynamics of the T4 phage proteome[31].

Given the challenge of determining the roles of uncharacterized T4 phage proteins through sequence analysis, structural analysis could serve as an alternative. Nevertheless, the attempts to elucidate the role of some T4 phage proteins through the determination of their structure have revealed unique protein folds that suggest entirely novel functions for the analyzed proteins[98].

This emphasizes the vastly unexplored infection mechanisms and interactions between T4 phage and *E. coli*.

Initial efforts were made in the 1970s to determine to which processes T4 phage proteins contribute or in which pathways they participate. Back then, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was employed to identify pre-replicative (proteins entering the cell along with the DNA) and the structural (proteins that constitute the phage virion) T4 proteins[99-101]. The assignment of some structural phage proteins to specific genes was confirmed by independent experiments. The association between a gene and a protein was validated by purifying proteins of interest from phage virions and their sequencing using limited proteolysis and Edman degradation[102-104].

The two-dimensional (2D) SDS-PAGE was also instrumental in the identification of T4 phage genes encoding for specific phage proteins. To approach this, experiments involving infections with both wild-type T4 phage and T4 phages with amber mutations within specific genes (mutation leading to premature polypeptide chain termination) were conducted. The dual-proteome, the entity of phage and host proteins at a particular time point of infection was then analyzed using 2D SDS-PAGE. The absence of specific protein bands in 2D gels of infection with the mutants, in contrast to their presence in 2D gels of infection with wild-type phage, made it possible to assign the proteins to specific phage genes[105]. The assignment led to the naming of T4 proteins based on gene names, which had previously often been represented by gene numbers due to their unknown functions.

While the insights from these 2D SDS-PAGE-based experiments have enhanced our understanding of gene-protein relations in T4 phage, the methodology is time-consuming, significantly limited by its detection threshold, resolution, and sensitivity, and is frequently leading to ambiguous results. Most of these limitations were overcome with the advent of liquid chromatography coupled with mass spectrometry (LC-MS)-based proteomics, which allows studying the entity of cellular proteins with high sensitivity and in-depth[106]. Its application in studying phage infection might provide insights into the host and phage proteome modulation upon infection.

The power of LC-MS-based proteomics to study phage-bacteria interactions has already been explored for marine bacteria, revealing both the temporal resolution of phage protein biosynthesis and the response of the host to phage infection on the protein level[83,84]. Therefore, LC-MS-based proteomics can be applied to study *E. coli* infection by T4 phage in a time-resolved manner to determine the changes in the relative abundance of the host and phage proteins throughout the infection and obtain the temporal patterns of phage protein biosynthesis. The temporarily resolved

proteome study of T4 phage infection could also provide initial hints about phage protein functionality based on their biosynthesis onset. More precisely, phage proteins synthesized early in the infection process are prime candidates for being involved in host takeover and initiation of phage replication, while the late proteins are likely involved in phage assembly and release of phage progeny from the bacterial host[107].

Apart from the advantages of LC-MS-based proteomics, it is important to note that this approach may have limitations for detecting small T4 phage proteins. Particularly, the proteolytic digestion step, which is critical for proteomics sample preparation, produces short peptides for small proteins, potentially rendering them undetectable by MS[108,109]. In addition, data processing in LC-MS analysis often includes standard filter criteria optimized for larger proteins, such as the “two unique peptides” filter being a criterion for confidential protein detection[108]. Consequently, small phage proteins may remain undetected due to their shorter sequence, resulting in fewer detectable peptides. Despite these limitations, LC-MS-based analysis of the proteome of *E. coli* infected with T4 phage represents a powerful approach that can provide new insights into infection mechanisms and reveal patterns that were previously undetected.

1.9. Post-translational protein modifications as an additional layer of infection regulation

The roles of some T4 phage proteins involved in host hijacking have been partially elucidated. Among such proteins are T4-derived ADP-ribosyltransferases (ARTs) that are instrumental in reprogramming host proteins (Chapter 1.4). T4 phage encodes three ARTs, Alt, ModA, and ModB, that catalyze the transfer of ADP-ribose from NAD to arginine residues of a target protein, leading to its ADP-ribosylation (Figure 3)[110]. The resulting post-translational protein modification alters the properties and functionality of the modified protein[111,112].

The Alt protein, which is a part of the T4 phage virion, is introduced into the bacterial cell with the phage DNA, acting as a pre-replicative HAF[113]. Inside the bacterial cell, Alt displays its catalytic activity by ADP-ribosylating α-, β-, and β’-subunits of *Ec*RNAP, as well as the σ^{70}-factor[114,115]. These modifications by Alt enhance the affinity of *Ec*RNAP for early T4 promoters, leading to up to 50% higher transcription of the corresponding genes[116]. Under control of T4 early promoters two other T4 ARTs are encoded, ModA and ModB. The target of ModA is the α-subunit of *Ec*RNAP. The ADP-ribosylation introduced by ModA shifts the transcription preference of *Ec*RNAP toward the genes under the control of the middle promoter[117]. Consequently, both Alt and ModA play a crucial role in reprogramming the transcriptional machinery of the T4 phage during the early phase of the infection by coordinating the temporal expression of phage genes.

Unlike Alt and ModA, which modify the *E. coli* transcriptional machinery, ModB specifically targets the translational apparatus by ADP-ribosylating the ribosomal protein S1 (rS1)[117]. However, the regulatory effect of rS1 ADP-ribosylation on translation – whether it enhances or inhibits the process – remains to be determined.

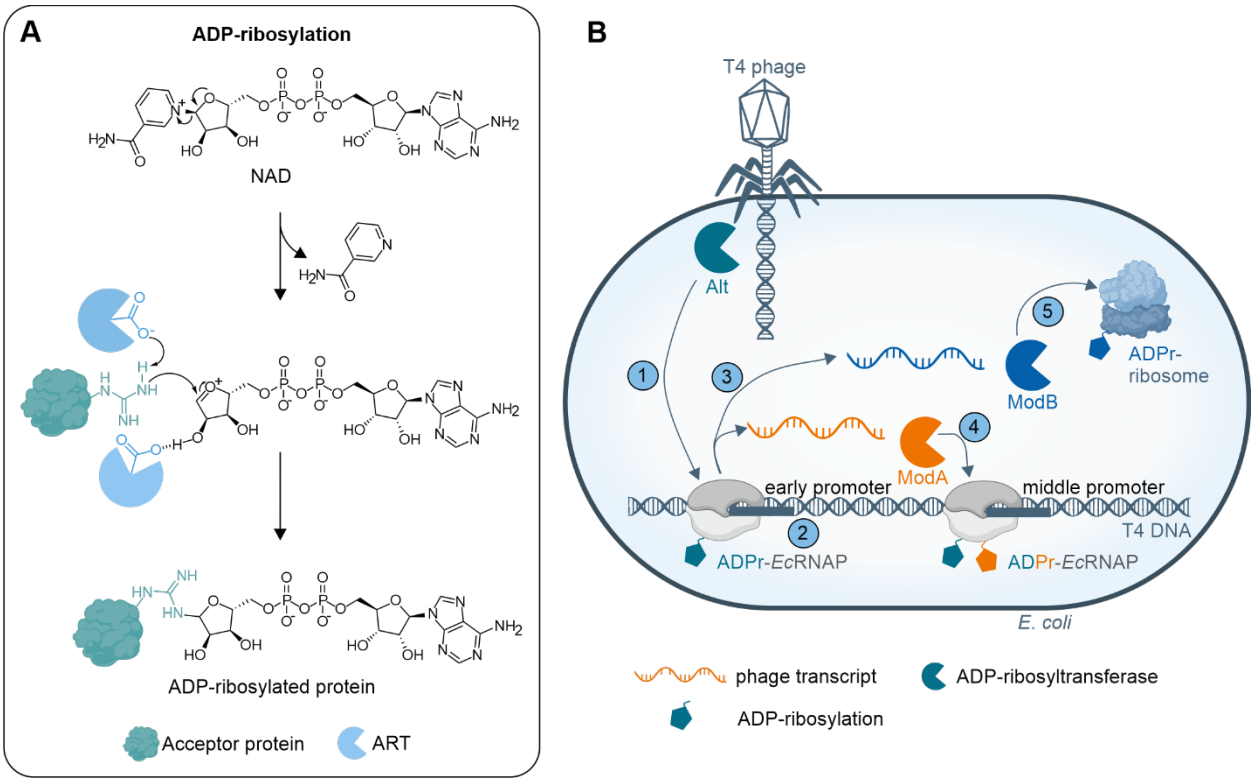


Figure 3: ADP-ribosylation of proteins catalyzed by T4 ARTs. A: Mechanism of ADP-ribosylation catalyzed by ARTs. First, a glutamate residue of ART destabilizes the N-glycosidic bond between ribose and nicotinamide in its substrate NAD, leading to the formation of the oxocarbenium ion of ADP-ribose, while nicotinamide serves as a leaving group. The arginine residue in the acceptor protein performs a nucleophilic attack on the oxocarbenium ion, which is facilitated by glutamate-mediated proton abstraction (mediated by ART), forming a new N-glycosidic bond[118]. **B:** Current stage of knowledge of involvement of T4 ARTs in hijacking *E. coli* gene expression machinery: Alt, a pre-replicative protein, enters the *E. coli* cell with phage DNA (1) and ADP-ribosylates *Ec*RNAP (2), enhancing its affinity for early T4 promoters and facilitating early gene expression (3). Subsequently, other ARTs like ModA and ModB are produced. ModA modifies *Ec*RNAP to shift its affinity toward T4 middle promoters (4), while ModB targets *E. coli* ribosomes (5), with the specific implications of this modification remaining to be elucidated.

*Ec*RNAP and rS1 represent only a fraction of T4 ART targets. Alt, for instance, was shown to modify at least ten other proteins, such as prolyl-tRNA-synthetase, GroEL, and pyruvate kinase I, among others[119]. ModB also ADP-ribosylates additional proteins, including trigger factor TF,

elongation factor EF-Tu, and ModB itself[119]. Detection of these ART targets was achieved by *in vitro* ADP-ribosylation assays using *E. coli* crude cells extracts and radiolabeled NAD, with subsequent identification of modified proteins by 2D SDS-PAGE, in-gel digestion and LC-MS/MS[119]. However, as previously discussed (Chapter 1.8), 2D SDS-PAGE is an approach limited in resolution and sensitivity, which could result in the misidentification of targets due to the co-isolation of co-migrating proteins during sample preparation for LC-MS/MS. For example, the identified ModB substrate EF-Tu is one of the most abundant bacterial proteins, making up to 6% of the total protein expressed in *E. coli*[120], which may be an artifact of the given experimental setup. Furthermore, the experimental setup only allowed the identification of *E. coli*-derived ART targets and did not consider possible T4 phage-derived proteins, as only the crude host extract was used.

Nevertheless, the results of this valuable study suggest that T4 ARTs target a range of proteins in addition to *Ec*RNAP and rS1, and, therefore, may have broader implications on phage infection than previously thought[119]. To verify the identified targets and to discover potentially unexplored ones, modern techniques such as immunoblotting and detection of modified proteins by LC-MS can be applied and provide insights into the unexplored roles of T4 ARTs. Exploring the role of ADP-ribosylation on post-translationally modified proteins can reveal novel strategies employed by T4 phage to hijack its bacterial host. Even more, the generation and characterization of T4 phage mutants with inactive Alt, ModA, or ModB proteins could provide insights into how ADP-ribosylation by specific ARTs affects the entire infection cycle, both at the phenotype and molecular level.

1.10. Mutagenesis as a key toward elucidation of biological function

A direct approach to understanding the biological role of specific phage enzymes, such as T4 ARTs, is to examine how their absence or inactivity affects the infection cycle. The resulting consequences can then be studied both at the molecular level and at the phenotypic level. Mutagenesis emerges as a crucial method for this purpose. Gene deletions have been preferred so far in evaluating the function of phage genes. However, due to the complex nature of phage genetics, complete deletion of a gene from the phage genome can significantly alter phage biology, potentially affecting the expression of downstream or upstream genes (Chapter 1.3). Moreover, in the case of studying solely catalytic roles of proteins in the infection cycle, the entire deletion of proteins might disrupt their other functionalities: Apart from being biocatalysts, proteins can be simultaneously involved in protein-protein interactions or serve as structural components of the phage.

To circumvent the throwbacks associated with whole gene deletions in the phage genome, a more specific strategy would be preferred. For example, in the case of enzymes, mutagenesis targeting only catalytically essential residues would be advantageous, thereby eliminating only one of the putative functionalities of the target protein and preserving its potential participation in other cellular processes.

However, targeted mutagenesis of lytic phages is challenging. In contrast to prophages with an integrated phage genome in the host genome, for which established host mutagenesis techniques can be used, lytic phages require tailor-made mutagenesis approaches. Historically, irradiation of phage DNA with UV light or ionizing radiation has been employed to introduce random mutations or gene deletions within the phage genome[121-123]. Phages exhibiting specific phenotypes or genotypes were then selected for detailed analysis. This approach has enabled initial discoveries regarding the critical roles of specific genes in the infection process, including assessments of gene essentiality. However, due to its randomness, this mutagenesis method is unsuitable for the site-specific introduction of mutations.

In recent years, several strategies have been developed for targeted mutagenesis of lytic phages. Most of these strategies rely on homologous recombination (HR)[124]. During phage infection, HR enables the integration of mutations into phage DNA from a donor DNA, which is flanked by homologous regions to the mutation insertion site. However, the frequency of successful mutagenesis by HR is relatively low, typically ranging from 10^{-10} to 10^{-4} [125]. This frequency can be lightly boosted by co-expression of phage-encoded recombination systems upon mutagenesis.

Another technique that is based on HR is “bacteriophage recombineering of electroporated DNA” (BRED). The approach involves the co-electroporation of phage DNA and donor DNA into the host, significantly increasing the efficiency with mutation rates reaching up to 10–15% [126]. Nevertheless, the applicability of BRED is constrained by the size of the bacteriophage genome and the transformability of the host.

Recently, the groundbreaking discovery of CRISPR-Cas has significantly advanced the field of genome engineering. Enabling precise mutagenesis in various organisms, CRISPR-Cas-based mutagenesis was also extended to phage mutagenesis[23,127,128]. For the latter, mutagenesis is based on the selection of a specific guide RNA sequence that guides the Cas nuclease to the intended mutagenesis site within the phage genome. Cas-mediated DNA targeting induces a double-strand break in phage DNA, preventing phage replication. Nevertheless, the break can be repaired via homologous recombination with a donor DNA present in the mutagenesis strain. The donor DNA features intended mutations flanked by homologous regions to the cleavage site. The recombination machinery of the phage directs the phage DNA repair via HR, thereby integrating

the mutation into the genome. Unfortunately, the CRISPR-Cas-based mutagenesis approach is frequently impeded by phage DNA modifications[127,128]. Particularly, to protect its DNA against host nucleases, some of the phages, including the T4 phage, extensively modify their DNA by introducing bulky functional groups, such as glycosylation, to escape recognition and degradation by the host (Chapter 1.3)[31]. Therefore, the efficiency of the mutagenesis of phage DNA with CRISPR-Cas is strongly impeded in the initial Cas-mediated DNA cleavage step[127,128]. The low targeting efficiency leads to a correspondingly low mutagenesis efficiency, which is up to 3% and highly dependent on the chosen spacer[128].

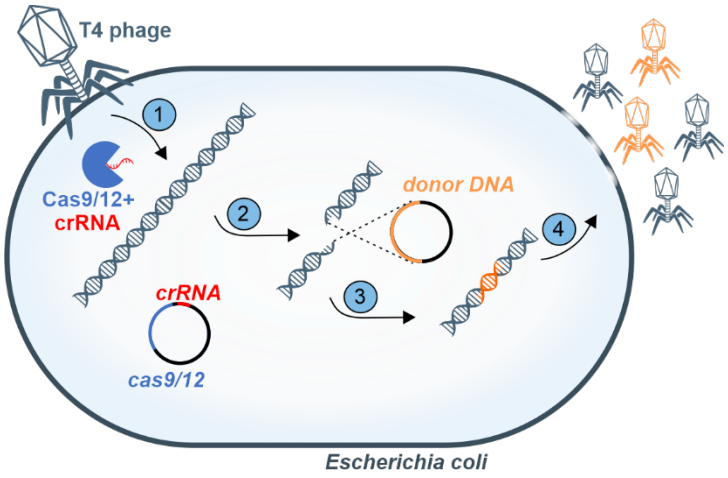


Figure 4: Simplified illustration of CRISPR-Cas-based *in vivo* phage mutagenesis using T4 phage as an example. Mutagenesis begins with the injection of phage DNA into the mutagenesis strain (1). Once inside the cell, the phage DNA is targeted by Cas nuclease (e.g., Cas9 or Cas12) at a specific position guided by a specifically designed guide RNA, resulting in a DNA double-strand break (2). The double-strand break is repaired by homologous recombination with the donor DNA, which is present on the plasmid within the mutagenesis strain and carries the intended mutations (3). The mutations are then introduced into the phage genome. However, due to low mutagenesis efficiency, cell lysis results in the release of both wild-type and mutant progeny.

In addition to *in vivo* approaches, phages can also be mutated *in vitro*. Although molecular cloning represents one of the options for the introduction of mutations into phage DNA, its effectivity is limited by the length of the DNA and the subsequent efficiency of its transformation to the cell[129]. Complete *in vitro* genome synthesis via polymerization was also reported for smaller phages[130]. Nonetheless, both presented *in vitro* approaches are inefficient and elaborate for introducing minor mutations into the phage genome.

Taken together, an efficient approach for site-specific mutagenesis of lytic phages is yet to be developed. The establishment of such an approach would make it possible to study the function of specific phage proteins. Once the underlying specific molecular mechanisms are understood,

these mutagenesis strategies can then be applied to customize and tailor phages, adjusting their properties for targeted applications.

1.11. Aims of this work

The T4 phage infection of *E. coli* is one of the most extensively studied bacterium-phage interactions, yet many facets of its molecular organization remain to be explored. Based on the current knowledge regarding T4 phage infection, several questions are arising: How exactly does the T4 phage maintain such an efficient transcription and translation takeover of the host? What is the basis for differentiation between phage and host macromolecules, such as RNA, during infection? Could there be another level of translation regulation, perhaps by RNA modifications? What is the role of post-translational protein modifications in the infection process? Furthermore, only 55% of 278 T4 phage proteins are assigned to specific functions. So what are the biological functions of the remaining uncharacterized 45% of T4 phage proteins? Answering these questions could reveal critical phage proteins involved in efficient host hijacking and possibly allow the identification of previously unknown mechanisms of host hijacking employed by T4 phage.

To assess the biological functions of specific proteins and their contribution to the processes raised in the questions above, an approach for efficient T4 phage mutagenesis is required. However, T4 phage mutagenesis remains a difficult endeavor to date due to extensive modifications within the phage DNA. Thus, overcoming the challenge of phage mutagenesis could provide a basis for further studies of specific phage proteins with unknown or only partially explored functions.

Therefore, this thesis aims to improve our molecular understanding of the T4 phage infection process in *E. coli*. First, a detailed picture of the regulation of phage infection at the molecular level needs to be obtained. This will provide insight into the molecular organization of infection and potentially lead to the identification of critical modulators of phage infection. To elucidate the biological function of specific phage proteins, a strategy for efficient phage mutagenesis is aimed to be developed.

Chapter I describes a multi-omics study with the aim to obtain a temporal resolution of the *E. coli* infection with T4 phage at the molecular level. The analysis of the infection process of *E. coli* with T4 phage at the transcriptome and proteome level allows the relative quantification of both host and phage transcripts and proteins throughout the infection. The kinetics of phage gene expression are studied, and consequently, phage transcripts and proteins are assigned to specific temporal groups of the infection. Furthermore, for the first time, the transcriptome and proteome of *E. coli* upon phage infection are analyzed, and the transcriptional and protein biosynthetic response of *E. coli* to T4 phage infection is obtained. This chapter provides insights into the molecular regulation of T4 phage infection, e.g., transcriptional and translational decoupling for specific phage genes and selective overall degradation of host transcripts.

The selective degradation of host RNA upon infection requires an explanation, of how the transcripts of host and phage can be differentiated from each other. One of the potential answers to this could be the presence of RNA modifications. **Chapter II** reviews the existing knowledge of the bacterial epitranscriptome with a focus on mRNA modifications, the methods existing for their identification, and their writers, readers, and erasers. This chapter highlights a significant gap in our knowledge about the presence of RNA modifications in bacterial mRNA, their regulatory roles, and biological significance.

Although initial insights into bacterial epitranscriptome were gained, the epitranscriptome of the phages is not explored to date. In **Chapter III**, the current knowledge about RNA modifications in bacteria is used to hypothesize how certain proteins from both bacteria and phages might influence and adapt the epitranscriptome of bacteriophages during infection.

Investigation of the biological roles of potential infection regulators requires analysis of how their absence or inactivity affects infection patterns. **Chapter IV** aims to develop a CRISPR-Cas-based approach for precise phage mutagenesis. The main burdens for this are currently the T4 DNA modifications that impede Cas nuclease-mediated DNA cleavage. The targeted approach includes temporal reduction of the abundance of phage DNA modifications to efficiently introduce point mutations into phage coding sequences, e.g., to mutate enzymes and therewith abolish their enzymatic activity. In such a way, the impact of their inactivity on phage infection can be studied, avoiding the deletion of entire genes and severely impacting the integrity and regulation of the phage genome.

Chapter V focuses on the ADP-ribosyltransferase ModB, a crucial regulator in the T4 phage infection process, and the identification of a novel post-translational modification RNAylation catalyzed by it. The potential role and molecular mechanisms of RNAylation are investigated. A T4 phage mutant with catalytically inactive ModB is generated to assess the impact of the post-translational protein modifications on the phage phenotype introduced by ModB.

The thesis aims to elucidate the intricate interplay between phage and host during infection at the single molecule and multi-omics level, thereby improving our fundamental understanding and facilitating the future customization of phages for therapeutic and biotechnological applications.

1.12. References

1 Clokie, M. R., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45 (2011). https://doi.org:10.4161/bact.1.1.14942

2 Suttle, C. A. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801-812 (2007). https://doi.org:10.1038/nrmicro1750

3 Hatfull, G. F. Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity. *J Virol* **89**, 8107-8110 (2015). https://doi.org:10.1128/JVI.01340-15

4 Dion, M. B., Oechslin, F. & Moineau, S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* **18**, 125-138 (2020). https://doi.org:10.1038/s41579-019-0311-5

5 Doore, S. M. & Fane, B. A. The microviridae: Diversity, assembly, and experimental evolution. *Virology* **491**, 45-55 (2016). https://doi.org:10.1016/j.virol.2016.01.020

6 Hatfull, G. F. & Hendrix, R. W. Bacteriophages and their genomes. *Curr Opin Virol* **1**, 298-303 (2011). https://doi.org:10.1016/j.coviro.2011.06.009

7 Simmonds, P. & Aiewsakun, P. Virus classification - where do you draw the line? *Arch Virol* **163**, 2037-2046 (2018). https://doi.org:10.1007/s00705-018-3938-z

8 Levin, B. R. & Bull, J. J. Population and evolutionary dynamics of phage therapy. *Nature Reviews Microbiology* **2**, 166-173 (2004). https://doi.org:10.1038/nrmicro822

9 Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *Fems Microbiol Rev* **38**, 916-931 (2014). https://doi.org:10.1111/1574-6976.12072

10 Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *P Natl Acad Sci USA* **96**, 2192-2197 (1999). https://doi.org:DOI 10.1073/pnas.96.5.2192

11 Haggardljungquist, E., Halling, C. & Calendar, R. DNA-Sequences of the Tail Fiber Genes of Bacteriophage P2 - Evidence for Horizontal Transfer of Tail Fiber Genes among Unrelated Bacteriophages. *J Bacteriol* **174**, 1462-1477 (1992). https://doi.org:DOI 10.1128/jb.174.5.1462-1477.1992

12 Sandmeier, H., Iida, S. & Arber, W. DNA Inversion Regions Min of Plasmid P15b and Cin of Bacteriophage-P1 - Evolution of Bacteriophage Tail Fiber Genes. *J Bacteriol* **174**, 3936-3944 (1992). https://doi.org:DOI 10.1128/jb.174.12.3936-3944.1992

13 Salmond, G. P. C. & Fineran, P. C. A century of the phage: past, present and future. *Nature Reviews Microbiology* **13**, 777-786 (2015). https://doi.org:10.1038/nrmicro3564

14 Ptashne, M. Lambda's switch: lessons from a module swap. *Curr Biol* **16**, R459-462 (2006). https://doi.org:10.1016/j.cub.2006.05.037

15 Janczuk, M., Niedziólka-Jönsson, J. & Szot-Karpinska, K. Bacteriophages in electrochemistry: A review. *J Electroanal Chem* **779**, 207-219 (2016). https://doi.org:10.1016/j.jelechem.2016.05.019

16 Crick, F. H., Brenner, S., Watstobi.Rj & Barnett, L. General Nature of Genetic Code for Proteins. *Nature* **192**, 1227-& (1961). https://doi.org:DOI 10.1038/1921227a0

17 Jacob, F. & Monod, J. Genetic Regulatory Mechanisms in Synthesis of Proteins. *J Mol Biol* **3**, 318-+ (1961). https://doi.org:Doi 10.1016/S0022-2836(61)80072-7

18 Ellis, E. L. & Delbruck, M. The Growth of Bacteriophage. *J Gen Physiol* **22**, 365-384 (1939). https://doi.org:10.1085/jgp.22.3.365

19 Luria, S. E. & Delbruck, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491-511 (1943). https://doi.org:10.1093/genetics/28.6.491

20 Zhu, B. Bacteriophage T7 DNA polymerase - sequenase. *Front Microbiol* **5**, 181 (2014). https://doi.org:10.3389/fmicb.2014.00181

21 Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009). https://doi.org:10.1126/science.1162986

22 Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* **117**, 119-128 (2015). https://doi.org:10.1016/j.biochi.2015.03.025

23 Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016). https://doi.org:10.1126/science.aad5147

24 Wilson, G. G. & Murray, N. E. Restriction and modification systems. *Annu Rev Genet* **25**, 585-627 (1991). https://doi.org:10.1146/annurev.ge.25.120191.003101

25 Pingoud, A., Wilson, G. G. & Wende, W. Type II restriction endonucleases--a historical perspective and more. *Nucleic Acids Res* **42**, 7489-7527 (2014). https://doi.org:10.1093/nar/gku447

26 Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat Rev Microbiol* **21**, 686-700 (2023). https://doi.org:10.1038/s41579-023-00934-x

27 Endersen, L. & Coffey, A. The use of bacteriophages for food safety. *Curr Opin Food Sci* **36**, 1-8 (2020). https://doi.org:10.1016/j.cofs.2020.10.006

28 Yahya, M., Hmaied, F., Jebri, S., Jofre, J. & Hamdi, M. Bacteriophages as indicators of human and animal faecal contamination in raw and treated wastewaters from Tunisia. *J Appl Microbiol* **118**, 1217-1225 (2015). https://doi.org:10.1111/jam.12774

29 Meile, S. *et al.* Engineered reporter phages for detection of Escherichia coli, Enterococcus, and Klebsiella in urine. *Nat Commun* **14**, 4336 (2023). https://doi.org:10.1038/s41467-023-39863-x

30 d'Herelle, F. Bacteriophage as a Treatment in Acute Medical and Surgical Infections. *Bull N Y Acad Med* **7**, 329-348 (1931).

31 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

32 Hinton, D. M., March-Amegadzie, R., Gerber, J. S. & Sharma, M. Bacteriophage T4 middle transcription system: T4-modified RNA polymerase; AsiA, a sigma 70 binding protein; and transcriptional activator MotA. *Methods Enzymol* **274**, 43-57 (1996). https://doi.org:10.1016/s0076-6879(96)74007-7

33 Rohrer, H., Zillig, W. & Mailhammer, R. ADP-ribosylation of DNA-dependent RNA polymerase of Escherichia coli by an NAD+: protein ADP-ribosyltransferase from bacteriophage T4. *Eur J Biochem* **60**, 227-238 (1975). https://doi.org:10.1111/j.1432-1033.1975.tb20995.x

34 Yu, F. & Mizushima, S. Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4. *J Bacteriol* **151**, 718-722 (1982). https://doi.org:10.1128/jb.151.2.718-722.1982

35 Montag, D., Hashemolhosseini, S. & Henning, U. Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. *J Mol Biol* **216**, 327-334 (1990). https://doi.org:10.1016/S0022-2836(05)80324-9

36 Franklin, J. L. & Mosig, G. Expression of the bacteriophage T4 DNA terminase genes 16 and 17 yields multiple proteins. *Gene* **177**, 179-189 (1996). https://doi.org:10.1016/0378-1119(96)00299-5

37 Barth, K. A., Powell, D., Trupin, M. & Mosig, G. Regulation of two nested proteins from gene 49 (recombination endonuclease VII) and of a lambda RexA-like protein of bacteriophage T4. *Genetics* **120**, 329-343 (1988). https://doi.org:10.1093/genetics/120.2.329

38 Zajanckauskaite, A., Malys, N. & Nivinskas, R. A rare type of overlapping genes in bacteriophage T4: gene 30.3' is completely embedded within gene 30.3 by one position downstream. *Gene* **194**, 157-162 (1997). https://doi.org:10.1016/s0378-1119(97)00127-3

39 Vaiskunaite, R., Miller, A., Davenport, L. & Mosig, G. Two new early bacteriophage T4 genes, repEA and repEB, that are important for DNA replication initiated from origin E. *J Bacteriol* **181**, 7115-7125 (1999). https://doi.org:10.1128/JB.181.22.7115-7125.1999

40 Landthaler, M. & Shub, D. A. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proc Natl Acad Sci U S A* **96**, 7005-7010 (1999). https://doi.org:10.1073/pnas.96.12.7005

41 Young, P., Ohman, M., Xu, M. Q., Shub, D. A. & Sjoberg, B. M. Intron-containing T4 bacteriophage gene sunY encodes an anaerobic ribonucleotide reductase. *J Biol Chem* **269**, 20229-20232 (1994).

42 Samatova, E., Konevega, A. L., Wills, N. M., Atkins, J. F. & Rodnina, M. V. High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide. *Nat Commun* **5**, 4459 (2014). https://doi.org:10.1038/ncomms5459

43 Kruger, D. H. & Bickle, T. A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**, 345-360 (1983). https://doi.org:10.1128/mr.47.3.345-360.1983

44 Weigele, P. & Raleigh, E. A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem Rev* **116**, 12655-12687 (2016). https://doi.org:10.1021/acs.chemrev.6b00114

45 Warren, R. A. Modified bases in bacteriophage DNAs. *Annu Rev Microbiol* **34**, 137-158 (1980). https://doi.org:10.1146/annurev.mi.34.100180.001033

46 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem J* **55**, 774-782 (1953). https://doi.org:10.1042/bj0550774

47 Lagerback, P. & Carlson, K. Amino acid residues in the GIY-YIG endonuclease II of phage T4 affecting sequence recognition and binding as well as catalysis. *J Bacteriol* **190**, 5533-5544 (2008). https://doi.org:10.1128/JB.00094-08

48 Carlson, K., Lagerback, P. & Nystrom, A. C. Bacteriophage T4 endonuclease II: concerted single-strand nicks yield double-strand cleavage. *Mol Microbiol* **52**, 1403-1411 (2004). https://doi.org:10.1111/j.1365-2958.2004.04062.x

49 Andersson, C. E., Lagerback, P. & Carlson, K. Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *J Mol Biol* **397**, 1003-1016 (2010). https://doi.org:10.1016/j.jmb.2010.01.076

50 Brody, E. N. *et al.* Old phage, new insights: two recently recognized mechanisms of transcriptional regulation in bacteriophage T4 development. *FEMS Microbiol Lett* **128**, 1-8 (1995). https://doi.org:10.1111/j.1574-6968.1995.tb07491.x

51 Wilkens, K. & Rüger, W. Characterization of Bacteriophage T4 Early Promotersin Vivowith a New Promoter Probe Vector. *Plasmid* **35**, 108-120 (1996).

52 Wilkens, K., Tiemann, B., Bazan, F. & Rüger, W. ADP-ribosylation and early transcription regulation by bacteriophage T4. *Adv Exp Med Biol* **419**, 71-82 (1997). https://doi.org:10.1007/978-1-4419-8632-0_8

53 Hinton, D. M. Transcriptional control in the prereplicative phase of T4 development. *Virol J* **7**, 289 (2010). https://doi.org:10.1186/1743-422X-7-289

54 Sommer, N., Salniene, V., Gineikiene, E., Nivinskas, R. & Rüger, W. T4 early promoter strength probed in vivo with unribosylated and ADP-ribosylated Escherichia coli RNA polymerase: a mutation analysis. *Microbiology (Reading)* **146 (Pt 10)**, 2643-2653 (2000). https://doi.org:10.1099/00221287-146-10-2643

55 Severinov, K. *et al.* A non-essential domain of Escherichia coli RNA polymerase required for the action of the termination factor Alc. *J Biol Chem* **269**, 14254-14259 (1994).

56 Westblade, L. F. *et al.* Rapid isolation and identification of bacteriophage T4-encoded modifications of Escherichia coli RNA polymerase: a generic method to study bacteriophage/host interactions. *J Proteome Res* **7**, 1244-1250 (2008). https://doi.org:10.1021/pr070451j

57 Kashlev, M., Nudler, E., Goldfarb, A., White, T. & Kutter, E. Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell* **75**, 147-154 (1993).

58 Pene, C. & Uzan, M. The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. *Mol Microbiol* **35**, 1180-1191 (2000). https://doi.org:10.1046/j.1365-2958.2000.01787.x

59 Marshall, P., Sharma, M. & Hinton, D. M. The bacteriophage T4 transcriptional activator MotA accepts various base-pair changes within its binding sequence. *J Mol Biol* **285**, 931-944 (1999). https://doi.org:10.1006/jmbi.1998.2373

60 Ouhammouch, M., Orsini, G. & Brody, E. N. The asiA gene product of bacteriophage T4 is required for middle mode RNA synthesis. *J Bacteriol* **176**, 3956-3965 (1994). https://doi.org:10.1128/jb.176.13.3956-3965.1994

61 Hinton, D. M. *et al.* Transcriptional takeover by sigma appropriation: remodelling of the sigma70 subunit of Escherichia coli RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. *Microbiology (Reading)* **151**, 1729-1740 (2005). https://doi.org:10.1099/mic.0.27972-0

62 Kassavetis, G. A., Elliott, T. & Geiduschek, E. P. Regulation of transcription in the segment of the bacteriophage T4 genome containing three of the head protein genes: plasmid-specific and phage chromosome-specific effects. *Virology* **139**, 283-294 (1984). https://doi.org:10.1016/0042-6822(84)90374-x

63 Wong, K., Kassavetis, G. A., Leonetti, J. P. & Geiduschek, E. P. Mutational and functional analysis of a segment of the sigma family bacteriophage T4 late promoter recognition protein gp55. *J Biol Chem* **278**, 7073-7080 (2003). https://doi.org:10.1074/jbc.M211447200

64 Tabib-Salazar, A., Mulvenna, N., Severinov, K., Matthews, S. J. & Wigneshweraraj, S. Xenogeneic Regulation of the Bacterial Transcription Machinery. *J Mol Biol* **431**, 4078-4092 (2019). https://doi.org:10.1016/j.jmb.2019.02.008

65 Geiduschek, E. P. & Kassavetis, G. A. Transcription of the T4 late genes. *Virol J* **7**, 288 (2010). https://doi.org:10.1186/1743-422X-7-288

66 Twist, K. A. *et al.* Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the beta-subunit flap domain of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **108**, 19961-19966 (2011). https://doi.org:10.1073/pnas.1113328108

67 Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134-141 (2004). https://doi.org:10.1016/j.virol.2004.08.001

68 Qi, D., Alawneh, A. M., Yonesaki, T. & Otsuka, Y. Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. *Genetics* **201**, 977-987 (2015). https://doi.org:10.1534/genetics.115.180364

69 Ueno, H. & Yonesaki, T. Recognition and specific degradation of bacteriophage T4 mRNAs. *Genetics* **158**, 7-17 (2001). https://doi.org:10.1093/genetics/158.1.7

70 Uzan, M. Bacteriophage T4 RegB endoribonuclease. *Methods Enzymol* **342**, 467-480 (2001). https://doi.org:10.1016/s0076-6879(01)42567-5

71 Young, E. T. *et al.* Bacteriophage T4 gene transcription studied by hybridization to cloned restriction fragments. *J Mol Biol* **138**, 423-445 (1980). https://doi.org:10.1016/s0022-2836(80)80011-8

72 Uzan, M. & Miller, E. S. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. *Virol J* **7**, 360 (2010). https://doi.org:10.1186/1743-422X-7-360

73 Marcaida, M. J., DePristo, M. A., Chandran, V., Carpousis, A. J. & Luisi, B. F. The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends Biochem Sci* **31**, 359-365 (2006). https://doi.org:10.1016/j.tibs.2006.05.005

74 Mudd, E. A., Prentki, P., Belin, D. & Krisch, H. M. Processing of Unstable Bacteriophage-T4 Gene-32 Messenger-Rnas into a Stable Species Requires Escherichia-Coli Ribonuclease-E. *Embo J* **7**, 3601-3607 (1988). https://doi.org:DOI 10.1002/j.1460-2075.1988.tb03238.x

75 Kennell, D. Processing endoribonucleases and mRNA degradation in bacteria. *J Bacteriol* **184**, 4645-4657; discussion 4665 (2002). https://doi.org:10.1128/JB.184.17.4645-4657.2002

76 Otsuka, Y., Ueno, H. & Yonesaki, T. Escherichia coli endoribonucleases involved in cleavage of bacteriophage T4 mRNAs. *J Bacteriol* **185**, 983-990 (2003). https://doi.org:10.1128/JB.185.3.983-990.2003

77 Kai, T., Selick, H. E. & Yonesaki, T. Destabilization of bacteriophage T4 mRNAs by a mutation of gene 61.5. *Genetics* **144**, 7-14 (1996). https://doi.org:10.1093/genetics/144.1.7

78 Otsuka, Y. & Yonesaki, T. A novel endoribonuclease, RNase LS, in Escherichia coli. *Genetics* **169**, 13-20 (2005). https://doi.org:10.1534/genetics.104.033290

79 Sanson, B., Hu, R. M., Troitskayadagger, E., Mathy, N. & Uzan, M. Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J Mol Biol* **297**, 1063-1074 (2000). https://doi.org:10.1006/jmbi.2000.3626

80 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009). https://doi.org:10.1038/nrg2484

81 Luke, K. *et al.* Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**, 182-191 (2002). https://doi.org:10.1006/viro.2002.1409

82 Guegler, C. K. & Laub, M. T. Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol Cell* **81**, 2361-2373 e2369 (2021). https://doi.org:10.1016/j.molcel.2021.03.027

83 Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *ISME J* **14**, 881-895 (2020). https://doi.org:10.1038/s41396-019-0580-z

84 Howard-Varona, C. *et al.* Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *Isme Journal* **12**, 1605-1618 (2018). https://doi.org:10.1038/S41396-018-0099-8

85 Bar-Yaacov, D. *et al.* RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res* **27**, 1696-1703 (2017). https://doi.org:10.1101/gr.222760.117

86 Nie, W. *et al.* A-to-I RNA editing in bacteria increases pathogenicity and tolerance to oxidative stress. *PLoS Pathog* **16**, e1008740 (2020). https://doi.org:10.1371/journal.ppat.1008740

87 Deng, X. *et al.* Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **43**, 6557-6567 (2015). https://doi.org:10.1093/nar/gkv596

88 Alseth, I., Dalhus, B. & Bjoras, M. Inosine in DNA and RNA. *Curr Opin Genet Dev* **26**, 116-123 (2014). https://doi.org:10.1016/j.gde.2014.07.008

89 Höfer, K. & Jäschke, A. Epitranscriptomics: RNA Modifications in Bacteria and Archaea. *Microbiol Spectr* **6** (2018). https://doi.org:10.1128/microbiolspec.RWR-0015-2017

90 Wang, J. *et al.* Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res* **47**, e130 (2019). https://doi.org:10.1093/nar/gkz751

91 Cahova, H., Winz, M. L., Höfer, K., Nubel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

92 Höfer, K. *et al.* Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol* **12**, 730-734 (2016). https://doi.org:10.1038/nchembio.2132

93 Zhu, T. T. *et al.* Host-mediated RNA editing in viruses. *Biol Direct* **18** (2023). https://doi.org:ARTN 1210.1186/s13062-023-00366-w

94 Volchkova, V. A., Dolnik, O., Martinez, M. J., Reynard, O. & Volchkov, V. E. RNA Editing of the GP Gene of Ebola Virus is an Important Pathogenicity Factor. *J Infect Dis* **212**, S226-S233 (2015). https://doi.org:10.1093/infdis/jiv309

95 Sherwood, A. V. *et al.* Hepatitis C virus RNA is 5'-capped with flavin adenine dinucleotide. *Nature* **619**, 811-818 (2023). https://doi.org:10.1038/s41586-023-06301-3

96 Li, H., Guo, Y., Qi, W. & Liao, M. N(6)-methyladenosine modification of viral RNA and its role during the recognition process of RIG-I-like receptors. *Front Immunol* **13**, 1031200 (2022). https://doi.org:10.3389/fimmu.2022.1031200

97 Jeffery, C. J. Current successes and remaining challenges in protein function prediction. *Front Bioinform* **3**, 1222182 (2023). https://doi.org:10.3389/fbinf.2023.1222182

98 Zhang, K. *et al.* Systemic Expression, Purification, and Initial Structural Characterization of Bacteriophage T4 Proteins Without Known Structure Homologs. *Front Microbiol* **12**, 674415 (2021). https://doi.org:10.3389/fmicb.2021.674415

99 King, J. & Laemmli, U. K. Polypeptides of the tail fibres of bacteriophage T4. *J Mol Biol* **62**, 465-477 (1971). https://doi.org:10.1016/0022-2836(71)90148-3

100 King, J. & Laemmli, U. K. Bacteriophage T4 tail assembly: structural proteins and their genetic identification. *J Mol Biol* **75**, 315-337 (1973). https://doi.org:10.1016/0022-2836(73)90024-7

101 Kutter, E. M. *et al.* Identification of bacteriophage T4 prereplicative proteins on two-dimensional polyacrylamide gels. *J Bacteriol* **176**, 1647-1654 (1994). https://doi.org:10.1128/jb.176.6.1647-1654.1994

102 Arisaka, F., Nakako, T., Takahashi, H. & Ishii, S. Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid sequence of its product. *J Virol* **62**, 1186-1193 (1988). https://doi.org:10.1128/JVI.62.4.1186-1193.1988

103 Boosman, A. Proteolysis of the major capsid protein T4 bacteriophage polyheads limited by quaternary structure. *J Biol Chem* **253**, 7981-7984 (1978).

104 Parker, M. L. *et al.* Nucleotide sequence of bacteriophage T4 gene 23 and the amino acid sequence of its product. *J Mol Biol* **180**, 399-416 (1984). https://doi.org:10.1016/0022-2836(84)90019-6

105 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970). https://doi.org:10.1038/227680a0

106 Cui, M., Cheng, C. & Zhang, L. High-throughput proteomics: a methodological mini-review. *Lab Invest* **102**, 1170-1181 (2022). https://doi.org:10.1038/s41374-022-00830-7

107 Weintraub, S. T. *et al.* Global Proteomic Profiling of Salmonella Infection by a Giant Phage. *J Virol* **93** (2019). https://doi.org:10.1128/JVI.01833-18

108 Bartel, J. *et al.* Optimized Proteomics Workflow for the Detection of Small Proteins. *J Proteome Res* **19**, 4004-4018 (2020). https://doi.org:10.1021/acs.jproteome.0c00286

109 Carr, S. *et al.* The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* **3**, 531-533 (2004). https://doi.org:10.1074/mcp.T400006-MCP200

110 Tiemann, B., Depping, R. & Rüger, W. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. *Gene Expression The Journal of Liver Research* **8**, 187-196 (1999).

111 Cohen, M. S. & Chang, P. J. N. c. b. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nat Chem Biol* **14**, 236-243 (2018).

112 Corda, D. & Di Girolamo, M. J. T. E. j. Functional aspects of protein mono-ADP-ribosylation. *EMBO* **22**, 1953-1958 (2003).

113 Mathews, C. K., Kutter, E. M., Mosig, G., Berget, P. B. Bacteriophage T4, Chapters 1 and 4. (1983).

114 Goff, C. G. Chemical structure of a modification of the Escherichia coli ribonucleic acid polymerase α polypeptides induced by bacteriophage T4 infection. *J Biol Chem* **249**, 6181-6190 (1974).

115 Rohrer, H., Zillig, W. & Mailhammer, R. ADP-Ribosylation of DNA-Dependent RNA Polymerase of Escherichia coli by an NAD+: Protein ADP-ribosyltransferase from Bacteriophage T4. *Eur J Biochem* **60**, 227-238 (1975).

116 Sommer, N., Salniene, V., Gineikiene, E., Nivinskas, R. & Rüger, W. T4 early promoter strength probed in vivo with unribosylated and ADP-ribosylated Escherichia coli RNA polymerase: a mutation analysis. *Microbiology* **146**, 2643-2653 (2000).

117 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004).

118 Tsuge, H. *et al.* Structural basis of actin recognition and arginine ADP-ribosylation by Clostridium perfringens iota-toxin. *Proc Natl Acad Sci U S A* **105**, 7399-7404 (2008). https://doi.org:10.1073/pnas.0801215105

119 Depping, R., Lohaus, C., Meyer, H. E. & Rüger, W. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochemical biophysical research communications* **335**, 1217-1223 (2005).

120 Furano, A. V. Content of elongation factor Tu in Escherichia coli. *Proc Natl Acad Sci U S A* **72**, 4780-4784 (1975). https://doi.org:10.1073/pnas.72.12.4780

121 Drake, J. W. Ultraviolet mutagenesis in bacteriophage T-4. I. Irradiation of extracellular phage particles. *J Bacteriol* **91**, 1775-1780 (1966). https://doi.org:10.1128/jb.91.5.1775-1780.1966

122 Drake, J. W. & Baltz, R. H. The biochemistry of mutagenesis. *Annu Rev Biochem* **45**, 11-37 (1976). https://doi.org:10.1146/annurev.bi.45.070176.000303

123 Mahler, M., Costa, A. R., van Beljouw, S. P. B., Fineran, P. C. & Brouns, S. J. J. Approaches for bacteriophage genome engineering. *Trends Biotechnol* **41**, 669-685 (2023). https://doi.org:10.1016/j.tibtech.2022.08.008

124 Chen, Y. *et al.* Genetic Engineering of Bacteriophages Against Infectious Diseases. *Front Microbiol* **10**, 954 (2019). https://doi.org:10.3389/fmicb.2019.00954

125 Pires, D. P., Cleto, S., Sillankorva, S., Azeredo, J. & Lu, T. K. Genetically Engineered Phages: a Review of Advances over the Last Decade. *Microbiol Mol Biol Rev* **80**, 523-543 (2016). https://doi.org:10.1128/MMBR.00069-15

126 Marinelli, L. J., Hatfull, G. F. & Piuri, M. Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage* **2**, 5-14 (2012). https://doi.org:10.4161/bact.18778

127 Dong, J. *et al.* Engineering T4 Bacteriophage for In Vivo Display by Type V CRISPR-Cas Genome Editing. *ACS Synth Biol* **10**, 2639-2648 (2021). https://doi.org:10.1021/acssynbio.1c00251

128 Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synth Biol* **6**, 1952-1961 (2017). https://doi.org:10.1021/acssynbio.7b00179

129 Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol Syst Biol* **1**, 2005 0018 (2005). https://doi.org:10.1038/msb4100025

130 Smith, H. O., Hutchison, C. A., 3rd, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci U S A* **100**, 15440-15445 (2003). https://doi.org:10.1073/pnas.2237126100

Chapter II

Integrated omics reveal time-resolved insights into T4 phage infection of *E. coli* on proteome and transcriptome levels

Authors:
Wolfram-Schauerte, M.*, **Pozhydaieva, N.***, Viering, M., Glatter, T., & Höfer, K.
*contributed equally

Published in:
Viruses, 2022, 14(11), 2502

Author contributions:

K.H., M.W.-S. and N.P. wrote the first draft of the manuscript and all authors contributed to reviewing, editing, and providing additional text for the manuscript. N.P. and M.V. developed and established the proteomics sample preparation. T.G. performed the LC-MS/MS analysis and initial data analysis with MaxQuant and helped to optimize the proteomics pipeline. N.P. analyzed and integrated the proteomics data. M.W.-S. performed NGS data analysis, established data analysis in R, integrated transcriptomics and proteomics data, and wrote the POTATO4 application. K.H. supervised the work. All authors have read and agreed to the published version of the manuscript.

2. Integrated omics reveal time-resolved insights into T4 phage infection of *E. coli* on proteome and transcriptome levels

2.1. Abstract

Bacteriophages are highly abundant viruses of bacteria. The major role of phages in shaping bacterial communities and their emerging medical potential as antibacterial agents has triggered a rebirth of phage research. To understand the molecular mechanisms by which phages hijack their host, omics technologies can provide novel insights into the organization of transcriptional and translational events occurring during the infection process. In this study, we apply transcriptomics and proteomics to characterize the temporal patterns of transcription and protein synthesis during the T4 phage infection of *E. coli*. We investigated the stability of *E. coli*-originated transcripts and proteins in the course of infection, identifying the degradation of *E. coli* transcripts and the preservation of the host proteome. Moreover, the correlation between the phage transcriptome and proteome reveals specific T4 phage mRNAs and proteins that are temporally decoupled, suggesting post-transcriptional and translational regulation mechanisms. This study provides the first comprehensive insights into the molecular takeover of *E. coli* by bacteriophage T4. This data set represents a valuable resource for future studies seeking to study molecular and regulatory events during infection. We created a user-friendly online tool, POTATO4, which is available to the scientific community and allows access to gene expression patterns for *E. coli* and T4 genes.

2.2. Introduction

Bacteriophages (phages) are highly abundant viruses that specifically interact with and infect bacteria. They are widespread in abundance and contribute to the largest proportion of biomass on Earth[1], thereby shaping bacterial community ecology[2]. The emergence of multi-antibiotic-resistant bacterial pathogens has led to a renaissance of phage research due to the potential application of phage-based therapies for treating bacterial infections[3,4]. Thus, the mechanisms of how phages specifically hijack their host's gene expression machinery are of enormous current interest. A variety of model phages are subjects for studying these systems, such as bacteriophage T4 (T4 phage). The T4 phage—a member of the *Straboviridae* family—belongs to the T-even phages, infecting the prokaryotic model organism *Escherichia coli*[5]. Early studies of the T4 phage have made valuable contributions to molecular biology tools, such as T4 polynucleotide kinase, T4 DNA, and RNA ligases[6], in addition to the discovery of fundamental biological processes[5]. Among others, these include the discovery of DNA as the genetic code, messenger RNA, and understanding the role of mutations or DNA restriction and modification[7,8].

T4 phage infection is temporally fine-tuned, highly efficient, and terminates with lysis of the host from 25 to 30 min[5]. The T4 phage possesses a double-stranded DNA (dsDNA) genome of approximately 169 kb in size, encoding 288 genes[5].

Notably, the T4 phage does not possess its own gene expression machinery and thus takes over the one of its host—*E. coli*. Consequently, the T4 phage needs to reprogram the host cell to promote the expression of its genes. To promote the expression of viral genes, the T4 phage manipulates the host's gene expression machinery using its proteins co-injected upon or expressed during infection, termed host acquisition factors (HAFs)[5,9]. These include ADP-ribosyltransferases which post-translationally modify host proteins[10-14], nucleases degrading host DNA and RNA[15,16], or transcription factors, which mediate the expression of phage genes from distinct promoters[5,17,18].

In combination, the action of T4-encoded HAFs leads to a temporally tightly controlled gene expression divided into an early, middle, and late infection phase[5,19]. Generally, it is accepted that *E. coli* mRNA is rapidly degraded and that host transcription is shut-off upon T4 phage infection[9,20,21]. Principally, phages are conceived to quickly takeover and dominate the transcriptome by expressing their genes in an infection-phase-specific manner and remodeling the host transcriptome to adapt to their specific needs[22-26]. Each infection phase is characterized by the expression of distinct sets of genes. Several studies have set out to characterize gene expression on a transcriptional level during T4 phage infection. A microarray study has focused on the time-resolved transcriptome of the T4 phage[19]. Biochemical studies have characterized specific host and phage mRNAs[9]. RNA sequencing was applied to describe the functional role of a toxin–antitoxin system during T4 infection[27]. Despite the characterization of specific gene sets, comprehensive transcriptomic analysis to investigate *E. coli* and T4 phage transcriptome is still lacking.

On the proteome level, the first attempts to elucidate the T4 phage proteome were performed via two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of phage proteins[28,29]. Applying these methods to identify small phage proteins or to characterize complex samples is time-consuming and challenging.

Recently, liquid chromatography–mass spectrometry (LC-MS)-based proteomics studies provided insights into the complex host–phage proteome (dual-proteome) rewiring of *Pseudoalteromonas* and *Bacteroidetes* and their specific phages in a time-dependent manner[30,31]. These studies shed light on essential infection regulation mechanisms from both host and virus perspectives. In addition, emerging omics technologies, such as GRAD-Seq, are powerful tools to study RNA–protein interactions during the viral predation of the bacterial cell[32].

In general, applying multi-omics techniques in a time-series context allows us to track the complex patterns of cellular information flow and to infer the underlying regulatory cascades[33]. This appears especially valuable in the context of phage–host interactions.

Using the power of transcriptomics and proteomics, we define the dual-phage and -host transcriptome and proteome of T4 phage infection in a time-resolved manner for the first time. Thereby, we characterized the temporal patterns of transcription and protein synthesis and their interconnection throughout T4 phage infection, not only for the T4 phage but also for the host *E. coli*. We show that most host transcripts, including tRNAs, are rapidly degraded upon infection. In contrast, four non-coding transcripts were found to be rather stable throughout infection. To the contrary, T4 phage genes are transcribed in an infection-phase-specific manner. On the proteome level, host proteins remain relatively stable, whereas the onset of phage protein synthesis occurs in distinct infection phases and corresponds to the functional protein classes needed in the respective infection phase. By comparing the time-transcribed early during infection but whose proteins are synthesized in the late infection phase. This indicates the presence of post-transcriptional regulatory mechanisms that control the translation of early phage mRNAs only in the late phase of infection. This work describes the first combinatorial and comprehensive study of the dual-transcriptome and -proteome of T4 phage/*E. coli* infection in a time-resolved manner. We highly appreciate all studies from the last few decades that have shaped our current understanding of T4 phage infection. Our data demonstrate that high-throughput technologies can help overcome laborious reductionist biochemical studies limited to distinct transcripts and proteins by studying a diverse population thereof at once. Moreover, by revealing the temporal coupling of RNA and protein synthesis during T4 phage infection, these data sets represent a valuable resource for future studies seeking to investigate molecular and regulatory events during infection. To enable broad community access to these data sets, we designed a web application to retrieve gene expression data for phage and/or host genes of interest called PrOteome TrAnscripTOme 4 (POTATO4).

2.3. Results and discussion

Time-Resolved Dual-RNA-Seq of T4 Phage Infection

T4 phage infection can be divided into three temporal phases: early (0–5 min), middle (5–10 min), and late (10–20 min), which are each characterized by the expression (transcription) of distinct sets of T4 phage genes (Figure 1a)[5,19]. In order to monitor the transcriptomic changes within these phases, total RNA was isolated from uninfected *E. coli* (t0) and at 1 (t1), 4 (t4), 7 (t7), and 20 min (t20) post-T4 phage infection. For the time course of infection, we observed relatively similar

yields of total RNA at all time points (Supplementary Figure S2). In order to monitor the time-resolved dual-transcriptome of phage and host during infection, rRNA-depleted total RNAs were subjected to Illumina RNA-Seq. PCA revealed the close clustering of replicates from the same time points, reflecting similar gene expression profiles except for t4, which had one outlier replicate (t4 R1) (Supplementary Figure S3).

To characterize the regulation of transcription during the course of infection, we tracked the expression of all remaining non-rRNA host genes as well as all annotated T4 phage genes (Figure 1). Following the fractions of reads per genome over the time course of infection, it becomes evident that the T4 phage quickly starts to dominate the transcriptome during the first seven minutes of infection (73.8% of T4 phage reads at t7) (Figure 1b). At the end of infection, the overall fraction of phage reads amounts to 82.8% (Figure 1b), similar to what was reported by Laub and colleagues [27]. In parallel, *E. coli* transcripts rapidly decline in abundance, which has been reported in several transcriptome studies[9,16,34], amounting to less than 20% among all reads at t20 (Figure 1b). This fast and nearly complete takeover of the non-rRNA transcriptome by T4 phage is a phenomenon commonly observed in other phage–host interactions, and is conceived to liberate nucleotide building blocks for phage transcription, DNA replication, or ATP metabolism[22,23,26,35].

E. coli Transcript Degradation Is Initiated during the First 4 min of T4 Phage Infection

During T4 phage infection, *E. coli* gene expression is globally shut off, whilst existing host transcripts are predominantly degraded[36]. This conception is mainly based on valuable studies describing the degradation of small subsets of host transcripts using elaborate rifampicin and Northern blot assays[9]. Thus, our dual-transcriptome approach provides the first comprehensive insights into all *E. coli* transcripts during T4 phage infection based on a single experiment.

Initially, we tracked the levels of all *E. coli* transcripts over the time course of infection (Figure 1c). For the vast majority of *E. coli* genes, transcript levels remain stable during 1 min post-infection – regardless of whether these transcripts are highly abundant or not. Subsequently, the levels of most transcripts drop immensely at 4 min post-infection and reach their minimum at the end of infection (t20). Only for a smaller set of genes we observed a slower decrease in transcript levels where the drop at t4 is less intense. This overall decline in host transcript levels is in good agreement with the decline of host mRNA read fractions observed throughout Pseudomonas phage LUZ19 and Acinetobacter baumannii phage *phiAbp1* infections[22,23]. During T4 phage infection, we observed an immense degree of host transcript degradation (75% reduction in host reads) already 7 min post-infection, representing the efficiency of phage-induced host

transcriptional takeover (Figure 1b,c, Supplementary Figure S4). Exemplarily, the mRNAs transcribed from the *lpp* and *ompA* genes are quite stable mRNAs in *E. coli*, with half-lives of 31 and 30 min, respectively[9]. Using rifampicin assays, Ueno and Yonesaki reported their rapid destabilization during T4 phage infection, reducing their half- lives to 2.3 and 2.5 min, respectively[9]. In accordance with these studies, our RNA-Seq data reveal a decline of transcript levels by approximately 50% for both lpp and ompA mRNA within the first 4 min of infection (Figure 1d). Interestingly, host-RNA degradation concerns not only mRNAs but also non-coding RNAs and transfer RNAs (tRNAs). For *E. coli* tRNAs, we detected steady and even increasing transcript levels within the first minute of infection, followed by a steep decrease in their abundances toward the end of infection (Supplementary Figure S5). This indicates predominant host tRNA decay during T4 phage infection, which has been described for a T4-like vibriophage so far[37]. Yang et al. speculate that early phage genes are translated using the host tRNA pool, which is successively degraded during infection followed by the transcription of phage tRNAs. These contribute to late phage mRNA translation, which prefers the phage tRNA code. Similar to the T4-like vibriophage, T4 might use host tRNAs for the initial translation of early mRNAs and subsequently as a nucleotide resource.

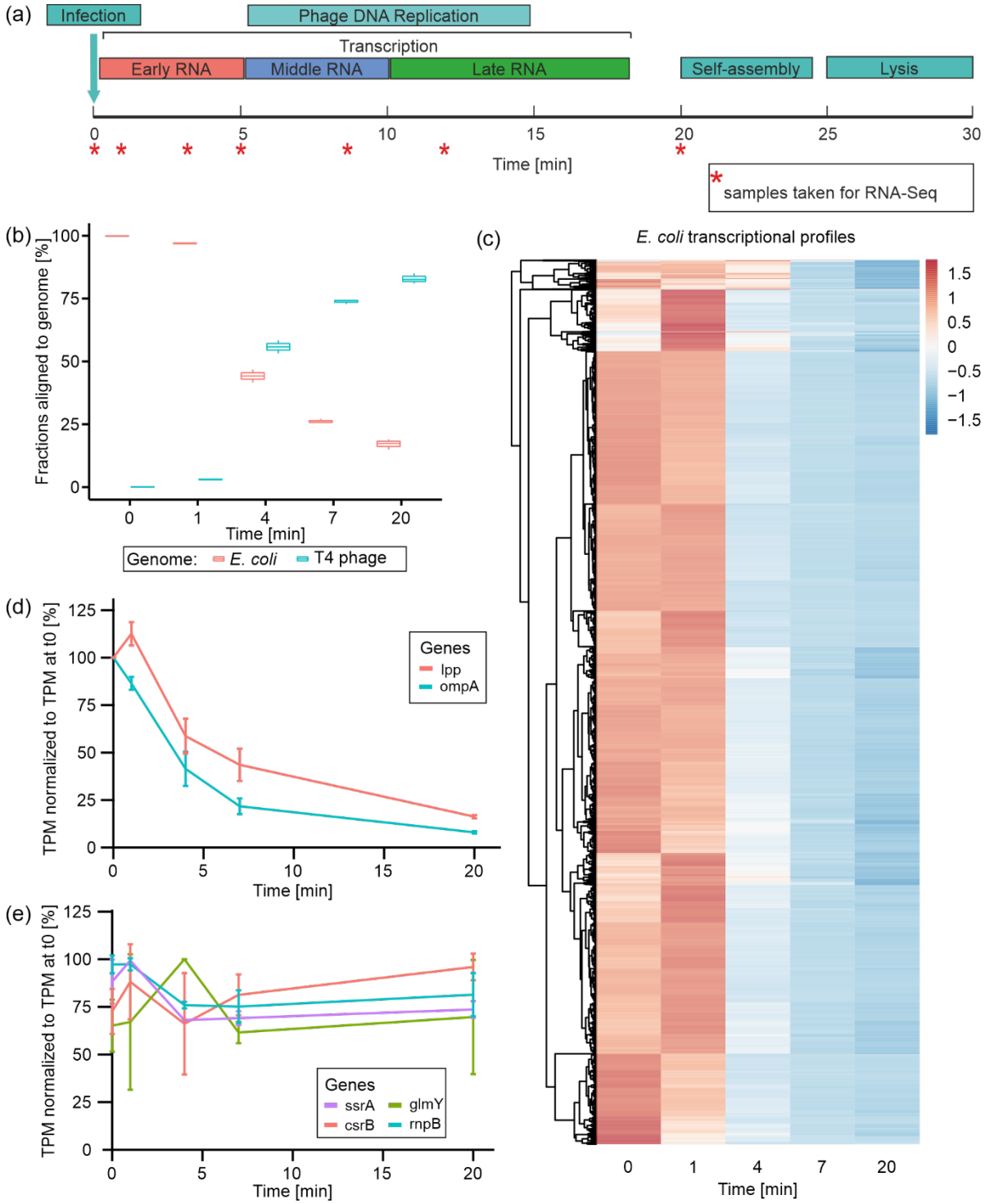


Figure 1. Time-resolved dual-RNA-Seq of T4 phage infection and the fate of host (*E. coli*) transcripts. (**a**) Schematic illustration of infection-phase-specific gene expression during T4 phage infection and indicated time points at which samples for dual-RNA-Seq have been taken. (**b**) Fractions of reads aligned to either *E. coli* or T4 phage genome over the time course of T4 phage infection calculated on the basis of TPM-normalized reads. (**c**) Heatmap of all *E. coli* genes' TPM values normalized by z-score over the time

course of infection. Genes are clustered according to expression profiles. (**d**) Plots of lpp and ompA mRNA levels over the time course of infection based on mean TPM values normalized to TPM value in uninfected *E. coli* (t0) [= 100%] for each RNA. (**e**) Plots of transcript levels of four comparably stable *E. coli* transcripts over the time course of infection based on mean TPM values, which were normalized to TPM value in uninfected *E. coli* (t0) [= 100%] for each RNA.

Despite observing massive host-RNA degradation, we questioned whether some *E. coli* transcripts might be comparably stable throughout infection. Therefore, we selected host genes with a mean expression accounting for at least 10 TPM and with at least 80% TPM at t20 compared with t0. Using these criteria, we identified four comparably stable transcripts (Figure 1e). Equal read distributions at these genes over all time points indicate that the determined TPM levels are not derived from degradation fragments accumulated during infection but rather from similar transcripts as detected before infection (Supplementary Figure S6). Among those stable host transcripts, we identified the highly abundant (mean of 67,858 TPM) transfer messenger RNA (tmRNA) SsrA, which was validated by Northern blotting (Supplementary Figure S7a). SsrA plays an important role in ribosome rescue and protein degradation[38,39]. Moreover, functional SsrA is required to induce the prophage of bacteriophage Mu and may act as a sensor for prophage activation[40]. This tmRNA has been reported as a highly stable RNA with a half-life of 89 min[41], which may explain its observed stability even during T4 phage infection. However, *E. coli* tRNAs that have similar RNA stability characteristics to SsrA are specifically degraded during T4 phage infection (Supplementary Figure S5a). One may speculate that SsrA is constitutively required during infection to keep the maximal amount of ribosomes available for the translation of T4 phage mRNAs, finally enabling the fast and efficient infection process.
Furthermore, the non-coding Rnase P RNA encoded by the *rnpB* gene was found to be relatively stable during infection. This RNA serves as a catalytic center in Rnase P[42], which plays a role in the maturation of tRNA by trimming the 5′-ends of tRNA precursors[43]. Potentially, this catalytic RNA may play a role in phage tRNA processing. Furthermore, GlmY (from *glmY* gene) and CsrB (from *csrB* gene) were identified as stable host transcripts, which are small regulatory RNAs (sRNAs) that affect gene expression by stabilizing or destabilizing target mRNAs through RNA–RNA and RNA–protein interactions, respectively (Figure 1e, Supplementary Figure S6b,c)[44,45]. Surprisingly, both sRNAs are usually unstable in exponentially growing *E. coli*, with half-lives of around 1.5 min[44,45]. As CsrB sequesters CsrA, a global activator of glycolysis (Sabnis et al., 1995), it stands to reason that the host may downregulate glycolysis to inhibit phage replication by temporally stabilizing CsrB. Additionally, the role of GlmY stabilization remains elusive. Its function

is linked to a susceptibility to cell envelope stress[46], which could be a strategy employed by the host to strengthen the cell envelope and protect against infection. However, due to predominant host mRNA degradation, the potential actions of these sRNAs may be aborted as a consequence of the lack of target RNAs.

In summary, apart from a few, surprisingly yet unreported stable RNAs during T4 phage infection, we observed global host transcript degradation.

As we also observed that a fraction of *E. coli* genes exhibit higher TPM values at t1 compared with t0 (Figure 1c), we speculated that these genes might be—temporarily—significantly upregulated as a response to the invading phage during the early phase of infection. Using DESeq2, we detected 1,050 genes differentially expressed at t1 compared with t0 (log2 fold change > 0 or < 0; adjusted *p*-value ≤ 0.05) (Supplementary Figure S8a, Supplementary Table S1). A total of 505 of these genes were upregulated, predominantly fulfilling functions in transcription, energy production, and conversion as well as carbohydrate transport and metabolism (Supplementary Figure S8b). We also observed the upregulation of the two host genes constituting the mcrBC system, which is known to counteract cytosine hydroxymethylation of T4 phage DNA as a defense mechanism[47]. Other phage defense systems present in this host's genome, including the *mazF*, *58aze* [48] and *lit*[49] genes, which were not differentially expressed. The set of 545 downregulated host genes is dominated by functions in translation and ribosome biogenesis, amino acid transport, and metabolism, as well as energy production and conversion (Supplementary Figure S8b). This initial gene regulatory alteration could resemble a host response to infection, a phage-induced change, or a combination thereof. Considering gene expression in other phages, one may speculate that these changes may originate from initial host cell reprogramming by the phage in order to create an optimal environment for phage infection[22-24,26].

In conclusion, we suggest that the host may initially try to adapt to T4 phage infection by gene expression changes, which are quickly interfered with by the phage that rapidly initiates the degradation of the vast majority of host transcripts.

T4 Phage Transcription Is Actively Controlled in an Infection-Phase-Dependent Manner

Apart from host transcriptional alterations, we obtained a time-resolved picture of T4 phage gene expression (Figure 2). We observed distinct gene expression patterns for different sets of T4 phage genes (Figure 2a). A large group of T4 phage genes (109 genes) is most strongly expressed at the end of infection, whereas the maximal transcript levels of similarly sized sets of genes are detected at t4 (62 genes) or t7 (103 genes) in Figure 2c. Furthermore, the expression

onsets, the initiation of degradation, and the decline of transcript levels vary highly among phage transcripts. Altogether, these findings vividly demonstrate the different classes of T4 phage genes, which have previously been reported[5,19]. Based on an already existing criterion for the classification of T4 phage genes[19], we defined criteria suiting our choice of time points for dual-RNA-Seq. Therefore, we classified a T4 phage gene based on its onset of expression, which resembles the period of time during which TPM values for a distinct gene are below 10% of its maximal TPM value. Accordingly, the expression of early genes starts during the first four minutes of infection (t0–t4), followed by middle genes (t4–t7) and late genes (t7–t20) (Figure 2b).

Based on this classification, most early genes show highest expression levels at t4 or t7, which decline towards t20 (Figure 2b). In total, 215 early genes were classified, which are predominantly associated with host cell reprogramming, DNA and RNA degradation, and some metabolic processes, as represented by gene functions and gene ontology (Supplementary Table S2). Additionally, the genes for the eight T4 phage-encoded tRNAs were classified as early genes. Intriguingly, this is the first characterization of all eight T4 tRNAs, as a previous microarray study had only focused on tRNA 2, 3, and 4 [19]. It appears that the expression of these eight tRNAs is initiated early during infection and continuously increases during the course of infection (Supplementary Figure S5b).

Middle and late genes are maximally expressed at t20 and differ by the onset of expression (either just after t4 or t7, respectively). We classified 21 middle and 38 late T4 phage genes, which overall encode structural phage proteins, viral release factors, and proteins associated with DNA replication. Overall, middle and late genes mediate processes that are important at the end of infection prior to phage particle release from the host cell[5,50] (Supplementary Table S2). Moreover, we also classified the small regulatory RNAs RNAC and RNAD – RNAs of yet unknown functions – as late RNAs, as reported earlier[19], and validated them by our Northern blot assay (Supplementary Figure S7b).

In summary, we detected the well-established infection-phase-specific T4 phage gene expression that progresses from host cell reprogramming over DNA replication to phage assembly and host cell lysis, which is a common feature shared by other transcriptionally characterized phages[19,22,23].

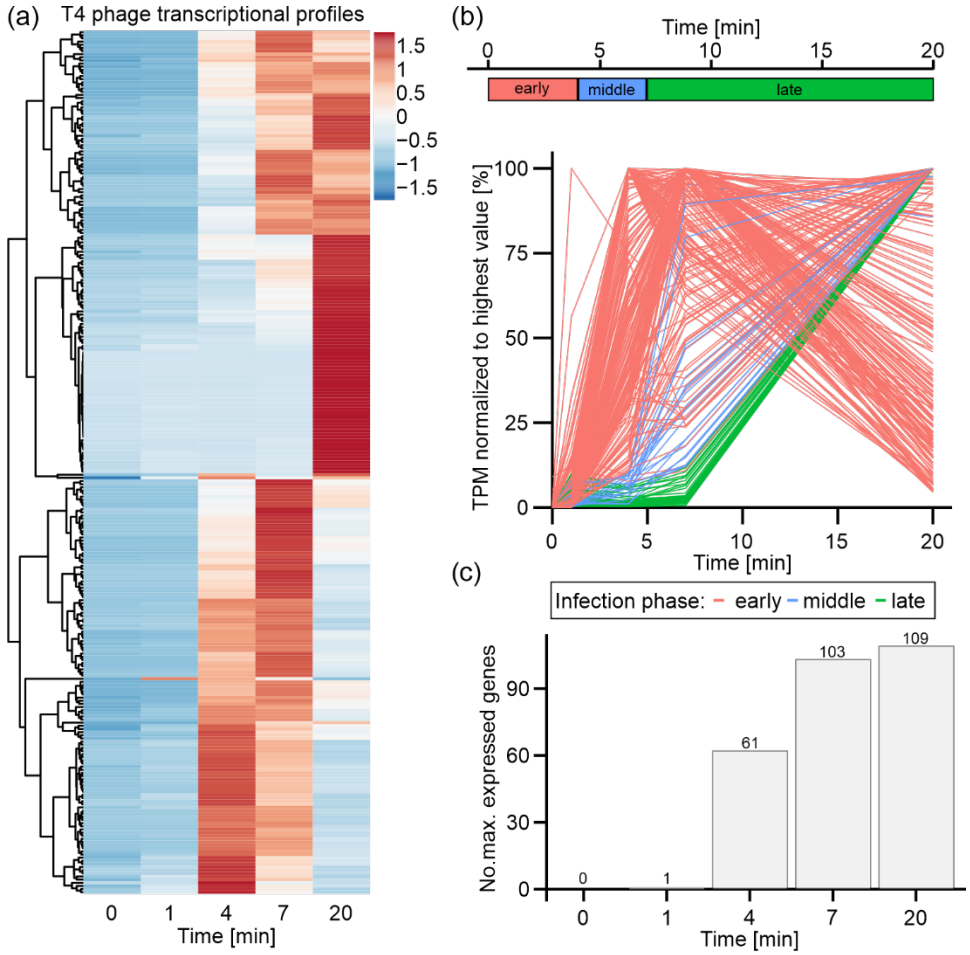


Figure 2. Expression and classification of T4 phage genes during infection of *E. coli*. T4 phage genes with mean TPM value below 1 were excluded from all analyses. (**a**) Heatmap of all T4 phage genes over time course of infection based on z-score-normalized TPM values. (**b**) Criteria for the classification of T4 phage genes: early (red), middle (blue), and late (green) based on the time frame of the onset of expression and maximal expression values (upper panel). Plots of all T4 phage genes over time course of infection based on mean TPM values normalized to the highest mean TPM value of each gene (lower panel). Genes colored by classification based on criteria depicted in upper panel. (**c**) Quantification of maximum expression of T4 phage genes per time point of infection.

Time-Resolved Dual-Proteome of T4 Phage Infection

The dynamics of the transcriptome usually correlate with the changes in the composition of the proteome. To systematically track the time course of changes in the quantity of proteins during the T4 phage infection of *E. coli*, we set out to apply proteomics. For this purpose, the total proteome was isolated from uninfected *E. coli* (t0) and 1 (t1), 3 (t3), 5 (t5), 8 (t8), 12 (t12), 20 (t20), and 30 (t30) min post-infection with T4 phage in biological triplicates (R1-R3) (Figure 3a). These time points cover the same infection phases as the ones analyzed in the transcriptomics

experiment (t0–t20) and additionally include time point t30. The rationale of the latter is to capture the maximal possible number of T4 phage proteins via proteomics.

Our proteomics workflow yielded a label-free quantified data set, allowing for the identification of 2,572 proteins in total over the entire time course of infection. A total of 2,326 proteins were assigned to *E. coli* and 246 to T4 phage. This results in 60% coverage of the known *E. coli* proteins and 85% of the annotated T4 phage proteins (Figure 3b,c) [5,51]. PCA and Pearson correlation revealed close clustering within biological replicates of the same time point, indicating the high consistency of the proteomics data (Supplementary Figures S9a, S10, and S11).

In order to determine the abundance of viral proteins during infection, we calculated the fractions of the LFQs contributed by T4 phage and *E. coli* over the time course of infection (Supplementary Figure S9b). During the infection, the LFQs are predominantly contributed by *E. coli* proteins. The fraction of the signals contributed by viral proteins increases throughout the course of infection, reaching its maximum of 14% at 30 min post-infection.

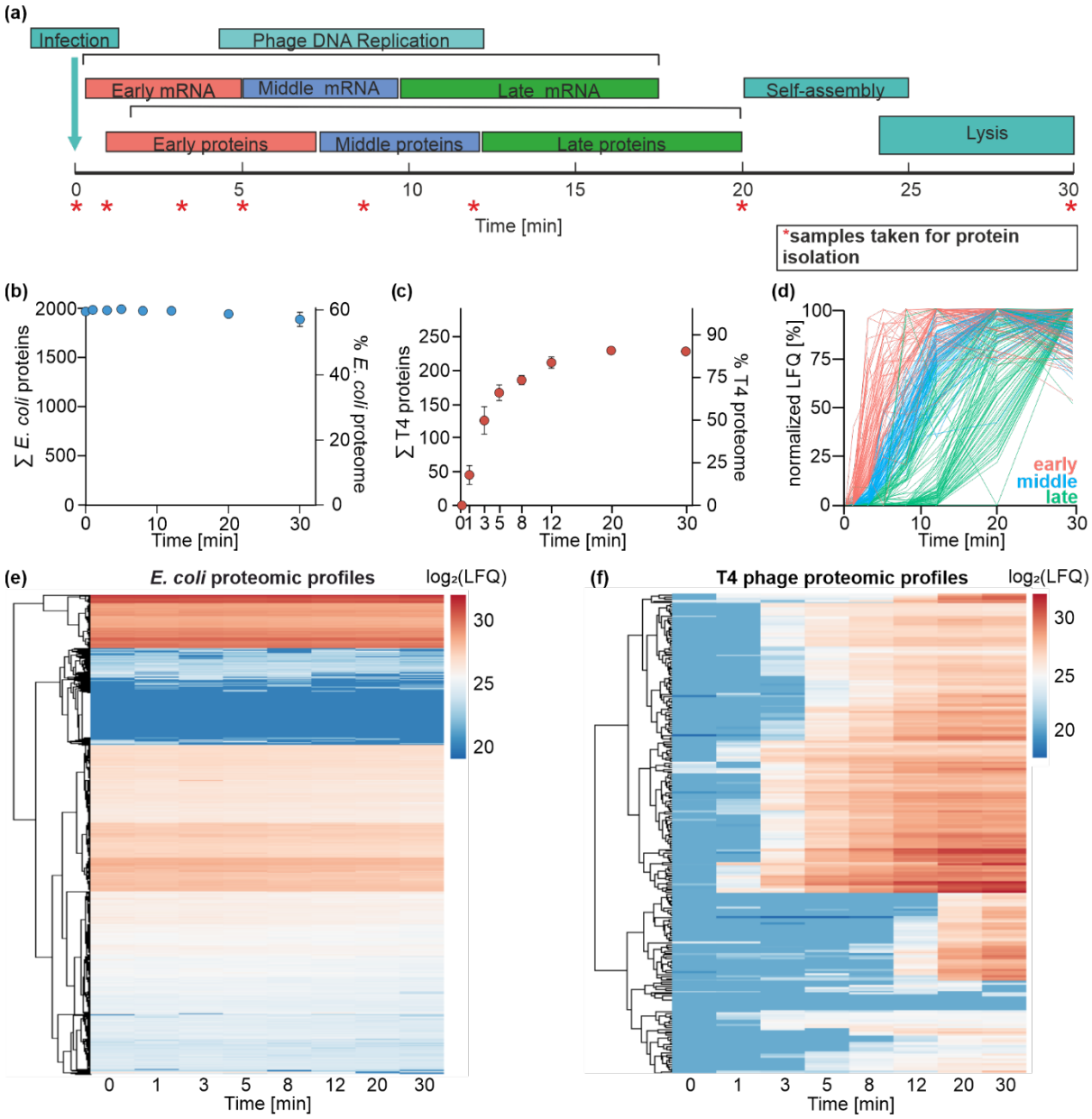


Figure 3. Time-resolved dual-proteome of T4 phage infection of *E. coli*. (**a**) Schematic illustration of infection-phase-specific gene expression during T4 phage infection. The time points at which samples were collected for the dual-proteome are indicated with an asterisk. (**b**) Monitoring of *E. coli* proteins over the time course of infection. The left *y*-axis identifies the absolute number of detected proteins, and the right *y*-axis depicts the coverage of the annotated *E. coli* proteome (n = 3). *E. coli* proteome is not significantly altered throughout the infection process. (**c**) Monitoring of T4 phage proteins over the time course of infection. The left *y*-axis identifies the absolute number of detected proteins, and the right *y*-axis depicts the coverage of the annotated T4 phage proteome (n = 3). The diversity of the T4 phage proteome increases throughout the infection.

E. coli Proteome Remains Stable during T4 Phage Infection

Transcriptome analysis has shown that *E. coli* transcripts are predominantly degraded in response to T4 phage infection. To analyze the host's response to phage infection on the level of the proteome, we examined the abundance of individual *E. coli* proteins.

In contrast to *E. coli* transcripts, we did not observe significant changes in *E. coli* protein abundance, indicating their stability throughout infection (Figure 3e), which was additionally confirmed by an SDS-PAGE analysis of the total proteome throughout infection (Supplementary Figure S12).

The first attempts to analyze the stability of host proteins during the infection with T4 phage were performed by Simon and Tomczak in 1978[52]. They reported that the maintenance of protein stability is a specific feature of T4 phage infection. Other *E. coli* phages, such as T7 and T5, utilize their own RNA polymerases for viral gene expression; however, T4 phage uses the *E. coli* RNA polymerase for this purpose. This dependence of T4 phage on host proteins might result in the preservation of the host proteome during the infection, which can be observed in our data set[53-55]. The hijacking of crucial *E. coli* protein complexes, such as ribosomes or the RNA polymerase, allows T4 to start with the rapid production of its transcripts and proteins and to lyse its host in up to 30 min.

T4 Phage Protein Synthesis Is Temporally and Functionally Regulated

In contrast to the *E. coli* proteome, we detected a highly dynamic T4 phage proteome throughout infection (Figure 3c,f). Viral proteins were observed already 1 min post-infection, and the number of detected T4 phage proteins increased exponentially until t12. The maximal number of identified T4 phage proteins was reached already at 20 min post-infection and remained stable until t30, confirming the detection of the maximal number of phage proteins. We obtained a coverage of up to 85% of the T4 phage genome. Proteins that were not identified in our proteomics study were classified as early transcripts in our transcriptome data set (Supplementary Table S2) and belong to a class of uncharacterized/hypothetical proteins (e.g., ProtID P13322, ProtID P13322). The lack of detection for some proteins might be due to their low expression levels during the early phase of infection or their small sizes (e.g., ProtID P39249 (5 kDa)), which limit their detection via LC-MS.

Our dual-proteome data set revealed the time-resolved onset of viral protein synthesis, confirming that T4 phage protein synthesis is a temporally highly regulated process (Figure 3e). A similar temporal regulation of protein biosynthesis has been observed in other studies characterizing viral proteomes, e.g., for marine phages[30,31].

In order to distinguish between temporal T4 phage protein classes (early (t0–t5), middle (t5–t8), and late (t8–t30)), we applied the same criteria as used for the classification of T4 phage transcripts. Briefly, we assigned T4 phage proteins to temporal groups based on the onset of protein detection (>10% of the maximal LFQ value of a specific protein). Based on this classification, we identified 61 early, 79 middle, and 105 late proteins (Figure 3d). We speculated that the temporal appearance of proteins could be linked to their functions, which we already observed on a transcriptome level.

Among 61 assigned early T4 phage proteins, T4 phage-encoded HAFs and regulatory proteins were identified, including ModA, ModB, MotB, and Dmd[10,18,56,57]. This observation is consistent with the concept that early infection phases are dominated by host cell reprogramming/adaptation processes [19,22,26] (Supplementary Table S2).

Furthermore, host-defense mechanisms are activated during the early phase of T4 phage infection. These include the inactivation of the host-derived proteases to prevent the degradation of the viral proteome[58]. In the genome of *E. coli,* more than 60 proteases and peptidases are encoded. In our data set, we identified 32 proteases and peptidases (Supplementary Table S3) expressed throughout phage infection. Most of them are involved in protein maturation and the cleavage of protein signal sequences and are described to be not of particular interest for phage infection[59,60]. However, two detected proteases were reported as crucial for infection: Lon protease and its predicted and uncharacterized homolog, LonH. In *E. coli*, Lon is responsible for cellular homeostasis, protein quality control, and metabolic regulation; however, it is also responsible for the selective degradation of short-lived regulatory proteins and abnormal proteins, such as the proteins of bacteriophages[61,62]. Our proteomics data reveal that Lon and LonH are consistently abundant during infection (Supplementary Table S3).

Nevertheless, T4 phage escapes Lon and potentially LonH activity by the expression of the Pin protein. The T4 phage Pin protein is described as a specific inhibitor of Lon[63]. The interaction of Pin and Lon leads to the complete inhibition of the degradation of T4 phage proteins[52]. Our data show that protease inhibitor Pin can be detected 3 min post-infection and that Pin appears to be 6.5-fold more abundant than Lon based on average LFQ values throughout infection (Supplementary Table S3). This provides novel insights into the regulation of host proteases by the T4 phage. Possibly by producing high amounts of anti-host factors, such as Pin, the host's phage defense by proteases is prevented.

A few T4 phage proteins involved in the metabolism of nucleic acids are also synthesized in the early phase of infection, such as the nucleases RegB and MobB, followed by other nucleases appearing at subsequent time points of infection (Supplementary Table S3). Nucleases are

involved in the degradation of the host transcripts, which can be confirmed by our transcriptomics data set (Figure 1c). Their activity leads to the generation of building blocks for the synthesis of viral DNA and RNA, restricting *E. coli* gene expression at the same time, as described above.

Besides HAFs and regulatory proteins, most early proteins (~70%) belong to a class of uncharacterized/hypothetical proteins. These uncharacterized proteins are primarily encoded in intergenic regions of the T4 phage genome[64]. Nevertheless, based on the functions of other early proteins, one may speculate that some of these uncharacterized proteins might be involved in host-hijacking or host-defense processes.

In total, 79 T4 phage proteins meet the criteria for middle proteins (Supplementary Table S3). At the middle phase of the infection, the number of proteins that are involved in T4 phage DNA replication (gene 61 (DNA primase), gene 43 (DNA polymerase)) and its protection against host nucleases (a-gt, b-gt) strongly increases. This indicates that infection proceeds from host adaptation (early phase) towards phage DNA replication (middle phase).

Finally, 105 T4 proteins were classified as late proteins. This class is mainly comprised highly abundant structural proteins and proteins involved in phage assembly and packaging, whose activity results in the formation and release of T4 phage progeny upon host lysis (Supplementary Figure S13, Supplementary Table S3).

Altogether, the data collected in this study confirm the previously assumed temporally resolved and overall highly organized T4 phage protein biosynthesis[29]. The temporal organization of the T4 phage proteome matches a clear pattern from early host cell reprogramming to DNA replication in the middle infection phase, ending with phage assembly and release. In addition, a direct link between the point in time of protein appearance and its function can be made. This fact might be beneficial for the elucidation of the biological functions of numerous uncharacterized T4 phage proteins that appear at various stages of infection.

Correlation of Transcriptomics and Proteomics Data Implicates Post-Transcriptional Mechanisms Governing T4 Phage Gene Expression

Our dual-transcriptomics and -proteomics study of T4 phage infection with *E. coli* revealed the temporal control of RNA transcription and degradation as well as protein synthesis during infection. On the transcriptomic level, it stands out that host RNAs are predominantly degraded during infection, including the large group of host mRNAs, whereas host proteins remained rather stable. Thus, host mRNA translation appears to be rapidly shut-off upon T4 phage infection. Mechanistically, it seems reasonable that the T4 phage takes over the entire transcription

capacities of the host whilst harnessing the existing cellular protein machinery, such as the gene expression apparatus for its propagation.

Whilst *E. coli* gene expression lacks temporal up- and downregulation patterns during infection, we observed the synthesis and degradation of T4 phage transcripts as well as the controlled onset of T4 phage protein synthesis in distinct infection phases. The power of our dual-omics approach not only enabled us to track time-resolved gene expression during infection but also allowed us to infer specific already-studied gene regulatory cascades within T4 phage infection. Here, we demonstrate the temporal order of T4 phage gene expression events for specific examples well-described in the literature. Conceptually, early T4 phage gene products mediate the transition to middle gene expression, which in turn activates the late infection phase (Figure 4).

For instance, the induction of the expression from T4 middle promoters is partially mediated by the MotA protein—an early gene product also classified as such in our data[65,66]. We observed the coordinated onset of *motA* gene transcription (transcriptome, t1) and translation (proteome, t1) early during infection followed by the expression onset of the *rIIB* gene—controlled by a T4 middle promoter—approximately 3 min afterwards (transcriptome, t4; proteome, t5) (Figure 4a). Similarly, production of AsiA—the co-activator of middle transcription during T4 phage infection[17,67]—is detected at early time points in our data sets (transcriptome, t1; proteome, t1-t3). On the other hand, late T4 phage transcription requires a sigma factor composed of the proteins gp55 and gp33[68,69]. In addition to the gp55–gp33 complex, the transcriptional activator gp45 is required to initiate the transcription from the late phage promoters[70,71]. As a consequence, late gene products are formed, which mainly fulfil roles in DNA packaging and capsid assembly. On both transcriptome and proteome levels, we detected the expression of the activators of late transcription—gp33, gp45, and gp55 (each: transcriptome, t1; proteome, t5)—in the early and middle phases of infection (Figure 4b,c). Upon the presence of all three gene products, we observed the expression of late genes, such as *alt* (transcriptome, t7; proteome, t8), whose gene product is part of the mature T4 phage progeny[72] (Figure 4b) or *68* (transcriptome, t7; proteome, t12), encoding a structural phage head protein[73] (Figure 4c). This emphasizes that our time-resolved dual-transcriptomics and -proteomics approach can reproduce T4 phage gene regulatory cascades (Figure 4d), which in combination contribute to phage assembly and release. Importantly, the above-described gene expression cascades were reported in previous studies with the focus on these particular proteins only[65,66,70,71]. In contrast to these highly valuable but laborious biochemical studies, our data enable the investigation of possibly any gene expression cascade during T4 phage infection simultaneously on the transcriptome and proteome levels. To enable broad community access to these correlated data sets, we designed a user-friendly web

application (POTATO4) to retrieve transcriptome and proteome information for T4 phage and *E. coli* genes of interest. Thereby, we hope to provide a valuable resource to build and support hypotheses and biochemical studies on T4 phage infection regulation and related subjects.

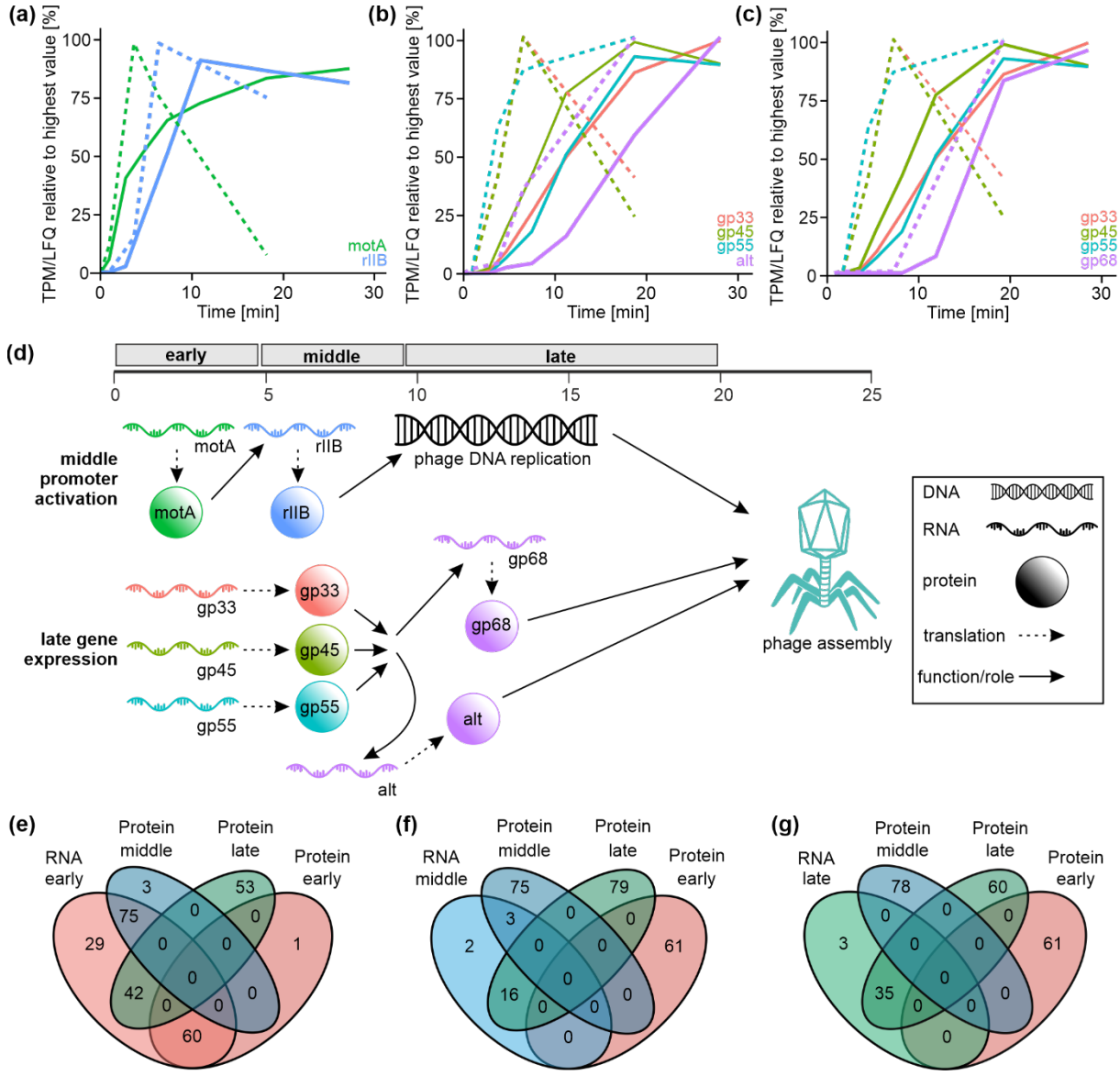


Figure 4. Analysis of correlation of transcriptome and proteome data of T4 phage infection. (**a**–**c**) Examples of the gene expression regulation during T4 phage infection based on the transcriptome (dashed line) and proteome data (continuous line). (**a**) Example of MotA expression which activates middle transcription of the *rIIB* gene. (**b**,**c**) Expression of the activators of late transcription – gp33, gp45 and gp55 triggers late expression of Alt (**b**) and gp68 (**c**). (**d**) Schematic representation of selected gene regulatory cascades during T4 phage infection. A similar color code as shown in (**a**–**c**) was applied. Infection-phase-specific transcripts and proteins are shown in the order of their appearance during infection based on the

data shown in (**a**–**c**). Arrows indicate the transition between gene regulatory and molecular events. (**e**–**g**) Venn diagrams displaying overlaps of T4 phage gene classes based on proteome data with either early genes (**e**), middle genes (**f**) or late genes (**g**) derived from transcriptome data set.

In addition, we set out to compare our transcriptome- and proteome-based classifications of T4 phage genes to obtain an overall impression of the correlation between the T4 phage transcriptome and proteome. To our surprise, we detected a large discrepancy between the fractions of T4 genes assigned to the individual infection phases. Most genes were classified as early transcripts based on our transcriptomics data, whereas our proteome-based classification yielded a comparably even distribution among T4 phage gene classes (Figure 4e–g). Thus, early T4 mRNAs encode early, middle, as well as late T4 proteins (Figure 4e). This is also emphasized in T4 genomic maps showing that clusters of early transcripts contribute to all classes of T4 phage proteins (Supplementary Figure S14). Late proteins are derived from all classes of T4 mRNAs (Figure 4e–g), while all late mRNAs only encode for a subset of late T4 proteins (Figure 4g). This discrepancy may be based on the expression principles of early T4 mRNAs in our transcriptomic data. For all early RNAs, moderate expression is detected already early during infection. However, early RNAs diverge into two distinct sets—one set with an expression peak in the early or middle infection phase and another set with a constitutively increasing expression to the end of the infection (Figure 2c). The finding of early mRNAs giving rise to late proteins (summarized in Supplementary Table S4) is in good agreement with the scientific concept of the post-transcriptional regulation of T4 phage gene expression, which has previously been built on reductionist biochemical studies[16,19]. Here, we were able to characterize this phenomenon globally. Still, it remains to be investigated how the entirety of early T4 mRNAs are regulated on a post-transcriptional and translational level to be translated late during infection. So far, it is well established that RegB specifically degrades early transcripts to shut down their translation[17,74]. This could also apply to the transcripts of late genes, which are transcribed early during infection. In order to further elucidate reasons for discrepancies in the temporal appearance of phage transcripts and respective proteins, translational regulatory factors, such as the position and composition of the Shine–Dalgarno sequence, the involvement of ribosome-binding proteins, or the presence of riboswitches, still need to be studied in detail in future studies.

2.4. Conclusions

In this work, we described the first time-resolved dual-transcriptome and -proteome study of the T4 phage infection of *E. coli*, allowing for comprehensive molecular insights into the infection process. In this study, we confirmed that *E. coli* transcriptome is largely degraded during the

infection, including both mRNAs and host tRNAs. Intriguingly, we identified four *E. coli* non-coding RNAs, which appear to possess unexpected stability throughout the infection. In stark contrast to the dual-transcriptome, the proteome of the host remains stable throughout infection, probably due to the utilization of the *E. coli* proteins by the T4 phage. Moreover, we could gain insights into the transcriptome and—for the first time—the comprehensive proteome of T4 phage infection, demonstrating the temporal control of T4 phage genes on two levels of gene expression. Correlating the phage transcriptome and proteome showed that specific T4 phage mRNAs and proteins are temporally decoupled, suggesting post-transcriptional and translational regulation mechanisms. Thus, we obtained the first global picture of T4 phage infection on the levels of gene expression, focusing on the phage and the host. Our data are in good agreement with multiple studies dedicated to the gene expression regulation of T4 phage infection.

Moreover, we speculated that the temporal appearance of T4 phage proteins might be linked to their function. Surprisingly, one-third of T4 phage proteins are annotated as uncharacterized/hypothetical proteins. Based on our data sets, these hypothetical proteins might be functionally classified.

Thereby, this work is a valuable resource for future studies focusing on yet unexplored phage–host interactions and gene regulatory events during T4 phage infection. Moreover, this work exemplifies the power of time-series transcriptomics and proteomics to obtain a comprehensive understanding of gene expression during phage infection. Our data sets can explain observed changes in protein and transcript abundances and provide insights into the causal flow of molecular information during phage infection.

2.5. Materials and Methods

Reagents

We purchased all reagents from Sigma-Aldrich (St. Louis, MO, USA) if not indicated differently.

Strains and Media

We obtained *Escherichia coli* strain B (*Escherichia coli* (Migula 1895)), Castellani and Chalmers 1919 (DSM 613, ATCC 11303; DSMZ, Braunschweig, Germany)) and T4 phage (Escherichia phage T4, DSM 4505; DSMZ, Braunschweig, Germany) from the DSMZ. We carried out *E. coli* strain B cultivation and T4 phage propagation in LB (Luria/Miller) medium supplemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$ for T4 phage infections.

RNA Isolation from T4 Phage Infected *E. coli*

We grew a culture of *E. coli* strain B to an OD_{600} of 0.5 at 37 °C in LB (Luria/Miller) medium supplemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$. We added T4 phage suspension to a multiplicity of infection (MOI) of 3.1 and subsequently it grew at room temperature. We took 5 mL of culture at 0 min (before infection), 1, 4, 7, and 20 min post-infection and immediately lysed using the hot lysis method [19]. Therefore, we incubated the culture with 1 volume of lysis solution (1% SDS, 4 mM EDTA) at 95 °C for 2 min each. We added 1 volume of water-saturated phenol (Roti-aqua phenol; Carl-Roth, Karlsruhe, Germany) to each sample and incubated at 67 °C for 10 min. We centrifuged samples at 10,000× *g* for 10 min, and then we added upper phase to 1 volume phenol/chloroform/isoamyl alcohol (Carl-Roth, Karlsruhe, Germany). We centrifuged samples again (10,000× *g*, 10 min). We precipitated the upper phase by centrifugation (14,000× *g*, 90 min, 4 °C) in the presence of 0.3 M sodium acetate and 1 volume isopropanol. We resuspended the RNA pellet in RNase-free water, and then we digested residual DNA with 2 µL DNase I (Roche, Basel, Switzerland) in 1 × DNase buffer at 37 °C for 30 min. We twice extracted the RNA with 1 volume phenol/chloroform/isoamylalcohol (Carl-Roth, Karlsruhe, Germany) and removed residual phenol by diethyl ether (ThermoFisher Scientific, Waltham, MA, USA) extraction. We again precipitated RNA in the presence of 0.3 M sodium acetate and 1 volume isopropanol by centrifugation (14,000× *g*, 4 °C, 90 min). We resuspended each RNA pellet in 50 µL RNase-free water. We performed RNA isolation in triplicates. We measured the RNA concentration using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) and a Qubit Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). We evaluated RNA integrity on a Bioanalyzer (Agilent, Santa Clara, CA, USA) using the RNA 6000 Nano Kit.

Preparation of RNA-Sequencing (RNA-Seq) Libraries and Illumina Sequencing

We conducted RNA sequencing at the Deep Sequencing Core Facility of the Bioquant Heidelberg (led by D. Ibberson). We subjected 1 µg total RNA to ribosomal RNA depletion (rRNA) by Ribo-Zero rRNA Removal Kit (Gram-negative bacteria). We randomly sheared rRNA-depleted RNA in 10 µL dH_2O at 94 °C for 10 min. We processed fragmented RNA using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. We barcoded cDNA by PCR amplification using Illumina TruSeq adapters for Illumina. We performed further cDNA size selection in the range of 300 to 500 bp employing the Agencourt RNAClean XP kit. We examined primer-depleted cDNA by Bioanalyzer, then we measured the concentration by Qubit. We sequenced multiplexed libraries on a NextSeq 500 platform (Illumina, San Diego, CA, USA).

Northern Blot Analysis

We generated in vitro transcription templates for ssrA and RNAC by PCR using partially hybridizing DNA oligos:
SsrA:TAATACGACTCACTATAGGACACGCCACTAACAAACTAGCCTGATTAAGTTTTAACGCTT, CGCGTGGAAGCCCTGCCTGGGGTTGAAGCGTTAAAACTTAATCAGGCTAGTTTG;
RNAC:TAATACGACTCACTATAGGGTTAAAAGGCCATATCTCAACCATATCCGAACGTTCCGTCAAAAACGC,TCGATTCGAGGAAATATCTTTGCCGTAAGCCGAGTAGCGTTTTTGACGGAACGTTCGG.
We transcribed radioactive RNA riboprobes for Northern blot in vitro from 1 µM DNA template in the presence of 40 mM Tris pH 8.1; 1 mM spermidine; 22 mM MgCl2; 0.01% Triton-X-100; 10 mM DTT; 5% DMSO; 0.1 mg/mL T7 RNA polymerase (purified in-house); 4 mM GTP, CTP, and GTP; 2 mM ATP; and 0.5 mCi/mL radioactive ATP (^{32}P-α-ATP, 3,000 Ci/mmol; Hartmann Analytics) at 37 °C for 4 h [75]. We digested and extracted in vitro transcription products using DNase I (Roche Diagnostics, Rotkreuz, Switzerland) with phenol/chloroform/isoamylalcohol (Carl Roth, Karlsruhe, Germany), as described above; then, we precipitated in the presence of 0.3 M NaOAc pH 5.5 and 1 volume isopropanol by centrifugation at 17,000× *g*, 4 °C, for 90 min. We resuspended RNA riboprobes in 50 µL MQ water.
We analyzed 10 µg total RNA per sample via 10% denaturing polyacrylamide gel electrophoresis (5 W, 60 min). We transferred RNA to a nylon membrane (GE Healthcare) in a Trans-Blot Turbo System (Bio-Rad) in the presence of 0.5 × TBE (50 mM Tris, 50 mM boric acid, 1.25 mM EDTA) at 250 mA for 2.5 h and UV crosslinked. We pre-hybridized membrane in 20 mL ROTI® Hybri-Quick (Carl Roth, Karlsruhe, Germany) for 30 min at 45 °C. We added 5 µL ^{32}P-labelled RNA riboprobe to the pre-hybridized membrane and incubated at 45 °C overnight. We subsequently washed blots twice with wash solution 1 (2× SSC, 0.1% SDS) and twice with wash solution 2 (0.25 × SSC, 0.1% SDS) for 5 min each. We visualized blots with storage phosphor screens (GE Healthcare, Chicago, IL, USA) at the Amersham Typhoon imaging system (GE Healthcare, Chicago, IL, USA). We quantified band intensities using ImageLab 6.1 (Bio-Rad, Hercules, CA, USA).

Proteome Samples Preparation

We grew *E. coli* strain B culture in LB medium at 37 °C and 180 rpm until OD_{600} of 0.8 was reached. We performed the T4 phage infection assay at room temperature and 120 rpm. We infected cells with T4 phage at a MOI of 5. We took 2 mL samples before infection (0 min) and 1, 3, 5, 8, 12, 20, and 30 min post-infection. We immediately harvested the cells by centrifugation at

17,000× *g* for 1 min; then, we directly resuspended RT and the pellet in 200 µL hot lysis buffer (Tris-HCl pH 7.5, 1% SLS, 2 mM TCEP, 95 °C) and boiled for 10 min at 95 °C. We performed a short sonication step to degrade nucleic acids present in the samples (10 sec, 20% amplitude, 0.5 pulse). We added 4 mM Iodoacetamide and incubated the samples for 30 min while protecting from light. We precipitated proteins via acetone and washed the pellets with 500 µL methanol (-80 °C) before air-drying and resuspending in 50 µL resuspension buffer (50 mM Tris-HCl, pH 7.5, 0.5% SLS). We determined the protein concentration by BCA assay (Pierce TM, BCA protein assay kit (reducing agent compatible), ThermoFisher Scientific, Waltham, MA, USA). We added 1 µg of sequencing-grade trypsin (Promega) to 20 µg of isolated proteins and digested o/n at 30 °C in the presence of 50 mM Tris-HCl, pH 7.5. We precipitated residual SLS by adding 1.5% TFA before separating precipitate by centrifugation at 4°C, 17,000× *g* for 10 min. We desalted the supernatant for mass spectrometric analysis using C_{18} solid phase columns (Chromabond C18 spin columns; Macherey Nagel, Düren, Germany).

Proteome LC-MS Analysis

We performed LC-MS analysis on an Exploris 480 instrument connected to an Ultimate 3000 rapid-separation liquid chromatography (RSLC) nano instrument and a nanospray flex ion source (all Thermo Scientific). We carried out peptide separation out on a reverse-phase high-performance liquid chromatography (HPLC) column (75 µm × 42 cm) packed in-house with C18 resin (2.4 µm; Dr. Maisch GmbH). For total proteome analysis, we performed peptide elution in backflush mode with a separating gradient from 98% solvent A (0.15% formic acid) and 2% solvent B (99.85% acetonitrile, 0.15% formic acid) to 25% solvent B over 40 min, followed by up to 60 min with 25% to 35% of solvent B at a flow rate of 300 nl/min. We performed label-free quantification (LFQ) data sets of total proteomes in data-dependent acquisition (DDA) mode. We acquired a high-resolution MS 1 scan at a resolution of 60,000 (at *m/z* 200) with a scan range from 350 to 1650 m/z, followed by MS/MS scans within 2s (Cycle 2s) of the most intense ions at a resolution of 15,000. We set charge state inclusion between 2 and 6. We set the ion accumulation time to 25 ms for MS and AUTO for MS/MS. We set the automatic gain control (AGC) to 300% for MS survey scans and 200% for MS/MS scans. The parameters of the measurements are summarized in Supplementary Table S3.

We performed DDA-LFQ analysis using MaxQuant [76] in standard settings using *E. coli* (Proteome: UP000000625) and bacteriophage T4 (Proteome: UP000009087) fusion database. We further evaluated the “proteinGroups.txt” MaxQuant output file with the SafeQuant R script updated to modify MaxQuant outputs[77].

Analysis and Visualization of RNA-Seq and Proteomics Data

We assessed the quality of reads obtained from Illumina RNA-Seq pre- and post-adapter trimming using FastQC (version 0.11.9). We processed Fastq files using the cutadapt tool (version 1.18) in order to remove reads containing Illumina TruSeq adapter sequences. We aligned reads to the genome of *E. coli* K12 (U00096.3) and bacteriophage T4 (NC_000866.4) using the hisat2 aligner (version 2.2.1) at default settings. Thereby, we successfully aligned 88.71 to 96.51% of reads to the reference genomes.

We applied Samtools (version 1.7) to select for primary alignments. We manually inspected BAM files as genomic maps using the Integrative Genomics Viewer (version 2.4.9). We quantified the reads mapped to individual features (annotated in gff3 files for U00096.3 and NC_000866.4 as gene) using featureCounts (Subread package version 2.0.1) with default settings while excluding reads overlapping multiple features.

Prior to further analysis, we manually removed *E. coli* genes annotated as rRNA (22 genes), which account for up to 0.6% of reads per sample (Supplementary Figure S1), from the counts table using R (version 4.1.2), because they were depleted from the total RNA before sequencing. We normalized the count data to transcripts per million (TPM) which allows to compare expression levels of genes between samples. We assessed sample clustering by principal component analysis (PCA) with the prcomp package. We calculated the fractions of TPM-normalized reads per sample and entity (T4 phage or *E. coli*) accordingly. We removed genes with low read counts (average read count below 1.5 across all samples) from the TPM-normalized count data, including 338 *E. coli* and 17 T4 phage genes. These genes would otherwise confuse the analysis due to low and variable read counts. We conducted further analyses in R focussed on data visualization using the pheatmap (1.0.12) and ggplot2 (3.3.6) package.

We assessed differential gene expression analysis during the early infection phase with DESeq2 by applying a Wald test at a log2fold change greater or smaller than 0 with a Benjamini–Hochberg-corrected p-value threshold of 0.05[78]. We did not assess other infection phases by DESeq2 as dramatic changes in the host and phage transcriptome have already been recorded 4 min post-infection. We conducted assignment of differentially expressed genes to Clusters of Orthologous Groups (COGs) in R using the COG database from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/research/cog/accessed on 12th April 2021).

Different from the RNA-Seq data, we further analyzed the proteomics data based on LFQ values, as they already express the approximate abundance of the respective proteins in the sample. We also assessed sample clustering in R by PCA analysis (prcomp package) and Pearson

correlation. We conducted further analyses in R focused on data visualization using the pheatmap (1.0.12) and ggplot2 (3.3.6) packages.

For both data sets, we based classification of T4 phage genes on the period of time during which the gene's expression was below 10% its maximal detected expression (transcriptomics: TPM, proteomics: LFQ). Therefore, we initiated early gene expression is during the first 4 (transcriptomics) or 5 (proteomics) minutes of infection, middle gene expression between 4 and 7 (transcriptomics) or 5 and 8 (proteomics) minutes, and late gene expression after 7 (transcriptomics) or 8 (proteomics) minutes post-infection.

We performed comparison and integration of time-series transcriptomics and proteomics data in R using data visualization tools as described above.

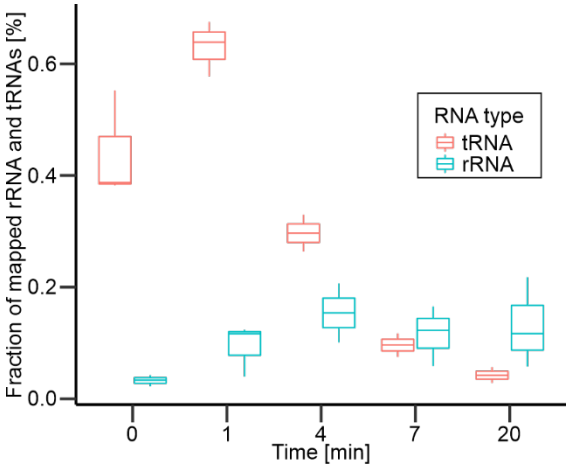
Data Availability Statement

Raw RNA-Seq data is deposited in GEO and accessible via identifier GSE211026. The MS Raw data can be accessed via the PRIDE/ProteomeXchange consortium under the project identifier PXD035873. POTATO4 is a user-friendly web application and accessible via https://rshiny.gwdg.de/apps/potato4/. The Supplementary Tables S1-S4 are deposited on Zenodo and accessible via https://zenodo.org/records/10855527.

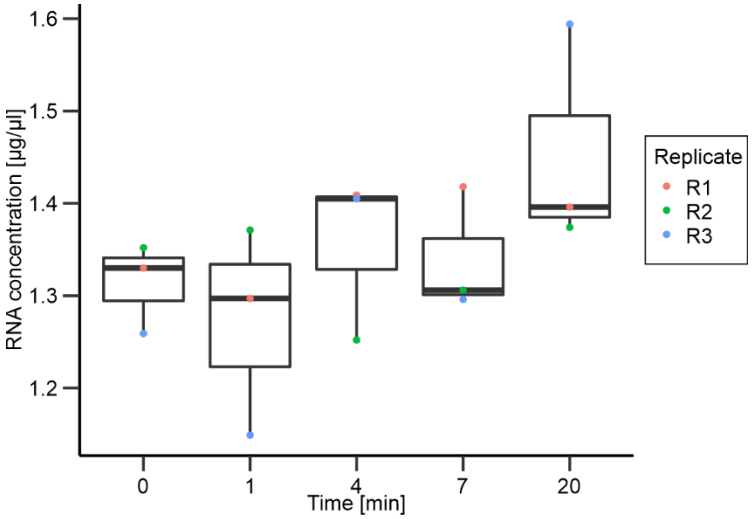
Acknowledgments

Processing and mapping of reads from fastq files was performed on the High-Performance Compute Cluster Cobra from the Max-Planck Compute and Data Facility. We thank Dominik Ricken for his support with the total RNA isolation from T4 phage-infected *E. coli*, David Ibberson (Deep Sequencing Lab, Bioquant Heidelberg) for preparing the RNA-Seq library and performing the Illumina Sequencing, and Carl Weile for the Northern blot analysis.

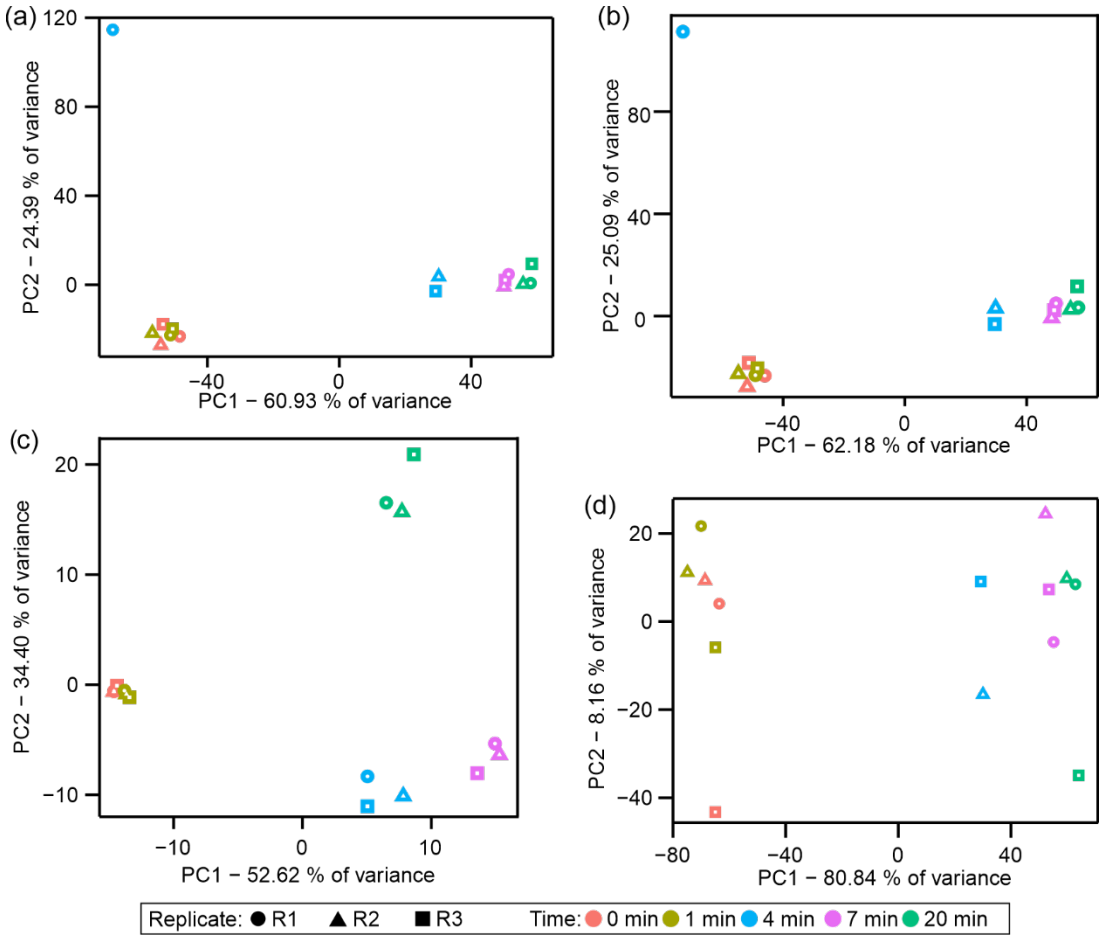
2.6. Supplementary Material



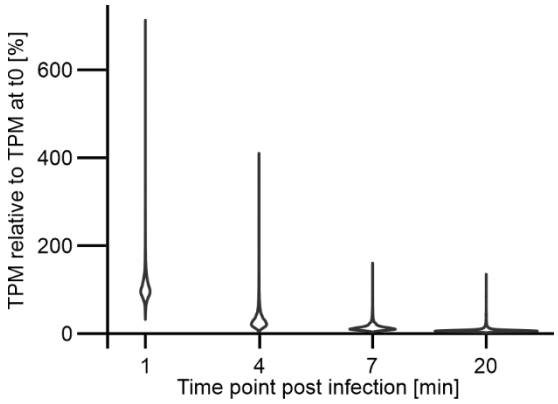
Supplementary Figure S5: Fractions of *E. coli* tRNAs and rRNAs among all reads in all samples. Fractions of *E. coli* rRNAs (22, blue) and *E. coli* tRNAs (84, red) among all reads per sample were calculated from TPM normalized count data and visualized in a boxplot. Fractions of respective RNA groups are shown for each time point (0 min (before infection) and 1, 4, 7, 20 min post infection). Data at each time point is based on TPM values from biological triplicates (n = 3).



Supplementary Figure S6: Total RNA concentration per time point and replicate after RNA isolation. Total RNA was isolated, DNase I digested and resuspended in 200 µl RNase-free water. RNA concentrations were determined with the NanoDrop. Here, the concentrations are presented in a boxplot in order to visualize overall RNA yields for all samples[µg/µl]. RNA samples reflect the total RNA of T4 phage infection at 0 (before infection) and 1, 4, 7, and 20 min post-infection in biological triplicates (n = 3).

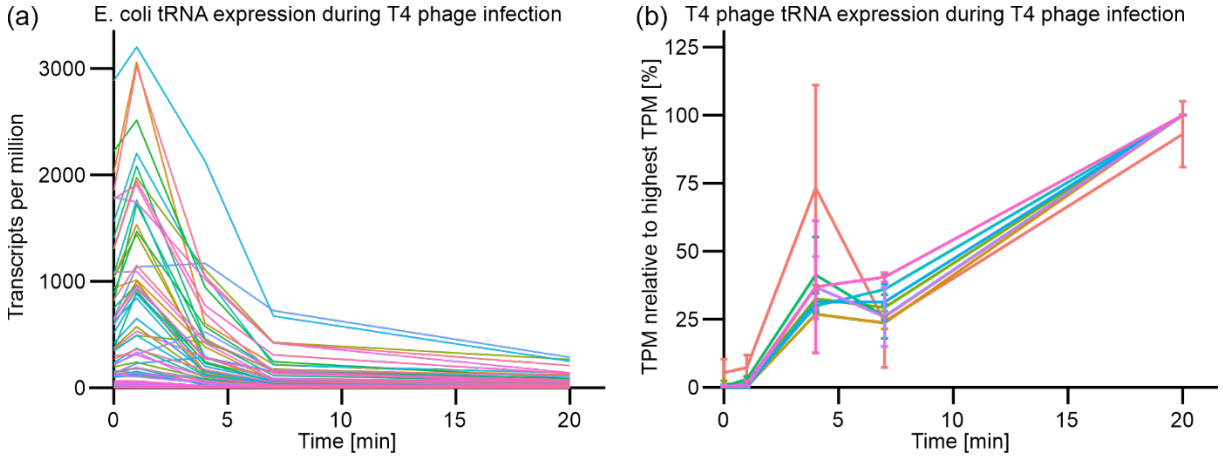


Supplementary Figure S7: Principal component analysis of dual-transcriptome RNA-Seq samples. Principal component analysis (PCA) of RNA-Seq samples based on read counts for *E. coli* and T4 phage genes. PCA analysis is shown for TPM normalized read counts for *E. coli* and T4 phage genes (a), *E. coli* genes only (b) and T4 phage genes only (c). Replicate 1 from 4 min time point (t4 R1) showed large discrepancies in *E. coli* gene expression compared to all other samples (a, b), whereas expression of T4 phage genes was in-line with the other replicates 2 and 3 from t4 (c). Thus, t4 R1 was excluded from further analyses which resulted in close clustering of replicates of the same time points in PCA of TPM normalized read counts for *E. coli* and T4 phage genes (d). Each PCA was performed on the basis of TPM normalized reads. Color indicates time point of infection, shape determines replicate. Principal components 1 (PC1) and 2 (PC2) with respective explained fraction of data variance are presented on x- and y-axis.



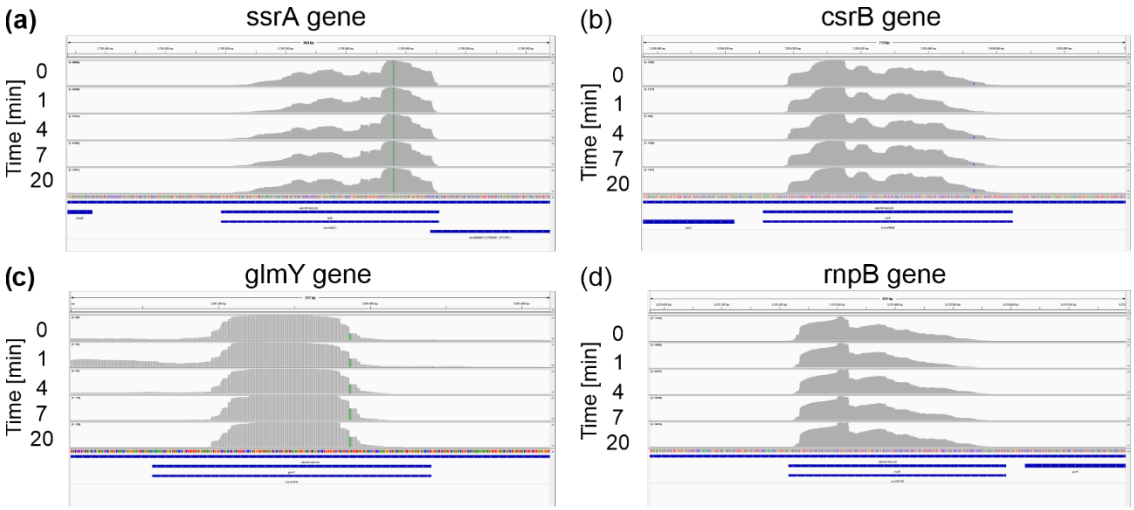
Supplementary Figure S8: Global degradation of host transcripts during T4 phage infection.

E. coli transcript degradation monitored with violin plots. Per time point (1, 4, 7, and 20 minutes post infection) the TPM for each *E. coli* gene and replicate were normalized by their mean TPM values at t0 (0 min post infection, uninfected *E. coli*). The violin plots display the distribution of these fractions per gene and time point post infection showing a decline of TPM values relative to t0. Calculations are based on mean TPM values from biological triplicates (n = 3) except for 4 min post-infection, where only duplicates (n = 2) were considered.

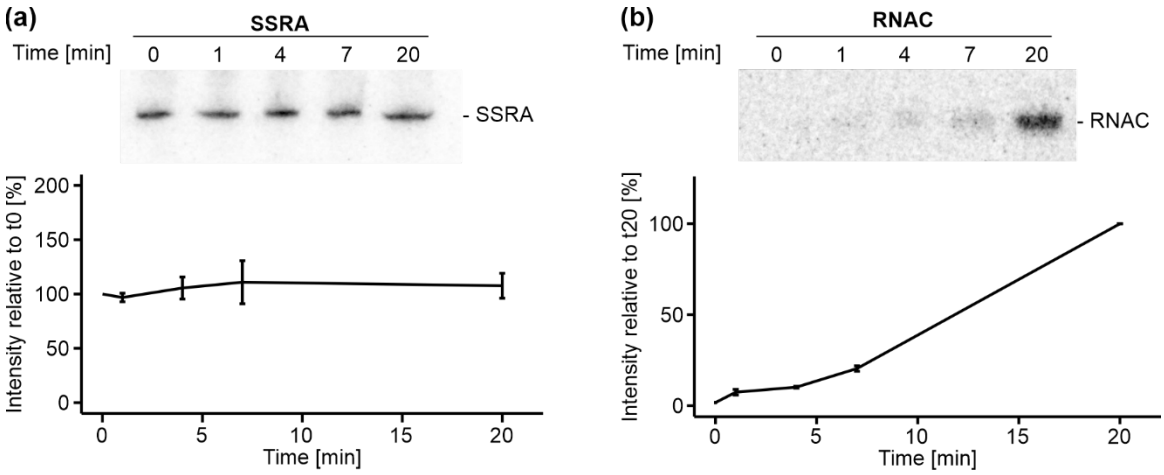


Supplementary Figure S9: Time-resolved expression of transfer RNAs during T4 phage infection.

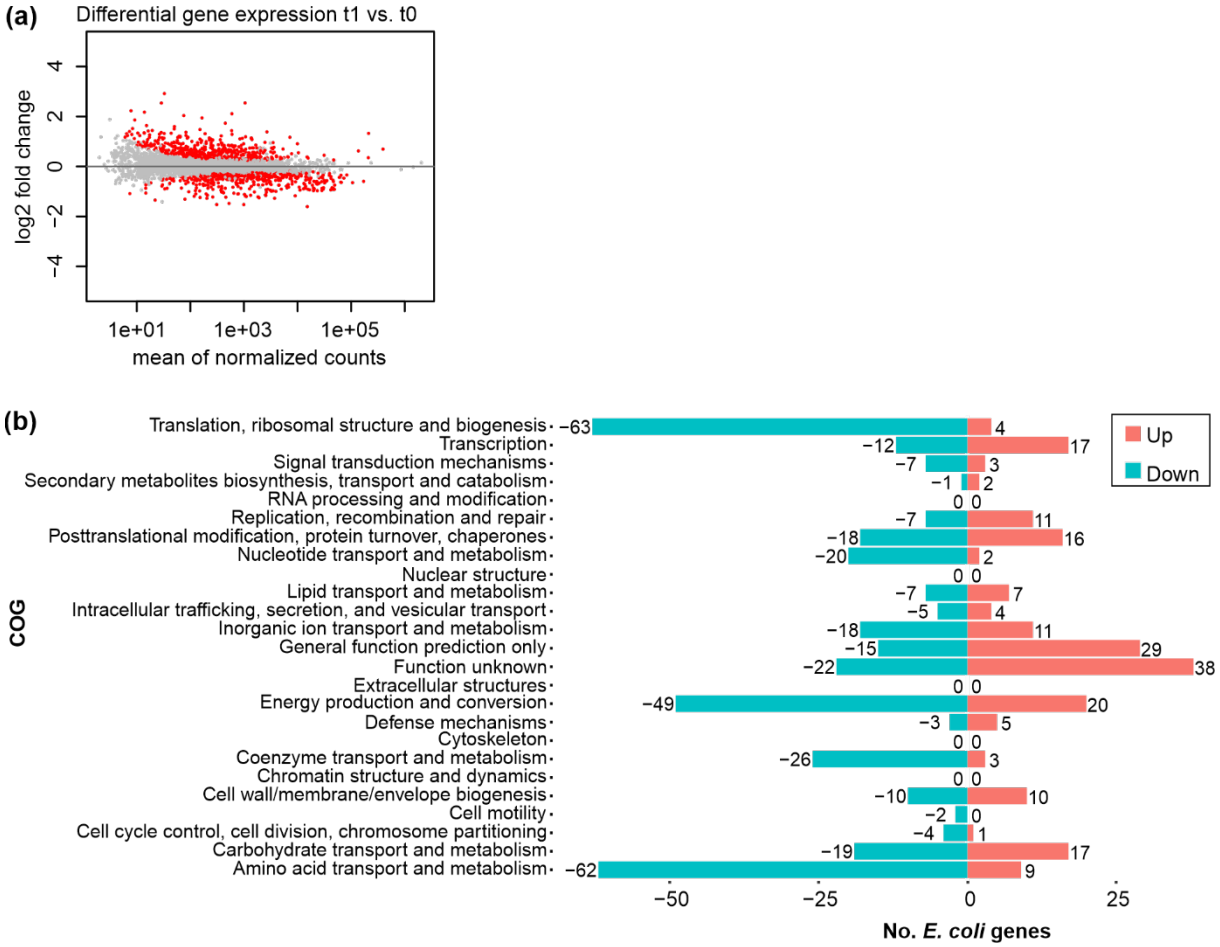
a) Expression of transfer RNAs (tRNAs) derived from the host *E. coli* during T4 phage infection plotted as nominal mean TPM values which were calculated on the basis of TPM values from biological triplicates (n = 3; except for 4 min post infection, n = 2). **b**) Expression of T4 phage tRNAs over the time course of infection measured as the fraction of TPM values relative to the highest TPM value per replicate per T4 phage tRNA. Data points represent mean, normalized TPM values, errorbars are based on the standard deviation of data in biological triplicates (n = 3; except for 4 min post infection, n = 2). Each color represents a single tRNA from either *E. coli* or T4 phage.



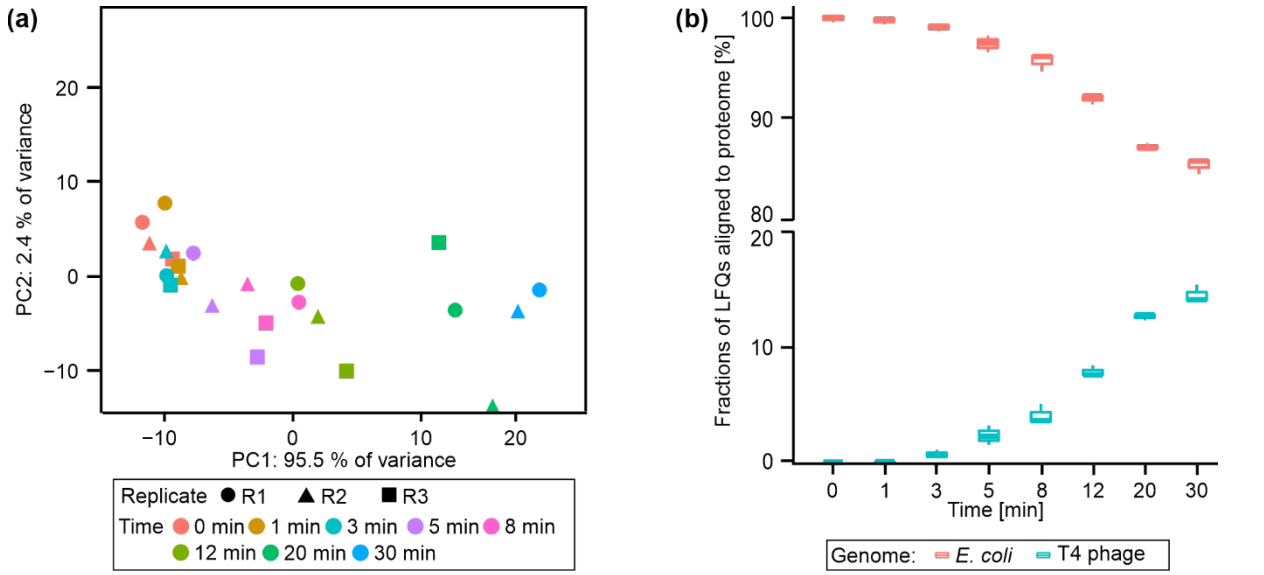
Supplementary Figure S10: **Read coverage on genes of comparably stable *E. coli* transcripts during T4 phage infection.** Read coverage on genes encoding for comparably *E. coli* stable transcripts under T4 phage infection. Coverages are depicted for all time points (0, 1, 4, 7 and 20 minutes) based on data for replicate 2 (R2) for the genes *ssrA* (a), *csrB* (b), *glmY* (c) and *rnpB* (d). Coverage visualization was performed using the Integrative Genomics Viewer at default settings.



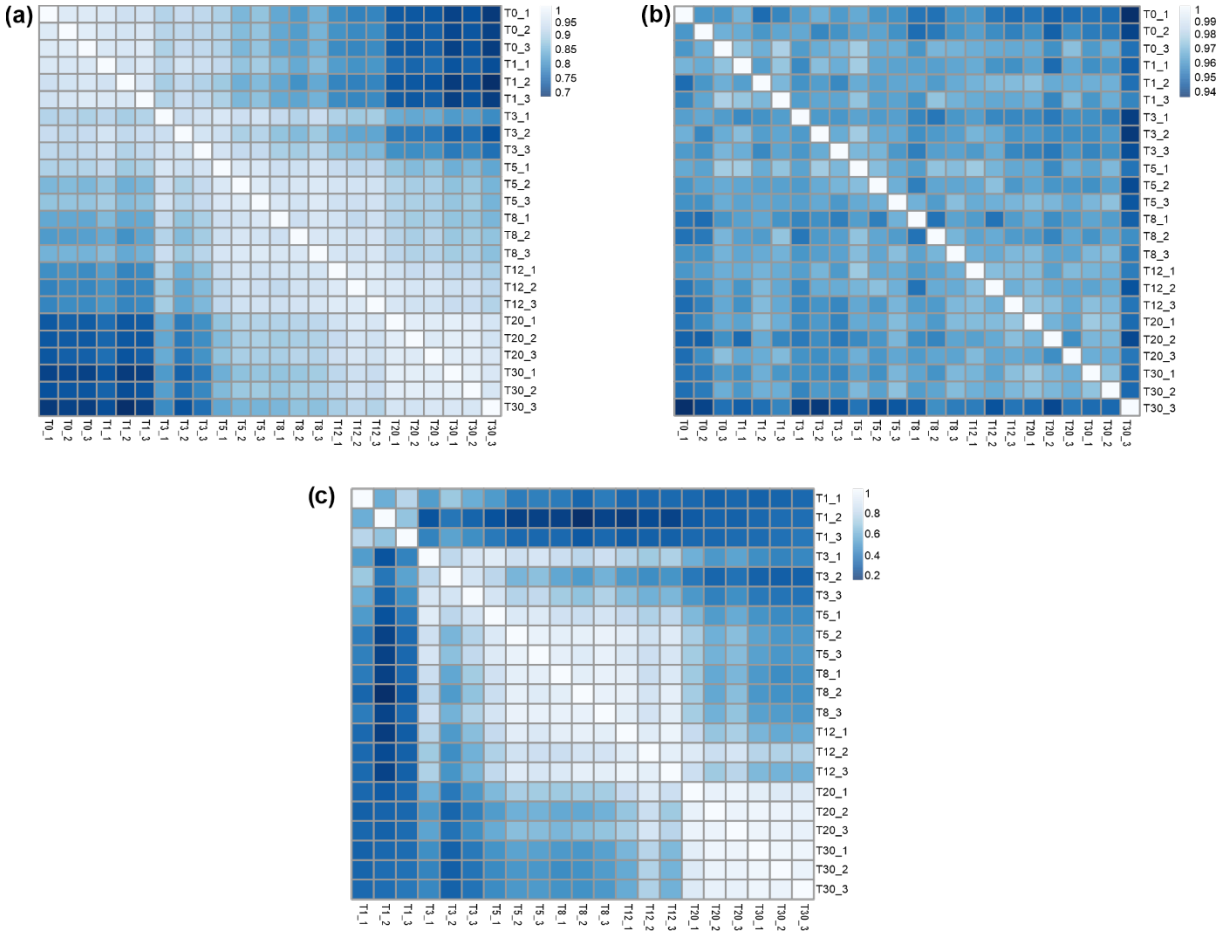
Supplementary Figure S11: Northern blot analysis of *E. coli* and T4 phage transcripts during infection. Northern blot analysis of SSRA transcripts over the time course of T4 phage infection at 0 min (before infection) and 1, 4, 7, and 20 minutes post infection shown for the *E. coli* SSRA transcript (**a**, upper panel) and the T4 phage sRNA RNAC (**b**, upper panel). Northern blot analysis was performed with 10 µg total RNA per time point. Band intensities were quantified using ImageLab 6.1 and normalized to either t0 (SSRA, **a**, lower panel) or t20 (RNAC, **b**, lower panel). Plots of normalized band intensities are based on three (n = 3, SSRA) or two (n = 2, RNAC) independent biological replicates of the Northern blots.



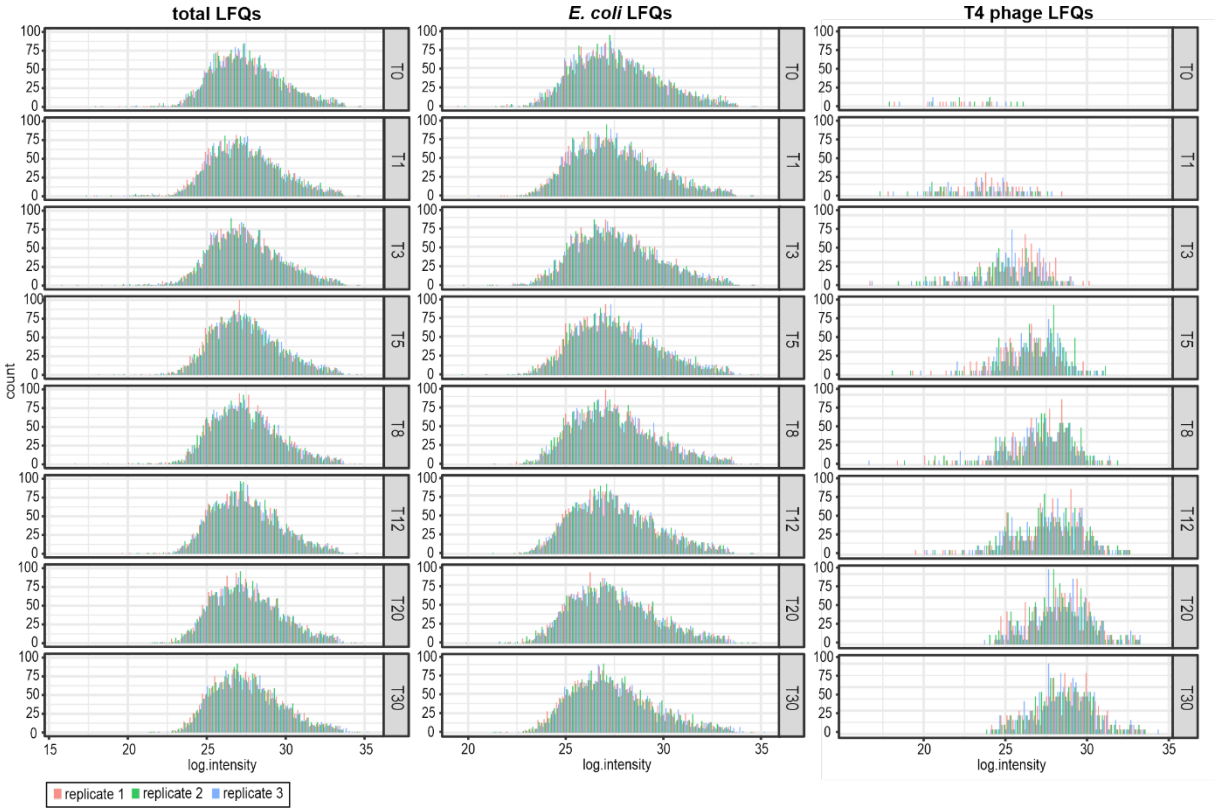
Supplementary Figure S12: **Differential expression of *E. coli* genes during the early phase of infection. a**) MA plot showing log2 fold changes of *E. coli* genes at 1 min post-infection compared to 0 min time point. Genes differentially expressed (LFC > 0 or > 0 and adjusted p-value < 0.05) are colored red. Differential expression analysis was performed with DESeq2. Therefore, T4 phage genes were removed from raw read count data, a pseudocount (0.5) was added and low expression and rRNA genes were removed for DESeq2 analysis. **b**) Assignment of significantly, differentially expressed *E. coli* genes to Clusters of Orthologous Groups (COGs) differentiating downregulated (cyan) or upregulated (red) *E. coli* genes during the early phase of T4 phage infection (1 min post-infection). The bar height determines the number of differentially expressed assigned to the respective COG.



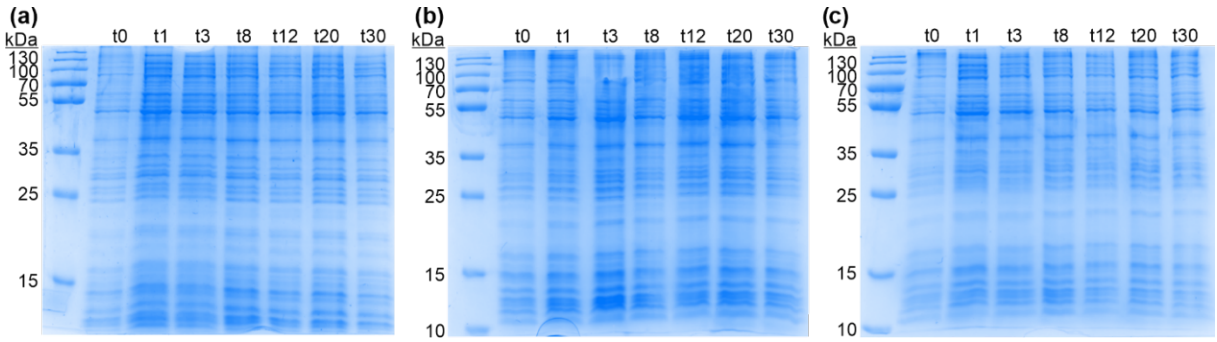
Supplementary Figure S13: Initial analysis of the time-resolved dual-proteome of T4 phage infection of *E. coli*. **a**) Principle component analysis (PCA) of proteomics samples based on LFQ values for individual *E. coli* and T4 phage proteins. The color indicates the time point of infection, the shape determines the replicate. Principal components 1 (PC1) and 2 (PC2) with a respective explained fraction of data variance are presented on the x- and y-axis. A close clustering of biological replicates per time point is observed (n = 3). **b**) Fractions of LFQ values derived from either the *E. coli* or the T4 phage proteome relative to the sum of all LFQ values per sample over the time course of T4 phage infection, based on biological triplicates (n = 3).



Supplementary Figure S14: Analysis of the correlation between all proteomics samples of this study. **a**) Pearson correlation between not processed samples (Pearson correlation coefficient: 0.7-1) presented in a heatmap. **b**) Pearson correlation between the analyzed samples in regard of *E. coli* proteins only presented in a heatmap. Low variation in Pearson correlation coefficient (0.94-1) confirms the stability of the *E. coli* proteome. **c**) Pearson correlation between the analyzed samples in regard of T4 phage proteins only presented in a heatmap. High variation in the Pearson correlation coefficient (0.2 -1) is in accordance with the dynamic expression of the T4 phage proteins. Samples are abbreviated by combining the respective time point (e.g. T12 = 12 min post-infection) and the biological replicate (e.g. R2 = replicate 2) to, e.g. T12_R2.

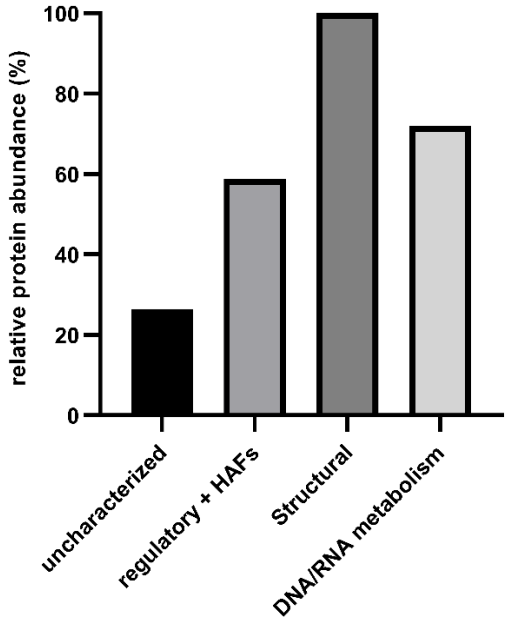


Supplementary Figure S15: Normalized histograms of LFQ values contributed by phage and host. Histograms of LFQ values are shown for all time points (T0, T1, T3, T5, T8, T12, T20, T30; e.g. T12 = 12 min post-infection) including all biological replicates (n = 3) per time point. Histograms are shown for all *E. coli* and host proteins (left panel), *E. coli* proteins only (middle panel) and T4 phage proteins only (right panel). All the histograms show normal distribution of the data, except for the histograms for T4 phage LFQs at early time points of infection (T0 – T3). This can be explained by the low abundance of phage proteins at the early stages of the infection.

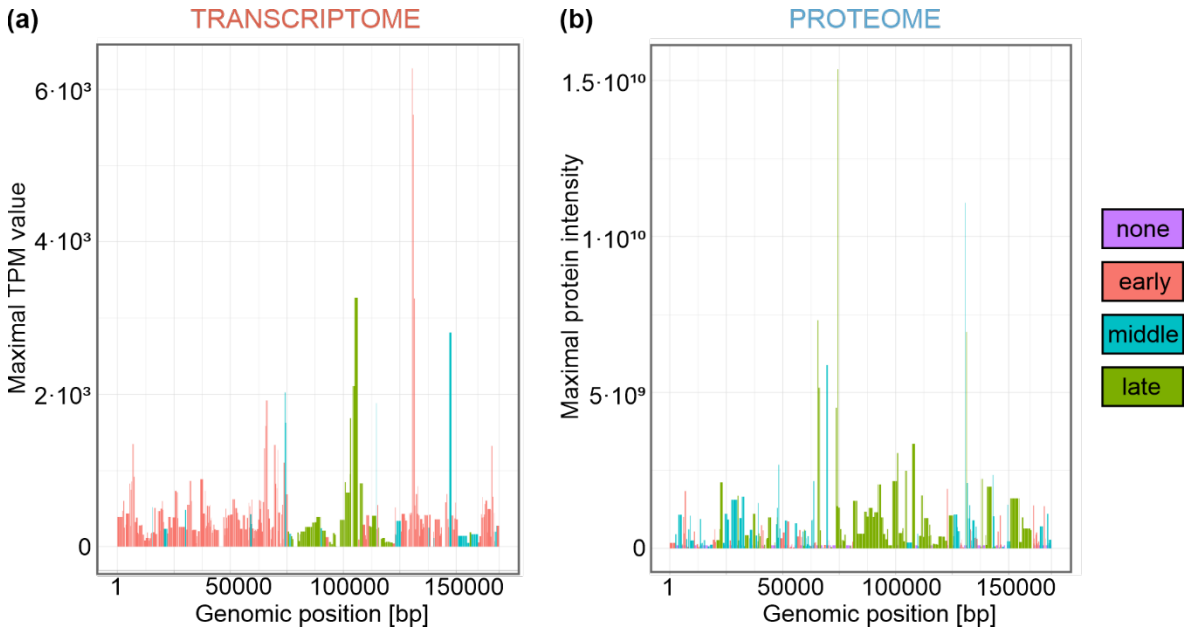


Supplementary Figure S16: SDS-PAGE analysis of the *E. coli* and T4 phage proteome over the time course of infection. Protein samples were taken prior to (t0) and 1 (t1), 3 (t3), 8 (t8), 12 (t12), 20 (t20) and 30 (t30) minutes after T4 phage infection of *E. coli* and subjected to SDS-PAGE analysis followed by protein

visualization by Coomassie staining. No significant changes over the time course of infection can be observed. The analysis was performed in biological triplicates (**a**, **b**, **c**).



Supplementary Figure S17: Relative abundance of T4 phage proteins from different functional groups. Bar plot illustrating the abundance of four major T4 phage protein groups (displayed on x-axis) relative to the most abundant group (structural T4 phage proteins). Absolute abundance of protein groups was calculated as the sum of mean LFQ values for all proteins belonging to a specific functional group. Absolute abundances were normalized to the absolute abundance of structural proteins. Presented data is based on LFQ values from biological triplicates (n = 3).



Supplementary Figure S18: **Genomic maps of T4 phage genome based on transcriptomic- and proteomic-based classifications of T4 phage genes. a**) Transcriptome-based classification showing bar height based on maximal TPM value for a T4 phage gene. **b**) Proteome-based classification showing bar height based on maximal LFQ value for a T4 phage gene-derived protein. X-axis represents the T4 phage genome coordinate in base pairs (bp). Each T4 phage gene is represented by a bar colored according to the infection phase it was classified to. Thus, early T4 transcripts contribute to all classes of T4 phage proteins and make up a major fraction of T4 phage genes. In contrast to that, the proteome is dominated by middle and late T4 phage proteins.

2.7. References

1 Salmond, G. P. & Fineran, P. C. A century of the phage: past, present and future. *Nat Rev Microbiol* **13**, 777-786 (2015). https://doi.org:10.1038/nrmicro3564

2 Chevallereau, A., Pons, B. J., van Houte, S. & Westra, E. R. Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol* **20**, 49-62 (2022). https://doi.org:10.1038/s41579-021-00602-y

3 Lu, T. K. & Koeris, M. S. The next generation of bacteriophage therapy. *Curr Opin Microbiol* **14**, 524-531 (2011). https://doi.org:10.1016/j.mib.2011.07.028

4 Brives, C. & Pourraz, J. Phage therapy as a potential solution in the fight against AMR: obstacles and possible futures. *Palgr Commun* **6** (2020). https://doi.org: https://doi.org/10.1057/s41599-020-0478-4

5 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

6 Rittie, L. & Perbal, B. Enzymes used in molecular biology: a useful guide. *J Cell Commun Signal* **2**, 25-45 (2008). https://doi.org:10.1007/s12079-008-0026-2

7 Gamkrelidze, M. & Dabrowska, K. T4 bacteriophage as a phage display platform. *Arch Microbiol* **196**, 473-479 (2014). https://doi.org:10.1007/s00203-014-0989-8

8 Kutter, E. *et al.* Evolution of T4-related phages. *Virus Genes* **11**, 285-297 (1995). https://doi.org:10.1007/BF01728666

9 Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134-141 (2004). https://doi.org:10.1016/j.virol.2004.08.001

10 Depping, R., Lohaus, C., Meyer, H. E. & Rüger, W. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochem Biophys Res Commun* **335**, 1217-1223 (2005). https://doi.org:10.1016/j.bbrc.2005.08.023

11 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004). https://doi.org:10.1128/JB.186.21.7262-7272.2004

12 Rohrer, H., Zillig, W. & Mailhammer, R. ADP-ribosylation of DNA-dependent RNA polymerase of Escherichia coli by an NAD+: protein ADP-ribosyltransferase from bacteriophage T4. *Eur J Biochem* **60**, 227-238 (1975). https://doi.org:10.1111/j.1432-1033.1975.tb20995.x

13 Koch, T., Raudonikiene, A., Wilkens, K. & Rüger, W. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. *Gene Expr* **4**, 253-264 (1995).

14 Wilkens, K., Tiemann, B., Bazan, F. & Rüger, W. ADP-ribosylation and early transcription regulation by bacteriophage T4. *Adv Exp Med Biol* **419**, 71-82 (1997). https://doi.org:10.1007/978-1-4419-8632-0_8

15 Uzan, M. RNA processing and decay in bacteriophage T4. *Progress in molecular biology and translational science* **85**, 43-89 (2009). https://doi.org:10.1016/S0079-6603(08)00802-7

16 Uzan, M. & Miller, E. S. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. *Virol J* **7**, 360 (2010). https://doi.org:10.1186/1743-422X-7-360

17 Hinton, D. M. Transcriptional control in the prereplicative phase of T4 development. *Virol J* **7**, 289 (2010). https://doi.org:10.1186/1743-422X-7-289

18 Patterson-West, J. *et al.* The Bacteriophage T4 MotB Protein, a DNA-Binding Protein, Improves Phage Fitness. *Viruses* **10** (2018). https://doi.org:10.3390/v10070343

19 Luke, K. *et al.* Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**, 182-191 (2002). https://doi.org:10.1006/viro.2002.1409

20 Kashlev, M., Nudler, E., Goldfarb, A., White, T. & Kutter, E. Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell* **75**, 147-154 (1993).

21 Drivdahl, R. H. & Kutter, E. M. Inhibition of transcription of cytosine-containing DNA in vitro by the alc gene product of bacteriophage T4. *J Bacteriol* **172**, 2716-2727 (1990). https://doi.org:10.1128/jb.172.5.2716-2727.1990

22 Brandao, A. *et al.* Differential transcription profiling of the phage LUZ19 infection process in different growth media. *RNA Biol* **18**, 1778-1790 (2021). https://doi.org:10.1080/15476286.2020.1870844

23 Yang, Z. *et al.* Global Transcriptomic Analysis of the Interactions between Phage phiAbp1 and Extensively Drug-Resistant Acinetobacter baumannii. *mSystems* **4**, e00068-00019 (2019). https://doi.org:10.1128/mSystems.00068-19

24 Mojardin, L. & Salas, M. Global Transcriptional Analysis of Virus-Host Interactions between Phage varphi29 and Bacillus subtilis. *J Virol* **90**, 9293-9304 (2016). https://doi.org:10.1128/JVI.01245-16

25 Leskinen, K., Blasdel, B. G., Lavigne, R. & Skurnik, M. RNA-Sequencing Reveals the Progression of Phage-Host Interactions between phiR1-37 and Yersinia enterocolitica. *Viruses* **8**, 111 (2016). https://doi.org:10.3390/v8040111

26 Kuptsov, N. *et al.* Global Transcriptomic Response of Staphylococcus aureus to Virulent Bacteriophage Infection. *Viruses* **14** (2022). https://doi.org:10.3390/v14030567

27 Guegler, C. K. & Laub, M. T. Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol Cell* **81**, 2361-2373 e2369 (2021). https://doi.org:10.1016/j.molcel.2021.03.027

28 Cowan, J., d'Acci, K., Guttman, B. & Kutter, E. *Gel analysis of T4 prereplicative proteins*. Vol. 4 (American Society of Microbiology, 1994).

29 Kutter, E. *et al.* From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of E. coli. *Viruses* **10**, 387 (2018). https://doi.org:10.3390/v10070387

30 Howard-Varona, C. *et al.* Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *Isme J* **12**, 1605-1618 (2018). https://doi.org:10.1038/s41396-018-0099-8

31 Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *Isme J* **14**, 881-895 (2020). https://doi.org:10.1038/s41396-019-0580-z

32 Gerovac, M. *et al.* A Grad-seq View of RNA and Protein Complexes in Pseudomonas aeruginosa under Standard and Bacteriophage Predation Conditions. *mBio* **12**, e03454-03420 (2021). https://doi.org:10.1128/mBio.03454-20

33 Kaur, S. *et al.* Temporal ordering of omics and multiomic events inferred from time-series data. *Npj Syst Biol Appl* **6**, 22 (2020). https://doi.org:10.1038/s41540-020-0141-0

34 Qi, D., Alawneh, A. M., Yonesaki, T. & Otsuka, Y. Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. *Genetics* **201**, 977-987 (2015). https://doi.org:10.1534/genetics.115.180364

35 Kim, J., Shen, R., Olcott, M. C., Rajagopal, I. & Mathews, C. K. Adenylate kinase of Escherichia coli, a component of the phage T4 dNTP synthetase complex. *J Biol Chem* **280**, 28221-28229 (2005). https://doi.org:10.1074/jbc.M502201200

36 Vigier, P. R. & Marcovich, H. The E. coli mRNA as a precursor for the T4 phage nucleic acids. *Mol Gen Genet* **133**, 353-362 (1974). https://doi.org:10.1007/BF00332711

37 Yang, J. Y. *et al.* Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Syst* **12**, 771-779 e775 (2021). https://doi.org:10.1016/j.cels.2021.05.019

38 Muto, A. *et al.* Structure and function of 10Sa RNA: trans-translation system. *Biochimie* **78**, 985-991 (1996). https://doi.org:10.1016/s0300-9084(97)86721-1

39 Christensen, S. K. & Gerdes, K. RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol Microbiol* **48**, 1389-1400 (2003). https://doi.org:10.1046/j.1365-2958.2003.03512.x

40 Ranquet, C., Geiselmann, J. & Toussaint, A. The tRNA function of SsrA contributes to controlling repression of bacteriophage Mu prophage. *Proc Natl Acad Sci U S A* **98**, 10220-10225 (2001). https://doi.org:10.1073/pnas.171620598

41 Keiler, K. C. & Shapiro, L. tmRNA in Caulobacter crescentus is cell cycle regulated by temporally controlled transcription and RNA degradation. *J Bacteriol* **185**, 1825-1830 (2003). https://doi.org:10.1128/JB.185.6.1825-1830.2003

42 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983). https://doi.org:10.1016/0092-8674(83)90117-4

43 Jarrous, N. & Reiner, R. Human RNase P: a tRNA-processing enzyme and transcription factor. *Nucleic Acids Res* **35**, 3519-3524 (2007). https://doi.org:10.1093/nar/gkm071

44 Babitzke, P. & Romeo, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol* **10**, 156-163 (2007). https://doi.org:10.1016/j.mib.2007.03.007

45 Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E. & Görke, B. The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli. *Nucleic Acids Res* **36**, 2570-2580 (2008). https://doi.org:10.1093/nar/gkn091

46 Hobbs, E. C., Astarita, J. L. & Storz, G. Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in Escherichia coli: analysis of a bar-coded mutant collection. *J Bacteriol* **192**, 59-67 (2010). https://doi.org:10.1128/JB.00873-09

47 Raleigh, E. A. & Wilson, G. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc Natl Acad Sci U S A* **83**, 9070-9074 (1986). https://doi.org:10.1073/pnas.83.23.9070

48 Alawneh, A. M., Qi, D., Yonesaki, T. & Otsuka, Y. An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin-antitoxin module. *Mol Microbiol* **99**, 188-198 (2016). https://doi.org:10.1111/mmi.13225

49 Bingham, R., Ekunwe, S. I., Falk, S., Snyder, L. & Kleanthous, C. The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J Biol Chem* **275**, 23219-23226 (2000). https://doi.org:10.1074/jbc.M002546200

50 Yap, M. L. & Rossmann, M. G. Structure and function of bacteriophage T4. *Future Microbiol* **9**, 1319-1327 (2014). https://doi.org:10.2217/fmb.14.91

51 Serres, M. H. *et al.* A functional update of the Escherichia coli K-12 genome. *Genome Biol* **2**, RESEARCH0035 (2001). https://doi.org:10.1186/gb-2001-2-9-research0035

52 Simon, L. D., Tomczak, K. & St John, A. C. Bacteriophages inhibit degradation of abnormal proteins in E. coli. *Nature* **275**, 424-428 (1978). https://doi.org:10.1038/275424a0

53 Mailhammer, R., Yang, H. L., Reiness, G. & Zubay, G. Effects of bacteriophage T4-induced modification of Escherichia coli RNA polymerase on gene expression in vitro. *Proc Natl Acad Sci U S A* **72**, 4928-4932 (1975). https://doi.org:10.1073/pnas.72.12.4928

54 von Gabain, A. & Bujard, H. Interaction of E. coli RNA polymerase with promotors of coliphage T5: the rates of complex formation and decay and their correlation with in vitro and in vivo transcriptional activity. *Mol Gen Genet* **157**, 301-311 (1977). https://doi.org:10.1007/BF00268667

55 Chamberlin, M., McGrath, J. & Waskell, L. New RNA polymerase from Escherichia coli infected with bacteriophage T7. *Nature* **228**, 227-231 (1970). https://doi.org:10.1038/228227a0

56 Tiemann, B., Depping, R. & Rüger, W. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. *Gene Expr* **8**, 187-196 (1999).

57 Otsuka, Y. & Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol Microbiol* **83**, 669-681 (2012). https://doi.org:10.1111/j.1365-2958.2012.07975.x

58 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336 (2020). https://doi.org:10.1038/s41586-019-1894-8

59 Maurizi, M. Proteases and protein degradation in Escherichia coli. *J Experientia* **48**, 178-201 (1992).

60 Gottesman, S. Proteases and their targets in Escherichia coli. *Annu Rev Genet* **30**, 465-506 (1996). https://doi.org:10.1146/annurev.genet.30.1.465

61 Skorupski, K., Tomaschewski, J., Rüger, W. & Simon, L. D. A bacteriophage T4 gene which functions to inhibit Escherichia coli Lon protease. *J Bacteriol* **170**, 3016-3024 (1988). https://doi.org:10.1128/jb.170.7.3016-3024.1988

62 Lee, I. & Suzuki, C. K. Functional mechanics of the ATP-dependent Lon protease- lessons from endogenous protein and synthetic peptide substrates. *Bba-Proteins Proteom* **1784**, 727-735 (2008). https://doi.org:10.1016/j.bbapap.2008.02.010

63 Hilliard, J. J., Simon, L. D., Van Melderen, L. & Maurizi, M. R. PinA inhibits ATP hydrolysis and energy-dependent protein degradation by Lon protease. *J Biol Chem* **273**, 524-527 (1998). https://doi.org:10.1074/jbc.273.1.524

64 Zhang, K. *et al.* Systemic Expression, Purification, and Initial Structural Characterization of Bacteriophage T4 Proteins Without Known Structure Homologs. *Front Microbiol* **12**, 674415 (2021). https://doi.org:10.3389/fmicb.2021.674415

65 Hinton, D. M. Transcription from a Bacteriophage-T4 Middle Promoter Using T4 Mota Protein and Phage-Modified RNA-Polymerase. *J Biol Chem* **266**, 18034-18044 (1991).

66 Guild, N. *et al.* Transcriptional activation of bacteriophage T4 middle promoters by the motA protein. *J Mol Biol* **199**, 241-258 (1988). https://doi.org:10.1016/0022-2836(88)90311-7

67 Pene, C. & Uzan, M. The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. *Mol Microbiol* **35**, 1180-1191 (2000). https://doi.org:10.1046/j.1365-2958.2000.01787.x

68 Kolesky, S. E., Ouhammouch, M. & Geiduschek, E. P. The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. *J Mol Biol* **321**, 767-784 (2002). https://doi.org:10.1016/s0022-2836(02)00732-5

69 Nechaev, S., Kamali-Moghaddam, M., Andre, E., Leonetti, J. P. & Geiduschek, E. P. The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **101**, 17365-17370 (2004). https://doi.org:10.1073/pnas.0408028101

70 Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M. & Kuriyan, J. Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J Mol Biol* **296**, 1215-1223 (2000). https://doi.org:10.1006/jmbi.1999.3511

71 Nossal, N. G. Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model. *Faseb J* **6**, 871-878 (1992). https://doi.org:10.1096/fasebj.6.3.1310946

72 Mullaney, J. M. & Black, L. W. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J Mol Biol* **261**, 372-385 (1996). https://doi.org:10.1006/jmbi.1996.0470

73 Keller, B., Sengstag, C., Kellenberger, E. & Bickle, T. A. Gene 68, a new bacteriophage T4 gene which codes for the 17K prohead core protein is involved in head size determination. *J Mol Biol* **179**, 415-430 (1984). https://doi.org:10.1016/0022-2836(84)90073-1

74 Sanson, B., Hu, R. M., Troitskayadagger, E., Mathy, N. & Uzan, M. Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J Mol Biol* **297**, 1063-1074 (2000). https://doi.org:10.1006/jmbi.2000.3626

75 Cahova, H., Winz, M. L., Höfer, K., Nubel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

76 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008). https://doi.org:10.1038/nbt.1511

77 Glatter, T. *et al.* Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res* **11**, 5145-5156 (2012). https://doi.org:10.1021/pr300273g

78 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

Chapter III

Shaping the bacterial epitranscriptome – 5′-terminal and internal RNA modifications

Authors:
Schauerte, M.*, **Pozhydaieva, N.***, & Höfer, K.
*contributed equally

Published in:
Advanced Biology, 2021, 5(8), 2100834

Author contributions:

M. S., and N.P.: Conceptualization, Writing – original draft, Writing – review & editing. K. H.: Conceptualization, Funding acquisition, Writing – review & editing.

3. Shaping the bacterial epitranscriptome – 5′-terminal and internal RNA modifications

3.1. Abstract

All domains of life utilize a diverse set of modified ribonucleotides that can impact the sequence, structure, function, stability and fate of RNAs as well as their interactions with other molecules. Today, we know more than 160 different RNA modifications that decorate the RNA at the 5’-terminus or internal RNA positions. The boost of next-generation sequencing technologies sets the foundation to identify and study the functional role of RNA modifications. The recent advances in the field of RNA modifications reveal a novel regulatory layer between RNA modifications and proteins, which has been central to developing a novel concept called “epitranscriptomics”.

The majority of RNA modifications studies focus on the eukaryotic epitranscriptome. In contrast, RNA modifications in prokaryotes are poorly characterized.

This review outlines the current knowledge of the prokaryotic epitranscriptome focusing on mRNA modifications. We describe several internal and 5´-terminal RNA modifications either present or likely present in prokaryotic mRNA. Thereby, the individual techniques to identify these epitranscriptomic modifications, their writers, readers and erasers, and their proposed functions are explored. Besides that, we point out still unanswered questions in the field of prokaryotic epitranscriptomics and outline its future perspectives in the dawn of next-generation sequencing technologies.

3.2. Introduction

The days when RNA was considered to be an oligomer that includes only four nucleosides - adenosine (A), guanosine (G), cytidine (C), and uridine (U)- are a thing of the past. Since Cohn and Volkin discovered the first modified RNA nucleoside pseudouridine in 1951, the list of detected modifications has constantly grown and includes most than 160 known RNA modifications today[1,2]. These nucleotide derivates vary in their complexity, ranging from simple methylations to cyclizations, large group additions, and glycosylations[3]. The versatile and numerous modifications of RNA within a cell are collectively termed “epitranscriptome”[4,5]. The complexity of the epitranscriptome is reflected by the various functions of the post- and co-transcriptional modifications, which can affect RNA structure, stability, and cellular fate[6]. Moreover, the functional role of a particular modification can vary between different RNA classes, such as mRNA, rRNA or tRNA. Understanding the RNA modifications' biological function and their discovery and localization in the transcriptome are of significant importance.

Initially, to characterize RNA modifications, the transcripts carrying a particular modification were digested to single nucleotides and analyzed by biophysical methods, such as chromatography-

and UV-based approaches[7]. However, these techniques do not allow for the identification of the position of such modified nucleotides in the transcriptome.

The emergence of DNA sequencing approaches, specifically the development of second-generation polymerase-based methods, referred to as next-generation sequencing (NGS), revolutionized the field of epitranscriptomics. In addition to DNA, RNA can be analyzed - termed RNA-Sequencing (RNA-Seq). Here, RNA is converted into cDNA by reverse transcription (RT), and the latter is subsequently subjected to high-throughput sequencing. In such a way, quantitative transcript profiling with single-base resolution can be achieved[8,9]. The sensitivity can be improved by antibody-, enzymes-, and chemical derivatization-based enrichment approaches applied before NGS. Also, first attempts have been made to adapt the Oxford Nanopore technology to identify several modifications within one transcript simultaneously. These technological advances give rise to study the epitranscriptome on a new level[10-12]. However, mostly all techniques were applied to study the eukaryotic epitranscriptome yet. In contrast, the role and distribution of prokaryotic RNA modifications remain poorly explored to this day.

In light of the discovery of the bacterial and eukaryotic epitranscriptome several proteins were identified to be the key players for the biosynthesis (writer proteins), the recognition (reader proteins) and the specific removal (eraser proteins) of RNA modifications. Their identification and characterization allow us to understand the regulatory processes controlling the presence of RNA modifications.

This review focuses on the internal and 5´-terminal RNA modifications in bacteria and archaea that exist or are very likely to be present in prokaryotic mRNA (Figure 1). Among the internal modifications, we present here N^6-methyladenosine (m^6A), and inosine (I) that are known to decorate prokaryotic mRNA[13-15]. Moreover, pseudouridine (Ψ), 5-methylcytosine (m^5C), and 2´-O-methylated nucleotides (Nm) are presented as highly abundant modifications of eukaryotic mRNA, whereupon their presence in prokaryotic mRNA was not verified yet. In the case of 5’-terminal RNA modifications, besides the well-known 5´-triphosphorylated (5´-PPP), 5´-diphosphorylated (5´-PP), 5´-monophosphorylated (5´-P), and 5´-hydroxylated (5´-OH) RNA, we address several *in vivo* or *in vitro* verified non-canonical RNA-caps. Here, Nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD), coenzyme A (CoA), Uridine diphosphate N-acteyl glucosamine (UPD-GlcNAc), thiamine, dinucleoside polyphosphate (Np_nN), and 5´-phospho-ADP-ribose capped RNAs are described. We present the current state of knowledge regarding the approaches developed to detect these 5’-terminal and internal modifications in a high-throughput and transcriptome-wide manner. Further, we summarize the known and suggested biological functions of each RNA modification and review its writers, readers and

erasers. Finally, we address the numerous open questions in the field of prokaryotic epitranscriptomics.

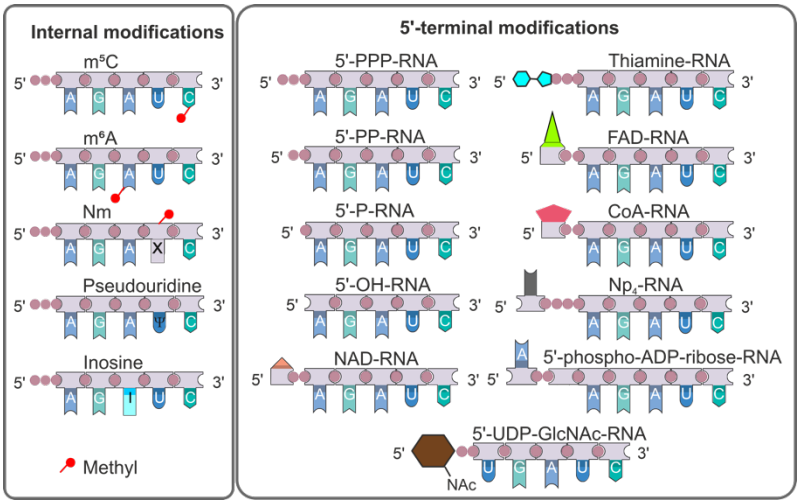


Figure 1: Overview of internal and 5’-terminal RNA modifications discussed in this review.

3.3. Internal RNA modifications

The non-canonical RNA nucleosides m^6A, I, Ψ, m^5C, and Nm are widely distributed in all kingdoms of life[16]. However, their role and distribution in different RNA classes (tRNA, mRNA, rRNA) are not entirely explored. All internal RNA modifications listed above were detected in eukaryotic mRNA, and their functions are partially elucidated. Only m^6A and inosine were identified in prokaryotic mRNA to date[13-15].

Over the past years, different approaches were established to detect these internal RNA modifications. The bottleneck for the detection of internal RNA modifications is usually their low abundance in mRNAs. Thus, sequencing strategies are needed that ensure high-sensitivity and single-base resolution to map the modifications to the transcriptome.

Such methods often include an enrichment step of modified transcripts prior to sequencing. Generally, enrichment methods are based on antibodies, interaction proteins (writer/reader/eraser) and chemical derivatization (Figure 2). Multiple sequencing approaches, including an enrichment or derivatization step, were established for internal RNA modifications discussed in this review. However, the developed strategies were mainly applied to detect the modified transcripts within the eukaryotic epitranscriptome.

In this section, we outline the current state of knowledge on internal RNA modifications. In particular, we highlight m^6A, I, Ψ, m^5C, and Nm mRNA modifications in prokaryotes. We discuss

the established high-throughput sequencing approaches that can be applied to explore the presence and functions of internal bacterial mRNA modifications.

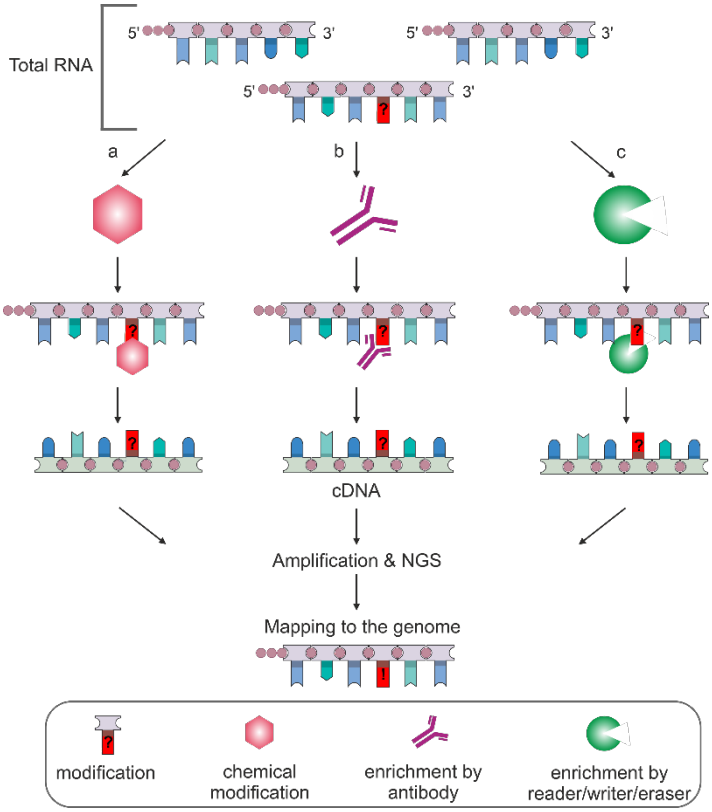


Figure 2: Strategies for the enrichment and detection of internal RNA modifications. Approaches such as (a) chemical derivatization, (b) enrichment by modification-specific antibodies or (c) writer/reader/eraser-proteins are widely applied to enrich for modified transcripts. The enriched transcripts are converted into cDNA, submitted to NGS, and the reads are mapped to the genome to identify the transcripts carrying an internal RNA modification as well as its position.

Pseudouridine Ψ (5-ribosyluracil)

Often referred to as the fifth nucleotide, Ψ is one of the most abundant and the first discovered RNA modification present in all domains of life[1,17] (Figure 3). This modification is equally distributed in eukaryotic cells throughout various RNA classes[18-20]. In prokaryotes, the appearance of Ψ is restricted to tRNA and rRNA[21,22].

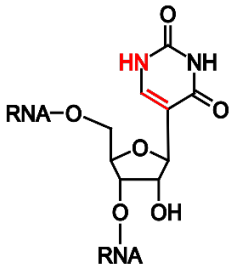
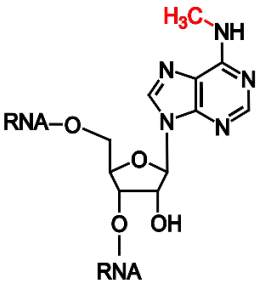
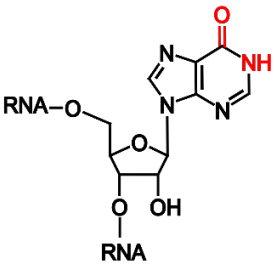
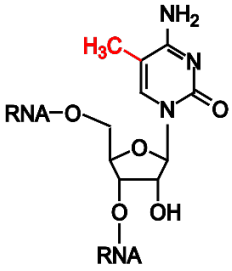
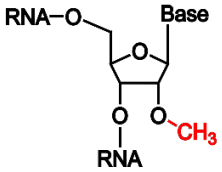
Modification	Structure	Detection	Writers, readers, erasers	Molecular roles
Pseudouridine (Ψ)		Pseudo-Seq Ψ–Seq Psi-Seq CeU-Seq RBS-Seq HydraPsiSeq	Writers: site-specific PUSs and snoRNA guided box H/ACA PNPs Eukaryotic reader: MetRS	Structure stabilization of tRNA and rRNA, translation modulation, alteration of codon translation, mRNA stability
N^6-methyladenosine (m^6A)		m^6A-Seq MeRIP-Seq m^6A-LAIC-Seq PA- m6A-Seq mi-CLIP m^6A-label-Seq DART-Seq NanoCompore	Writers: RlmF, RlmJ (prokaryotes), METTL3, METTL4, WTAP (eukaryotes) Eukaryotic erasers: FTO, ALKBH5 (eukaryotes) Eukaryotic readers - YTHDF1-3ly proteins	In eukaryotes involved in mRNA stability, RNA-protein interactions, miRNA processing, splicing, translation initiation and efficiency regulation.
Inosine (I)		ICE-Seq EndoVIPER-Seq	Writers: adenosine deaminases that act on rRNA (ADAR) or tRNA (ADAT) in eukaryotes Prokaryotic homologous – TadA, acts on tRNA and mRNA Reader: endonucleaseV	Codon alterations, alternative splicing, regulation of stress response, alternative splicing and 3´-UTR-variation in eukaryotes
5-methylcytidine (m^5C)		Bisulfite sequencing RIP-Seq Aza-IP NSUN2-RNA cross-linking based miCLIP TAWO-Seq	Writers: eukaryotic NSUN and DNMT2 methyltransferases Erasers - eukaryotic TET enzymes remove the modification by its further oxidation to hm^5C Eukaryotic readers: ALYREF, YBX1, YPS	tRNA structure stabilization, translational fidelity of ribosome
2´-O-methylation (Nm)		RT with limited dNTPs 2OMe-Seq RiboMeth-Seq Nm-Seq RibOxi-Seq	Writers - stand-alone methyltransferases or snoRNA-guided 2´-O-methylases	transcript stability, inhibition of ribosomal protein translation, translation efficiency

Figure 3: An overview about internal RNA modifications in eukaryotes and prokaryotes.

Identification, quantification and validation of pseudouridylation

Initially, the screening of RNA for Ψ was a two-step procedure – RNA was digested with a nuclease and analyzed via chromatography-based approaches[23,24]. In such a way, the specific pseudouridylation sites were identified for the first time within yeast tRNA[25]. Nonetheless, such methods are applicable only for abundant and purified transcripts but are insufficient to determine the cellular Ψ landscape. Triggered by the development of NGS techniques, post-transcriptional modifications such as Ψ were detected at single-base resolution.

Several NGS approaches for Ψ detection were established, which are based on the formation of covalent adducts between Ψ and 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate (CMCT), among them Pseudo-Seq[26], Ψ-Seq[27], PSI-Seq[28], and CeU-Seq[29]. The derivatized N^3-CMC-Ψ residues terminate the RT and cause the formation of the abortive cDNA products. The subsequent NGS of the cDNA enables the identification of positions where RT stop was mediated by the presence of N^3-CMC-Ψ. Thus, CMCT derivatization of RNA was successfully performed to confirm the presence of Ψ35 in *Escherichia coli* $tRNA^{Tyr}$ [30]. However, besides all advantages of CMCT derivatization, low abundant pseudouridylation events may be missed. The presence of N^3-CMC-Ψ terminates the RT. Therefore, only one pseudouridylation site can be determined within one transcript.

A quantitative transcriptome-wide Ψ profiling with single base-resolution is achieved by applying the RBS (RNA bisulfite)-Seq approach[31]. Khoddami *et al.* use a Ψ-monobisulfite addition with subsequent heat-induced ribose ring-opening and Mg^{2+}-assisted reorientation that causes base-skipping during cDNA synthesis. Further RNA modifications, such as m^5C and m^1A, can be identified simultaneously with Ψ in this approach. However, RBS-Seq was so far applied for the evaluation of the eukaryotic transcriptome. Due to this approach's high sensitivity, the application of RBS-Seq on the prokaryotic transcriptome can be considered. Hence, the presence of Ψ in prokaryotic tRNA, mRNA and rRNA needs to be investigated.

The quantitative information about the pseudouridylation levels at a specific position of the RNA molecule can improve the understanding of the functional role of Ψ. Recently, the so-called HydraPsiSeq was established and verified for mapping and absolute quantification of Ψ residues in *Saccharomyces cerevisiae* and *Homo sapiens*[32]. This approach is based on hydrazine mediated RNA cleavage at uridine followed by subsequent RNA hydrolysis at abasic sites using an aniline treatment. Ψ-modified residues are resistant towards hydrazine-dependent cleavage. After sequencing, U cleavage profiles are generated and used to determine the position of Ψ and the pseudouridylation level. The comparison of the latter parameter in different samples enables

the determination of constitutive and variable Ψ sites. Interestingly, HydraPsiSeq requires low amounts of RNA (as low as 10–50 ng), making it compatible with single-cell RNA sequencing strategies.

Based on the knowledge gained from all the NGS-based methods developed for the detection of Ψ, several bioinformatical tools for the prediction of Ψ were reported within the last few years[33-37]. However, the majority of methods were developed for the identification of Ψ in eukaryotes. Studies focusing on prokaryotic pseudouridylation are still missing.

Writers, erasers and functions of Ψ

The isomerization of uracil to Ψ is post-transcriptionally catalyzed by so-called pseudouridine synthases (PUSs) in bacteria[38]. For example, the prokaryotic model organism *E. coli* contains 11 pseudouridine synthases, of which four are tRNA specific, six – rRNA specific, and one is observed to catalyze the pseudouridylation on both tRNA and rRNA[39].

Independent of the pseudouridine synthase class, generally, two distinct reactions occur during U to Ψ isomerization – the cleavage of N^1-$C^{1'}$ glycosidic bond that allows for a base rotation by 180° as well as the subsequent formation of C^5-$C^{1'}$ glycosidic bond[40]. The resulting nucleotide possess partially changed chemical properties and altered biophysical characteristics. As a result, the introduction of Ψ into rRNA and tRNA provides stabilization and enables the formation of highly ordered RNA structures. First, the phosphodiester backbone's thermodynamic stability increases through the water-mediated hydrogen bond network[41]. Second, the RNA duplex stability is improved by enhanced base stacking between the Ψ-A base pair[42,43]. Within rRNA, Ψ is usually located at functionally essential areas. The abolishment of Ψ in rRNA results in reduced translation rates and increased stop-codon readthrough in yeast[44]. Moreover, in *E. coli*, abolished Ψ levels are known to inhibit the translation[44,45]. In eukaryotes, pseudouridylation of mRNA can modulate important cellular processes, in particular translation. Ψ-containing codons affect the translation of cognate codons[46]. Interestingly, the translation efficiency of proteins increases in the presence of Ψ containing mRNA in eukaryotes[47]. On the other hand, in a fully reconstituted *E. coli* translation system, the presence of Ψ in mRNA was observed to reduce the overall protein production [46]. However, the presence of Ψ in bacterial mRNA was not verified yet. Thus, future studies have to analyze the influence of pseudouridylation on the bacterial translation *in vivo*.

Although the Ψ writers are known, its erasers which convert Ψ to U or remove it from RNA are not reported[46]. However, the first Ψ reader protein, methionine aminoacyl $tRNA^{Met}$ synthetase (MetRS), was recently discovered in *S. cerevisiae*[48]. Here, MetRS interacts with pseudouridylated

tRNA, which triggers the interaction between the tRNA charging and mRNA translation. Hence, it would be exciting to investigate the presence of analogous readers within prokaryotes.

N^6-methyladenosine (m^{6}A)

N^6-methyladenosine (m^{6}A) RNA methylation was first observed by Perry and Kelley in 1974 and later recognized to be one the most abundant post-transcriptional mRNA modifications. m^{6}A is widely distributed in the eukaryotic, prokaryotic, and the viral transcriptome[49,50] (Figure 3). While in eukaryotes, m^{6}A modification is present in most classes of RNA, including mRNA, ncRNA, rRNA, and tRNA[51], in bacteria, only rRNA, mRNA, and ncRNA were observed carrying m^{6}A modification so far[13].

Identification, quantification, and validation of m^{6}A

The m^{6}A has almost identical chemical properties as its precursor adenosine. Therefore, selective chemical derivatization of either A or m^{6}A, which induce RT aberration, is not possible. For this reason, methods for transcriptome-wide m^{6}A mapping based on principles other than chemical derivatization are required.

The first developed transcriptome-wide adenosine methylome mapping approaches, such as m^{6}A-Seq, MeRIP (m^{6}A methylated RNA immunoprecipitation)-Seq, and m^{6}A-LAIC (m^{6}A-level and isoform-characterization)-Seq, were based on the immunoprecipitation of m^{6}A containing transcripts with an anti-m^{6}A antibody followed by subsequential high-throughput sequencing[52-54]. However, all three mentioned approaches need high amounts of isolated total RNA and are strongly dependent on the specificity of the antibody. Thus, the unspecific binding of the antibody can result in false-positive m^{6}A mapping. Furthermore, the antibody-based approaches have a resolution of ~100-200 nt, which hampers the m^{6}A detection at single-base level. The low resolution of m^{6}A mapping can be improved to ~23 nt by the application of PA-m^{6}A (photo-crosslinking-assisted m^{6}A)-Seq[55]. Here, the artificial ribonucleoside 4-thiouridine (4SU) is supplied to the cells and subsequently incorporated into the RNA *in vivo*. The RNAs containing 4SU are covalently cross-linked to the in proximity bound anti-m^{6}A-antibody. Afterwards, RNA, which is not protected by the antibody, is removed by RNase treatment. Subsequently, the antibody is cleaved from the RNA, which is subjected to NGS. However, this approach is limited by the specificity of the anti-m^{6}A-antibody and the incorporation efficiency of 4SU into the RNA. The detection of m^{6}A at single-base level can be improved by a photo-crosslinking-based method, the so-called miCLIP (m^{6}A individual-nucleotide-resolution cross-linking and immunoprecipitation)-Seq approach[56]. Here, the cross-linked antibody is partially removed from

the target transcript by protease digest. The remaining covalent adduct causes truncated cDNAs or the transition of m^6A-adjacent C to T during RT. Similar to the described methods, mi-CLIP requires an anti-m^6A-antibody. Thus the specificity of the latter is crucial for the correct identification of m^6A sites.

2015, Deng *et al.* combined the PA-m^6A approach with ultra-high-performance liquid chromatography coupled with subsequential mass spectrometry[13]. This technique allowed for the specific identification of m^6A sites within prokaryotic mRNA as well as for the quantification of m^6A/A ratios. The latter varied between 0.02 %−0.28 % in different bacterial strains being generally higher in Gram-negative bacteria. Moreover, the authors identified a consensus motif (GCCAG) for prokaryotic mRNA m^6A modifications different from the eukaryotic one (RRACU). Furthermore, in bacteria, the m^6A/A ratio was observed to be stable at various growth conditions. This observation led to the suggestion that the m^6A mRNA modification in bacteria can alter the stability of the mRNA rather than playing a role in regulatory processes[13]. However, no further studies were performed that determines the precise role of m^6A in prokaryotic mRNA to date. The lack of m^6A data for prokaryotic organisms might be due to low m^6A levels. The detection of low m^6A levels is challenging as the N^6- methyl group is chemically unreactive and difficult to label. In contrast to other bulky RNA modifications, m^6A does not cause nucleotide misincorporation or termination during RT, which would allow for the determination of m^6A at single-base resolution.

In 2020, Shu *et al.* developed the m^6A-label-Seq approach to address these issues[57]. m^6A-label-Seq is a single-base resolution technique. Cells are grown in the presence of Se-allyl-l-selenohomocysteine (SeAM). Eukaryotic m^6A writers accept SeAM as a methyl donor cofactor, thereby transferring the allyl group to adenosine, generating N^6-allyladenosine (a^6A) instead of m^6A. Chemically induced cyclization of a^6A results in a nucleotide misincorporation during RT allowing for the detection of m^6A at single-base level[57]. So far, m^6A-label-Seq was applied to human cell lines. By feeding the cells with SeAM, only ten percent of the naturally occurring m^6A sites were substituted with a^6A[57]. In contrast to other m^6A detection methods, m^6A-label-Seq is RNA motif-independent and can be used to identify clustered m^6A sites compared to the methods described before. This approach can potentially be applied to study the role of prokaryotic m^6A mRNA. However, the acceptance of the SeAM by prokaryotic methyl transferases and the uptake of this artificial cosubstrate by the prokaryotic cell has to be studied first.

The constraints coincide with metabolic m6A labeling strategies that can be circumvented by transiting to chemical labeling approaches, such as m^6A-SEAL-Seq[58]. m^6A-SEAL-Seq utilizes the human originated m^6A eraser FTO, which oxidizes m^6A to N^6-hydroxymethyladenosine (hm^6A). The latter can be further converted to N^6-dithiolsitolmethyladenosine (dm^6A) by a dithiothreitol

(DTT)-mediated thiol-addition reaction. The resulting dm^{6}A possesses a free sulfhydryl group that facilitates the simple introduction of functional application tags such as biotin, allowing for the subsequent enrichment of modified transcripts.

Other antibody-free methods to detect m^{6}A containing mRNAs are based on the application of m^{6}A-sensitive RNA endoribonucleases. For instance, such approaches as MAZTER-Seq and m^{6}A-sensitive RNA-endoribonuclease–facilitated (m^{6}A-REF)-Seq are both utilizing an *E. coli* originating RNA endonuclease and toxin MazF[59,60]. MazF specifically cleaves RNA at unmethylated sites upstream of the ACA sequence motif[61]. At the same time, the methylated counterparts “m^{6}A-CA” remain unrecognized and intact. Both approaches allow for the quantification of the methylation levels at single-nucleotide resolution. For instance, m^{6}A-REF-Seq works sufficiently on minimal amounts of input material (ng-pg of RNA)[60]. However, the major limitation of both MAZTER-Seq and m^{6}A-REF-Seq is the restriction m^{6}A modification sites located at ACA motifs.

All the described methods require an RT step and a PCR amplification prior to high-throughput sequencing, which introduces bias in sample composition[62]. Thus, technologies are needed that allow for the direct sequencing of RNA as well as RNA modifications. Recently, Nanopore direct RNA sequencing (DRS) was developed. In the case of Nanopore sequencing, a single RNA molecule is guided through a protein pore embedded in a synthetic membrane. A nucleotide or its derivative specifically affects the ion flow through the membrane by passing across, causing a change of the measured electric signal[11]. Recently, the DRS method was applied to identify post-transcriptional RNA modifications[11,12]. Leger *et al.* developed a NanoCompore approach for m^{6}A mapping in the mammalian transcriptome[12]. In this approach, two DRS datasets (non-treated control sample versus demethylated sample) are compared. The changes in signal levels within both datasets indicate the presence or absence of RNA modifications. Unfortunately, to date, only low throughput sequencing was achievable using this approach. Improved DRS based techniques can revolutionize the epitranscriptomics as they enable the analysis of the modification at RNA level.

Writers, erasers and functions of m^{6}A

While in eukaryotes, m^{6}A is the best-studied mRNA modification, in prokaryotes, the functional role of m^{6}A is significantly less explored. The m^{6}A abundance within mRNA and its regulation differ in the bacterial and eukaryotic world.

To date, only two bacterial m^{6}A writers RlmF and RlmJ are described. These enzymes methylate adenosines in rRNA and tRNA[13]. The deletion of *rlmF* and *rlmJ* genes does not affect the level of

mRNA methylation in bacteria. Hence, RlmF and RlmJ do not belong to the class of mRNA methyltransferases[13]. In prokaryotes, the m^6A modification is prevalently abundant within open reading frames (ORFs) (72 %) and rarely present at the termini of mRNA[13]. Both the function and the regulation of m^6A modifications in bacterial mRNA are entirely unexplored.

In contrast, eukaryotic m^6A writers, erasers and readers are well studied, and the function of m^6A is well characterized. In eukaryotic mRNA, m^6A is enriched around stop codons, within 3‘-untranslated regions (3´-UTRs) and long internal exons[52,63]. Eukaryotic m^6A modifications are described to be involved in mRNA stability, RNA-protein interactions, miRNA processing, splicing, initiation, and regulation of translational efficiency[8,13,64,65]. The methylation of adenosine in eukaryotes is highly regulated by writers METTL3 and METTL14[66,67]. FTO and ALKBH5 were identified being the eukaryotic erasers of m^6A[68,69]. Eukaryotic YTHDF1-3 proteins are the readers and selectively bind to m^6A mRNA and promote its decay and regulate the translation[70,71]. The discovery of m^6A regulatory proteins enables the development of novel methods to study m^6A modified transcripts. For instance, the eukaryotic m^6A reader, YTH, was successfully applied for m^6A detection in an approach termed DART (deamination adjacent to RNA modification targets)-Seq[72]. DART-Seq is an antibody-free method that detects global m^6A modification, using a fusion of the cytidine deaminase APOBEC1 to the m^6A-binding YTH domain. APOBEC1-YTH expression results in the deamination of C to U adjacent to the m^6A sites. Subsequently, the total RNA is isolated and subjected to RNA-Seq. m^6A sites are identified that are close to C-to-U mutations. The main advantage of DART-Seq is the low amount of RNA (approx. 10 ng) needed for the analysis of m^6A. However, this method is still restricted to eukaryotic organisms as prokaryotic m^6A reader proteins have not been discovered yet.

Inosine (I)

Inosine (I) is one of the most prevalent post-transcriptional RNA modifications and was identified as a biosynthetic precursor of a purine by Warren *et al.* in 1957[73]. However, in addition to its appearance in the cell as a biosynthetic intermediate, inosine was identified as a deamination product of adenine within DNA or RNA (Figure 3). While the presence of inosine in DNA results in genetic material damage[74], I-to-A modification in RNA is significant and essential for its biological function.

Identification, quantification and validation of inosine

Initially, the mapping of inosine was based on the modification-induced change in base-pairing. Inosine forms a Watson-Crick type base pair and a wobble pair with uridine and is read as guanine

by reverse transcriptase[75]. According to this property, high-throughput sequencing of cDNA is performed, allowing for identifying inosine sites. However, in this approach, the efficiency of inosine identification is limited by the abundance of modified adenosines and the rate of I/A interconversion. A background noise resulting from single-nucleotide polymorphisms, somatic mutations, pseudogenes, and sequencing errors impede inosine positioning as well[76].

Higher accuracy in the mapping of inosine was achieved by ICE (inosine chemical erasing)-Seq[77]. The main principle of this approach is the derivatization of inosine with acrylonitrile to N^1-cyanoethylinosine that inhibits Watson-Crick base pairing and subsequently terminates the RT reaction. The truncated cDNAs are not amplified by PCR and therefore not in the NGS analysis. However, ICE-Seq is limited in its sensitivity. This issue can be solved by enrichment of transcripts containing inosine prior to RNA-Seq.

Recently, the enrichment strategy for inosine containing RNA was established as a part of so-called EndoVIPER (endonuclease V inosine precipitation enrichment)-Seq[78,79]. In this approach, Knutson *et al.* applied the *E. coli* inosine reader EndonucleaseV (eEndoV) to enrich inosine-containing transcripts. eEndoV is known to be a conserved nucleic acid repair enzyme that specifically recognizes and binds to inosine in DNA and promotes the base excision repair [80]. Interestingly, eEndoV is highly specific for inosine in RNA and exhibits a low nanomolar binding affinity without displaying the endonuclease activity in the presence of calcium. The specificity of eEndoV for inosine containing transcripts was successfully applied to identify inosine in the eukaryotic transcriptome. However, EndoVIPER-Seq has not yet been applied to investigate the presence of inosine in the prokaryotic transcriptome . As EndoVIPER-Seq utilizes the prokaryotic inosine reader eEndoV, the application of this technique can enable the identification of previously unknown inosine sites.

Writers, erasers and functions of inosine

The C6 deamination of adenosine to inosine in RNA can occur either by spontaneous hydrolysis or enzymatic conversion[81]. Spontaneous deamination can be caused by environmental factors, such as exposure to nitrosative compounds[81]. The enzymatic mediated A-to-I editing is specifically catalyzed by adenosine deaminases[81]. Within eukaryotes, adenosine deaminases are classified into adenosine deaminases acting on tRNA (ADAT) and adenosine deaminases acting on mRNA (ADAR)[81-83]. Both deaminase classes possess double-stranded RNA binding domains [84,85]. Prokaryotic TadA protein was initially identified as a bacterial ADAT analogue. For a long time, its deamination activity was assumed to be restricted to tRNA[86], precisely $tRNA^{Arg}_{ACG}$ and $tRNA^{Leu}_{AAG}$ [87]. However, a few years ago, the targets of TadA in *E. coli* were expanded by 15

mRNA sites, which confirmed the presence of inosine in prokaryotic mRNA for the first time[14]. One of the identified TadA targets, *hokB*, is coding for a self-killing toxin in *E. coli*. The A-to-I modification of the hokB transcript results in the translation of a HokB protein isoform that displayed enhanced toxicity connected to higher bacterial persistence and antibiotic resistance[14]. In 2020, Nie *et al.* proceeded with the discovery of inosine in prokaryotes [15]. In this study, 30 inosine sites in the mRNA of *Xanthomonas oryzae pv. oryzicola* were identified by bioinformatical approaches. Later these sites were verified by immunoprecipitation of modified RNA and its subsequential sequencing (iRIP-Seq)[15]. During oxidative stress conditions, A-to-I editing within the fliC transcript was observed. This post-transcriptional modification caused an S128P mutation in the respective protein that is essential for flagellum formation. Surprisingly, the S128P mutation in FliC was identified to increase bacterial tolerance towards oxidative stress and induce biofilm formation.

The recent studies analyzing A-to-I editing in prokaryotic mRNA demonstrate that adenosine deamination in bacterial RNA has a significant regulatory function.

5-methylcytidine (m^5C)

5-methylcytidine (m^5C) is a modification initially identified in DNA and later observed in RNA[88]. The methylation of the cytosine does not affect the Watson-Crick base pairing. Still, it impacts the base's biophysical properties, increasing its hydrophobicity and affecting the base stacking[89,90] (Figure 3). While the m^5C modification of rRNA is distributed over all domains of life, the methylation of cytidine within tRNA and mRNA was observed only in eukaryotes and archaea[91-95]. In eukaryotic mRNA, m^5C was described to be involved in the modulation of mRNA nuclear export and regulation of protein translation[96,97].

Identification, quantification and validation of 5-methylcytidine

The first discovery of m^5C was triggered by the application of different biophysical technologies, among them chromatography-based approaches, mass spectrometry (MS), and nuclear magnetic resonance (NMR)[98]. However, all these methods are not applicable for high-throughput sequencing and single-base m^5C mapping.

The development of NGS approaches triggered the establishment of methods to explore the cytidine methylome. In 2009, the bisulfite sequencing technique, earlier applied to study m^5C in DNA, was adapted by Schäfer *et al.* to specifically map m^5C RNA modifications[99,100]. In the presence of the nucleophilic reagent bisulfite, cytosine residues are selectively deaminated to uracil, while methylated cytosines remain unaffected. cDNA synthesis followed by PCR

amplification and NGS allowed for the detection of unmethylated and methylated cytosines in tRNAs, rRNAs and mRNAs. Using this method, a base-resolved localization of the m^5C can be achieved. However, other methylcytidine modifications such as 3-methylcytidine, N4-methylcytidine, N4,2′-O-dimethylcytidine and N4-acetylated variants can prevent the C-to-U conversion by bisulfite[100]. Bisulfite-Seq was applied to identify m^5C modification in archaeal mRNA (*Solfolobus solfataricus*), whereupon the presence of m^5C in bacterial mRNA could not be verified with this approach[92].

To improve the quality of m^5C detection, alternative m^5C RNA mapping approaches based on the selective enrichment of modified transcripts before NGS were developed. Approaches such as Aza-IP (5-azacytidine-mediated RNA immunoprecipitation)[101] and miCLIP (methylation individual-nucleotide-resolution crosslinking and immunoprecipitation)[102] utilize eukaryotic m^5C RNA writers NSUN2 and DNMT2. The Aza-IP method includes metabolic RNA labelling[101] and is based on the ability of m^5C methyltransferases to form a covalent enzyme-substrate intermediate with their targets[103]. In this approach, 5-azacytidine (5-azaC) is incorporated into the RNA *in vivo*. The applied cytidine analogue is a suicide inhibitor of the m^5C RNA methyltransferase. Thereby it induces the formation of a stable RNA methyltransferase-RNA adduct, which can be enriched by the application of a writer-specific antibody. Aza-IP allows for the identification of RNA substrates of m^5C RNA methyltransferases NSUN2 and DNMT2. However, m^5C modifications sites that are independent of NSUN2 and DNMT2 are not detected. Furthermore, 5-azaC is toxic to cells, and the substitution rate of cytidine to the 5azaC is low[104,105].

miCLIP does not require metabolic labeling of RNA and is based on the application of a modified m^5C writer[102]. Here, an irreversible covalent bond between the m^5C writer and the RNA is formed by crosslinking. Similar to Aza-IP, the resulting complex can be immunoprecipitated with a specific antibody.

Nevertheless, the applicability of both approaches for the analysis of the prokaryotic transcriptome must be verified. To successfully apply the Aza-IP or miCLIP approach, the writer protein that forms a covalent bond with their substrate cytidine has to be identified. However, to date, the writers for the m^5C modification in prokaryotes remain unknown.

Apart from the immunoprecipitation of the writer-RNA complex, the m^5C containing RNA can be directly enriched by RIP (RNA immunoprecipitation)-Seq[92]. In this approach, an m^5C-specific antibody is applied. However, similar to all other antibody-based techniques, RIP-Seq is limited by the antibody specificity towards m^5C modification[94]. Also, the m^5C antibody cannot distinguish between m^5C present in DNA and RNA.

To date, several approaches were established to analyze the cellular cytidine methylome. However, their application to explore the presence and biological role of the m^5C modification in the prokaryotic transcriptome is still under investigation.

Writers, erasers and functions of 5-methylcytidine

The introduction of m^5C modification into eukaryotic RNA is catalyzed by the SAM-dependent methyltransferases NSUN (NSUN1-NSUN7) and DNMT2[103,106,107]. Next to these writers, the proteins involved in the removal of m^5C are identified in eukaryotes. For instance, ten-eleven translocation (TET) enzymes were reported to erase the m^5C from RNA by its oxidation to 5-hydroxymethyl cytidine (hm^5C)[108]. Recently, eukaryotic m^5C readers ALYREF, YBX1 and YPS were identified[96,109]. All of them exhibit preferential binding to m^5C-containing mRNA and promote its cytoplasmic export. Based on these results, m^5C modification of RNA was assumed to be involved in regulating the cellular RNA localization and translation in eukaryotes.

However, in contrast to eukaryotes, the m^5C existence in bacterial mRNA was still not proved. Therefore, writers, readers and erasers of m^5C were not identified in bacteria so far.

2´-O-Methylation (Nm)

Compared to the already described post-transcriptional RNA modifications, 2´-O-methylation (Nm) is an RNA base-unspecific modification and the most common modification of the RNA ribose moiety (Figure 3). As described for the other internal RNA modifications, 2´-O-methylation remained cryptic for a long time. Recently, the analysis of NGS data sets gave rise to the physiological function of Nm. So far, the involvement of Nm in RNA structure modulation, innate immunity and translation regulation is suggested[110-113]. The methylation of the ribose 2´-OH of tRNA and rRNA occurs widely in all domains of life. In eukaryotes, this modification can be found in mRNA, rRNA, tRNA, and snRNA, whereas in prokaryotes the presence of Nm is restricted to rRNA and tRNA[111,114,115].

Identification, quantification and validation of 2´-O-methylation

The Nm modification influences the biophysical properties of a modified nucleoside. The modification of 2´-OH to 2´-O-Me group increases the hydrophobicity of transcripts but decreases their nucleophilicity[116]. The increase of RNA stability in the presence of alkaline conditions was utilized to develop a high-throughput Nm mapping approach called RiboMeth (ribose methylation)-Seq[117]. This protocol is based on random fragmentation of phosphodiester bonds under mild alkaline conditions. The presence of a 2′-O-Me group protects the 3′-adjacent

phosphodiester bond from nucleolytic hydrolysis, resulting in a changed cleavage pattern. RiboMeth-Seq was successfully applied to identify 2’-O-Me-sites in rRNA of the budding yeast and to explore the stress-induced modulation of 2´-O-methylation in *E. coli* tRNA[118]. In combination with LC-MS/MS, the 2´-O-methylation levels were determined for different growth conditions. In this study, the abundance of 2´-O-methylation increased during starvation and antibiotic stress conditions. Interestingly, the introduction of Nm into tRNAs improved their stability during stress conditions[118].

In 2017, two Nm mapping technologies based on the different chemical properties of 2´-OH nucleosides and Nm were reported, namely Nm-Seq and RibOxi (ribose oxidation sequencing)-Seq[119,120]. The Nm-Seq approach allows for the detection of Nm at single-base resolution. This method is based on iterative oxidation-elimination-dephosphorylation (OED) reaction cycles. In each round, one 2´-OH nucleoside is removed from the 3´-end of the RNA. The Nm is resistant to the performed oxidation. Therefore, the chemical RNA degradation process is interrupted in the presence of Nm. The resulting RNA library is subsequently analyzed by NGS. This approach was applied to map Nm sites in human mRN[120].

In contrast to Nm-Seq, in RibOxi-Seq, the RNA is randomly digested with Benzonase, which results in RNA fragments with 2´,3´-OH ends[119]. Afterwards, the combination of oxidation and β-elimination is performed to expose Nm to the 3´-end of the RNA. Similar to Nm-Seq, 3’-Nm modified RNA fragments are protected from oxidation and can be analyzed by NGS. This method was applied to identify the Nm sites in mammalian rRNA[119]. However, this approach can not be used quantitatively to compare methylation intensity between different sites.

Furthermore, the presence of Nm in RNA affects the reverse transcriptase. Interestingly, under limited dNTPs concentrations, Nm modified residues trigger specific pausing of the reverse transcriptase[121]. In this approach, called 2OMe-Seq[121], RT reactions are performed in the presence of high and low dNTP concentrations using the same RNA as a template for the reverse transcriptase. The subsequential bioinformatical analysis of both sequencing outputs enables the identification of Nm positions. However, to date, Nm modifications in prokaryotes were observed only in rRNA and tRNA.

Writers, erasers and functions of 2´-O-methylation

Generally, the introduction of Nm modification into RNA is performed either by a stand-alone methyltransferase[122] or a snoRNA-guided 2´-O-methylase[123-125]. The latter is known to catalyze the introduction of Nm into eukaryotic and archaeal RNA, while in bacteria, the site-specific methyltransferase-mediated catalysis is common[126,127]. For instance, the Nm modifications of

tRNA at the positions 18, 32 and 34 are conserved in all domains of life. They are introduced in bacteria by site-specific Nm writers TrmH, TrmJ, and TrmL[127-129]. One of these modification sites in bacterial tRNA, particularly Gm18, was involved in suppressing the innate human immune activation[111].

In contrast to prokaryotes, the presence of Nm in eukaryotic mRNA was already proved, and its functional role partially explored. Nm containing mRNA possesses increased stability and inhibits ribosomal protein translation[47,113,130]. The position of Nm within eukaryotic mRNA affects the extent of translation inhibition[130]. It has been shown that the presence of Nm within the second codon in mRNA almost completely abolishes the translation.

Finally, to date, the presence of Nm in prokaryotic mRNA is not verified. Therefore, a functional role of this modification is not reported.

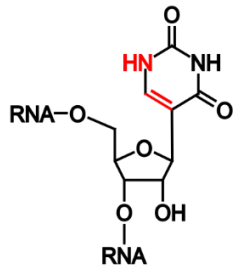
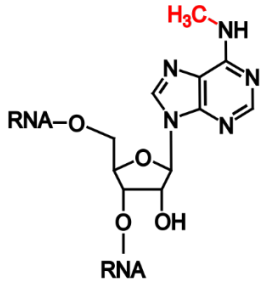
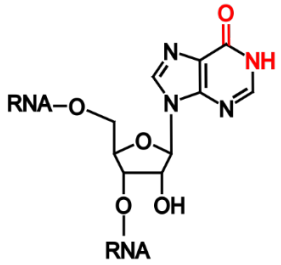
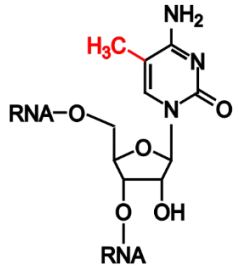
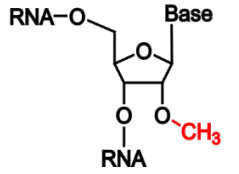
Modification	Structure	Detection	Writers, readers, erasers	Molecular roles
Pseudouridine (Ψ)		Pseudo-Seq Ψ–Seq Psi-Seq CeU-Seq RBS-Seq HydraPsiSeq	Writers: site-specific PUSs and snoRNA guided box H/ACA PNPs Eukaryotic reader: MetRS	Structure stabilization of tRNA and rRNA, translation modulation, alteration of codon translation, mRNA stability
N^6-adenosine (m^6A)		m^6A-Seq MeRIP-Seq m^6A-LAIC-Seq PA- m6A-Seq mi-CLIP m^6A-label-Seq DART-Seq NanoCompore	Writers: RlmF, RlmJ (prokaryotes), METTL3, METTL4, WTAP (eukaryotes) Eukaryotic erasers: FTO, ALKBH5 (eukaryotes) Eukaryotic readers - YTHDF1-3ly proteins	In eukaryotes involved in mRNA stability, RNA-protein interactions, miRNA processing, splicing, translation initiation and efficiency regulation.
Inosine (I)		ICE-Seq EndoVIPER-Seq	Writers: adenosine deaminases that act on rRNA (ADAR) or tRNA (ADAT) in eukaryotes Prokaryotic homologous – TadA, acts on tRNA and mRNA Reader: endonucleaseV	Codon alterations, alternative splicing, regulation of stress response, alternative splicing and 3´-UTR-variation in eukaryotes
5-methylcytidine (m^5C)		Bisulfite sequencing RIP-Seq Aza-IP NSUN2-RNA cross-linking based miCLIP TAWO-Seq	Writers: eukaryotic NSUN and DNMT2 methyltransferases Erasers - eukaryotic TET enzymes remove the modification by its further oxidation to hm^5C Eukaryotic readers: ALYREF, YBX1, YPS	tRNA structure stabilization, translational fidelity of ribosome
2´-O-methylation (Nm)		RT with limited dNTPs 2OMe-Seq RiboMeth-Seq Nm-Seq RibOxi-Seq	Writers - stand-alone methyltransferases or snoRNA-guided 2´-O-methylases	transcript stability, inhibition of ribosomal protein translation, translation efficiency

Figure 3: An overview about internal RNA modifications in eukaryotes and prokaryotes.

3.4. 5’-Terminal RNA modifications

Typically, transcription in prokaryotes and eukaryotes is initiated with a nucleoside triphosphate (NTP) creating a 5’-triphosphorylated RNA (5’-PPP-RNA). These primary transcripts are distinguished from their processed variants such as 5’-diphosphorylated (5’-PP-RNA), 5’-monophosphorylated (5’-P-RNA), or 5’-hydroxylated RNA (5’-OH-RNA) (Figure 4A). These different 5’-termini can determine their recognition by enzymes with RNA processing or binding activity such as RNase E. This endonuclease specifically recognizes a 5’-monophosphate in order to hydrolyze 5’-P-RNA[131,132].

Apart from these different triphosphate-derived 5’-ends of RNA, viral and eukaryotic mRNAs possess a 5’-cap[133,134]. This RNA cap comprises an N7-methylguanosine linked to the first nucleotide of the RNA via a 5’-to-5’ triphosphate (5’-m^7GpppN). The cap is deposited on the mRNA merely co-transcriptionally. After processing to 5’-PP-RNA by an RNA triphosphatase, the RNA guanylyltransferase covalently attaches guanosine monophosphate in 5’-to-5’-direction. This is subsequently methylated at the N7-position[135]. The functions of this eukaryotic cap are associated with pre-mRNA splicing, poly(A)-tailing, nuclear mRNA export, translation, and mRNA stability[135]. However, in bacteria, primary transcripts were believed to harbor phosphorylated 5’-ends only and are thereby distinguished from the majority of eukaryotic mRNAs. In addition to NTP-driven transcription initiation, primer-dependent transcription initiation can also create primary transcripts in bacteria[136,137]. These primary transcripts are characterized by a 5’-hydroxyl group (5’-OH-RNAs).

5’-PPP/PP/P/OH: Primary transcripts and their processed forms

As mentioned above, the 5’-phosphorylation state of an RNA is important to define its origin (primary or processed transcript). Primary transcripts are valuable RNAs as they indicate transcription start sites (TSSs) and 5’-UTRs and provide crucial information on the operon architecture in bacteria.

The detection of primary transcripts originating from the initiating NTP (5’-PPP-RNAs) is accomplished by differential RNA-seq (dRNA-seq) (Figure 4B). Here, total RNA from a bacterial culture is split into two fractions[138]: One sample is subjected to treatment with 5’-dependent terminator nuclease (TEX), which degrades processed transcripts and leaves primary transcripts intact (+TEX). The other sample is untreated and contains all types of 5’-phosphorylated RNAs (-TEX)[138]. Afterwards, tobacco acid pyrophosphatase (TAP) processes 5’-PPP-RNAs to 5’-P-RNAs that are subsequently ligated to a 5’-RNA linker and poly(A)-tailed prior to cDNA synthesis[138] (Figure 4B). After NGS (typically Illumina sequencing), reads in the +TEX sample typically show

enrichment at the 5’-end of genes compared to the -TEX sample (Figure 4B). Based on the enrichment patterns, TSSs can be computed bioinformatically[9]. This dRNA-seq approach has already been applied to various bacterial species such as *Staphylococcus aureus* and *Bacillus subtilis* [9] or *E. coli*[139]. Further, it has already been used to identify TSSs in host and phage genomes during infection[140,141].

Cappable-seq is another sequencing technique to detect TSSs[142]. Here, the vaccinia virus capping enzyme uses 3’-biotinylated GTP as a substrate to specifically cap 5’-PPP- and 5’-PP-RNA, whilst 5’-P-RNA remains unmodified. Capped and biotinylated RNA is subsequently enriched by streptavidin pull-down. Comparing enriched and non-enriched samples bioinformatically allows for the identification of primary transcripts and their TSSs. Thereby, 41 % of novel TSSs in *E. coli* could be annotated[142]. Also, Cappable-seq has proven suitable for meta-transcriptome analysis, as it enabled the mapping of TSSs in the mouse gut microbiome[142]. SMRT (Single Molecule, Real Time)-Cappable-seq is adjusted to sequence the intact, full-length primary transcripts[143]. The classical Cappable-seq relies on the fragmentation of cDNA before Illumina sequencing. To determine TSSs of entire primary transcripts, these are polyA-tailed during SMRT-Cappable-seq library preparation before cDNA synthesis. Then, unfragmented cDNA (> 1 kbp) is sequenced by PacBio sequencing[143]. This technology allows for the determination of full-length transcripts. SMRT-Cappable-seq revealed that 40 % of genes in *E. coli* are read-through at the transcription termination sites (TTSs)[143]. These findings suggest the existence of more operon variants than previously annotated, which may be determined by properties of the respective TTSs and growth conditions. The latter could resemble a gene regulatory mechanism that defines the number of cistrons per operon to fine-tune metabolic activities[143].

Interestingly, transcription cannot only be initiated with canonical NTPs but also with so-called nanoRNAs. These are short oligonucleotides (2 – 5 nts) that arise from RNA degradation[136,144]. They are used by eukaryotic, prokaryotic as well as archaeal RNA Polymerases (RNAPs) as substrates *in vitro*[145]. The first demonstration of nanoRNA-driven transcription initiation *in vivo* has been provided by studies of *Pseudonomas aeruginosa*. Here, nanoRNAs are frequently used for transcription initiation upon their stabilization[136]. Further studies have demonstrated their presence in *E. coli* and *Vibrio cholera*[137]. Transcripts derived from nanoRNA-mediated transcription initiation typically carry a 5’-OH-group[137]. nanoRNA-derived transcripts can be identified by the ectopic expression of an oligoRNase coupled with 5’-RNA-seq[146]. The expression of an oligoRNase reduces nanoRNA levels resulting in decreased levels of transcripts initiated by nanoRNAs. Comparing the reads obtained from ectopic expression of the oligoRNase to the wild-type condition indicates transcripts subjected to this non-canonical transcription initiation[146].

Another approach to identify transcripts that possess a 5’-OH involves *E. coli* RNA ligase RtcB. It ligates a 5’-desthiobiotinylated and 3’-phosphorylated oligonucleotide to 5’-OH-RNA in total RNA[147]. This allows for the capture of these transcripts via streptavidin that can be identified by NGS[147].

Notably, this primer-dependent transcription initiation occurs during stationary phase growth of *E. coli*[137] – even in the presence of an ectopically expressed oligoRNase that degrades nanoRNAs[146]. Here, nanoRNA-priming appears to mediate biofilm formation highlighting its biological significance[146].

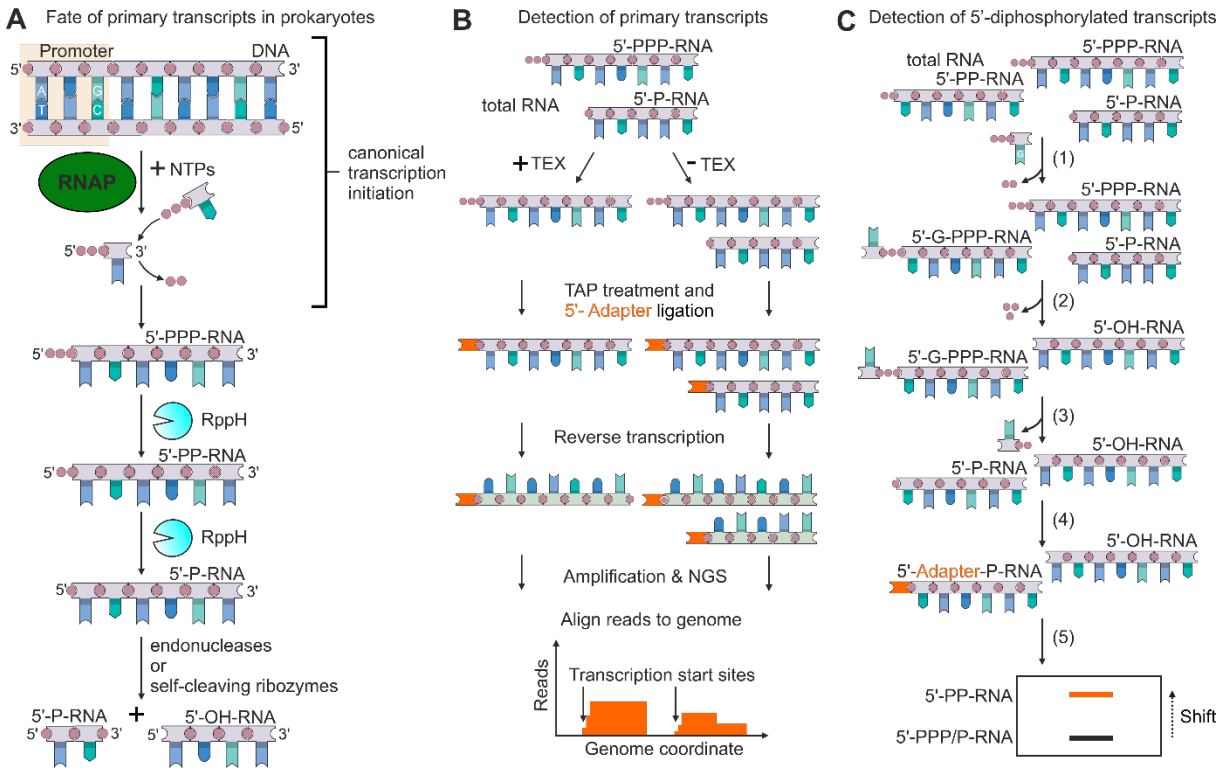


Figure 4: Synthesis, degradation and identification of primary transcripts and their derivatives. A) Canonical transcription is initiated by RNA Polymerase (RNAP) with NTPs resulting in 5’-PPP-RNA. It can be processed to 5’-PP-RNA and 5’-P-RNA by RppH. 5’-OH-RNA can be derived from internal RNA cleavage by endonucleases or self-cleaving ribozymes. B) Differential RNA-Seq for the enrichment and identification of primary transcripts. C) Detection of 5’-PP-RNAs with the PACO assay relies on the specific protection of 5’-PP-RNAs and the ligation of adapters. 5’-PP-RNAs are m^7G-capped by the Guanylyltransferase (1). Afterwards uncapped RNAs are dephosphorylated to 5’-OH-RNA by alkaline phosphatase (2). Pyrophosphatase converts capped RNAs into 5’-P-RNAs (3). 5’-P-RNA are ligated to an adapter (PABLO-assay) (4). Shifted bands in the Northern blot indicate 5’-PP-RNAs (5).

Once a particular phosphorylation state of an RNA's 5'-end has been discovered, one should set out to validate these findings. The phosphorylation assay by ligation of oligonucleotides (PABLO) is useful to assess the levels of a distinct 5'-monophosphorylated transcript relative to all other 5'-phosphorylation states[148] (Figure 4C). Here, the T4 DNA ligase covalently links the 3'-OH of a DNA oligo to the 5'-phosphate of an RNA. A splinted ligation ensures specific ligation of the DNA oligo to the 5'-end of the RNA. In this case, a "splint", a ssDNA complementary to the 5'-end of the RNA and the 3'-end of the DNA oligo, enables the specific ligation of the adapter[148]. Afterwards, the ligation product and the non-ligated product can be analyzed by Northern blotting to evaluate the levels of 5'-P-RNA.

The fraction of 5'-PP-RNA for a particular transcript can be quantified using the phosphorylation assay of capping outcome (PACO)[140]. In the first step, 5'-PP-RNA is capped by the eukaryotic guanylyltransferase whilst other RNA 5'-termini remain unmodified (Figure 4C). Afterwards, the RNA mixture is subjected to alkaline phosphatase treatment. Thereby, all RNA 5'-termini are converted into 5'-OH except for the newly capped 5'-PP-RNA. Subsequent pyrophosphatase treatment removes the cap from the former 5'-PP-RNA and releases 5'-P-RNA, which can be quantified via the PABLO assay[140]. In *E. coli*, the cellular 5'-PP-RNA levels are estimated around 50 % for distinct RNAs and can range up to 90 % in RppH defective cells[140]. Interestingly, 5'-PP-RNA arises from degradation of 5'-PPP-RNA to 5'-P-RNA by RppH and does not originate from the incorporation of NDP at the 5'-end of RNA during transcription initiation (Figure 4A)[140].

5'-cap-like structures on bacterial RNA

For decades, the absence of 5'-capped RNAs was regarded as a key feature of prokaryotic gene expression. About a decade ago, MS studies of bacterial total RNA by the Liu lab revealed that adenosine-containing enzymatic cofactors could decorate the 5'-end of prokaryotic RNAs. These cofactors include NAD or CoA [149,150]. Several studies have indicated that these cofactors are used as so-called non-canonical initiating nucleotides (NCINs) by the bacterial RNAP[151,152]. Therefore, the cofactors' adenosine moiety base pairs with the complementary strand at the +1 position if the transcript initiates with an adenosine[151]. Consequently, enzymatic cofactors can occupy the 5'-termini of various RNAs as cap-like structures. The following section summarizes the current knowledge of 5'-terminal RNA modifications in bacteria (Figure 5).

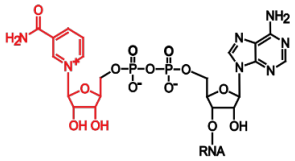
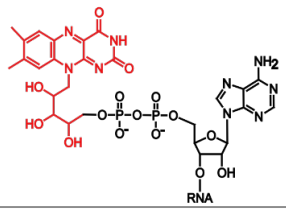
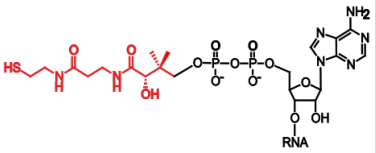
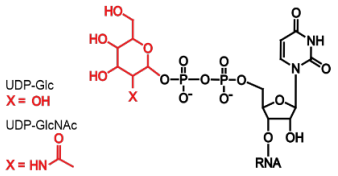
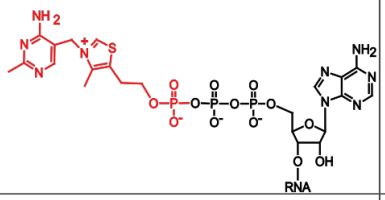
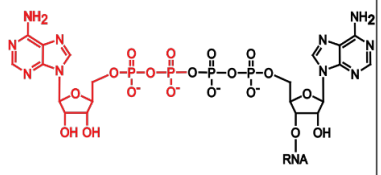
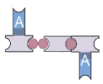
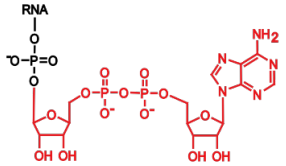
Modification	Structure	Detection	Writers, readers, erasers	Molecular roles
NAD 		NAD captureSeq NAD-tagSeq Mass spectrometry NAD-capQ	Incorporation by RNAP during transcription initiation Decapping by nudix hydrolases NudC in *E. coli* and RppH in *B. subtilis*; DeNADing via DXO/Rai1 family proteins in eukaryotes	Decoration of 5'-termini of sRNAs and mRNAs Stabilization against 5'-end dependent RNA processing in bacteria Association with ribosomes in *Arabidopsis*
FAD 		Mass spectrometry FAD-capQ	Incorporation by RNAP during transcription initiation in vitro DeFADing in eukaryotes by DXO/Rai1 family proteins	Decoration of 5'-termini of predominantly sRNAs; RNA sequences remain unidentified
CoA 		Mass spectrometry	Incorporation by RNAP during transcription initiation in vitro Decapping by nudix hydrolase NudC; DeCoAping via eukaryotic DXO/Rai1 family proteins	Decoration of 5'-termini of predominantly sRNAs; RNA sequences remain unidentified RNA protection against 5'-end dependent RNA processing and nudix hydrolase RppH
UDP-GlcNac/ UPD-Glc 		Mass spectrometry	Incorporation by RNAP during transcription initiation in vitro	RNA sequences remain unidentified
AThTP (Thiamine) 		None	Incorporation by T7 RNAP during transcription initiation in vitro	Existence in vivo and RNA sequences remain unidentified
Ap_4A (Dinucleoside tetraphosphates) 		Mass spectrometry Northern blotting	Incorporation by RNAP during transcription initiation; capping by tRNA synthetase in vitro Decapping by ApaH	About 15 transcripts analyzed by Northern blotting; the entirety of RNA sequences remains unidentified Protection against NudC
5'-phospho-ADP-ribose 		None	Deposition post-transcriptionally by Tpt1 from 5'-P-RNA and NAD	Existence in vivo and RNA sequences remain unidentified

Figure 5: On overview about 5’-terminal RNA modifications in prokaryotes.

Nicotinamide adenine dinucleotide

NAD (here, NAD refers to its oxidized form NAD^+) is an essential redox cofactor in all domains of life and one of the most widely studied NCINs (Figure 5). It has initially been proposed as an NCIN due to the ability of T7 RNAP to initiate transcription with this cofactor[153]. Six years later, mass spectrometric analyses of total RNA from *E. coli* and *Streptomyces venezuelae* digested to single nucleotide level have indicated NAD as a potential building block at the 5’-end of RNA as well[149]. The abundance of these 5’-NAD-capped-RNAs (NAD-RNAs) was estimated to 3,000 copies per cell.

Identification, quantification and validation of NAD-RNAs

Subsequent studies aimed to identify the transcripts in *E. coli* that are subjected to NAD-capping by the chemo-enzymatic capture and sequencing of NAD-RNAs[154]. This method, termed NAD captureSeq (Figure 6A), initially labels NAD-capped RNA with an alkyne moiety from pentyn-1-ol in an ADP-ribosyl cyclase catalyzed transglycosylation reaction. In a subsequent reaction step, a biotin-azide is covalently linked to the alkyne-modified RNA by copper-catalyzed azide–alkyne cycloaddition (CuAAC). Harnessing the streptavidin affinity to biotin, NAD-RNA can thereby be isolated from total RNA. Finally, adapters are ligated and cDNA is synthesized which is subjected to NGS (Illumina)[154,155]. In *E. coli*, mainly small regulatory RNAs (sRNAs) were identified to carry the redox cofactor NAD at their 5’-end as a cap-like structure[154].

Since the development of the NAD captureSeq technology, this method has enabled the discovery of NAD-RNA in other organisms from the prokaryotic world[156,157]. In *B. subtilis*, mostly mRNAs are NAD-capped. The overall abundance of these 5’-modified RNAs amounts to approximately 6.3 fmol/µg RNA, which corresponds to 220 copies per cell[156]. In *Staphylococcus aureus*, mRNAs and sRNAs are NAD-capped that account for 25 fmol/µg RNA, approximately 900 NAD-RNAs per cell[157].

It should be noted that, apart from prokaryotes, NAD-RNAs also exist in the eukaryotic kingdom of life. The application of the NAD captureSeq technology to various total RNA isolates revealed that NAD-capped RNAs are present in *S. cerevisiae*[158], *Arabidopsis thaliana*[159,160] and human embryonic kidney (HEK) cell lines[161].

Due to copper-induced RNA fragmentation, NAD-RNAs’ full-length sequence information is lost during NAD captureSeq library preparation. The recently developed SPAAC(strain-promoted azide-alkyne cycloaddition)-NAD-Seq provides a copper-free method to capture full-length NAD-RNAs[162]. Further, the combination of SPAAC-NAD-Seq with m^7G-RNA depletion tackles the issue of the NAD captureSeq reactivity towards these canonically capped eukaryotic mRNAs[162].

Recently, a modified version of NAD captureSeq to identify NAD-capped RNA has been developed (Figure 6B). NAD-tagSeq similarly relies on an initial transglycosylation reaction of NAD-RNA with pentyn-1-ol but uses a 3'-azide-RNA instead of a biotin-azide in the subsequent CuAAC reaction[160,163]. Former NAD-RNA is thereby 5'-labelled with a specific RNA-tag. This tag can be used for the specific enrichment of NAD-RNAs via complementary biotinylated oligonucleotides (Figure 6B). Then, full-length RNAs are sequenced using the Oxford Nanopore. Importantly, this approach allows for direct multiplexing of samples via the RNA-tags and the evaluation of NAD-RNA abundance in non-enriched samples[160,163].

In addition, NAD tagSeqII – a combination of NAD tagSeq and SPAAC-NAD-Seq – recently identified that NAD-capping in *E. coli* is dependent on its growth phases[164].

CapZyme-Seq can be used to assess the 5'-terminal modifications of transcripts. Therefore, RNA is processed by NudC, RppH or Rai1 and sequenced. These 5'-end processing enzymes either cleave within the pyrophosphate of the cofactor (catalyzed by NudC), remove the entire cofactor (catalyzed by Rai1) or trim 5'-PPP-ends (catalyzed by RppH). In detail, transcripts that are vulnerable to the respective Nudix hydrolase are converted to 5'-P-RNAs which are subsequently ligated to a 5'-adapter. This adapter allows for barcoding of samples and for the accurate determination of 5'-ends of the respective transcripts[165]. Thus, NCIN-RNAs (NAD-, CoA-, FAD and NADH-capped RNAs) can be detected and quantified. However, CapZyme-Seq is not entirely specific towards NAD-RNAs only, as other 5'-terminal RNA modifications (e.g CoA, FAD) can also be targeted by these 5'-RNA processing enzymes. Nevertheless, CapZyme-Seq allows for the detection of capped-RNA species, even if it cannot differentiate the types of NCIN-RNAs.

The quantification and estimation of NAD-capping levels or abundance of NAD-RNAs is mostly achieved by MS measurements of digested total RNA[149] (Figure 6C). Recently, the absolute quantification of NAD-RNA levels has been improved by a method called CapQuant. It detects NAD-caps but also other adenosine-containing cofactors[166]. This method relies on the digest of total RNA with nuclease P1, which hydrolyzes phosphodiester bonds. As pyrophosphate bonds are not affected by this treatment, caps such as NAD remain intact and are detected by high-resolution MS[149,166]. NAD-capQ is another recently developed method for the quantification of NAD-RNA among total RNA that does not rely on MS[167] (Figure 6E). Here, NAD, released from nuclease P1 digested total RNA, functions as a redox cofactor for an enzymatic cycling reaction. Its product subsequently reacts with NADH. The unique absorption of the resulting compound is harnessed for the quantification of NAD via a fluorometric assay (Figure 6E). Thereby, NAD-RNA concentrations of about 120 fmol/µg total *E. coli* RNA were determined, which amounts to approx. 4,000 molecules per cell[167].

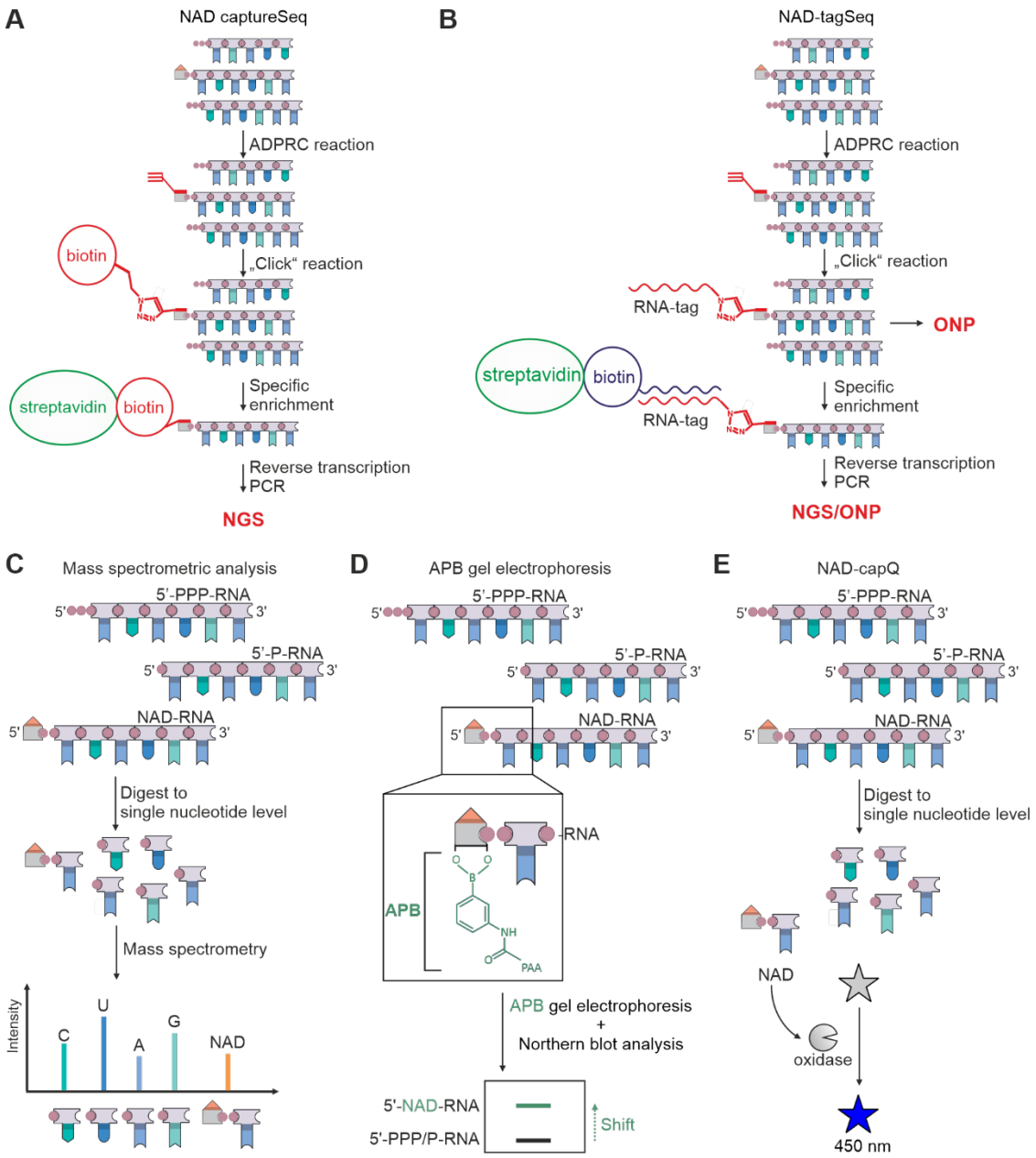


Figure 6: Identification, quantification and validation of NAD-RNAs. A) The NAD captureSeq workflow starts with the specific fusion of biotin to NAD-RNAs via click chemistry. The former NAD-RNAs are enriched and finally sequenced via Illumina (NGS). B) NAD-tagSeq modifies NAD-RNAs with a specific RNA-tag. This can serve for the enrichment of NAD-RNAs or the differentiation of NAD-RNAs from other RNA species by Oxford Nanopore sequencing (ONP). C) Total RNA is digested to single nucleotide level. NAD is detected via LC-MS/MS. D) The interaction of APB with the *cis*-diol of the NAD-cap results in a shift of NAD-RNA relative to 5’-PPP-RNA in an APB-PAGE gel. Thereby, NAD-capping levels of an RNA can be quantified. E) Quantification of NAD-RNAs in total RNA with NAD-capQ. Released NAD serves as a cofactor for an oxidase creating a colorimetric product with absorption at 450 nm.

Whilst these methods can only indicate the existence of NAD-RNA in total RNA, NGS-based approaches detect the actual NAD-capped transcripts. However, NAD-RNAs identified by sequencing techniques require validation which includes quantitative reverse transcription PCR (qRT-PCR) or aminophenyl boronic acid polyacrylamide gel electrophoresis (APB-PAGE) (Figure 6D). The latter harnesses the ability of boronic acid to interact with 1,2-*cis*-diols. These are exhibited by the ribose from NAD and the 3’-terminal nucleoside of the RNA. Consequently, NAD-RNA interacts more strongly with APB than 5’-PPP-RNA and is thereby retarded during APB-PAGE relative to 5’-PPP-RNA. After APB-PAGE, Northern Blotting with a probe specific towards the RNA of interest is applied to validate and quantify the extent of NAD capping for a particular transcript[168] (Figure 6D).

Such assays require a standard of NAD-RNA which is typically created by *in vitro* transcription with NAD and radiolabeled NTPs. The synthesis of NAD-RNA can also be achieved by a simple chemo-enzymatic approach. Here, 5’-P-RNA, initiating with adenosine, is coupled to nicotinamide mononucleotide (NMN) via a nucleophilic substitution reaction[169]. This reaction may be performed with any 5’-P-RNA yielding up to 50 % NAD-RNA and can overcome the restraint that the RNAP incorporates NAD at the +1 position with varying efficiency.

Writers, erasers and functions of NAD-RNAs

The writer of the NAD-cap as an epitranscriptomic modification is the RNAP. So far, three bacterial RNAPs have been demonstrated to accept NAD as an NCIN (Figure 7A)[151,156,157]. This fact is supported by structural studies of the *E. coli* RNAP, which binds NAD similar to ATP[151,152]. Thereby, NAD-capping by the RNAP occurs *ab initio* during transcription. As the adenosine moiety of NAD is directly involved in base pairing with the complementary strand at the +1 position, NAD-capping of transcripts by the RNAP requires transcripts that initiate with adenosine[151]. Consequently, ATP and NAD compete for the incorporation at the 5’-end of RNA. However, not all transcripts initiating with adenosine were actually found to be NAD-capped in *E. coli*[154]. The different extents of NAD-capping suggest a cellular mechanism by which transcription initiation efficiency at these genes with either NAD or ATP is regulated. Whilst NAD capping is not affected by sigma factors[156], preferentially A and T are most frequently located at the -1 position of respective NAD-capped transcripts in *B. subtilis* and *S. aureus* [156,157]. It has been shown that the -1 position influences the incorporation efficiency of NAD by the respective RNAP[156,157]. This is supported by structural studies of the RNAP during transcription initiation[151].

In the eukaryotic world, the RNA polymerase II and the mitochondrial RNAP (mtRNAP) have been described to mediate the NAD-capping of transcripts in a similar manner[170]. This also implicates

eukaryotic RNAPs as writers of NAD-RNAs. In light of the endosymbiotic theory, especially the mtRNAP – a relative of the T7 RNAP[171] – suggests a well-conserved feature of the prokaryotic RNAP to accept NAD and also other NCINs. Furthermore, it explains how mitochondrial RNAs could be NAD-capped[161].

Next to the discovery of the writers of NAD-capped RNAs, researchers were engaged to study the biological function of NAD-capped RNAs. Compared to 5’-P-RNA, NAD-RNA is stabilized against 5’-end-dependent nucleases such as RNase E in *E. coli* or RNase J1 in *B. subtilis*[154,156]. The *E. coli* Nudix hydrolase NudC functions as a decapping enzyme of NAD-RNA (Figure 7B). NudC catalyzes the pyrophosphate hydrolysis in the NAD cap which releases NMN and a 5’-P-RNA that initiates with an adenosine[172]. This RNA-cleavage product may be degraded more easily by nucleases such as RNase E[154,172]. Homologs of this decapping enzyme exist in yeast (NPY1)[173] but remain unidentified in other bacteria such as *B. subtilis* or *S. aureus*. Instead, *B. subtilis* RppH catalyzes the decapping of NAD-RNA analogous to NudC. The depletion of these decapping enzymes resulted in an increased stability of the NAD-RNAs *in vivo*[154,156]. Interestingly, eukaryotic decapping enzymes of NAD-RNA exhibit different processing activity. Enzymes of the DXO/Rai1 family release intact NAD and 5’-P-RNA which is subsequently degraded due to their exonuclease activity (Figure 7C)[161,174]. This form of NAD-RNA processing is referred to as deNADing. Further, the ADP-ribosyl cyclase CD38 can convert NAD-RNA in ADP-ribose-capped-RNA by liberating nicotinamide *in vitro*[175].

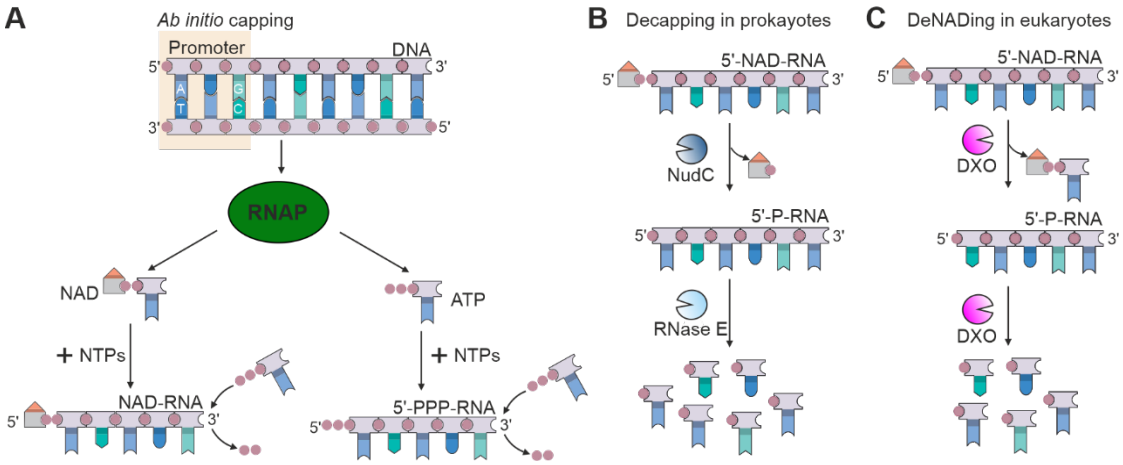


Figure 7: Writers and erasers of NAD-RNAs. A) Synthesis of NAD-RNAs *ab initio* during transcription initiation. The RNA polymerase uses NAD instead of ATP in order to start transcription of +1A transcripts. B) NAD-RNA is decapped by NudC generating NMN and a 5’-P-RNA which can subsequently be degraded by 5’-dependent nucleases such as RNase E. C) In eukaryotes, enzymes of the DXO/Rai1 family perform

deNADing of NAD-RNA liberating intact NAD. The residual 5’-P-RNA is degraded by the exonuclease activity of the enzyme.

The decapping activity of enzymes on NAD-RNA is usually assessed using radiolabeled NAD-RNA *in vitro*. A recent study used a NAD-RNA, furanylated at the adenosine, as a valuable substrate to screen for novel decapping enzymes[175]. Upon releasing NMN or nicotinamide by a decapping enzyme, the fluorescence of the furanylated adenosine is no longer quenched. Thereby, decapping activity can be monitored by simple fluorometric measurements over time. This assay allows for the characterization of known decapping enzymes of NAD-RNAs and the identification of new erasers[175].

Apart from the protective effect of the NAD-cap, the functions of NAD-RNAs remain unknown. The identity of NAD-RNAs in different bacterial species varies. Whereas mainly sRNAs are NAD-capped in *E. coli*, mRNAs are NAD-capped in *B. subtilis* and *S. aureus*[154,156,157]. In the latter bacterium, *S. aureus*, RNAIII is NAD-capped, and increasing NAD-capping levels coincide with reduced production of toxins encoded by this RNA[157]. The mechanistic basis of this observation remains elusive.

To this day, the functional consequences of NAD-capping in eukaryotes are not well understood. In *A. thaliana*, NAD-RNAs were found to be associated with translating ribosomes. Thus, it was suggested that NAD caps might be involved in the process of translation[159]. Nevertheless, NAD-RNA functions beyond stabilization remain to be elucidated in prokaryotes[176].

Flavin adenine dinucleotide

Flavin adenine dinucleotide (FAD) is an important enzymatic redox cofactor of multiple essential metabolic processes. FAD is composed of adenosine diphosphate and flavin mononucleotide (FMN), which contains an isoalloxazine ring that mediates the redox reaction. Whilst FAD refers to its fully oxidized state, $FADH_2$ corresponds to the fully reduced form of FAD after the acceptance of two electrons and two protons.

FAD has first been reported as a potential non-canonical cap of RNA in 2003 (Figure 5). Huang *et al.* showed that the T7 RNAP accepts FAD as an NCIN[153]. Analogous to NAD-capping, the adenosine moiety of FAD could base pair with the nucleobase at +1 position of the complementary strand. Thereby, the RNAP can FAD-cap RNA *ab initio*. Moreover, *E. coli* RNAP accepts FAD as an NCIN *in vitro* as well[152]. Consequently, FAD-capped RNA (FAD-RNA) may also exist in *E. coli*[177].

The existence of FAD-RNA in *E. coli* total RNA has already been verified by CapQuant[166]. Thereby, the cellular FAD-RNA concentration was quantified to 0.17 fmol/µg *E. coli* RNA – more than 10-fold lower than the amount of NAD-RNA[166]. Another method capable of quantifying the FAD-cap levels in total RNA is FAD-capQ which has only been applied to eukaryotic RNA so far[174]. This technique quantifies FAD-RNA in total RNA using *Schizosaccharomyces pombe* Rai1 (SpRai1) which removes intact FAD from FAD-RNA *in vitro* (Figure 8A)[174]. Thereby, FAD is released from RNA and serves as a substrate for an oxidase reaction of a commercially available fluorometric assay[174]. In HEK 293T cells, FAD-capQ identified 0.33 fmol/µg short RNA (< 200 nt). In the absence of DXO, the mammalian decapping enzyme of FAD-RNA, FAD-RNA levels were doubled[174].

These methods are only capable of assessing the overall FAD-capping levels in total RNA. However, no method is available that captures FAD-capped transcripts. If FAD-RNAs are present *in vivo*, APB-PAGE can be applied analogously to NAD-RNA[168]. This electrophoretic method retards *in vitro* transcribed FAD-RNA relative to PPP-RNA[168]. Interestingly, NAD-RNA is more efficiently retarded than FAD-RNA during APB-PAGE. The ribitol group in FAD allows for rotation of hydroxyl groups, resulting in a decreased retardation of FAD-RNA[168]. Northern Blotting of the APB-gel with a probe specific towards the RNA of interest allows for the validation and quantification of the FAD-RNA.

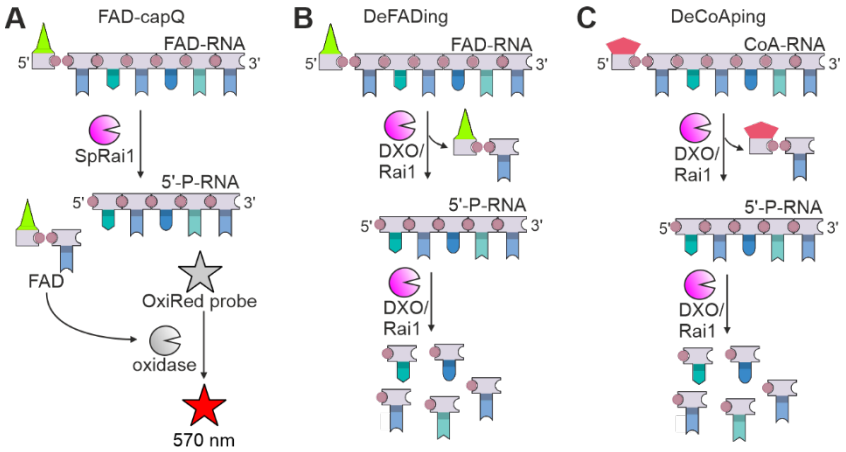


Figure 8: Erasers of FAD- and CoA-RNAs in eukaryotes and FAD-RNA quantification assay. A) Workflow of FAD-capQ which applies SpRai1 on total RNA which thereby removes FAD from FAD-RNAs. FAD then serves as a cofactor for an oxidase which converts an OxiRed probe in a product with absorption at 570 nm. B) DeFADing by DXO/Rai1 family enzymes in eukaryotes releases intact FAD from FAD-RNA. The resulting 5’-P-RNA is subsequently degraded. C) DeCoAping by DXO/Rai1 family enzymes in

eukaryotes remove intact 3’-dephospho-CoA from CoA-RNA. Then, exonucleolytic activity leads to 5’-P-RNA decay.

Erasers of FAD-RNA in bacteria such as the Nudix hydrolase NudC that decaps NAD-RNA in *E. coli* are not reported to this day. The only known decapping enzymes of FAD-RNA exist in eukaryotes. Mouse DXO (mDXO) and *Kluyveromyces lactis* Dxo1 (KlDxo1) are capable of deFADing FAD-capped RNA *in vitro*. DeFADing releases intact FAD and 5’-P-RNA. This cleavage product can subsequently be degraded by the decapping enzymes (Figure 8B) that make use of their intrinsic 5’-3’-exonuclease activity[174]. SpRai1 does not harbor exonuclease activity and only removes FAD, which maintains 5’-P-RNA integrity[174].

Coenyzme A

Coenzyme A (CoA) is composed of 3’-phosphorylated adenosine diphosphate and pantetheine. The latter contains a terminal thiol group that activates building blocks for fatty acids via thioester formation. Thereby, CoA plays a central role as an enzymatic cofactor for fatty acid metabolism.
The existence of CoA-capped RNA (CoA-RNA) was initially suggested by *in vitro* transcription experiments with T7 RNAP using 3’-dephospho-CoA as substrate (Figure 5). Instead of ATP, T7 RNAP incorporates 3’-dephospho-CoA at the 5’-end of primary transcripts[153]. Importantly, only 3’-dephospho-CoA serves as an NCIN. In contrast, CoA which originates from phosphorylation of 3’-dephospho-CoA by 3’-dephospho-CoA kinase cannot be used for *ab initio* RNA capping. The 3’ phosphate on the adenosine moiety of CoA prevents transcriptional elongation[153].
Proof for the existence of CoA-RNA *in vivo* has been provided by LC-MS measurements using digested total RNA[150]. Thereby, 3’-dephospho-CoA and its thioester derivates (e.g. acetyl-thioester) were identified to decorate the 5’-end of bacterial RNA in *S. venezuelae* (13 fmol/µg total RNA) and *E. coli* (8 fmol/µg total RNA)[150]. Using size fractionation of total RNA, the approximate maximum size of CoA-RNAs was estimated to 200 nucleotides per RNA[150]. However, due to the lack of a specific capturing technique for CoA-RNA, the exact transcripts carrying a CoA-cap are unidentified to this day.

Writers, erasers and functions of CoA-RNAs

Apart from T7 RNAP initiating transcription with 3’-dephospho-CoA *in vitro*[153], *E. coli* RNAP can also start transcription of adenosine initiating transcripts with 3’-dephospho-CoA *in vitro*[151]. In the eukaryotic world, the mitochondrial RNAP accepts 3’-dephospho-CoA as substrate *in vitro* as well[170].

Decapping of CoA-RNA is exhibited by several enzymes from different kingdoms of life. Bird and colleagues showed decapping of CoA-RNA by the Nudix hydrolase NudC *in vitro*[151]. Upon decapping, the pyrophosphate moiety is cleaved, releasing pantetheine phosphate and a 5'-P-RNA initiating with adenosine. Thereby, NudC is the only bacterial enzyme with reported decapping activity on CoA-RNA. Interestingly, the Nudix hydrolase RppH is not able to hydrolyze the CoA-cap[151]). The same is observed for the Nudix hydrolase NudL from *E. coli* which has recently been reported to hydrolyze CoA whilst being inactive on CoA-RNA[178].

In addition, three eukaryotic erasers of the CoA-cap have recently been reported. mDXO and SpRai1 are both capable of deCoAping CoA-RNA *in vitro* (Figure 8C)[174]. DeCoAping refers to the 5'-terminal processing of CoA-RNA that releases 3'-dephospho-CoA and 5'-P-RNA (Figure 8C)[174]. KlDxo1 combines deCoAping and exonuclease activity which leads to immediate RNA degradation after decapping of CoA-RNA *in vitro*[174]. Considering that 13 Nudix hydrolases – the minority of which properly characterized – do exist in *E. coli*, it remains open whether bacterial Nudix hydrolases could also exhibit deCoAping activity.

Similar to FAD-RNAs, the sequences and the functions of CoA-RNAs remain entirely unidentified to this day.

UDP-GlcNAc and UDP-Glc

The precursors of cell wall building blocks UDP-glucose (UDP-Glc) and UDP-GlcNAc contain a uridine and can thereby serve as NCINs similar to FAD or NAD (Figure 5). Here, transcripts initiating with uridine can be subjected to UDP-GlcNAc- or UDP-Glc-capping. These cell wall precursors can be incorporated by *E. coli* RNAP at the 5'-end of RNA instead of UTP[152]. Thus, UDP-GlcNAc-/UDP-Glc-RNAs are capped co-transcriptionally – analogous to RNAs capped with adenosine-derived cofactors[152]. The concentrations of UDP-Glc (2.5 mM) and especially UDP-GlcNAc (9.2 mM) amount to the same order of magnitude as concentrations of UTP (8.3 mM) [179]. Thus, capping of RNA in *E. coli* with these cell wall precursors is likely to occur. It can be speculated that the capping levels can be linked to the cellular cell wall synthesis. Depending on the cellular UDP-Glc and UDP-GlcNAc concentrations, capping with these NCINs could be affected as well[180]. The analysis of RNA isolated from *E. coli* DH5α by CapQuant revealed the existence of UDP-GlcNAc- and UDP-Glc-RNAs. The concentration of UDP-GlcNAc-RNA (2.5 fmol/µg RNA) exceeds the concentration of UDP-Glc-RNA (0.22 fmol/µg RNA) by 10-fold. Thus, UDP-GlcNAc-RNAs and NAD-RNAs (2.2 fmol/µg RNA) are present in similar concentrations[166]. Interestingly, these non-canonical caps also exist in the eukaryotic world (humans, mice and yeast)[166].

More complex 5’-UDP-derived caps, such as UDP-GlcNAc-pentapeptide, do not serve as NCINs for *E. coli* RNAP *in vitro*[152]. *In vivo*, this could prevent RNA from sequestration to the cell membrane/ cell wall and retain these more expensive building blocks for their actual purpose of cell wall synthesis[152]. Apart from these speculations, the sequences and functions of UDP-GlcNAc- and UDP-Glc-RNAs remain elusive.

Thiamine

Vitamin B1 – known as thiamine – is present in all domains of life and can be found in different phosphorylation states: thiamine monophosphate (ThMP), thiamine diphosphate (ThDP) and thiamine triphosphate (ThTP)[181]. ThDP – also known as thiamine pyrophosphate- is required for cellular aldehyde group transfer and the most abundant form of thiamine in various species including bacteria[181]. Interestingly, ThTP synthesis is elevated under certain stress conditions such as hypoxia in *E. coli*[181]. Further, thiamine exists in conjugation with adenine – adenosine thiamine triphosphate (AThTP) and adenosine thiamine diphosphate (AThDP) – in *E. coli* and eukaryotes[182]. Interestingly, upon carbon starvation of *E. coli*, AThTP can accumulate (79 – 120 pmol/mg protein)[183].

As described above, bacterial and T7 RNAP can use adenosine-containing cofactors as NCINs for the initiation of transcription instead of ATP (Figure 5). Thereby, these NCINs compete for their incorporation into RNA with the canonical nucleotide ATP. A recent study has shown that AThDP and AThTP – both containing an adenosine moiety – can similarly be incorporated into RNA by T7 RNAP *in vitro*[184]. Thus, it is speculated that 5’-thiamine-capped RNA (thiamine-RNA) could be produced by RNAPs from the prokaryotic world and probably also by the related mtRNAP *in vitro*. However, its existence and functions in bacteria remain elusive. In order to examine the presence and identity of thiamine-RNA in bacteria, capturing approaches similar to NAD captureSeq could be suitable. The thiazole ring of thiamine supplies a hub for the chemical capture of thiamine-RNA, as its thiazole ring opens upon nucleophilic attack of hydroxide under alkaline conditions[184]. By using a nucleophilic linker molecule, thiamine-RNA can be modified with an alkyne moiety which can subsequently be linked to a biotin-azide conjugate. Such a biotin-thiamine conjugate may be used for the specific enrichment of thiamine-RNA from total RNA in the future[184]. The functionality of these potential thiamine-RNAs is as obscure as the sequence of the transcripts subjected to capping.

Np_nN (Dinucleoside polyphosphates)

Dinucleoside polyphosphates (Np_nNs) are unusual nucleotides existing in all domains of life. They are composed of two nucleosides whose 5’-ends are linked to each other by a polyphosphate group of three to six phosphates[185]. These molecules are known to act as second messengers – frequently referred to as alarmones – whose cellular concentrations increase upon different stress stimuli such as cadmium that induces disulfide stress[186]. However, their functions in bacteria remained unknown for more than 50 years.

Recently, these alarmones were identified to serve as a bacterial RNA cap (Figure 5). Here, Np_nNs could similarly act as NCINs by complementary base pairing of one nucleobase with the +1 position of the antisense strand during transcription initiation. Thereby, an Np_n-cap is created whose chemical nature is determined by the length of the polyphosphate (n) and the identity of the nucleoside (N). Various adenosine-derived dinucleoside tetraphosphates – Np_4As (Ap_4A, Cp_4A, Gp_4A, Up_4A) – were identified as non-canonical 5’-caps of transcripts in *E. coli*[187]. Exemplary studies on the yeiP RNA, which encodes the translational elongation factor EF-Tu, showed that this RNA is heavily Np_4A-capped upon cadmium-induced stress in *E. coli*[187]. Here, boronate gel electrophoresis which specifically retards Np_n-capped RNA due to an additional *cis*-diol followed by Northern blotting was used to detect Np_n-capped transcripts and other 5’-modifications[187]. Moreover, calculations based on 14 randomly investigated transcripts with median Np_n-capping levels over 40 % suggest towards a higher abundance of Np_n-capped RNA than NAD-RN[187]. Interestingly, disulfide stress induced by cadmium or diamide dramatically increases the Np_n-capping levels [187]. It has to be noted that boronate gel electrophoresis is directed towards the identification of individual transcripts and its capping levels. However, this technique cannot reveal the type of nucleoside and the number of phosphates that constitute the Np_n-cap (e.g. Cp_n or Ap_n). Hence, an LC-MS/MS approach – similar to the one used to prove the existence of NAD- and CoA-RNA – was developed for the identification of Np_n-capped RNA. Briefly, the small RNA fraction (< 200 nucleotides) is purified from *E. coli* and subjected to nuclease P1[188]. Consequently, RNA is degraded to single nucleotide level, whereas Np_n-caps remain intact and can be detected by LC-MS/MS analysis[188]. Thereby, Ap_3A, Ap_5A and Ap_3G as well as methylated forms of Np_nNs were detected as 5’-caps of RNA in *E. coli*[188]. Still, the scientific community is missing a specific capturing approach for Np_n-capped RNA in order to determine the exact transcripts. Also, the promoter sequences that might influence Np_n-capping on a transcriptome-wide scale in various organisms are unidentified.

Writers, erasers and functions of Np_n-capped RNA

In order to synthesize Np_n-capped transcripts, the RNAP must accept these molecules as NCINs and initiate transcription with a respective Np_nN. *In vitro*, both T7 and *E. coli* RNAP have been demonstrated as writers of several Np_n-caps in the presence of various Np_nNs (Figure 9A) (e.g. $Ap_{3-6}A$, $Ap_{4-5}G$, Gp_4G)[187-190]. Independent studies have shown that these alarmones are more efficiently used as NCINs by these RNAPs *in vitro* than NAD and CoA [188,189]. Importantly, the length of the polyphosphate connecting the two nucleosides seems to affect the efficiency of transcription initiation. Tetraphosphates appear to be optimal linkers as demonstrated by molecular dynamics simulation and *in vitro* transcription experiments with T7 RNAP[189,190]. Increasing linker length coincides with lower incorporation efficiency[189]. Moreover, the promoter influences the efficiency of transcription initiation with Np_nNs. A purine at -1 position increases transcription initiation with Np_4A *in vivo.* The nucleobase of the Np_n-cap also appears to be involved in base pairing with the base at the -1 position at the antisense strand[189,190]. *In vivo*, dinucleoside polyphosphates such as Ap_4A are synthesized by tRNA synthetases such as *E. coli* lysyl-tRNA synthetase LysU[191]. Interestingly, this enzyme also generates Ap_4-capped RNA *in vitro* using ATP and 5’-P-RNA as substrates (Figure 9A)[187]. Whether this tRNA synthetase or only the RNAP performs Np_n-capping in *E. coli* remains elusive.

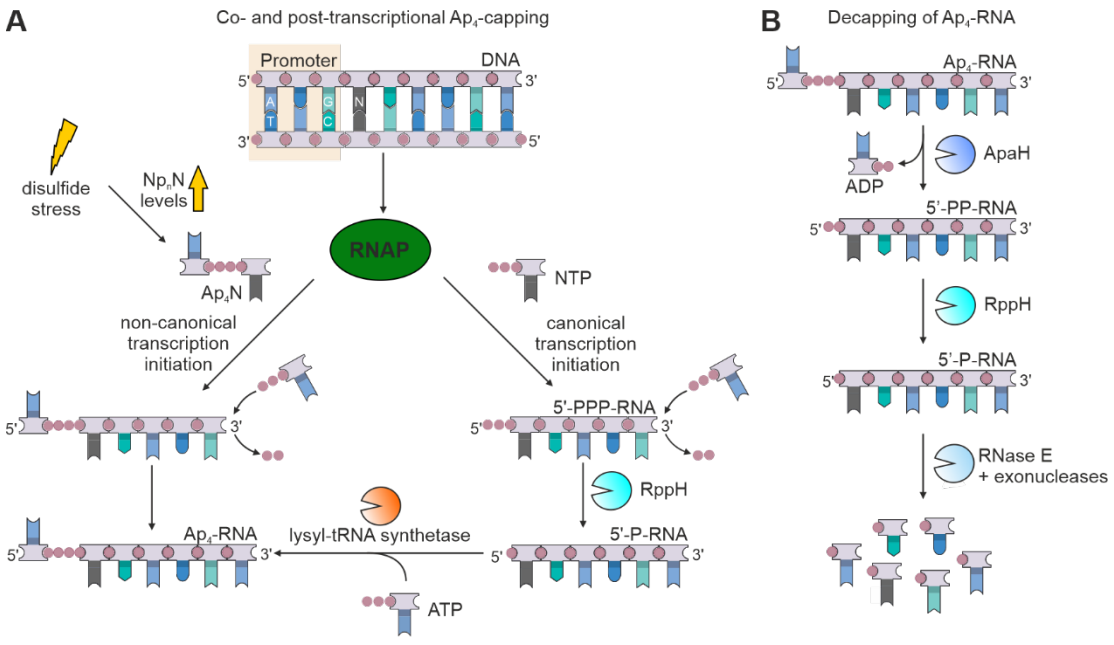


Figure 9: Dinucleoside tetraphosphates in prokaryotes. Writers, erasers and biological context of Np_4-caps. A) Disulfide stress increases cellular Np_nN concentrations in *E. coli*. The RNA Polymerase either incorporates Ap_4N *ab initio* during transcription or the lysyl-tRNA synthetase post-transcriptionally creates this cap from ATP and 5’-P-RNA. B) Ap_4-RNA is decapped by ApaH which is followed by 5’-PP-RNA conversion to 5’-P-RNA. This RNA is degradable by 5’-dependent endonucleases, e.g. RNase E.

As elaborated above, 5’-cofactor capped RNAs such as NAD-RNA or CoA-RNA can be processed by the Nudix hydrolase NudC. Whilst Np_n-capped RNA is protected against NudC, RppH and bis(5-nucleosyl)-tetraphosphatase (ApaH) both remove the Np_n-cap[187,188]. In *E. coli*, ApaH is the major regulator of the stability of Np_4-capped RNA[187]. In its absence, Np_4-capped RNAs are stabilized, and their levels are elevated. Further, ApaH appears to be inactivated upon disulfide stress resulting in an increase of Np_n-capped transcript levels. Thereby, ApaH governs the lifetime of Np_4-capped RNA in *E. coli*. Upon decapping, 5’-PP-RNA is released, which could subsequently be processed to 5’-P-RNA by RppH (Figure 9B). These cleavage steps could finally induce RNase E-mediated RNA degradation of Np_n-capped RNA[187].

Phospho-ADP-ribose

The NAD-dependent enzyme Tpt1 is known to remove an internal 2’-phosphate from RNA. This creates a 2’-hydroxyl group in the RNA and simultaneously releases nicotinamide from NAD. Thereby, the cofactor is irreversibly converted to ADP-ribose cyclic phosphate[192]. Interestingly, *Aeropyrum pernix* Tpt1 (ApeTpt1) can ADP-ribosylate 5’-P-RNA using NAD as a substrate, which results in 5’-phospho-ADP-ribosylated RNA (Figure 5). As typical for ADP-ribosylation, nicotinamide serves as a leaving group. Thereby, a covalent bond between oxygen of the phosphate group and the C1 atom of ADP-ribose is formed. To this day, the occurrence of this phenomenon is restricted to *in vitro* experiments. Consequently, its occurrence *in vivo* as well as its readers, erasers and ultimately functions remain elusive.

3.5. Conclusion and Outlook

The last few years have largely increased our understanding of internal and 5’-terminal RNA modifications in prokaryotes and eukaryotes. This process was primarily driven by the development of various novel technologies (e.g. RNA-seq based methods) with increased specificities and sensitivities with particular regard to internal RNA modifications. These methods allow for the identification and validation of 5’-terminal and internal RNA modifications. They comprise the epitranscriptome which likely provides additional regulatory layers of gene expression and potentially harbours various still unidentified functions. However, the characterization of the bacterial epitranscriptome is still lacking behind. Many technologies have been developed and applied to analyze RNA modifications in eukaryotes.

Despite the large set of methods to study internal RNA modifications, these techniques have only rarely been applied to prokaryotes. The knowledge about internal RNA modifications in bacteria is consequently limited. Whilst the RNA modifications m^5C, m^6A, I and Ψ are merely studied in

rRNA and tRNA, internal modifications of bacterial mRNAs are poorly understood regarding their existence, location and functions as well as their writers, readers and erasers. Furthermore, for the majority of the known RNA modifications, no NGS approaches for efficient mapping are available to date. For this reason, their presence in mRNA is tedious to prove.

Till today, more than 160 different RNA modifications are described[2]. The majority of RNA modifications was identified in tRNA but was not yet observed in mRNAs. For instance, the modification of guanine to queuosine was ubiquitously observed in prokaryotic and eukaryotic tRNA and is suggested to be involved in many cellular processes, including translation, aerobic/anaerobic metabolism control, and bacterial virulence[193]. Even though the biosynthesis pathways of queuosine modified RNA in bacteria are described, the existence of this modification in mRNA remains unexplored[194].

In contrast to the barely studied bacterial internal RNA modifications, which can be investigated by a variety of methods, the prokaryotic kingdom of life provides pioneering organisms for the study of 5’-terminal RNA modifications. Here, various non-canonical initiating nucleotides, including enzymatic cofactors such as NAD, FAD or CoA decorate the 5’-end of sRNAs and mRNAs. 5’-terminal modifications in bacteria are created via non-canonical transcription initiation by the RNA Polymerase and can be erased by enzymes of the Nudix hydrolase family. However, the exact transcripts that carry such 5’-caps were only identified in a small set of bacteria covering only a few modifications. Here, mainly NAD-RNAs are properly characterized in terms of RNA sequences. However, the identity of transcripts decorated by other 5’-terminal modifications such as FAD, CoA, AThTP, Np_nN, UDP-GlcNAc remains elusive.

In general, the detection of bacterial and also eukaryotic RNA modifications faces major challenges. These include the probably low abundance of some modifications and the lack of sensitive methods for their detection in reasonable amounts of biological material. This requires the improvement of existing and the development of novel specific capturing or enrichment strategies for RNA modifications other than an NAD-cap or 5’-phosphates or internal modifications. It has to be noted that here presented enrichment approaches – either chemical (e.g. NAD captureSeq) or biological (e.g. antibody-mediated enrichment) – can result in biased hits after sequencing. Consequently, RNA modifications detected on the basis of such sequencing data require validation via different approaches as presented before (MS, gel retardation or 5’-phosphorylation assays). In addition, these tools can provide quantitative measures for the respective RNA modifications.

Further, methods that can detect several RNA modifications in a parallel manner might be of incredible value. Single molecule sequencing technologies with sensitivities towards non-

canonical nucleobases could allow for the simultaneous detection of several RNA modifications in a single transcript. The Oxford Nanopore and SMRT sequencing using the PacBio sequencing technology have both been demonstrated as sequencing tools for the detection of m^6A in total RNA isolates[10,195]. However, their applicability to parallelly monitor multiple RNA modifications on the same transcript has not been shown to this day.

Apart from major challenges regarding the detection of RNA modifications, their biological functions in bacteria remain to be elucidated. In the light of studies analyzing the meta-transcriptome – transcriptomes of microbiomes- the potential biological role of RNA modifications becomes even more fascinating[196,197]. Which roles might RNA modifications play in these contexts? How are their abundances affected under physiological conditions such as in the microbiome? Exemplarily, dual RNA-Seq can define the transcriptome of both a pathogen and its host[198]. So far, it has not been combined with studies of the epitranscriptome, which may fulfil functions in the context of pathogen-host interactions. Further, it appears that 5’-terminal RNA modifications could correlate to the cellular concentrations of the respective cofactors or nucleoside analogues. These can be affected by growth conditions and growth stages of bacterial cell populations. Consequently, studies analyzing the epitranscriptome under different growth conditions and growth stages may help to detect RNA modifications. Here, one should consider that the expression of the writers and erasers and the presence of the respective RNA modifications may correlate to the distinct physiological conditions. Also, these RNA modifications could be linked to various metabolic processes and stress conditions in bacteria[180,190].

Moreover, the current picture of bacterial epitranscriptomics reflects an entire cell population. Single-cell transcriptomics may be valuable to investigate how RNA modifications are correlated to different bacterial phenotypes within such populations[141].

In summary, the discrepancy between the available detection methods and the knowledge about internal RNA modifications in bacteria is striking. The modifications as well as their writers, readers and erasers are well described in eukaryotes but lack proper characterization in prokaryotes. In contrast to that, 5’-terminal RNA modifications are predominantly understood in bacteria but sufficient techniques for their identification are rare.

Common for both internal and 5’-terminal RNA modifications, the readers and functions of these diverse modifications are still obscure. The ultimate goal is to identify the exact functions of these various modifications in bacteria (and eukaryotes) and to determine their biological significance. Provided the rapid advances in the field of (epi-)transcriptomics, the elucidation of the relevance of these RNA modifications in bacteria seems promising.

Acknowledgements

K.H. acknowledges funding from the Eliteprogramm of the Baden-Württemberg Stiftung and the Max Planck Society.

3.6. References

1 Cohn, W. E. & Volkin, E. Nucleoside-5'-Phosphates from Ribonucleic Acid. *Nature* **167**, 483-484 (1951). https://doi.org:DOI 10.1038/167483a0

2 Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**, D303-D307 (2018). https://doi.org:10.1093/nar/gkx1030

3 McCown, P. J. *et al.* Naturally occurring modified ribonucleosides. *Wiley Interdiscip Rev RNA* **11**, e1595 (2020). https://doi.org:10.1002/wrna.1595

4 Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635-1646 (2012). https://doi.org:10.1016/j.cell.2012.05.003

5 Saletore, Y. *et al.* The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol* **13**, 175 (2012). https://doi.org:10.1186/gb-2012-13-10-175

6 Lewis, C. J., Pan, T. & Kalsotra, A. RNA modifications and structures cooperate to guide RNA-protein interactions. *Nat Rev Mol Cell Biol* **18**, 202-210 (2017). https://doi.org:10.1038/nrm.2016.163

7 Lane, B. G. in *Modification Editing of RNA* 1-20 (1998).

8 Wang, X. *et al.* -methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388-1399 (2015). https://doi.org:10.1016/j.cell.2015.05.014

9 Sharma, C. M. & Vogel, J. Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol* **19**, 97-105 (2014). https://doi.org:10.1016/j.mib.2014.06.010

10 Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201-206 (2018). https://doi.org:10.1038/nmeth.4577

11 Lorenz, D. A., Sathe, S., Einstein, J. M. & Yeo, G. W. Direct RNA sequencing enables m(6)A detection in endogenous transcript isoforms at base-specific resolution. *RNA* **26**, 19-28 (2020). https://doi.org:10.1261/rna.072785.119

12 Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *BioRxiv*, 843136 (2019).

13 Deng, X. *et al.* Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **43**, 6557-6567 (2015). https://doi.org:10.1093/nar/gkv596

14 Bar-Yaacov, D. *et al.* RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res* **27**, 1696-1703 (2017). https://doi.org:10.1101/gr.222760.117

15 Nie, W. *et al.* A-to-I RNA editing in bacteria increases pathogenicity and tolerance to oxidative stress. *PLoS Pathog* **16**, e1008740 (2020). https://doi.org:10.1371/journal.ppat.1008740

16 McCown, P. J. *et al.* Naturally occurring modified ribonucleosides. *Wires Rna* **11** (2020). https://doi.org:ARTN e1595 10.1002/wrna.1595

17 Davis, F. F. & Allen, F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *Journal of Biological Chemistry* **227**, 907-915 (1957).

18 Sun, W. J. *et al.* RMBase: a resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res* **44**, D259-265 (2016). https://doi.org:10.1093/nar/gkv1036

19 Spenkuch, F., Motorin, Y. & Helm, M. Pseudouridine: still mysterious, but never a fake (uridine)! *RNA Biol* **11**, 1540-1554 (2014). https://doi.org:10.4161/15476286.2014.992278

20 Penzo, M., Guerrieri, A. N., Zacchini, F., Trere, D. & Montanaro, L. RNA Pseudouridylation in Physiology and Medicine: For Better and for Worse. *Genes (Basel)* **8**, 301 (2017). https://doi.org:10.3390/genes8110301

21 Ofengand, J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514**, 17-25 (2002). https://doi.org:10.1016/s0014-5793(02)02305-0

22 Roovers, M. *et al.* Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Res* **34**, 4293-4301 (2006). https://doi.org:10.1093/nar/gkl530

23 Cohn, W. E. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *J Biol Chem* **235**, 1488-1498 (1960).

24 Choi, Y. C. & Busch, H. Modified nucleotides in T1 RNase oligonucleotides of 18S ribosomal RNA of the Novikoff hepatoma. *Biochemistry* **17**, 2551-2560 (1978). https://doi.org:10.1021/bi00606a015

25 Holley, R. W. *et al.* Structure of a ribonucleic acid. *Science*, 1462-1465 (1965).

26 Carlile, T. M., Rojas-Duran, M. F. & Gilbert, W. V. Pseudo-Seq: Genome-Wide Detection of Pseudouridine Modifications in RNA. *Methods Enzymol* **560**, 219-245 (2015). https://doi.org:10.1016/bs.mie.2015.03.011

27 Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014). https://doi.org:10.1016/j.cell.2014.08.028

28 Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS One* **9**, e110799 (2014). https://doi.org:10.1371/journal.pone.0110799

29 Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**, 592-597 (2015). https://doi.org:10.1038/nchembio.1836

30 Addepalli, B. & Limbach, P. A. Pseudouridine in the anticodon of Escherichia coli tRNATyr (QΨA) is catalyzed by the dual specificity enzyme RluF. *J Biol Chem* **291**, 22327-22337 (2016).

31 Khoddami, V. *et al.* Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* **116**, 6784-6789 (2019).

32 Marchand, V. *et al.* HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res* **48**, e110 (2020). https://doi.org:10.1093/nar/gkaa769

33 Li, Y. H., Zhang, G. & Cui, Q. PPUS: a web server to predict PUS-specific pseudouridine sites. *Bioinformatics* **31**, 3362-3364 (2015). https://doi.org:10.1093/bioinformatics/btv366

34 Chen, W., Tang, H., Ye, J., Lin, H. & Chou, K. C. iRNA-PseU: Identifying RNA pseudouridine sites. *Mol Ther Nucleic Acids* **5**, e332 (2016). https://doi.org:10.1038/mtna.2016.37

35 He, J. *et al.* PseUI: Pseudouridine sites identification based on RNA sequence information. *BMC Bioinformatics* **19**, 306 (2018). https://doi.org:10.1186/s12859-018-2321-0

36 Lv, Z., Zhang, J., Ding, H. & Zou, Q. RF-PseU: A Random Forest Predictor for RNA Pseudouridine Sites. *Front Bioeng Biotechnol* **8**, 134 (2020). https://doi.org:10.3389/fbioe.2020.00134

37 Song, B. *et al.* PIANO: A Web Server for Pseudouridine-Site (Psi) Identification and Functional Annotation. *Front Genet* **11**, 88 (2020). https://doi.org:10.3389/fgene.2020.00088

38 Johnson, L. & Söll, D. In vitro biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from Escherichia coli. *Proc Natl Acad Sci U S A* **67**, 943-950 (1970).

39 Wrzesinski, J., Nurse, K., Bakin, A., Lane, B. G. & Ofengand, J. A Dual-Specificity Pseudouridine Synthase - an Escherichia-Coli Synthase Purified and Cloned on the Basis of Its Specificity for Psi-746 in 23s Rna Is Also Specific for Psi-32 in Trna(Phe). *Rna* **1**, 437-448 (1995).

40 Hamma, T. & Ferre-D'Amare, A. R. Pseudouridine synthases. *Chem Biol* **13**, 1125-1135 (2006). https://doi.org:10.1016/j.chembiol.2006.09.009

41 Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**, 5020-5026 (1995). https://doi.org:10.1093/nar/23.24.5020

42 Kierzek, E. *et al.* The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* **42**, 3492-3501 (2014). https://doi.org:10.1093/nar/gkt1330

43 Desaulniers, J. P. *et al.* Pseudouridines in rRNA helix 69 play a role in loop stacking interactions. *Org Biomol Chem* **6**, 3892-3895 (2008). https://doi.org:10.1039/b812731j

44 Liang, X. H., Liu, Q. & Fournier, M. J. rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol Cell* **28**, 965-977 (2007). https://doi.org:10.1016/j.molcel.2007.10.012

45 King, T. H., Liu, B., McCully, R. R. & Fournier, M. J. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-435 (2003). https://doi.org:10.1016/s1097-2765(03)00040-6

46 Eyler, D. E. *et al.* Pseudouridinylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* **116**, 23068-23074 (2019). https://doi.org:10.1073/pnas.1821754116

47 Hoernes, T. P. *et al.* Eukaryotic Translation Elongation is Modulated by Single Natural Nucleotide Derivatives in the Coding Sequences of mRNAs. *Genes (Basel)* **10**, 84 (2019). https://doi.org:10.3390/genes10020084

48 Levi, O. & Arava, Y. S. Pseudouridine-mediated translation control of mRNA by methionine aminoacyl tRNA synthetase. *Nucleic Acids Res* **49**, 432-443 (2021). https://doi.org:10.1093/nar/gkaa1178

49 Beemon, K. & Keith, J. Localization of N6-methyladenosine in the Rous sarcoma virus genome. *J Mol Biol* **113**, 165-179 (1977). https://doi.org:10.1016/0022-2836(77)90047-x

50 Dimock, K. & Stoltzfus, C. M. Sequence specificity of internal methylation in B77 avian sarcoma virus RNA subunits. *Biochemistry* **16**, 471-478 (1977). https://doi.org:10.1021/bi00622a021

51 Ma, H. *et al.* N 6-Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nat Chem Biol* **15**, 88-94 (2019).

52 Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206 (2012). https://doi.org:10.1038/nature11112

53 Meng, J. *et al.* A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. *Methods* **69**, 274-281 (2014). https://doi.org:10.1016/j.ymeth.2014.06.008

54 Molinie, B. *et al.* m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. *Nat Methods* **13**, 692-698 (2016). https://doi.org:10.1038/nmeth.3898

55 Chen, K. *et al.* High-Resolution-Methyladenosine (m6A) Map Using Photo-Crosslinking-Assisted mA Sequencing. *Angewandte Chemie-International Edition* **54**, 1587-1590 (2015). https://doi.org:10.1002/anie.201410647

56 Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767-772 (2015). https://doi.org:10.1038/nmeth.3453

57 Shu, X. *et al.* A metabolic labeling method detects m(6)A transcriptome-wide at single base resolution. *Nat Chem Biol* **16**, 887-895 (2020). https://doi.org:10.1038/s41589-020-0526-9

58 Wang, Y., Xiao, Y., Dong, S., Yu, Q. & Jia, G. Antibody-free enzyme-assisted chemical approach for detection of N(6)-methyladenosine. *Nat Chem Biol* **16**, 896-903 (2020). https://doi.org:10.1038/s41589-020-0525-x

59 Garcia-Campos, M. A. *et al.* Deciphering the "mA Code" via Antibody-Independent Quantitative Profiling. *Cell* **178**, 731-+ (2019). https://doi.org:10.1016/j.cell.2019.06.013

60 Zhang, Z. *et al.* Single-base mapping of m(6)A by an antibody-independent method. *Sci Adv* **5**, eaax0250 (2019). https://doi.org:10.1126/sciadv.aax0250

61 Imanishi, M., Tsuji, S., Suda, A. & Futaki, S. Detection of N(6)-methyladenosine based on the methyl-sensitivity of MazF RNA endonuclease. *Chem Commun (Camb)* **53**, 12930-12933 (2017). https://doi.org:10.1039/c7cc07699a

62 Conesa, A. *et al.* Erratum to: A survey of best practices for RNA-seq data analysis. *Genome Biol* **17**, 181 (2016). https://doi.org:10.1186/s13059-016-1047-4

63 Schwartz, S. *et al.* High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409-1421 (2013). https://doi.org:10.1016/j.cell.2013.10.047

64 Meyer, K. D. *et al.* 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999-1010 (2015). https://doi.org:10.1016/j.cell.2015.10.012

65 Roignant, J. Y. & Soller, M. m(6)A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. *Trends Genet* **33**, 380-390 (2017). https://doi.org:10.1016/j.tig.2017.04.003

66 Liu, J. *et al.* A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat Chem Biol* **10**, 93-95 (2014). https://doi.org:10.1038/nchembio.1432

67 Li, R., Yang, Y. G., Gao, Y., Wang, Z. Q. & Tong, W. M. A distinct response to endogenous DNA damage in the development of Nbs1-deficient cortical neurons. *Cell Res* **22**, 859-872 (2012). https://doi.org:10.1038/cr.2012.3

68 Jia, G. *et al.* N 6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology* **7**, 885-887 (2011).

69 Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29 (2013). https://doi.org:10.1016/j.molcel.2012.10.015

70 Wang, X. & He, C. Reading RNA methylation codes through methyl-specific binding proteins. *RNA Biol* **11**, 669-672 (2014). https://doi.org:10.4161/rna.28829

71 Edens, B. M. *et al.* FMRP Modulates Neural Differentiation through m6A-Dependent mRNA Nuclear Export. *Cell Reports* **28**, 845-+ (2019). https://doi.org:10.1016/j.celrep.2019.06.072

72 Meyer, K. D. DART-seq: an antibody-free method for global m(6)A detection. *Nat Methods* **16**, 1275-1280 (2019). https://doi.org:10.1038/s41592-019-0570-0

73 Warren, L., Flaks, J. G. & Buchanan, J. M. Biosynthesis of the purines. XX. Integration of enzymatic transformylation reactions. *J Biol Chem* **229**, 627-640 (1957).

74 Krokan, H. E. & Bjoras, M. Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583 (2013). https://doi.org:10.1101/cshperspect.a012583

75 Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F. & Ochoa, S. Synthetic polynucleotides and the amino acid code. V. *Proc Natl Acad Sci U S A* **48**, 613-616 (1962). https://doi.org:10.1073/pnas.48.4.613

76 Marceca, G. P. *et al.* Detecting and Characterizing A-To-I microRNA Editing in Cancer. *Cancers (Basel)* **13**, 1699 (2021). https://doi.org:10.3390/cancers13071699

77 Sakurai, M., Yano, T., Kawabata, H., Ueda, H. & Suzuki, T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat Chem Biol* **6**, 733-740 (2010). https://doi.org:10.1038/nchembio.434

78 Knutson, S. D., Ayele, T. M. & Heemstra, J. M. Chemical Labeling and Affinity Capture of Inosine-Containing RNAs Using Acrylamidofluorescein. *Bioconjugate Chem* **29**, 2899-2903 (2018). https://doi.org:10.1021/acs.bioconjchem.8b00541

79 Knutson, S. D. & Heemstra, J. M. EndoVIPER-seq for Improved Detection of A-to-I Editing Sites in Cellular RNA. *Curr Protoc Chem Biol* **12**, e82 (2020). https://doi.org:10.1002/cpch.82

80 Yao, M., Hatahet, Z., Melamede, R. J. & Kow, Y. W. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli. *J Biol Chem* **269**, 16260-16268 (1994).

81 Alseth, I., Dalhus, B. & Bjoras, M. Inosine in DNA and RNA. *Curr Opin Genet Dev* **26**, 116-123 (2014). https://doi.org:10.1016/j.gde.2014.07.008

82 Gerber, A., Grosjean, H., Melcher, T. & Keller, W. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *Embo J* **17**, 4780-4789 (1998). https://doi.org:10.1093/emboj/17.16.4780

83 Zinshteyn, B. & Nishikura, K. Adenosine-to-inosine RNA editing. *Wiley Interdiscip Rev Syst Biol Med* **1**, 202-209 (2009). https://doi.org:10.1002/wsbm.10

84 Keller, W., Wolf, J. & Gerber, A. Editing of messenger RNA precursors and of tRNAs by adenosine to inosine conversion. *FEBS Lett* **452**, 71-76 (1999). https://doi.org:10.1016/s0014-5793(99)00590-6

85 Lehmann, K. A. & Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875-12884 (2000). https://doi.org:10.1021/bi001383g

86 Wolf, J., Gerber, A. P. & Keller, W. tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *Embo J* **21**, 3841-3851 (2002). https://doi.org:10.1093/emboj/cdf362

87 Rafels-Ybern, A. *et al.* The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes. *Mol Biol Evol* **36**, 650-662 (2019). https://doi.org:10.1093/molbev/msy245

88 Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* **175**, 315-332 (1948).

89 Wang, S. & Kool, E. T. Origins of the large differences in stability of DNA and RNA helices: C-5 methyl and 2'-hydroxyl effects. *Biochemistry* **34**, 4125-4132 (1995). https://doi.org:10.1021/bi00012a031

90 Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339-346 (2017). https://doi.org:10.1038/nature21351

91 Dubin, D. T. & Taylor, R. H. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res* **2**, 1653-1668 (1975). https://doi.org:10.1093/nar/2.10.1653

92 Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *Plos Genet* **9**, e1003602 (2013). https://doi.org:10.1371/journal.pgen.1003602

93 Amort, T. *et al.* Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol* **18**, 1 (2017). https://doi.org:10.1186/s13059-016-1139-1

94 Cui, X. *et al.* 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol Plant* **10**, 1387-1399 (2017). https://doi.org:10.1016/j.molp.2017.09.013

95 David, R. *et al.* Transcriptome-Wide Mapping of RNA 5-Methylcytosine in Arabidopsis mRNAs and Noncoding RNAs. *Plant Cell* **29**, 445-460 (2017). https://doi.org:10.1105/tpc.16.00751

96 Yang, X. *et al.* 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res* **27**, 606-625 (2017). https://doi.org:10.1038/cr.2017.55

97 Xing, J. *et al.* NSun2 Promotes Cell Growth via Elevating Cyclin-Dependent Kinase 1 Translation. *Mol Cell Biol* **35**, 4043-4052 (2015). https://doi.org:10.1128/MCB.00742-15

98 Agris, P. F., Sierzputowska-Gracz, H. & Smith, C. Transfer RNA contains sites of localized positive charge: carbon NMR studies of [13C]methyl-enriched Escherichia coli and yeast tRNAPhe. *Biochemistry* **25**, 5126-5131 (1986). https://doi.org:10.1021/bi00366a022

99 Susan, J. C., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic acids research* **22**, 2990-2997 (1994).

100 Schaefer, M., Pollex, T., Hanna, K. & Lyko, F. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res* **37**, e12 (2009). https://doi.org:10.1093/nar/gkn954

101 Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol* **31**, 458-464 (2013). https://doi.org:10.1038/nbt.2566

102 Hussain, S., Aleksic, J., Blanco, S., Dietmann, S. & Frye, M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol* **14**, 215 (2013). https://doi.org:10.1186/gb4143

103 Motorin, Y., Lyko, F. & Helm, M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. *Nucleic Acids Res* **38**, 1415-1430 (2010). https://doi.org:10.1093/nar/gkp1117

104 Flatau, E., Gonzales, F. A., Michalowsky, L. A. & Jones, P. A. DNA methylation in 5-aza-2'-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells. *Mol Cell Biol* **4**, 2098-2102 (1984). https://doi.org:10.1128/mcb.4.10.2098-2102.1984

105 Jüttermann, R., Li, E. & Jaenisch, R. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc Natl Acad Sci U S A* **91**, 11797-11801 (1994).

106 Reid, R., Greene, P. J. & Santi, D. V. Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Res* **27**, 3138-3145 (1999). https://doi.org:10.1093/nar/27.15.3138

107 Bujnicki, J. M., Feder, M., Ayres, C. L. & Redman, K. L. Sequence-structure-function studies of tRNA:m5C methyltransferase Trm4p and its relationship to DNA:m5C and RNA:m5U methyltransferases. *Nucleic Acids Res* **32**, 2453-2463 (2004). https://doi.org:10.1093/nar/gkh564

108 Fu, L. *et al.* Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *J Am Chem Soc* **136**, 11582-11585 (2014). https://doi.org:10.1021/ja505305z

109 Zou, F. *et al.* Drosophila YBX1 homolog YPS promotes ovarian germ line stem cell development by preferentially recognizing 5-methylcytosine RNAs. *Proc Natl Acad Sci U S A* **117**, 3603-3609 (2020). https://doi.org:10.1073/pnas.1910862117

110 Gehrig, S. *et al.* Identification of modifications in microbial, native tRNA that suppress immunostimulatory activity. *J Exp Med* **209**, 225-233 (2012).

111 Rimbach, K., Kaiser, S., Helm, M., Dalpke, A. H. & Eigenbrod, T. 2'-O-Methylation within Bacterial RNA Acts as Suppressor of TLR7/TLR8 Activation in Human Innate Immune Cells. *J Innate Immun* **7**, 482-493 (2015). https://doi.org:10.1159/000375460

112 Hoernes, T. P. *et al.* Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Research* **44**, 852-862 (2016). https://doi.org:10.1093/nar/gkv1182

113 Choi, J. *et al.* 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat Struct Mol Biol* **25**, 208-216 (2018). https://doi.org:10.1038/s41594-018-0030-z

114 Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187-1200 (2017). https://doi.org:10.1016/j.cell.2017.05.045

115 Dennis, P. P., Tripp, V., Lui, L., Lowe, T. & Randau, L. C/D box sRNA-guided 2'-O-methylation patterns of archaeal rRNA molecules. *BMC Genomics* **16**, 632 (2015). https://doi.org:10.1186/s12864-015-1839-z

116 Krogh, N., Birkedal, U. & Nielsen, H. in *RNA Methylation* 189-209 (Springer, 2017).

117 Birkedal, U. *et al.* Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew Chem Int Ed Engl* **54**, 451-455 (2015). https://doi.org:10.1002/anie.201408362

118 Galvanin, A. *et al.* Bacterial tRNA 2'-O-methylation is dynamically regulated under stress conditions and modulates innate immune response. *Nucleic Acids Res* **48**, 12833-12844 (2020). https://doi.org:10.1093/nar/gkaa1123

119 Zhu, Y., Pirnie, S. P. & Carmichael, G. G. High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA* **23**, 1303-1314 (2017). https://doi.org:10.1261/rna.061549.117

120 Dai, Q. *et al.* Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods* **14**, 695-698 (2017). https://doi.org:10.1038/nmeth.4294

121 Incarnato, D. *et al.* High-throughput single-base resolution mapping of RNA 2´-O-methylated residues. *Nucleic Acids Res* **45**, 1433-1441 (2017). https://doi.org:10.1093/nar/gkw810

122 Somme, J. *et al.* Characterization of two homologous 2'-O-methyltransferases showing different specificities for their tRNA substrates. *RNA* **20**, 1257-1271 (2014). https://doi.org:10.1261/rna.044503.114

123 Cavaille, J., Nicoloso, M. & Bachellerie, J. P. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* **383**, 732-735 (1996). https://doi.org:10.1038/383732a0

124 Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M. & Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077-1088 (1996). https://doi.org:10.1016/s0092-8674(00)81308-2

125 Shubina, M. Y., Musinova, Y. R. & Sheval, E. V. Nucleolar Methyltransferase Fibrillarin: Evolution of Structure and Functions. *Biochemistry (Mosc)* **81**, 941-950 (2016). https://doi.org:10.1134/S0006297916090030

126 Bachellerie, J. P., Cavaille, J. & Huttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775-790 (2002). https://doi.org:10.1016/s0300-9084(02)01402-5

127 Persson, B. C., Jager, G. & Gustafsson, C. The spoU gene of Escherichia coli, the fourth gene of the spoT operon, is essential for tRNA (Gm18) 2'-O-methyltransferase activity. *Nucleic Acids Res* **25**, 4093-4097 (1997). https://doi.org:10.1093/nar/25.20.4093

128 Hori, H. *et al.* Identification and characterization of tRNA (Gm18) methyltransferase from Thermus thermophilus HB8: domain structure and conserved amino acid sequence motifs. *Genes Cells* **7**, 259-272 (2002). https://doi.org:10.1046/j.1365-2443.2002.00520.x

129 Benítez-Páez, A., Villarroya, M., Douthwaite, S., Gabaldón, T. & Armengod, M.-E. YibK is the 2′-O-methyltransferase TrmL that modifies the wobble nucleotide in Escherichia coli tRNALeu isoacceptors. *RNA* **16**, 2131-2143 (2010).

130 Hoernes, T. P. *et al.* Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res* **44**, 852-862 (2016). https://doi.org:10.1093/nar/gkv1182

131 Bandyra, K. J. & Luisi, B. F. RNase E and the High-Fidelity Orchestration of RNA Metabolism. *Microbiol Spectr* **6** (2018). https://doi.org:10.1128/microbiolspec.RWR-0008-2017

132 Mackie, G. A. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**, 720-723 (1998). https://doi.org:10.1038/27246

133 Furuichi, Y., Muthukrishnan, S. & Shatkin, A. J. 5'-Terminal m-7G(5')ppp(5')G-m-p in vivo: identification in reovirus genome RNA. *Proc Natl Acad Sci U S A* **72**, 742-745 (1975). https://doi.org:10.1073/pnas.72.2.742

134 Wei, C. M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell* **4**, 379-386 (1975). https://doi.org:10.1016/0092-8674(75)90158-0

135 Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA* **2**, 277-298 (2011). https://doi.org:10.1002/wrna.52

136 Goldman, S. R. *et al.* NanoRNAs prime transcription initiation in vivo. *Mol Cell* **42**, 817-825 (2011). https://doi.org:10.1016/j.molcel.2011.06.005

137 Vvedenskaya, I. O. *et al.* Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. *Genes Dev* **26**, 1498-1507 (2012). https://doi.org:10.1101/gad.192732.112

138 Borries, A., Vogel, J. & Sharma, C. M. in *Tag-Based Next Generation Sequencing* 109-121 (2011).

139 Thomason, M. K. *et al.* Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J Bacteriol* **197**, 18-28 (2015). https://doi.org:10.1128/JB.02096-14

140 Luciano, D. J., Vasilyev, N., Richards, J., Serganov, A. & Belasco, J. G. A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in Escherichia coli. *Mol Cell* **67**, 44-54 e46 (2017). https://doi.org:10.1016/j.molcel.2017.05.035

141 Wicke, L. *et al.* Introducing differential RNA-seq mapping to track the early infection phase for Pseudomonas phage ϕKZ. *RNA Biol* **18**, 1099-1110 (2021). https://doi.org:10.1080/15476286.2020.1827785

142 Ettwiller, L., Buswell, J., Yigit, E. & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**, 199 (2016). https://doi.org:10.1186/s12864-016-2539-z

143 Yan, B., Boitano, M., Clark, T. A. & Ettwiller, L. SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nat Commun* **9**, 3676 (2018). https://doi.org:10.1038/s41467-018-05997-6

144 Liao, H., Liu, M. & Guo, X. The special existences: nanoRNA and nanoRNase. *Microbiol Res* **207**, 134-139 (2018). https://doi.org:10.1016/j.micres.2017.11.014

145 Nickels, B. E. & Dove, S. L. NanoRNAs: a class of small RNAs that can prime transcription initiation in bacteria. *J Mol Biol* **412**, 772-781 (2011). https://doi.org:10.1016/j.jmb.2011.06.015

146 Druzhinin, S. Y. *et al.* A Conserved Pattern of Primer-Dependent Transcription Initiation in Escherichia coli and Vibrio cholerae Revealed by 5' RNA-seq. *Plos Genet* **11**, e1005348 (2015). https://doi.org:10.1371/journal.pgen.1005348

147 Peach, S. E., York, K. & Hesselberth, J. R. Global analysis of RNA cleavage by 5'-hydroxyl RNA sequencing. *Nucleic Acids Res* **43**, e108 (2015). https://doi.org:10.1093/nar/gkv536

148 Celesnik, H., Deana, A. & Belasco, J. G. in *RNA Turnover in Bacteria, Archaea and Organelles* Vol. 447 *Methods in Enzymology* 83-98 (Academic Press, 2008).

149 Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y. & Liu, D. R. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol* **5**, 879-881 (2009). https://doi.org:10.1038/nchembio.235

150 Kowtoniuk, W. E., Shen, Y., Heemstra, J. M., Agarwal, I. & Liu, D. R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A* **106**, 7768-7773 (2009). https://doi.org:10.1073/pnas.0900528106

151 Bird, J. G. *et al.* The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. *Nature* **535**, 444-447 (2016). https://doi.org:10.1038/nature18622

152 Julius, C. & Yuzenkova, Y. Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic Acids Res* **45**, 8282-8290 (2017). https://doi.org:10.1093/nar/gkx452

153 Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res* **31**, e8 (2003). https://doi.org:10.1093/nar/gng008

154 Cahova, H., Winz, M. L., Höfer, K., Nubel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

155 Winz, M. L. *et al.* Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq. *Nature protocols* **12**, 122-149 (2017). https://doi.org:10.1038/nprot.2016.163

156 Frindert, J. *et al.* Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in B. subtilis. *Cell Rep* **24**, 1890-1901 e1898 (2018). https://doi.org:10.1016/j.celrep.2018.07.047

157 Morales-Filloy, H. G. *et al.* The 5' NAD Cap of RNAIII Modulates Toxin Production in Staphylococcus aureus Isolates. *J Bacteriol* **202** (2020). https://doi.org:10.1128/JB.00591-19

158 Walters, R. W. *et al.* Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114**, 480-485 (2017). https://doi.org:10.1073/pnas.1619369114

159 Wang, Y. *et al.* NAD(+)-capped RNAs are widespread in the Arabidopsis transcriptome and can probably be translated. *Proc Natl Acad Sci U S A* **116**, 12094-12102 (2019). https://doi.org:10.1073/pnas.1903682116

160 Zhang, H. *et al.* NAD tagSeq reveals that NAD(+)-capped RNAs are mostly produced from a large number of protein-coding genes in Arabidopsis. *Proc Natl Acad Sci U S A* **116**, 12072-12077 (2019). https://doi.org:10.1073/pnas.1903683116

161 Jiao, X. *et al.* 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. *Cell* **168**, 1015-1027 e1010 (2017). https://doi.org:10.1016/j.cell.2017.02.019

162 Hu, H. *et al.* SPAAC-NAD-seq, a sensitive and accurate method to profile NAD(+)-capped transcripts. *Proc Natl Acad Sci U S A* **118**, e2025595118 (2021). https://doi.org:10.1073/pnas.2025595118

163 Shao, X. *et al.* NAD tagSeq for transcriptome-wide identification and characterization of NAD(+)-capped RNAs. *Nature protocols* **15**, 2813-2836 (2020). https://doi.org:10.1038/s41596-020-0363-z

164 Zhang, H. *et al.* Use of NAD tagSeq II to identify growth phase-dependent alterations in E. coli RNA NAD(+) capping. *Proc Natl Acad Sci U S A* **118**, e2026183118 (2021). https://doi.org:10.1073/pnas.2026183118

165 Vvedenskaya, I. O. *et al.* CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD<sup/>. *Mol Cell* **70**, 553-564 e559 (2018). https://doi.org:10.1016/j.molcel.2018.03.014

166 Wang, J. *et al.* Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res* **47**, e130 (2019). https://doi.org:10.1093/nar/gkz751

167 Grudzien-Nogalska, E., Bird, J. G., Nickels, B. E. & Kiledjian, M. "NAD-capQ" detection and quantitation of NAD caps. *RNA* **24**, 1418-1425 (2018). https://doi.org:10.1261/rna.067686.118

168 Nubel, G., Sorgenfrei, F. A. & Jaschke, A. Boronate affinity electrophoresis for the purification and analysis of cofactor-modified RNAs. *Methods* **117**, 14-20 (2017). https://doi.org:10.1016/j.ymeth.2016.09.008

169 Höfer, K., Abele, F., Schlotthauer, J. & Jäschke, A. Synthesis of 5'-NAD-Capped RNA. *Bioconjugate Chem* **27**, 874-877 (2016). https://doi.org:10.1021/acs.bioconjchem.6b00072

170 Julius, C., Riaz-Bradley, A. & Yuzenkova, Y. RNA capping by mitochondrial and multi-subunit RNA polymerases. *Transcription* **9**, 292-297 (2018). https://doi.org:10.1080/21541264.2018.1456258

171 Masters, B. S., Stohl, L. L. & Clayton, D. A. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **51**, 89-99 (1987). https://doi.org:10.1016/0092-8674(87)90013-4

172 Höfer, K. *et al.* Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol* **12**, 730-734 (2016). https://doi.org:10.1038/nchembio.2132

173 Zhang, D. *et al.* Structural basis of prokaryotic NAD-RNA decapping by NudC. *Cell Res* **26**, 1062-1066 (2016). https://doi.org:10.1038/cr.2016.98

174 Doamekpor, S. K. *et al.* DXO/Rai1 enzymes remove 5'-end FAD and dephospho-CoA caps on RNAs. *Nucleic Acids Res* **48**, 6136-6148 (2020). https://doi.org:10.1093/nar/gkaa297

175 Abele, F. *et al.* A Novel NAD-RNA Decapping Pathway Discovered by Synthetic Light-Up NAD-RNAs. *Biomolecules* **10** (2020). https://doi.org:10.3390/biom10040513

176 Höfer, K. & Jäschke, A. Epitranscriptomics: RNA Modifications in Bacteria and Archaea. *Microbiol Spectr* **6** (2018). https://doi.org:10.1128/microbiolspec.RWR-0015-2017

177 Jäschke, A., Höfer, K., Nubel, G. & Frindert, J. Cap-like structures in bacterial RNA and epitranscriptomic modification. *Curr Opin Microbiol* **30**, 44-49 (2016). https://doi.org:10.1016/j.mib.2015.12.009

178 Spangler, J. R. & Huang, F. The E. coli NudL enzyme is a Nudix hydrolase that cleaves CoA and its derivatives. *bioRxiv*, 2020.2001.2031.929182 (2020). https://doi.org:10.1101/2020.01.31.929182

179 Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **5**, 593-599 (2009). https://doi.org:10.1038/nchembio.186

180 Julius, C. & Yuzenkova, Y. Noncanonical RNA-capping: Discovery, mechanism, and physiological role debate. *Wiley Interdiscip Rev RNA* **10**, e1512 (2019). https://doi.org:10.1002/wrna.1512

181 Makarchikov, A. F. *et al.* Thiamine triphosphate and thiamine triphosphatase activities: from bacteria to mammals. *Cell Mol Life Sci* **60**, 1477-1488 (2003). https://doi.org:10.1007/s00018-003-3098-4

182 Frederich, M. *et al.* Thiaminylated adenine nucleotides. Chemical synthesis, structural characterization and natural occurrence. *FEBS J* **276**, 3256-3268 (2009). https://doi.org:10.1111/j.1742-4658.2009.07040.x

183 Bettendorff, L. *et al.* Discovery of a natural thiamine adenine nucleotide. *Nat Chem Biol* **3**, 211-212 (2007). https://doi.org:10.1038/nchembio867

184 Mohler, M., Höfer, K. & Jäschke, A. Synthesis of 5'-Thiamine-Capped RNA. *Molecules* **25**, 5492 (2020). https://doi.org:10.3390/molecules25235492

185 Fraga, H. & Fontes, R. Enzymatic synthesis of mono and dinucleoside polyphosphates. *Biochim Biophys Acta* **1810**, 1195-1204 (2011). https://doi.org:10.1016/j.bbagen.2011.09.010

186 Coste, H., Brevet, A., Plateau, P. & Blanquet, S. Non-Adenylylated Bis(5'-Nucleosidyl) Tetraphosphates Occur in Saccharomyces-Cerevisiae and in Escherichia-Coli and Accumulate Upon Temperature Shift or Exposure to Cadmium. *Journal of Biological Chemistry* **262**, 12096-12103 (1987). https://doi.org:https://doi.org/10.1016/S0021-9258(18)45321-5

187 Luciano, D. J., Levenson-Palmer, R. & Belasco, J. G. Stresses that Raise Np(4)A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Mol Cell* **75**, 957-966 e958 (2019). https://doi.org:10.1016/j.molcel.2019.05.031

188 Hudecek, O. *et al.* Dinucleoside polyphosphates act as 5'-RNA caps in bacteria. *Nat Commun* **11**, 1052 (2020). https://doi.org:10.1038/s41467-020-14896-8

189 Benoni, R., Culka, M., Hudecek, O., Gahurova, L. & Cahova, H. Dinucleoside Polyphosphates as RNA Building Blocks with Pairing Ability in Transcription Initiation. *ACS Chem Biol* **15**, 1765-1772 (2020). https://doi.org:10.1021/acschembio.0c00178

190 Luciano, D. J. & Belasco, J. G. Np(4)A alarmones function in bacteria as precursors to RNA caps. *Proc Natl Acad Sci U S A* **117**, 3560-3567 (2020). https://doi.org:10.1073/pnas.1914229117

191 Zamecnik, P. C., Stephenson, M. L., Janeway, C. M. & Randerath, K. Enzymatic synthesis of diadenosine tetraphosphate and diadenosine triphosphate with a purified lysyl-sRNA synthetase. *Biochem Biophys Res Commun* **24**, 91-97 (1966). https://doi.org:10.1016/0006-291x(66)90415-3

192 Munir, A., Banerjee, A. & Shuman, S. NAD+-dependent synthesis of a 5'-phospho-ADP-ribosylated RNA/DNA cap by RNA 2'-phosphotransferase Tpt1. *Nucleic Acids Res* **46**, 9617-9624 (2018). https://doi.org:10.1093/nar/gky792

193 Vinayak, M. & Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Bioscience Reports* **30**, 135-148 (2010). https://doi.org:10.1042/Bsr20090057

194 Yuan, Y. *et al.* Discovery of novel bacterial queuine salvage enzymes and pathways in human pathogens. *Proc Natl Acad Sci U S A* **116**, 19126-19135 (2019). https://doi.org:10.1073/pnas.1909604116

195 Vilfan, I. D. *et al.* Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *J Nanobiotechnology* **11**, 8 (2013). https://doi.org:10.1186/1477-3155-11-8

196 Schwartz, M. H. *et al.* Microbiome characterization by high-throughput transfer RNA sequencing and modification analysis. *Nat Commun* **9**, 5353 (2018). https://doi.org:10.1038/s41467-018-07675-z

197 Shakya, M., Lo, C. C. & Chain, P. S. G. Advances and Challenges in Metatranscriptomic Analysis. *Front Genet* **10**, 904 (2019). https://doi.org:10.3389/fgene.2019.00904

198 Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* **10**, 618-630 (2012). https://doi.org:10.1038/nrmicro2852

Chapter IV

The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections

Authors:
Pozhydaieva, N.*, Wolfram-Schauerte, M.*, Keuthen, H., & Höfer, K.

*contributed equally

Published in:

Current Opinion in Microbiology, 2024, 77, 102417

Author contributions:

N. P., M. W.-S. and H. K.: Conceptualization, Writing – original draft, Writing – review & editing.

K. H.: Conceptualization, Funding acquisition, Writing – review & editing.

4. The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections

4.1. Abstract

RNA modifications play essential roles in modulating RNA function, stability, and fate across all kingdoms of life. The entirety of the RNA modifications within a cell is defined as the epitranscriptome. While eukaryotic RNA modifications are intensively studied, understanding bacterial RNA modifications remains limited, and knowledge about bacteriophage RNA modifications is almost nonexistent. In this review, we shed light on known mechanisms of bacterial RNA modifications and propose how this knowledge might be extended to bacteriophages. We build hypotheses on enzymes potentially responsible for regulating the epitranscriptome of bacteriophages and their host. This review highlights the exciting prospects of uncovering the unexplored field of bacteriophage epitranscriptomics and its potential role to shape bacteriophage–host interactions.

4.2. Introduction

Since the discovery of bacteriophages (phages) more than 100 years ago, bacteriophage research has significantly impacted our understanding of fundamental biological processes[1]. Phages have been pivotal as model systems for understanding fundamental principles in molecular biology and discovering their biotechnological potential[2].

Lytic bacteriophages efficiently infect their bacterial host, completing the infection cycle with the release of new bacteriophage progeny through cell lysis. The infection process is highly regulated on the molecular level and typically exhibits a high degree of specificity for a given host–bacteria pair. Regardless of the specific host–bacteria pair, bacteriophage infections have consistently been observed as temporally highly regulated processes in various studies, revealing the precise timing of gene expression[3]. Therefore, the bacteriophage genes are classified into early, middle and late genes, signifying their timing in the infection cycle. To maintain efficient gene expression and, consequently, phage replication and propagation, lytic bacteriophages take control over the host's cellular machinery or its specific components.

Phage gene transcription is catalysed by either bacteriophage RNA polymerase (RNAP) or host-encoded RNAP. Based on our current textbook knowledge, RNAPs utilise the host-provided nucleotide pool consisting of uridine triphosphate (UTP), adenosine triphosphate (ATP), guanosine triphosphate (GTP) and cytidine triphosphate (CTP) to generate phage transcripts during the different infection phases. It is assumed that RNAs transcribed during each phase of phage infection are directly translated by ribosomes, resulting in proteins from the respective infection phase. However, recent multi-omics studies have revealed that the appearance of

transcripts and proteins in bacteriophages does not always coincide, showing, for instance, that early transcripts contribute to the translation of late proteins[4]. This observation suggests the presence of mechanisms that allow to distinguish between bacteriophage RNAs reflected by their time point of translation.

Moreover, recent studies show that the stability and processing of host and bacteriophage RNA differ strongly during infection, indicating precise distinction and selective degradation of transcripts[4-6].

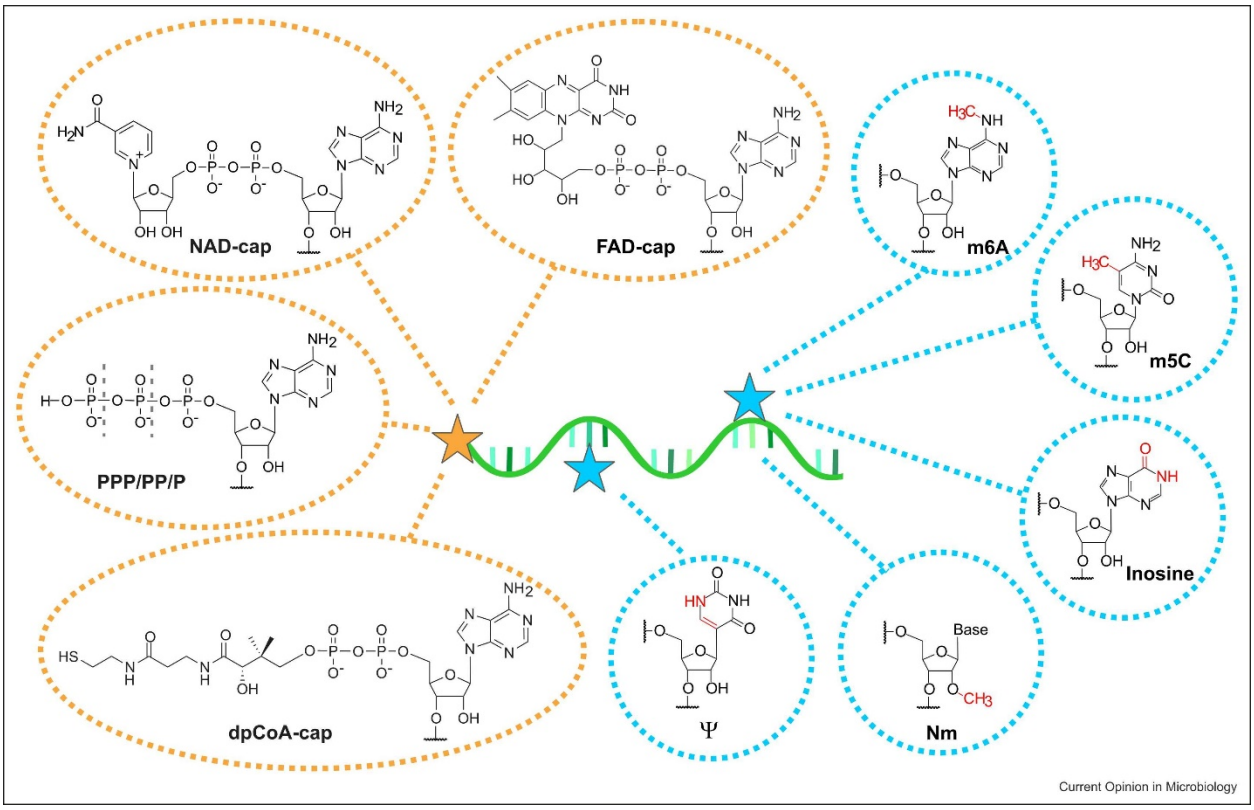
These findings suggest the existence of so-far-unknown additional mechanisms that enable the specific differentiation between phage and host transcripts and define their processing during infection, raising numerous questions, such as:

- How do RNA modifications influence the precise processing of bacteriophage and host transcripts during infection, despite their shared genetic building blocks?
- Could RNA modifications offer an additional mechanism for regulating phage infections?
- Which enzymes, supplied by both bacteriophages and bacteria, have the potential to shape the presence and function of RNA modifications?

Besides known factors influencing RNA stability and fate, such as RNA secondary structure or RNA-binding proteins, RNA modifications have been shown to regulate RNA processing in all domains of life. Incorporating chemical modifications into RNA strongly affects its biochemical properties, stability and function in cellular and biological processes[7]. These modifications can be categorised into internal (modifications on bases or nucleosides) and terminal (cap-like modifications at 5'-terminus or 3'-terminus) RNA modifications (Box 1) (reviewed in [8]). The diversity of the RNA modifications within a cell is collectively defined as the epitranscriptome. Proteins known as writers (biosynthesis of the modifications), readers (recognition) and erasers (removal) shape the epitranscriptome and its function(s) (reviewed in [8]). Their interplay results in a wide variety of more than 170 reported RNA modifications[9]. RNA modifications are found in all domains of life, with a significant focus on eukaryotes. However, the exploration and functional characterisation of RNA modifications are still evolving in bacteria, and even less is known about bacteriophages in this context.

Box1

Terminal (orange) and internal (blue) RNA modifications in bacteria.



For bacterial RNA, the 5´-terminus defines its origin. In primary transcripts, the initiating nucleoside triphosphate gives rise to the triphosphorylated 5´-end of the RNA. Through nucleolytic RNA processing, 5'-P-RNA is formed, and recent research has revealed the formation of diphosphorylated RNA as well[10]. However, the 5'-end of RNA can carry functional groups different from phosphates, due to transcription initiation with non-canonical initiating nucleotides. Here, nucleotide-based cofactors such as NAD (here used to refer to the oxidised NAD+), FAD and dpCoA can be accepted by RNAPs to initiate transcription, leading to the generation of 5'-capped RNAs. These 5'-RNA-caps, such as the NAD-cap, are assumed to protect the modified RNA from degradation by 5´-end-dependent endonucleases such as RNase E[11,12].

The role of internal RNA modifications differs strongly based on the type of the modified RNA. For bacterial RNA, m6A, 5-methylcytidine (m5C), inosine (I), pseudouridine (Ψ) and 2´-O-methylation (Nm) are the most common internal RNA modifications[8]. However, only two of them were detected to date in bacterial mRNA–m6A and inosine[13].

Whilst the role of m6A in bacterial rRNA is well-studied and has been determined to play a role in folding and stability, translational control and cell fitness[14-16], the presence of m6A in bacterial mRNA, where m6A is predominantly found within open reading frames (ORFs), remains relatively unexplored[13].

Inosine is a product of C6 deamination of adenosine, which occurs either by spontaneous hydrolysis or enzymatic conversion[17]. Inosine has been identified in mRNA of several bacteria, where it exerts a regulatory function. Its presence has a significant regulatory impact in bacteria and has been described to be involved in oxidative stress tolerance and the induction of biofilm formation[18,19].

Nm was described to have an impact on transcript stability and translation efficiency. Both, m5C and Ψ, were described as being present in tRNA, contributing to its structural stabilisation and influence on the translational fidelity of the ribosome[20].

Here, we review the current knowledge about selected RNA modifications in bacteria and explore the potential roles of known RNA modification mechanisms, encompassing writers, readers and erasers, in the context of bacteriophage infection. We focus on bacteriophage infections, with an emphasis on bacterial viruses, while excluding viruses that infect archaea. Further, we speculate on bacteriophage-encoded factors that may have potential roles in shaping bacteriophage and host epitranscriptomes during the infection (Figure 1).

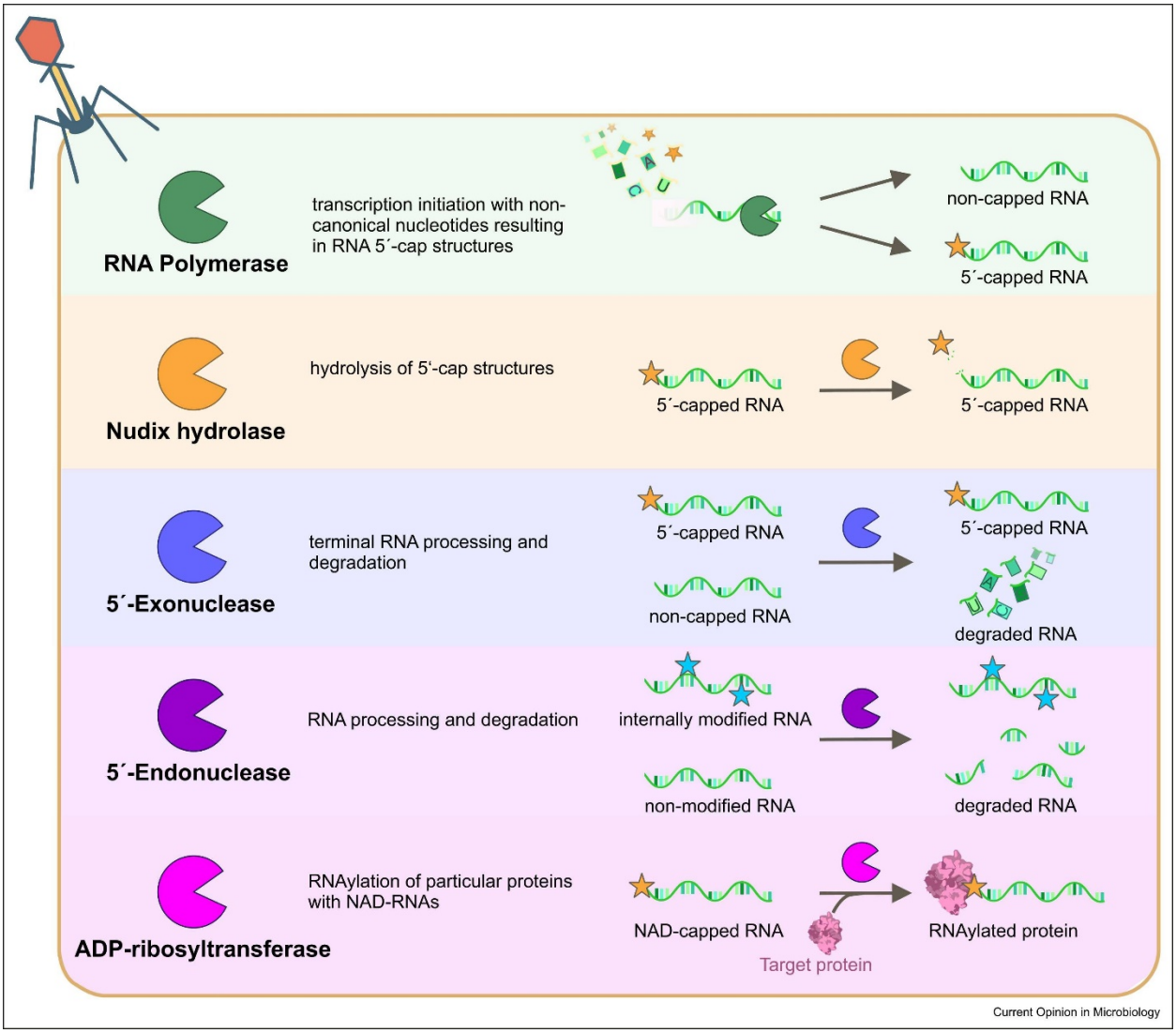


Figure 1. Putative modulators of the bacteriophage and host epitranscriptome during infection.

4.3. Decorating RNAs: RNA modification by host and bacteriophage RNA polymerases

Multi-subunit RNAPs play a pivotal role in shaping the epitranscriptome. In the bacterial hosts — during transcription initiation (*ab initio*) — RNAPs can incorporate non-canonical nucleotides at the 5'-end of RNAs, including nicotinamide adenine dinucleotide (NAD), flavine adenine dinucleotide (FAD) or dephospho-coenzyme A (dpCoA) — giving rise to cofactor-capped RNAs (Box 1, Table 1)[21,22]. RNAPs from *Escherichia coli*, *Bacillus subtilis* and *Staphylococcus aureus* have been implicated in non-canonical transcription initiation[21,23,24]. Thus, a broad range of bacterial hosts may be equipped with cofactor-capped RNA species before a potential infection by a bacteriophage (reviewed in [8,25]).

Table 1. Comparison of (hypothetical) biological roles of RNA modifications in bacteriophage infections.

A	***Biosynthesis of 5'-modified transcripts***	
	RNA Polymerase(s)	
	Host	**Bacteriophage**
Biological role	• Non-canonical transcription initiation (e.g. NAD, FAD and dephospho-CoA) *in vitro* and *in vivo*[21,22]	• Non-canonical transcription initiation (e.g. NAD, FAD and dephospho-CoA) *in vitro* [26] • Incorporation of modified nucleotides (e.g. pseudouridine) *in vitro*[26]
Potential role during bacteriophage infection	• Capping of host transcripts to protect them from degradation by bacteriophage enzymes • Exploitation of host RNAP for transcription of bacteriophage genes and capping of bacteriophage RNA to increase their stability	• Capping of bacteriophage transcripts to protect them from degradation by the host • Incorporation of the modified nucleotides during bacteriophage infection to enhance stability or avoid immune recognition of bacteriophage transcripts
B	***RNA processing and degradation***	
	Nudix hydrolases	
	Host	***Bacteriophage***
Biological role	• Decapping of capped RNAs (e.g. NAD-capped RNA) *in vitro* and *in vivo* by Nudix hydrolases[12] • Hydrolysis of PP from the 5´-PPP primary transcripts by pyrophosphohydrolases (e.g. RppH)[10,27]	• No knowledge about bacteriophage Nudix hydrolases involved in RNA processing in infections
Potential role during bacteriophage infection	• Decapping of bacteriophage transcripts by Nudix hydrolyses to destabilise bacteriophage transcripts and induce their degradation • Recruitment of host Nudix hydrolases by bacteriophages to induce the degradation of (capped) host transcripts	• Expression of own bacteriophage Nudix hydrolases to control the stability and degradation of 5´-capped host and bacteriophage transcripts

	Endo- and exonucleases	
Biological role	• Cofactor-caps as shields for RNA to protect against 5'-P-end-dependent nucleases such as RNase E[12]	• Unknown impact of RNA modifications on bacteriophage nucleases
Potential role during bacteriophage infection	• RNA modifications as potential epitranscriptomic marks for nucleases to distinguish own (host) from invader (bacteriophage) RNAs	• RNA modifications as distinct features of host or bacteriophage RNAs to trigger their cleavage
C	***Post-translational protein modification***	
	ADP-ribosyltransferase(s)	
Biological role	• No RNAylation observed in any bacteria	• RNAylation of the host's translational apparatus by the bacteriophage T4 ADP-ribosyltransferase ModB[28]
Potential role during bacteriophage infection	• Potential functions of bacterial ARTs in RNAylation to counteract bacteriophage infection	• RNAylation to modulate protein function during infection • Bacteriophage-mediated RNAylation of host proteins as a means to take control over the host cell

A) Biosynthesis of 5'-modified transcripts, B) RNA processing and degradation and C) post-translational modifications.
Modifier B (ModB).

Bacteriophages have different ways of controlling transcription through RNAPs (reviewed in [29,30]). On the one hand, bacteriophages such as T4 make use of the host's transcriptional apparatus throughout infection[31], as their genome does not encode an enzyme that can catalyse RNA biosynthesis. On the other hand, bacteriophages can encode their own (set of) RNAP(s) that they use to transcribe their own genes[29,30]. In the latter case, one can distinguish between virion-associated[32] and non-virion-associated RNAPs[30]. The virion-associated RNAP is co-injected with the bacteriophage genome into the host cell and ensures early transcription of bacteriophage genes not excluding its involvement in later stages of infection. Subsequently, non-virion RNAP

is synthesised from early bacteriophage genes during infection and drives middle and late transcription. Bacteriophages that only partially rely on the host RNAP may encode the non-virion RNAP only, which is transcribed during infection by the host RNAP.

Drawing from the elucidated mechanisms of RNA synthesis during phage infection, one can investigate whether host and/or bacteriophage RNAPs incorporate non-canonical building blocks, alongside the canonical RNA bases U, A, C and G, into phage transcripts.

Some bacteriophages, such as bacteriophage T4, utilise the host RNAP for the transcription of their own genes[31]. In this particular case, *E. coli* RNAPs can incorporate cofactor-caps upon transcription initiation, as described above, thereby likely defining the bacteriophage/host epitranscriptome during infection. Here, one may consider the various strategies of bacteriophages to hijack the host's transcriptional apparatus and its yet-unknown effect on the epitranscriptional regulation[29]. For instance, three adenosine diphosphate (ADP)-ribosyltransferases (ARTs) of bacteriophage T4 post-translationally modify host proteins with ADP-ribose from the substrate NAD to modulate cellular processes such as transcription[33-35]. The ARTs Alt and ModA ADP-ribosylate the host RNAP to direct its specificity towards bacteriophage genes[33,36]. It is so far unknown whether this post-translational protein modification (ADP-ribosylation) also influences the host RNAP's ability to initiate transcription with cofactors. One could imagine that ADP-ribosylation of RNAP could provide a means to incorporate RNA modifications to distinguish newly synthesised bacteriophage messenger RNAs (mRNAs) from host RNA.

The T7 bacteriophage is a well-studied example for a phage that requires the host RNAP only during the early phase of phage infection — in particular — until its own RNAP (T7 RNAP) is generated. T7 RNAP transcribes T7 genes in middle- and late- infection phase[37]. The T7 RNAP is probably one of the most famous single-subunit RNAPs (ssRNAPs) and is widely used for *in vitro* transcription applications and protein expression systems[37,38]. Moreover, the T7 RNAP stands out with its capability to incorporate non-canonical building blocks co-transcriptionally[26]. It caps RNAs with various cofactors, for example, NAD, with up to 50% efficiency *in vitro*[26] and accepts modified nucleotides such as pseudouridine as substrates to generate internal RNA modifications (Box 1, Table 1)[39]. Pseudouridine is a critical building block of the Covid-19 mRNA vaccines that are generated using large-scale *in vitro* transcriptions with T7 RNAP[40]. Besides its ability to incorporate non-canonical nucleotides into RNA, T7 RNAP exhibits several other advantageous properties for biotechnological applications. These include characteristics such as its single-subunit nature (unlike bacterial multi-subunit RNAPs), high specificity for the short T7 promoter (18 nt), higher transcriptional speed, independence from auxiliary transcription factors,

ease of engineering, possible application as a parallel expression system and the ability to produce long transcripts[41-43]. This emphasises the potential and broad applicability of bacteriophage RNAPs in regard to synthesising and modifying RNA. T7 RNAP can even replicate small RNAs or use deoxynucleotides as artificial RNA building blocks (reviewed in [44]). In addition, single-point mutations in T7 RNAP or its homologues can cause acceptance of deoxynucleotides or 2'-fluoro-ribonucleotides as alternative substrates for transcription[45-47]. However, *in vivo*, evidence for T7 RNAP function in installing RNA modifications during T7 infection is still missing. T7-like ssRNAPs are found in diverse bacteriophages with various host ranges. These include T7-like RNAPs in Pseudomonas aeruginosa bacteriophages[48], Klebsiella bacteriophages K11[49] and KP34[47], Salmonella bacteriophage SP[50] and Synechococcus bacteriophage Syn5[51]. However, whether these RNAPs may exhibit similar activities as T7 RNAP towards installing RNA modifications *in vitro*, remains elusive.

The larger bacteriophages, so-called Jumbo phages (genome usually > 200 kb), display additional interesting features of transcription (reviewed in [30]). Some Jumbo bacteriophages form a nucleus-like structure (pseudonucleus), enabling the compartmentalisation of phage DNA from the bacterial cytoplasm. This results in locally separated phage gene transcription within the host[52,53]. Upon infection, a pseudonucleus is formed to protect the bacteriophage DNA from bacterial nucleases and to allow transcription of the phage transcripts by the phage-encoded RNAP[54-56]. To the best of our knowledge, these nucleus-like compartments have not yet been investigated in terms of their exact molecular composition. It is plausible that this compartment created by the phage differs from the bacterial cytosol in terms of the abundance of nucleotides and epitranscriptional writers. This difference may create distinct transcriptional environments that either promote or hinder the incorporation of specific RNA modifications by the relevant RNAPs. On the other hand, one may argue that transcriptional environments in the cytosol and pseudonuclei may be similar to each other allowing to equip phage transcripts similar to host transcripts. These hypotheses might be exciting prospects for future studies of RNA biosynthesis in Jumbo bacteriophages.

Altogether, bacterial RNAPs can shape cofactor-capped transcriptomes of bacterial hosts and — depending on the bacteriophage's transcriptional strategy — might contribute to cofactor-capping of bacteriophage transcripts, although evidence is lacking so far. *In vitro*, bacteriophage RNAPs possess the capabilities to cofactor-cap transcripts and directly incorporate internal RNA modifications. It is likely that our current knowledge only scratches the surface of bacteriophage RNAP diversity such as single- and multi-subunit organisation, infection phase-specific occurrence as virion and non-virion RNAPs. Given the diverse features of RNAPs and their

capabilities to install RNA modifications, it is possible that RNA modifications may exist in and even regulate bacteriophage infections.

4.4. Cleaning up: removal of modifications by Nudix hydrolases

Bacterial Nudix hydrolases have been described to play an important function in the removal of bacterial RNA modifications (reviewed in [8])[57]. Nudix enzymes generally hydrolyse nucleoside diphosphates linked to a moiety X within their diphosphate moiety, thereby releasing a nucleoside monophosphate and a monophosphate-X group. In *E. coli*, 13 Nudix hydrolases are described, which vary in their substrate spectrum, ranging from nucleoside-based cofactors to modified RNA species[58]. A well-characterised Nudix hydrolase known to interact with RNA *in vivo* is *E. coli* RppH (EcRppH). It processes primary transcripts harbouring a 5'-triphosphate to 5'-monophosphorylated RNAs (5'-P-RNAs) (Table 1), favouring their degradation by 5'-end-dependent nucleases such as RNase E[27,58].

Bacteriophage T4 infection of *E. coli* represents a well-studied scenario for this. Here, EcRppH was suggested to promote mRNA decay by generating 5'-P-RNA that activates RNase E-mediated RNA decay[27,59]. Experimental evidence showing how far this mechanism regulates the presence of host and bacteriophage transcripts is missing.

Another bacterial Nudix hydrolase, described to interact with bacterial transcripts, is *E. coli* NudC (EcNudC). EcNudC is known to hydrolyse cofactor-caps from RNA, thereby decapping NAD-RNAs *in vivo*[11] and several cofactor-capped RNAs, such as NAD-, FAD- or CoA-RNAs *in vitro*[12,60,61]. Interestingly, in other bacteria, different enzymes are involved in NAD-RNA decapping. For instance, due to a missing NudC homologue in Bacillus subtilis, BsRppH performs decapping of NAD-RNAs[24].

In the context of bacteriophage infection, one could speculate that Nudix hydrolases may also be involved in processing both bacteriophage and host RNA modifications during infection. It needs to be investigated whether distinct host Nudix hydrolases may positively or negatively contribute to bacteriophage infections, for example, as a required prerequisite in the host cell or as a host defence strategy. Bacterial hosts employ various strategies to defend themselves from bacteriophage infection. For instance, the Thoeris system senses and aborts bacteriophage infection by depleting essential NAD from the cellular metabolite pool via conversion to cyclic-ADP-ribose[62]. One may hypothesise that the Thoeris and related systems may also act on NAD-capped RNA or may influence the NAD-capping of transcripts by RNAPs during phage infections by reducing the NAD pool. Also, it could be speculated that bacteriophages could sequester NAD in NAD-capped RNA to protect a minimum level of NAD in the cell.

Given the widespread conservation of Nudix hydrolases across all domains of life, one might question whether these enzymes are also found in bacteriophages that could potentially influence the epitranscriptome during the phage–host interaction. Interestingly, Nudix hydrolases have been predicted in various bacteriophages[63]. However, only two bacteriophage-derived Nudix hydrolases have been characterised[63,64]. The T4-like Vibrio bacteriophage KVP40 possesses the Nudix hydrolase domain protein NatV active on NAD, NADH and ADP-ribose, thereby regulating cellular NAD levels during infection[63]. The bacteriophage T4-encoded enzyme NudE.1 — named based on its sequence homology to *E. coli* Nudix hydrolase NudE — has been described to hydrolyse substrates such as ADP-ribose and FAD, both cofactor(-derived) compounds, as well as Ap3A, an alarmone involved in stress signalling[64]. Notably, the activity of bacteriophage Nudix hydrolases on cofactors may indicate that they could similarly be active on cofactor-capped RNAs, such as in the case of EcNudC (Table 1). However, it cannot be generalised that a Nudix hydrolase, active on a cofactor, can equally process a cofactor-capped RNA. For instance, EcNudE hydrolyses NAD, but is inactive on NAD-capped RNA[12], demonstrating distinct substrate specificities. Further, the occurrence of Nudix hydrolases in various bacteriophage genomes suggests yet-unexplored roles in nucleotide metabolism and potential functions in processing or removal of RNA modifications.

4.5. Taking control: RNA modifications in host take-over

As mentioned above, the identity, synthesis and removal of cofactor-caps (terminal modifications) such as the NAD-cap are characterised in bacteria. The general notion persists that cofactor-caps such as NAD-caps stabilise host transcripts compared with their 5’-triphosphorylated counterparts that are more easily degraded by 5’-end-dependent nucleases in bacteria[11,12,23,24]. The existence of cofactor-capped bacteriophage mRNAs has not been studied yet. In general, one might speculate that both internal and external RNA modifications could provide an additional regulatory layer for the bacteriophage to evade bacterial defence systems or to take-over the host cell. A recent study has indicated that bacteriophages may also use existing RNA modifications of the host to regulate cellular processes during infection[28]. Modifier B (ModB) — one of the ARTs from bacteriophage T4 — not only accepts NAD but also NAD-capped RNA as a substrate to attach entire RNA chains to host proteins in a covalent manner[28]. Through ModB-mediated 'RNAylation', bacteriophage T4 targets the host’s translational apparatus resulting in an efficient bacteriophage infection (Table 1)[28]. The concept of 'RNAylation' suggests a direct connection between RNA modification and post-translational protein modification, which may be present in additional bacteriophage–host interactions. ModB homologues appear to exist in various other

bacteriophages targeting Escherichia, Klebsiella or Salmonella as indicated by Blast search (Supplementary Table 1).

These findings exemplify that bacteriophages may exploit their proteins, such as in the case of ModB, to 'read' RNA modifications, facilitating host take-over or possibly immunity against bacteria, thereby underlying the epitranscriptome as an important aspect of bacteriophage infection.

Another RNA modification that could significantly impact RNA fate and stability during phage infection involves the addition of polyA tails at the 3´-end of RNA. In contrast to eukaryotes, where polyA tails are important features of mature mRNAs, polyA tails in prokaryotes actively promote RNA degradation (reviewed in [65]). Bacterial polyA polymerases such as Ec polyA RNAP I attach multiple adenosines to the 3'-end of transcripts, thereby destabilising the RNA (reviewed in [65]). PolyA tails of phage RNAs can occur on primary transcripts and after initial nucleolytic processing of the transcripts. Thus, polyA tails are located within or at the ends of protein-coding regions of RNA, indicating that they are added to a later stage in the life cycle of a phage RNA[66,67]. Importantly, polyA tailing is vital in order for some RNA fragments to be degraded[66]. However, it is unclear whether polyA tails are directly involved in overall destabilisation or even stabilisation of phage RNAs during infection[67]. One may speculate that the polyA tailing machinery of the host may either be used by the host to counteract phage infection by initiating RNA degradation. Alternatively, the phage may recruit this machinery to selectively enhance the degradation or stabilisation of phage and/or host transcripts.

4.6. Molecular duel: potential role of internal RNA modifications in bacteriophage infection

Until now, the functions of most RNA modifications in bacteria remain largely unknown — especially for mRNAs[8]. Ribosomal RNA (rRNA) and transfer RNA (tRNA) modifications play key roles in regulation and fine-tuning of translation. In particular, rRNA modifications impact the mRNA decoding efficiency[68-70], whereupon tRNA modifications are crucial for ensuring the stability, abundance and optimal affinity of tRNAs for the ribosomes[71-73]. tRNAs are highly modified RNA species decorated with various modifications, including — amongst others — pseudouridine and inosine (Box 1), which are installed by more than 20 different modifying enzymes in *E. coli* (reviewed in [74]). Intriguingly, hypomodification of tRNAs (modification at lower levels than usual) triggers their degradation in Vibrio cholera, exemplifying the importance of RNA modifications for RNA stability and decay[75]. Despite such important roles for rRNA and tRNA modifications, they have not yet been studied in bacteriophage infections. Host tRNA pools are often downregulated upon bacteriophage infections[4,76,77]. Could this be triggered by mechanisms

of tRNA hypomodification or downregulation of tRNA-modifying enzymes in the host as a response to infection? Some bacteriophages even encode their own tRNAs[76,78] that are expressed during infection[4,77]. Do these bacteriophage-encoded tRNAs also contain modifications such as pseudouridine or inosine? Do host enzymes install these modifications, do bacteriophages direct these enzymes to specifically act on these tRNAs or do they encode their own modifying enzymes? These questions only exemplify which variety and mechanisms of RNA modifications could play a role in bacteriophage infections and how they might shape central processes such as translation.

Pseudouridine may have become one of the most famous internal RNA modifications (Box 1). It is an integral part of mRNA-based vaccines, which prevents recognition of the RNA by the innate immune system of human cells[40,79]. Vaccines without RNA modification trigger the innate immune response and are thus less effective[40]. Key players in the innate immune response are toll-interleukin-1 receptor (TIR) domain proteins that sense the immunogenic material and activate signalling cascades (reviewed in [80]). Interestingly, homologues of eukaryotic TIR domains are also found in bacterial defence systems, as recently demonstrated for the Thoeris system[62]. Based on the conservation of such innate immune recognition systems across all domains of life, one may speculate that immune recognition of RNA modifications may take place in bacteriophage–host interactions as well. For instance, when bacteriophages with single- or double-stranded RNA genomes[81] (reviewed in [82]) infect their bacterial hosts. One could imagine that these genomes are decorated with RNA modifications that trigger anti-phage defence mechanisms. On the other hand, bacteriophages may install host-like RNA modifications on their transcripts/genomes in order to avoid immune recognition and anti-bacteriophage defence.

Another RNA-modifying event in bacteria is A-to-I editing (Box 1). This has been observed in various bacterial species, including *E. coli*. For example, in *E. coli*, the enzyme TadA is responsible for converting adenosine 34 in the anticodon of tRNAs to inosine[83]. Interestingly, TadA was shown to perform A-to-I editing in a small set of mRNAs, too, which changes distinct codons, as inosine is read as guanosine[18]. As a result, different amino acids are incorporated into the encoded proteins impacting protein function, for instance, increasing the toxicity of the HokB protein[18]. Blast search indicates that bacteriophages may possess TadA homologues in their genomes (Supplementary Table 1). Thus, one may hypothesise that bacteriophages could fine-tune protein expression and diversity through targeted A-to-I editing of mRNAs with self- or host-encoded factors. Additionally, it can be assumed that inosine is also present within phage mRNAs. While no information is currently available for bacteriophages, A-to-I conversions have been observed in eukaryotic viruses. These conversions are host-dependent deamination and

advantageous for the virus in evading the host immune response and reducing virus toxicity[84,85]. Interestingly, A-to-I editing also occurs as an anti-phage defence mechanism in some bacteria. The restriction by an adenosine deaminase acting on RNA (RADAR) system senses phage infection and converts (deoxy)ATP to (deoxy)inosine triphosphate (ITP) by deamination of adenosine to inosine[86]. This limits phage infection by inhibiting phage DNA replication and creating an imbalance in the cellular nucleotide pool[86,87]. The RADAR system is able to bind and translocate RNA[87], however, A-to-I editing of RNA by the RADAR system has not yet been observed *in vitro* and *in vivo*[86,87]. Nevertheless, this exemplifies the importance of nucleotides and their modifications during phage infections.

Along this line, also other RNA modifications have been detected in eukaryotic viruses. For instance, N6-methyladenosine (m6A) has been found in transcripts of eukaryotic DNA viruses and in the genomes of RNA viruses[88]. These modifications play a role in the regulation of viral replication and protection from the innate immune response of the host[88]. Similar protective mechanisms against the host's innate immune response can be expected in bacteriophages, given their constant evolutionary race with bacteria[89].

In summary, although the presence and impact of RNA modifications within the phage transcriptome have not been reported, the existence of RNA modifications in eukaryotic viruses and the potential discovery of homologues of bacterial and eukaryotic RNA-modifying enzymes, such as TadA, Mettl3 and RluF in phages (Supplementary Table 1), suggest that internal RNA modifications are likely present in phages. In such a scenario, the presence of internal modifications within phage transcripts and possibly varying levels of these modifications would help to address one of initial questions of this review: how do RNA modifications influence the precise processing and degradation of bacteriophage and host transcripts during infection?

4.7. Housekeeping nucleases: RNA modulation in phage–host crosstalk

RNA synthesis, processing and its selective degradation are key processes during bacteriophage infection. Based on available time-resolved transcriptomic studies during phage infection, a rapid degradation of host RNA is observed, while phage transcripts are actively transcribed and remain preserved[28,90]. As the abundance of host transcripts is already strongly decreased within the first minutes of infection, it suggests the presence of selective RNA degradation mechanisms during the infection, for example, nuclease-based degradation. The degradation of RNA by bacterial nucleases has been reported to be impacted by terminal RNA modifications. Nevertheless, research regarding the influence of RNA modifications in the context of phage–host interaction is currently lacking[11]. Both bacteriophage and host possess their own set of nucleases, which might

selectively process and degrade RNA throughout the infection process. Here the following questions arise: how do these nucleases distinguish between phage and bacterial RNA when both types of transcripts are composed of the same four nucleotide building blocks? And what constitutes the molecular basis for discerning between host and phage RNA? In addition to the sequence specificity of nucleases and the impact of RNA secondary structure motifs on recognition and cleavage, RNA modifications can exert an impact on RNA stability and its susceptibility to nucleases[11,91,92].

Studies that characterise the processing of transcripts during phage infection in the context of RNA modifications are missing to date. However, initial insights into the potential impact of RNA modifications on RNA processing have been gleaned from previous studies, particularly in the case of certain nucleases. For example, in *E. coli*, RNase E plays a central role in RNA processing and has also been implicated in the bacteriophage T4 infection cycle through the processing of gene-32 mRNA, which is crucial for T4 DNA replication and repair[66,90,93-96]. Nevertheless, cap structures, such as the NAD-cap, can protect transcripts from degradation mediated by RNase E, thereby providing another level of post-transcriptional regulation[11]. As described above, EcRNAP performs NAD-capping of RNA [21] and bacteriophage T4 relies on EcRNAP to express its genes[31]. Thus, it is conceivable that bacteriophages may utilise cofactor-capping of their transcripts to modulate RNA stability against host nucleases such as RNase E.

Another potential strategy employed by bacteriophages to counteract host nucleases through RNA modifications is illustrated by bacteriophage tRNAs. It is conceivable that phage-encoded tRNAs have evolved to withstand the impact of host anticodon nucleases that deactivate tRNAs by cleaving within the anticodon. Notably, the anticodon of tRNAs is a heavily modified RNA region[74] and one might speculate that RNA modifications might influence its cleavage by nucleases, including VapC, PrrC, Colicin D and Colicin E5. The latter might play a fundamental role in the host's defence against viral infections[76].

With numerous instances of nucleases actively processing both phage and host RNA during the infection process, it becomes increasingly clear that understanding how these nucleases differentiate between phage and host transcripts is of paramount importance. This inquiry is particularly promising when considering the role of RNA modifications in this selective process, as it sheds light on the intricate mechanisms at play during phage infection.

4.8. Conclusions

In this review, we illuminated the known and potential writers, readers and erasers of RNA modifications in both bacteria and bacteriophages. The field of epitranscriptomics during phage

infection remains understudied and the presence of RNA modifications in bacteriophages has yet to be proved. This review highlights that both bacteria and bacteriophages harbour a variety of genes that could encode potential writers, erasers and readers of internal and terminal RNA modifications.

The observed resilience of bacteriophages in maintaining infections and effectively hijacking their host, regardless of the presence of numerous anti-phage defence systems, strongly suggests the involvement of RNA modifications in bacteriophage infections. The modifications of phage RNA potentially offer an additional protective layer to phage RNA, making it less susceptible to degradation by the bacterial immune system. Moreover, phage RNA modifications can act as a factor distinguishing between phage and host RNA during infection, a process observed but not yet explored on a molecular level[4]. On the other hand, it is plausible that phages do not just modify their RNA but also target host RNA, potentially altering its function through these modifications. It is also conceivable that bacteriophages may reprogram host nucleases to hinder the introduction of modifications or exploit them for their own advantage. Therefore, both terminal and internal modifications, as discussed in this review, could significantly impact bacteriophage infections, enabling phages to manipulate their host or, conversely, contributing to anti-phage defence mechanisms.

All these possibilities become more credible and of higher relevance of investigation, when considering the current research in the field of the epitranscriptome of eukaryotic viruses. For eukaryotic viruses, the impact of RNA modifications on the regulation of host take-over during the infection was already observed, strongly contributing to efficient infection, for example, the substitution of uridine by pseudouridine enhances RNA stability and decreases anti-RNA immune response[97], or the introduction of the FAD-cap protects viral RNA from innate immune recognition[98]. Therefore, given the features of the RNA modifications and their prevalence across all domains of life, and across eukaryotic viruses, it is reasonable to speculate their existence in bacteriophages and potential roles in infection regulation.

Identifying modified RNA building blocks using well-established methods (reviewed in [8]) is the key to addressing this question. Such studies will reveal valuable insights into bacteriophage infections, provide mechanistic details of infection regulation and potentially unveil novel bacteriophage and host immune systems. Moreover, the rapid advancements in sequencing technologies, especially third-generation methods such as direct RNA-sequencing[99], hold the promise of simultaneously determining both the transcriptome and epitranscriptome in the future, which would be immensely beneficial for this research field.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: K.H. is in the process of applying for a patent (PCT/EP2021/071295) covering the RNAylation that lists K.H. as inventor. The remaining authors declare no competing interests.

Acknowledgements

This project has received funding from the European Research Council, Belgium under the European Union's Horizon 2020 research and innovation programme (ERC-2023-STG grant 101114948, NAD-ART to K.H.), from the German Research Council (DFG: SPP 2330 project number 464500427, RTG 2355 project 11 and RTG 2927 project 06 to K.H.), from the Max Planck Society (Max Planck Research Group Leader funding to KH), from the Studienstiftung des Deutschen Volkes e.V. (PhD scholarship to M.W.S.) and from the Joachim Herz Stiftung (Add-on Fellowships for Interdisciplinary Life Science to M.W.S.).

Supplementary material

Supplementary Table 1: Results of the BLAST search for selected enzymes.

Supplementary Table 1 is available at: https://zenodo.org/records/10866140

Data Availability

NCBI BLAST+ was used to perform a homology search for Modifier B (ModB) (UniProtID: P39423), TadA (UniProtID: P68398), METTL3 (UniProtID: F1R777) and RluF (UniProtID: P32684) on the EMBL website (https://www.ebi.ac.uk/Tools/sss/ncbiblast/, accessed on 26/10/2023). UniProtKB Viruses database was used for the search.

4.9. References

1 Clokie, M. R., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31-45 (2011). https://doi.org:10.4161/bact.1.1.14942

2 Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. *Bioessays* **37**, 6-9 (2015). https://doi.org:10.1002/bies.201400152

3 Yang, H. *et al.* Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered* **5**, 300-304 (2014). https://doi.org:10.4161/bioe.32110

4 Wolfram-Schauerte, M., Pozhydaieva, N., Viering, M., Glatter, T. & Höfer, K. Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of E. coli on Proteome and Transcriptome Levels. *Viruses-Basel* **14**, 2502 (2022). https://doi.org:10.3390/v14112502

5 Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134-141 (2004). https://doi.org:10.1016/j.virol.2004.08.001

6 Wicke, L. *et al.* Introducing differential RNA-seq mapping to track the early infection phase for Pseudomonas phage ϕKZ. *RNA Biol* **18**, 1099-1110 (2021). https://doi.org:10.1080/15476286.2020.1827785

7 Shi, H., Chai, P., Jia, R. & Fan, X. Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation. *Mol Cancer* **19**, 78 (2020). https://doi.org:10.1186/s12943-020-01194-6

8 Schauerte, M., Pozhydaieva, N. & Höfer, K. Shaping the Bacterial Epitranscriptome-5'-Terminal and Internal RNA Modifications. *Adv Biol (Weinh)* **5**, e2100834 (2021). https://doi.org:10.1002/adbi.202100834

9 Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res* **50**, D231-D235 (2022). https://doi.org:10.1093/nar/gkab1083

10 Luciano, D. J., Vasilyev, N., Richards, J., Serganov, A. & Belasco, J. G. A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in Escherichia coli. *Mol Cell* **67**, 44-54 e46 (2017). https://doi.org:10.1016/j.molcel.2017.05.035

11 Cahova, H., Winz, M. L., Höfer, K., Nubel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

12 Höfer, K. *et al.* Structure and function of the bacterial decapping enzyme NudC. *Nat Chem Biol* **12**, 730-734 (2016). https://doi.org:10.1038/nchembio.2132

13 Deng, X. *et al.* Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **43**, 6557-6567 (2015). https://doi.org:10.1093/nar/gkv596

14 Golovina, A. Y. *et al.* The last rRNA methyltransferase of E. coli revealed: the yhiR gene encodes adenine-N6 methyltransferase specific for modification of A2030 of 23S ribosomal RNA. *RNA* **18**, 1725-1734 (2012). https://doi.org:10.1261/rna.034207.112

15 Sergiev, P. V., Serebryakova, M. V., Bogdanov, A. A. & Dontsova, O. A. The ybiN gene of Escherichia coli encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. *J Mol Biol* **375**, 291-300 (2008). https://doi.org:10.1016/j.jmb.2007.10.051

16 Ma, H. *et al.* N(6-)Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nat Chem Biol* **15**, 88-94 (2019). https://doi.org:10.1038/s41589-018-0184-3

17 Alseth, I., Dalhus, B. & Bjoras, M. Inosine in DNA and RNA. *Curr Opin Genet Dev* **26**, 116-123 (2014). https://doi.org:10.1016/j.gde.2014.07.008

18 Bar-Yaacov, D. *et al.* RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res* **27**, 1696-1703 (2017). https://doi.org:10.1101/gr.222760.117

19 Nie, W. *et al.* A-to-I RNA editing in bacteria increases pathogenicity and tolerance to oxidative stress. *PLoS Pathog* **16**, e1008740 (2020). https://doi.org:10.1371/journal.ppat.1008740

20 Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **9**, e1003602 (2013). https://doi.org:10.1371/journal.pgen.1003602

21 Bird, J. G. *et al.* The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. *Nature* **535**, 444-447 (2016). https://doi.org:10.1038/nature18622

22 Julius, C. & Yuzenkova, Y. Bacterial RNA polymerase caps RNA with various cofactors and cell wall precursors. *Nucleic Acids Res* **45**, 8282-8290 (2017). https://doi.org:10.1093/nar/gkx452

23 Morales-Filloy, H. G. *et al.* The 5' NAD Cap of RNAIII Modulates Toxin Production in Staphylococcus aureus Isolates. *J Bacteriol* **202** (2020). https://doi.org:10.1128/JB.00591-19

24 Frindert, J. *et al.* Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in B. subtilis. *Cell Rep* **24**, 1890-1901 e1898 (2018). https://doi.org:10.1016/j.celrep.2018.07.047

25 Wolfram-Schauerte, M. & Höfer, K. NAD-capped RNAs - a redox cofactor meets RNA. *Trends Biochem Sci* **48**, 142-155 (2023). https://doi.org:10.1016/j.tibs.2022.08.004

26 Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res* **31**, e8 (2003). https://doi.org:10.1093/nar/gng008

27 Deana, A., Celesnik, H. & Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**, 355-358 (2008). https://doi.org:10.1038/nature06475

28 Wolfram-Schauerte, M. *et al.* A viral ADP-ribosyltransferase attaches RNA chains to host proteins. *Nature* **620**, 1054-1062 (2023). https://doi.org:10.1038/s41586-023-06429-2

29 Tabib-Salazar, A., Mulvenna, N., Severinov, K., Matthews, S. J. & Wigneshweraraj, S. Xenogeneic Regulation of the Bacterial Transcription Machinery. *J Mol Biol* **431**, 4078-4092 (2019). https://doi.org:10.1016/j.jmb.2019.02.008

30 Sokolova, M. L., Misovetc, I. & K, V. S. Multisubunit RNA Polymerases of Jumbo Bacteriophages. *Viruses-Basel* **12** (2020). https://doi.org:10.3390/v12101064

31 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

32 Falco, S. C., Laan, K. V. & Rothman-Denes, L. B. Virion-associated RNA polymerase required for bacteriophage N4 development. *Proc Natl Acad Sci U S A* **74**, 520-523 (1977). https://doi.org:10.1073/pnas.74.2.520

33 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004). https://doi.org:10.1128/JB.186.21.7262-7272.2004

34 Tiemann, B., Depping, R. & Rüger, W. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. *Gene Expr* **8**, 187-196 (1999).

35 Depping, R., Lohaus, C., Meyer, H. E. & Rüger, W. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochem Biophys Res Commun* **335**, 1217-1223 (2005). https://doi.org:10.1016/j.bbrc.2005.08.023

36 Koch, T., Raudonikiene, A., Wilkens, K. & Rüger, W. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. *Gene Expr* **4**, 253-264 (1995).

37 Summers, W. C. & Siegel, R. B. Control of template specificity of E. coli RNA polymerase by a phage-coded protein. *Nature* **223**, 1111-1113 (1969). https://doi.org:10.1038/2231111a0

38 Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int J Biol Macromol* **118**, 49-56 (2018). https://doi.org:10.1016/j.ijbiomac.2018.05.198

39 Chen, T. H., Potapov, V., Dai, N., Ong, J. L. & Roy, B. N(1)-methyl-pseudouridine is incorporated with higher fidelity than pseudouridine in synthetic RNAs. *Sci Rep* **12**, 13017 (2022). https://doi.org:10.1038/s41598-022-17249-1

40 Morais, P., Adachi, H. & Yu, Y. T. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. *Front Cell Dev Biol* **9**, 789427 (2021). https://doi.org:10.3389/fcell.2021.789427

41 Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393-404 (2004). https://doi.org:10.1016/s0092-8674(04)00120-5

42 Wang, W. *et al.* Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnol Adv* **36**, 2129-2137 (2018). https://doi.org:10.1016/j.biotechadv.2018.10.001

43 Kortmann, M., Kuhl, V., Klaffl, S. & Bott, M. A chromosomally encoded T7 RNA polymerase-dependent gene expression system for Corynebacterium glutamicum: construction and comparative evaluation at the single-cell level. *Microb Biotechnol* **8**, 253-265 (2015). https://doi.org:10.1111/1751-7915.12236

44 McAllister, W. T. & Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol Microbiol* **10**, 1-6 (1993). https://doi.org:10.1111/j.1365-2958.1993.tb00897.x

45 Zhu, B. *et al.* Synthesis of 2'-Fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Res* **43**, e94 (2015). https://doi.org:10.1093/nar/gkv367

46 Padilla, R. & Sousa, R. Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2'-groups using a mutantT7 RNA polymerase (RNAP). *Nucleic Acids Res* **27**, 1561-1563 (1999). https://doi.org:10.1093/nar/27.6.1561

47 Lu, X. *et al.* Klebsiella Phage KP34 RNA Polymerase and Its Use in RNA Synthesis. *Front Microbiol* **10**, 2487 (2019). https://doi.org:10.3389/fmicb.2019.02487

48 Lammens, E. M., Feyaerts, N., Kerremans, A., Boon, M. & Lavigne, R. Assessing the Orthogonality of Phage-Encoded RNA Polymerases for Tailored Synthetic Biology Applications in Pseudomonas Species. *Int J Mol Sci* **24** (2023). https://doi.org:10.3390/ijms24087175

49 Rong, M., Castagna, R. & McAllister, W. T. Cloning and purification of bacteriophage K11 RNA polymerase. *Biotechniques* **27**, 690, 692, 694 (1999). https://doi.org:10.2144/99274bm13

50 Krieg, P. A. & Melton, D. A. In vitro RNA synthesis with SP6 RNA polymerase. *Methods Enzymol* **155**, 397-415 (1987). https://doi.org:10.1016/0076-6879(87)55027-3

51 Zhu, B. *et al.* The RNA polymerase of marine cyanophage Syn5. *J Biol Chem* **288**, 3545-3552 (2013). https://doi.org:10.1074/jbc.M112.442350

52 Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425-431 (2020). https://doi.org:10.1038/s41586-020-2007-4

53 Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194-197 (2017). https://doi.org:10.1126/science.aal2130

54 Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat Microbiol* **5**, 48-55 (2020). https://doi.org:10.1038/s41564-019-0612-5

55 Lavysh, D. *et al.* The genome of AR9, a giant transducing Bacillus phage encoding two multisubunit RNA polymerases. *Virology* **495**, 185-196 (2016). https://doi.org:10.1016/j.virol.2016.04.030

56 Takahashi, I. & Marmur, J. Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for Bacillus subtilis. *Nature* **197**, 794-795 (1963). https://doi.org:10.1038/197794a0

57 Sharma, S. *et al.* Mammalian Nudix proteins cleave nucleotide metabolite caps on RNAs. *Nucleic Acids Res* **48**, 6788-6798 (2020). https://doi.org:10.1093/nar/gkaa402

58 McLennan, A. G. The Nudix hydrolase superfamily. *Cell Mol Life Sci* **63**, 123-143 (2006). https://doi.org:10.1007/s00018-005-5386-7

59 Qi, D., Alawneh, A. M., Yonesaki, T. & Otsuka, Y. Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. *Genetics* **201**, 977-987 (2015). https://doi.org:10.1534/genetics.115.180364

60 Zhou, W. *et al.* Structural insights into dpCoA-RNA decapping by NudC. *RNA Biol* **18**, 244-253 (2021). https://doi.org:10.1080/15476286.2021.1936837

61 Vvedenskaya, I. O. *et al.* CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD<sup/>. *Mol Cell* **70**, 553-564 e559 (2018). https://doi.org:10.1016/j.molcel.2018.03.014

62 Ofir, G. *et al.* Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116-120 (2021). https://doi.org:10.1038/s41586-021-04098-7

63 Lee, J. Y., Li, Z. & Miller, E. S. Vibrio Phage KVP40 Encodes a Functional NAD(+) Salvage Pathway. *J Bacteriol* **199** (2017). https://doi.org:10.1128/JB.00855-16

64 Xu, W., Gauss, P., Shen, J., Dunn, C. A. & Bessman, M. J. The gene e.1 (nudE.1) of T4 bacteriophage designates a new member of the Nudix hydrolase superfamily active on flavin adenine dinucleotide, adenosine 5'-triphospho-5'-adenosine, and ADP-ribose. *J Biol Chem* **277**, 23181-23185 (2002). https://doi.org:10.1074/jbc.M203325200

65 Hajnsdorf, E. & Kaberdin, V. R. RNA polyadenylation and its consequences in prokaryotes. *Philos Trans R Soc Lond B Biol Sci* **373** (2018). https://doi.org:10.1098/rstb.2018.0166

66 Goodrich, A. F. & Steege, D. A. Roles of polyadenylation and nucleolytic cleavage in the filamentous phage mRNA processing and decay pathways in Escherichia coli. *RNA* **5**, 972-985 (1999). https://doi.org:10.1017/s1355838299990398

67 Johnson, M. D., Popowski, J., Cao, G. J., Shen, P. & Sarkar, N. Bacteriophage T7 mRNA is polyadenylated. *Mol Microbiol* **27**, 23-30 (1998). https://doi.org:10.1046/j.1365-2958.1998.00649.x

68 Burakovsky, D. E. *et al.* Impact of methylations of m2G966/m5C967 in 16S rRNA on bacterial fitness and translation initiation. *Nucleic Acids Res* **40**, 7885-7895 (2012). https://doi.org:10.1093/nar/gks508

69 Valadon, C. & Namy, O. The Importance of the Epi-Transcriptome in Translation Fidelity. *Noncoding RNA* **7** (2021). https://doi.org:10.3390/ncrna7030051

70 Hughes, K. J., Chen, X., Burroughs, A. M., Aravind, L. & Wolin, S. L. An RNA Repair Operon Regulated by Damaged tRNAs. *Cell Rep* **33**, 108527 (2020). https://doi.org:10.1016/j.celrep.2020.108527

71 Noguchi, S., Nishimura, Y., Hirota, Y. & Nishimura, S. Isolation and characterization of an Escherichia coli mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA. *J Biol Chem* **257**, 6544-6550 (1982).

72 Vicens, Q., Paukstelis, P. J., Westhof, E., Lambowitz, A. M. & Cech, T. R. Toward predicting self-splicing and protein-facilitated splicing of group I introns. *RNA* **14**, 2013-2029 (2008). https://doi.org:10.1261/rna.1027208

73 Motorin, Y. & Helm, M. tRNA stabilization by modified nucleotides. *Biochemistry* **49**, 4934-4944 (2010). https://doi.org:10.1021/bi100408z

74 de Crecy-Lagard, V. & Jaroch, M. Functions of Bacterial tRNA Modifications: From Ubiquity to Diversity. *Trends Microbiol* **29**, 41-53 (2021). https://doi.org:10.1016/j.tim.2020.06.010

75 Kimura, S. & Waldor, M. K. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *Proc Natl Acad Sci U S A* **116**, 1394-1403 (2019). https://doi.org:10.1073/pnas.1814130116

76 van den Berg, D. F., van der Steen, B. A., Costa, A. R. & Brouns, S. J. J. Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations. *Elife* **12** (2023). https://doi.org:10.7554/eLife.85183

77 Yang, J. Y. *et al.* Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Syst* **12**, 771-779 e775 (2021). https://doi.org:10.1016/j.cels.2021.05.019

78 Bailly-Bechet, M., Vergassola, M. & Rocha, E. Causes for the intriguing presence of tRNAs in phages. *Genome Res* **17**, 1486-1495 (2007). https://doi.org:10.1101/gr.6649807

79 Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165-175 (2005). https://doi.org:10.1016/j.immuni.2005.06.008

80 O'Neill, L. A. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol* **7**, 353-364 (2007). https://doi.org:10.1038/nri2079

81 Yang, Y. *et al.* Corrigendum: Characterization of the first double-stranded RNA bacteriophage infecting Pseudomonas aeruginosa. *Sci Rep* **7**, 46853 (2017). https://doi.org:10.1038/srep46853

82 Callanan, J. *et al.* RNA Phage Biology in a Metagenomic Era. *Viruses-Basel* **10** (2018). https://doi.org:10.3390/v10070386

83 Rafels-Ybern, A., Torres, A. G., Grau-Bove, X., Ruiz-Trillo, I. & Ribas de Pouplana, L. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol* **15**, 500-507 (2018). https://doi.org:10.1080/15476286.2017.1358348

84 Zhu, T. *et al.* Host-mediated RNA editing in viruses. *Biol Direct* **18**, 12 (2023). https://doi.org:10.1186/s13062-023-00366-w

85 Volchkova, V. A., Dolnik, O., Martinez, M. J., Reynard, O. & Volchkov, V. E. RNA Editing of the GP Gene of Ebola Virus is an Important Pathogenicity Factor. *J Infect Dis* **212 Suppl 2**, S226-233 (2015). https://doi.org:10.1093/infdis/jiv309

86 Duncan-Lowey, B. *et al.* Cryo-EM structure of the RADAR supramolecular anti-phage defense complex. *Cell* **186**, 987-998 e915 (2023). https://doi.org:10.1016/j.cell.2023.01.012

87 Gao, Y. *et al.* Molecular basis of RADAR anti-phage supramolecular assemblies. *Cell* **186**, 999-1012 e1020 (2023). https://doi.org:10.1016/j.cell.2023.01.026

88 Li, H., Guo, Y., Qi, W. & Liao, M. N(6)-methyladenosine modification of viral RNA and its role during the recognition process of RIG-I-like receptors. *Front Immunol* **13**, 1031200 (2022). https://doi.org:10.3389/fimmu.2022.1031200

89 Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat Rev Microbiol* **21**, 686-700 (2023). https://doi.org:10.1038/s41579-023-00934-x

90 Uzan, M. & Miller, E. S. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. *Virol J* **7**, 360 (2010). https://doi.org:10.1186/1743-422X-7-360

91 Feng, Y., Vickers, T. A. & Cohen, S. N. The catalytic domain of RNase E shows inherent 3' to 5' directionality in cleavage site selection. *Proc Natl Acad Sci U S A* **99**, 14746-14751 (2002). https://doi.org:10.1073/pnas.202590899

92 Zhang, D. *et al.* Structural basis of prokaryotic NAD-RNA decapping by NudC. *Cell Res* **26**, 1062-1066 (2016). https://doi.org:10.1038/cr.2016.98

93 Lambrecht, S. J., Stappert, N., Sommer, F., Schroda, M. & Steglich, C. A Cyanophage MarR-Type Transcription Factor Regulates Host RNase E Expression during Infection. *Microorganisms* **10** (2022). https://doi.org:10.3390/microorganisms10112245

94 Kokoska, R. J. & Steege, D. A. Appropriate expression of filamentous phage f1 DNA replication genes II and X requires RNase E-dependent processing and separate mRNAs. *J Bacteriol* **180**, 3245-3249 (1998). https://doi.org:10.1128/JB.180.12.3245-3249.1998

95 Mudd, E. A., Prentki, P., Belin, D. & Krisch, H. M. Processing of unstable bacteriophage T4 gene 32 mRNAs into a stable species requires Escherichia coli ribonuclease E. *EMBO J* **7**, 3601-3607 (1988). https://doi.org:10.1002/j.1460-2075.1988.tb03238.x

96 Uzan, M. RNA processing and decay in bacteriophage T4. *Prog Mol Biol Transl Sci* **85**, 43-89 (2009). https://doi.org:10.1016/S0079-6603(08)00802-7

97 Martinez Campos, C. *et al.* Mapping of pseudouridine residues on cellular and viral transcripts using a novel antibody-based technique. *RNA* **27**, 1400-1411 (2021). https://doi.org:10.1261/rna.078940.121

98 Sherwood, A. V. *et al.* Hepatitis C virus RNA is 5'-capped with flavin adenine dinucleotide. *Nature* **619**, 811-818 (2023). https://doi.org:10.1038/s41586-023-06301-3

99 Jain, M., Abu-Shumays, R., Olsen, H. E. & Akeson, M. Advances in nanopore direct RNA sequencing. *Nat Methods* **19**, 1160-1164 (2022). https://doi.org:10.1038/s41592-022-01633-w

Chapter V

Temporal epigenome modulation enables efficient bacteriophage engineering and functional analysis of phage DNA modifications

Authors:
Pozhydaieva, N., Billau, F. A., Wolfram-Schauerte, M., Ramírez Rojas, A. A., Paczia, N., Schindler, D., Höfer, K.

Published on:
bioRxiv 2024.01.28.577628

Author contributions:
K.H. and Na.P. designed the study and wrote the initial draft. F.A.B, Na.P. cloned, expressed, analyzed the NgTET system, and prepared samples for LC-MS. Na.P. established a mutagenesis pipeline, evaluated experimental results. Ni.P. and Na.P. established an LC-MS pipeline to analyze phage DNA composition. A.A.R.R. and D.S. performed library preparation and ONT-sequencing. M.W.S. established the pipeline for the evaluation of the sequencing results. All authors reviewed and edited the manuscript.

5. Temporal epigenome modulation enables efficient bacteriophage engineering and functional analysis of phage DNA modifications

5.1. Abstract

Lytic bacteriophages hold substantial promise in medical and biotechnological applications. CRISPR-Cas systems offer a way to explore these mechanisms via site-specific phage mutagenesis. However, phages can resist Cas-mediated cleavage through extensive DNA modifications like cytosine glycosylation, hindering mutagenesis efficiency. Our study utilizes the eukaryotic enzyme NgTET to temporarily reduce phage DNA modifications, facilitating Cas nuclease cleavage and enhancing mutagenesis efficiency. This approach enables precise DNA targeting and seamless point mutation integration, exemplified by deactivating specific ADP-ribosyltransferases crucial for phage infection. Furthermore, by temporally removing DNA modifications, we elucidated the effects of these modifications on T4 phage infections without necessitating gene deletions.

Our results present a strategy enabling the investigation of phage epigenome functions and streamlining the engineering of phages with cytosine DNA modifications. The described temporal modulation of the phage epigenome is valuable for synthetic biology and fundamental research to comprehend phage infection mechanisms through the generation of mutants.

5.2. Main

Bacteriophages (phages) are viruses that specifically infect prokaryotic hosts. The high potential of the application of lytic phages in both medical and industrial settings has boosted phage research in recent years[1-4]. However, a comprehensive understanding of the molecular mechanisms that underlie efficient phage infections, which is essential for their targeted utilization, remains significantly underexplored. This knowledge gap is evident even in extensively studied phages like the *Escherichia coli* bacteriophage T4 (phage T4). Approximately half of the 273 encoded proteins in phage T4 are associated with known functions, while the roles of the other proteins remain unclear[5].

To explore the biological functions of phage-derived proteins, gene deletion has been the primary method historically employed, providing critical insights into gene essentiality for the phage infection process[6,7]. However, mutagenesis targeting only the catalytic residues rather than deleting the entire protein offers a possibility to elucidate the enzyme's role within a specific molecular context. The generation of catalytically inactive proteins preserves the other potential functions of the studied protein, such as involvement in protein-protein or protein-nucleic acid interactions. Particularly in phages with complex genomes, like phage T4, which has numerous

overlapping genes and gene splicing arrangements[7], targeted mutagenesis is crucial to avoid unintended effects on other genes' expression.

Prophages, whose genomes are integrated into the host chromosome and passively replicate alongside the host genome, can be mutated *in vivo* using the same genetic tools as for the mutagenesis of their host. However, this does not apply to lytic phages, as their genetic material exists separately from the host genome within bacterial cells and undergoes rapid replication during a limited time of infection. Therefore, targeted mutagenesis of lytic phages is highly challenging, as evidenced by the diverse strategies developed in recent years to approach it (reviewed in Mahler *et al.* 2023 [8]). Many developed approaches rely on recombination between phage DNA and a donor sequence, e.g., homologous recombination. This approach allows for gene replacements, deletions, or insertions. However, the efficiency of this method is low (>0.05%), resulting in the requirement of extensive screening to identify phage mutants[9,10]. To streamline the screening, incorporating reporter genes alongside the mutation to simplify the detection of the mutant phages has been employed[10,11]. Yet, the insertion of reporter genes may affect the complex gene expression of the phage, like the implications of phage gene deletions as described above. Additionally, inserting an extra gene may affect the packaging of phage DNA into the capsid, which has a fixed size[10].

Another approach, building on homologous recombination principles, is the bacteriophage recombineering of electroporated DNA (BRED) method, which offers a relatively high mutagenesis rate (10-15%)[12]. However, its effectiveness depends on the successful co-transformation of both phage DNA and donor DNA into the same cell, which might be limited by the host's transformation efficiency or the phage genome size[13].

The discovery of the antiphage clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein (Cas) has revolutionized the field of genome editing. The successful application of targeted CRISPR-Cas-based mutagenesis across various organisms[14], has also raised the interest in its application for phage engineering. In this context, CRISPR-Cas can be employed to target a specific position within the phage genome during infection, generating a DNA double-strand break. This break can be further repaired via homologous recombination with donor DNA (DNA carrying the desired mutation) present within the infected cell. However, the application of the CRISPR-Cas-based phage mutagenesis on the model phages, such as phage T4, has revealed an overall low mutagenesis efficiency (0-3%) and being strongly dependent on the spacer selected for the mutagenesis[15-17]. This strongly impedes the specific targeting of the phage genome via CRISPR-Cas and requires pre-screening for

efficient spacer before the mutagenesis, strongly reducing the applicability of the approach for efficient phage mutagenesis.

The reason for the strong spacer dependence of Cas targeting efficiency is attributed to the extensive DNA modifications - present on phage DNA - that protect it from cleavage. The DNA modifications among the phages are widely distributed and are also present in the phage T4[18,19]. The enzyme deoxycytidylate 5-hydroxymethyltransferase (*gene 42*) originating from T4 phage catalyzes the conversion of 2´-deoxycytidine 5´-monophosphate (dCMP) into 5-hydroxymethyl-2´-deoxycytidine 5´-monophosphate, which is subsequently converted to into 5-hydroxymethyl-2´-deoxycytidine 5'-triphosphate and incorporated into T4 DNA during replication (further referred to as 5hmdC). Next, the 5hmdC is glycosylated by α- or β-glycosyltransferases (α-/β-gt) to 5-α-/β-glycosyl hydroxymethyl-2´-deoxycytidine (5ghmdC) within T4 DNA (Fig. 1a)[20,21]. The T4 phage epigenome plays a crucial role in phage fitness, exemplified by the essential *gene 42*. Notably, amber mutants for the gene have been reported. However, in this scenario, phage infections must be carried out using *E. coli* strains that harbor additional plasmids and thus do not represent the wild-type *E. coli*. Furthermore, *α-/β-gt* serve as auxiliary genes[7,17,22], as the encoded proteins are pivotal in protecting phage DNA through glycosylation from host defense systems and host nucleases, including Cas nucleases[15,23]. Thus, DNA modifications such as 5hmdC and 5ghmdC prevent the effective use of CRISPR-Cas for targeted phage mutagenesis and counterselection.

In this study, we apply a eukaryotic ten-eleven translocation (TET) methylcytosine dioxygenase[24] to temporally reduce the abundance of 5ghmdC within the phage T4 genome to enable the specific and efficient targeting of phage DNA with CRISPR-Cas. This results in the facilitated introduction of point mutations into the phage genome in a scarless manner, exemplarily shown for two specific T4 phage ADP-ribosyltransferases crucial for phage infection. The increased targeting efficiency of phage DNA by Cas nucleases allows for sequence-specific spacer selection. Simultaneously, the improved targeting efficiency of phage DNA by Cas nucleases can be exploited for efficient counterselection, streamlining the enrichment process for the generated phage mutant. We utilized a high-throughput screening approach based on next-generation sequencing (NGS) to identify the mutants, facilitating the identification and validation of phage T4 mutants. This method is compatible with automation, simplifying the phage genome engineering process and eliminating the need to introduce reporter genes into the phage genome. The scarless nature of this approach allows for the precise study of the impact of introduced mutations on phage infection without the side effects associated with gene deletion or reporter gene insertion. Overall, our findings propose an efficient strategy for introducing point mutations into genomes of lytic phages that possess cytosine-based DNA modifications. These advantages

collectively position this technique as a valuable tool in synthetic biology and biotechnology for creating "designer phages" as well as in fundamental research to explore the mechanisms underlying phage infections. Moreover, this technique offers the opportunity to study the influence of phage epigenome on phage infection dynamics without necessitating the deletion of the essential phage DNA modifying genes.

5.3. Results

Eukaryotic NgTET modulates the abundance of T4 DNA modifications

Phage DNA modifications, such as 5ghmdC, impede the recognition and targeting of phage DNA by host nucleases and CRISPR-Cas systems[25]. The absence of these modifications would enhance phage DNA accessibility to host nucleases, resulting in phage genome degradation and the prevention of successful propagation[26,27]. No host- or phage-originating enzymes are currently known to specifically remove 5ghmdC modifications from the phage T4 genome.

In contrast to the 5ghmdC modification, enzymes that act on its direct precursor, 5hmdC, are well-known within the eukaryotic realm and play crucial roles in epigenetic regulations. Eukaryotic TET dioxygenases *are involved in active DNA demethylation* via iterative oxidation of 5-methyl-2´-deoxycytydine (5mdC) to 5hmdC, 5-formyl-2´-deoxycytidine (5fdC), and 5-carboxy-2´-deoxycytidine (5cadC) (Fig. 1b)[28,29]. In the next step, 5cadC undergoes non-enzymatic decarboxylation, leading to the formation of unmodified 2´-deoxycytidine (dC)[30,31]. Therefore, we hypothesized that eukaryotic TET enzymes could potentially be harnessed to convert 5hmdC modifications in phage DNA into 5fdC and 5cadC (Fig. 1c). Leveraging the demethylation activity of TET enzymes on phage DNA presents an opportunity to diminish 5hmdC levels and compete effectively with T4 phage α/β-gt enzymes for the substrate 5hmdC during T4 phage infection. This approach has the potential to decrease the prevalence of 5ghmdC in phage DNA, thereby enhancing the efficacy of CRISPR-Cas-based editing, as depicted in Figure 1c.

To test this hypothesis, we selected the well-characterized TET enzyme from the single-cell protist *Naegleria gruberi* (NgTET), which has already been expressed in *E. coli* in a catalytically active form[32,33]. First, we evaluated the influence of heterologous NgTET expression on *E. coli* growth to confirm that the dioxygenase expression does not impede host growth and, consequently, does not affect the host's susceptibility to phage infection (see Extended Data Fig. 1a-b). Notably, the expression of NgTET in *E. coli* did not significantly impact the bacterial growth rate or bacterial lysis by T4 phage. Moreover, phages derived from the lysis of the NgTET-expressing strain were still found to be infectious and cause complete lysis when infecting wild-type *E. coli* (Extended Data Fig. 1b).

Next, we investigated if NgTET is active on phage T4 DNA. For this, we evaluated if the expression of NgTET in *E. coli* during phage T4 infection alters the presence of 2´-deoxycytidine modifications such as 5hmdC and 5ghmdC in the phage DNA. To do this, we infected an *E. coli* strain expressing NgTET with T4 wild-type (T4 WT) phages. We conducted a comprehensive LC-MS-based analysis of the DNA composition of the resulting T4 phages, which we refer to as "NgTET-treated T4" phages. This analysis enabled us to compare the DNA composition of the phage progeny obtained from the NgTET-expressing *E. coli* strain, NgTET-treated T4, with that of the phage progeny resulting from the infection of *E. coli* wild-type cells, T4 WT (see Figs. 1d-g and Extended Data Fig. 2). In agreement with previous studies that characterize the T4 WT DNA composition[21], we observed the complete absence of unmodified 2´-deoxycytidine in T4 WT DNA. According to our LC-MS analysis, >99% (expressed as relative abundance (r.a.)) of all 2´-deoxycytidines were present as 5ghmdC and <1% - as 5hmdC (Fig. 1d-e).

The analysis of NgTET-treated T4 DNA revealed a reduction of 5 ghmdC from 99% r.a. (T4 WT) to 55% r.a. (NgTET-treated T4) (Fig. 1f). In addition, 5hmdC (10.5% r.a.), 5fdC (2.3% r.a.), 5cadC (0.9% r.a.), and unmodified 2´-deoxycytidines, dC (34.4% r.a.) were detected in the NgTET-treated T4 DNA (Fig. 1b). 5fdC and 5cadC are known oxidation products of 5hmdC[32,33] (Fig. 1b). The significant reduction in 5ghmdC levels (Extended Data Fig. 2e), coupled with the presence of unmodified dC, which is the result of the decarboxylation of 5cadC[30,31], confirms the catalytic activity of NgTET on T4 DNA.

To unambiguously link the altered modifications abundance in T4 DNA with NgTET enzymatic activity, we generated a catalytically inactive NgTET D234A mutant as a negative control[34]. LC-MS analysis of DNA isolated from the phage progeny released from *E. coli* strain overexpressing inactive NgTET D234A (referred to as "NgTET D234A-treated T4") showed a DNA composition similar to that of T4 WT (Fig. 1g) (>1% r.a. for 5hmdC and 99% r.a. for 5ghmdC in NgTET D234A-treated T4 DNA). These findings make us confident that the observed reduction of 5ghmdC is due to the catalytic activity of NgTET, expressed in *E. coli* during T4 phage infection, efficiently oxidizing 5hmdC within phage T4 DNA.

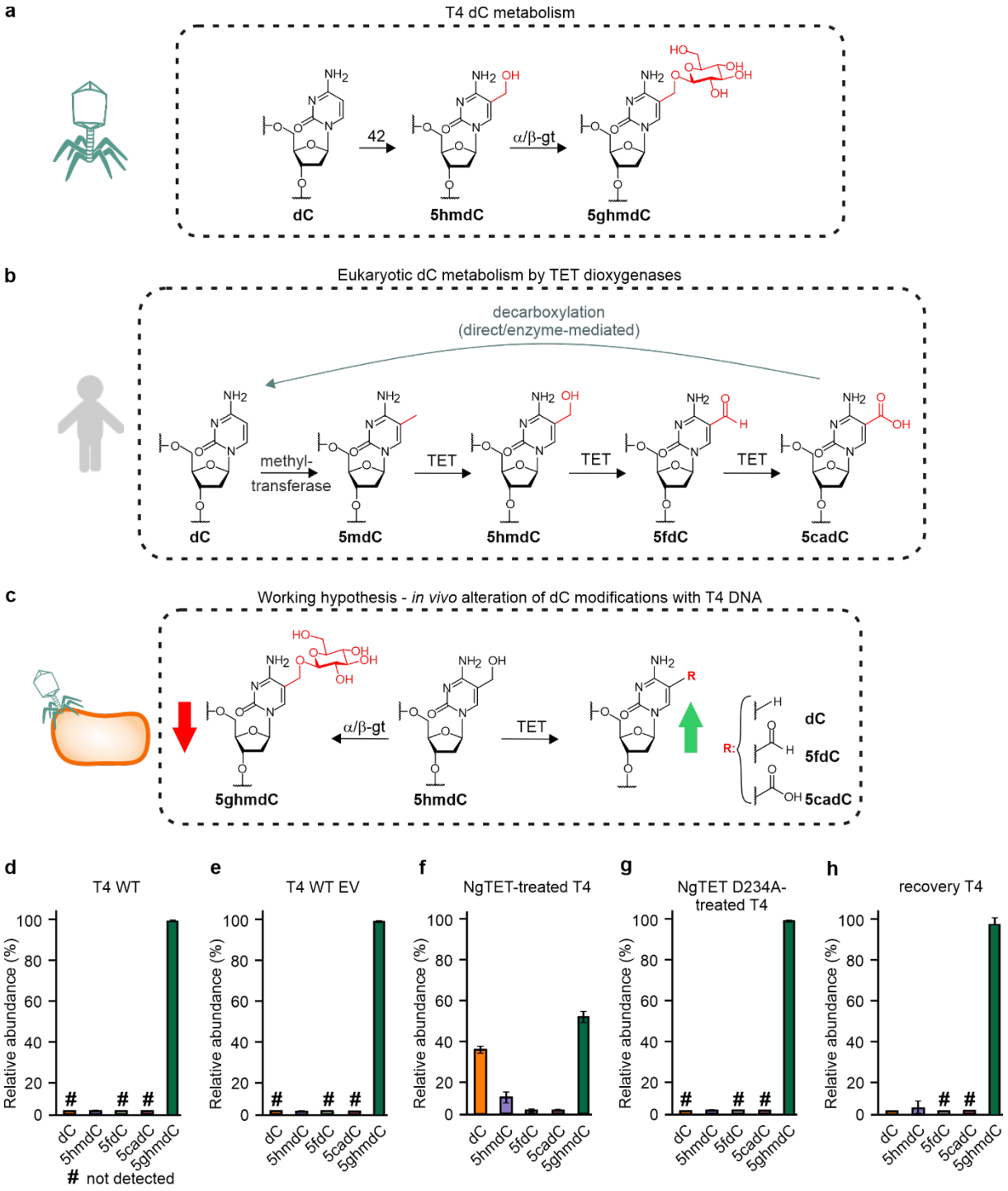


Fig. 19: T4 phage DNA is extensively modified. a, phage T4 DNA is extensively modified by deoxycytidylate 5-hydroxymethyltransferase (42) (oxidation of dC to 5hmdC), and α/β-gt (glycosylation of 5hmdC to 5ghmdC). **b,** TET dioxygenase plays a crucial role in eukaryotic epigenetic regulation by demethylating 5mdC through its iterative oxidation. **c,** 5hmdC present in phageT4 is one of the natural substrates of TET. Therefore, the glycosylation of 5hmdC to 5ghmdC by α-gt/β-gt is expected to be downregulated in the presence of TET due to substrate competition between TET and α-gt/β-gt. **d-h,**

Relative abundance (%) of 2´-deoxycytidine metabolites determined via LC-MS analysis in different T4 strains: T4 WT (**d**), T4 WT propagated in the presence of empty vector (EV) (T4 WT EV) (**e**), NgTET-treated T4 (**f**), NgTET D234A-treated T4 (**g**), and recovery T4 (**h**). The presence of dC traces in recovery T4 DNA (**h**) may be attributed to residual NgTET-treated T4 phages that did not infect *E. coli* and therefore were not recovered. Hashtag highlights the nucleosides not detected in a sample. *n* = 3 biological replicates. The significance of specific cytosine modification changes is shown in Extended Data Fig. 2e.

Modulation of phage DNA modifications by NgTET alters bacterial cell lysis

As emphasized earlier, DNA modifications significantly contribute to phage fitness, and their absence can potentially alter the phage infection phenotype. Based on this, we aimed to investigate whether the phenotype of NgTET-treated T4 phages might be altered due to reduced abundance of 5ghmdC modification. Specifically, we hypothesized the host cell lysis being affected for NgTET-treated T4 compared to the lysis by T4 WT phages, as less modified phage DNA would be more susceptible to degradation by host nucleases (Fig. 2a). Consequently, the replication and gene expression machinery of the phage will be impeded, affecting the lysis of the bacterial host.

To answer this question, we infected wild-type *E. coli* with either T4 WT or NgTET-treated T4 (Fig. 2b) and monitored bacterial cell lysis. In line with our hypothesis, NgTET-treated T4 phages exhibited a notably slower lysis rate than the T4 WT phages, resulting in an approximately 15-minute delay in the onset of lysis. Nevertheless, both T4 WT and NgTET-treated T4 phages ultimately induced complete lysis of the bacterial culture. Furthermore, the phages released during the lysis with NgTET-treated T4 were confirmed to be infectious, as evidenced by their ability to infect and lyse wild-type *E. coli* (Extended Data Fig. 1b).

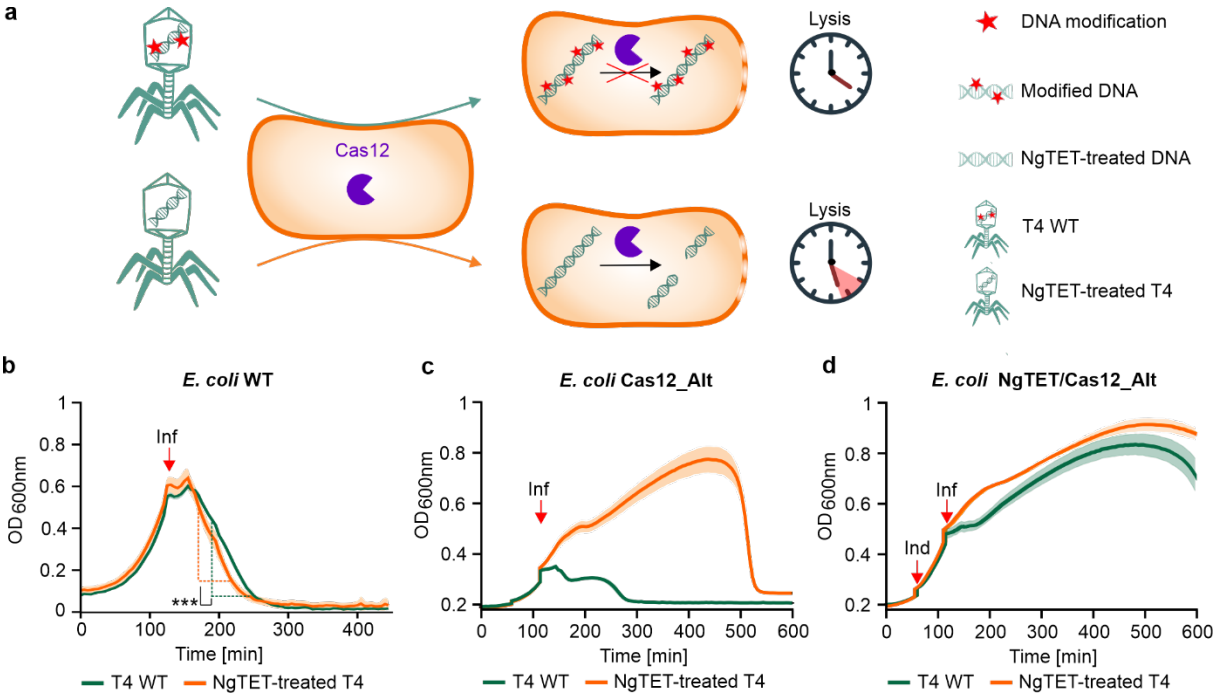


Fig. 2: Impact of T4 DNA modifications on T4 phage lysis behaviour in the presence and absence of CRISPR-Cas12. a, T4 DNA modifications protect phage DNA against Cas12 nuclease targeting, ensuring efficient phage propagation cell lysis. Reduced DNA modifications increase Cas12 nuclease susceptibility, leading to impeded and therewith delayed lysis of bacterial culture. **b**, Lysis kinetics of *E. coli* infected by T4 WT and NgTET-treated T4. Red arrow highlights the time point of the phage addition (Inf) (two-sided Student's t-test, $P = 0.0096$ at $P_{signif} < 0.05$). $n = 3$ biological replicates. **c**, Lysis kinetics of *E. coli* Cas12_Alt by T4 WT and NgTET-treated T4. Red arrows highlight the time point of NgTET expression induction (Ind) and the addition of the phage (Inf). $n = 3$ biological replicates. **d**, Lysis kinetics of *E. coli* NgTET/Cas12_Alt by T4 WT and NgTET-treated T4. Red arrows highlight the time point of NgTET expression induction (Ind) and the addition of the phage (Inf). $n = 3$ biological replicates.

Temporal removal of DNA modifications enhances CRISPR-Cas12 targeting of phage DNA

The previous experiment showed that treating phage T4 with NgTET led to diminished lysis of *E. coli*, showing the importance of phage DNA modifications for its fitness. Building on this, we aimed to investigate whether CRISPR-Cas targeting of the phage DNA would also be enhanced due to the NgTET treatment. We hypothesized that the reduction in DNA modifications could increase the susceptibility of phage DNA to Cas nuclease targeting, as demonstrated by the more effective targeting of non-glycosylated phage T4 DNA by CRISP-Cas12[35]. Therefore, NgTET treatment of T4 DNA would potentially result in Cas-mediated DNA double-strand breaks. Such a break would impede phage replication, consequently causing delayed or even absent lysis by

NgTET-treated T4 compared to T4 WT. Given the extensive modifications in T4 WT DNA, we expected its lysis not to be significantly delayed by the expression of the CRISPR-Cas system.

To follow this hypothesis, we conducted a lysis experiment in which we infected an *E. coli* strain, heterologously expressing the CRISPR-Cas12, with either T4 WT or NgTET-treated T4 phages (Fig. 2c). For a proof of principle study, we designed a spacer targeting the T4 gene *alt* which encodes the T4 Alt ADP-ribosyltransferase.

First, we verified that overexpression of CRISPR-Cas12 without a spacer does not negatively affect phage lysis and bacterial growth (Extended Data Fig. 3). Next, we conducted the infection of the *E. coli* strain expressing the CRISPR-Cas12_Alt (spacer targeting *alt* gene) with T4 WT. As expected, we did not observe any negative impact of CRISPR-Cas12_Alt expression on phage lysis. This observation demonstrates that DNA modifications protect the genome of phage T4 from nucleolytic cleavage by Cas12, enabling efficient phage lysis.

Next, we analyzed the lysis by NgTET-treated T4 in the same experimental settings. While infecting *E. coli* expressing CRISPR-Cas12_Alt, a delay of approximately 250 min in lysis compared to infection with T4 WT was observed. This delay supports our hypothesis that the reduced abundance of 5ghmdC in NgTET-treated T4 enables Cas nuclease targeting of phage DNA. Consequently, phage propagation is impeded, which is reflected in the delayed onset of lysis.

To confirm that this delay is attributed to the targeting of T4 DNA by CRISPR-Cas12_Alt, we infected *E. coli* expressing the CRISPR-Cas12 without any spacer, thus lacking any target, with either T4 WT or NgTET-treated T4. Since there was no spacer in CRISPR-Cas12, no targeting of phage DNA and no additional impact on phage lysis was expected. Consistent with our expectations, we observed the same lysis behaviour for NgTET-treated T4 infection of the CRISPR-Cas12 expressing *E. coli* strain as for the infection of wild-type *E. coli* (Extended Data Fig. 1d). This confirms that the observed delay in Fig. 2c is a direct consequence of CRISPR-Cas12_Alt efficiently targeting the *alt* gene in DNA of NgTET-treated T4 phage.

Continuous removal of T4 DNA modifications prevents bacterial lysis

Despite the observed lysis delay in *E. coli* expressing CRISPR-Cas12_Alt when infected by NgTET-treated T4 phages in the previous experiment, the eventual lysis of the culture indicates that the phages ultimately succeeded in propagating efficiently (Fig. 2c). We hypothesized, that NgTET-treated T4 phages regain their lysis capacity during infection. We speculated that the progeny of NgTET-treated T4 phages, which evaded CRISPR-Cas12_Alt targeting, restored wild-

type-like DNA modification levels, as NgTET dioxygenase was not heterologously expressed upon infection.

To prove our hypothesis, we conducted a further lysis experiment. This time, we infected *E. coli* cells that were simultaneously expressing both CRISPR-Cas12_Alt and NgTET dioxygenase. The constant overexpression of the NgTET was aimed to maintain a reduction in DNA modifications in the phage progeny generated during the experiment. Notably, upon expression of NgTET dioxygenase, we did not observe the onset of lysis for up to 600 min post-infection in both T4 WT and NgTET-treated T4 infections (Fig. 2d), which strongly supports our hypothesis.

In conclusion, these findings underscore the critical role of T4 DNA modifications in protecting the DNA from host anti-phage defense systems, particularly those operating at DNA level, such as CRISPR-Cas12. While the reduction in 5ghmdC levels in NgTET-treated T4 seems to enhance phage DNA susceptibility to CRISPR-Cas12, it simultaneously leads to decreased phage fitness and a substantial alteration of the phage phenotype (Fig. 2d).

The removal of DNA modifications by NgTET is reversible

To preserve the integrity of phage phenotypes in genetic studies, particularly when utilizing phage strains with decreased DNA modifications for mutagenesis, it is crucial to restore wild-type (WT) DNA modification levels. In this study, we aimed to temporally alter T4 DNA modification levels during mutagenesis and restore them to T4 WT levels after successfully introducing the mutation of interest. In such a way, only the impact of the introduced mutations can be studied. The reversible alteration of the T4 epigenome represents a notable advantage compared to employing Δα-/β-gt T4 or respective amber T4 mutants, where permanent alterations of phage DNA modifications can affect the phage phenotype[22].

We hypothesized that treating T4 DNA with NgTET could be used to decrease 5ghmdC levels temporarily. Hence, our investigation aimed to determine if the reduction in the 5ghmdC fraction within the NgTET-treated T4 DNA, is reversible to wild-type levels in the subsequent phage generation, termed as "recovery T4", produced in the absence of NgTET dioxygenase.

To confirm that the effects of phage DNA treatment with NgTET are indeed reversible and do not impact DNA modifications in the subsequent progeny, we infected *E. coli* WT with NgTET-treated T4. The DNA composition of recovery T4 was analyzed by LC-MS (Fig. 1h). Our data show that already for the first generation of recovery T4 phages, the modification levels were comparable to the modification levels in T4 WT DNA (3% r.a. for 5hmdC, 97% r.a. for 5ghmdC, and traces of >0.3% dC). The absence of the NgTET-mediated oxidation products, 5fdC, and 5cadC, confirms

that the altered abundance of DNA modifications in NgTET-treated T4 is transient and can be restored in the next phage generation.

Next, the recovery phage was also analyzed regarding its infectivity of the host, demonstrating same lysis rate as T4 WT phage (Extended Data Fig. 1b).

NgTET treatment of T4 DNA allows efficient Cas-mediated phage DNA targeting *in vivo*

To assess the impact of NgTET treatment of phages on Cas-mediated genome engineering efficiency, we selected two T4 phage-encoded genes, *modA,* and *alt*, as mutation targets. Alt and ModA belong to the enzyme class of ADP-ribosyltransferases. During infection, they have been described to play a crucial role in the takeover of control over *E. coli* cells by T4 phage, although the exact mechanisms behind this process are still not fully understood[36-38].

We aimed to generate T4 phage variants carrying catalytically inactive versions of Alt or ModA by exchanging single amino acids essential for their ADP-ribosyltransferase activity[39]. Based on the previous studies, we generated Alt E577A and ModA E165A mutants and confirmed them to be catalytically inactive *in vivo* (Fig. 3a, Extended Data Fig. 4).

To generate T4 phages carrying Alt E577A and ModA E165A mutations, we employed the heterologously expressed CRISPR-Cas9 and Cas12 systems, which were previously evaluated for the application to engineer phage T4[15,35]. For both CRISPR-Cas systems, the targeting of WT T4 DNA was reported to be strongly hindered by phage DNA modifications, leading to overall low mutagenesis efficiency[15,35].

To introduce the site-specific point mutations into the phage DNA, we designed the spacers to precisely target the intended mutation insertion sites within *alt* and *modA*. We conducted targeting efficiency (TE) tests to evaluate Cas nucleases' efficiency in targeting phage DNA. Therefore, we infected an *E. coli* strain that heterologously expressed one of the analyzed CRISPR-Cas systems with a spacer designed to target the *alt* or *modA* gene with a specific number of T4 phages. To determine the number of phages capable of efficiently infecting the host in the presence of CRISPR-Cas, we conducted a plaque assay (Fig. 3b). The number of plaques formed on the plate represented the phages that successfully evaded CRISPR-Cas targeting and propagated effectively. Based on this number, we calculated the reduction in phage numbers due to CRISPR-Cas targeting, which corresponds to the TE for the specific CRISPR-Cas system and the spacer. First, we determined for each construct TE after T4 infection of wild-type *E. coli* expressing respective CRISPR-Cas systems (Fig. 3b). In this case, the 5ghmdC modifications were still present on >99% of T4 cytosines (Fig. 1d), potentially hindering the activity of CRISPR-Cas systems. In the absence of NgTET, most constructs showed a TE of 0 to 0.1, except for

Cas9_ModA, which showed a TE of ~0.5. These results confirm the generally low targeting efficiency of T4 WT DNA by CRISPR-Cas9 and Cas12, which is additionally influenced by the specific site targeted within the phage genome, as also reported in earlier studies[15,35].

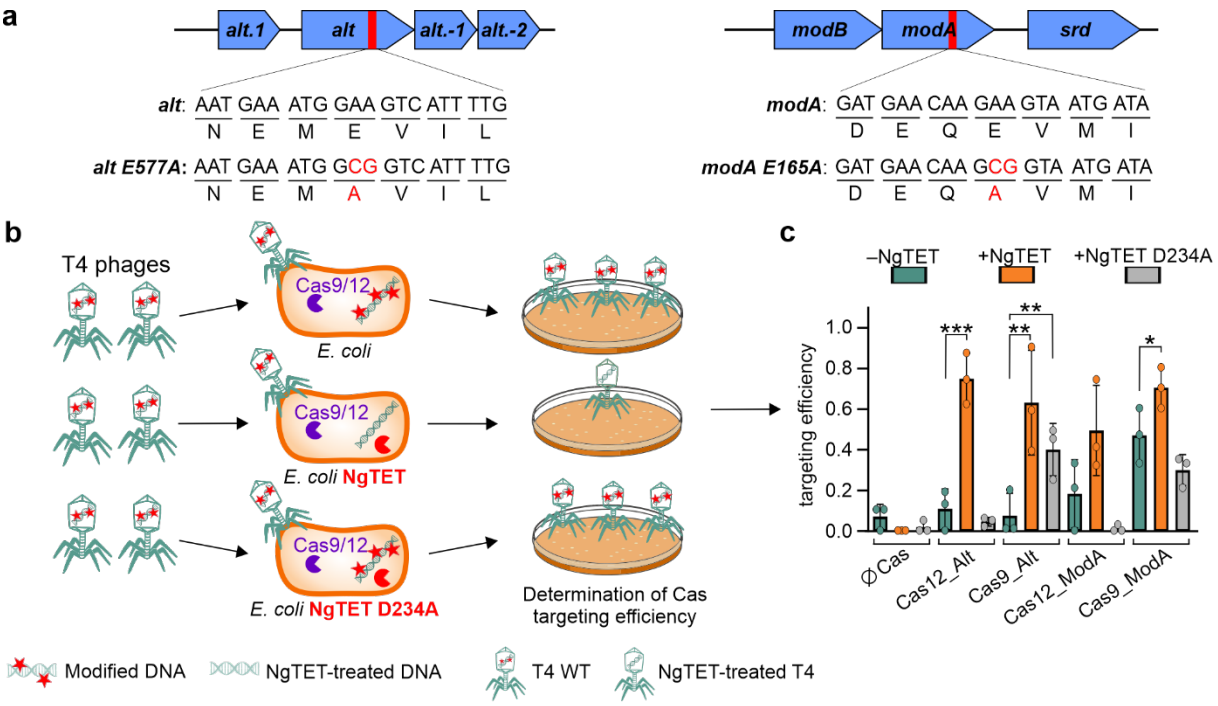


Fig. 3: Effect of NgTET overexpression on CRISPR-Cas targeting efficiency *in vivo*. a, Target mutations in *alt* and *modA* genes that lead to the abolishment of ADP-ribosylation activity. **b,** Design of TE experiment. **c,** Evaluation of Cas-mediated T4 DNA TE *in vivo* in the presence or absence of NgTET/NgTET D234A (two-sided Student's t-test, * - P_{signif} < 0.05, ** - P_{signif} < 0.025, *** - P_{signif} < 0.0125). *n* = 3 biological replicates.

Next, we investigated whether the presence of NgTET enzyme, leading to a reduction of DNA modifications on T4 DNA, can increase the Cas targeting efficiency of T4 DNA *in vivo*. We infected *E. coli* recombinantly expressing the respective CRISPR-Cas system and NgTET simultaneously. The expression of NgTET increased the TE for all analyzed targets (Fig. 3c). For example, for CRISPR-Cas12 targeting *alt*, we observed a significant 7-fold increase in TE from 0.11 to 0.74. As for CRISPR-Cas9 targeting *modA*, which already had a high TE before NgTET treatment (TE of 0.47), the TE increased by 1.5-fold due to NgTET treatment of DNA. To confirm the connection between the increased Cas cleavage activity on T4 DNA and altered DNA modification levels associated with NgTET activity, we performed a parallel experiment using the inactive mutant

NgTET D234A for TE determination[32]. As expected, the expression of the inactive NgTET mutant did not improve targeting efficiency in all analyzed settings (Fig. 3c).

In summary, the NgTET dioxygenase activity on T4 DNA directly enhances its TE by CRISPR-Cas9 and Cas12 and strongly diminishes the TE's reliance on the specific target site within the phage genome. This affords flexibility in spacer design and the ability to target the desired genome region, which is a critical factor for mutagenesis and the introduction of SNPs into the modified genome.

Temporarily removal of T4 DNA modifications enables efficient phage T4 mutagenesis

The treatment of T4 DNA with NgTET significantly enhances its targeting efficiency by CRISPR-Cas9 and -Cas12 systems (Fig. 3c). To validate that NgTET treatment indeed enhances phage mutagenesis through improved CRISPR-Cas targeting, we conducted mutagenesis experiments on both the *alt* and *modA* genes, using T4 WT or NgTET-treated T4. To prove the hypothesis, we employed four distinct mutagenesis setups: (1) using CRISPR-Cas9 alone, (2) the CRISPR-Cas9 combined with NgTET dioxygenase, (3) CRISPR-Cas12 alone, and (4) CRISPR-Cas12 combined with NgTET dioxygenase.

In the settings where NgTET dioxygenase was expressed during the mutagenesis (settings 2 and 4), the phages were pretreated with NgTET before infecting the mutagenesis strain (Fig. 4). This pretreatment aimed to reduce the abundance of 5ghmdC and enhance the targeting of phage DNA by CRISPR-Cas, as demonstrated for TE experiments previously (Fig. 3c). Nevertheless since the TEs for both CRISPR-Cas systems and genes exhibited variability ranging between 0.5 and 0.75 (Fig. 3c), we anticipated that a fraction of the phages used for mutagenesis would evade CRISPR-Cas targeting. Consequently, mutagenesis was expected to yield a mixed population of wild-type and mutant T4 phages.

As the NgTET treatment increases DNA susceptibility to specific CRISPR-Cas targeting, it suggests a potential strategy for enhanced counterselection with the same CRISPR-Cas system used for the mutagenesis. Accordingly, we conducted a counterselection round in the absence of donor DNA.

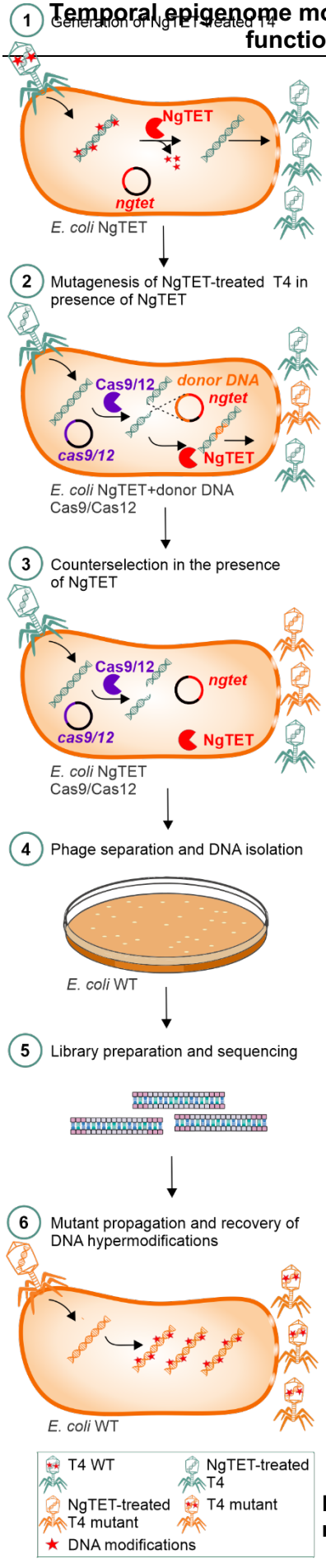


Fig. 4: Established workflow for T4 phage mutagenesis and mutants screening.

Following mutagenesis, the resulting phage population was isolated by plating with *E. coli* B strain (wild-type) and then screened for mutations, enabling the identification of specific genetic changes.

As mentioned above, a mixed population of the phages (wild-type and mutant phages) was expected. Although reporter genes could be introduced along the mutations to streamline mutant detection, they can potentially impact phage physiology, as outlined in the introduction. To avoid it, we adopted a multiplexed ONT-based approach[40] for high-throughput screening of phage mutants. The mutation region was amplified from the phage with plaque-specific barcodes, which enables simultaneous screening of thousands of phages. Combining the mutagenesis with ONT screening, we determined that CRISPR-Cas9 and Cas12-based mutagenesis in the absence of NgTET did not result in the introduction of point mutations for all tested CRISPR-Cas systems, spacers, and targeted genes (Table 1). However, expression of NgTET in the infected *E. coli* host strain boosted the mutagenesis efficiency and resulted in 2.6-5.6% positive hits for targeted *modA* and *alt* genes (Extended Data Fig. 5-6). Here, the utilization of the CRISPR-Cas12 combined with NgTET resulted in successful mutagenesis of both analyzed genes, while the application of the CRISPR-Cas9 in combination with NgTET yielded the desired mutant only for *modA* gene.

In conclusion, coexpressing NgTET dioxygenase in the mutagenesis strain to reduce the T4 DNA modification levels and enhance DNA targeting by Cas nucleases (Fig. 4), efficiently facilitated the introduction point mutations to phage T4 genome. These mutants could not be generated in the experimental setup without NgTET. Therefore, NgTET treatment of phage DNA eliminates the need for screening for the most efficient spacer, enabling the targeting of any desired position within the phage genome and the introduction of point mutations.

Table 1: Summary of phage genome editing efficiency in the presence and absence of NgTET (in green: conditions resulting in positive clones, n=1)

Edit name	Edited gene	Mutagenesis strain (*E. coli* BL21(DE3)...)	Plaques screened	Mutants identified	Mutation success rate (%)
alt E577A	*alt*	*alt* E577A, Cas9_Alt	47	0	0
		NgTET_*alt* E577A, Cas9_Alt	45	0	0
		alt E577A, Cas12_Alt	38	0	0
		NgTET_*alt* E577A, Cas12_Alt	44	1	2.6
modA E165A	*modA*	*modA* E165A, Cas9_ModA	36	0	0
		NgTET_*modA* E165A, Cas9_ModA	36	1	2.8
		modA E165A, Cas12_ModA	35	0	0
		NgTET_*modA* E165A, Cas12_ModA	36	2	5.6

Epigenetic cytosine modifications are common in phage genomes

To evaluate the applicability of the NgTET system for phage engineering beyond phage T4, we examined the occurrence of homologs of T4 DNA-modifying enzymes in other phages. We identified a total of 494 homologs for T4 gene 42, 131 for α-glycosyl-transferase, and 40 for β-glycosyl-transferase (Fig. 5, Supplementary Table 1).

The prevalent identification of the homologous of gene 42 responsible for the initial hydroxymethylation of dC to 5hmdC suggests the presence of related modifications in several other phages, potentially involving glycosylations, arabinosylations, and other sugar modifications[18,41,42]. Notably, homologous enzymes were found in phages infecting a broad spectrum of bacterial genera. A considerable number of these bacterial hosts are pathogenic and belong to World Health Organization (WHO) priority strains[43]. Therefore, given the widespread presence of homologous DNA modifiers among phages and the potential of NgTET treatment

demonstrated in this study on phage T4 DNA — both for investigating the role of DNA modifications and for efficient phage mutagenesis — the application of NgTET could be expanded to other phages. This extension would hold significant promise for both fundamental research and phage engineering.

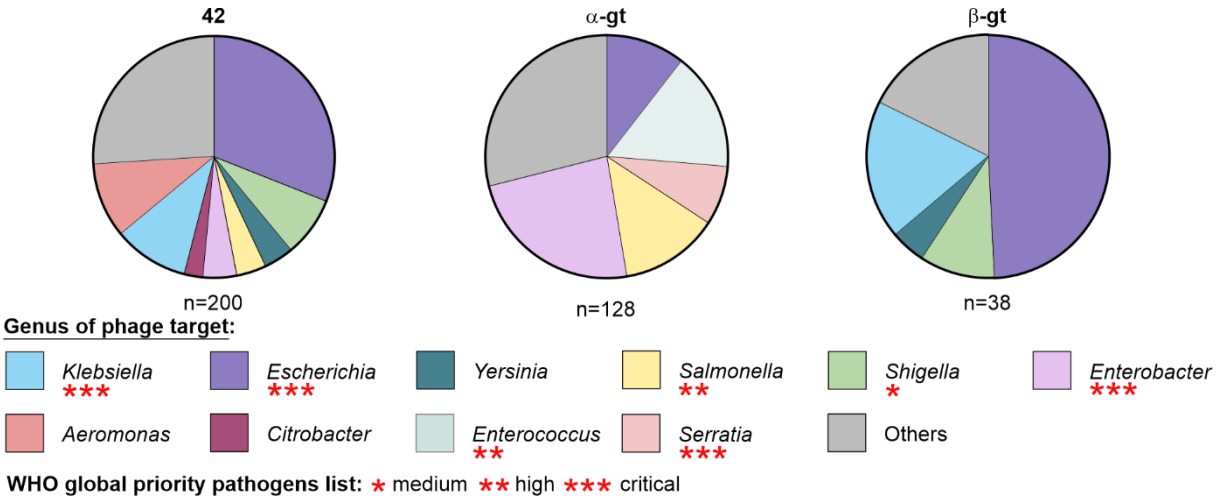


Fig. 5: Distribution of cytosine modifying enzymes homologous to T4-originating enzymes 42, α-gt and β-gt among phages based on the genus of infected bacteria (BLAST score >80). Red asterisks highlight the global priority pathogens classification of the bacterial genus according to the WHO.

5.4. Discussion

CRISPR-Cas has emerged as an extremely valuable tool for genome engineering. The effectiveness of CRISPR-Cas systems in phage engineering has been limited by extensive DNA modifications up to this point. In particular, specific Cas-mediated targeting and induction of DNA double-strand breaks, essential for the mutagenesis, is prevented by DNA modifications[15,25]. The low Cas targeting efficiency of phage DNA is further reflected in the low mutagenesis rate, necessitating the screening of a large phage population to identify the phage mutants. Our study circumvents previous limitations and presents a novel approach for precise, site-specific phage mutagenesis based on temporarily reducing phage DNA modifications. This enhances DNA accessibility for Cas nucleases, boosting the efficiency of phage DNA double-strand breaks – the first and essential step of CRISPR-Cas-based mutagenesis. Thus, we eliminate the need for pre-screening for the most efficient spacer, significantly facilitating the introduction of scarless mutations at selected positions within the phage genome.

Moreover, the phage mutagenesis efficiency of up to 6% reported in this study, combined with the ONT-based high-throughput screening method, simplifies the detection of the point mutations within the phage genome, thereby reducing the need to screen large phage populations. This

approach avoids additional genomic changes associated with reporter gene insertion and the potential drawbacks on phage DNA packaging[10], allowing for the efficient, scarless introduction of point mutations into the phage genome.

To demonstrate the feasibility of our approach, we successfully introduced point mutations to inactivate T4 ARTs Alt and ModA. ARTs have been shown to be important for efficient host hijacking, which is achieved by introducing post-translation protein modifications such as ADP-ribosylation and RNAylation[44]. In addition to its catalytic function, Alt has been reported as an important T4 phage structural protein[7]. Therefore, deleting the entire gene would not solely allow the study of ADP-ribosylation's role, but might also have a more pronounced impact on the phage infection. Thus, introducing a point mutation via our NgTET tool to inactivate the enzyme while preserving its role as a structural protein facilitates the functional characterization of the enzyme.

The modulation of T4 phage DNA modifications by NgTET treatment is possible due to the decoupled introduction of hydroxymethylation and glycosylation on T4 cytosines. The biosynthesis of modifications within phage T4 DNA occurs at both the nucleotide and polynucleotide chain levels. Initially, the hydroxymethylation is formed on dCMP, which is next converted to 5-hydroxymethylated 2'-deoxycytidine 5'-triphosphate and is integrated into phage DNA during replication. The glycosylation of so-formed 5hmdC occurs at the DNA level[18]. The fact, that glycosylation happens directly on the DNA, and not on the nucleotide levels, allows NgTET, which also acts on DNA level, to oxidize 5hmdC into 5fdC and 5cadC, thereby preventing the glycosylation. The propagation of the resulting phage progeny in the strain without NgTET, allows the glycosyltransferases to introduce the glycosylations back and thereby recover the wild-type-like modifications of phage T4 DNA. Such a recovery of DNA modifications in the subsequent phage generation highlights the major advantage of NgTET-coupled CRISPR-Cas mutagenesis over performing it with *α/β-gt* deletion or amber strains. In such deletion or amber strains, DNA glycosylations are permanently absent, making the phage consistently more susceptible to nucleases and thereby reducing its fitness.

Additionally, NgTET treatment results in a fraction of non-modified 2´-deoxycytidines (approximately 35% for dC). To our knowledge, this was not possible by amber mutations or deletion of gene 42 due to its essentiality for phage infection[7,22]. Therefore, the presence of non-modified dC allows for efficient phage mutagenesis and also holds promise for studying the biological roles of various phage modifications or improving phage genome sequencing. Specifically - for sequencing approaches - NgTET treatment could be applied to reduce modification-mediated errors[45], improving sequencing accuracy by decreasing the DNA modifications.

An important feature of the NgTET/CRISPR-Cas-coupled gene editing technique - presented in this study - is its potential transferability to other bacteriophages. Our homology search results show the cytosine-modifying enzymes being distributed among various phages (Fig. 5). Particularly, homologs of gene 42, responsible for the conversion of dC to 5hmdC, were detected in numerous phages. Gene 42 introduces a chemically active hydroxymethyl group to 2´-deoxycytidines, allowing further functionalization with various sugars and other functional groups. As a result, the modified DNA becomes impervious to degradation by DNA-binding nucleases, enabling the phages to effectively evade the host defense mechanisms. The capability of NgTET to act on the precursors 5mdC and 5hmdC, generated by homologous of gene 42, broadens the utility of our approach to engineer and to study phages carrying such DNA modifications. Notably, such phages are known to infect clinically and biotechnologically relevant bacterial genera, including *Klebsiella*, *Salmonella*, and *Serratia*. Therefore, NgTET-mediated mutagenesis has the potential to become a valuable tool for studying such phages by facilitating their mutagenesis, and taking us a step closer to the generation of "designer phages".

Lastly, it is worth considering the possibility of homologous enzymes to eukaryotic TET dioxygenases existing within bacteria. These enzymes may potentially be involved in counteracting phage infections through mechanisms analogous to those synthetically applied in this study. Reducing the extent of DNA modifications could enable nucleolytic cleavage of the invader's DNA, which would be a powerful anti-phage defense system. On the other hand, in the case of phages, homologs of TET dioxygenases have been observed46. Yet, their involvement is typically associated with the modification of 2´-deoxycytidines, involving oxidation from 5mdC to 5hmdC. In contrast to NgTET, no following oxidation to 5fdC, 5cadC and dC has been described to date.

Taken together, the field of bacteriophage epigenetics remains widely unexplored and, therefore, represents an intriguing subject for future research, providing the opportunity to gain profound insights into the ongoing arms race between bacteria and bacteriophages.

5.5. Methods

Cloning of NgTET, sgRNAs and sequences for homologous recombination

The gene encoding for the NgTET from *Naegleria gruberi* was purchased from IDT as a gblock and amplified by PCR. XhoI and NcoI restriction sites were introduced during the PCR amplification of the vector. The resulting PCR product was digested with XhoI and NcoI and introduced into the pET-28a vector (Merck Millipore). The ModA E165A, Alt E577A, and NgTET D243A were generated by site-directed mutagenesis. The insertion of the sgRNA sequences into pCPf1 (#122185, Addgene) and DS-SPCas (#48645, Addgene) plasmids was performed via complete plasmid amplification with the primers carrying the respective 5'-overhangs. The linearised plasmid was circularised via blunt-end ligation. All resulting plasmids were sequenced by Sanger sequencing and transformed into chemically competent *E. coli* BL21 (DE3). All primers and strains used in this work can be found in Supplementary Table 2. The plasmid maps are deposited at https://github.com/MaikTungsten/CRISPRT4

Phage T4 propagation

For phage T4 propagation, *E. coli* BL21 (DE3) cells were used. The initial culture was set to OD_{600} ~0.1 and grown at 37°C, 160 rpm until OD_{600} ~0.8 was reached. Phage T4 was added to a multiplicity of infection (MOI) of 0.8 together with 1 mM $MgCl_2$ and 1 mM $CaCl_2$, and the infection was run for 3 – 4 h at room temperature, 120 rpm. The lysate was centrifuged at 1,200 x *g* for 5 min and the supernatant was filtered through Steritop filters (pore size 0.45 µm). The concentration of the phages was determined via plaque assay and the phage suspension was stored at 4°C.

For the propagation of NgTET-treated T4 phages, the infection was performed in the same way using *E. coli* BL21 (DE3) pET28a_NgTET. The medium was supplied with 50 µM kanamycin. As the culture reached OD_{600} ~0.4, expression of NgTET was induced by the addition of 50 µM Isopropyl-β-d-thiogalactopyranosid (IPTG). The culture was grown for another 2 h at 37°C and 160 rpm and infected with phage T4 as described above. To recover NgTET-treated T4, the infection of *E. coli* BL21 (DE3) was repeated with NgTET-treated T4 as described above.

Plaque assay and efficiency of plaquing

E. coli culture of interest was grown to OD_{600} ~0.8 in the presence of strain-specific antibiotics (Supplementary Table 2). Subsequently, 300 µL of the culture were infected with T4 WT phage or NgTET–treated T4 phage with either defined or unknown MOI. The bacteria-phage suspension

was incubated at 37°C for 7 min, followed by transfer to 4 ml of LB-soft-agar (0.75%), thoroughly mixed. After mixing, the suspension was poured onto an LB-agar plate. According to the respective strains, the agar was supplemented with corresponding antibiotics. The plates were incubated at 37°C overnight and subsequently assessed for plaque-forming units (PFU) determination. To determine the efficiency of plaquing, the number of resulting plaque-forming units (pfu) in analyzed settings was divided by the input PFU.

Growth and Lysis assays

The growth and lysis of the bacterial cells were assessed by measuring the optical density of the bacterial cultures at 600 nm. This was done either by manually withdrawing samples for OD_{600} measurements or using the Tecan Spark plate reader (Tecan Group, Männedorf, Switzerland). To determine the growth curves, initial cultures were inoculated at an OD_{600} of 0.1, and the growth was monitored until the stationary growth phase was reached. For the lysis assays, the bacterial culture of interest was grown starting from OD_{600} 0.1, with the addition of antibiotics if necessary. In the case of NgTET expression, induction occurred at OD_{600} of 0.4 by adding 100 µM IPTG. Phage infection was carried out at OD_{600} of 0.8 with MOI of 3. The lysis was conducted at 23°C.

Detection of inactive mutants of Alt and ModA

The *in vivo* activity of the ARTs mutants was assessed by expressing the target proteins in *E. coli* BL21 (DE3) strains of interest (Supplementary Table 2). To confirm the presence of ADP-ribosylation events, the cells were grown at 37°C and were sampled first at OD_{600} of 0.8 for Alt and 1.2 for ModA. Afterwards, the induction was performed with 1 mM IPTG, and proteins were expressed at 37°C (for Alt and its mutants) and 4°C (for ModA and its mutants). After 1 h of expression, the cultures were sampled again and submitted to 12% SDS-PAGE. For protein visualization prior to blotting, the gels were supplemented with 5%(v/v) 2,2,2-trichloroethanol (TCE) and proteins were visualized under UV-transillumination (300 nm) for 60 s[47]. The gels were then equilibrated in transfer buffer (25 mM Tris (pH 8.3), 192 mM glycine, 0.1% (v/v) SDS, and 20% (v/v) methanol), and proteins transferred to nitrocellulose membrane (NCM) in a semi-dry manner at 400 mA for 50 min. After blotting, membranes were washed 3 times with TBS-Tween (TBS-T; 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% [v/v] Tween® 20). Subsequently, the membranes were blocked in 5% [w/v] milk powder in TBS-T for 1 h at room temperature. For the detection of ADP-ribosylated proteins, the membranes were incubated o/n at 4°C in 10 ml washing buffer (1% [w/v] milk powder in TBS-T) containing a 1:10,000 dilution of anti-pan-ADP-ribose binding reagent MABE1016 (Merck) at 4°C[48]. After washing, the membranes were incubated with

10 ml of a 1:10,000 dilution of horseradish-peroxidase-(HRP)-goat-anti-rabbit-IgG secondary antibody (Advansta) in washing buffer at room temperature for 1 h. Following another wash with PBS, the ADP-ribosylated proteins were visualized using chemiluminescence using the SignalFire ECL Reagent or the SignalFire Elite ECL Reagent (Cell Signaling Technology) according to the manufacturer's instructions.

Phage DNA isolation

Phage suspension of interest with a concentration of $>10^{10}$ PFU/ml was pretreated with 20 U DNase I (ThermoFisher Scientific, MA, USA) and 2 µL RNase A/T1 Mix (4 µg RNase A, 10 U RNase T1, ThermoFisher Scientific, MA, USA) at 37°C for 30 min to remove host-originating nucleic acids. Next, the phage was purified in a 0-45% sucrose gradient, generated in TM buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, pH 7.5). Therefore, 500 µL of the phage solution was loaded on top of the gradient and centrifugated at 70,000 x *g*, 20 min, 4°C. The fraction of the gradient containing the phages was removed with a blunt cannula and transferred into a new ultracentrifugation tube. 30 ml of ice-cold TM buffer were added and the phages were pelleted at 100,000 x *g* for 1 h at 4°C. The supernatant was discarded and the pellet was resuspended in 500 µL TM buffer and incubated at 4°C overnight. 1 µg of Proteinase K (Carl Roth, Karlsruhe, Germany) was added and the samples were incubated for another 30 minutes at 37°C. For DNA isolation, a phenol/chloroform/isoamyl alcohol (P/C/I) (Carl Roth, Karlsruhe, Germany) extraction was performed three times. The remaining phenol was removed by chloroform back-extraction (3x). The DNA was precipitated with 0.1 volume of 3 M NaOAc (pH 5.5) and 2.5 volumes of absolute ethanol at -20°C overnight. The DNA was pelleted at 15,000 x *g* at 4°C for 1 h and the pellet was washed twice with 70% ethanol. The purified DNA was resuspended in Millipore water and stored at -20°C until further use. For LC-MS analysis, the isolated DNA was processed to single nucleosides by applying Nucleoside Digestion Mix (NEB, MA, USA).

Phage T4 mutagenesis

E. coli pET-28a_NgTET_*x* + pCpf1_*x*/DS_SPcas_*x* (x: donor DNA and respective sgRNA) strains were used for mutagenesis (Supplementary Table 2). The cells were cultivated at 37°C and 160 rpm until OD_{600} ~0.4, after which NgTET expression was induced by the addition of 50 µM IPTG. The cultivation was proceeded until OD_{600} of 0.8. At this point, the cultures were adjusted to room temperature and 120 rpm and supplemented with 1 mM $MgCl_2$ and 1 mM CaCl. The cells were infected with T4 NgTET phage set to MOI of 0.5. The infection was performed for 3 h, 130 rpm, and 23°C after which the cells were pelleted and supernatant filtered through Steritop filters (pore

size 0.22 µm). The titer of the newly generated phages was determined via plaque assay and the phages were used for the counterselection via infection of *E. coli E. coli* pET-28a_NgTET_*x* + pCpf1_*x*/DS_SPcas_*x* strain. The counterselection was performed under the same conditions as the mutagenesis. The counterselected phages were filtered and used for a plaque assay with *E. coli* B strain. Single plaques were picked and transferred to 100 µl Pi-Mg buffer (26 mM Na_2HPO_4, 68 mM NaCl, 22 mM KH_2PO_4, 1 mM $MgSO_4$, pH 7.5) supplemented with 2% chloroform. The suspension was incubated for 1 h at room temperature and was further used for mutagenesis analysis via sequencing, infection of *E. coli*, or stored at 4°C until further use.

LC-MS

Relative determination of dA, dT, dC, 5hmdC, 5fdC, 5cadC, and 5ghmdC was performed using HRES LC-MS. The chromatographic separation was performed on a ThermoFisher Scientific Vanquish HPLC System using a Atlantis T3 C18 column (150 x 2.1mm, 100 A, 3 µm, Waters, MA, USA) equipped with a 20 X 2.1 mm guard column of similar specificity at a constant eluent flow rate of 0.2 ml/min and a column temperature of 40 °C with eluent A being 10 mM Ammonium Acetate in water at a pH of 4.5 and eluent B being 0.1 % of formic acid in MeOH (Honeywell, NC, USA).

The injection volume was 5 µl for standards and T4 WT samples and 20 µl for NgTET-treated samples.

The mobile phase profile consisted of the following steps and linear gradients: 0 – 1 min constant at 5% B; 1 – 5 min from 5 to 90% B; 5 – 7 min constant at 90% B; 7 – 7.1 min from 90 to 5% B; 7.1 to 12 min constant at 5% B.

A Thermo Scientific ID-X Orbitrap mass spectrometer was used in negative and positive ionization mode (separate injections) with a High-temperature electrospray ionization source and the following conditions: H-ESI spray voltage at 3400 V(+), 2400 V (-) sheath gas at 35 arbitrary units, auxiliary gas at 7 arbitrary units, sweep gas at 0 arbitrary units, ion transfer tube temperature at 300°C, svaporiser Temperature at 275°C.

Detection was performed in full scan mode using the orbitrap mass analyzer at a mass resolution of 120 000 in the mass range 200 – 450 (m/z).

Extracted ion chromatograms of the $[M-H]^-$ (dA, dT, dC, 5hmdC, 5fdC, 5cadC)/$[M+H]^+$ (5ghmdC) forms were integrated using Tracefinder software (ThermoFisher Scientific, MA, USA). Relative abundance in each sample was calculated by normalizing the peak area of each peak by the peak area of the dG signal in each specific sample, thus using dG as a sample-specific internal standard.

Screening for phage T4 mutants by multiplexed Nanopore amplicon sequencing

The method for highly multiplexed Nanopore sequencing is based on Ramírez Rojas *et al*. 2024 [40]. Briefly, the method relies on a series of two PCRs to generate the multiplex amplicon DNA for Nanopore sequencing. The first PCR attaches standardized overhangs (M13 fwd/rev sequences in this instance), serving as an amplification sequence for the second PCR attaching the barcodes. 1 µL of isolated phage T4 in Pi-Mg buffer was used as a PCR template in the first PCR with the following reaction mix: 0.125 µM each primer (Supplementary Table 2) in 1x High Fidelity Master Mix (NEB) with a total volume of 10 µl. PCR settings: 98 °C for 30 s followed by 30 cycles 98 °C 20 s, 69 °C 30 s, and 72 °C 2 min with a final hold at 72 °C 5 min. The dual barcodes were attached in a second PCR with the following reaction mix using KAPA HiFi HotStart ReadyMix (Roche), using 1 µL of a 1:10 dilution of the initial PCR as a template and 0.3 µM of the barcoding primers (Supplementary Table 2) in 7 µl reactions. PCR settings: 95 °C for 3 min followed by 20 cycles 98 °C 20 s, 66 °C 15 s, and 72 °C 60 s with a final extension 72 °C 5 min at hold at 12 °C. All the barcoded PCR reactions were pooled and purified using NucleoMag kit (Macherey Nagel) for NGS library preps. Briefly, DNA was bound to magnetic beads, washed twice with 80% ethanol, and eluted in 100 µL elution buffer (5 mM Tris-HCl pH 8.5). Concentration was determined with Nanodrop (ThermoFisher Scientific) and Qubit (Invitrogen) using the broad range and/or high-sensitivity assay. Sequencing libraries were generated with the SQK-LSK109 Ligation Sequencing kit (Oxford Nanopore Technologies) according to the manufacturer guidelines starting with 1 µg of input DNA. Sequencing was performed on Flongle flow cells (R9.4.1 chemistry) on a MinION device.

Analysis of long-read sequencing data

Nanopore raw reads were basecalled using guppy (v6.1.2 to v6.4.2), basecalled raw reads are deposited under BioProject PRJNA952186. In a miniconda environment, reads were demultiplexed using minibar[49] and mapped to the phage T4 reference genome (NC_000866.4) using minimap2 (version 2.24)[50]. The resulting SAM files were converted to BAM files, sorted, and indexed with samtools (version 1.4.1)[51]. Variant calling in the target region for mutagenesis was subsequently performed using longshot (version 0.4.1)[52] and resulting VCF files were inspected for desired point mutants with Integrative Genomics Viewer (IGV 2.16.0)[53]. Further, read counts per sample and target region were obtained from sorted BAM files using featureCounts (subread, version 2.0.1)[54] and a custom R script. Detailed code and an easy command line application for these data analysis steps are provided to the community with CRISPRT4 (code available: https://github.com/MaikTungsten/CRISPRT4)

Phylogenetic analysis

The identification of homologs for proteins 42, α-, and β-glycosyltransferases was conducted through a protein BLAST search against the NCBI database (data collected in May 2023) (Supplementary Table 1). Only sequences originating from phages were selected for further analysis to maintain specificity.

Acknowledgments

We thank Tristan Krause, Petra Mann, Ruzica Sedic and Madita Viering for their experimental assistance. K.H. is supported by funding from the Max Planck Society and the German Research Council (DFG SPP 2330, Project number 464500427). A.A.R.R. and D.S. are supported by the Max Planck Society within the framework of the MaxGENESYS project. M.W.S. is supported by funding from the Joachim Herz Foundation and the Studienstiftung des deutschen Volkes e.V.

Competing interests

K.H. and Na.P. filed a PCT application for "Engineering of Phages", European Patent Application No. 23 175 257.7. The other authors declare no competing interests.

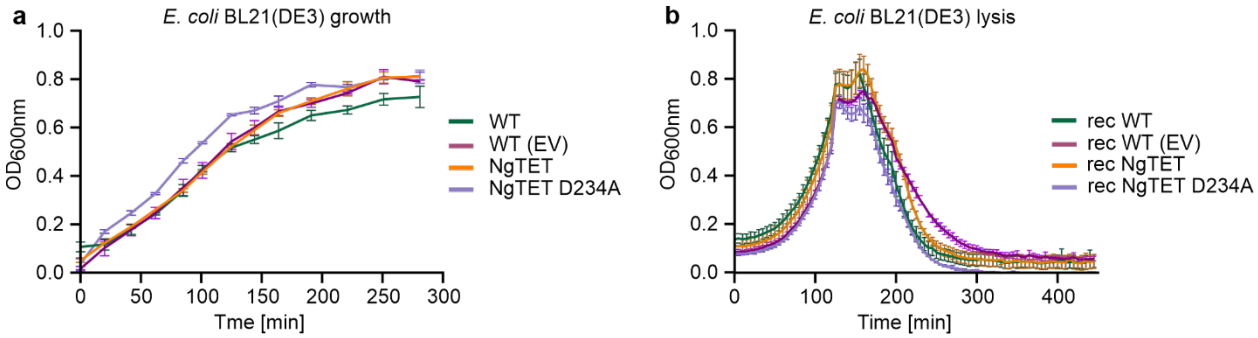
Data availability statements

https://github.com/MaikTungsten/CRISPRT4

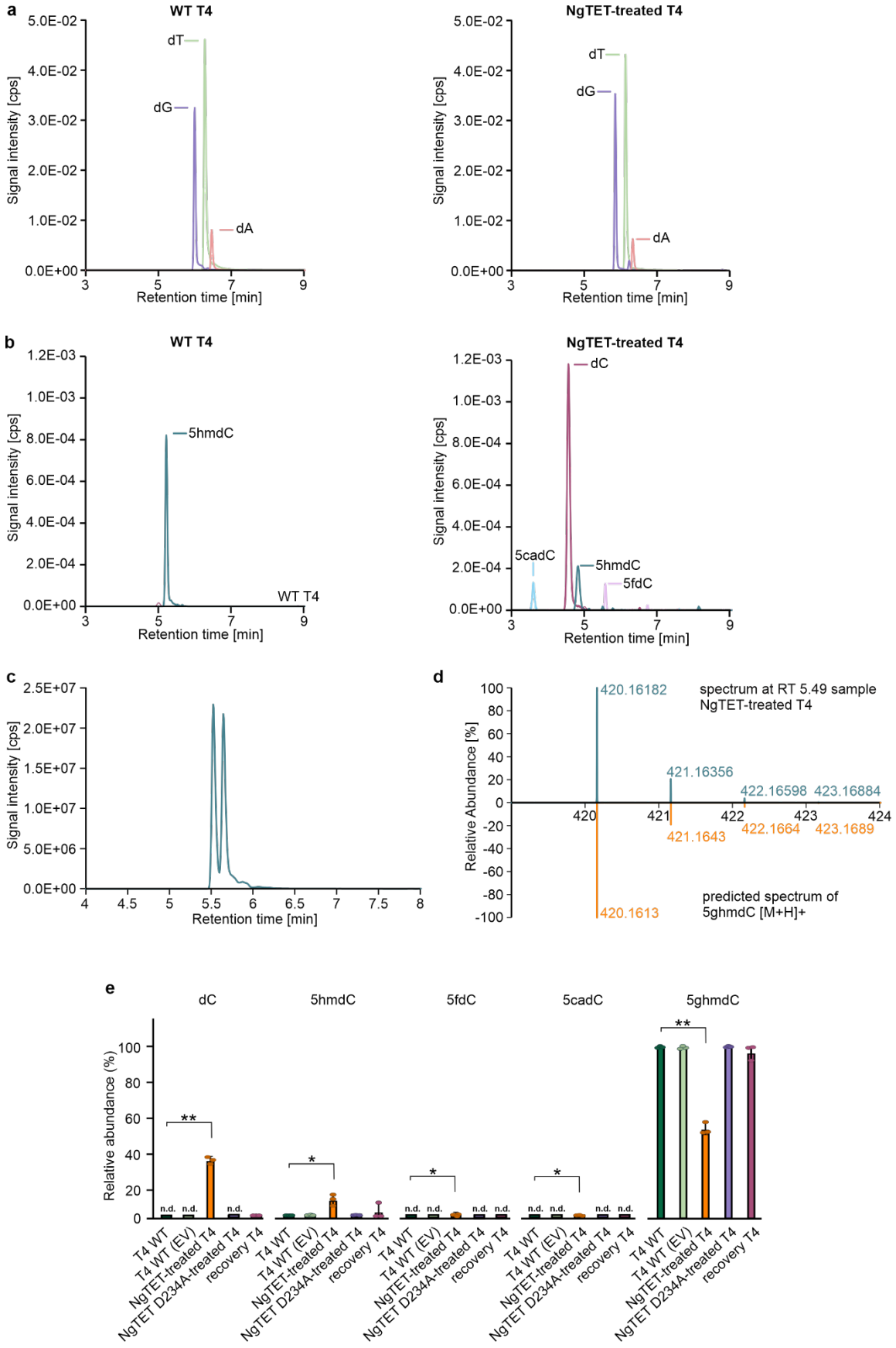
Code availability statements

https://github.com/MaikTungsten/CRISPRT4

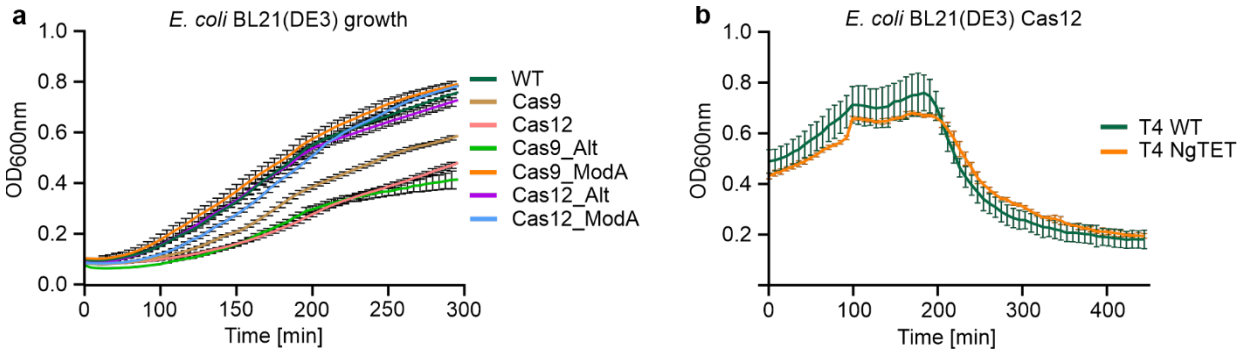
5.6. Supplementary Material



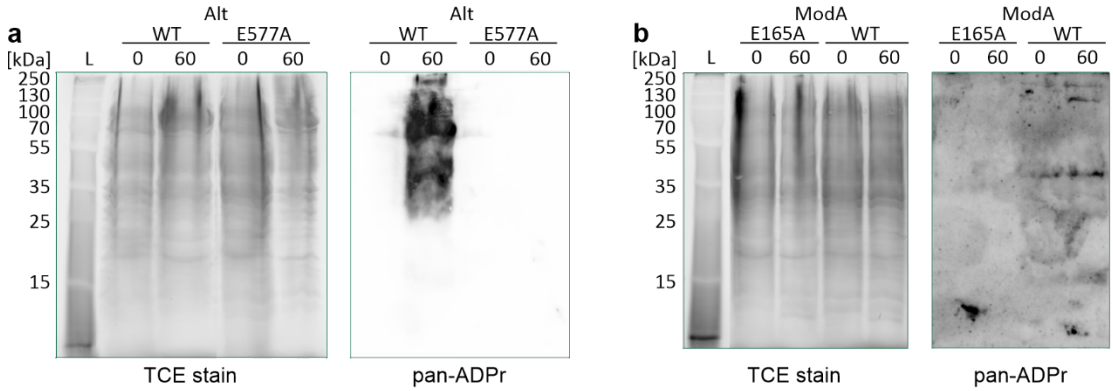
Extended Data Fig. 1: Growth and lysis of *E. coli* upon different conditions. a, Impact of NgTET recombinant expression on *E. coli* growth. *n* = 3 biological replicates. **b**, Lysis of *E. coli* by the phages recovered from different conditions. *n* = 3 biological replicates.


a
WT T4
NgTET-treated T4
dT
dG
dA
Signal intensity [cps]
Retention time [min]
b
WT T4
NgTET-treated T4
5hmdC
dC
5cadC
5fdC
c
d
spectrum at RT 5.49 sample NgTET-treated T4
420.16182
421.16356
422.16598
423.16884
421.1643
422.1664
423.1689
420.1613
predicted spectrum of 5ghmdC [M+H]+
Relative Abundance [%]
e
dC
5hmdC
5fdC
5cadC
5ghmdC
Relative abundance (%)
n.d.
T4 WT
T4 WT (EV)
NgTET-treated T4
NgTET D234A-treated T4
recovery T4

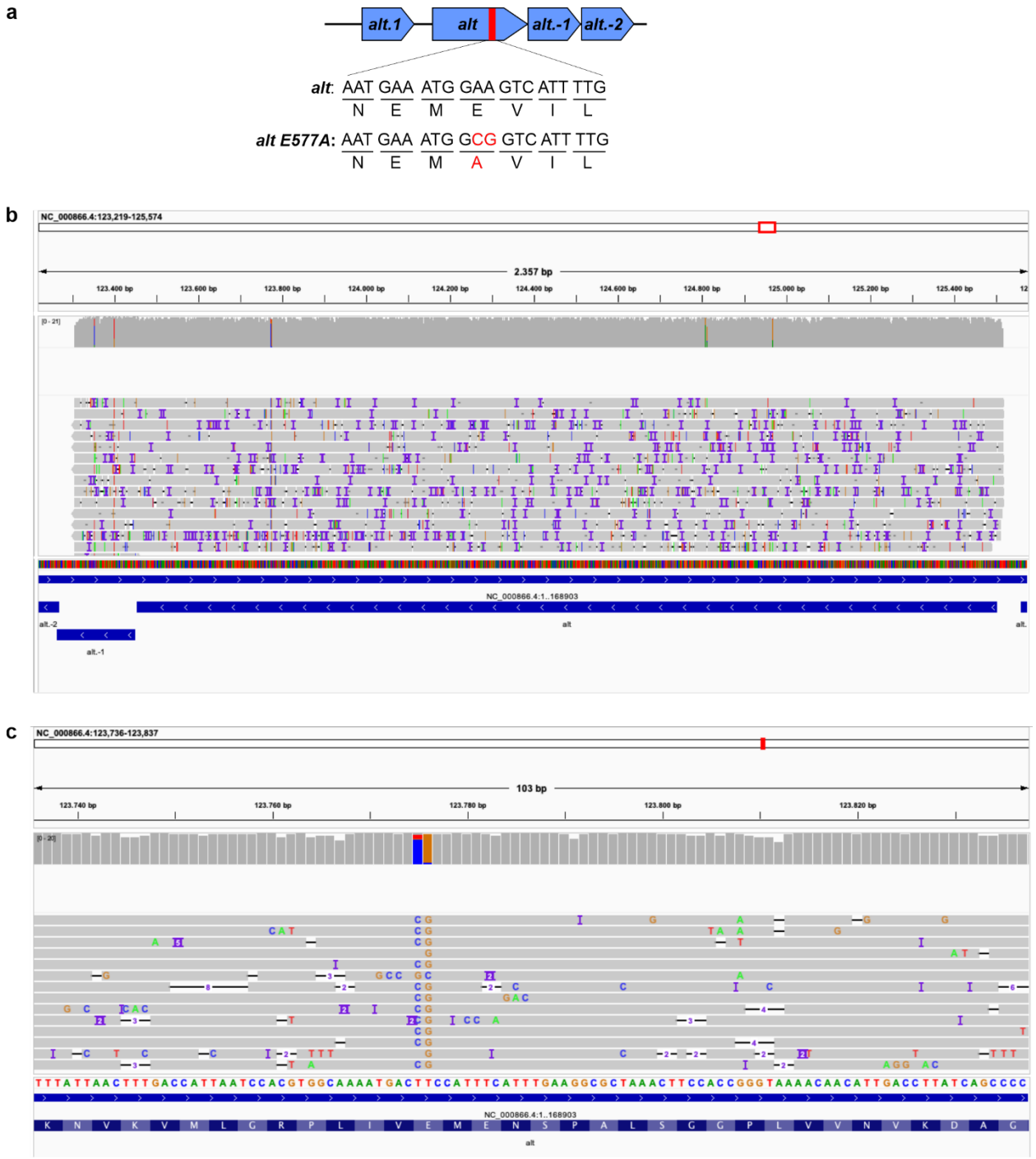
Extended Data Fig. 2: LC-MS analysis of T4 DNA composition. a-b, Extracted ion chromatogram overlay of nucleotide mass traces in T4 WT DNA (front) and NgTET-treated T4 DNA (back). Data exemplifies differences in relative abundances of cysteine nucleotides (dC, 5hmdC, 5fdC, 5cadC) (b) while dG, dT and dA (a) show similar trends. Signal intensities have been normalized against the overall nucleotide signal in the respective sample to correct for differences in sample concentration and injection volume. **c-d,** Extracted ion chromatogram of presumed 5ghmdC (c). The fragmentation pattern of presumed 5ghmdC corresponds to the predicted one (d). **e**, Relative abundance of cytosine derivatives in T4 phage isolated from different strains (two-sided Student´s t-test, * - P_{signif} < 0.05, ** - P_{signif} < 0.025; n.d.: not detected). *n* = 3 biological replicates.



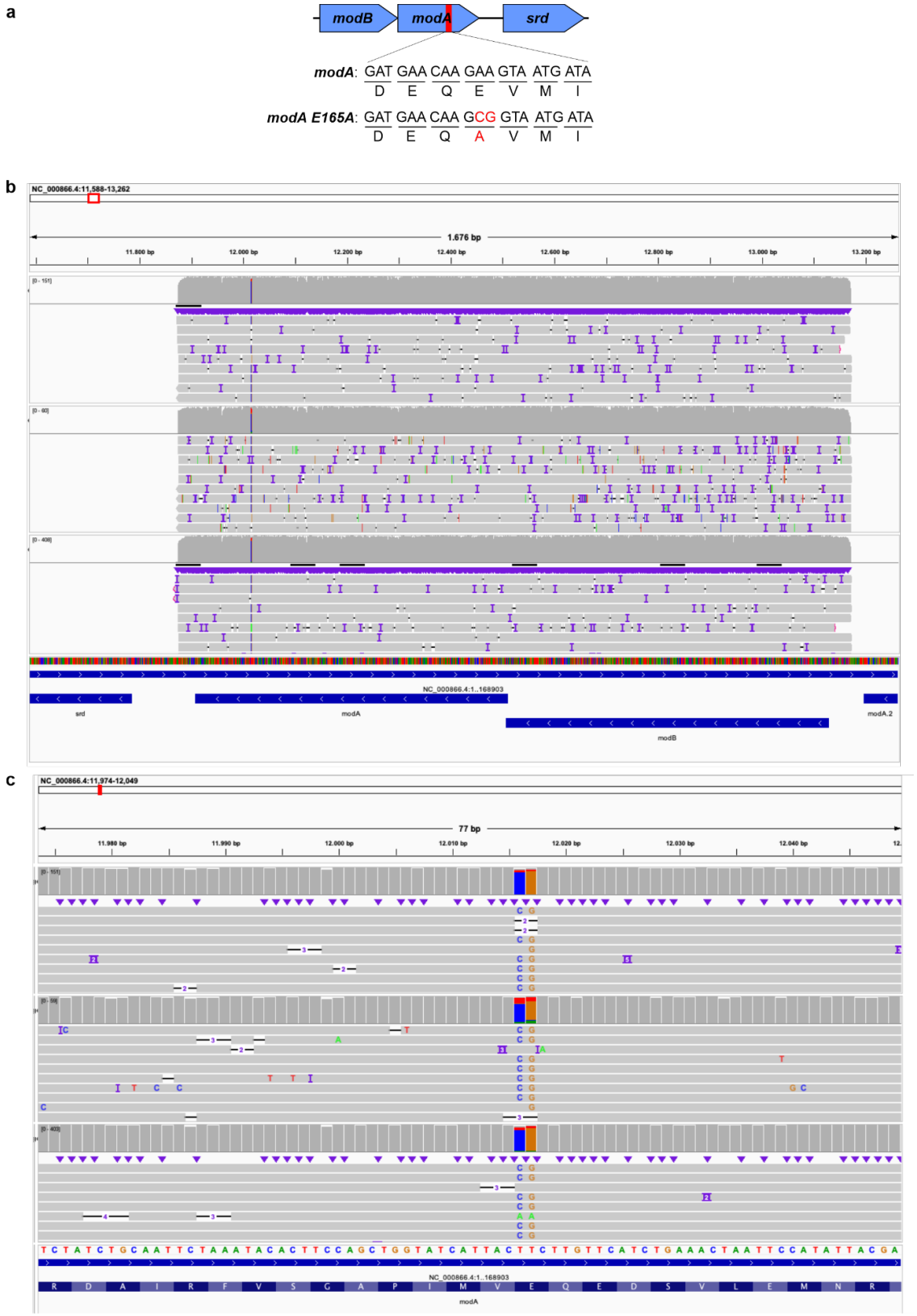
Extended Data Fig. 3: Effect of Cas expression on growth and lysis of *E. coli*. a, Impact of CRISPR/Cas9 or Cas12 systems recombinant expression on *E. coli* growth. *n* = 3 biological replicates. **b**, Impact of CRISPR/Cas12 system recombinant expression on T4 WT and T4 NgTET lysis efficiency. *n* = 3 biological replicates.



Extended Data Fig. 4: Validation of Alt and ModA ARTs inactivation via Alt E577A and ModA E165A mutations. a-b, Stain-free scan (TCE stain, loading control, left) and Western blot analysis (pan-ADPr antibody for ADP-ribosylation detection, right) to identify ADP-ribosylation events by ARTs and their mutants. Both mutants demonstrate complete abolishment of ADP-ribosylation. *n* = 3 biological replicates, a representative example is shown.



Extended Data Fig. 5: *Alt* mutants sequencing summary. a, The targeted mutation site in *alt* gene. **b**, Sequence coverage at the mutation site. **c**: Alt E577A mutant.



Extended Data Fig. 6: ModA mutants sequencing summary. a, The targeted mutation site in modA gene. **b**, Sequence coverage at the mutation site. c, ModA E165A mutant.

5.7. References

1 Moye, Z. D., Woolston, J. & Sulakvelidze, A. Bacteriophage Applications for Food Production and Processing. *Viruses-Basel* **10** (2018). https://doi.org:10.3390/v10040205

2 Schroven, K., Aertsen, A. & Lavigne, R. Bacteriophages as drivers of bacterial virulence and their potential for biotechnological exploitation. *Fems Microbiol Rev* **45** (2021). https://doi.org:10.1093/femsre/fuaa041

3 Vandamme, E. J. & Mortelmans, K. A century of bacteriophage research and applications: impacts on biotechnology, health, ecology and the economy! *J Chem Technol Biot* **94**, 323-342 (2019). https://doi.org:10.1002/jctb.5810

4 Elois, M. A., Silva, R. D., Pilati, G. V. T., Rodriguez-Lazaro, D. & Fongaro, G. Bacteriophages as Biotechnological Tools. *Viruses-Basel* **15** (2023). https://doi.org:10.3390/v15020349

5 Wolfram-Schauerte, M., Pozhydaieva, N., Viering, M., Glatter, T. & Höfer, K. Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of E. coli on Proteome and Transcriptome Levels. *Viruses-Basel* **14** (2022). https://doi.org:10.3390/v14112502

6 Drake, J. W. Ultraviolet mutagenesis in bacteriophage T-4. I. Irradiation of extracellular phage particles. *J Bacteriol* **91**, 1775-1780 (1966). https://doi.org:10.1128/jb.91.5.1775-1780.1966

7 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

8 Mahler, M., Costa, A. R., van Beljouw, S. P. B., Fineran, P. C. & Brouns, S. J. J. Approaches for bacteriophage genome engineering. *Trends Biotechnol* **41**, 669-685 (2023). https://doi.org:10.1016/j.tibtech.2022.08.008

9 Oda, M., Morita, M., Unno, H. & Tanji, Y. Rapid detection of Escherichia coli O157:H7 by using green fluorescent protein-labeled PP01 bacteriophage. *Appl Environ Microbiol* **70**, 527-534 (2004). https://doi.org:10.1128/AEM.70.1.527-534.2004

10 Srikant, S., Guegler, C. K. & Laub, M. T. The evolution of a counter-defense mechanism in a virus constrains its host range. *Elife* **11** (2022). https://doi.org:10.7554/eLife.79549

11 Tanji, Y. *et al.* Escherichia coli detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. *J Biotechnol* **114**, 11-20 (2004). https://doi.org:10.1016/j.jbiotec.2004.05.011

12 Marinelli, L. J. *et al.* BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS One* **3**, e3957 (2008). https://doi.org:10.1371/journal.pone.0003957

13 Chen, Y. *et al.* Genetic Engineering of Bacteriophages Against Infectious Diseases. *Front Microbiol* **10**, 954 (2019). https://doi.org:10.3389/fmicb.2019.00954

14 Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016). https://doi.org:10.1126/science.aad5147

15 Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synth Biol* **6**, 1952-1961 (2017). https://doi.org:10.1021/acssynbio.7b00179

16 Duong, M. M., Carmody, C. M., Ma, Q., Peters, J. E. & Nugen, S. R. Optimization of T4 phage engineering via CRISPR/Cas9. *Sci Rep* **10**, 18229 (2020). https://doi.org:10.1038/s41598-020-75426-6

17 Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* **6**, e00648 (2015). https://doi.org:10.1128/mBio.00648-15

18 Weigele, P. & Raleigh, E. A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem Rev* **116**, 12655-12687 (2016). https://doi.org:10.1021/acs.chemrev.6b00114

19 Wyatt, G. R. & Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem J* **55**, 774-782 (1953). https://doi.org:10.1042/bj0550774

20 Kornberg, S. R., Zimmerman, S. B. & Kornberg, A. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. *J Biol Chem* **236**, 1487-1493 (1961).

21 Flodman, K., Correa, I. R., Jr., Dai, N., Weigele, P. & Xu, S. Y. In vitro Type II Restriction of Bacteriophage DNA With Modified Pyrimidines. *Front Microbiol* **11**, 604618 (2020). https://doi.org:10.3389/fmicb.2020.604618

22 Kutter, E. M. & Wiberg, J. S. Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4. *J Virol* **4**, 439-453 (1969). https://doi.org:10.1128/JVI.4.4.439-453.1969

23 Dong, J. *et al.* Engineering T4 Bacteriophage for In Vivo Display by Type V CRISPR-Cas Genome Editing. *ACS Synth Biol* **10**, 2639-2648 (2021). https://doi.org:10.1021/acssynbio.1c00251

24 Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916-919 (2010). https://doi.org:10.1126/science.1186366

25 Vlot, M. *et al.* Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. *Nucleic Acids Res* **46**, 873-885 (2018). https://doi.org:10.1093/nar/gkx1264

26 Vasu, K. & Nagaraja, V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev* **77**, 53-72 (2013). https://doi.org:10.1128/MMBR.00044-12

27 Wang, S. *et al.* Landscape of New Nuclease-Containing Antiphage Systems in Escherichia coli and the Counterdefense Roles of Bacteriophage T4 Genome Modifications. *J Virol* **97**, e0059923 (2023). https://doi.org:10.1128/jvi.00599-23

28 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009). https://doi.org:10.1126/science.1170116

29 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133 (2010). https://doi.org:10.1038/nature09303

30 Liutkeviciute, Z. *et al.* Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J Am Chem Soc* **136**, 5884-5887 (2014). https://doi.org:10.1021/ja5019223

31 Feng, Y. *et al.* Direct decarboxylation of ten-eleven translocation-produced 5-carboxylcytosine in mammalian genomes forms a new mechanism for active DNA demethylation. *Chem Sci* **12**, 11322-11329 (2021). https://doi.org:10.1039/d1sc02161c

32 Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391-395 (2014). https://doi.org:10.1038/nature12905

33 Pais, J. E. *et al.* Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. *P Natl Acad Sci USA* **112**, 4316-4321 (2015). https://doi.org:10.1073/pnas.1417939112

34 Hashimoto, H. *et al.* Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. *Nucleic Acids Res* **43**, 10713-10721 (2015). https://doi.org:10.1093/nar/gkv870

35 Liu, Y. *et al.* Covalent Modifications of the Bacteriophage Genome Confer a Degree of Resistance to Bacterial CRISPR Systems. *J Virol* **94** (2020). https://doi.org:10.1128/JVI.01630-20

36 Depping, R., Lohaus, C., Meyer, H. E. & Rüger, W. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochem Bioph Res Co* **335**, 1217-1223 (2005). https://doi.org:10.1016/j.bbrc.2005.08.023

37 Koch, T. & Rüger, W. The ADP-ribosyltransferases (gpAlt) of bacteriophages T2, T4, and T6: sequencing of the genes and comparison of their products. *Virology* **203**, 294-298 (1994). https://doi.org:10.1006/viro.1994.1487

38 Tiemann, B., Depping, R. & Rüger, W. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. *Gene Expression* **8**, 187-196 (1999).

39 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004). https://doi.org:10.1128/JB.186.21.7262-7272.2004

40 Rojas, A. A. R., Brinkmann, C. K. & Schindler, D. Validation of Golden Gate assemblies using highly multiplexed Nanopore amplicon sequencing. (2024). https://doi.org:https://doi.org/10.48550/arXiv.2401.14191

41 Hutinet, G., Lee, Y. J., de Crecy-Lagard, V. & Weigele, P. R. Hypermodified DNA in Viruses of E. coli and Salmonella. *EcoSal Plus* **9**, eESP00282019 (2021). https://doi.org:10.1128/ecosalplus.ESP-0028-2019

42 Thomas, J. A., Orwenyo, J., Wang, L. X. & Black, L. W. The Odd "RB" Phage-Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-Like Phage RB69. *Viruses-Basel* **10** (2018). https://doi.org:10.3390/v10060313

43 Asokan, G. V., Ramadhan, T., Ahmed, E. & Sanad, H. WHO Global Priority Pathogens List: A Bibliometric Analysis of Medline-PubMed for Knowledge Mobilization to Infection Prevention and

Control Practices in Bahrain. *Oman Med J* **34**, 184-193 (2019). https://doi.org:10.5001/omj.2019.37

44 Wolfram-Schauerte, M. *et al.* A viral ADP-ribosyltransferase attaches RNA chains to host proteins. *Nature* **620**, 1054-1062 (2023). https://doi.org:10.1038/s41586-023-06429-2

45 Chiou, C. S. *et al.* Correcting modification-mediated errors in nanopore sequencing by nucleotide demodification and reference-based correction. *Commun Biol* **6**, 1215 (2023). https://doi.org:10.1038/s42003-023-05605-4

46 Burke, E. J. *et al.* Phage-encoded ten-eleven translocation dioxygenase (TET) is active in C5-cytosine hypermodification in DNA. *P Natl Acad Sci USA* **118** (2021). https://doi.org:10.1073/pnas.2026742118

47 Ladner, C. L., Yang, J., Turner, R. J. & Edwards, R. A. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem* **326**, 13-20 (2004). https://doi.org:10.1016/j.ab.2003.10.047

48 Gibson, B. A., Conrad, L. B., Huang, D. & Kraus, W. L. Generation and Characterization of Recombinant Antibody-like ADP-Ribose Binding Proteins. *Biochemistry-Us* **56**, 6305-6316 (2017). https://doi.org:10.1021/acs.biochem.7b00670

49 Krehenwinkel, H. *et al.* Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience* **8** (2019). https://doi.org:10.1093/gigascience/giz006

50 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018). https://doi.org:10.1093/bioinformatics/bty191

51 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org:10.1093/gigascience/giab008

52 Edge, P. & Bansal, V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat Commun* **10**, 4660 (2019). https://doi.org:10.1038/s41467-019-12493-y

53 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011). https://doi.org:10.1038/nbt.1754

54 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014). https://doi.org:10.1093/bioinformatics/btt656

Chapter VI

A viral ADP-ribosyltransferase attaches RNA chains to host proteins

Authors:
Wolfram-Schauerte, M., **Pozhydaieva, N.**, Grawenhoff, J., Welp, L. M., Silbern, I., Wulf, A., Billau, F. A., Glatter, T., Urlaub, H., Jäschke, H., Höfer, K.

Published in:
Nature, 2023, 620(7976), 1054-1062

Author contributions:
K.H. and A.J. designed the study. K.H., M.W.-S., J.G., F.A.B. and N.P. cloned, expressed, purified and analysed the ARTs and their target proteins. K.H., I.S., L.M.W., A.W. and M.W. prepared samples for mass spectrometry. I.S., L.M.W., A.W. and H.U. developed an LC–MS/MS pipeline to study ADP-ribosylation and RNAylation, and analysed the data. T.G. performed mass spectrometry analysis of rS1. M.W.-S. developed the RNAylomeSeq pipeline and analysed the data. N.P. created and characterized the ModB mutant phage. K.H., H.U. and A.J. supervised the work. K.H., M.W.-S. and A.J. wrote the first draft, and all authors contributed to reviewing, editing and providing additional text for the manuscript.

6. A viral ADP-ribosyltransferase attaches RNA chains to host proteins

6.1. Abstract

The mechanisms by which viruses hijack the genetic machinery of the cells they infect are of current interest. When bacteriophage T4 infects *Escherichia coli*, it uses three different adenosine diphosphate (ADP)-ribosyltransferases (ARTs) to reprogram the transcriptional and translational apparatus of the host by ADP-ribosylation using nicotinamide adenine dinucleotide (NAD) as a substrate[1,2]. NAD has previously been identified as a 5′ modification of cellular RNAs[3-5]. Here we report that the T4 ART ModB accepts not only NAD but also NAD-capped RNA (NAD–RNA) as a substrate and attaches entire RNA chains to acceptor proteins in an 'RNAylation' reaction. ModB specifically RNAylates the ribosomal proteins rS1 and rL2 at defined Arg residues, and selected *E. coli* and T4 phage RNAs are linked to rS1 *in vivo*. T4 phages that express an inactive mutant of ModB have a decreased burst size and slowed lysis of *E. coli*. Our findings reveal a distinct biological role for NAD–RNA, namely the activation of the RNA for enzymatic transfer to proteins. The attachment of specific RNAs to ribosomal proteins might provide a strategy for the phage to modulate the host's translation machinery. This work reveals a direct connection between RNA modification and post-translational protein modification. ARTs have important roles far beyond viral infections[6], so RNAylation may have far-reaching implications.

6.2. Main

ARTs catalyse the transfer of one or multiple ADP–ribose (ADPr) units from NAD to target proteins[7]. Bacterial and archaeal ARTs act as toxins and are involved in host defence or drug-resistance mechanisms[8], whereas eukaryotic ARTs have roles in distinct processes ranging from DNA damage repair to macrophage activation and stress response[9]. Viruses use ARTs as weapons to reprogram the host's gene-expression system6. Mechanistically, a nucleophilic residue of the target protein (usually Arg, Glu, Asp, Ser or Cys) attacks the glycosidic carbon atom in the nicotinamide riboside moiety of NAD, forming a covalent bond as N-, O- or S-glycoside[7] (Fig. 1a). As the adenosine moiety of NAD is not involved in this reaction, we speculated that elongation of the adenosine to long RNA chains (by means of regular 5′–3′ phosphodiester bonds) might be tolerated by ARTs, potentially leading to the formation of covalent RNA–protein conjugates (Fig. 1b). RNAs that have a 5′-NAD cap have previously been found in bacteria (including E. coli[3,10,11]), archaea[12,13] and eukaryotes[5,14-19], with NAD–RNA concentrations ranging from 1.9 to 7.4 fmol µg−1 RNA16. This modification was observed in different types of RNA, including mRNA and small regulatory RNA (sRNA)[20]. However, little is known about the biological functions of this RNA cap[21].

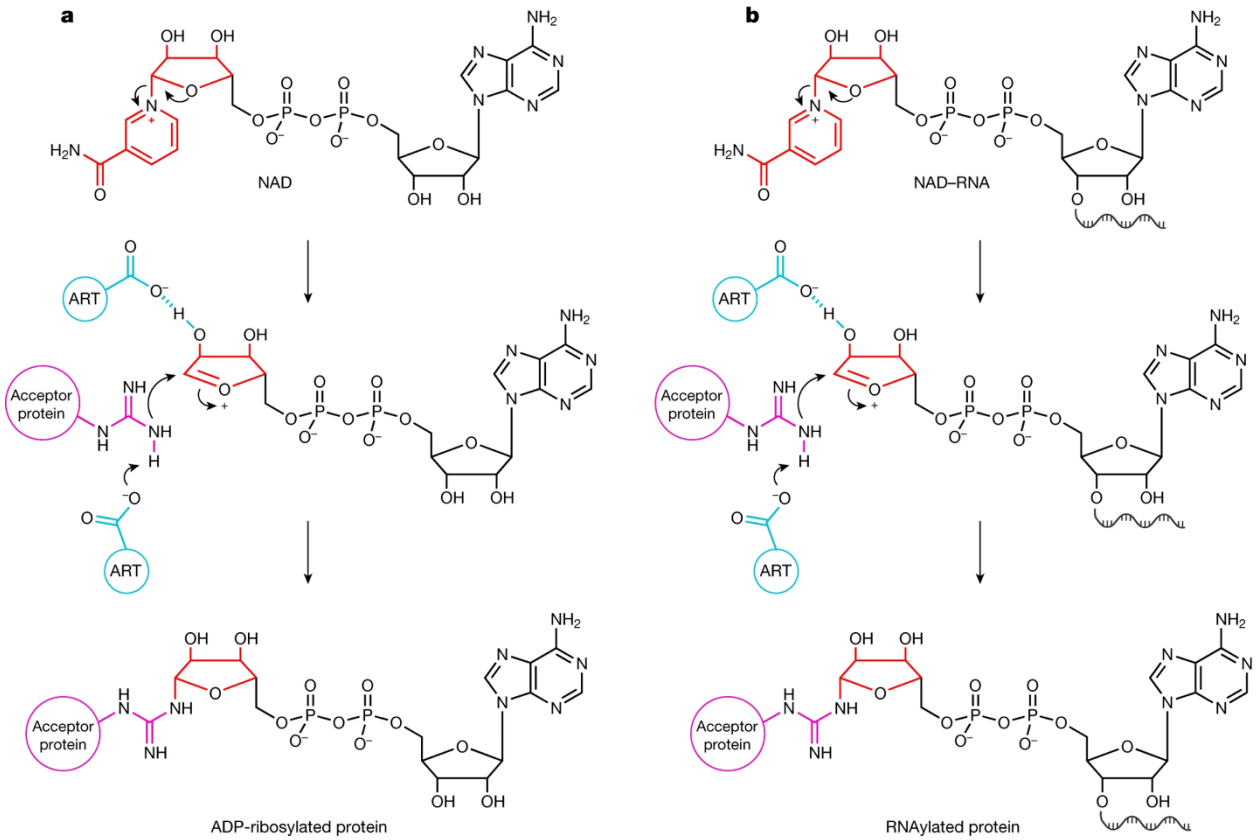


Fig. 1: Mechanisms of ADP-ribosylation and proposed RNAylation. a) The mechanism of ADP-ribosylation for Arg. Initially, the N-glycosidic bond between the ribose and nicotinamide is destabilized by a Glu residue of an ART. This leads to the formation of an oxocarbenium ion of ADPr, with nicotinamide as the leaving group. This electrophilic ion is attacked by a nucleophilic Arg residue of the acceptor protein after Glu-mediated proton abstraction, leading to the formation of an N-glycosidic bond[22]. **b)** Our proposed RNAylation-reaction mechanism. In a similar way to ADP-ribosylation in the presence of NAD, we propose that ARTs might use NAD–RNA to catalyse an RNAylation reaction, thereby covalently attaching an RNA to an acceptor protein. Red, nicotinamide riboside of NAD and NAD-RNA; blue, catalytic residues of the ART; purple, nucleophilic Arg residue of the acceptor protein.

The infection cycle of bacteriophage T4 relies on the sequential expression of early, middle and late phage genes that are transcribed by *E. coli* RNA polymerase (RNAP)[23]. For the specific temporal reprogramming of the *E. coli* transcriptional and translational apparatus, the T4 phage uses 3 ARTs that modify more than 30 host proteins. Upon infection, one of these ARTs, Alt, is injected into the bacterium with the phage DNA and immediately ADP-ribosylates *E. coli* RNAP at different residues, which is thought to result in the preferential transcription of phage genes from early promoters[24,25]. Two early phage genes encode the ARTs ModA[26] and ModB[1,27]. ModA completes the ADP-ribosylation of RNAP, whereas ModB is thought to modify the host protein rS1

(refs. [1,27]). However, it is still not known how ADP-ribosylation changes the properties of the target proteins, or whether other proteins are also modified during T4 infection.

6.3. Results

ModB catalyses RNAylation *in vitro*

To test our idea that ARTs may accept NAD–RNAs as substrates, we purified Alt, ModA and ModB. We incubated them with either a synthetic, site-specific ^{32}P-labelled 5′-NAD–RNA 8-base oligonucleotide (8-mer) or a 3′-fluorophore-labelled 5′-NAD–RNA 10-mer to test for either self-modification or the modification of target proteins. Whereas both Alt and ModA showed only a small amount of target RNAylation (Extended Data Fig. 1a), ModB rapidly RNAylated its known ADP-ribosylation target protein, rS1, without detectable self-RNAylation (Fig. 2a and Extended Data Fig. 1b). By contrast, ModB-mediated ADP-ribosylation in the presence of ^{32}P-NAD resulted in the modification of both proteins (ModB and rS1) with similar intensity (Fig. 2b and Extended Data Fig. 1c). No signal was evident when either ModB or rS1 was missing, or when a 5′-^{32}P-monophosphate–RNA (5′-^{32}P–RNA) of the same sequence was used as a substrate for ModB (Extended Data Fig. 1d). Moreover, a mutated active site (R73A, G74A) of ModB also prevented the RNAylation of rS1 (ref. [1]) (Extended Data Fig. 2a,b). This mutation similarly affected both the ADP-ribosylation and the RNAylation activity of ModB.

RNAylation follows an ADP-ribosylation-like mechanism

ModB-catalysed RNAylation of rS1 was strongly inhibited by the ART inhibitor 3-methoxybenzamide (3-MB)[28], which is thought to mimic the nicotinamide moiety (Extended Data Fig. 2c), confirming an ADP-ribosylation-like mechanism. Moreover, RNAylated rS1 proteins that carry a ^{32}P-labelled ADPr moiety were treated with the ribonuclease (RNase) T1 to determine whether the RNA and the protein are covalently linked (Extended Data Fig. 2d). This treatment would remove the ^{32}P label if the RNA were non-covalently bound to rS1 or covalently linked at any position other than the 5′-terminal positions. The ^{32}P-rS1 signal did not disappear after treatment with T1, but it disappeared entirely after treatment with trypsin, which breaks down rS1 (Extended Data Fig. 2e). Collectively, these data indicate that the RNA is covalently linked to rS1 at its 5′ end, as shown in Fig. 1b.

RNAylation assays using short linear or hairpin-forming NAD–RNAs (Fig. 2c and Extended Data Fig. 3a) revealed that ModB has a preference for unstructured NAD–RNAs as a substrate, although it also accepted longer, biologically relevant NAD-capped RNAs as substrates, such as a NAD-capped Qβ RNA fragment of around 100 nucleotides[29] (Fig. 2d and Extended Data Fig.

3b). RNAylation with NAD-capped 100-nucleotide RNA caused the modified rS1 protein to migrate with an apparent mass of 100 kDa (Fig. 2e). Treatment of the RNAylated protein with nuclease P1, which hydrolyses 3′–5′ phosphodiester bonds but does not attack the pyrophosphate bond of the 5′-ADPr, reversed this shift, and the ^{32}P-labelled product migrated in a similar way to unmodified rS1 or ADPr–rS1 (Fig. 2e), confirming the proposed nature of the covalent linkage.

To exclude the possibility that ModB removes only the nicotinamide moiety from the NAD–RNA by hydrolysis, thereby generating a highly reactive ribosyl moiety that could (through its masked aldehyde group) spontaneously react with nucleophiles in its vicinity[30], we prepared ADPr-modified RNA and tested it as a substrate for ModB. No modification could be detected (Extended Data Fig. 3c), providing no support for spontaneous RNAylation.

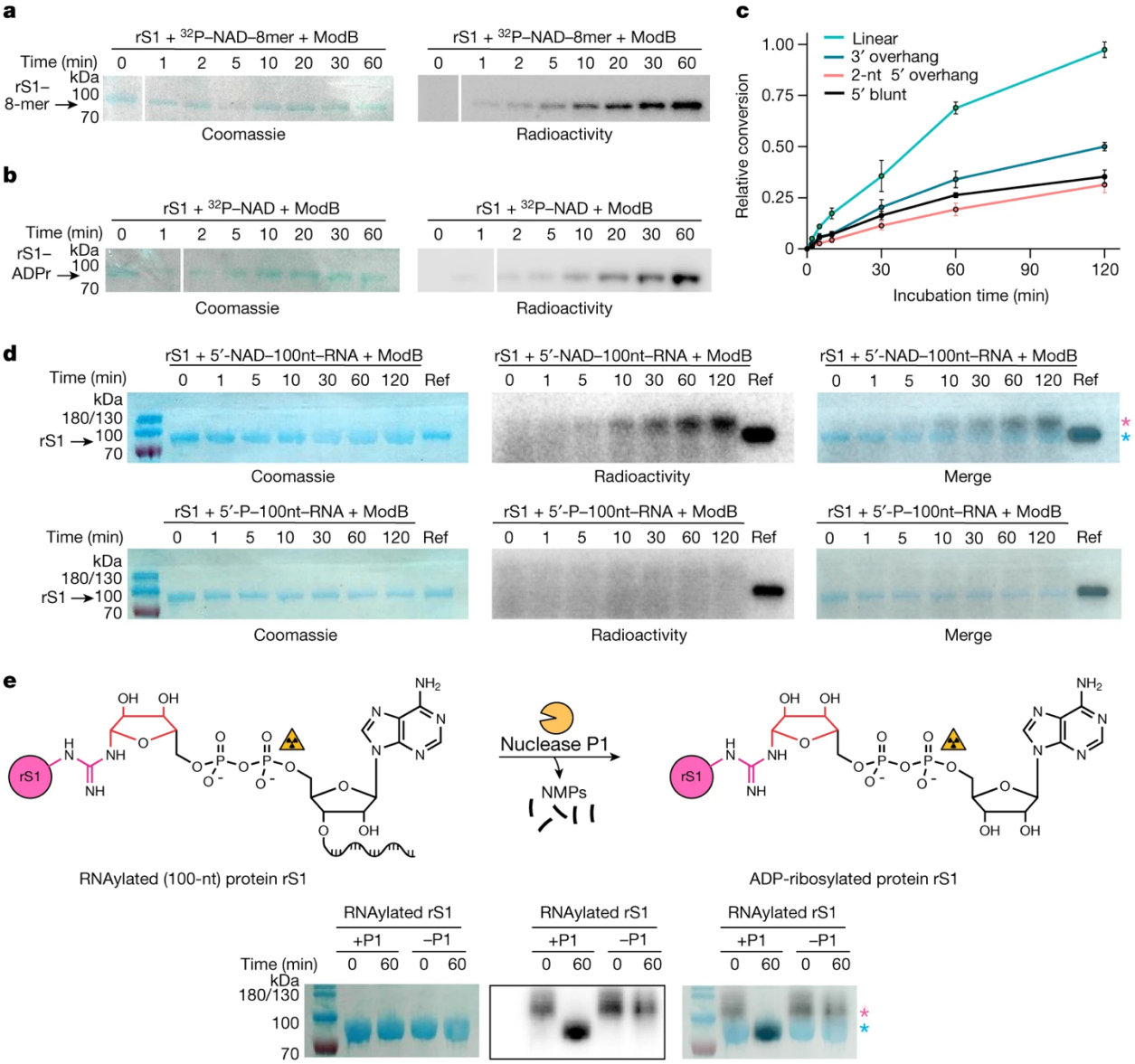


Fig. 2: Post-translational protein modification of rS1 by ModB *in vitro*.

a) Time course of the RNAylation of rS1 by ModB (n = 3). SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gels are shown for rS1 + ^{32}P–NAD–8-mer + ModB. Complete gels and a reaction schematic are shown in Extended Data Fig. 1b. **b)** Time course of the ADP-ribosylation of rS1 by ModB (n = 3), showing rS1 + ^{32}P–NAD + ModB. Complete gels and a reaction schematic are shown in Extended Data Fig. 1c. rS1 RNAylation (a) and ADP-ribosylation (b) are indicated by the acquisition of a radioactive signal overlapping with the Coomassie stain. **c)** The role of RNA secondary structure on RNAylation reaction. Four different 3′ Cy5-labelled NAD-capped RNAs were tested, including a linear 10-mer NAD-capped RNA and three structured NAD-capped RNAs with a 3′ overhang, a dinucleotide 5′ overhang or a blunt end. SDS–PAGE analysis is shown in Extended Data Fig. 3a. Relative conversion refers to the intensity of the RNAylated rS1 band relative to the maximal RNAylation intensity observed among all four tests. Data points represent mean ± s.d. values based on quantification of fluorescence Cy5 signals (n = 3 biologically independent replicates). **d)** *In vitro* kinetics of the RNAylation of rS1 by ModB using 5′-NAD–100-nucleotide (100-nt) RNA as the substrate (top), analysed by SDS–PAGE. The pink asterisk indicates shifted RNAylated rS1; the blue asterisk indicates ADP-ribosylated rS1. ADP-ribosylated rS1 serves as a reference (Ref). The mass of 100 nucleotides is around 30 kDa; RNAylated rS1 has a mass of around 100 kDa (70 kDa from rS1, 30 kDa from RNA). 5′-P–100nt RNA was used as a negative control (bottom, n = 2). The two bands above the 100 kDa band are denoted 180/130. **e)** The nuclease P1 breaks down RNAylated protein rS1. The covalently attached 100-nucleotide-long RNA results in a shift of the RNAylated protein rS1 (which has a mass of around 100 kDa) in SDS–PAGE. Nuclease P1 cleaves the phosphodiester bond, resulting in degradation of the attached RNA into mononucleotides. Nuclease P1 converts RNAylated rS1 into ADP-ribosylated rS1 (mass of around 70 kDa), which can be seen by the presence of a downshifted protein band in the SDS–PAGE gel (n = 1). Red, ribose moiety of RNAylated/ADP-ribosylated protein; NMPs, nucleoside monophosphates; radioactivity symbol indicates site of ^{32}P-label; pacman symbolizes nuclease P1. The pink and blue asterisks are the same as in **d**.

To exclude the degradation of RNA during RNAylation, we supplied ModB with an NAD–RNA 10-mer that carried a fluorescent dye (Cy5) at the 3′ terminus (Extended Data Figs. 2a and 3a). The time-course analysis of the RNAylation indicates that intact oligonucleotide chains were attached to rS1 for a variety of NAD-capped RNAs (Extended Data Fig. 3a).

ModB modifies Arg residues in rS1

To identify the amino acid residues in protein rS1 to which RNA chains are covalently linked during RNAylation, we used tools developed to analyse protein ADP-ribosylation.

The radioactive signal of ^{32}P-RNAylated protein rS1 and ^{32}P–ADP-ribosylated rS1 did not change after treatment with $HgCl_2$ (which cleaves S-glycosides at Cys residues), NH_2OH (which hydrolyses O-glycosides at Asp and Glu) (Extended Data Fig. 4a) or recombinant enzyme ARH3

(which hydrolyses O-ADPr glycosides specifically at Ser residues) (Extended Data Fig. 4b), although it was efficiently removed by treatment with human ARH1 (Fig. 3a,b and Extended Data Fig. 4c,d). These findings indicate that the main products of ModB-catalysed RNAylation are linked as N-glycosides by Arg residues (Extended Data Fig. 4c,d).

To establish that ModB-mediated ADP-ribosylation or RNAylation also occurs at Arg residues *in vivo*, we isolated genomically His-tagged rS1 from non-infected or T4-infected *E. coli*. Analysis using liquid chromatography with tandem mass spectrometry (LC–MS/MS) confirmed that there was specific modification of Arg residues in rS1 with ADPr. These ADPr modifications were present only in the T4-infected sample (Extended Data Table 1 and Supplementary Table 1). R139 was identified as a modified residue, as confirmed by site-directed mutagenesis to Lys or Ala; rS1(R139K) and rS1(R139A) mutants were expressed in T4-infected *E. coli*, purified and analysed, revealing that these mutations prevent modification at those positions (Extended Data Table 2 and Supplementary Table 2).

LC–MS/MS analysis verifies RNAylation

The LC–MS/MS analysis above did not show unambiguously that the modification of rS1 was derived from RNAylated or ADP-ribosylated rS1. We therefore optimized LC–MS/MS to detect the covalent attachment of RNA to rS1. For this analysis, *in vitro* RNAylated, truncated rS1 protein was subjected to an RNase A/T1 and tryptic digest. The obtained mixture was directly subjected to LC–MS/MS analysis, and MS data were evaluated using the RNPXL software tool[31], on the assumption that the RNAylated rS1 peptide still has a trinucleotide (ADPr–cytidine) attached. The LC–MS/MS analysis this time showed the covalent attachment of a trinucleotide (ADPr–cytidine) to an rS1 peptide encompassing amino acid positions 129–150. Strikingly, the precursor mass ($[M + 3H]^{3+}$ with a mass-to-charge ratio (m/z) = 1,115.81, expected molecular mass = 3,344.41 Da) plus the gas-phase b- and y-type fragmentation pattern, which shows the characteristic neutral loss of CH_2N_2 (derived from a modified Arg[32]) or ribose, ADPr or ADPr-A′ adducts, revealed that the RNA is attached by an N-glycosidic bond to R139 and/or R142 (Fig. 3c–e, Extended Data Fig. 5 and Supplementary Table 3). We could not unambiguously assign the modified Arg because of the low intensity of the respective fragment ions and the occurrence of mixed spectra containing ion fragments of the same peptide species modified at different sites (Fig. 3c–e).

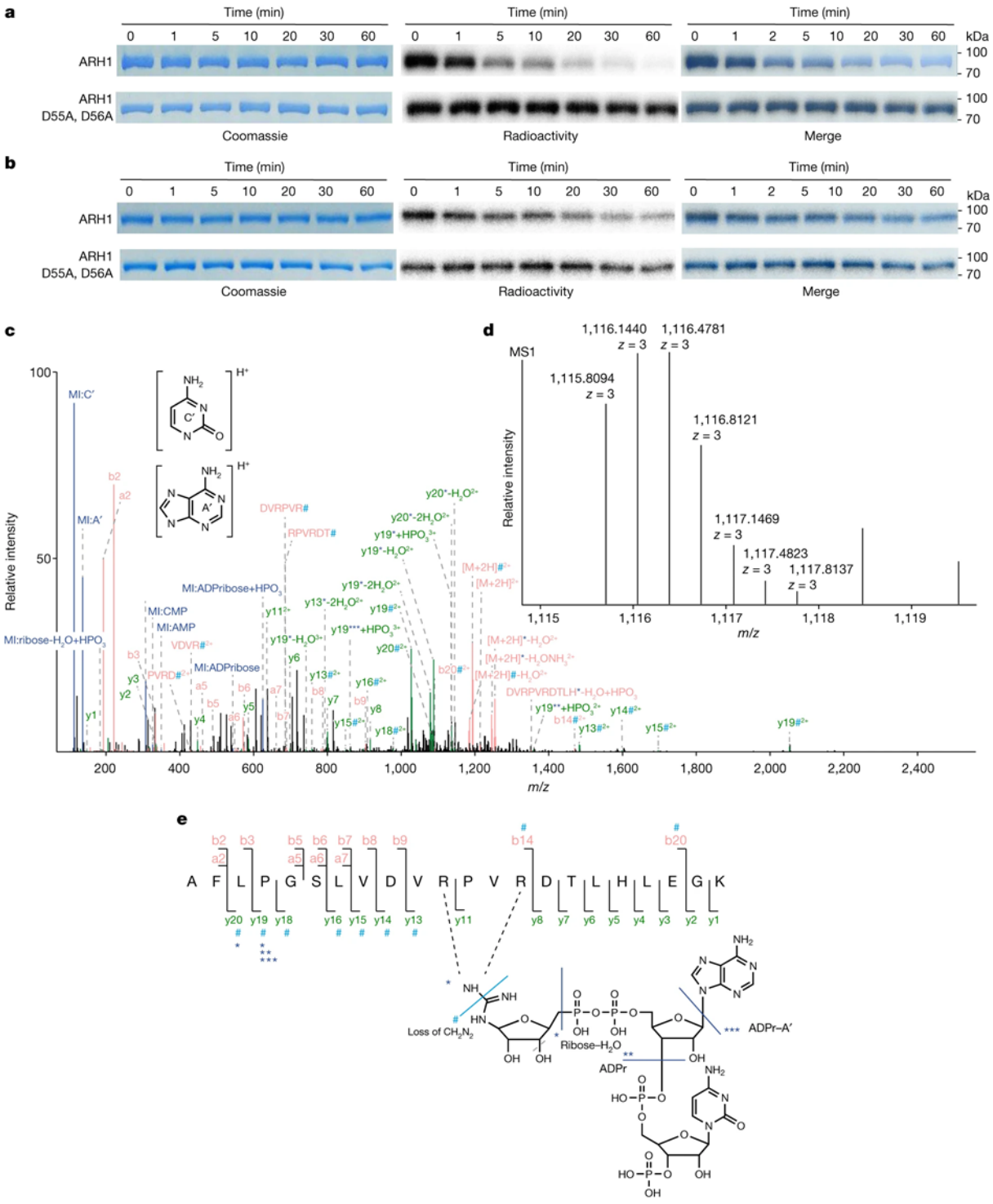


Fig. 3: Identification of RNAylation sites of rS1.

a,b) Specific removal of ADP-ribosylation and RNAylation by ARH1 (n = 3). Schematics of the reaction are shown in Extended Data Fig. 4c,d. Enzyme kinetics of ARH1 in the presence of ADP-ribosylated (a) or RNAylated (b) protein rS1 were analysed by SDS–PAGE. Mutation of the catalytically important residues D55 and D56 abolished the removal of ADP-ribosylation and RNAylation. **c-e)** Tandem MS-based

identification of RNAylated rS1 peptide. **c)** The MS/MS fragment ion spectrum (spectrum ID: 23723) of RNAylated rS1 peptide AFLPGSLVDVRPVRDTLHLEGK carrying ADPr plus cytidine monophosphate and a 3′ phosphate group. The spectrum shows marker ions (MI) of adenine (A′) and cytosine (C′), adenosine monophosphate (AMP), cytidine monophosphate (CMP), ribose–H2O and ADPr. The precursor ion ($[M + 2H]^{2+}$) and fragment ions y13–y16, y18–y20, b14 and b20 show a specific loss of mass of 42.021798 Da (#), which can be explained by the loss of CH_2N_2 at the modified Arg^{32}. Precursor ions, y13, y19 and y20 are shifted by the mass of ribose–H_2O (*). The spectrum also shows precursor ions and y19 being shifted by ADPr with (**) and without (***) the loss of adenine. Blue, MI; red, precursor ions, internal fragment ions, b-type fragment; green, y-type fragment ions. **d)** Isotopic peak pattern of the precursor ion as detected in the MS precursor ion scan for the MS/MS spectrum shown in c. **e)** Sequence and RNA adduct representation of the RNAylated peptide shown in c and d, including annotations of unshifted fragment ions and fragment ions showing arginine loss (#), as well as ribose–H_2O (*), ADPr (**) and ADPr–adenine (***). The fragmentation products of the ADPr + CMP + 3′-phosphate adduct observed in the MS/MS spectrum shown in c are indicated in the structure by light blue (mass loss) and dark blue (mass adducts) lines.

rS1 is RNAylated and ADP-ribosylated *in vivo*

To distinguish quantitatively between ADP-ribosylation and RNAylation *in vivo*, we used immunoblotting with an antibody-like ADPr-binding reagent (pan-ADPr) that specifically recognizes ADP-ribosylated proteins but detects RNAylated proteins only after treatment with nuclease P1 (Fig. 4a and Extended Data Fig. 6a). rS1 was expressed in non-infected or T4-infected *E. coli*, affinity-purified and its ADP-ribosylation was analysed with pan-ADPr. We found extensive ADP-ribosylation of rS1 only in the T4-infected sample. After treatment with nuclease P1, the pan-ADPr signal intensity of the rS1 band increased (Fig. 4b and Extended Data Fig. 6b), indicating RNAylation of rS1. Thus rS1 was found to be both ADP-ribosylated and RNAylated *in vivo*, with RNAylation accounting for around 30 % of the modifications. It remained unclear, however, whether the two modifications are mutually exclusive or can occur simultaneously in the same molecule at different sites. Moreover, the signal for ADPr disappeared after ARH1 treatment, further confirming the nature of the RNA–protein linkage (Fig. 4b and Extended Data Fig. 6b). We found that the ADP-ribosylation and RNAylation of rS1 occur in parallel *in vivo*.

ModB RNAylates proteins with selected RNAs

To identify the RNAs linked to rS1 by ModB during infection by the T4 phage, we developed an RNAylomeSeq approach (Extended Data Fig. 6c) in which genomically His-tagged rS1 was isolated from T4-infected *E. coli* and captured on Ni-NTA beads. In a similar way to NAD

captureSeq[33], RNA was reverse-transcribed 'on-bead' and the resulting cDNA was amplified by PCR and analysed using next-generation sequencing.

We applied this workflow to *E. coli* treated with wild-type (WT) T4 phage. As a negative control, we used CRISPR–Cas9 technology to generate a T4 phage that expressed the catalytically inactive mutant ModB(R73A, G74A) (ref. [34]). We compared the abundance of reads mapped to individual RNA species and identified specific *E. coli* and T4 phage RNAs enriched in WT T4 phage samples (Fig. 4c, Extended Data Fig. 6d,e, Supplementary Table 4 and Supplementary Fig. 3). Several of the *E. coli* transcripts (mRNAs and sRNAs) have been reported to be 5′-NAD-capped in *E. coli*[3,35], including RNAs of the genes acpP, glmY, mcaS, oxyS, aspA and rob, which makes them suitable substrates for ModB. We also identified phage transcripts, such as ipIII (internal head protein III), that were enriched in our datasets (Fig. 4c, Extended Data Fig. 6d,e and Supplementary Table 4). The enriched RNAs do not share any common features apart from adenosine (+1A) at the transcription start site, which is crucial for the biosynthesis of NAD-capped RNAs *in vivo*[36].

ModB RNAylates OB-fold proteins

To understand how ModB identifies its target proteins, we analysed the structural features of known target proteins. rS1 contains oligonucleotide-binding (OB)-fold domains[29]. One structural variant of OB folds is the S1 domain, which is present in rS1 in six copies that vary in sequence (Extended Data Fig. 7a). RNAylated R139 and R142 are located in domain 2 of rS1. We speculated that the S1 domain might be important for substrate recognition by ModB. To characterize the specificity of ModB for different S1 domains, we cloned, expressed and purified each S1 domain of rS1 (D1–D6) and tested them in an RNAylation assay (Fig. 4d and Extended Data Fig. 7b). In agreement with the mass spectrometry (MS) data (Extended Data Table 1 and Supplementary Table 1), we detected strong RNAylation signals for rS1 D2 and D6, whereas rS1 D1, D3, D4 and D5 were modified to a lesser extent. Multiple sequence alignment of rS1 D2 and D6, and the S1 domain of *E. coli* PNPase, revealed that these S1 domains share an Arg residue as part of the loop that connects strands 3 and 4 of the β-barrel[37] (Extended Data Fig. 7c). This loop is packed on the top of the β-barrel and might therefore be accessible to ModB. For rS1 D2, the residues R139 and R142 are the sites of RNAylation identified by MS (Fig. 3e–g and Supplementary Tables 1–3). Mutation analysis confirmed that the RNAylation level of D2 is significantly reduced if R139 is replaced by Ala or Lys (Extended Data Fig. 8a,b). *E. coli* RNase E also has an S1 domain in its active site with an Arg in the loop between strands 3 and 4. In the RNAylation *in vitro* assays, RNase E was modified by ModB, whereas control proteins without the

S1 domain (such as BSA and the NudC inactive mutant) were not. These data suggest that OB folds such as S1 domains with an embedded Arg are RNAylation target motifs (Fig. 4e).

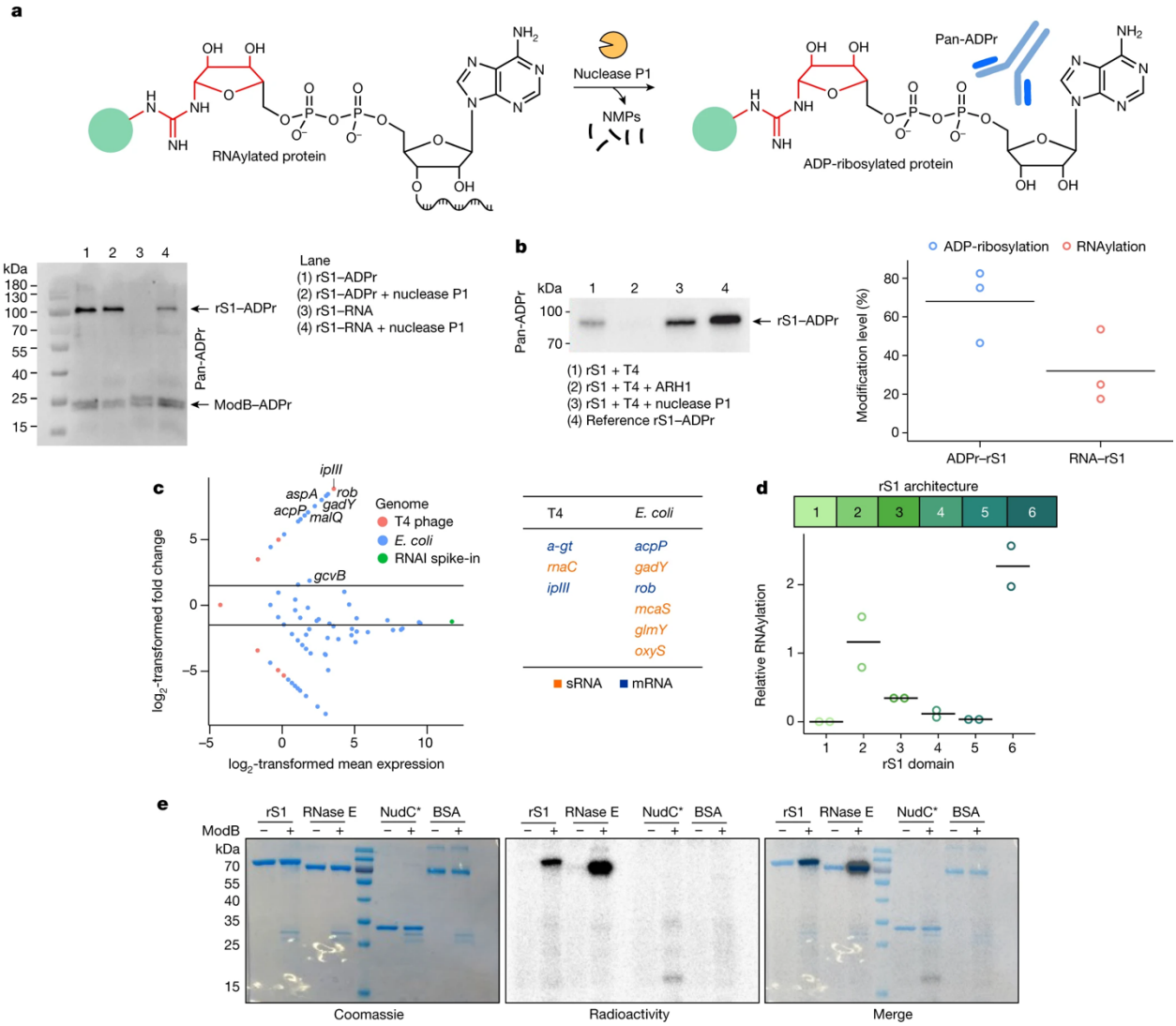


Fig. 4: *In vivo* characterization of ADP-ribosylation and RNAylation.

a) Quantification of the RNAylation of rS1 using a nuclease P1 digest and western blot analysis. Green circle represents the protein. **b)** Quantification of rS1 RNAylation *in vivo* based on biological triplicates (n = 3). Data are shown as mean (grey bar) and individual data points. Complete blots and intensity normalization are shown in Extended Data Fig. 6b. **c)** Identification of RNA substrates of ModB using RNAylomeSeq. The MA plot shows data for one of three biological replicates (n = 3). Further details are given in Extended Data Fig. 6c,d. **d)** Quantification of the RNAylation of rS1. Modification of rS1 domains 1–6 (n = 2 biologically independent replicates; black lines show the mean). **e)** SDS–PAGE analysis of the RNAylation of protein rS1, RNase E, inactive NudC mutant (NudC*: V157A, E174A, E177A, E178A) and bovine serum albumin (BSA) by ModB (n = 2 biologically independent replicates).

rL2 is a target for RNAylation by ModB

To discover additional RNAylation target proteins of ModB, a cell lysate, prepared from exponentially growing *E. coli*, was incubated with purified ModB and an NAD–10-mer RNA with a fluorescent 3′ Cy5 label (Fig. 5a and Extended Data Fig. 8c). We approximated the cellular conditions with respect to the presence of proteins, nucleic acids and various small molecules, including NAD[38].

Kinetic analysis of the ModB activity in these lysates showed that several *E. coli* proteins were RNAylated (Extended Data Figs. 8c and 9a), including rS1 (which migrates in a similar way to an RNAylated rS1 we added as a marker) and a protein with a mass of around 35 kDa. Notably, this pattern was not observed in the presence of 5′-monophosphorylated RNA–Cy5. We also characterized the simultaneous ADP-ribosylation in the same lysates showing different patterns of ADP-ribosylation targets and RNAylation targets of ModB (Extended Data Fig. 9b). In *E. coli*, NAD–RNA concentrations amount to around 5 μM (ref. [4]), compared with an approximately 700-fold excess of NAD (2.6 mM; ref. [38]). To simulate this molar excess of NAD over NAD–RNA in the lysate assay, we added NAD to our lysates. This showed that with a 700-fold excess of NAD, RNAylation still occurs with an efficiency of approximately 67 % (Extended Data Fig. 9c). We then assessed the intensity of ModB relative to *E. coli* proteins by proteomics, which revealed that a 100-fold dilution, relative to our standard assay conditions, may resemble relative ModB intensity during infection[39] (Extended Data Table 3). In lysates with ModB concentrations closer to those in cellular conditions, similar ADP-ribosylation and RNAylation patterns were observed as under standard conditions (Extended Data Fig. 9d).

These results indicate that in cellular conditions in which NAD is much more abundant than NAD–RNA, ModB RNAylates specific target proteins (Extended Data Figs. 8c and 9c). Because ModB was previously assumed to preferentially ADP-ribosylate proteins involved in translation[1], we monitored the RNAylation patterns of isolated *E. coli* ribosomes (Fig. 5a) and observed a similar pattern to that for the lysates (Extended Data Figs. 8c and 9).

To identify the RNAylated proteins, we RNAylated the *E. coli* ribosome with a 40-nucleotide-long NAD–RNA, resulting in a gel shift of RNAylated ribosomal proteins. MS analysis of the isolated gel band identified the ribosomal protein L2 (rL2) as a target for RNAylation by ModB (Extended Data Fig. 10a,b). rL2 is a protein with a mass of around 35 kDa and is probably the target observed in the lysates (Extended Data Figs. 8c and 9). It is evolutionarily highly conserved and is required for the association of the 30S and 50S subunits, involved in tRNA binding to both the A and P sites, and important for peptidyltransferase activity[40]. Similar to rS1, PNPase and RNase E, rL2 contains an RNA-binding domain that is homologous to the OB fold[41]. *In vitro* RNAylation assays

found that about 80 % of the rL2 was RNAylated by ModB in the presence of NAD–RNA (Extended Data Fig. 10c). *In vitro* RNAylation sites of rL2 were identified using the LC–MS/MS approach, including an MS data search with RNPxl, as described above. Trinucleotides (ADPr–C) were found to be attached to R217 and R221 (Extended Data Fig. 10d–g and Supplementary Table 6). R221 is located close (11 Å away) to H229, which is indispensable for ribosomal peptidyltransferase activity[40]. Future studies will reveal whether the RNAylation of rL2 and rS1 influences the translation efficiency of the ribosome (Fig. 5b).

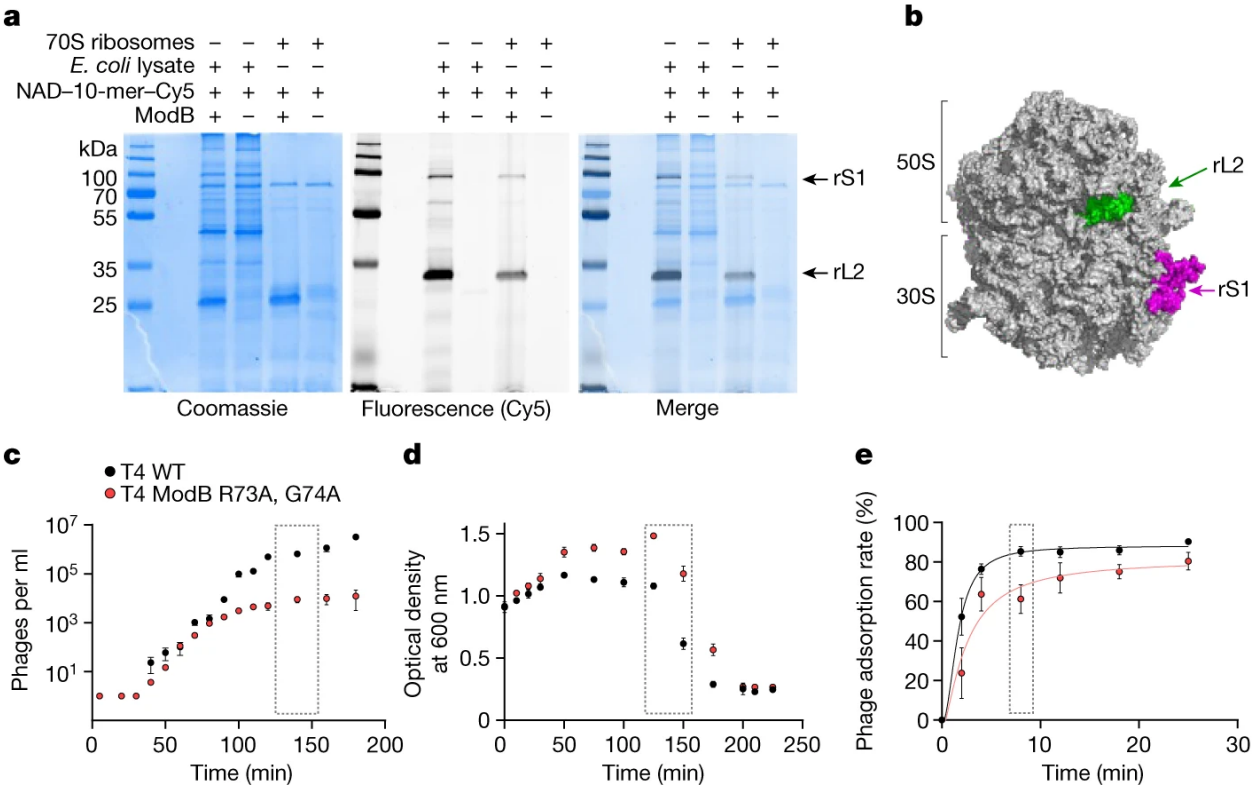


Fig. 5: RNAylation of the ribosome and phenotype of a ModB mutant T4 phage.

a) Characterization of ModB substrate specificity. RNAylation of two ribosomal proteins (rS1 and rL2) in cell lysates and 70S ribosome assemblies (n = 3). **b)** Illustration of the RNAylated proteins rS1 and rL2 in the context of the 70S ribosome, based on the cryo-electron microscopy structure of the hibernating 70S *E. coli* ribosome (PDB: 6H4N)[42]. **c–e)** Characterization of the T4 ModB R73A, G74A mutant phenotype, showing the burst size (c), *E. coli* lysis (d) and phage adsorption (e) of WT T4 phages and T4 ModB(R73A, G74A) (n = 3 biologically independent replicates for each). Data points with error bars represent mean ± s.d. Grey dotted boxes indicate time points used for assessing statistical significance in the case of burst size (c, 140 min after infection; two-sided Student's t-test, $P = 0.0015$ at $P_{signif} < 0.05$) and phage adsorption (e, 8 min after infection; t-test, two-sided, $P = 0.029$ at $P_{signif} < 0.05$) but indicate the delayed lysis without a statistical test in d. Statistical tests are shown in Supplementary Fig. 4.

ModB is important for phage infection

To investigate the functional role of ModB during phage infection, we compared the phenotypes of WT T4 and T4 ModB(R73A, G74A). We observed that the burst size (the number of virions released per infected *E. coli* cell) of T4 ModB(R73A, G74A) was decreased fourfold by 50 min after infection (15 ± 3 progeny per cell) compared with WT T4 (60 ± 32 progeny per cell) (Fig. 5c). By 140 min after infection, phages produced by WT T4 (6.6 × 105 ± 1.3 × 105 progeny) significantly exceeded the number of progeny from T4 ModB(R73A, G74A) (5.5 × 104 ± 3.1 × 104) (Fig. 5c and Supplementary Fig. 4a). At 140 min after infection, a 12-fold decrease in the progeny number compared with the WT T4 phage was observed for T4 ModB(R73A, G74A). Thus, ModB inactivation noticeably affects phage propagation properties.

We also observed a delay in lysis of approximately 20 min for the *E. coli* culture grown in the presence of the mutant phages (Fig. 5d). To determine whether ModB affects the infection cycle at the intra- or extracellular stage of infection, we measured the kinetics of phage adsorption to the cell (Fig. 5e). We observed a significantly lower adsorption rate for mutant phages. At 8 min after infection, around 61.3 ± 7.3 % of the T4 ModB(R73A, G74A) mutants successfully entered *E. coli*, compared with 85.3 ± 2.4 % for WT T4 phages (Fig. 5e and Supplementary Fig. 4b). These results indicate that phages are generated in the presence of inactivated ModB are less effective in the first stages of the infection, namely the attachment to, and penetration of, the host. This finding is consistent with the delayed host lysis.

6.4. Discussion

Most of the interactions between RNA and proteins are non-covalent[43], but there are some exceptions[44]. These include the peptidyl–tRNA intermediates in protein biosynthesis[45] (which are esters) and the adenoviral VPg proteins that form a phosphodiester bond (by means of a tyrosine OH group) with a nucleotide, which is then used to initiate transcription[46,47]. Here we show that an ART can attach NAD-capped RNAs to target proteins post-transcriptionally through the formation of glycosidic bonds. This finding represents a distinct biological function of the NAD cap on RNAs in bacteria, namely the activation of the RNA for enzymatic transfer to an acceptor protein. We discovered that the RNAylation of target proteins (a previously undescribed post-translational protein modification) has a role in the infection of the bacterium *E. coli* by bacteriophage T4. We discovered that ModB is a target-specific ART that RNAylates proteins that are part of the translational apparatus. We found that rS1 and rL2 are RNAylated at specific Arg residues in their RNA-binding regions. Moreover, we identified predominantly *E. coli* transcripts that are linked to rS1 during T4 phage infection. Inactivation of ModB caused a delay in bacterial lysis during phage

infection and decreased the number of progeny released. It remains unclear how the mutation of ModB (a non-capsid protein) will affect phage adsorption to the host cell. Precisely defining phage composition and architecture in future studies might help to explain this phenomenon.

Our findings introduce a molecular mechanism by which the T4 phage targets the translational machinery of its host and indicate that RNAylation might have a role in bacteriophage pathogenicity. It remains to be determined, however, whether ADP-ribosylation or RNAylation is the more important function of ModB. The T4 mutant ModB(R73A, G74A) abolished not only RNAylation but also ADP-ribosylation activity. This makes it difficult to determine whether the observed effects on T4 infection are due to RNAylation specifically or to the loss of ADP-ribosylation activity.

ModB was known to be an enzyme that uses NAD as a substrate to ADP-ribosylate host proteins during T4 infection. During this study, it became clear that ModB accepts not only NAD as a substrate, but also NAD–RNA. Enzymes typically have high specificity for their substrates and tolerate only limited chemical modifications. It was therefore surprising that ModB tolerates the attachment of a bulky RNA chain to the 3′ OH group of NAD (NAD–RNA) for the modification of a specific subset of target proteins. Remarkably, all four of the proteins (rS1, rL2, RNase E and PNPase) identified here as RNAylation targets of ModB are well known to interact with RNA. We therefore assume that both the ability of ModB to accept NAD–RNA as a substrate and the RNA affinity of the target protein determine RNAylation specificity. We did not succeed in generating a mutant of ModB that only ADP-ribosylates or RNAylates. RNAylation occurs by an ADP-ribosylation-like mechanism that involves the same catalytic residues as ADP-ribosylation, but the RNA affinity of the target protein might determine RNAylation specificity.

We considered why a phage ART would attach specific RNAs to proteins involved in translation. When a T4 phage infects E. coli it aims to reprogram the host ribosome to translate its mRNAs[48]. One way to achieve this may be a controlled shutdown of ribosomes that do not participate in the translation of T4 mRNAs. The discovery of crucial ribosomal proteins, rS1 and rL2, as RNAylation targets leads us speculate that RNAylation might impair their functionality, such as modulating peptidyltransferase activity. The fact that mostly E. coli transcripts are linked to rS1 *in vivo* suggests that undesired host gene-expression events are stopped by RNAylation. In this way, the phage might exploit RNAylation to inactivate distinct host ribosomes.

Future studies could show whether ribosomes that translate *E. coli* transcripts are blocked by RNAylation. This proposed mechanism would enable the phage to regulate the activity of the ribosome throughout the infection cycle and to stop the translation of host proteins.

Why only one of the three known T4 ARTs carries out efficient RNAylation is not understood. ModA and ModB both contain characteristic features of Arg-specific ARTs, such as the active-site motif R-S-EXE1. Differences in substrate specificity are therefore probably due to sequence differences (ModA and ModB are 25 % identical and have 47 % homologous amino acids)[1].

ARTs are not limited to phages. ADP-ribosylated proteins have been detected in hosts following infection by various viruses, including influenza, coronaviruses and HIV. As well as viruses using ARTs as weapons, the mammalian antiviral defence system uses host ARTs to inactivate viral proteins. Moreover, mammalian ARTs and poly-(ADPr) polymerases are regulators of critical cellular pathways and are known to interact with RNA[49]. Thus ARTs might catalyse RNAylation reactions in different organisms, making RNAylation a phenomenon of broad biological relevance. Finally, RNAylation may be considered as both a post-translational protein modification and a post-transcriptional RNA modification. Our findings challenge the established views of how RNAs and proteins interact with each other. The discovery of these previously undescribed RNA–protein conjugates comes at a time when the structural and functional boundaries between different classes of biopolymer are becoming increasingly blurred[50,51].

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request. NGS data are accessible via GEO record GSE214431. LC–MS/MS raw data for the measurements of rS1 ADP-ribosylation *in vivo*, in-gel digest and estimation of ModB abundance have been deposited in PRIDE with the accession code PXD041714. LC–MS/MS raw data for measurements of *in vitro* ADP-ribosylated and RNAylated rS1 and rL2 have been deposited in PRIDE with the accession code PXD038910. Reference genomes for E. coli (U00096.3) and T4 phage (NC_000866.4) were retrieved from NCBI. Protein structures (2MFI, 2MFL, 2KHI, 5XQ5, 2KHJ, 7K00 and 6H4N) were downloaded from PDB using the indicated accession code (https://www.rcsb.org/). *E. coli* K12 pan proteome (UP000000625) and selected protein sequences were retrieved from Uniprot (https://www.uniprot.org/). Supplementary information is available, including raw gel and blot images. Source data are provided with this paper.

Acknowledgements

We thank N. Beumer, J. Hoff, S. J. Keding, J. Kahnt, J. Koch, N. Lichti, P. Mann, N. Moskalchuk, M. Raabe, E. Tamerler, M. Viering and M. Weber for experimental assistance. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant 882789 RNACoenzyme, to A.J.) and from the German Research Council (DFG; project 439669440, TRR319, project A02, to A.J.). M.W.-S. is supported by the Studienstiftung des Deutschen Volkes and the Joachim Herz Stiftung. K.H. is supported by the Max Planck Society, Baden-Württemberg Stiftung, Carl-Zeiss-Stiftung and the German Research Council (grant DFG-SPP2330). H.U. is supported by the Max Planck Institute for Multidisciplinary Sciences and by the German Research Council (grants DFG-SPP1935, DFG-SFB1286 and DFG-SFB1565 (project number 469281184)).

6.5. Methods

General

Reagents were purchased from Sigma-Aldrich and used without further purification. Oligonucleotides, DNA and RNA were purchased from Integrated DNA Technologies (Supplementary Tables 7–10). Concentrations of DNA and RNA were determined by measurements using the NanoDrop ND-1000 spectrophotometer. Radioactively labelled proteins and nucleic acids were visualized using storage phosphor screens (GE Healthcare) and a Typhoon 9400 imager (GE Healthcare). Uncropped gel and blot images are provided (Supplementary Fig. 1).

Preparation of 5′ppp–RNA, 5′p–RNA and 5′-NAD–RNA by *in vitro* transcription

DNA templates for Qβ RNA (100-nucleotide RNA) and *E. coli* RNAI were amplified by PCR (primer sequences are listed in Supplementary Table 9), and PCR products were analysed by 2 % agarose gel electrophoresis and purified using the QIAquick PCR purification kit (QIAGEN). 5′-Triphosphate (ppp) Qβ RNA and RNAI were synthesized by *in vitro* transcription in the presence of 1× transcription buffer (40 mM Tris, pH 8.1, 1 mM spermidine, 10 mM $MgCl_2$, 0.01 % Triton X-100), 5 % DMSO, 10 mM DTT, 4 mM of each NTP, 20 µg T7 RNA polymerase (2 mg ml^{-1}, purified in our laboratory) and 200 nM DNA template. NAD–RNAI was made under similar conditions using 2 mM ATP and 4 mM NAD. The same conditions were applied for the synthesis of a mixture of α-^{32}P-labelled 5′-NAD and pppQβ RNAs, except we used 2 mM ATP, 80 µCi ^{32}P-α-ATP and 4 mM NAD instead of 4 mM ATP. The *in vitro* transcription reactions were incubated at 37 °C for 4 h and digested with DNase I (Roche). RNA was purified by denaturing PAGE, isopropanol-precipitated and resuspended in Millipore water. RNA sequences are listed in Supplementary Table 7.

To convert 5′ppp–RNAs into 5′-monophosphate–RNAs (5′p–RNAs), 250 pmol Qβ RNA was treated with 60 U RNA 5′-polyphosphatase (Epicentre) in 1× polyphosphatase reaction buffer at 37 °C for 70 min. Protein was removed from 5′p–RNAs by phenol–chloroform extraction and residual phenol–chloroform was removed by three rounds of diethyl ether extraction. 5′p–RNAs were isopropanol precipitated and resuspended in Millipore water.

5′-radiolabelling of 5′-monophosphate and NAD-capped RNAs

We treated 120 pmol 5′p-Qβ RNA or 6.25 nmol 5′p–RNA 8-mer (Supplementary Table 7) with 50 U T4 polynucleotide kinase in 1× reaction buffer B and 1,250 µCi ^{32}P-γ-ATP. The reaction was incubated at 37 °C for 2 h. The resulting 5′-^{32}P-RNA 8-mer and 5′-^{32}P-Qβ RNA were separated from residual protein by phenol–chloroform extraction. The remaining ^{32}P-γ-ATP was removed by

washing with three column volumes of Millipore water and centrifugation in 10 kDa (for Qβ RNA) or 3 kDa (for the 8-mer) Amicon filters (Merck Millipore) at 14,000 rpm at 4 °C four times. RNA sequences are listed in Supplementary Table 7. To convert the purified 5′-^{32}P-RNAs into 5′-^{32}P-NAD-capped RNA, 800 pmol 5′-^{32}P-RNA 8-mer or 30 pmol 5′-^{32}P-Qβ RNA was incubated in 50 mM $MgCl_2$ in the presence of a spatula tip of nicotinamide mononucleotide phosphorimidazolide, synthesized as described[52], at 50 °C for 2 h. RNAs were purified by washing with Millipore water and centrifugation in 10 kDa (Qβ RNAs) or 3 kDa (8-mer) Amicon filters at 14,000 rpm at 4 °C four times. The concentrations of the 5′-^{32}P-RNAs were measured using a NanoDrop ND-1000 spectrophotometer and were used to calculate the approximate concentrations of yielded 5′-NAD-capped ^{32}P-RNAs, assuming an approximate yield of the imidazolide reaction of 50 % (ref. [52]). The 5′-^{32}P-ADPr–RNA 8-mer was synthesized by incubating 8 µM 5′-^{32}P-NAD–RNA 8-mer and 0.08 µM ADP-ribosyl cyclase CD38 (R&D Systems) in 1× degradation buffer at 37 °C for 4 h. The reaction was purified by P/C/I-diethyl ether extraction and filtration through 3 kDa filters and washing with four column volumes of Millipore water.

Cloning of ADP-ribosyltransferases, ADP-ribose hydrolases and target proteins

To amplify bacteriophage T4 genes modA (GeneID: 1258568; Uniprot: P39421), modB (GeneID: 1258688; Uniprot: P39423) and alt (GeneID: 1258760; Uniprot: P12726), a single plaque from bacteriophage T4 revitalization was resuspended in Millipore water and used in a ‘plaque’ PCR, analogous to bacterial-colony PCR. The gene encoding the ADP-ribosylhydrolase ARH1 (GeneID: 141; Uniprot: P54922) was purchased from IDT as gBlocks and amplified by PCR. E. coli genes coding for rS1 (GeneID: 75205313; Uniprot: P0AG67), rL2 (GeneID: 947820; Uniprot: P60422) and PNPase (GeneID: 947672; Uniprot: P05055) were PCR-amplified from genomic DNA of *E. coli* K12, which was isolated using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Nucleotide sequences are listed in Supplementary Table 8. XhoI and NcoI restriction sites were introduced during amplification using appropriate primers (Supplementary Table 9). The resulting PCR product was digested with XhoI and NcoI (Thermo Fisher Scientific) and cloned into the pET–28c vector (Merck Millipore). After Sanger sequencing, the resulting plasmids were transformed into *E. coli* One Shot BL21 (DE3) (Life Technologies). The ARH1 D55,56A, ModB(R73A) and rS1 mutants were generated by site-directed mutagenesis using a procedure based on the Phusion Site-Directed Mutagenesis Kit (Thermo Scientific). The resulting plasmids were sequenced and transformed into *E. coli* One Shot BL21 (DE3). All strains used and generated in this work are summarized in Supplementary Table 10.

Purification of rS1, rS1 domains and variants, rL2, the PNPase S1 domain, RNase E(1–529), Alt, NudC, NudC*(V157A, E174A, E177A, E178A) and NudC(E178Q)

Isopropyl beta-d-thiogalactoside (IPTG)-induced *E. coli* One Shot BL21 (DE3) containing the respective plasmid (Supplementary Table 10) was cultured in LB medium at 37 °C. Protein expression was induced at an optical density at 600 nm (OD600) of 0.8, bacteria were collected after centrifugation for 3 h at 37 °C and lysed by sonication (30 s at 50 % power, five times) in HisTrap buffer A (50 mM Tris-HCl, pH 7.8, 1 M NaCl, 1 M urea, 5 mM $MgSO_4$, 5 mM β-mercaptoethanol, 5% glycerol, 5 mM imidazole, one tablet per 500 ml complete EDTA-free protease inhibitor cocktail (Roche)). The lysate was cleared by centrifugation (37,500 g for 30 min at 4 °C) and the supernatant was applied to a 1 ml Ni-NTA HisTrap column (GE Healthcare). The protein was eluted with an imidazole gradient using an analogous gradient of HisTrap buffer B (HisTrap buffer A with 500 mM imidazole added) and analysed by SDS–PAGE.

Further protein purification was achieved by size-exclusion chromatography (SEC) through a Superdex 200 10/300 GL column (GE Healthcare) using SEC buffer containing 0.5 M NaCl and 25 mM Tris-HCl, pH 8. Fractions of interest were analysed by SDS–PAGE, pooled and concentrated in Amicon Ultra-4 centrifugal filters (molecular weight cut-off (MWCO) 10 kDa with centrifugation at 2,000 rpm and 4 °C). Protein concentration was measured with a NanoDrop ND-1000 spectrophotometer. Finally, proteins were stored in SEC buffer supplemented with 50 % glycerol at −20 °C.

Purification of ARH1 and ARH1(D55A, D56A)

E. coli BL21 DE3 pET28-ARH1 and BL21-pET28-ARH1 D55A, D56A (Supplementary Table 10) were grown to an OD600 = 0.6 at 37 °C and 175 rpm. Afterwards, bacteria were allowed to cool to room temperature for 30 min. Expression was induced with 1 mM IPTG, and bacteria were finally grown overnight at room temperature while shaking at 175 rpm. Bacteria were collected by centrifugation and proteins were purified in a similar way to rS1 variants.

Purification of ModA

E. coli BL21 DE3 pET28-ModA (Supplementary Table 10) was grown to an OD600 = 1 at 37 °C with shaking at 175 rpm. Protein expression was induced with 0.5 mM IPTG and bacteria were collected by centrifugation after 3 h at 37 °C. Pelleted bacteria were resuspended in 50 mM NaH_2PO_4, pH 8, 300 mM NaCl, 1 mM DTT with one tablet per 500 ml complete EDTA-free protease inhibitor cocktail (Roche) and lysed by sonication (3 × 1 min at 5 % power). Lysates were centrifuged at 3,000 g at 4 °C for 20 min. Sediments were washed by resuspension in 30 ml 50 mM

Tris-HCl, pH 7.5, 2 mM EDTA, 100 mM NaCl, 1 M urea, 1 mM DTT and one tablet EDTA-free protease inhibitor (Roche), and centrifuged at 10,000 g at 4 °C for 20 min. Pellets containing inclusion bodies were resuspended in 40 ml 100 mM Tris, pH 11.6, 8 M urea, transferred to 12–14 kDa MWCO dialysis bags (Roth) and dialysed overnight against 50 mM NaH2PO4, 300 mM NaCl. Protein solutions were centrifuged at 20,000g at 4 °C for 30 min. Supernatants were batch purified using disposable 10 ml columns (Thermo Fisher Scientific) packed with 2 ml Ni-NTA agarose (Jena Bioscience) and equilibrated with 10 column volumes of 50 mM NaH_2PO_4 (pH 8), 300 mM NaCl. Proteins were purified by washing the columns with 30 column volumes of 50 mM NaH_2PO_4, 300 mM NaCl, 15 mM imidazole, eluted with 5 ml 50 mM NaH2PO4, 300 mM NaCl, 300 mM imidazole and concentrated in Amicon (Merck Millipore) filters (MWCO 10 kDa with centrifugation at 2,000 rpm and 4 °C). Finally proteins were purified by SEC, as described for rS1.

Purification of ModB and ModB(R73A, G74A)

E. coli BL21 DE3 pET28–ModB and *E. coli* BL21 DE3 pET28–ModB(R73A, G74A) (Supplementary Table 10) were grown to OD600 = 2.0 at 37 °C with shaking at 185 rpm and cooled to 4 °C while being shaken at 160 rpm for at least 30 min. Protein expression was induced by the addition of 1 mM IPTG. The cultures were then incubated for 120 min at 4 °C, with shaking at 160 rpm and bacteria were collected by centrifugation (4,000 rpm at 4 °C for 25 min). The ModB protein was purified from the supernatant as described for rS1 variants.

Alphafold prediction of ModB structure

The Alphafold prediction of ModB structure was performed with AlphaFold2.ipynb (v.1.3.0, https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) with default parameters (use_templates = false, use_amber = false; msa_mode = MMseqs2 (UniRef+Environmental), model_type = “AlphaFold2-ptm”, max_msa = null, pair_mode = unpaired+paired, auto advanced settings). The ModB protein sequence was retrieved from Uniprot (primary accession: P39423). The ModB structure prediction model from rank_1 was further assessed using PyMol.

In vitro ADP-ribosylation and RNAylation of rS1 and rL2 with ^{32}P-labelled NAD, NAD–8-mer, NAD–Qβ RNA or NAD–10-mer–Cy5

rS1 (0.3 μM) was ADP-ribosylated in the presence of 0.25 μCi μl−1 ^{32}P-NAD or RNAylated in the presence of one of 0.6 μM ^{32}P-NAD–8-mer, 0.03 μM ^{32}P-NAD–Qβ RNA or 0.8 μM NAD–10-mer–Cy5 (Supplementary Table 7) by 1.4 μM ModB and in 1× transferase buffer (10 mM $Mg(OAc)_2$, 22 mM NH_4Cl, 50 mM Tris-acetate pH 7.5, 1 mM EDTA, 10 mM β-mercaptoethanol and 1 %

glycerol) at 15 °C for at least 120 min. Samples (5 µl) were taken before the addition of ModB and after 1, 2, 5, 10, 30, 60 and 120 min, and mixed with 5 µl 2× Laemmli buffer to stop the reaction. Reactions were assessed by 12 % SDS–PAGE and gels were stained in Instant Blue solution (Sigma-Aldrich) for 10 min. Radioactive signals were visualized using storage phosphor screens and a Typhoon 9400 imager. The intensity of the radioactive bands was quantified using ImageQuant 5.2 (GE Healthcare). The RNAylation with NAD-capped Cy5-labelled RNA was visualized with the ChemiDoc (Bio-Rad) Cy5 channel. Gels were then stained by Coomassie solution and imaged using the same system. In some cases, stain-free imaging of proteins in SDS gels was performed by 2,2,2-trichloroethanol (TCE) incorporated in the gel. TCE binds to tryptophan residues of the proteins, which enhances their fluorescence under ultraviolet light and thereby enables their detection[53].

rL2 was ADP-ribosylated or RNAylated at the same settings using either 6.4 µM NAD or 6.4 µM NAD–8-mer as a substrate to modify 4.6 µM rL2 in the presence of 1.57 µM ModB for 4 h for LC–MS/MS measurements. For shift assays, 538 nM rL2 was RNAylated by 2.61 µM ModB in the presence of 6 µM NAD–8-mer. 12 % SDS– PA gels were fixed with a solution of 40 % ethanol and 10 % acetic acid overnight and stained using Flamingo fluorescent protein dye (Bio-Rad) for up to 6 h and imaged with the ChemiDoc (Bio-Rad). Signal intensity was quantified in ImageLab (Bio-Rad). Where indicated, statistical tests were performed using two-sided t-tests in R (v.4.2.2) implemented in the ggpubr package (v.0.6.0) using a significance level of 0.05.

In vitro RNAylation of *E. coli* RNA polymerase with NAD–10-mer–Cy5

We incubated 0.8 µM NAD–10-mer–Cy5 (Supplementary Table 7) with 0.5 µM of protein *E. coli* RNA polymerase (New England Biolabs) and 3 µM Alt or ModA in the presence of 1 × transferase buffer at 15 °C for 60 min. Samples were taken before the addition of Alt or ModA and after 60 min incubation. The reactions were stopped by the addition of 1 volume of 2 × Laemmli buffer. Reactions were analysed by 10 % SDS–PAGE with rS1 RNAylated by ModB with NAD–10-mer–Cy5 as a reference protein. RNAylated proteins were visualized using the ChemiDoc (Bio-Rad) Cy5 channel. Afterwards, gels were stained in Coomassie solution and imaged using the same system.

Analysis of protein rS1 self-RNAylation

In 20-µl reactions, 3.6 µM ^{32}P-ADPr–8-mer (Supplementary Table 7) was incubated with either 2.6 µM rS1, 3.9 µM ModB or both 2.59 µM rS1 and 3.9 µM ModB in 1 × transferase buffer. As a positive control, equal amounts of protein rS1 and ModB were incubated with 0.6 µM ^{32}P-NAD–8-

mer. All reactions were incubated at 15 °C for 60 min. Samples were taken before the addition of ModB or after 60 min, and reactions were stopped by adding one volume of 2 × Laemmli buffer. Reactions were analysed by 12 % SDS–PAGE and autoradiography imaging.

RNAylation of protein rS1 with Qβ RNA (100-nucleotide–RNA) and specificity for the 5′-NAD cap

0.05 µM ^{32}P-NAD–Qβ RNA, 0.15 µM 5′-^{32}P-Qβ RNA or 0.15 µM 5′-^{32}PPP-Qβ RNA (Supplementary Table 7) was incubated with 2.3 µM rS1 and 1.4 µM ModB in the presence of 1 × transferase buffer at 15 °C for 60 min. Samples were taken before the addition of ModB and after 60 min, and reactions were stopped by adding 1 volume 2 × Laemmli buffer. Reactions were analysed by 10 % SDS–PAGE, applying rS1–^{32}P-ADPr in 1 × Laemmli buffer as a reference, and subsequent autoradiography imaging.

Preparation of RNAylated and ADP-ribosylated rS1 for enzymatic treatments

ADP-ribosylation or RNAylation reactions were performed with radio-labelled substrates, washed and equilibrated in 1 × transferase or 1 × degradation buffer for further enzymatic treatments. The reactions were washed with four column volumes of the corresponding buffer by centrifugation at 10,000 g at 4 °C in 10 kDa Amicon (Merck Millipore) filters. Proteins RNAylated with Cy5-labelled RNA were equilibrated in the same buffers using Zeba Spin desalting columns (7 kDa MWCO, 0.5 ml) (Thermo Fisher Scientific) according to the manufacturer's instructions.

Nuclease P1 digest of protein rS1 RNAylated with 100-nucleotide–RNA (rS1-100-nucleotide–RNA)

An rS1–100-nucleotide-RNA (^{32}P) mixture (19 µl) was equilibrated in 1× transferase buffer and incubated with either 1 µl nuclease P1 or 1 µl Millipore water at 37 °C for 60 min. Samples were taken at the beginning and after 60 min, and reactions were stopped by adding one volume of 2 × Laemmli buffer. Reactions were analysed by 10 % SDS–PAGE, applying rS1–^{32}P-ADPr in 1 × Laemmli buffer as a reference, and subsequent autoradiography imaging.

Tryptic digest of ^{32}P-labelled rS1–8-mer and rS1–ADPr

Mixtures (19 µl) of both rS1 and rS1–8-mer (^{32}P) and of rS1 and rS1–ADPr (^{32}P) in 1 × degradation buffer were incubated with either 0.2 µg Trypsin (Sigma, EMS0004, mass-spectrometry grade) or Millipore water as a negative control at 37 °C. Samples were taken before the addition of Trypsin/Millipore water and after 120 min. Reactions were stopped by adding one volume 2 × Laemmli buffer to samples and were analysed by 12 % SDS–PAGE and autoradiography imaging.

Chemical removal of ADP-ribosylation and RNAylation *in vitro*

Aliquots from washed and equilibrated ADP-ribosylated (1 µl) and RNAylated (2 µl) (^{32}P) rS1 were treated with either 10 mM $HgCl_2$ or 500 mM NH_2OH (refs. [54,55]) at 37 °C for 1 h. Reactions were stopped by adding 2 × Laemmli buffer and analysed by 12 % SDS–PAGE.

Enzymatic removal of ADP-ribosylation and RNAylation *in vitro*

Aliquots from washed and equilibrated (in 1× degradation buffer) ADP-ribosylated (1 µl) and RNAylated (2 µl) rS1 (^{32}P) were treated with 0.5 U endonuclease P1 (Sigma-Aldrich)[56] or 0.95 µM ARH1 or ARH3 (human recombinant, Enzo Life Science)[57] in the presence of 10 mM $Mg(OAc)_2$, 22 mM NH_4Cl, 50 mM HEPES, 1 mM EDTA, 10 mM β-mercaptoethanol and 1 % (v/v) glycerol in a total volume of 20 µl at 37 °C for 1 h. Enzymatic reactions were stopped by adding 2 × Laemmli buffer and analysed by 12 % SDS–PAGE.

Inhibition of RNAylation and ADP-ribosylation with 3-methoxybenzamide

Reactions (20 µl) of 1.4 µM ModB and 2.3 µM protein rS1 with either 1 µM ^{32}P-NAD–8-mer or 3 µM 5′-^{32}P–8-mer (Supplementary Table 7) were incubated in the presence of 2 mM 3-MB (50 mM stock in DMSO) or the absence of the inhibitor (DMSO only) at 15 °C (ref. [58]). Samples were taken before the addition of ModB and after 60 min. Reactions were stopped by the addition of 1 volume 2× Laemmli buffer and analysed by 12 % SDS–PAGE.

Effect of RNA secondary structure on RNAylation efficiency

We incubated 1.1 µM NAD–RNA–Cy5 (linear, 5′ overhang, 3′ overhang and blunt ends; Supplementary Table 7) with 0.9 µM rS1 and 0.4 µM ModB in 1× transferase buffer. Samples of 5 µl were taken before the addition of ModB protein and 2, 5, 10, 30, 60 and 120 min after the start of the reaction. The samples were directly mixed with one volume of 2 × Laemmli buffer to stop the reaction. The conversion of the substrates was analysed by 12% SDS–PAGE, following visualization on ChemiDoc (Bio-Rad) in the Cy5 channel. The maximum observed signal intensity of RNAylated rS1 protein was used to determine the relative conversion for each of the analysed substrates at distinct time points.

Culture of the *E. coli* B strain and infection with T4 phages

Precultures of *E. coli* B strain pTAC-rS1 (Supplementary Table 10) were incubated in LB medium with 100 µg ml^{-1} ampicillin at 37 °C and 185 rpm overnight. For the main cultures, 150 ml LB medium with 100 µg ml^{-1} ampicillin were inoculated with preculture to an OD600 = 0.1. At OD600 = 0.4, protein expression was induced by the addition of 1 mM IPTG. At OD600 = 0.8,

cultures were either infected with bacteriophage T4 at a multiplicity of infection (MOI) of 10 (20 ml phage solution) (DSM 4505, Leibniz Institute DSMZ) or not infected by adding 20 ml LB medium instead (negative control). Cultures were incubated for 20 min at 37 °C with shaking at 240 rpm. Bacteria were collected by centrifugation at 4,000 g at room temperature for 15 min. Pellets were stored at −80 °C.

Purification of His-tagged rS1 from infected *E. coli* strain B pTAC-rS1

Bacterial pellets were resuspended in 10 ml buffer A and lysed via sonication (1 × 5 min, cycle 2 at 50 % power). Lysates were centrifuged at 37,500 g at 4 °C for 30 min. The supernatant was filtered through 0.45-µm filters (Sarstedt). rS1 from bacteriophage T4-infected or non-infected E. coli B strain was purified from the supernatant by gravity Ni-NTA affinity chromatography. Ni-NTA agarose slurry (1 ml, Thermo Fisher Scientific) was added to a 10 ml propylene column and equilibrated in buffer A. The supernatant was loaded onto the column twice. The column was washed with a mixture of 95 % buffer A and 5 % buffer B containing 29.75 mM imidazole. Protein was eluted from the column with 10 ml buffer B.

His-tagged-protein rS1 from T4-infected or uninfected *E. coli* B strain pTAC-rS1 was washed with two filter volumes of 1× degradation buffer (12.5 mM Tris-HCl, pH 7.5, 25 mM NaCl, 25 mM KCl, 5 mM $MgCl_2$) by centrifugation in 10-kDa Amicon filters at 5,000 g at 4 °C and concentrated to a final volume of 120 µl. The fractions were analysed by 12 % SDS–PAGE analysis and the gel was stained in Instant Blue solution for 10 min and imaged immediately.

Purification of His-tagged rS1 and rL2 for LC–MS/MS analysis

E. coli B strain with endogenously His-tagged rS1 and *E. coli* B strain expressing His-tagged rS1 WT, R139A or R139K were infected with T4 to an MOI of 5.0, as described above for 8 min. 100 ml culture was collected and the pellet resuspended in 1.5 ml Ni-NTA buffer A with 15 mM imidazole (50 mM Tris-HCl, pH 7.8, 1 M NaCl, 1 M urea, 5 mM $MgSO_4$, 5 mM β-mercaptoethanol, 5 % glycerol, 15 mM imidazole, one tablet per 500 ml complete EDTA-free protease inhibitor cocktail (Roche)). Cells were lysed by sonication (three times for 2 min at 80 % power) and supernatant was cleared by centrifugation at 17,000 g at 4 °C for 30 min. The supernatant was incubated with 75 µl Ni-NTA magnetic beads (Jena Bioscience) equilibrated in Ni-NTA buffer A with 15 mM imidazole for 1 h at 4 °C. Magnetic beads were washed seven times with 1 ml Ni-NTA buffer A with 15 mM imidazole and three times with Ni-NTA buffer without imidazole but with 4 M urea. Finally, protein was eluted by addition of Ni-NTA elution buffer (50 mM Tris-HCl, pH 7.8, 1 M NaCl, 1 M Urea, 5 mM $MgSO_4$, 5 mM β-mercaptoethanol, 5 % glycerol, 300 mM imidazole, one tablet per

500 ml complete EDTA-free protease inhibitor cocktail (Roche)). Protein was equilibrated in 1× transferase buffer with Zeba columns (7 kDa MWCO, 0.5 ml) according to the manufacturer's instructions, and protein was digested with trypsin in a 1:20 ratio (w/w) at 37 °C for 3 h. Peptides were C18-purified using 50 mM triethylamine-acetate (pH 7.0) buffer in combination with 0–90 % acetonitrile and Chromabond C18 WP spin columns (20 mg, Macherey Nagel). Purified peptides were dissolved in HPLC-grade H2O and subjected to LC–MS/MS analysis (see below).

In vitro RNAylated rS1 (D2) reactions in 1 × transferase buffer were directly digested (without further purification) with 1 µg RNase A (Thermo Fisher Scientific) and 100 U RNase T1 (Thermo Fisher Scientific) at 37 °C for 1 h, following tryptic digest at 37 °C for 3 h in the same buffer with trypsin (Promega) in a 1:30 ratio (w/w) relative to the total protein content per sample. Peptides were purified with Chromabond C18 WP spin columns as described above and used for LC–MS/MS analysis (see below).

In vitro RNAylation reactions of rL2 with NAD–8-mer and ADP-ribosylation reactions were purified at similar settings to the proteins from T4 phage-infected *E. coli*. Here, reactions (200 µl) were incubated with 100 µl Ni-NTA beads equilibrated in 800 µl Ni-NTA buffer A with 10 mM imidazole and 40 U murine RNase inhibitor (New England Biolabs) at 4 °C for 1 h. Beads were washed eight times with 1 ml streptavidin wash buffer (50 mM Tris-HCl, pH 7.4, 8 M urea) at room temperature and protein was eluted with 130 µl Ni-NTA elution buffer. Purified proteins were rebuffered in 100 mM NH4OAc using Zeba spin desalting columns (7 kDa MWCO, 0.5 ml) according to the manufacturer's instructions. rL2 samples were dissolved in 4 M urea in 50 mM Tris-HCl (pH 7.5) and incubated for 30 min at room temperature, followed by dilution to 1 M urea with 50 mM Tris-HCl (pH 7.5). 10 µg RNase A (Thermo Fisher Scientific) and 1 kU RNase T1 (Thermo Fisher Scientific) were added, following incubation for 4 h at 37 °C. For protein digestion, 0.5 µg trypsin (Promega) was added to each sample and digestion was performed overnight at 37 °C. Samples were adjusted to 1 % acetonitrile (ACN) and to pH 3 using formic acid. Samples were cleaned up using C18 columns (Harvard Apparatus) according to the manufacturer's instructions.

LC–MS/MS analysis of His-tagged, *in vitro* RNAylated rS1 and rL2

Cleaned-up rS1 and rL2 peptide samples were dissolved in 2 % ACN, 0.05 % trifluoroacetic acid and subjected to LC–MS/MS analysis using an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex Ultimate 3000 RSLCnano system. Peptides were loaded on a Pepmap 300 C18 trap column (Thermo Fisher Scientific) (flow rate, 10 µl min^{-1}) in buffer A (0.1 % (v/v) formic acid) and washed for 3 min with buffer A. Peptide separation was performed on an in-house-packed C18 column (30 cm; ReproSil-Pur 120 Å, 1.9 µm, C18 AQ; inner diameter,

75 µm; flow rate 300 nl min^{-1}) by applying a linear gradient of buffer B (80 % (v/v) ACN, 0.08 % (v/v) formic acid). The main column was equilibrated with 5 % buffer B for 18 s, the sample was applied and the column was washed for 3 min with 5 % buffer B.

A linear gradient of 10–45 % buffer B over 44 min was applied to elute peptides, followed by 4.8 min washing at 90 % buffer B and 6 min at 5 % buffer B. Eluting rS1 and rL2 peptides were analysed for 58 min in positive mode using a data-dependent top-20 acquisition method. The resolution for MS1 and MS2 were set to 120,000 and 30,000 full-width at half-maximum, respectively, and automatic gain control (AGC) targets were set to 106 (MS1) and 105 (MS2). The MS1 scan range was set to m/z = 350–1,600. Precursors were fragmented using 28 % normalized, higher-energy collision-induced dissociation fragmentation. Other analysis parameters were set as follows: isolation width, 1.6 m/z; dynamic exclusion, 9 s; maximum injection times for MS1 and MS2, 60 ms and 120 ms, respectively.

For all measurements, the lock mass option (m/z 445.120025) was used for internal calibration.

Analysis of *in vitro* RNAylated rS1 and rL2 MS data

MS data were analysed and validated manually using the OpenMS pipeline RNPxl and OpenMS TOPPASViewer30. Precursor mass tolerance was set to 6 ppm. MS/MS mass tolerance was set to 20 ppm. A neutral loss of 42.021798 Da ($C1H_2N_2$) at Arg residues was defined, as well as adducts of ribose minus H_2O (78.010565 Da, C_5H_2O), ADP-ribose (541.06111 Da, $C_{15}H_{21}N_5O_{13}P_2$) and ADPr without adenine (485.97295 Da; $C_{10}H_{17}O_{16}P_3$)[32]. Results were filtered for a 1 % false discovery rate on peptide spectrum match level. Ion chromatograms for rS1 peptides were extracted and visualized using Skyline (v.21.2.0.369)[59].

LC–MS/MS analysis of His-tagged rS1 isolated from T4-phage-infected *E. coli*

LC–MS/MS analysis of protein digests was performed on an Exploris 480 mass spectrometer connected to an electrospray ion source (Thermo Fisher Scientific). Peptide separation was done using the Ultimate 3000 nanoLC-system (Thermo Fisher Scientific), equipped with a packed-in-house C18 resin column (Magic C18 AQ 2.4 µm, Dr. Maisch). The peptides were eluted from a precolumn in backflush mode with a gradient from 98% solvent A (0.15% formic acid) and 2% solvent B (99.85 % ACN, 0.15 % formic acid) to 35 % solvent B over 40 min and 90 min, respectively. The flow rate was set to 300 nl min^{-1}. The data-dependent acquisition mode for label-free quantification was set to obtain one high-resolution MS scan at a resolution of 60,000 (m/z of 200) with scanning range from 350 to 1,650 m/z. MS/MS scans were acquired for the 20 most-intense ions (90 min gradient) and for the most-intense ions detected within 2 s (cycle 1 s, 40 min

gradient). To increase the efficiency of MS/MS attempts, the charged-state screening mode was adjusted to exclude unassigned and singly charged ions. The ion accumulation time was set to 25 ms for MS and 'auto' for MS/MS scans. The AGC was set to 300 % for MS survey scans and 200 % for MS/MS scans.

Raw MS spectra were analysed using MaxQuant (v.1.6.17.0 and 2.0.3.0) using a fasta database of the targets proteins and a set of common contaminant proteins. The following search parameters were used: full tryptic specificity required (cleavage after Lys or Arg residues); three missed cleavages allowed; carbamidomethylation (C) set as a fixed modification; and oxidation (M; +16 Da), deamidation (N, Q; +1 Da) and ADP-ribosylation (K; +541 Da) set as variable modifications. MaxQuant was executed in the default setting. All MaxQuant parameters are listed in Supplementary Tables 1 and 2. The MS proteomics data have been deposited with the ProteomeXchange Consortium by the PRIDE partner repository under the dataset identifier PXD041714.

Generation of *E. coli* B strain with endogenously His-tagged rS1

The *E. coli* B strain with endogenously His-tagged rS1 was created by homologous recombination of linear transforming DNA (tDNA) using the pRET/ET plasmid in the *E. coli* B strain. The linear tDNA was generated by fusion PCR aligning four fragments: 156 base pairs (bp) of the rpsA gene with an additional His-tag amplified from the pET28 rS1 vector (serving as the left homologous flank), a 70-bp fragment of the native rpsA terminator, the Flp-flanked kanamycin cassette from pKD4 and 140 bp of the 3′ flanking region of the rpsA gene (the right homologous flank). The primers used are indicated in Supplementary Table 9. The subsequent procedure for recombination is based on the protocol for the *E. coli* Gene Deletion Kit by RET/ET Recombination (Gene Bridges). In brief, *E. coli* B strain containing the pRED/ET plasmid was grown in LB medium supplemented with 100 μg ml^{-1} ampicillin at 30 °C. At OD600 = 0.35, L-arabinose was added to 0.33 % (w/v) to induce expression of the RED/ET recombination system during growth at 37 °C for 1 h. Next, 1.4 ml culture was collected by centrifugation at 3,000 g at 4 °C for 1 min, and cells were washed twice with 1 ml cold 10% glycerol and finally resuspended in 50 μl 10 % glycerol. Cells were electroporated with 1 μg tDNA using a MicroPulser Electroporator (Bio-Rad) at standard settings (Ec1). Electroporated cells were immediately resuspended in 1 ml prewarmed LB medium and incubated at 37 °C with shaking at 600 rpm for 3 h. Finally, cells were plated on kanamycin (30 μg ml^{-1}) LB–agar plates. Cells took 2 days to recover and grow. Successful recombination was evaluated by Sanger sequencing and correct protein expression was validated by pull-down and proteomics.

RNAylomeSeq

Cultures (100 ml) of *E. coli* B strain with endogenously His-tagged rS1 (Supplementary Table 10) in LB medium supplied with 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 30 μg ml^{-1} kanamycin were grown at 37 °C in 250 ml baffled Erlenmeyer flasks to an OD600 of around 0.8. T4 phage WT or T4 phage ModB(R73A, G74A) were added to an MOI of 5.0. For the uninfected negative control, the same volumes of LB medium were added to the cultures. Cultures were then incubated at 37 °C for 8 min and *E. coli* was collected by centrifugation at 3,000 g for 13 min. Dried pellets were stored at −80 °C.

Pellets from the 100 ml culture infected with either WT T4 phage, T4 phage ModB(R73A, G74A) or the uninfected control (LB) were resuspended in 2 ml Ni-NTA wash buffer (10 mM imidazole, 50 mM Tris-HCl, pH 7.5, 1 M NaCl, 1 M urea, 5 mM $MgSO_4$, 5 mM β-mercaptoethanol, 5 % glycerol, pH 8.0, EDTA-free protease inhibitor (Roche, one tablet per 500 ml)) on ice and lysed by sonication (6 min, 50 % power, 0.5 s pulse). The lysate was cleared from the cell debris by centrifugation at 21,000 g at 4 °C for 30 min. Supernatant (1.9 ml), 50 μl Ni-NTA agarose beads (Jena Bioscience, equilibrated in Ni-NTA wash buffer), 80 U murine RNase inhibitor (New England Biolabs) and 4.72 μg rS1 D2 RNAylated with NAD-capped RNAI were combined and incubated at 4 °C in a rotary mixer for 30 min. Entire samples were transferred to Mobicol mini spin columns (MoBiTec). Beads were washed four times with 200 μl Ni-NTA wash buffer and subsequently eight times with 200 μl streptavidin wash buffer (50 mM Tris-HCl, pH 7.5, 8 M urea). Beads were equilibrated in standard ligation buffer (10 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.4) and blocked with bovine serum albumine (BSA) before 3′ RNA-adapter ligation at 4 °C overnight in the presence of standard ligation buffer, 50 mM β-mercaptoethanol, 0.05 μg μl−1 BSA, 15 % (v/v) DMSO, 5 μM adenylated RNA-3′-adapter, 0.5 U μl^{-1} T4 RNL1 (New England Biolabs) and 10 U μl^{-1} T4 RNL2, tr. K227Q (New England Biolabs). Protein was rebound by the addition of NaCl to 1.5 M and incubation at 20 °C, with shaking at 400 rpm for 20 min. Beads were subsequently washed six times with streptavidin wash buffer and equilibrated in first strand buffer (50 mM Tris-HCl, pH 8.3, 3 mM $MgCl_2$, 75 mM KCl) and blocked with BSA. Reverse transcription of protein-bound RNA was done in a 30-μl scale for 1 h at 40 °C using 10 U μl^{-1} Superscript IV Reverse Transcriptase (Invitrogen) in the presence of 5 μM RT primer, first strand buffer, 25 mM β-mercaptoethanol, 0.05 μg μl^{-1} BSA and 0.5 mM dNTPs. After incubation, NaCl was added to 1.5 M and the solution was incubated at 20 °C, with shaking at 400 rpm for 1 h to rebind RNA–cDNA hybrids. Beads were subsequently washed five times with 0.25 × streptavidin wash buffer (2 M urea, 50 mM Tris-HCl, pH 7.5), equilibrated in ExoI buffer (10 mM Tris-HCl, pH 7.9, 5 mM β-mercaptoethanol, 10 mM $MgCl_2$, 50 mM NaCl) and blocked with BSA. Residual RT primer was removed by ExoI digest with

1 U µl^{-1} *E. coli* ExoI (New England Biolabs) in ExoI buffer at 37 °C for at least 30 min. Finally, beads were washed with 200 µl 0.25 × streptavidin wash buffer five times and subsequently with 200 µl immobilization buffer (10 mM Na-HEPES, pH 7.2, 1 M NaCl) three times. cDNA was eluted by incubation of beads in 100 µl 150 mM NaOH at 55 °C for 25 min and by washing with 100 µl MQ water. Eluate pH was neutralized by the addition of 0.05 volumes of 3 M NaOAc, pH 5.5. cDNA was removed from the residual protein by phenol–chloroform extraction and precipitated with 2.5 volumes of ethanol in the presence of 0.3 M NaOAc, pH 5.5 overnight. Precipitated cDNA was C-tailed using 1 U µl−1 TdT (Thermo Fisher) in the presence of 1.25 mM CTP and 1× TdT buffer at 37 °C for 30 min and subsequently inactivated at 70 °C for 10 min. 5 µM cDNA anchor (hybridization of forward and reverse anchor, Supplementary Table 9) was ligated to C-tailed cDNA in standard ligation buffer in the presence of 10 µM ATP and 1.5 U µl^{-1} T4 DNA Ligase (Thermo Fisher Scientific) at 4 °C overnight. Ligation reactions were inactivated at 70 °C for 10 min and cDNA was ethanol precipitated.

For the preparation of the Illumina RNAylomeSeq library, cDNA was amplified by PCR using 2 U Phusion Polymerase (Thermo Fisher Scientific) in the presence of 5 % (v/v) DMSO, 200 µM dNTPs and 2,500 nM New England Biolabs Next Universal and Index Primer each (Primer Set 1, New England Biolabs). PCR products were purified by native PAGE and ethanol-precipitated. The double-stranded DNA (dsDNA) concentration was determined using a Quantus fluorometer (Promega) and library size was determined with the Bioanalyzer (Agilent). Equimolar amounts of each library were sequenced on a MiniSeq system (Illumina) using the MiniSeq High-Output Kit (150 cycles, Illumina) generating 20 million 151-bp single-end reads.

Analysis of next-generation sequencing data

Next-generation sequencing (NGS) data were demultiplexed using bcl2fastq (v.2.20.0, Illumina). Fastq files were assessed using FastQC (v.0.11.9) and Illumina sequencing adapters were trimmed from reads using cutadapt (v.1.18). Reads were aligned to a reference genome composed of an *E. coli* K12 (U00096.3), bacteriophage T4 (NC_000866.4) and RNAI (our design) with hisat2 (v.2.2.1). Primary alignments were selected using samtools (v.1.7) and reads per genomic feature were counted with featureCounts (v.2.0.1 from Subread package). The resulting counts table was subjected to further analysis and data visualization in R (v.4.1.2). Read counts were normalized to the overall number of mapped reads per sample and to the respective read counts for the RNAI spike-in as follows:

$$normreadcount_{i,j} = \frac{readcount_{i,j} * readcount\ (RNAI_j)}{\sum_i readcount_{i,j}}$$

Data visualization was done with a custom R script[60] and alignments were manually inspected in Integrative Genomics Viewer (IGV v.2.4.9). Hits were identified based on the following criteria: log2-transformed fold change (LFC) ≥ 1.5 comparing WT T4 and the T4 R73A, G74A mutant and log2-normalized mean expression among WT and R73A, G74A sample of one replicate ≥ −0.5.

Quantitative PCR validation of NGS data

cDNAs from RNAylomeSeq were diluted 1:30 in Millipore water. Quantitative PCR was performed on 1 µl diluted cDNA in 10 µl scale in technical duplicates amplifying regions of 100–150 bp with the iTaq Universal SYBR Green Supermix (Bio-Rad), according to the manufacturer's instructions, using the primers indicated in Supplementary Table 9. The log2-transformed difference in cycle-threshold values for WT T4 and T4 R73A, G74A infected samples from corresponding replicates was computed and an LFC ≥ 1 was set as a threshold for cDNA enrichment.

Ribosome RNAylation and proteomic analysis of RNAylated proteins

70S ribosomes (4.3 µg µl^{-1}) were RNAylated in transferase buffer in the presence of either 1 µM NAD–10-mer–Cy5 or 1 µM NAD–40-mer–Cy5 (Supplementary Table 7) by 0.05 µg µl^{-1} ModB at 15 °C for 90 min. RNAylated and non-RNAylated control samples were analysed using 12 % SDS–PAGE. To identify RNAylated proteins, SDS–PAGE-separated protein bands were excised and proteins were digested in gel as described previously[61]. LC-MS was carried out on an Exploris 480 mass spectrometer connected to an Ultimate 3000 RSLCnano system with a Proflow upgrade and a nanospray flex ion source (all Thermo Scientific). Peptide mixtures were then analysed on the LC-MS system described above with a peptide-separating gradient of 30 min from 2 % to 35 % buffer B. Peptide separation was performed on a reverse-phase HPLC column (75 µm × 42 cm) packed in-house with C18 resin (2.4 µm, Dr. Maisch). Peptides were ionized at 2.3 kV spray voltage with a heated capillary temperature at 275 °C and funnel RF level at 40. MS survey scans were acquired with a resolution of 120.000 at m/z 200 and full MS AGC target of 300 % with a maximal IT of 50 ms. The mass range was set to 350–1,650. Fragment spectra were acquired in data-dependent acquisition mode with a quadrupole isolation window of m/z = 1.5, an AGC target value of 200 % and a resolution of 15.000, and fragmentation was induced with a normalized higher-energy collision-induced dissociation collision energy of 27 %. MS raw data were searched with SEQUEST embedded in Proteome Discoverer 2.2 (Thermo Scientific) against a Uniprot *E. coli* protein database containing the bacteriophage T4 protein ModB. Spectral counts were exported from Scaffold Viewer and total spectral counts per sample were used to normalize

spectral counts for all other proteins by division in Microsoft Excel 2016 followed by calculation of the ratio of normalized spectral counts from modified and unmodified bands.

RNAylation of proteins in *E. coli* lysates

A fresh pellet from 40 ml *E. coli* B strain culture at an OD600 of around 0.8 was resuspended in 2 ml transferase buffer (10 mM $Mg(OAc)_2$, 22 mM NH_4Cl, 50 mM Tris-acetate, pH 7.5, 1 mM EDTA, 10 mM 2-mercaptoethanol, 1 % glycerol). Cells were lysed by sonication (3 × 2 min at 50 % power, 0.5 s pulse) and the lysate was cleared from the cell debris by centrifugation at 27,670 g at 4 °C for 30 min. The supernatant was used in RNAylation assays.

Lysate (100 µl) was incubated in the presence of 0.93 µM NAD–10-mer–Cy5 (0.47 µM with reference to the NAD-capped) or 0.93 µM P–10-mer–Cy5 (Supplementary Table 7), 0.37 U murine RNase inhibitor (New England Biolabs) and 0.69 µM ModB at 15 °C. Samples of 10 µl were taken before the addition of ModB and after 2, 5, 10, 20, 30 and 60 min, and were immediately resuspended in one volume of 2× Laemmli buffer. Samples were analysed by 12 % SDS–PAGE applying the same reference (rS1 RNAylated with NAD–10-mer–Cy5) to each gel. The Cy5 signal was recorded using the Cy5 blot option of the ChemiDoc Imaging System at a manual exposure of 90 s. Gels were then stained in Coomassie solution and imaged with the same system.

E. coli lysates with various concentrations of ModB were processed and analysed by proteomics as described previously[39].

Determination of NAD concentrations in *E. coli* lysates

A dilution series of *E. coli* cell lysate was prepared in PBS. NAD was diluted in PBS starting from a 100 mM stock creating NAD solutions of 1,000 nM to 3.125 nM. The NAD solutions, the lysate dilution series and a PBS blank were assessed for their NAD concentrations using the NAD/NADH-Glo Assay (Promega), according to the manufacturer's instructions in triplicates. Luminescence measurements were carried out on a Tecan plate reader (Spark) in a 384-well flat white plate. A linear fit ($R^2 = 0.9836$) was performed for NAD concentrations between 400 nM and 4 nM with a linear correlation to intensity. The equation was used to calculate NAD concentrations for the *E. coli* lysate as the mean of the technical triplicates.

Western blotting

Proteins were separated by 10 % SDS–PAGE and gels were equilibrated in transfer buffer (25 mM Tris, pH 8.3, 192 mM glycine, 20 % (v/v) methanol). Polyvinylidene difluoride membranes with a pore size of 0.2 µm (GE Healthcare) were activated in methanol for 1 min and equilibrated in transfer buffer. Proteins were transferred from gels to the membranes in a semi-dry manner at

300 mA for 1.5 h, unless indicated differently. After the transfer, membranes were dehydrated by soaking in methanol and washed twice with TBS-Tween (TBS-T; 10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05 % (v/v) Tween 20). Afterwards, 10 ml blocking buffer (5 % (w/v) milk powder in TBS-T) were added to the membranes and incubated at room temperature for 1 h. To detect ADP-ribosylated proteins, membranes were incubated with a 1:10,000 dilution of anti-pan-ADPr binding reagent MABE1016 (Merck) in 10 ml washing buffer (1 % (w/v) milk powder in TBS-T) at 4 °C overnight[62]. Membranes were washed and incubated with 10 ml of a 1:10,000 dilution of the horseradish peroxidase–goat-anti-rabbit IgG secondary antibody (Advansta) in washing buffer at room temperature for 1 h. Afterwards, membranes were washed with PBS. ADP-ribosylated proteins were visualized by chemiluminescence using the SignalFire ECL Reagent or the SignalFire Elite ECL Reagent (Cell Signaling Technology), according to the manufacturer's instructions.

If proteins in SDS–PAGE gels needed to be visualized before blotting, a TCE staining method[53] was used. Resolving gels were supplemented with 0.5 % (v/v) TCE. For visualization, gels were activated by ultraviolet transillumination (with a wavelength of 300 nm) for 60 s. Proteins then showed fluorescence in the visible spectrum.

Quantification of RNAylation

rS1 proteins were isolated from *E. coli* strain B pTAC rS1 bacteria (Supplementary Table 10) that were either uninfected or infected with bacteriophage T4. rS1 (1.5 µM) was treated with 1 µM ARH1 in the presence of 12.5 mM Tris-HCl, pH 7.5, 25 mM NaCl, 25 mM KCl and 5 mM $MgCl_2$. Alternatively, rS1 (1.5 µM) was treated with 0.5 U endonuclease P1 in 100 mM $Mg(OAc)_2$, 220 mM NH_4Cl, 500 mM HEPES, pH 7.5, 10 mM EDTA, 100 mM β-mercaptoethanol and 10 % glycerol. Digests were incubated at 37 °C for 2 h. Afterwards, digests were precipitated by the addition of nine volumes of ethanol and precipitated by centrifugation (14,000 pm) at 4 °C for 1 h. Protein pellets were resuspended in 10 µl 1× Laemmli buffer and analysed by Western blotting. ADPr modifications were detected by the primary antibody MABE1016 (Merck) as described above. The pan-ADPr signals for ADP-ribosylated rS1 were normalized to the corresponding band intensities in the TCE stain. Normalized intensities for untreated rS1 were then divided by the intensity for P1-treated rS1 to yield the fractions of ADP-ribosylated and RNAylated rS1 for the two modifications.

Phage mutagenesis

The CRISPR–Cas9 spacer plasmids were generated by introducing the modB spacer sequence into the DS-SPCas plasmid (Addgene, 48645) (Supplementary Table 10). The modB-carrying vector pET28_ModB was used as a donor DNA for homologous recombination in CRISPR–Cas9-mediated mutagenesis. The pET28_ModB plasmid was modified by site-directed mutagenesis, during which point mutations R73A and G74A were exposed to modB. The R73A mutation led to the inactivation of ADP-ribosyltransferase activity. The G74A mutation was located in the protospacer adjacent motif and was required to prevent the cleavage of donor DNA by Cas9 nuclease. The applied primers are listed in Supplementary Table 9. The resulting plasmids were sequenced and transformed into *E. coli* BL21 (DE3).

The CRISPR–Cas9-mediated mutagenesis was based on previous work[34]. The DS_SPCas_ModB plasmid with the target spacer sequence and the donor plasmid pET28a_ModB_R73A/G74A were co-transformed into *E. coli* DH5α. The cells were further infected by bacteriophage T4 (1:10,000 phages:cells), and the plaque assay was done. The plates were incubated overnight at 37 °C and the resulting plaques were screened for mutants. Single plaques were picked by sterile pipet tips and transferred into 200 µl Pi–Mg buffer (26 mM Na_2HPO_4, 68 mM NaCl, 22 mM KH_2PO_4, 1 mM $MgSO_4$, pH 7.5) supplied with 2 µl CCl_3H. The samples were incubated at room temperature for 1 h. Next, 2 µl of the sample was transferred to a new PCR tube and heated to 95 °C for 10 min. The sample was further used for DNA amplification using PCR (primers used are listed in Supplementary Table 9). The amplified DNA was purified by agarose gel electrophoresis and submitted for Sanger sequencing.

Plaque assay

The *E. coli* culture of interest was grown to an OD600 of around 0.8–1.0. Next, 300 µl of the culture was infected with 100 µl of WT T4 phage or T4 ModB(R73A, G74A) (Supplementary Table 10) mutant, with either defined or unknown MOI. The bacteria–phage suspension was incubated at 37 °C for 7 min and subsequently transferred to 4 ml LB soft agar (0.75 %), mixed and poured onto an LB-agar plate (1.5 % LB agar). The plates were incubated at 37 °C overnight and validated the following day.

Time course of T4-mediated lysis of *E. coli*

LB medium (100 ml in 500-ml baffled flasks) was inoculated with *E. coli* B culture overnight to OD600 = 0.1 and was then incubated at 37 °C with shaking at 180 rpm until OD600 = 0.8 was reached. The culture was cooled to room temperature and infected by either WT T4 phages or T4

ModB(R73A, G74A) mutants (Supplementary Table 10) to an MOI of 5. The culture was further incubated at room temperature with shaking at 150 rpm. Cell lysis was tracked by measuring the OD600 at different times of infection (0–200 min after infection). The experiment was run in biological triplicates.

Burst-size determination

LB medium (100 ml in 500-ml baffled flasks) was inoculated with *E. coli* B culture overnight to OD600 = 0.1 and was then incubated at 37 °C with shaking at 180 rpm until OD600 = 0.8 was reached, as above. The culture was infected either by WT T4 phages or T4 ModB(R73A, G74A) mutant (Supplementary Table 10) to an MOI of 0.01 and further incubated at 37 °C without shaking.

To determine the total number of infective centres, T0 (comprising unadsorbed and already adsorbed phages), at 5 min after infection, 100 µl of infected culture was used to reinfect 300 µl *E. coli* B cells (OD600 = 1.0) with a subsequent plaque assay. The number of unadsorbed phages (U) was determined by transferring 1 ml infected culture to 50 µl CCl_3H. In this way, *E. coli* cells were disrupted, after which the unadsorbed phages remained intact and were used for plaque assay. T0−U, represents the number of initially infected centres. The number of unadsorbed phages (Uxmin) was continuously traced during infection and used to calculate the number of T4-phage progeny (T4-phage progeny = Uxmin/(T0−U5min). The time point at which the first increase in phage number was observed was treated as the first burst time point and was used to calculate the phage burst size (burst size = Uburst1/(T0−U5min)).

Data were plotted using OriginPro 2020b software. Error bars represent s.d. of the means for three biological replicates. For selected time points, statistical tests were done as two-sided t-tests in R (v.4.2.2) implemented in the ggpubr package (v.0.6.0) using a significance level of 0.05.

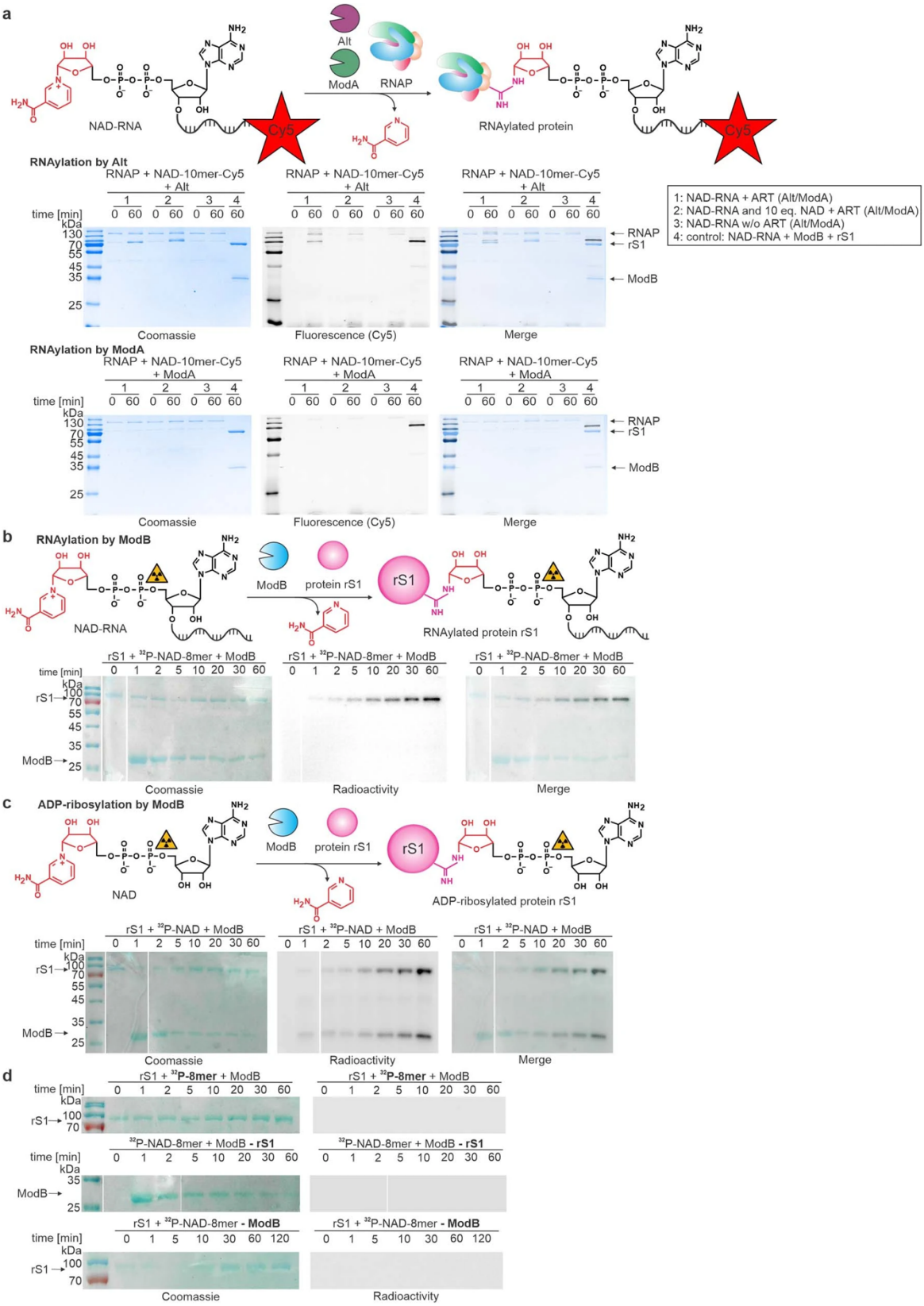
Phage adsorption kinetics

LB medium (100 ml in 500-ml baffled flasks) was inoculated with *E. coli* B culture overnight to an OD600 = 0.1 and incubated at 37 °C with shaking at 180 rpm until OD600 = 0.8 was reached, as above. The culture was cooled to room temperature and infected by either WT T4 phages or T4 ModB(R73A, G74A) mutants (Supplementary Table 10) to an MOI of 0.1. Immediately after infection, 100 µl of the culture was used to determine the number of total infective centres, T0, by plaque assay. Then 100 µl of the culture was taken at different time points of infection (0–25 min after infection) and 5 µl CCl_3H was added to disrupt *E. coli* cells. This suspension was subsequently used to determine the number of unadsorbed phages (Uxmin) by plaque assay. The

calculation of the adsorption rate was performed as follows: adsorption rate (%) = 100 % − (Uxmin/T× 100 %).

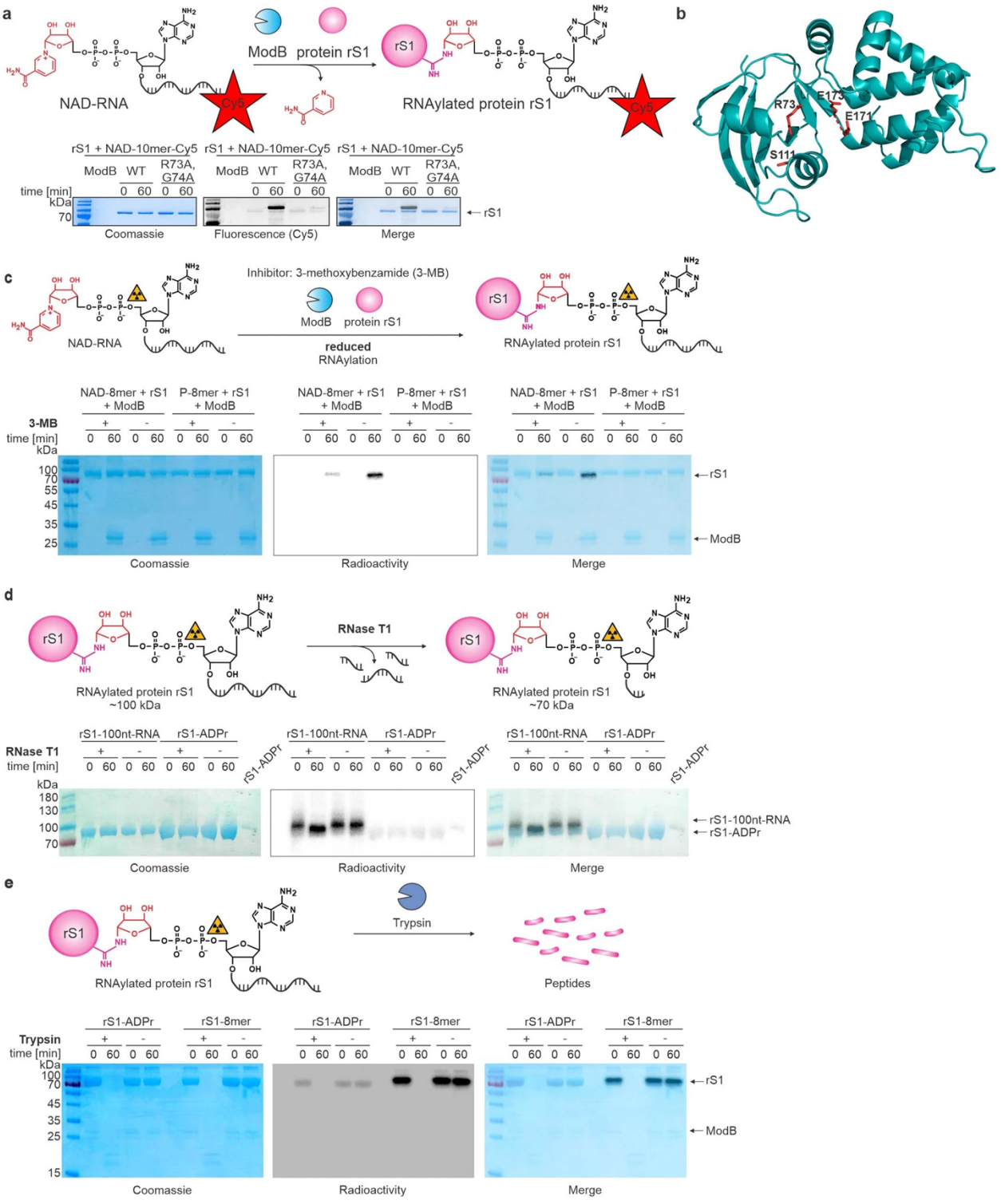
Data were plotted using OriginPro 2020b software. Error bars represent s.d. of the means for three biological replicates. For selected time points, statistical tests were done as two-sided t-tests in R (v.4.2.2) implemented in the ggpubr package (v.0.6.0) using a significance level of 0.05.

6.6. Extended Data and Supplementary Information



Extended Data Fig. 1 ADP-ribosylation and RNAylation by T4 ARTs.

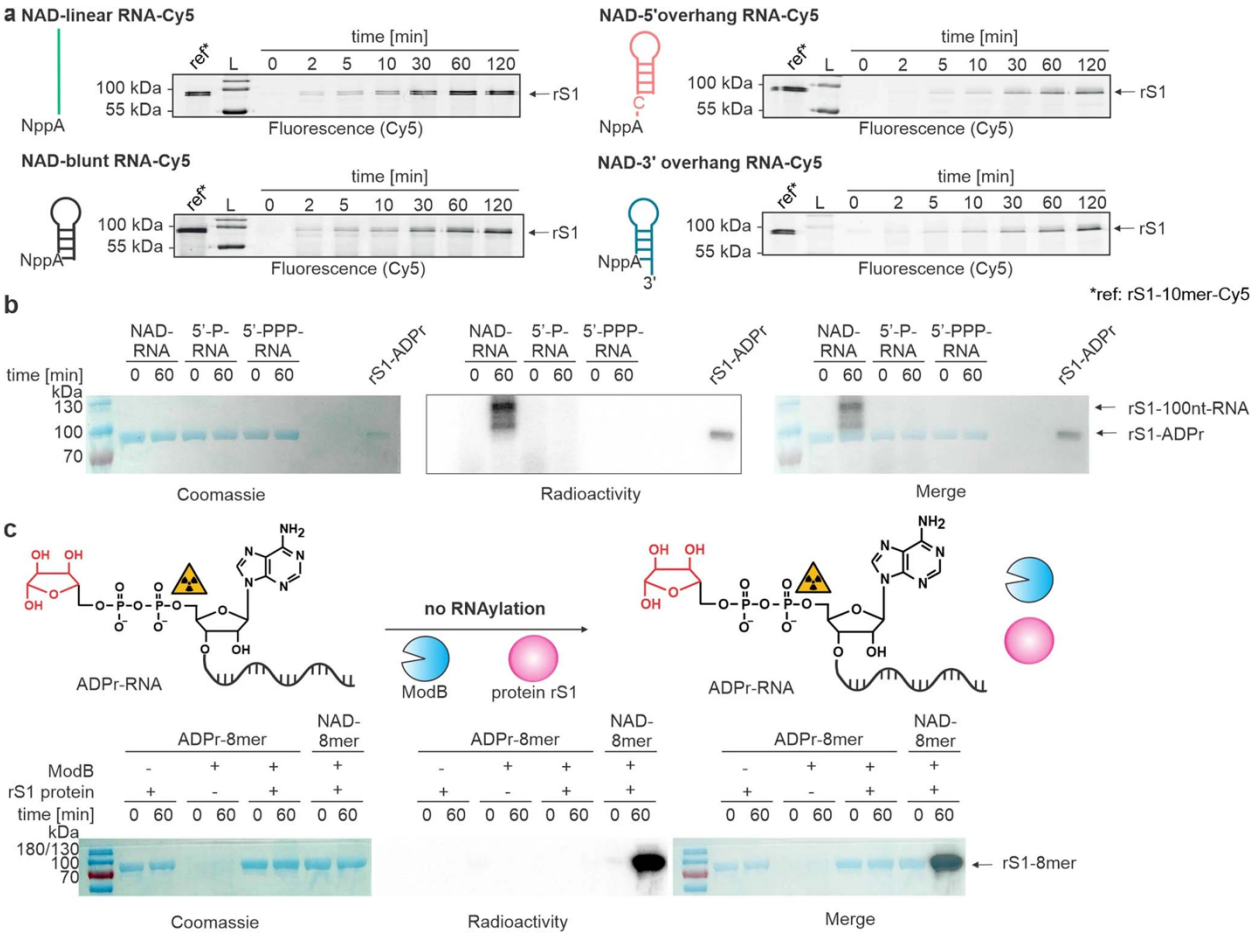
a, Characterisation of RNAylation of the RNA polymerase (RNAP) by the ARTs Alt or ModA in the presence of NAD-10mer-Cy5 (1), additional 10 equivalents of NAD (2) or in the absence of the respective ART (3) (n = 3). rS1 RNAylated with NAD-10mer-Cy5 by ModB serves as a reference (4). The RNAP is a well-established target protein of Alt and ModA and was thus chosen to assess RNAylation by Alt and ModA. Alt slightly RNAylates the RNAP *in vitro* which is abolished in the presence of 10 equivalents of NAD relative to NAD-10mer-Cy5. Protein load is visualised by Coomassie staining and RNAylated protein is visualised in the fluorescent Cy5 channel. b,c, Time course analysis of the ModB-mediated RNAylation (b) and ADP-ribosylation (c) of rS1 analysed by SDS-PAGE (n = 3 each). RNAylated or ADP-ribosylated protein is visualised by radioactivity scan and protein load confirmed by Coomassie staining. d, Negative controls for RNAylation of rS1 with ModB analysed by SDS-PAGE. RNAylation assay was performed in the presence of ^{32}P-RNA lacking the NAD-cap (upper panel) and in the absence of either rS1 (- rS1) (middle panel) or ModB (-ModB) (lower panel) (n = 3). In these experiments, no RNAylation was detected in the radioactive scan of the SDS-PAGE gel.



Extended Data Fig. 2 Characterisation of the RNAylation of protein rS1 by ModB.

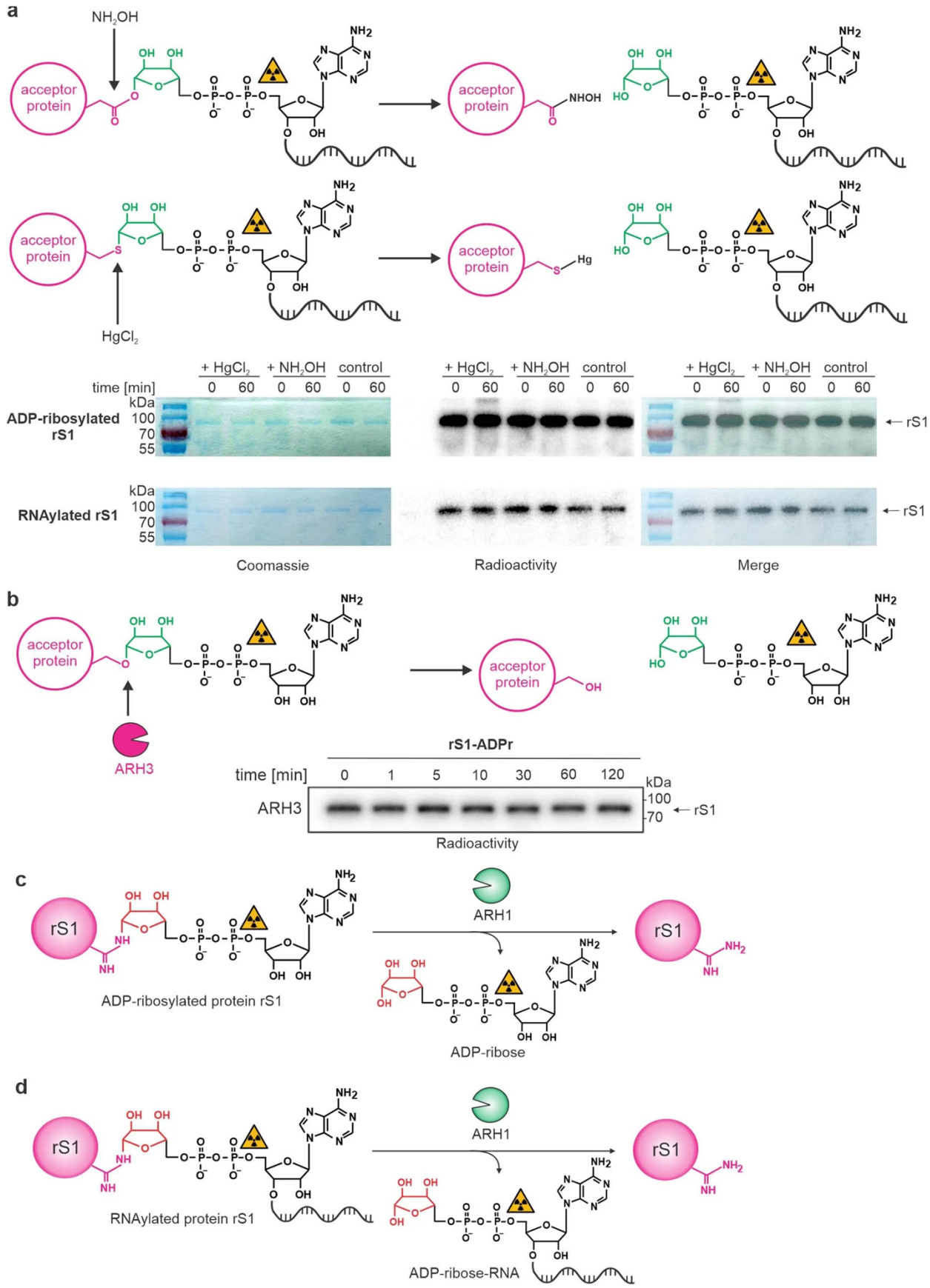
a, RNAylation of rS1 in the presence of catalytically active ModB and catalytically inactive ModB R73A, G74A with NAD-10mer-Cy5 (n = 3). In addition to the catalytically important residue R73, we mutated G74 as well. Mutation of G74 results in an altered PAM region, which is important for CRISPR-Cas9 gene editing

of the T4 phage genome. b, AlphaFold prediction[63] of the structure of ModB. Active site residues of the R-S-EXE motif1 are highlighted in red. Corresponding confidence metrics are shown in Supplementary Fig. 2. c, Inhibition of *in vitro* RNAylation of protein rS1 by ModB via ART inhibitor 3-MB. Reactions were performed with ^{32}P-NAD-RNA 8mer (^{32}P-NAD-8mer) as well as ^{32}P-RNA 8mer (negative control) (n = 3). 3-MB reduces the yield of RNAylated rS1. d, *in vitro* digest of RNAylated and ADP-ribosylated protein rS1 by RNase T1. Reactions performed in the absence of RNase T1 (-) serve as negative controls. Protein rS1 ADP-ribosylated in the presence of ^{32}P-NAD was applied as a reference (S1-ADPr) (n = 2). Upon T1 digest, the 100nt-RNA at rS1 is shortened, and the molecular weight of RNAylated rS1 is reduced. This leads to similar electrophoretic mobility as for ADP-ribosylated rS1. e, Tryptic digest of ADP-ribosylated and RNAylated protein rS1 (n = 2). The protein is degraded in the presence of trypsin, and RNAylation and ADP-ribosylation signals are lost. All samples were analysed by 12 % SDS-PAGE, protein was visualised by Coomassie staining and RNAylation was assessed via a radioactivity scan.



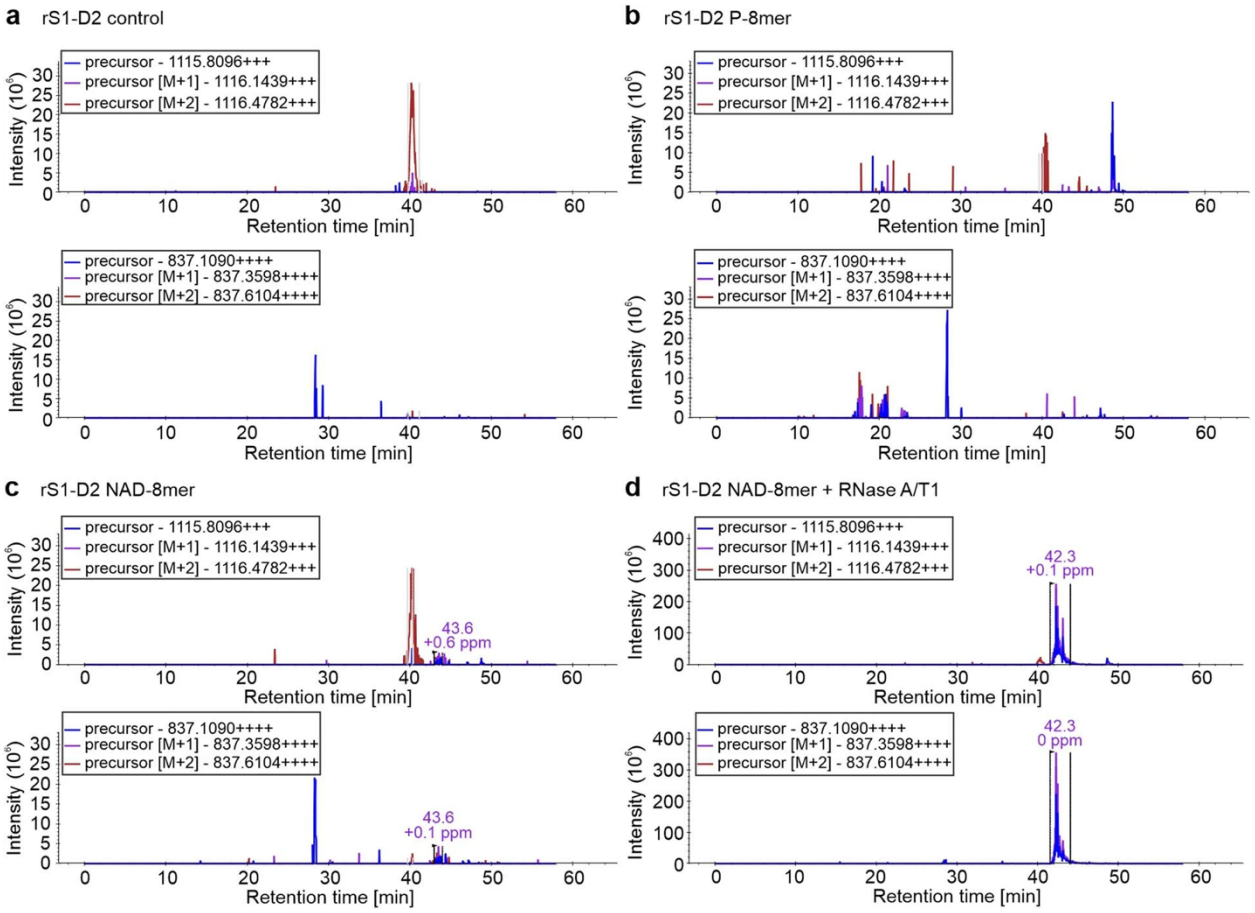
Extended Data Fig. 3 Characterisation of ModB mediated RNAylation *in vitro*.

a, Analysis of the role of RNA secondary structure on RNAylation reaction. Four different 3′-Cy5-labelled NAD (NppA)-capped RNAs were tested including a linear (green) NAD-capped RNA (10mer) and three structured RNAs with either a 3′-overhang (blue), a 5′-overhang (red) or a blunt end (black) (n = 3). SDS-PAGE analyses of the time course of RNAylation are shown. Quantification of the signal intensities (Cy5 scan) relative to the reference is shown in Fig. 2c. ModB prefers linear 5′-ends of NAD-capped RNAs. L = ladder. b, Analysis of the RNAylation dependency on the presence of a 5′-NAD-cap of the RNA. 10 % SDS-PAGE analysis of *in vitro* RNAylation of the protein rS1 by ModB in the presence of either 5′-NAD-capped (NAD-RNA), 5′-monophosphate- (5′-P-RNA) or 5′-triphosphate-100nt-RNA (5′-PPP-RNA) (n = 2). RNAylation with radiolabelled RNA is detected by radioactivity scan and protein load visualised by Coomassie staining. *In vitro* RNAylation of rS1 is only observed in the presence of NAD-RNA. RNAylated rS1 cannot be detected by Coomassie staining due to low sensitivity. c, Characterisation of ADPr-RNA (which is lacking the nicotinamide moiety compared to NAD-RNA) as a substrate for ModB (n = 2). As a positive control, NAD-8mer was applied. All reactions were analysed by 12 % SDS-PAGE. RNAylation with radiolabelled RNA is detected by radioactivity scan and protein load visualised by Coomassie staining. ADPr-RNA is not accepted as a substrate for ModB-catalysed RNAylation *in vitro*.


a
NH2OH
acceptor protein
NHOH
HgCl2
Hg
+ HgCl2
+ NH2OH
control
time [min]
0 60 0 60 0 60
kDa
100
70
55
ADP-ribosylated rS1
RNAylated rS1
rS1
Coomassie
Radioactivity
Merge
b
ARH3
rS1-ADPr
time [min]
0 1 5 10 30 60 120
kDa
100
70
rS1
Radioactivity
c
rS1
ADP-ribosylated protein rS1
ARH1
ADP-ribose
d
RNAylated protein rS1
ADP-ribose-RNA

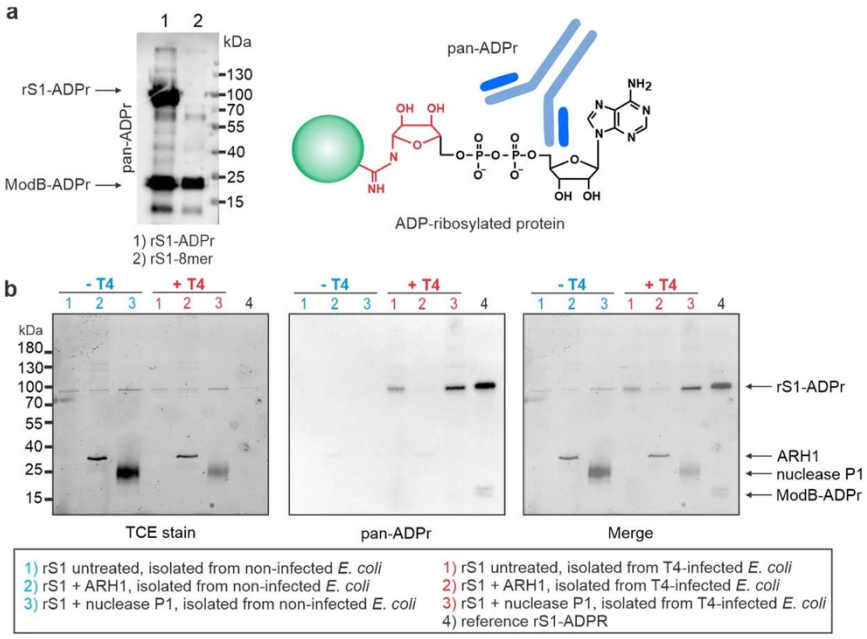
Extended Data Fig. 4 Specific removal of RNAylation using chemical and enzymatic treatments.

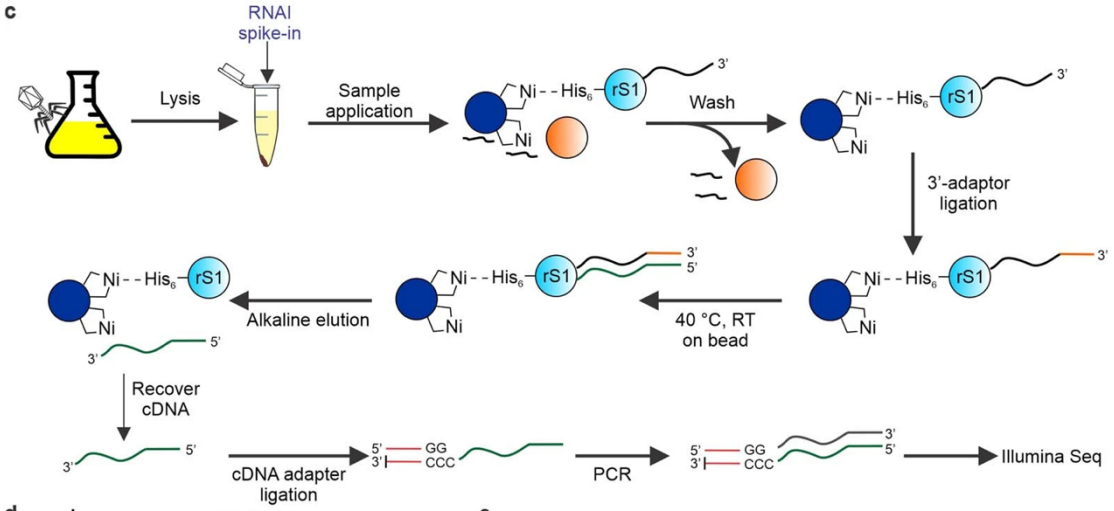
a, Different ADP-ribose-protein linkages have been shown to be either stable or unstable in the presence of $HgCl_2$ and neutral hydroxylamine (NH_2OH), which represents a relatively straightforward and fast approach to identify ADP-ribosylation sites. Treatment with NH_2OH hydrolyses linkages between glutamate/aspartate and ADP-ribose. $HgCl_2$ specifically cleaves thiol-glycosidic bonds. ADP-ribosylated and RNAylated protein rS1 were treated with NH_2OH or $HgCl_2$. The removal of ADPr or RNA by these chemicals would result in a decrease of the radioactive signal of protein rS1. All samples were analysed by 12 % SDS-PAGE, stained in Coomassie (protein loading control) and RNAylation assessed as radioactivity. A decrease of the radioactive signal in comparison to the control (untreated) was not determined (n = 1). b, *in vitro* time course of the stability of rS1 ADP-ribosylation in the presence of ARH3 analysed by 12 % SDS-PAGE (n = 1). The autoradiography scan is presented. ARH3 did not remove the ADP-ribosylation. c-d, Reaction schematics for the removal of the ADP-ribosylation (c) and RNAylation (d) of rS1 by ARH1.

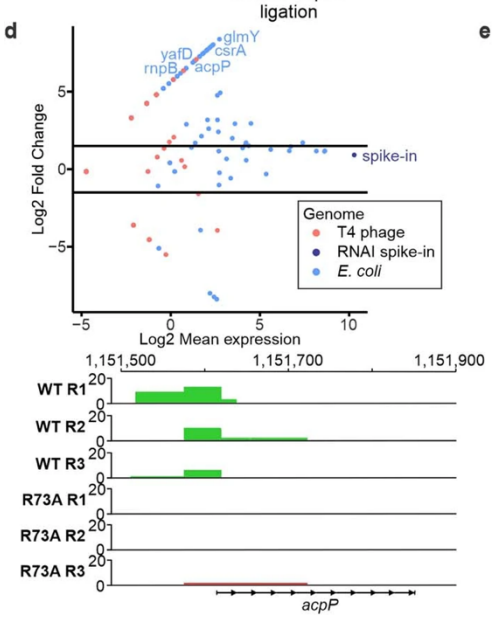


Extended Data Fig. 5 Ion chromatograms of unmodified rS1 and *in vitro* RNAylated rS1 extracted from LC-MS/MS data.

Extracted ion chromatograms (XICs) for triply and quadruply charged precursor ions (monoisotopic masses 1115.8096 and 837.1090, respectively). XICs were extracted using Skyline59, an open source document editor for creating and analysing targeted proteomics experiments. The masses correspond to an rS1 peptide AFLPGSLVDVRPVRDTLHLEGK with an attached ADPr-cytidine. Recombinant S1 domain 2 was *in vitro* incubated with ModB and one of the following components: a, no other supplements, b, uncapped RNA-8mer, c, NAD-RNA-8mer, d, NAD-RNA-8mer treated with RNase A and T1 (results in ADPr-cytidine adducts). An elution peak at 42.3 min is observable in d and corresponds to the peptide modified with ADPr-cytosine. Spurious intensities can be observed in c and might represent a degradation product. a and b show only background/contaminant peaks. A contaminant peak at 40 min can be also observed in d (consider the difference in the intensity scale between d and a-c).



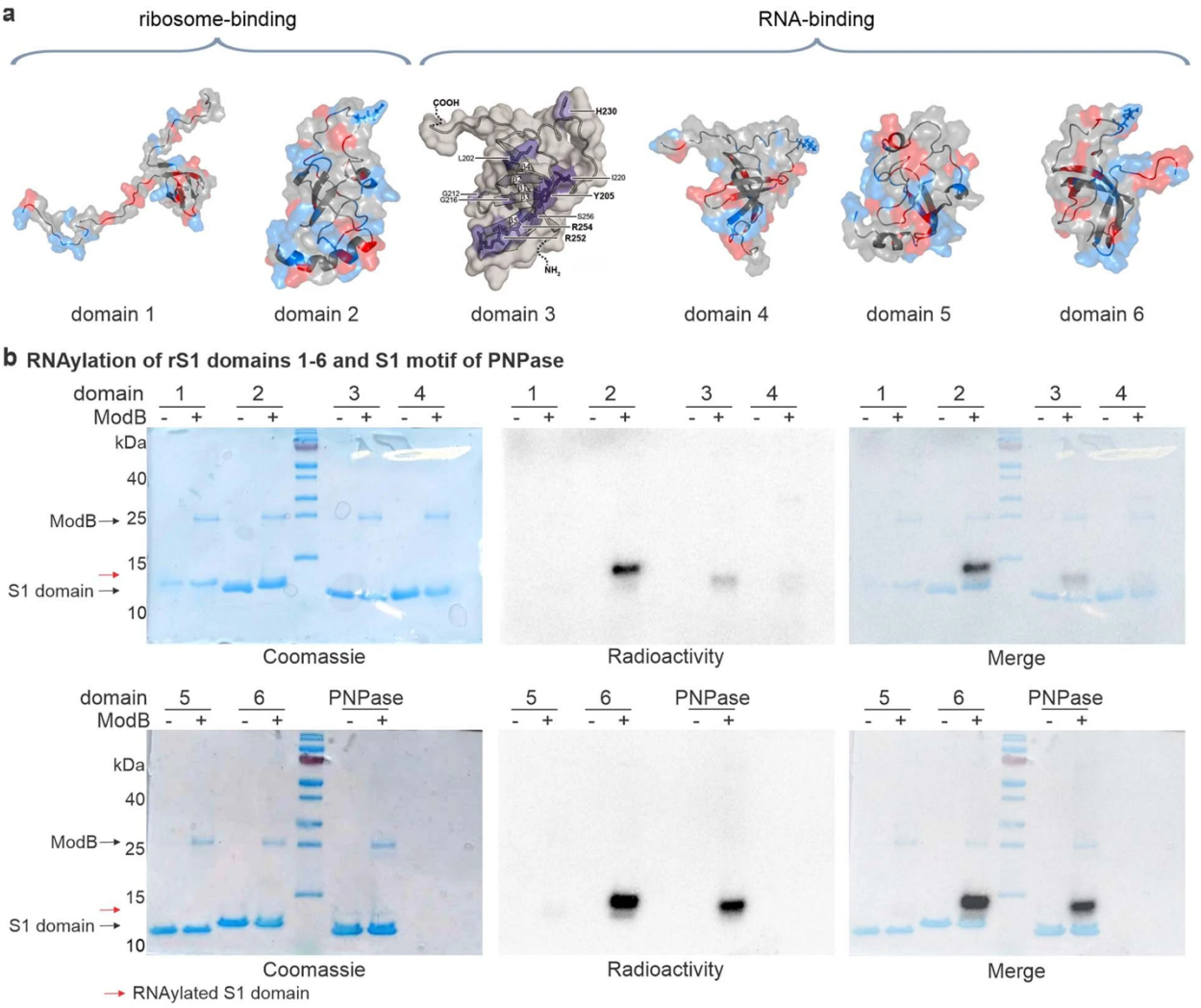




e

Gene	RNAylomeSeq R1	RNAylomeSeq R2	RNAylomeSeq R3	qPCR R1	qPCR R2	qPCR R3	Genome	Read distribution	NAD-RNA
acpP	+	+	+	+	+	+	*E. coli*	5'-UTR and 5'-fragment	Cahova et al., 2015 Zhang et al., 2021
gadY	+	-	-	+	-	+	*E. coli*	full coverage	Cahova et al., 2015
rob	+	-	-	n.d.	n.d.	n.d.	*E. coli*	5'-UTR and 5'-fragment	Zhang et al., 2021
ipIII	+	-	-	n.d.	n.d.	n.d.	T4 phage	within gene body	n.d.
mcaS	-	+	-	-	+	-	*E. coli*	full coverage	Cahova et al., 2015
glmY	-	+	-	n.d.	n.d.	n.d.	*E. coli*	full coverage	Cahova et al., 2015 Zhang et al., 2021
oxyS	-	+	-	-	+	-	*E. coli*	full coverage	Cahova et al., 2015
a-gt	-	-	+	n.d.	n.d.	n.d.	T4 phage	within gene body	n.d.
rnaC	-	-	+	-	-	(+)	T4 phage	full coverage and 3'-UTR	n.d.

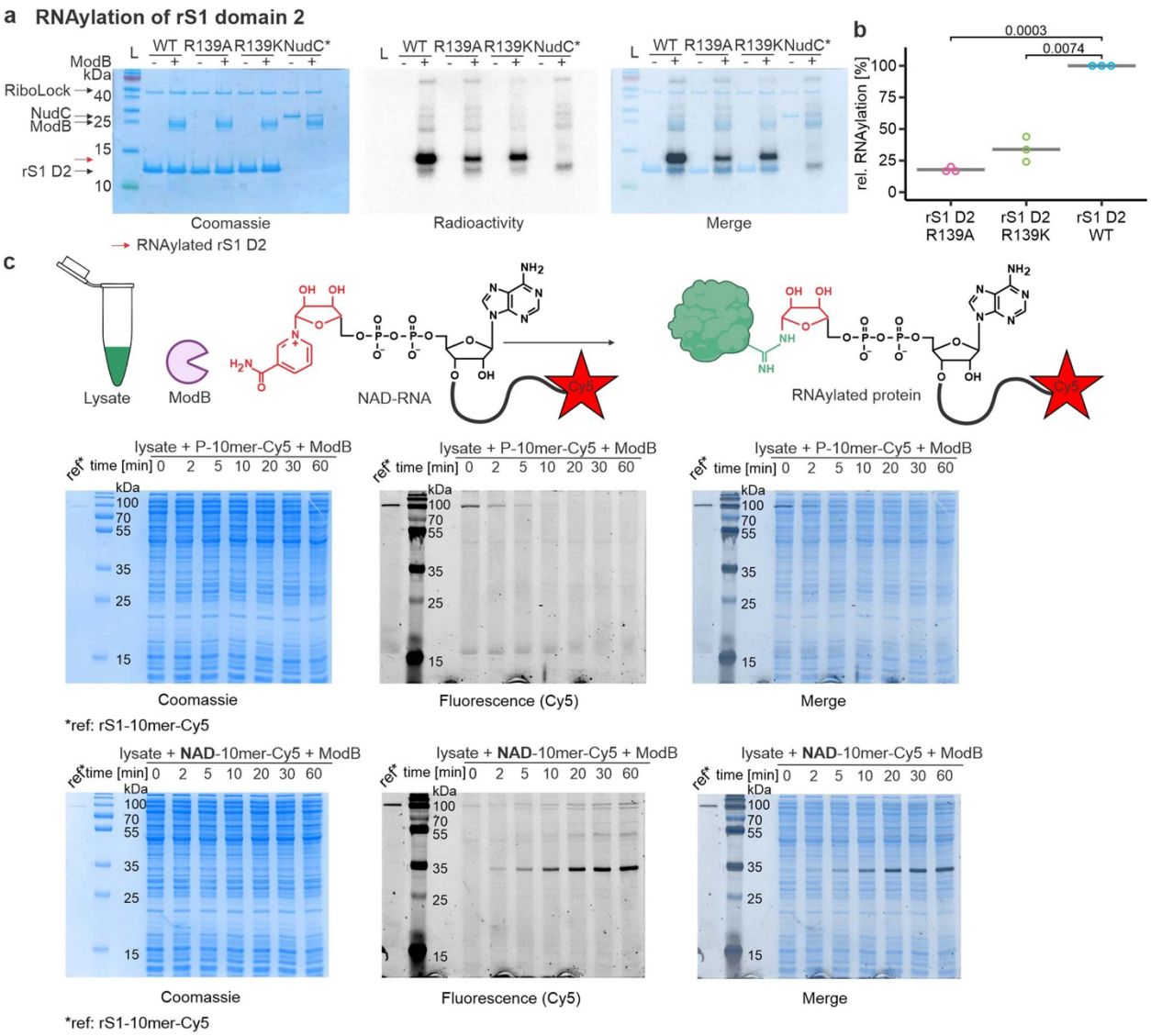
Extended Data Fig. 6 *In vivo* characterisation of the RNAylation by Western blot and RNAylomeSeq. a, Analysis of the substrate specificity of the pan-ADPr antibody. *In vitro* prepared ADP-ribosylated or RNAylated protein rS1 was applied to evaluate the specificity of the antibody (n = 3). The pan-ADPr antibody detected ADP-ribosylated proteins rS1 and ModB (lane 1). In contrast, RNAylated rS1 is not detected by pan-ADPr (lane 2). However, a signal for ADP-ribosylated ModB was observed due to self-ADP-ribosylation in its expression host *E. coli* (lane 2). b, Quantification of RNAylation using the combination of nuclease P1 digest and detection of protein-linked ADP-ribose by Western blot. Visualisation of protein load by TCE stain. Removal of the ADP-ribose signal by ARH1 treatment. pan-ADPr signals for ADP-ribosylated rS1 were normalised to corresponding band intensities in the TCE stain. Normalised intensities for untreated rS1 were then divided by the intensity for P1-treated rS1 to yield the fractions of ADP-ribosylated and RNAylated rS1 among the two modifications. The corresponding dot plot is shown in Fig. 4b (n = 3 biologically independent replicates). c, Schematic illustration of the RNAylomeSeq protocol: Identification of RNAylated RNAs which are covalently attached to rS1 *in vivo*. Briefly, endogenously His-tagged rS1 is isolated from T4 phage infected *E. coli* with Ni-NTA beads. A spike-in - rS1 domain 2 RNAylated with NAD-RNAI - (RNAI spike-in) is added to the lysate which is meant to be enriched via the RNAylomeSeq workflow. rS1 captured on Ni-NTA beads is intensively washed with 8 M urea in order to remove RNA non-covalently bound to rS1. Similar to NAD captureSeq[33], an RNA 3′-adapter is ligated to covalently linked RNAs and RNA is reverse transcribed “on-bead”. cDNA is then eluted by alkaline digest of RNA and an additional adapter is ligated to the 3′-terminus of the cDNA. cDNA is amplified by PCR and sequenced by next-generation single-end sequencing (Illumina). Importantly, the RNAI spike-in is not meant to be enriched in any sample but rather to be found in each sample in similar amounts. Thereby, read counts can be normalised to the RNAI counts in each sample allowing for their comparison. d, MA plot showing RNAs enriched in the T4 phage WT infected sample compared to T4 phage ModB R73A, G74A control identified by RNAylomeSeq for replicate 2 (total of n = 3 biological replicates). Read counts per sample have been normalised to RNAI spike-in read counts which serves as an enrichment control for each sample. Thus, RNAI is not found enriched comparing T4 WT and T4 ModB R73A, G74A. Mean expression values (T4 WT and T4 ModB R73A, G74A condition) have been normalised by Log2 (x-axis) for each replicate separately. T4 WT and T4 ModB R73A, G74A read counts were compared via log2 fold change (y-axis). Read coverage on identified RNAylated RNAs as analysed in IGV is exemplarily shown for acpP in the lower panel depicting reads in T4 WT samples (green) vs. T4 ModB R73A, G74A samples (red). RNAylomeSeq merely identifies 5’-termini of mRNAs or, if 200nt or smaller, entire sRNA sequences. This is due to the application of single-end Illumina-Seq which automatically only captures the 5’-end of the respective read/transcript. e, Selected hits of RNAs identified by RNAylomeSeq comparing T4 phage WT and T4 ModB R73A, G74A. acpP was identified in all three replicates. However, some transcripts were only detected in one or two replicates. Enrichments have been further validated on cDNA level by qPCR. +: enriched; −: not enriched; (+): enriched, but Log2 fold change <= 1; n.d.: not defined.



Extended Data Fig. 7 RNAylation of rS1 domains D1 – D6 and S1 motif of PNPase by ModB *in vitro*.

a, Illustration of the rS1-motifs of rS1 based on crystal structures (PDB) of domains 1 (2MFI), 2 (2MFL), 4 (2KHI), 5 (5XQ5) and 6 (2KHJ) as well as an NMR structure of domain 3[64]. b, *in vitro* RNAylation of S1 domains and PNPase by ModB using a ^{32}P-NAD-8mer. ModB and S1 domains (D1-6) are marked with black arrows. RNAylated rS1 domains, characterised by a shift compared to the non-modified proteins, are highlighted with red arrows. n = 2 of biologically independent replicates. Reactions were analysed by

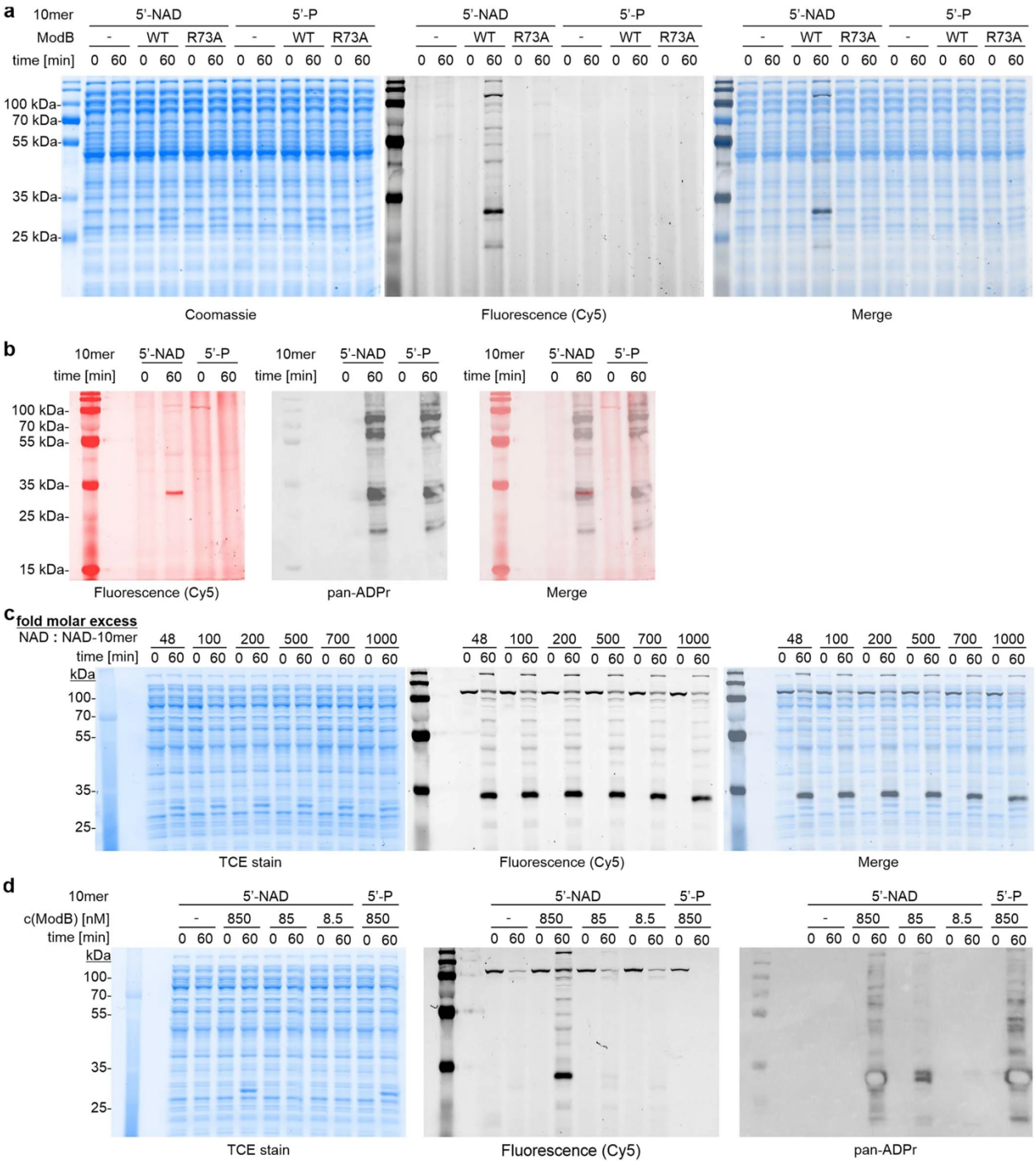
16 % Tricine-SDS-PAGE, stained in Coomassie and RNAylation recorded by autoradiography imaging (Radioactivity). c, Local alignment of rS1 D2 and D6 as well as the S1 domain of PNPase using T-coffee expresso[65]. R139 of D2 (highlighted with an arrow) is conserved in PNPase and D6.



Extended Data Fig. 8 Characterisation and identification of RNAylation target proteins of ModB.

a, Analysis of the *in vitro* RNAylation of rS1 domain 2 and its mutants R139A and R139K by 16 % Tricine-SDS-PAGE. An inactive NudC mutant (NudC*; V157A, E174A, E177A, E178A) was used as a negative control (n = 3). Radioactivity indicates RNAylation, Coomassie scan visualises protein load. b, Quantification of relative intensities of RNAylation of rS1 domain 2 and its mutants R139A and R139K based on radioactivity in 16 % Tricine-SDS-PAGE analysis. Per replicate, intensities were normalised to the rS1 D2 WT band intensity. A two-sided t-test was performed at psignif. < 0.05 indicating significantly decreased RNAylation of R139 mutants of rS1 domain 2 (p-value = 0.0003 (WT vs. R139A) and 0.0074 (WT vs. R139K)). n = 3 of biologically independent replicates. c, RNAylation of *E. coli* cell lysate by ModB using 3′-Cy5-labelled NAD-RNA (schematically shown in upper panel). A time course of *E. coli* cell lysate RNAylation by ModB in the presence of either a 5′-monophosphorylated RNA 10mer (P-10mer-Cy5, middle panel) or 5′-NAD-capped RNA 10mer (NAD-10mer-Cy5, lower panel), each with a 3′-fluorescent (Cy5)

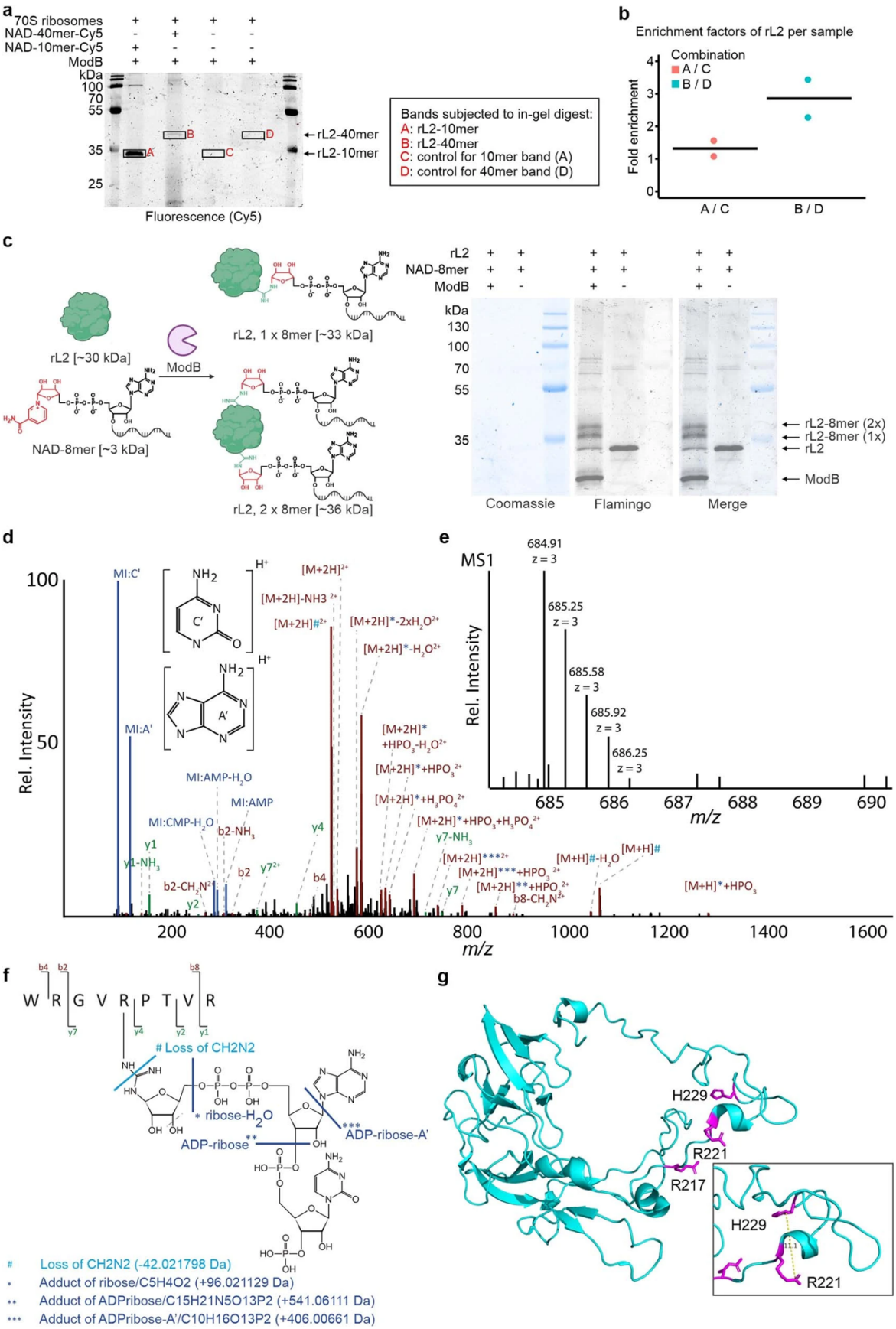
label. rS1 RNAylated with an NAD-10mer-Cy5 is applied as a reference (ref). The time course of lysate RNAylation was analysed by 12 % SDS-PAGE, protein visualised by Coomassie staining and RNAylation recorded via fluorescence (Cy5). n = 3 of biologically independent replicates. NAD concentration in the lysates exceeds the utilised NAD-10mer concentration by 48-fold. NAD concentration in the lysates of 22.5 µM (n = 1 biologically independent replicates, n = 3 technical replicates) was determined using the NAD/NADH-Glo assay (Promega). The schematic protein and tube in c were created using BioRender (https://biorender.com).



Extended Data Fig. 9 Characterisation of the specificity of ModB-mediated RNAylation in *E. coli* lysates.

a, RNAylation of *E. coli* cell lysate in the presence of ModB WT or inactive ModB R73A, G73A or the absence of ModB using 3′-Cy5-labelled NAD-10mer or P-10mer. Time point 0 shows lysate before addition of ModB, 60 min shows RNAylation after 60 min incubation with ModB. Reactions were analysed by 12 % SDS-PAGE, protein visualised by Coomassie staining and RNAylation recorded via fluorescence

(Cy5). n = 2 of biologically independent replicates. b, Samples from lysate RNAylation with Cy5-labelled 5′-NAD- or 5′-P-10mer (as presented in Extended Data Fig. 8c) before addition of ModB (0 min) and after 60 min incubation in the presence of ModB (60 min) were analysed by 10 % SDS-PAGE and RNAylation monitored by fluorescence (Cy5, here shown in red). Subsequently, Western blotting was performed and ADP-ribosylation was detected using pan-ADPr binding reagent (MABE1016, shown in grayscale). n = 2 of biologically independent replicates. Different band patterns were observed for ModB-mediated RNAylation and ADP-ribosylation in *E. coli* lysates indicating a distinct target specificity of ModB for RNAylation. NAD concentration in the lysates exceeds utilised NAD-10mer concentration by 48-fold. NAD concentration in the lysates of 22.5 µM (n = 1 biologically independent replicates, n = 3 technical replicates) was determined using the NAD/NADH-Glo assay (Promega). c, Lysate RNAylation by ModB in the presence of various molar excesses of NAD over NAD-10mer-Cy5 ranging from 48-fold (native lysate) to 1000-fold via additional spike-in NAD (n = 2). Cy5 represents RNAylation. TCE stain indicates protein load, which is enabled by binding of trichloroethanol in the gel to tryptophan residues in proteins which enhances their fluorescence under UV light and thereby enables their detection[53]. 700-fold molar excess of NAD reduces RNAylation to 67 % (n = 2 biologically independent replicates) compared to “native” lysate. Total Cy5 signals for each lane were quantified to determine and compare RNAylation levels. d, Lysate RNAylation and ADP-ribosylation in the presence of various ModB concentrations (850, 85 and 8.5 nM) monitored by fluorescence (Cy5 for RNAylation) and Western blot (pan-ADPr for ADP-ribosylation). TCE stain indicates protein load. In average, RNAylation is reduced to 8.6 % and ADP-ribosylation is reduced to 6.9 % in lysates with ModB concentrations that approximate the cellular conditions (85 nM). Total Cy5 or pan-ADPr signals (excluding ModB ADP-ribosylation signal) for each lane were quantified to determine and compare RNAylation or ADP-ribosylation levels, respectively. n = 2 biologically independent replicates.


a
70S ribosomes
NAD-40mer-Cy5
NAD-10mer-Cy5
ModB
kDa
100
70
55
35
25
A
B
C
D
rL2-40mer
rL2-10mer
Fluorescence (Cy5)
Bands subjected to in-gel digest:
A: rL2-10mer
B: rL2-40mer
C: control for 10mer band (A)
D: control for 40mer band (D)
b
Enrichment factors of rL2 per sample
Combination
A / C
B / D
Fold enrichment
A / C
B / D
c
rL2 [~30 kDa]
ModB
NAD-8mer [~3 kDa]
rL2, 1 x 8mer [~33 kDa]
rL2, 2 x 8mer [~36 kDa]
rL2
NAD-8mer
ModB
kDa
130
100
70
55
35
rL2-8mer (2x)
rL2-8mer (1x)
rL2
ModB
Coomassie
Flamingo
Merge
d
Rel. Intensity
m/z
MI:C'
MI:A'
MI:AMP-H2O
MI:AMP
MI:CMP-H2O
b2-NH3
y4
y1
y1-NH3
b2-CH2N2
b2
y7 2+
b4
y2
[M+2H]2+
[M+2H]-NH3 2+
[M+2H]#2+
[M+2H]*-2xH2O2+
[M+2H]*-H2O2+
[M+2H]* +HPO3-H2O2+
[M+2H]*+HPO3 2+
[M+2H]*+H3PO4 2+
[M+2H]*+HPO3+H3PO4 2+
y7-NH3
[M+2H]***2+
[M+2H]***+HPO3 2+
[M+2H]**+HPO3 2+
y7
b8-CH2N2
[M+H]#-H2O
[M+H]#
[M+H]*+HPO3
e
MS1
Rel. Intensity
m/z
684.91
z = 3
685.25
z = 3
685.58
z = 3
685.92
z = 3
686.25
z = 3
f
W R G V R P T V R
b4 b2 b8
y7 y4 y2 y1
Loss of CH2N2
* ribose-H2O
** ADP-ribose
*** ADP-ribose-A'
# Loss of CH2N2 (-42.021798 Da)
* Adduct of ribose/C5H4O2 (+96.021129 Da)
** Adduct of ADPribose/C15H21N5O13P2 (+541.06111 Da)
*** Adduct of ADPribose-A'/C10H16O13P2 (+406.00661 Da)
g
H229
R221
R217
H229
R221

Extended Data Fig. 10 Scope of ModB RNAylation targets in *E. coli*.

a, RNAylation of *E. coli* ribosomes by ModB. RNAylated protein is shifted upon incubation with NAD-40mer compared to NAD-10mer which itself increases protein weight by approx. 3 kDa. Relative enrichment of RNAylated target protein was assessed by subjecting RNAylated protein bands and respective control bands generated in the absence of RNA to in-gel digest and LC-MS/MS analysis (n = 2). b, Plot of the enrichment of fractional spectral counts for 50S ribosomal protein L2 (rL2) based on in-gel-digest and LC-MS/MS analysis presented in a. Enrichment is calculated for RNAylation with NAD-10mer (A/C) or NAD-40mer (B/D), relative to the respective, non-RNAylated control bands based on spectral counts from Scaffold (n = 2). c, Analysis of the *in vitro* RNAylation of rL2 by ModB in the presence of NAD-8mer. RNAylated rL2 proteins have reduced electrophoretic mobility during SDS-PAGE. Protein was visualised by fluorescent protein stain (Flamingo) and protein ladder visualised by Coomassie staining. Signals were quantified using ImageLab indicating that about 80 % of rL2 is RNAylated by ModB *in vitro* (n = 3). Band patterns indicated that rL2 can be RNAylated once or even twice *in vitro*. d–f, Tandem MS-based identification of RNAylated rL2 peptide. d, MS/MS fragment ion spectrum (spectrum ID: 8679) of RNAylated rL2 peptide WRGVRPTVR carrying ADP-ribose plus cytidine-monophosphate and a 3′-phosphate group. The spectrum shows marker ions of adenine (A') and cytosine (C') as well as AMP and CMP. The precursor ion ($[M+xH]^{x+}$) is detected unshifted, shifted by the mass of ADP-ribose (*) and by ADP-ribose with adenine loss (**). Also, precursor ions show a specific loss of 42.021798 Da, which can be explained by a loss of CH_2N_2 at the modified arginine. e, Isotopic peak pattern of the precursor ion shown in d, as detected in the corresponding MS precursor ion scan. f, Schematic sequence and RNA adduct representation of the RNAylated peptide shown in d and e including annotations of fragment ions. The fragmentation products observed in the MS/MS spectrum, shown in d, of the ADP-ribose+CMP+3′-phosphate adduct are indicated in the structure by light blue (mass loss) and dark blue (mass adducts) lines. g, Selected RNAylated residues of rL2 identified by LC-MS/MS. The catalytically important H229 is 11.1 Å apart from R221. rL2 structure derived from a 1.98 Å cryo-EM structure (7K00)[66]. The schematic protein in c was created using BioRender (https://biorender.com).

Extended Data Table 1: ADP-ribosylation of endogenously His-tagged rS1 during T4 phage infection

MaxQuant intensities are presented for T4 phage-infected and uninfected samples in biologically independent triplicates (n = 3). R139/R142 located in rS1 domain 2 and R485/R487 in rS1 domain 6 appear as ADP-ribosylation sites on rS1 *in vivo* in all three replicates. The ratio comparing intensity of ADP-ribosylated and unmodified species of the same peptide is computed for each sample and peptide.

Arginine (domain)	Sequence	Modifications	Intensity uninfected replicate 1	Intensity uninfected replicate 2	Intensity uninfected replicate 3	Intensity T4 infected replicate 1	Intensity T4 infected replicate 2	Intensity T4 infected replicate 3
R139/R142 (rS1 domain 2)	AFLPGSLVDVRPVRDTLHLEGK	ADP-ribosyl	0.00E+00	0.00E+00	0.00E+00	2.72E+08	2.50E+08	3.61E+08
	AFLPGSLVDVRPVRDTLHLEGK	Unmodified	1.41E+11	1.48E+11	3.50E+10	4.13E+09	1.23E+10	2.64E+10
	Ratio: intensity of ADP-ribosylated relative to unmodified peptide	**ADP-ribosyl/ unmodified**	0.00E+00	0.00E+00	0.00E+00	**6.60E-02**	**2.03E-02**	**1.37E-02**
R485/R487 (rS1 domain 6)	ASEASRDRVEDATLVLSVGDEVEAK	ADP-ribosyl	0.00E+00	0.00E+00	0.00E+00	3.51E+07	2.30E+07	3.50E+07
	ASEASRDRVEDATLVLSVGDEVEAK	Unmodified	2.37E+10	6.59E+10	2.36E+10	7.84E+08	6.52E+08	1.23E+10
	Ratio: intensity of ADP-ribosylated relative to unmodified peptide	**ADP-ribosyl/ unmodified**	0.00E+00	0.00E+00	0.00E+00	**4.47E-02**	**3.52E-02**	**2.83E-03**
R19 (rS1 domain 1)	EIETRPGSIVR	ADP-ribosyl	0.00E+00	0.00E+00	0.00E+00	7.73E+06	0.00E+00	0.00E+00
	EIETRPGSIVR	Unmodified	5.76E+10	7.17E+10	7.30E+10	4.47E+10	3.06E+10	2.19E+10
	Ratio: intensity of ADP-ribosylated relative to unmodified peptide	**ADP-ribosyl/ unmodified**	0.00E+00	0.00E+00	0.00E+00	**1.73E-04**	**0.00E+00**	**0.00E+00**

Extended Data Table 2 ADP-ribosylation of rS1-WT, -R139K and -R139A during T4 phage infection

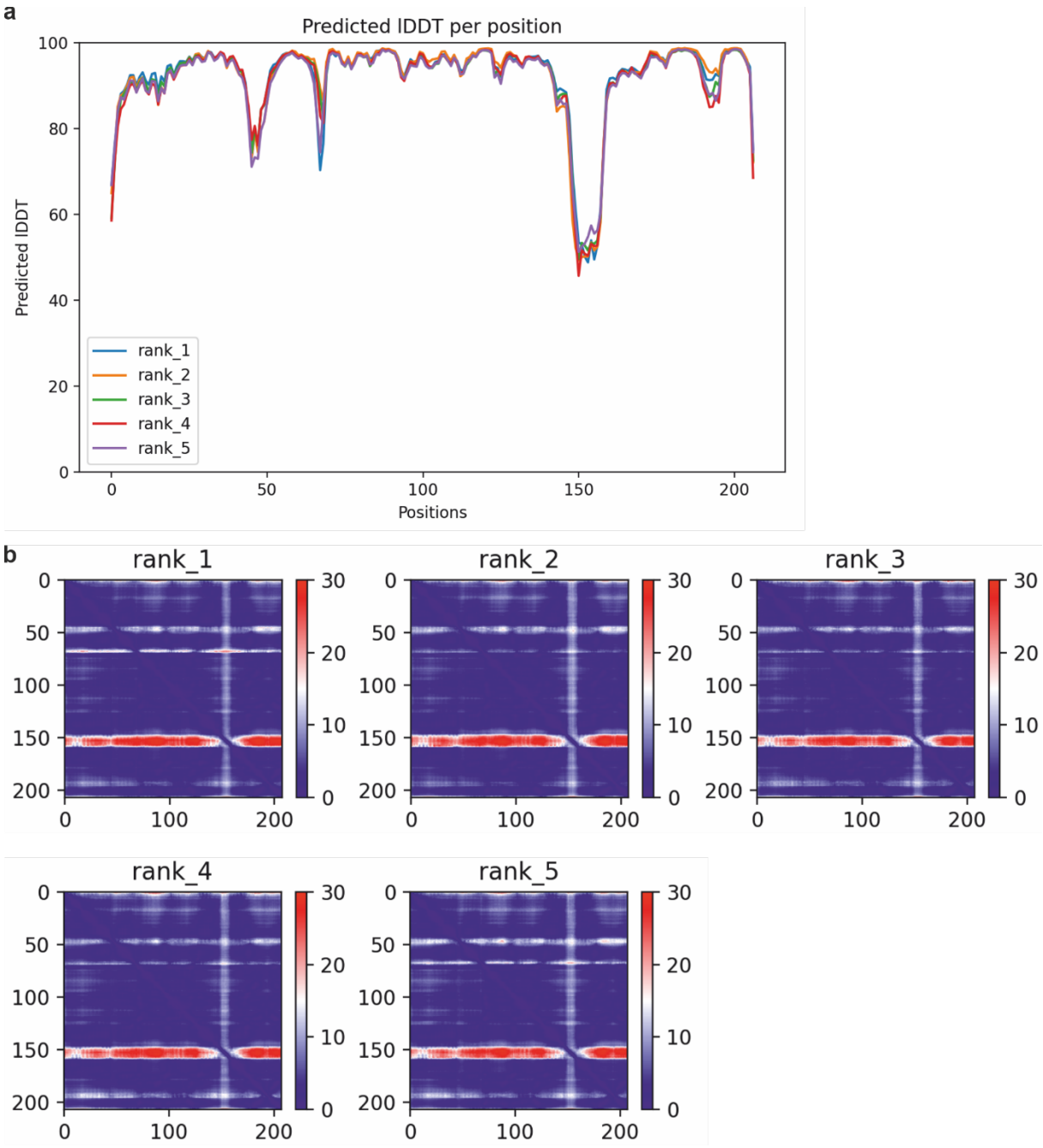
MaxQuant intensities are presented for T4 phage-infected and uninfected samples in biologically independent triplicates (n = 3) only for the respective peptide of the R139 mutation site which is expected for the respective rS1 version. ADP-ribosylation of the peptide in rS1 is observed *in vivo* in all three replicates. However, ADP-ribosylation at position 139 is abolished by R139A or R139K mutations (mutation indicated in red). The intensity of ADP-ribosylated peptide relative to the intensity of the corresponding unmodified peptide species is at least 3-fold reduced upon R139 mutation. One may speculate that R142 is nevertheless ADP-ribosylated in the mutated rS1 proteins but overall ADP-ribosylation yield at the peptide may be reduced as the potentially predominant ADP-ribosylation site (R139) is not available for modification.

rS1 R139K – peptide AFLPGSLVDVKPVRDTLHLEGK							
Sequence	**Modifications**	**Intensity rS1-R139K uninfected replicate 1**	**Intensity rS1-R139K uninfected replicate 2**	**Intensity rS1-R139K uninfected replicate 3**	**Intensity rS1-R139K T4 infected replicate 1**	**Intensity rS1-R139K T4 infected replicate 2**	**Intensity rS1-R139K T4 infected replicate 3**
AFLPGSLVDV**K**PVRDTLHLEGK	ADP-ribosyl	0.00E+00	0.00E+00	0.00E+00	1.61E+09	4.87E+09	1.50E+09
AFLPGSLVDV**K**PVRDTLHLEGK	Unmodified	2.27E+10	4.88E+10	4.70E+10	4.17E+10	3.52E+10	6.60E+10
Ratio: intensity of ADP-ribosylation relative to unmodified peptide		**0.00E+00**	**0.00E+00**	**0.00E+00**	**3.86E-02**	**1.38E-01**	**2.27E-02**
rS1 R139A – peptide AFLPGSLVDVAPVRDTLHLEGK							
Sequence	**Modifications**	**Intensity rS1-R139A uninfected replicate 1**	**Intensity rS1-R139A uninfected replicate 2**	**Intensity rS1-R139A uninfected replicate 3**	**Intensity rS1-R139A T4 infected replicate 1**	**Intensity rS1-R139A T4 infected replicate 2**	**Intensity rS1-R139A T4 infected replicate 3**
AFLPGSLVDV**A**PVRDTLHLEGK	ADP-ribosyl	0.00E+00	0.00E+00	0.00E+00	5.27E+09	8.05E+09	9.17E+09
AFLPGSLVDV**A**PVRDTLHLEGK	Unmodified	8.33E+10	9.93E+10	7.37E+10	7.50E+10	7.96E+10	7.32E+10
Ratio: intensity of ADP-ribosylation relative to unmodified peptide		**0.00E+00**	**0.00E+00**	**0.00E+00**	**7.02E-02**	**1.01E-01**	**1.25E-01**
rS1 WT – peptide AFLPGSLVDVRPVRDTLHLEGK							
Sequence	**Modifications**	**Intensity rS1-WT uninfected replicate 1**	**Intensity rS1-WT uninfected replicate 2**	**Intensity rS1-WT uninfected replicate 3**	**Intensity rS1-WT T4 infected replicate 1**	**Intensity rS1-WT T4 infected replicate 2**	**Intensity rS1-WT T4 infected replicate 3**
AFLPGSLVDV**R**PVRDTLHLEGK	ADP-ribosyl	0.00E+00	1.17E+08	0.00E+00	1.27E+10	1.12E+10	5.59E+09
AFLPGSLVDV**R**PVRDTLHLEGK	Unmodified	1.12E+10	3.34E+09	2.27E+10	6.85E+09	2.28E+10	1.51E+10
Ratio: intensity of ADP-ribosylation relative to unmodified peptide		**0.00E+00**	**3.49E-02**	**0.00E+00**	**1.85E+00**	**4.91E-01**	**3.71E-01**

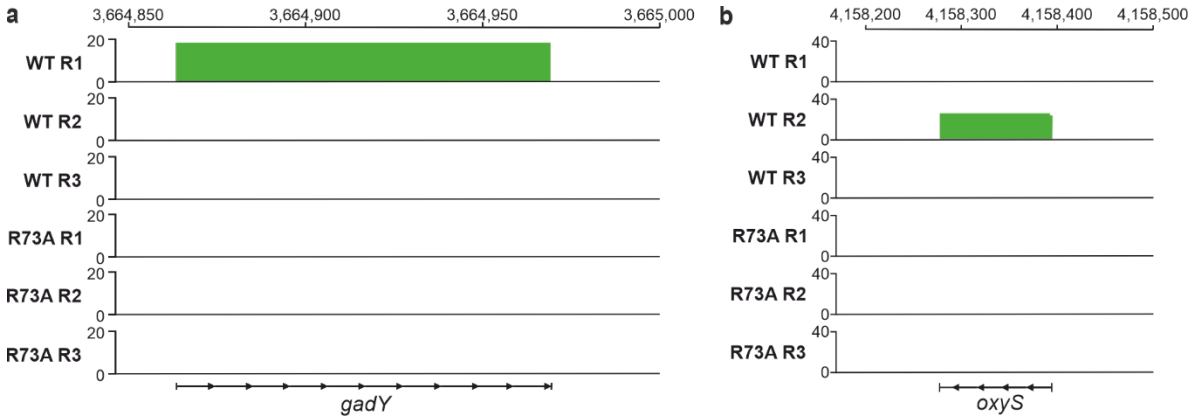
Extended Data Table 3: Comparison of ModB protein intensities in lysate assay and *in vivo* via proteomics

Normalised Log2 intensities for selected *E. coli* proteins and ModB found in proteomic analysis of *E. coli* cell lysates for *in vitro* RNAylation (n = 1) and in a previously published data set of the *E. coli* and T4 phage proteome 5 min post-infection38 (n = 3). Intensity of ModB is divided by the intensity for various *E. coli* proteins. At 8.5 nM ModB, the ratios approximate conditions found *in vivo*. Raw data is presented in Supplementary Table 5.

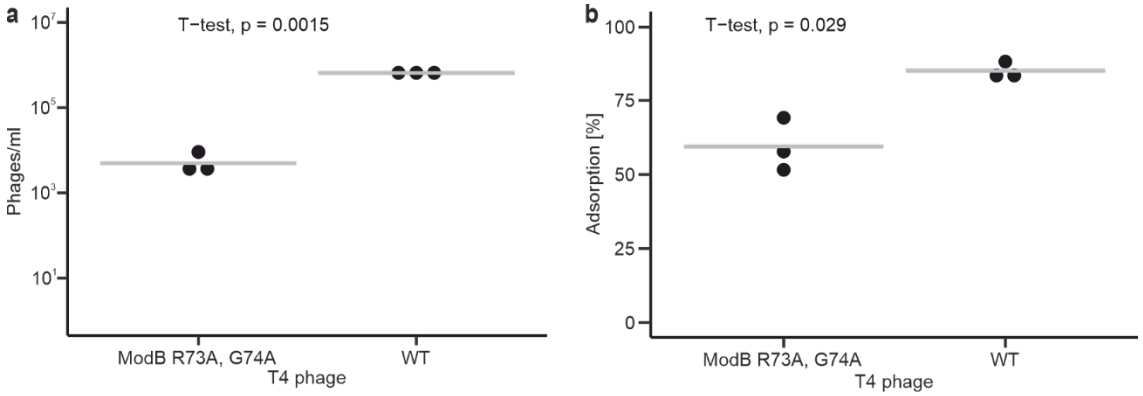
Protein	ModB at 85 nM in lysate		ModB at 8.5 nM in lysate		ModB *in vivo* (T4 phage infection)	
	Normalised Log2 Quantity	*Ratio Quantity ModB : Protein*	*Normalised Log2 Quantity*	*Ratio Quantity ModB : Protein*	*Mean of normalised Log2 Quantity*	*Ratio Quantity ModB : Protein*
rL2	31.71178244	0.895197071	31.47566823	0.812044472	32.78901	0.769619903
rS1	32.01051275	0.886842862	31.86823455	0.802041367	31.95659484	0.789667114
ModB	28.38829475	1	25.5596424	1	25.23507203	1
RpoB	31.36204058	0.905180091	31.20527758	0.819080758	30.90054714	0.816654537
Hfq	24.45952114	1.160623488	24.35407369	1.049501727	28.57508535	0.883114494
NudC	22.74631504	1.248039285	22.78526997	1.121761666	25.21463515	0.99919014



Supplementary Figure 2: Confidence metrics of the Alphafold prediction model of ModB structure shown in Extended Data Fig. 1f. Plots of predicted local-distance difference test (pLDDT) (**a**) and predicted aligned error (PAE) (**b**) are shown as Alphafold prediction metrics [63]. Confidence metrics for the model used here are represented by “rank_1”.



Supplementary Figure 3: IGV coverage plots for RNAylomeSeq data. Read coverage on identified RNAylated RNAs as analysed in IGV is exemplarily shown for *gadY* (**a**) and *oxyS* (**b**) depicting reads in T4 WT samples (green) vs. T4 ModB R73A, G74A samples (red) for each replicate. RNAylomeSeq merely identifies 5'-termini of mRNAs or, if 200nt or smaller, entire sRNA sequences. This is due to the application of single-end Illumina-Seq which automatically only captures the 5'-end of the respective read/transcript.



Supplementary Figure 4: Statistical tests for phenotype of T4 phage ModB R73A, G74A mutant shown in Fig. 5 c-e. Dotplots are shown for burst size 140 min post-infection (**a**) and phage adsorption to the host cell 8 min post-infection (**b**). Grey bars represent mean, black dots individual data points of n=3 biologically independent replicates each. Two-sided t-tests found both differences in burst size (140 min post infection; t-test, two-sided, p-value = 0.0015 at $p_{signif.} < 0.05$) and phage adsorption between T4 phage WT and T4 phage R73A, G74A (8 min post infection; t-test, two-sided, p-value = 0.029 at $p_{signif.} < 0.05$) to be statistically significant on a significance level of 0.05. T4 phage ModB R73A, G74A produces less progeny and adsorbs less efficiently to the host cell.

Supplementary Tables 1-6 are larger Excel tables that can be accessed via the following link: https://zenodo.org/records/10866162

Supplementary Table 1: MaxQuant Output for LC-MS/MS analysis of endogenously His-tagged rS1 from T4 phage-infected *E. coli* B strain. Endogenously His-tagged rS1 was isolated from T4 phage-infected (inf) and -uninfected *E. coli* B strain and subjected to LC-MS/MS analysis in biological triplicates (n=3). Intensities from MaxQuant are only shown for rS1 (1A; modificationSpecificPeptides). ADP-ribosylation is detected only for a small subset of rS1 peptides from T4 phage-infected samples whilst absent in uninfected samples. Predominantly, R139/R142 and R485/R487 were identified as ADP-ribosylation sites in all three replicates. ADP-ribosylated peptides are listed by rS1 domain and with the respective arginine residues in 1B; ADPr peptides. Modifications occur at R485/R487 (domain 6) and R139/R142 (domain 2) in all three replicates. Comparing intensities of ADP-ribosylated and unmodified peptides (1C; ADPr vs. unmodified peptides) shows ratios varying from 1.4 % to 6.6 %. Based on this data, one may speculate that R139/R142 and R485/R487 might be major ADP-ribosylation sites on rS1 *in vivo*. MaxQuant parameters for the presented data are presented in 1D; Parameters MaxQuant.

Supplementary Table 2: MaxQuant Output for LC-MS/MS analysis of His-tagged rS1-WT, -R139A and -R139K mutants from T4 phage-infected *E. coli*. MaxQuant Output filtered for rS1 protein and sorted according to ADP-ribosylation (ADP-ribosylwoDP) is presented (2A; modificationSpecificPeptides). T4 phage-infected samples (T4 phage) and -uninfected control (LB control) per rS1 version (WT, R139A or R139K mutant) are presented. A total of three biological replicates (n=3) were analysed per rS1 version. Peptides which were found ADP-ribosylated are listed and are assigned to their respective location (rS1 domain) and the modified arginines in the rS1 protein (2B; ADPr peptides). Intensities of ADP-ribosylated peptides are compared to their unmodified counterpart each by dividing the respective intensities (2C; ADPr vs. unmodified peptides). The peptide AFLPGSLVDVR(K/A)PVRDTLHLEGK is found ADP-ribosylated for rS1 WT, rS1 R139A and R139K only in T4 phage-infected samples across all three replicates. It becomes obvious that for the WT peptide high intensities of the ADP-ribosylated peptide relative to the unmodified peptide are detected across all three T4 phage-infected replicates. For the mutant peptides R139A and R139K, these intensities are at least 3-fold lower. Based on this finding, one may speculate that R139 mutation might reduce ADP-ribosylation of the AFLPGSLVDVRPVRDTLHLEGK peptide in rS1. MaxQuant parameters for the presented data are presented in 2D; Parameters MaxQuant.

Supplementary Table 3: *In vitro* ADP-ribosylation and RNAylation sites in rS1 protein as identified by LC-MS/MS analysis. Peptide spectrum match (PSM) information for ADP-ribosylated and/or RNAylated rS1 peptides are given in summarised form (pivot) and as complete output from OpenMS tool RNPxl (PSMs). Results were filtered for 1 % FDR on PSM level and q-values (scores) are given. Spectrum IDs, precursor m/z values, charge states, best localisation of modification within the peptide sequence, localisation score and mass errors (in ppm) are provided in "PSMs" sheet.

Supplementary Table 4: Genes identified to contribute to the RNAylome by RNAylomeSeq. An excerpt from the counts table is presented. Hits are calculated based on the mean read counts for each gene among T4 WT and R73A, G74A (MUT) samples for each replicate individually. For a hit, the log2 Fold Change (LFC) between WT and MUT sample is to be greater than 1.5 and the log2 transformed mean expression greater than -0.5. Hits are indicated as "+" for individual replicates. Read distribution for hits is presented in column "IGV" and the existence of corresponding NAD-capped transcripts is indicated in column "NAD-capped RNA?". Raw read counts are shown in WT_R1 – MUT_R3. Some hits are present in all replicates, some in one or two replicates only. Importantly, for the majority of protein_coding and ncRNA genes, reads initiate with an adenosine or contain an adenosine no more than 2 nt away from the read start. tRNA and rRNA (which more likely represent the background) hits are more abundant in replicates 2 and 3. Especially in replicate 3, the fraction of RNAI reads varies comparing WT and MUT samples, which may explain this variation from the background.

Supplementary Table 5: MaxQuant Output for LC-MS/MS analysis of *E. coli* cell lysate with addition of ModB to various concentrations. Samples 1 and 3 represent lysates without ModB, whilst samples 2 and 4 contain 85 nM and 8.5 nM ModB, respectively. The data serve as raw data for values shown in Extended Data Table 3, where log2 transformed protein intensity values are shown for each sample.

Supplementary Table 6: *In vitro* ADP-ribosylation and RNAylation sites in rL2 protein as identified by LC-MS/MS analysis. Peptide spectrum match (PSM) information for ADP-ribosylated and/or RNAylated rL2 peptides are given in summarised form (pivot) and as complete output from OpenMS tool RNPxl (PSMs). Results were filtered for 1 % FDR on PSM level and q-values (scores) are given. Spectrum IDs, precursor m/z values, charge states, best localisation of modification within the peptide sequence, localisation score and mass errors (in ppm) are provided in "PSMs" sheet.

Supplementary Table 7: RNAs used in this study.

RNA	RNA sequence
8mer	ACAGUAUU
RNAI	ACAGUAUUUGGUAUCUGCGCUCUGCUGAAGCCAGUUACCUUCGGAAA AAGAGUUGGUAGCUCUUGAUCCGGCAAACAAACCACCGCUGGUAGCG GUGGUUUUUUUGUU
100nt-RNA (Qβ)	AUCUUGAUACUACCUUUAGUUCGUUUAAACACGUUCUUGAUAGUAUC UUUUUAUUAACCCAACGCGUAAAGCGUUGAAACUUUGGGUCAAUUUG AUCAUG
10mer-Cy5	ACAGUAUUUG
2nt-5'overhang-Cy5	ACAGACUUCGGUCU-Cy5
3'overhang-Cy5	AGACUUCGGUCUA-Cy5
5'P-blunt-Cy5	AGACUUCGGUCU-Cy5
linear-Cy5	AGACUUCGAC-Cy5
40mer-Cy5	ACAGUAUUUGGUAUCUGCGCUCUGCUGAAGCCAGUUACUU-Cy5

Supplementary Table 8: Genomic DNA sequence of ARTs, rS1 variants and ADP-ribose hydrolases. Start codon in italic; thrombin cleavage site in bold; mutations in red and bold; restriction sites underlined.

Gene [5', 3' restriction site]	DNA sequence
Alt [*NcoI*, *XhoI*]	CC*ATGG*GAGAACTTATTACAGAATTATTTGACGAAGATACTACTCTTCCAATTACAAA CTTATATCCAAAGAAGAAAATACCGCAAATTTTTTCAGTTCATGTTGATGATGCAATT GAACAACCAGGCTTTCGTTTATGTACCTATACATCTGGAGGTGATACTAATCGTGAT TTAAAGATGGGCGATAAAATGATGCATATTGTTCCTTTTACATTAACTGCTAAAGGTT CAATTGCTAAATTAAAAGGTCTTGGTCCAAGCCCAATTAATTATATCAATTCAGTTTTT ACTGTTGCAATGCAAACAATGCGCCAGTATAAAATTGATGCCTGTATGCTCCGTATT CTTAAGTCTAAAACTGCTGGCCAAGCTCGACAAATTCAAGTTATTGCTGATAGACTT ATCCGTAGTCGTTCAGGTGGTAGATACGTCCTTCTTAAGGAACTCTGGGATTACGA TAAAAAGTATGCATATATTCTTATACATCGCAAAAATGTATCACTAGAAGACATTCCAG GAGTTCCGGAAATTAGTACCGAGCTCTTTACTAAAGTTGAATCGAAGGTCGGTGAT GTTTATATCAATAAAGATACTGGGGCTCAAGTAACTAAAAATGAGGCAATTGCAGCA TCTATTGCGCAAGAAAATGATAAACGTTCTGACCAAGCTGTAATCGTTAAAGTTAAA ATTTCCCGTAGAGCAATTGCGCAAAGTCAGTCATTGGAATCTTCTAGATTTGAAACA CCAATGTTTCAAAAATTTGAGGCTTCAGCGGCCGAATTAAATAAACCAGCGGACGC GCCTTTAATTTCTGATTCTAATGAATTAACGGTAATTTCTACTTCAGGATTTGCACTA GAGAATGCTCTTAGCAGTGTTACAGCTGGGATGGCATTCAGAGAAGCTTCTATAATT CCTGAAGATAAAGAATCCATTATTAACGCAGAAATAAAAAATAAAGCTTTAGAAAGAT TACGAAAAGAATCTATTACTTCAATAAAAACCTTAGAAACTATTGCTTCTATCGTCGA TGATACTTTAGAAAAATATAAGGGTGCTTGGTTTGAAAGAAATATTAACAAACATTCG CATTTAAACCAAGATGCTGCAAATGAGTTAGTACAAAATTCTTGGAATGCAATAAAAA CAAAGATTATTCGAAGAGAATTACGTGGATATGCTCTTACCGCTGGATGGTCATTAC ATCCTATAGTCGAAAATAAAGATTCATCTAAATACACACCAGCGCAAAAACGCGGAA TTCGTGAATACGTAGGTTCAGGATATGTAGACATAAATAATGCTCTTTTGGGATTATA TAATCCAGATGAGCGTACAAGTATTTTGACAGCATCTGACATAGAAAAAGCTATTGA TAATTTAGATTCAGCCTTTAAAAATGGTGAACGATTACCAAAAGGTATTACTTTGTAT CGTTCACAACGAATGTTACCTTCAATATACGAAGCAATGGTAAAAAATCGAGTTTTTT ATTTTAGAAACTTTGTGTCAACATCATTATATCCAAATATTTTTGGTACTTGGATGACT GATTCATCTATAGGTGTTTTACCAGACGAAAAGCGTTTAAGCGTTTCTATTGATAAAA CTGATGAAGGACTTGTAAATTCTAGCGATAATTTAGTTGGAATTGGATGGGTTATTAC TGGGGCTGATAAGGTCAATGTTGTTTTACCCGGTGGAAGTTTAGCGCCTTCAAATG AAATGGAAGTCATTTTGCCACGTGGATTAATGGTCAAAGTTAATAAAATAACCGATG CATCTTACAATGATGGAACAGTTAAAACTAACAACAAGCTTATTCAAGCTGAAGTTAT GACCACAGAAGAACTCACCGAATCGGTAATCTATGACGGAGACCATTTAATGGAAA CTGGTGAATTGGTTACAATGACAGGTGATATAGAAGATAGAGTTGACTTTGCATCAT TTGTTTCATCAAATGTTAAACAGAAAGTAGAATCATCTCTTGGAATTATTGCGTCTTG CATAGATATTGCAAACATGCCTTACAAGTTCGTTCAAGGACTGGTGCCGCGCGGCA GC<u>CTCGAG</u>
ModA [*NcoI*, *XhoI*]	C*ATGG*GAAAATACTCAGTAATGCAACTAAAAGATTTTAAAATAAAATCAATGGATGC ATCGGTGCGTGCTTCTATTCGTGAAGAATTACTTTCTGAAGGGTTTAATTTATCTGA AATTGAACTTTTAATTCATTGTATTACTAATAAACCAGATGACCATTCTTGGTTAAATG AAATAATCAAATCTCGTTTGGTTCCAAACGATAAACCTCTTTGGAGAGGTGTTCCAG CTGAGACTAAACAAGTATTAAATCAAGGAATTGATATTATTACATTTGATAAAGTCGTA TCAGCTTCATATGATAAAAATATAGCTCTACATTTTGCTTCTGGTTTAGAGTATAACAC ACAAGTTATTTTTGAATTCAAAGCTCCTATGGTATTCAATTTCCAGGAGTATGCTATA AAAGCTCTACGCTGTAAAGAATACAATCCAAACTTTAAGTTTCCGGATAGTCATCGT

	TATCGTAATATGGAATTAGTTTCAGATGAACAAGAAGTAATGATACCAGCTGGAAGT GTATTTAGAATTGCAGATAGATATGAGTATAAAAAGTGTTCAACATACACTATCTATAC TCTTGATTTTGAAGGATTTAATCTACTGGTGCCGCGCGGCAGCCTCGAG
ModB [*NcoI*, *XhoI*]	CC*ATGG*GAATTATTAATCTTGCAGATGTTGAACAGTTATCTATAAAAGCTGAAAGCGT TGATTTTCAATATGATATGTATAAAAAGGTCTGTGAAAAATTTACTGACTTTGAGCAG TCTGTTCTTTGGCAATGTATGGAAGCCAAAAAGAATGAAGCTCTTCATAAGCATTTA AATGAAATCATTAAAAAGCATTTAACTAAATCGCCTTATCAATTATATCGTGGTATATC AAAATCGACAAAAGAACTCATTAAAGATTTACAAGTTGGAGAAGTGTTTTCAACGAA CAGGGTAGATTCATTTACTACTAGTTTGCATACAGCGTGTTCTTTTTCTTATGCTGAA TATTTCACTGAAACAATACTTCGTTTAAAAACTGATAAAGCTTTTAATTATTCTGACCA TATCAGCGATATTATACTTTCTTCTCCTAATACTGAGTTTAAGTACACGTATGAAGATA CTGATGGATTAGATTCAGAGCGTACTGATAACTTAATGATGATTGTGCGTGAACAAG AATGGATGATTCCAATTGGAAAGTATAAAATAACTTCTATTTCAAAAGAAAAATTACA CGATTCATTTGGAACATTTAAAGTTTATGATATTGAGGTAGTTGAACTGGTGCCGCG CGGCAGCCTCGAG
ModB R73A, G74A [*NcoI*/*XhoI*]	*CCATGG*GAATTATTAATCTTGCAGATGTTGAACAGTTATCTATAAAAGCTGAAAGCGT TGATTTTCAATATGATATGTATAAAAAGGTCTGTGAAAAATTTACTGACTTTGAGCAG TCTGTTCTTTGGCAATGTATGGAAGCCAAAAAGAATGAAGCTCTTCATAAGCATTTA AATGAAATCATTAAAAAGCATTTAACTAAATCGCCTTATCAATTATATGCGGCAATATC AAAATCGACAAAAGAACTCATTAAAGATTTACAAGTTGGAGAAGTGTTTTCAACGAA CAGGGTAGATTCATTTACTACTAGTTTGCATACAGCGTGTTCTTTTTCTTATGCTGAA TATTTCACTGAAACAATACTTCGTTTAAAAACTGATAAAGCTTTTAATTATTCTGACCA TATCAGCGATATTATACTTTCTTCTCCTAATACTGAGTTTAAGTACACGTATGAAGATA CTGATGGATTAGATTCAGAGCGTACTGATAACTTAATGATGATTGTGCGTGAACAAG AATGGATGATTCCAATTGGAAAGTATAAAATAACTTCTATTTCAAAAGAAAAATTACA CGATTCATTTGGAACATTTAAAGTTTATGATATTGAGGTAGTTGAACTGGTGCCGCG CGGCAGCCTCGAGCACCACCACCACCACCACTGA
pET28-rS1 [*NcoI*, *XhoI*]	CC*ATGG*GAACTGAATCTTTGCGGCATGCTCAACTCTTTGAAGAGTCCTTAAAAGAA ATCGAAACCCGCCCGGGTTCTATCGTTCGTGGCGTTGTTGTTGCTATCGACAAAGA CGTAGTACTGGTTGACGCTGGTCTGAAATCTGAGTCCGCCATCCCGGCTGAGCAG TTCAAAAACGCCCAGGGCGAGCTGGAAATCCAGGTAGGTGACGAAGTTGACGTTG CTCTGGACGCAGTAGAAGACGGCTTCGGTGAAACTCTGCTGTCCCGTGAGAAAG CTAAACGTCACGAAGCCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAAC TGTTACCGGTGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAAC GGTATTCGTGCGTTCCTGCCAGGTTCTCTGGTAGACGTTCGTCCGGTGCGTGACA CTCTGCACCTGGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAG CGCAACAACGTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAG AGCGCGATCAGCTGCTGGAAAACCTGCAGGAAGGCATGGAAGTTAAAGGTATCGT TAAGAACCTCACTGACTACGGTGCATTCGTTGATCTGGGCGGCGTTGACGGCCTG CTGCACATCACTGACATGGCCTGGAAACGCGTTAAGCATCCGAGCGAAATCGTCA ACGTGGGCGACGAAATCACTGTTAAAGTGCTGAAGTTCGACCGCGAACGTACCCG TGTATCCCTGGGCCTGAAACAGCTGGGCGAAGATCCGTGGGTAGCTATCGCTAAA CGTTATCCGGAAGGTACCAAACTGACTGGTCGCGTGACCAACCTGACCGACTACG GCTGCTTCGTTGAAATCGAAGAAGGCGTTGAAGGCCTGGTACACGTTTCCGAAAT GGACTGGACCAACAAAAACATCCACCCGTCCAAAGTTGTTAACGTTGGCGATGTA GTGGAAGTTATGGTTCTGGATATCGACGAAGAACGTCGTCGTATCTCCCTGGGTCT GAAACAGTGCAAAGCTAACCCGTGGCAGCAGTTCGCGGAAACCCACAACAAGGG CGACCGTGTTGAAGGTAAAATCAAGTCTATCACTGACTTCGGTATCTTCATCGGCTT GGACGGCGGCATCGACGGCCTGGTTCACCTGTCTGACATCTCCTGGAACGTTGC AGGCGAAGAAGCAGTTCGTGAATACAAAAAAGGCGACGAAATCGCTGCAGTTGTT

	CTGCAGGTTGACGCAGAACGTGAACGTATCTCCCTGGGCGTTAAACAGCTCGCAG AAGATCCGTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCGTAACCGG TAAAGTAACTGCAGTTGACGCTAAAGGCGCAACCGTAGAACTGGCTGACGGCGTT GAAGGTTACCTGCGTGCTTCTGAAGCATCCCGTGACCGCGTTGAAGACGCTACCC TGGTTCTGAGCGTTGGCGACGAAGTTGAAGCTAAATTCACCGGCGTTGATCGTAA AAACCGCGCAATCAGCCTGTCTGTTCGTGCGAAAGACGAAGCTGACGAGAAAGAT GCAATCGCAACTGTTAACAAACAGGAAGATGCAAACTTCTCCAACAACGCAATGGC TGAAGCTTTCAAAGCAGCTAAAGGCGAG**CTGGTGCCGCGCGGCAGC**CTCGAG
pET28-rS1 R139A [*NcoI*, *XhoI*]	CC*ATGG*GAACTGAATCTTTGCGGCATGCTCAACTCTTTGAAGAGTCCTTAAAAGAA ATCGAAACCCGCCCGGGTTCTATCGTTCGTGGCGTTGTTGTTGCTATCGACAAAGA CGTAGTACTGGTTGACGCTGGTCTGAAATCTGAGTCCGCCATCCCGGCTGAGCAG TTCAAAAACGCCCAGGGCGAGCTGGAAATCCAGGTAGGTGACGAAGTTGACGTTG CTCTGGACGCAGTAGAAGACGGCTTCGGTGAAACTCTGCTGTCCCGTGAGAAAG CTAAACGTCACGAAGCCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAAC TGTTACCGGTGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAAC GGTATTCGTGCGTTCCTGCCAGGTTCTCTGGTAGACGTT**GCC**CCGGTGCGTGACA CTCTGCACCTGGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAG CGCAACAACGTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAG AGCGCGATCAGCTGCTGGAAAACCTGCAGGAAGGCATGGAAGTTAAAGGTATCGT TAAGAACCTCACTGACTACGGTGCATTCGTTGATCTGGGCGGCGTTGACGGCCTG CTGCACATCACTGACATGGCCTGGAAACGCGTTAAGCATCCGAGCGAAATCGTCA ACGTGGGCGACGAAATCACTGTTAAAGTGCTGAAGTTCGACCGCGAACGTACCCG TGTATCCCTGGGCCTGAAACAGCTGGGCGAAGATCCGTGGGTAGCTATCGCTAAA CGTTATCCGGAAGGTACCAAACTGACTGGTCGCGTGACCAACCTGACCGACTACG GCTGCTTCGTTGAAATCGAAGAAGGCGTTGAAGGCCTGGTACACGTTTCCGAAAT GGACTGGACCAACAAAAACATCCACCCGTCCAAAGTTGTTAACGTTGGCGATGTA GTGGAAGTTATGGTTCTGGATATCGACGAAGAACGTCGTCGTATCTCCCTGGGTCT GAAACAGTGCAAAGCTAACCCGTGGCAGCAGTTCGCGGAAACCCACAACAAGGG CGACCGTGTTGAAGGTAAAATCAAGTCTATCACTGACTTCGGTATCTTCATCGGCTT GGACGGCGGCATCGACGGCCTGGTTCACCTGTCTGACATCTCCTGGAACGTTGC AGGCGAAGAAGCAGTTCGTGAATACAAAAAAGGCGACGAAATCGCTGCAGTTGTT CTGCAGGTTGACGCAGAACGTGAACGTATCTCCCTGGGCGTTAAACAGCTCGCAG AAGATCCGTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCGTAACCGG TAAAGTAACTGCAGTTGACGCTAAAGGCGCAACCGTAGAACTGGCTGACGGCGTT GAAGGTTACCTGCGTGCTTCTGAAGCATCCCGTGACCGCGTTGAAGACGCTACCC TGGTTCTGAGCGTTGGCGACGAAGTTGAAGCTAAATTCACCGGCGTTGATCGTAA AAACCGCGCAATCAGCCTGTCTGTTCGTGCGAAAGACGAAGCTGACGAGAAAGAT GCAATCGCAACTGTTAACAAACAGGAAGATGCAAACTTCTCCAACAACGCAATGGC TGAAGCTTTCAAAGCAGCTAAAGGCGAG**CTGGTGCCGCGCGGCAGC**CTCGAG
pET28-rS1 R139K [*NcoI*, *XhoI*]	CC*ATGG*GAACTGAATCTTTGCGGCATGCTCAACTCTTTGAAGAGTCCTTAAAAGAA ATCGAAACCCGCCCGGGTTCTATCGTTCGTGGCGTTGTTGTTGCTATCGACAAAGA CGTAGTACTGGTTGACGCTGGTCTGAAATCTGAGTCCGCCATCCCGGCTGAGCAG TTCAAAAACGCCCAGGGCGAGCTGGAAATCCAGGTAGGTGACGAAGTTGACGTTG CTCTGGACGCAGTAGAAGACGGCTTCGGTGAAACTCTGCTGTCCCGTGAGAAAG CTAAACGTCACGAAGCCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAAC TGTTACCGGTGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAAC GGTATTCGTGCGTTCCTGCCAGGTTCTCTGGTAGACGTT**AAA**CCGGTGCGTGACA CTCTGCACCTGGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAG CGCAACAACGTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAG AGCGCGATCAGCTGCTGGAAAACCTGCAGGAAGGCATGGAAGTTAAAGGTATCGT

	TAAGAACCTCACTGACTACGGTGCATTCGTTGATCTGGGCGGCGTTGACGGCCTG CTGCACATCACTGACATGGCCTGGAAACGCGTTAAGCATCCGAGCGAAATCGTCA ACGTGGGCGACGAAATCACTGTTAAAGTGCTGAAGTTCGACCGCGAACGTACCCG TGTATCCCTGGGCCTGAAACAGCTGGGCGAAGATCCGTGGGTAGCTATCGCTAAA CGTTATCCGGAAGGTACCAAACTGACTGGTCGCGTGACCAACCTGACCGACTACG GCTGCTTCGTTGAAATCGAAGAAGGCGTTGAAGGCCTGGTACACGTTTCCGAAAT GGACTGGACCAACAAAAACATCCACCCGTCCAAAGTTGTTAACGTTGGCGATGTA GTGGAAGTTATGGTTCTGGATATCGACGAAGAACGTCGTCGTATCTCCCTGGGTCT GAAACAGTGCAAAGCTAACCCGTGGCAGCAGTTCGCGGAAACCCACAACAAGGG CGACCGTGTTGAAGGTAAAATCAAGTCTATCACTGACTTCGGTATCTTCATCGGCTT GGACGGCGGCATCGACGGCCTGGTTCACCTGTCTGACATCTCCTGGAACGTTGC AGGCGAAGAAGCAGTTCGTGAATACAAAAAAGGCGACGAAATCGCTGCAGTTGTT CTGCAGGTTGACGCAGAACGTGAACGTATCTCCCTGGGCGTTAAACAGCTCGCAG AAGATCCGTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCGTAACCGG TAAAGTAACTGCAGTTGACGCTAAAGGCGCAACCGTAGAACTGGCTGACGGCGTT GAAGGTTACCTGCGTGCTTCTGAAGCATCCCGTGACCGCGTTGAAGACGCTACCC TGGTTCTGAGCGTTGGCGACGAAGTTGAAGCTAAATTCACCGGCGTTGATCGTAA AAACCGCGCAATCAGCCTGTCTGTTCGTGCGAAAGACGAAGCTGACGAGAAAGAT GCAATCGCAACTGTTAACAAACAGGAAGATGCAAACTTCTCCAACAACGCAATGGC TGAAGCTTTCAAAGCAGCTAAAGGCGAG**CTGGTGCCGCGCGGCAGC**CTCGAG
pTAC-rS1 [*XhoI*, *SphI*]	*ATG*AAGCTTCCTCGAGAGACTGAATCTTTTGCTCAACTCTTTGAAGAGTCCTTAAAA GAAATCGAAACCCGCCCGGGTTCTATCGTTCGTGGCGTTGTTGTTGCTATCGACAA AGACGTAGTACTGGTTGACGCTGGTCTGAAATCTGAGTCCGCCATCCCGGCTGAG CAGTTCAAAAACGCCCAGGGCGAGCTGGAAATCCAGGTAGGTGACGAAGTTGAC GTTGCTCTGGACGCAGTAGAAGACGGCTTCGGTGAAACTCTGCTGTCCCGTGAG AAAGCTAAACGTCACGAAGCCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTG AAACTGTTACCGGTGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCT GAACGGTATTCGTGCGTTCCTGCCAGGTTCTCTGGTAGACGTTCGTCCGGTGCGT GACACTCTGCACCTGGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCA GAAGCGCAACAACGTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGC GCAGAGCGCGATCAGCTGCTGGAAAACCTGCAGGAAGGCATGGAAGTTAAAGGT ATCGTTAAGAACCTCACTGACTACGGTGCATTCGTTGATCTGGGCGGCGTTGACG GCCTGCTGCACATCACTGACATGGCCTGGAAACGCGTTAAGCATCCGAGCGAAAT CGTCAACGTGGGCGACGAAATCACTGTTAAAGTGCTGAAGTTCGACCGCGAACGT ACCCGTGTATCCCTGGGCCTGAAACAGCTGGGCGAAGATCCGTGGGTAGCTATCG CTAAACGTTATCCGGAAGGTACCAAACTGACTGGTCGCGTGACCAACCTGACCGA CTACGGCTGCTTCGTTGAAATCGAAGAAGGCGTTGAAGGCCTGGTACACGTTTCC GAAATGGACTGGACCAACAAAAACATCCACCCGTCCAAAGTTGTTAACGTTGGCG ATGTAGTGGAAGTTATGGTTCTGGATATCGACGAAGAACGTCGTCGTATCTCCCTG GGTCTGAAACAGTGCAAAGCTAACCCGTGGCAGCAGTTCGCGGAAACCCACAAC AAGGGCGACCGTGTTGAAGGTAAAATCAAGTCTATCACTGACTTCGGTATCTTCAT CGGCTTGGACGGCGGCATCGACGGCCTGGTTCACCTGTCTGACATCTCCTGGAA CGTTGCAGGCGAAGAAGCAGTTCGTGAATACAAAAAAGGCGACGAAATCGCTGCA GTTGTTCTGCAGGTTGACGCAGAACGTGAACGTATCTCCCTGGGCGTTAAACAGC TCGCAGAAGATCCGTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCGT AACCGGTAAAGTAACTGCAGTTGACGCTAAAGGCGCAACCGTAGAACTGGCTGAC GGCGTTGAAGGTTACCTGCGTGCTTCTGAAGCATCCCGTGACCGCGTTGAAGACG CTACCCTGGTTCTGAGCGTTGGCGACGAAGTTGAAGCTAAATTCACCGGCGTTGA TCGTAAAAACCGCGCAATCAGCCTGTCTGTTCGTGCGAAAGACGAAGCTGACGAG

	AAAGATGCAATCGCAACTGTTAACAAACAGGAAGATGCAAACTTCTCCAACAACGC AATGGCTGAAGCTTTCAAAGCAGCTAAAGGCGAGTGCATGCACGTAGAG
S1 D1 [*NcoI*, *XhoI*]	CC*ATGG*AGTCCTTAAAAGAAATCGAAACCCGCCCGGGTTCTATCGTTCGTGGCGTT GTTGTTGCTATCGACAAAGACGTAGTACTGGTTGACGCTGGTCTGAAATCTGAGTC CGCCATCCCGGCTGAGCAGTTCAAAAACGCCCAGGGCGAGCTGGAAATCCAGGT AGGTGACGAAGTTGACGTTGCTCTGGACGCAGTAGAAGACGGCTTCGGTGAAAC TCTGCTGTCCCGTGAGAAAGCTAAACGTCACGAAGCC**CTGGTGCCGCGCGGCAG** **C**CTCGAG
S1 D2 [*NcoI*, *XhoI*]	CC*ATGG*CCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAACTGTTACCGG TGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAACGGTATTCGTG CGTTCCTGCCAGGTTCTCTGGTAGACGTTCGTCCGGTGCGTGACACTCTGCACCT GGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAGCGCAACAAC GTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAGAG**CTGGTGC** **CGCGCGGCAGC**CTCGAG
S1 D2 R139A [*NcoI*, *XhoI*]	CC*ATGG*CCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAACTGTTACCGG TGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAACGGTATTCGTG CGTTCCTGCCAGGTTCTCTGGTAGACGTT**GCC**CCGGTGCGTGACACTCTGCACCT GGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAGCGCAACAAC GTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAGAG**CTGGTGC** **CGCGCGGCAGC**CTCGAG
S1 D2 R139K [*NcoI*, *XhoI*]	CC*ATGG*CCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAACTGTTACCGG TGTTATCAACGGCAAAGTTAAGGGCGGCTTCACTGTTGAGCTGAACGGTATTCGTG CGTTCCTGCCAGGTTCTCTGGTAGACGTT**AAA**CCGGTGCGTGACACTCTGCACCT GGAAGGCAAAGAGCTTGAATTTAAAGTAATCAAGCTGGATCAGAAGCGCAACAAC GTTGTTGTTTCTCGTCGTGCCGTTATCGAATCCGAAAACAGCGCAGAG**CTGGTGC** **CGCGCGGCAGC**CTCGAG
S1 D3 [*NcoI*, *XhoI*]	CC*ATGG*CCCGCGATCAGCTGCTGGAAAACCTGCAGGAAGGCATGGAAGTTAAAG GTATCGTTAAGAACCTCACTGACTACGGTGCATTCGTTGATCTGGGCGGCGTTGAC GGCCTGCTGCACATCACTGACATGGCCTGGAAACGCGTTAAGCATCCGAGCGAAA TCGTCAACGTGGGCGACGAAATCACTGTTAAAGTGCTGAAGTTCGACCGCGAACG TACCCGTGTATCCCTGGGCCTGAAACAGCTGGGCGAAGATCCG**CTGGTGCCGCG** **CGGCAGC**CTCGAG
S1 D4 [*NcoI*, *XhoI*]	CC*ATGG*CCTGGGTAGCTATCGCTAAACGTTATCCGGAAGGTACCAAACTGACTGGT CGCGTGACCAACCTGACCGACTACGGCTGCTTCGTTGAAATCGAAGAAGGCGTTG AAGGCCTGGTACACGTTTCCGAAATGGACTGGACCAACAAAAACATCCACCCGTC CAAAGTTGTTAACGTTGGCGATGTAGTGGAAGTTATGGTTCTGGATATCGACGAAG AACGTCGTCGTATCTCCCTGGGTCTGAAACAGTGCAAAGCTAACCCG**CTGGTGCC** **GCGCGGCAGC**CTCGAG
S1 D5 [*NcoI*, *XhoI*]	CC*ATGG*CCTGGCAGCAGTTCGCGGAAACCCACAACAAGGGCGACCGTGTTGAAG GTAAAATCAAGTCTATCACTGACTTCGGTATCTTCATCGGCTTGGACGGCGGCATC GACGGCCTGGTTCACCTGTCTGACATCTCCTGGAACGTTGCAGGCGAAGAAGCA GTTCGTGAATACAAAAAAGGCGACGAAATCGCTGCAGTTGTTCTGCAGGTTGACG CAGAACGTGAACGTATCTCCCTGGGCGTTAAACAGCTCGCAGAAGATCCG**CTGGT** **GCCGCGCGGCAGC**CTCGAG
S1 D6 [*NcoI*, *XhoI*]	CC*ATGG*CCTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCGTAACCGG TAAAGTAACTGCAGTTGACGCTAAAGGCGCAACCGTAGAACTGGCTGACGGCGTT GAAGGTTACCTGCGTGCTTCTGAAGCATCCCGTGACCGCGTTGAAGACGCTACCC TGGTTCTGAGCGTTGGCGACGAAGTTGAAGCTAAATTCACCGGCGTTGATCGTAA AAACCGCGCAATCAGCCTGTCTGTTCGTGCGAAAGACGAAGCTGACGAGAAA**CTG** **GTGCCGCGCGGCAGC**CTCGAG

S1 domain of PNPase [*NcoI*, *XhoI*]	CC*ATGG*CAGAAATCGAAGTGGGCCGCGTCTACACTGGTAAAGTGACCCGTATCGT TGACTTTGGCGCATTTGTTGCCATCGGCGGCGGTAAAGAAGGTCTGGTCCACATC TCTCAAATCGCTGACAAACGCGTTGAGAAAGTGACCGATTACCTGCAGATGGGTC AGGAAGTACCGGTGAAAGTTCTGGAAGTTGATCGCCAGGGCCGTATCCGTCTGAG CATTAAAGAAGCGACTGAGCAGTCTCAACCTGCTGCA**CTGGTGCCGCGCGGCAG** **C**CTCGAG
pET28 ARH1 [*NcoI*, *XhoI*]	CC*ATGG*AAAAATACGTCGCCGCGATGGTTTTGTCAGCTGCTGGCGATGCTTTGGG ATATTATAATGGAAAGTGGGAATTTCTTCAGGACGGGGAGAAAATTCATCGTCAACT GGCTCAATTAGGGGGGCTGGATGCTCTGGACGTTGGCCGTTGGCGTGTGTCTGA TGATACTGTCATGCACTTGGCAACAGCCGAGGCTTTGGTCGAGGCCGGAAAGGCT CCAAAACTGACTCAGCTTTATTATTTGTTAGCCAAGCACTATCAGGATTGCATGGAA GATATGGACGGTCGCGCACCCGGGGGTGCGTCTGTACACAACGCGATGCAGCTT AAACCTGGGAAACCGAATGGCTGGCGTATCCCATTTAACTCGCATGAAGGAGGGT GTGGCGCGGCGATGCGCGCGATGTGTATCGGTTTGCGTTTTCCGCATCACTCTCA ATTAGACACACTGATCCAAGTATCGATCGAGTCAGGACGTATGACCCATCATCACC CGACAGGGTACCTTGGCGCACTTGCGTCCGCCTTATTCACGGCCTATGCGGTAAA TAGCCGCCCTCCATTGCAGTGGGGTAAGGGACTTATGGAGCTTTTGCCAGAGGCT AAAAAATACATTGTCCAATCCGGGTACTTTGTGGAAGAAAATTTACAGCATTGGTCT TATTTTCAAACGAAGTGGGAAAACTATCTTAAACTGCGTGGAATCTTGGACGGCGA GAGTGCTCCAACATTCCCTGAATCTTTTGGCGTTAAAGAGCGCGACCAGTTCTACA CTTCGTTGTCATATAGTGGCTGGGGCGGTTCATCTGGGCATGATGCCCCCATGATC GCGTATGACGCGGTGCTGGCGGCGGGAGACTCCTGGAAAGAGCTTGCGCACCG CGCCTTCTTTCACGGAGGTGACTCGGATTCGACCGCAGCCATTGCTGGATGTTGG TGGGGCGTCATGTACGGATTTAAGGGCGTCAGCCCCAGCAACTACGAAAAATTAG AGTATCGCAATCGCCTTGAGGAAACAGCTCGCGCACTTTACTCGCTGGGTAGTAAA GAAGACACTGTTATCTCGCTG**CTGGTGCCGCGCGGCAGC**CTCGAG
pET ARH1 D55A, D56A [*NcoI*, *XhoI*]	CC*ATGG*AAAAATACGTCGCCGCGATGGTTTTGTCAGCTGCTGGCGATGCTTTGGG ATATTATAATGGAAAGTGGGAATTTCTTCAGGACGGGGAGAAAATTCATCGTCAACT GGCTCAATTAGGGGGGCTGGATGCTCTGGACGTTGGCCGTTGGCGTGTGTCT**GC** **GGCG**ACTGTCATGCACTTGGCAACAGCCGAGGCTTTGGTCGAGGCCGGAAAGGC TCCAAAACTGACTCAGCTTTATTATTTGTTAGCCAAGCACTATCAGGATTGCATGGA AGATATGGACGGTCGCGCACCCGGGGGTGCGTCTGTACACAACGCGATGCAGCT TAAACCTGGGAAACCGAATGGCTGGCGTATCCCATTTAACTCGCATGAAGGAGGG TGTGGCGCGGCGATGCGCGCGATGTGTATCGGTTTGCGTTTTCCGCATCACTCTC AATTAGACACACTGATCCAAGTATCGATCGAGTCAGGACGTATGACCCATCATCAC CCGACAGGGTACCTTGGCGCACTTGCGTCCGCCTTATTCACGGCCTATGCGGTAA ATAGCCGCCCTCCATTGCAGTGGGGTAAGGGACTTATGGAGCTTTTGCCAGAGGC TAAAAAATACATTGTCCAATCCGGGTACTTTGTGGAAGAAAATTTACAGCATTGGTC TTATTTTCAAACGAAGTGGGAAAACTATCTTAAACTGCGTGGAATCTTGGACGGCG AGAGTGCTCCAACATTCCCTGAATCTTTTGGCGTTAAAGAGCGCGACCAGTTCTAC ACTTCGTTGTCATATAGTGGCTGGGGCGGTTCATCTGGGCATGATGCCCCCATGAT CGCGTATGACGCGGTGCTGGCGGCGGGAGACTCCTGGAAAGAGCTTGCGCACC GCGCCTTCTTTCACGGAGGTGACTCGGATTCGACCGCAGCCATTGCTGGATGTTG GTGGGGCGTCATGTACGGATTTAAGGGCGTCAGCCCCAGCAACTACGAAAAATTA GAGTATCGCAATCGCCTTGAGGAAACAGCTCGCGCACTTTACTCGCTGGGTAGTA AAGAAGACACTGTTATCTCGCTG**CTGGTGCCGCGCGGCAGC**CTCGAG
pET rL2 [*NcoI, XhoI*]	CCATGGGCGCAGTTGTTAAATGTAAACCGACATCTCCGGGTCGTCGCCACGTAGT TAAAGTGGTTAACCCTGAGCTGCACAAGGGCAAACCTTTTGCTCCGTTGCTGGAA AAAAACAGCAAATCCGGTGGTCGTAACAACAATGGCCGTATCACCACTCGTCATAT CGGTGGTGGCCACAAGCAGGCTTACCGTATTGTTGACTTCAAACGCAACAAAGAC

	GGTATCCCGGCAGTTGTTGAACGTCTTGAGTACGATCCGAACCGTTCCGCGAACA TCGCGCTGGTTCTGTACAAAGACGGTGAACGCCGTTACATCCTGGCCCCTAAAGG CCTGAAAGCTGGCGACCAGATTCAGTCTGGCGTTGATGCTGCAATCAAACCAGGT AACACCCTGCCGATGCGCAACATCCCGGTTGGTTCTACTGTTCATAACGTAGAAAT GAAACCAGGTAAAGGCGGTCAGCTGGCACGTTCCGCTGGTACTTACGTTCAGATC GTTGCTCGTGATGGTGCTTATGTCACCCTGCGTCTGCGTTCTGGTGAAATGCGTAA AGTAGAAGCAGACTGCCGTGCAACTCTGGGCGAAGTTGGCAATGCTGAGCATATG CTGCGCGTTCTGGGTAAAGCAGGTGCTGCACGCTGGCGTGGTGTTCGTCCGACC GTTCGCGGTACCGCGATGAACCCGGTAGACCACCCACATGGTGGTGGTGAAGGT CGTAACTTTGGTAAGCACCCGGTAACTCCGTGGGGCGTTCAGACCAAAGGTAAGA AGACCCGCAGCAACAAGCGTACTGATAAATTCATCGTACGTCGCCGTAGCAAACTC GAG

Supplementary Table 9: Primers used in this study. Corresponding restriction site in bold, underlined; mutation in bold and red

Primer	Sequence (5' to 3')
Fwd Qβ T7	TAATACGACTCACTATTATCTTGATACTACCTTTAG
Rev Qβ	CATGATCAAATTGACCCAAAGTTTCAACGCTTTACGCG
Fwd RNAI T7	TAATACGACTCACTATAACAGTATTTGGTATC
Rev RNAI	ACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
Fwd Alt NcoI	ATCGAC**CCATGG**GAGAACTTATTACAGAATTATTTGACG
Rev Alt XhoI	ATTCGA**CTCGAG**GCTGCCGCGCGGCACCAGTCCTTGAACGAACTTGTAAGGCATG
Fwd ModA NcoI	ATCGA**CCATGG**GAAAATACTCAGTAATGCAACTAAAAG
Rev ModA XhoI	ATCGTA**CTCGAG**GCTGCCGCGCGGCACCAGTAGATTAAATCCTTCAAAATCAAG
Fwd ModB NcoI	ATCGAC**CCATGG**GAATTATTAATCTTGCAGATGTTG
Rev ModB XhoI	ACTTAG**CTCGAG**GCTGCCGCGCGGCACCAGTTCAACTACCTCAATATCATAAAC
Fwd rS1 NcoI	ATCGAC**CCATGG**GAACTGAATCTTTTGCTCAACTCTTTGAAGAGTCC
Rev rS1 XhoI	ATTCGA**CTCGAG**GCTGCCGCGCGGCACCAGCTCGCCTTTAGCTGCTTTG
Fwd rS1-pTAC XhoI	ATGAAGCTTC**CTCGAG**AGACTGAATCTTTTGCTCAACTCTTTGAAGAGTCC
Rev rS1-pTAC SphI	CTCTACGT**GCATGC**ACTCGCCTTTAGCTGCTTTGAAAGCTTCAGCC
Fwd NcoI rS1 D1	ATCGAC**CCATGG**AGTCCTTAAAAGAAATCGAAACCCGCCCGGG
Rev XhoI rS1 D1	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGGGCTTCGTGACGTTTAGCTTTCTCACGGG
Fwd NcoI rS1 D2	ATCGAC**CCATGG**CCTGGATCACGCTGGAAAAAGCTTACGAAGATGCTGAAAC
Rev XhoI rS1 D2	GGTG**CTCGAG**GCTGCCGCGCGGCACCAGCTCTGCGCTGTTTTCGGATTCGATAACGGCAC
Fwd NcoI rS1 D3	ATCGAC**CCATGG**CCCGCGATCAGCTGCTGGAAAACCTGCAGGAAGG
Rev XhoI rS1 D3	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGCGGATCTTCGCCCAGCTGTTTCAGGCCCAGG
Fwd NcoI rS1 D4	ATCGAC**CCATGG**CCTGGGTAGCTATCGCTAAACGTTATCCGGAAGG
Rev XhoI rS1 D4	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGCGGGTTAGCTTTGCACTGTTTCAGACCCAGGGAG
Fwd NcoI rS1 D5	ATCGAC**CCATGG**CCTGGCAGCAGTTCGCGGAAACCCACAACAAGGGCGACCGTGTTG
Rev XhoI S1 D5	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGCGGATCTTCTGCGAGCTGTTTAACGCCCAGGGAGATACG
Fwd NcoI rS1 D6	ATCGAC**CCATGG**CCTTCAACAACTGGGTTGCTCTGAACAAGAAAGGCGCTATCG
Rev XhoI rS1 D6	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGTTTCTCGTCAGCTTCGTCTTTCGCACGAACAGACAGG
Fwd NcoI PNPase rS1 binding	ATCGAC**CCATGG**CAGAAATCGAAGTGGGCCGCGTCTACACTGGTAAAGTGACCCG
Rev XhoI PNPase rS1 binding	TGGTG**CTCGAG**GCTGCCGCGCGGCACCAGTGCAGCAGGTTGAGACTGCTCAGTCGCTTC
Fwd ARH1 NcoI	TGCAG**CCATGG**AAAAATACGTCGCCGCGATG
Rev ARH1 XhoI	GTGGTG**CTCGAG**GCTGCCGCGCGGCACCAG

Fwd rS1 R139A	CTGGTAGACGTT**GCC**CCGGTGCGTGACACTC
Fwd rS1 R139K	CTGGTAGACGTT**AAA**CCGGTGCGTGACACTC
Rev rS1 R139	AGAACCTGGCAGGAACGCACGAATACCG
Fwd ARH1 D55,56A	GGCCGTTGGCGTGTGTCT**GCGGCG**ACTGTCATGCACTTGGC
Rev ARH1 D55,56A	AACGTCCAGAGCATCCAGCCCCCCTAA
Fwd ModB R73A	CCTTATCAATTATAT**GCG**GGTATATCAAAATCG
Rev ModB R73A	CGATTTAGTTAAATGCTTTTTAATGATTTC
Fwd ModB G74A	GACAAAAGAACTCATTAAAGATTTAC
Rev ModB G74A	GATTTTGATAT**TGCCGC**ATATAATTGATAAGGCG
Fwd ModB DS_SPCas	AATTATATCGGTTTTAGAGCTATGCTGTTTTGAATGGTCC
Rev ModB DS_SPCas	GATAAGGCGAGCTAGCACTGTACCTAGGACTGAGC
Fwd ModB amplification T4 genome	CCAAGAATGGTCATCTGGTTTATTAG
Rev ModB amplification T4 genome	CCGCCTTGGGCTCCCTGG
Fwd sequencing ModB T4 genome	CAGTTATCTATAAAAGCTGAAAG
Rev sequencing ModB T4 genome	CTTTCCAATTGGAATCATCCATTC
rpsA homologous downstream fwd	TTCTCTGACTCTTCGGGATTTTTATTC
rpsA homologous downstream rev	AGGCAAATTAAGCGGCTGCTG
Terminator region fwd	TTCTCTGACTCTTCGGGATTTTTATTC
Terminator region rev	AGGACGAAACCTGCAATCTGTC
FRT pKD4 fwd	TCGGAATAAAAATCCCGAAGAGTCAGAGAAGTCCATATGAATATCCTCC TTAGTTC
FRT pKD4 rev	GTTTACTTGACAGATTGCAGGTTTCGTCCTGTGTAGGCTGGAGCTGCT TC
5_70 left rev rpsA	AGGACGAAACCTGCAATCTGTC
5_fwd_rS1 amplification	GGCGTTGATCGTAAAAACCGC
Fwd NcoI rL2	**CCATGG**GCGCAGTTGTTAAATGTAAACCG
Rev XhoI rL2	**CTCGAG**TTTGCTACGGCGACGTACGATG
adenylated RNA-3'-adapter	/5rApp/CNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/
RT primer	CAGACGTGTGCTCTTCCGAT
cDNA anchor fwd	ACACGACGCTCTTCCGATCTGGG
cDNA anchor rev	/5Phos/CAGATCGGAAGAGCGTCGTGTCCC/3SpC3/
qPCR acpP fwd	CGTGGTAAGACCTGCCGG
qPCR acpP rev	CTCAACGGTGTCAAGAGAATCCAAAAC
qPCR gadY fwd	GAGCACAAAGTTTCCCGTGC
qPCR gadY rev	AAACCCGGCATAGGGGACC
qPCR mcaS fwd	AAAATAGAGTCTGTCGACATCCGC

qPCR mcaS rev	CACCGGCGCAGAGGAGAC
qPCR oxyS fwd	AAAAGCGGATCCTGGAGATCC
qPCR oxyS rev	GAAACGGAGCGGCACCTC
qPCR rnaC fwd	CGTTGCGGCAACCTTGTC
qPCR rnaC rev	AAAAATATTGAGTAGCGTCAACTAC

Supplementary Table 10: Strains and plasmids used in this study.

Name	Description	Reference or resource
***E. coli* strain B**	*E. coli* strain applied for bacteriophage T4 infection	DMSZ, *Escherichia coli* (Migula 1895) Castellani and Chalmers 1919 (DSM 613, ATCC 11303)
***E. coli* strain B pTAC rS1**	*E. coli* strain B expressing His-tagged rS1 under the control of *E. coli* RNA polymerase promoter	This study
***E. coli* BL21 (DE3) pET16 RNase E (1-529)**	*E. coli* strain expressing His-tagged catalytic domain of RNase E (1-529)	Plasmid was a kind gift from Prof. Dr. Ben Luisi [67]
***E. coli* BL21 (DE3) pET28 NudC V157A, E174A, E177A, E178A**	*E. coli* strain expressing His-tagged inactive Mutant of NudC	[3]
***E. coli* BL21 (DE3) pET28 rS1**	*E. coli* strain expressing His-tagged rS1	This study
***E. coli* BL21 (DE3) pET28 rS1 R139K**	*E. coli* strain expressing His-tagged rS1 R139K variant	This study
***E. coli* BL21 (DE3) pET28 rS1 R139A**	*E. coli* strain expressing His-tagged rS1 R139A variant	This study
***E. coli* BL21 (DE3) pET28 rS1 D1**	*E. coli* strain expressing His-tagged rS1 D1	This study
***E. coli* BL21 (DE3) pET28 rS1 D2**	*E. coli* strain expressing His-tagged rS1 D2	This study
***E. coli* BL21 (DE3) pET28 rS1 D2 R139K**	*E. coli* strain expressing His-tagged rS1 D2 R139K	This study
***E. coli* BL21 (DE3) pET28 rS1 D2 R139A**	*E. coli* strain expressing His-tagged rS1 D2 R139A	This study
***E. coli* BL21 (DE3) pET28 rS1 D3**	*E. coli* strain expressing His-tagged rS1 D3	This study
***E. coli* BL21 (DE3) pET28 rS1 D4**	*E. coli* strain expressing His-tagged rS1 D4	This study
***E. coli* BL21 (DE3) pET28 rS1 D5**	*E. coli* strain expressing His-tagged rS1 D5	This study
***E. coli* BL21 (DE3) pET28 rS1 D6**	*E. coli* strain expressing His-tagged rS1 D6	This study
***E. coli* BL21 (DE3) pET28 Alt**	*E. coli* strain expressing His-tagged Alt	This study
***E. coli* BL21 (DE3) pET28 ModA**	*E. coli* strain expressing His-tagged ModA	This study
***E. coli* BL21 (DE3) pET28 ModB**	*E. coli* strain expressing His-tagged ModB	This study
***E. coli* BL21 (DE3) pET28 ModB R73A, G74A**	*E. coli* strain expressing His-tagged ModB with point mutations R73A and G74A	This study
***E. coli* BL21 (DE3) pET28 NudC**	*E. coli* strain expressing His-tagged NudC	[3]

***E. coli* BL21 (DE3) pET28 PNPase S1 domain**	*E. coli* strain expressing His-tagged PNPase S1 domain	This study
***E. coli* BL21 (DE3) pET28 ARH1**	*E. coli* strain expressing His-tagged ARH1	This study
***E. coli* BL21 (DE3) pET28 ARH1 D55A, D56A**	*E. coli* strain expressing His-tagged ARH1 D55A, D56A	This study
***E. coli* DHα DS_SPCas_ModB**	*E. coli* strain expressing CRISPR-Cas9 system for cleavage of *modB*	This study
***E. coli* DHα DS_SPCas_ModB pET28 ModB R73A, G74A**	*E. coli* strain for editing of *modB* within T4 phage genome	This study
***E. coli* BL21 (DE3) pET28 rL2**	*E. coli* strain expressing His-tagged rL2	This study
***E. coli* B strain with endogenously His-tagged rS1**	*E. coli* strain with endogenous expression of rS1 with a His-tag fusion at the C-terminus	This study
T4 WT	Wild-type bacteriophage T4	Escherichia phage T4, DSM 4505, DSMZ, Braunschweig, Germany)
T4 ModB R73A, G74A	T4 phage mutant carrying inactive ModB version ModB R73A, G74A	This study

6.7. References

1 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004). https://doi.org:10.1128/JB.186.21.7262-7272.2004

2 Koch, T., Raudonikiene, A., Wilkens, K. & Rüger, W. Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of E. coli RNA polymerase modulates T4 "early" transcription. *Gene Expr* **4**, 253-264 (1995).

3 Cahova, H., Winz, M. L., Höfer, K., Nübel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

4 Chen, Y. G., Kowtoniuk, W. E., Agarwal, I., Shen, Y. & Liu, D. R. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol* **5**, 879-881 (2009). https://doi.org:10.1038/nchembio.235

5 Jiao, X. *et al.* 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. *Cell* **168**, 1015-1027 e1010 (2017). https://doi.org:10.1016/j.cell.2017.02.019

6 Fehr, A. R. *et al.* The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. *Genes Dev* **34**, 341-359 (2020). https://doi.org:10.1101/gad.334425.119

7 Cohen, M. S. & Chang, P. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nat Chem Biol* **14**, 236-243 (2018). https://doi.org:10.1038/nchembio.2568

8 Simon, N. C., Aktories, K. & Barbieri, J. T. Novel bacterial ADP-ribosylating toxins: structure and function. *Nat Rev Microbiol* **12**, 599-611 (2014). https://doi.org:10.1038/nrmicro3310

9 Lüscher, B. *et al.* ADP-Ribosylation, a Multifaceted Posttranslational Modification Involved in the Control of Cell Physiology in Health and Disease. *Chem Rev* **118**, 1092-1136 (2018). https://doi.org:10.1021/acs.chemrev.7b00122

10 Morales-Filloy, H. G. *et al.* The 5' NAD Cap of RNAIII Modulates Toxin Production in Staphylococcus aureus Isolates. *J Bacteriol* **202** (2020). https://doi.org:10.1128/JB.00591-19

11 Frindert, J. *et al.* Identification, Biosynthesis, and Decapping of NAD-Capped RNAs in B. subtilis. *Cell Rep* **24**, 1890-1901 e1898 (2018). https://doi.org:10.1016/j.celrep.2018.07.047

12 Ruiz-Larrabeiti, O. *et al.* NAD+capping of RNA in Archaea and Mycobacteria. *bioRxiv*, 2021.2012.2014.472595 (2021). https://doi.org:10.1101/2021.12.14.472595

13 Gomes-Filho, J. V. *et al.* Identification of NAD-RNAs and ADPR-RNA decapping in the archaeal model organisms Sulfolobus acidocaldarius and Haloferax volcanii. *bioRxiv*, 2022.2011.2002.514978 (2022). https://doi.org:10.1101/2022.11.02.514978

14 Walters, R. W. *et al.* Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114**, 480-485 (2017). https://doi.org:10.1073/pnas.1619369114

15 Zhang, Y. *et al.* Extensive 5'-surveillance guards against non-canonical NAD-caps of nuclear mRNAs in yeast. *Nat Commun* **11**, 5508 (2020). https://doi.org:10.1038/s41467-020-19326-3

16 Wang, J. *et al.* Quantifying the RNA cap epitranscriptome reveals novel caps in cellular and viral RNA. *Nucleic Acids Res* **47**, e130 (2019). https://doi.org:10.1093/nar/gkz751

17 Zhang, H. *et al.* NAD tagSeq reveals that NAD(+)-capped RNAs are mostly produced from a large number of protein-coding genes in Arabidopsis. *Proc Natl Acad Sci U S A* **116**, 12072-12077 (2019). https://doi.org:10.1073/pnas.1903683116

18 Wang, Y. *et al.* NAD(+)-capped RNAs are widespread in the Arabidopsis transcriptome and can probably be translated. *Proc Natl Acad Sci U S A* **116**, 12094-12102 (2019). https://doi.org:10.1073/pnas.1903682116

19 Dong, H. *et al.* NAD(+)-capped RNAs are widespread in rice (Oryza sativa) and spatiotemporally modulated during development. *Sci China Life Sci* **65**, 2121-2124 (2022). https://doi.org:10.1007/s11427-021-2113-7

20 Wolfram-Schauerte, M. & Höfer, K. NAD-capped RNAs - a redox cofactor meets RNA. *Trends Biochem Sci* (2022). https://doi.org:10.1016/j.tibs.2022.08.004

21 Höfer, K. & Jäschke, A. Epitranscriptomics: RNA Modifications in Bacteria and Archaea. *Microbiol Spectr* **6** (2018). https://doi.org:10.1128/microbiolspec.RWR-0015-2017

22 Tsuge, H. *et al.* Structural basis of actin recognition and arginine ADP-ribosylation by Clostridium perfringens iota-toxin. *Proc Natl Acad Sci U S A* **105**, 7399-7404 (2008). https://doi.org:10.1073/pnas.0801215105

23 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

24 Rohrer, H., Zillig, W. & Mailhammer, R. ADP-ribosylation of DNA-dependent RNA polymerase of Escherichia coli by an NAD+: protein ADP-ribosyltransferase from bacteriophage T4. *Eur J Biochem* **60**, 227-238 (1975). https://doi.org:10.1111/j.1432-1033.1975.tb20995.x

25 Depping, R., Lohaus, C., Meyer, H. E. & Rüger, W. The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: target proteins identified. *Biochem Biophys Res Commun* **335**, 1217-1223 (2005). https://doi.org:10.1016/j.bbrc.2005.08.023

26 Skorko, R., Zillig, W., Rohrer, H., Fujiki, H. & Mailhammer, R. Purification and properties of the NAD+: protein ADP-ribosyltransferase responsible for the T4-phage-induced modification of the alpha subunit of DNA-dependent RNA polymerase of Escherichia coli. *Eur J Biochem* **79**, 55-66 (1977). https://doi.org:10.1111/j.1432-1033.1977.tb11783.x

27 Tiemann, B., Depping, R. & Rüger, W. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. *Gene Expr* **8**, 187-196 (1999).

28 Rankin, P. W., Jacobson, E. L., Benjamin, R. C., Moss, J. & Jacobson, M. K. Quantitative studies of inhibitors of ADP-ribosylation in vitro and in vivo. *J Biol Chem* **264**, 4312-4317 (1989).

29 Goelz, S. & Steitz, J. A. Escherichia coli ribosomal protein S1 recognizes two sites in bacteriophage Qbeta RNA. *J Biol Chem* **252**, 5177-5179 (1977).

30 Cervantes-Laurean, D., Jacobson, E. L. & Jacobson, M. K. Glycation and glycoxidation of histones by ADP-ribose. *J Biol Chem* **271**, 10461-10469 (1996). https://doi.org:10.1074/jbc.271.18.10461

31 Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* **11**, 1064-1070 (2014). https://doi.org:10.1038/nmeth.3092

32 Gehrig, P. M. *et al.* Gas-Phase Fragmentation of ADP-Ribosylated Peptides: Arginine-Specific Side-Chain Losses and Their Implication in Database Searches. *J Am Soc Mass Spectrom* **32**, 157-168 (2021). https://doi.org:10.1021/jasms.0c00040

33 Winz, M. L. *et al.* Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq. *Nat Protoc* **12**, 122-149 (2017). https://doi.org:10.1038/nprot.2016.163

34 Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *Acs Synth Biol* **6**, 1952-1961 (2017). https://doi.org:10.1021/acssynbio.7b00179

35 Zhang, H. *et al.* Use of NAD tagSeq II to identify growth phase-dependent alterations in E. coli RNA NAD(+) capping. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2026183118

36 Bird, J. G. *et al.* The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. *Nature* **535**, 444-447 (2016). https://doi.org:10.1038/nature18622

37 Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M. & Murzin, A. G. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* **88**, 235-242 (1997). https://doi.org:10.1016/s0092-8674(00)81844-9

38 Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* **5**, 593-599 (2009). https://doi.org:10.1038/nchembio.186

39 Wolfram-Schauerte, M., Pozhydaieva, N., Viering, M., Glatter, T. & Höfer, K. Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of E. coli on Proteome and Transcriptome Levels. *Viruses* **14**, 2502 (2022).

40 Diedrich, G. *et al.* Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding to A and P sites and peptidyl transfer. *Embo J* **19**, 5241-5250 (2000). https://doi.org:10.1093/emboj/19.19.5241

41 Nakagawa, A. *et al.* The three-dimensional structure of the RNA-binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. *Embo J* **18**, 1459-1467 (1999). https://doi.org:10.1093/emboj/18.6.1459

42 Beckert, B. *et al.* Structure of a hibernating 100S ribosome reveals an inactive conformation of the ribosomal protein S1. *Nat Microbiol* **3**, 1115-1121 (2018). https://doi.org:10.1038/s41564-018-0237-0

43 Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341 (2018). https://doi.org:10.1038/nrm.2017.130

44 Drygin, Y. F. Natural covalent complexes of nucleic acids and proteins: some comments on practice and theoryon the path from well-known complexes to new ones. *Nucleic Acids Res* **26**, 4791-4796 (1998). https://doi.org:10.1093/nar/26.21.4791

45 Ibba, M. & Soll, D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **69**, 617-650 (2000). https://doi.org:10.1146/annurev.biochem.69.1.617

46 Rekosh, D. M., Russell, W. C., Bellet, A. J. & Robinson, A. J. Identification of a protein linked to the ends of adenovirus DNA. *Cell* **11**, 283-295 (1977). https://doi.org:10.1016/0092-8674(77)90045-9

47 Rothberg, P. G., Harris, T. J., Nomoto, A. & Wimmer, E. O4-(5'-uridylyl)tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proc Natl Acad Sci U S A* **75**, 4868-4872 (1978). https://doi.org:10.1073/pnas.75.10.4868

48 Uzan, M. & Miller, E. S. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. *Virol J* **7**, 360 (2010). https://doi.org:10.1186/1743-422X-7-360

49 Ke, Y., Zhang, J., Lv, X., Zeng, X. & Ba, X. Novel insights into PARPs in gene expression: regulation of RNA metabolism. *Cell Mol Life Sci* **76**, 3283-3299 (2019). https://doi.org:10.1007/s00018-019-03120-6

50 Otten, E. G. *et al.* Ubiquitylation of lipopolysaccharide by RNF213 during bacterial infection. *Nature* **594**, 111-116 (2021). https://doi.org:10.1038/s41586-021-03566-4

51 Flynn, R. A. *et al.* Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* **184**, 3109-3124 e3122 (2021). https://doi.org:10.1016/j.cell.2021.04.023

52 Höfer, K., Abele, F., Schlotthauer, J. & Jäschke, A. Synthesis of 5'-NAD-Capped RNA. *Bioconjugate Chem* **27**, 874-877 (2016). https://doi.org:10.1021/acs.bioconjchem.6b00072

53 Ladner, C. L., Yang, J., Turner, R. J. & Edwards, R. A. Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal Biochem* **326**, 13-20 (2004). https://doi.org:10.1016/j.ab.2003.10.047

54 Hsia, J. A. *et al.* Amino acid-specific ADP-ribosylation. Sensitivity to hydroxylamine of [cysteine(ADP-ribose)]protein and [arginine(ADP-ribose)]protein linkages. *J Biol Chem* **260**, 16187-16191 (1985).

55 McDonald, L. J., Wainschel, L. A., Oppenheimer, N. J. & Moss, J. Amino acid-specific ADP-ribosylation: structural characterization and chemical differentiation of ADP-ribose-cysteine adducts formed nonenzymatically and in a pertussis toxin-catalyzed reaction. *Biochemistry-Us* **31**, 11881-11887 (1992). https://doi.org:10.1021/bi00162a029

56 Silberklang, M., Gillum, A. M. & RajBhandary, U. L. The use of nuclease P1 in sequence analysis of end group labeled RNA. *Nucleic Acids Res* **4**, 4091-4108 (1977). https://doi.org:10.1093/nar/4.12.4091

57 Abplanalp, J. *et al.* Proteomic analyses identify ARH3 as a serine mono-ADP-ribosylhydrolase. *Nat Commun* **8**, 2055 (2017). https://doi.org:10.1038/s41467-017-02253-1

58 Banasik, M., Komura, H., Shimoyama, M. & Ueda, K. Specific inhibitors of poly(ADP-ribose) synthetase and mono(ADP-ribosyl)transferase. *J Biol Chem* **267**, 1569-1575 (1992).

59 MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966-968 (2010). https://doi.org:10.1093/bioinformatics/btq054

60 Wolfram-Schauerte, M., Pozhydaieva, Nadiia, Grawenhoff, Julia, Welp, Luisa, Silbern, Ivan, Wulf, Alexander, Billau, Franziska, Glatter, Timo, Urlaub, Henning, Jäschke, Andres, Höfer, Katharina. (Zenodo, 2023).

61 Dell'Aquila, G. *et al.* Mobilization and Cellular Distribution of Phosphate in the Diatom Phaeodactylum tricornutum. *Front Plant Sci* **11**, 579 (2020). https://doi.org:10.3389/fpls.2020.00579

62 Gibson, B. A., Conrad, L. B., Huang, D. & Kraus, W. L. Generation and Characterization of Recombinant Antibody-like ADP-Ribose Binding Proteins. *Biochemistry-Us* **56**, 6305-6316 (2017). https://doi.org:10.1021/acs.biochem.7b00670

63 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

64 Qureshi, N. S., Bains, J. K., Sreeramulu, S., Schwalbe, H. & Fürtig, B. Conformational switch in the ribosomal protein S1 guides unfolding of structured RNAs for translation initiation. *Nucleic Acids Res* **46**, 10917-10929 (2018). https://doi.org:10.1093/nar/gky746

65 Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13-17 (2011). https://doi.org:10.1093/nar/gkr245

66 Watson, Z. L. *et al.* Structure of the bacterial ribosome at 2 A resolution. *Elife* **9**, e60482 (2020). https://doi.org:10.7554/eLife.60482

67 Callaghan, A. J. *et al.* Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature* **437**, 1187-1191 (2005). https://doi.org:10.1038/nature04084

Chapter VII

7. Discussion and Outlook

Despite the T4 phage being one of the most extensively studied phages, it is striking that nearly half of its proteome (45%) remains of unknown function. This perfectly highlights the incompleteness of our knowledge of the T4 phage's efficient infection mechanisms. This PhD thesis aimed to contribute new insights to narrow this knowledge gap on the molecular mechanisms of T4 phage infection of *E. coli*. Re-exploration of phage gene expression, host reprogramming mechanisms, and the host response to infection presented in this thesis has enhanced our understanding of the molecular organization of T4 phage infection and revealed its novel concepts for host hijacking. Apart from broadening our knowledge, the findings have raised several questions that future research needs to address.

7.1. Multi-omics study reveals unexplored mechanisms regulating T4 phage infection

The first objective of this work was to acquire a time-resolved overview of the T4 phage infection of *E. coli* at the molecular scale. This included elucidating the kinetics of phage gene transcription and protein biosynthesis while simultaneously studying the molecular response of the host *E. coli* to infection.

This was achieved in a comprehensive multi-omics study using high-throughput techniques to investigate the dual proteome and transcriptome during T4 phage infection of *E. coli* (Chapter 2)[1]. This work not only summarized and verified findings from various previous studies on the molecular regulation of T4 phage infection, targeting specific transcripts and proteins, but it also provided comprehensive molecular insights into the response of *E. coli* to infection for the first time. Additionally, this study identified transcription-translation decoupling for a group of T4 phage genes, providing novel perspectives on the infection dynamics.

Beyond obtaining a highly resolved overview of infection organization, the study's outcome has also raised several questions: How is the host transcriptome selectively shifted and arrested upon T4 phage infection? Which mechanisms underlie the decoupling between transcription and translation of phage genes? Furthermore, what are the biological functions of the uncharacterized phage proteins, which constitute about 45% of the T4 phage proteome, in the infection process?

7.1.1. What is the driver of rapid arrest and depletion of host transcriptome?

Our data showed a rapid arrest of the *E. coli* transcriptome upon T4 phage infection, characterized by a significant decline in the abundance of host transcripts within the first 4 min of infection and domination of T4 transcriptome latest at 7 min post-infection[1] (Chapter 2.3: Figure 1).

Understanding which host and phage proteins could contribute to the degradation of the host transcriptome upon T4 infection, as well as elucidating the mechanisms allowing distinction between phage and host transcripts, is crucial to comprehend transcriptional control takeover.

One of the factors contributing to the decrease in host transcript levels is the inability of *E. coli* to synthesize new transcripts. This is due to the rapid degradation of *E. coli* DNA in the early phase of infection by T4 phage nucleases[2], a process reported also for some other phages[3-5]. However, the arrest of the transcription via host DNA degradation would not result in a simultaneous decrease in the abundance of all bacterial transcripts due to their variable half-lives. Notably, in fast-growing bacteria with a doubling time of less than one hour, which includes *E. coli*, the average RNA half-life is 2-10 min[6]. The exact half-life of RNA can be additionally influenced by, e.g., RNA sequence, secondary structure, its interaction with other proteins, or the presence of RNA modifications[6,7]. Therefore, the varying transcripts half-lives would be expected to result in a non-uniform reduction of transcript abundance. Moreover, even the abundance of stable *E. coli* transcripts like lpp and ompA, with longer half-lives of around 30 min, rapidly decreased within the first few minutes of T4 phage infection in our study[8] (Chapter 2.3: Figure 1). These observations align with previous research indicating half-lives of lpp and ompA transcripts being reduced to approximately 2 min upon T4 phage infection[8]. Thus, the universal decline across various *E. coli* transcripts, including those that typically show a long half-life, underscores the active and targeted degradation of *E. coli* transcripts during T4 phage infection.

Influencing host RNA turnover is a common strategy among several phages, involving disruption of RNA biosynthesis pathways, actively degrading host RNA, or redirecting host nucleases to degrade specific transcripts[9-13]. Nevertheless, specifically for *E. coli*, the exact mechanisms underlying its transcriptome arrest upon T4 phage infection are not entirely explored. For instance, it was shown that the T4 phage-originated Srd protein is associated with *E. coli* RNase E and stimulates its nuclease activity to cleave host mRNA early in the infection[14]. Srd was shown to be crucial for T4 phage infection by contributing to the efficient transition from host to phage gene expression[14]. However, the mechanism that allows the Srd-RNase E complex to target specific *E. coli* RNAs and preserve the ones of the phage was not explored.

In contrast to host RNA transcription, phage RNA transcription is consistently maintained in a time-specific pattern throughout the infection. While *E. coli* nuclease RNase LS targets phage middle and late transcripts to prevent cell lysis, it is actively blocked by phage protein Dmd[15,16] (Chapter 1.5). Further engagement of host nucleases targeting the phage transcripts to stop the infection is unexplored. On the other hand, phage-derived RNase RegB is actively processing phage transcripts during infection. In particular, the shift in phage expression from the early to middle phage genes is partially modulated by the T4 RegB endonuclease. This nuclease recognizes the

GGAC motif in early phage transcripts, one of the most common Shine-Dalgarno sequences in T4 phage transcripts[17-19]. The cleavage by RegB initiates the degradation of transcripts by *E. coli* nucleases[17], and therefore, regulates phage infection.

Thus, the observations made in the dual-transcriptome study combined with the previous knowledge on RNA metabolism upon T4 phage infection strongly suggest the presence of unexplored mechanisms for selective degradation of host RNA upon infection. The initial step toward clarifying these mechanisms could involve identifying other proteins and nucleases that may be involved in this process. To determine whether other host or phage nucleases are involved in host RNA degradation or anti-phage defense mechanisms (e.g., targeting phage RNA to prevent infection), studying infection transcriptome applying *E. coli* and T4 phage mutants with deactivated or deleted nucleases could provide initial insights. Promising protein candidates showing significant impact on modulation of transcriptome upon infection can then be identified and their catalytic and regulatory roles in the infection process can be investigated on a molecular level.

Furthermore, enhancing the temporal resolution of transcriptome analysis – by increasing the frequency of transcriptome sampling during the early stages of infection – can provide ideas on how specific nucleases affect the stability of particular transcripts in the early phase. Given that our study observed a uniform decrease in *E. coli* transcript abundance as early as 4 min post-infection, a higher temporal resolution of RNA-seq could allow for tracking differential transcript degradation patterns at the initial steps of transcription arrest during T4 phage infection.

The resolution of the first minutes of infection can uncover the potential response of *E. coli* to phage invasion. For instance, after 1 min of infection, the response of *E. coli* to infection is indicated by the increase in TPM values for specific bacterial genes, featured by the genes in the clusters of orthologous groups (COGs) related to “transcription”, “energy production and conversion”, “unknown function”, and “general function predicted”, among others. Also, it is possible to speculate that the anti-phage defense systems might be encoded among the differentially expressed host transcripts appearing early in the infection. Conducting transcriptome analysis between 1 and 4 min post-infection could reveal previously unidentified dynamic host-phage molecular interactions that occur before the ultimate T4 phage’s transcriptional hijacking. Additionally, it may reveal critical phage genes, the expression of which marks a point beyond which *E. coli* can no longer maintain its own metabolism and gene expression effectively.

To gain such insights, bulk RNA-seq can be performed[20]. This involves sequencing of the RNA extracted from a population of cells, as it was done in the given study[1]. However, the output would represent an average of transcripts abundance across a population of cells subjected to the experiment. The potential heterogeneity (e.g., the different onset of infection) of the molecular organization of T4 phage infection is therefore only partially captured by this method. Thus, single-

cell RNA-seq can be a valuable alternative to capture infections of distinct cells, enabling transcript profiling at the single cell level[21-23]. The recently developed M3-seq was shown to be a powerful approach to study phage infection at the transcriptional level[24]. Thus, its application to study T4 phage infection of *E. coli* could deepen our understanding of how the host responds to phage infection on the transcriptional level within the first few minutes of infection and how heterogeneous this response is within a population of infected cells.

7.1.2. RNA modifications as a potential means for differentiation of phage and host RNA

Identifying the nucleases responsible for the targeted degradation of the *E. coli* transcripts is the first crucial step toward understanding transcriptome regulation during T4 phage infection. However, it is equally important to understand how the selective degradation toward host versus phage transcripts is achieved. For instance, ssRNA phage MS2 was shown to have evolved RNA folds to protect its RNA from degradation by nucleases[25]. While the presence of RNA secondary structures was proven for some T4 phage transcripts and studied in the context of their impact on translation, their role in RNA stability has not been elucidated so far[26,27].

Another potential factor affecting host and phage RNA's differential stability could be RNA modifications on the respective transcripts. Chapters III and IV of this thesis compile the existing knowledge on RNA modifications in bacteria and phages, highlighting the limited knowledge on bacterial RNA modifications and the absolutely unexplored field of phage RNA modifications[28,29]. In Chapter IV potential mechanisms were discussed, by which specific phage and host enzymes could potentially introduce, recognize, or remove RNA modifications. Selective RNA modifications could occur during phage infection and might regulate the role and stability of RNA. Like T4 phage DNA, which is extensively modified to preserve its integrity and avoid nucleolytic degradation by host nucleases[30,31], RNA modifications may also protect phage RNA from degradation. Therefore, investigating and potentially discovering RNA modifications in T4 phage and *E. coli* transcripts upon infection might unravel a new layer of infection regulation and provide the basis for transcript differentiation between host and virus.

To analyze the spectrum of RNA modifications and their abundance during infection, methods targeting a specific RNA modification can be utilized[28,32]. In such a way, a confirmation of the presence of particular RNA modifications on analyzed transcripts is possible. However, targeting individual modifications could be labor-intensive, as over 170 different RNA modifications have been reported to date[33]. A comprehensive method capable of simultaneous identification of multiple modifications in RNA mixtures or even within a single transcript would be beneficial. In this context, Nanopore sequencing emerges as a powerful tool. Its advancements have already facilitated the mapping of specific modification sites, e.g., m^6A sites in RNA[34-36], and with ongoing

technological progress, it is conceivable that Nanopore sequencing could be expanded to examine the prevalence and dynamics of RNA modifications during phage infection. The potential discovery of RNA modifications on specific groups of transcripts, whether from host or phage, can allow us to relate their presence to RNA susceptibility to degradation by host and phage nucleases.

7.1.3. Transcription-translation decoupling in T4 phage

In our study, we noted a significant decoupling between transcription and translation for a substantial number of T4 phage transcripts. In contrast to this observation, the transcription and translation of the host *E. coli* are known to occur in temporal proximity[37,38]. Of the 206 T4 transcripts identified as early T4 RNAs, only 60 encode for early proteins, while the majority of 75 RNAs encode for middle proteins and a further 42 RNAs encode for late proteins (Chapter 2: Figure 4e). Although some instances of transcription-translation decoupling may be attributed to the analytical parameters of the correlation analysis, which assigns temporal groups based on TPM (for transcriptome) or LFQ (for proteome) exceeding 10% of their maximum values, a significant observation remains: Specifically, 117 phage proteins are synthesized considerably later in the infection cycle than their corresponding RNAs. This observation implies the existence of post-transcriptional or translational regulatory mechanisms during T4 phage infection, which remain unexplored.

For two T4 phage genes, *soc* and *endolysin* (*e*, also known as lysozyme), a transcription-translation decoupling has been reported[27]. Notably, *endolysin* transcription begins early in the infection and continues in the late infection phase. In the T4 genome, the *endolysin* gene is under the dual control of two independent promoters: a late promoter upstream of the *endolysin* nucleotide sequence and an additional early promoter upstream of the mentioned late promoter. Consequently, the transcript synthesized in the early phase of the infection cycle is longer than the transcript that emerges in the late phase. While the longer RNA sequence of the early transcript features the ability to form a secondary RNA structure that inhibits the translation initiation site of endolysin, the translation of late endolysin transcript is not affected by RNA secondary structure[27]. It has been hypothesized that the phage might require the translation of endolysin under certain conditions for survival in the early stages of infection[27]. The same mode of regulation was also described for the soc transcript. Our dual-omics data correlate with this observation, showing transcription and translation for *endolysin* and *soc* genes being decoupled upon infection (Figure 1A).

Secondary RNA structures have been identified in other T4 phage transcripts, such as those of gp49 and I-TevI, where they appear to impact translation negatively[39]. RNA sequencing data do not offer precise temporal classification for these transcripts due to their low abundance and noisy

temporal profiles. Nevertheless, previous microarray experiments categorized T4 transcripts of *gp49* and *I-TevI* as delayed early and early, respectively[40]. Our proteomics analysis classifies I-TevI and Gp49 as late proteins[1] (Figure 1B). Thus, internal RNA structures may act as negative regulators for the translation of certain T4 RNAs and thereby cause the observed transcription-translation decoupling.

However, the observation of the negative impact of secondary structures on T4 mRNA translation cannot be generalized to all T4 transcripts exhibiting a secondary structure. For instance, intramolecular RNA structures in the T4 transcripts gp38 and gp25 have been shown to enhance their translation[41], demonstrating the opposite, positive effect of RNA secondary structures on the translation of the respective T4 transcripts.

Once the correlation between phage RNA secondary structure and its enhanced or reduced translation is understood, this knowledge can be exploited for synthetic means and phage engineering. In particular, understanding how RNA translation can be actively modulated by its structure can be applied as a regulatory element to control protein biosynthesis.

Beyond physical structure, chemical RNA modifications, which are discussed in detail in Chapter III and IV, could potentially influence the translation of phage RNA. However, as mentioned previously (Chapter 7.1.2), the existence of RNA modifications on phage transcripts has not yet been confirmed. If present, mRNA modifications may indeed affect the translation rate, as it has been demonstrated for mRNA with pseudouridine modifications, whose translation was negatively impacted in an artificially reconstituted *E. coli* translation system[42]. Therefore, the presence of RNA modifications on phage transcripts must be investigated and if found, their effect on translation can be studied through *in vitro* translation systems, followed by *in vivo* studies using phage or host mutants unable to introduce specific RNA modifications.

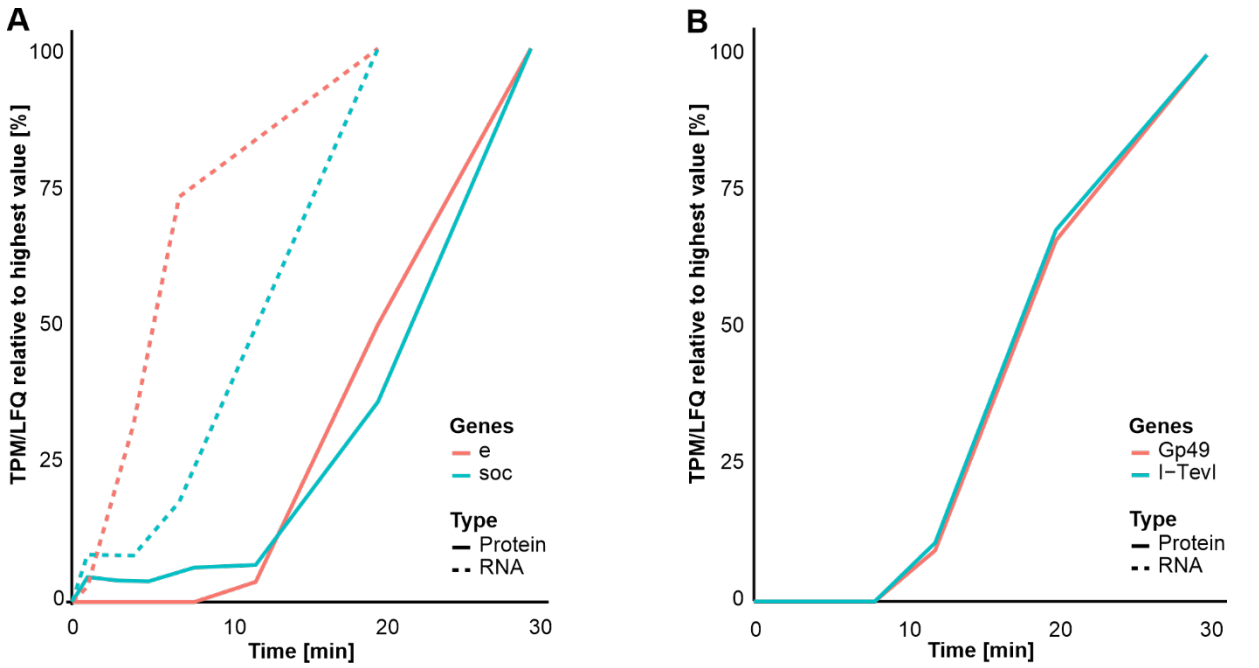


Figure 20: Time-series analysis of the relative abundance of selected T4 phage transcripts and their respective proteins. A: *Endolysin (e)* and *soc* expression during T4 phage infection, showing the transcription and translation being decoupled as was reported earlier[27]. **B:** Time-series of T4 Gp49 and I-TevI proteins abundance in the time-course of the infection, leading to the classification of both proteins as late. Transcript detection for both genes is low, impeding their temporal classification. Nevertheless, another study showed the early transcription for both genes[40] and reported the presence of RNA secondary structures impacting their translation[39]. These findings support the hypothesis that RNA secondary structures could regulate the decoupling of transcription and translation during T4 phage infection.

7.1.4. Identifying the role of uncharacterized T4 phage proteins

Elucidating the functional and biological roles of uncharacterized proteins represents a challenge in the post-genomic era[43]. This extends also to the T4 phage, half of whose proteome remains of unknown function. The persistence of homologous uncharacterized proteins across various T4-like phages implies that they could be crucial in determining host specificity, bypassing host defense mechanisms, or enabling phage replication under particular growth conditions[44]. Understanding the roles of these proteins is essential for overall comprehension of T4 phage infection mechanisms.

The dual-omics study can provide an initial insight into the function of an uncharacterized protein, especially considering the timing of its appearance in the infection. The T4 phage infection is a highly regulated and organized process that begins with the phage taking control over its bacterial host, followed by phage DNA replication and culminates in the assembly and release of phage progeny[45]. Interestingly, a big fraction of early T4 proteins, ~65%, is contributed by uncharacterized T4 proteins. While their particular biological roles are yet unknown, their early

appearance may suggest their involvement in host takeover, as being for instance HAFs[45,46] or being involved in countering host immunity toward the phages[47].

Bacterial immunity toward phages is a rapidly expanding research field, substantially enriching our understanding of the complex bacterium-phage interactions[48]. It shows that bacteria employ chemical, enzymatic, and physical strategies to evade or terminate phage infections[49-53]. The phages, in turn, evolved countermeasures, such as the synthesis of novel proteins that challenge bacterial defense mechanisms[47,54,55]. In order to counteract the bacterial anti-phage systems, phages need to react fast at the early stage of the infection, achieved by the expression of various proteins that block bacterial immunity[44,56,57]. Based on the number of phage proteins with unknown functions, our knowledge regarding the interaction of the phage with bacteria and on the phage mechanisms to escape bacterial immunity is still preliminary. Therefore, identifying the function of early-expressed, uncharacterized phage proteins could give insights into novel molecular strategies of the phage to hijack its host. The determination or prediction of the biological function of the uncharacterized phage proteins using sequence homology-based methods generally falls short due to their small size and rapid evolutionary turnover[47].

Although sequence homology has its limitations in the prediction of the function of these small uncharacterized proteins, structural homology based on three-dimensional protein conformation can be a powerful tool[58]. Structural homology, in contrast to protein sequence homology, can be retained across long evolutionary timescales[59]. It has also recently been shown that the annotation of metagenomic proteins can be significantly improved by up to 70% by incorporating structural features[60]. Furthermore, the search for structural homology is strongly streamlined by advancements in computational protein structure prediction with the possibility to gain high-confidence structural models, for instance via AlphaFold[54,61]. Although the prediction of protein structures does not displace the need for experimental structure determination, a computational model can become a starting point for elucidation of the function of uncharacterized proteins.

The utility of structural prediction followed by structural homology search for phage proteins was demonstrated in the identification of anti-CRISPR (Acr) phage proteins across a number of phages by combining these methods[62]. Thus, the uncharacterized T4 phage proteins can be approached in the same way to gain first insights into their functionality. Once the potential function of a protein has been identified and verified through *in vitro* studies, its effect on phage infection efficiency can be evaluated. This can be accomplished by engineering the T4 phage to express a mutant version of the protein with impaired functionality. The engineered phage can be further studied to assess the importance of the studied protein in the context of phage infection.

7.2. Current advancements and future perspectives in streamlined phage mutagenesis

Robust phage engineering techniques are crucial for both – fundamental phage research and the tailoring of phages for specific applications. The application of phage mutagenesis to study the role of the specific viral proteins in the infection cycle was often connected to a deletion of the studied gene. However, deleting an entire gene from the highly complexly arranged T4 phage genome (discussed in Chapter 1.3), can have severe consequences, e.g., impeding the expression of downstream and upstream genes. Additionally, a phage protein can fulfill multiple functions during infection. For example, it can be a core protein[63], potentially essential for virion assembly, while also acting as a catalyst, as in the case of T4 ART Alt (Chapter 1.9). Therefore, deleting an entire gene may affect multiple infection-relevant processes simultaneously, making it impossible to attribute the effect to a specific protein functionality, as in the case of Alt and its ADP-ribosylation activity. To avoid this issue, site-specific mutations can be inserted to abolish one of the protein functionalities, which opens up new opportunities for studying the biological function of phage proteins.

An efficient and scarless approach is required to generate such mutants with a minimal number of inserted mutations. Scarless refers to the relinquishment of reporter genes or selection markers that are introduced to validate the success of mutagenesis and, moreover means that surrounding coding sequences are not affected in their integrity. CRISPR-Cas-based phage engineering offers the best approach to meet these requirements. Nevertheless, its applicability for T4 phage mutagenesis was shown to be strongly restricted due to extensive phage DNA modifications[64,65]. The T4 phage engineering strategy outlined in Chapter V involves pretreatment of phage DNA with eukaryotic NgTET dioxygenase *in vivo*, followed by CRISPR-Cas-mediated mutagenesis[66]. The NgTET treatment serves for temporal modulation of cytosine modifications within T4 phage DNA. This leads to significantly enhanced efficiency of the first and crucial mutagenesis step – the generation of a double-strand break of phage DNA at the intended mutation site via Cas nuclease targeting. It was shown that NgTET-mediated reduction of T4 DNA modifications resulted in an up to 7-fold increase in Cas nuclease cleavage efficiency of phage DNA *in vivo*. This increase in Cas12-mediated DNA cleavage enabled a mutagenesis success rate of 6% for the studied targets *alt* and *modA*[66], while no mutagenesis occurred for the same targets using the original CRISPR-Cas12 mutagenesis systems[64,67] where NgTET was not present. These results show that the established approach significantly improves the efficiency of T4 phage mutagenesis allowing the introduction of as small mutations as point mutations. Furthermore, there is potential to increase further the mutagenesis rate beyond the 6% achieved in this study by optimizing additional steps in the mutagenesis procedure.

In particular, after Cas-mediated cleavage of phage DNA, the second step of mutagenesis is the introduction of the mutation into the phage genome via homologous recombination (Chapter 1: Figure 4). The introduction of the mutation into the DNA relies in our approach on the T4 homology-directed repair (HDR) machinery, components of which are known to be highly expressed during phage infection[68]. The HDR system of T4 phage includes Gp46 and Gp47 proteins involved in the formation of single-stranded DNA, the phage ssDNA-binding protein Gp32, the recombination mediator protein UvsY, the recombinase UvsX, and the DNA helicase UvsW[68-71]. Based on the obtained dual-proteome data of the T4 phage infection (Chapter 2)[1], these proteins are predominantly identified as late proteins, except Gp47, that has been categorized as middle protein. The late expression of proteins involved in HDR could negatively affect the overall efficiency of mutagenesis. This is mainly due to the fact that non-modified host DNA, including the donor DNA for HR, is rapidly degraded by T4 phage nucleases early in the infection process[72,73]. Consequently, the donor DNA may no longer be present in the cell by the time the proteins required for homologous recombination are expressed. Incorporating heterologous recombination systems into the mutagenesis strain could be an effective strategy to enhance the synchronization between Cas targeting of phage DNA and homologous recombination. Pre-expressing a heterologous recombination system prior to phage infection, e.g., the widely used Lambda red recombineering system for *E. coli*[74], could enable the HDR directly after Cas-mediated phage DNA cleavage during the early stages of infection. Therefore, it can be hypothesized that implementing a heterologous recombination system into the procedure of our phage mutagenesis approach, as presented in Chapter V, might even further boost phage mutagenesis efficiency.

Moreover, phage mutant detection can also be streamlined in the future. Given the current maximum mutagenesis rate of 6% achieved in the published mutagenesis setup[66], the majority of the phage population remains wild-type after mutagenesis. Long-read sequencing, applied in the presented study[66], offers a way to screen large phage populations for mutants without the need to introduce reporter genes, which, while effective in detecting phage mutants, could potentially disrupt phage genetics or impede DNA packaging[31,44]. Nevertheless, the process of screening can be further simplified by reducing the number of phages that need to be screened. This can be accomplished by counterselection of wild-type phages in a post-mutagenesis phage population. Specifically, this counterselection can occur at the DNA level by employing the Cas nuclease, the same as used for mutagenesis, to target and cleave the DNA of wild-type phages at the intended mutation site. The recognition and cleavage of wild-type phage DNA would prevent their propagation and thus minimize their overall proportion in the phage population after mutagenesis[67]. Since T4 phage DNA modifications impede Cas nuclease targeting of the phage

DNA[30,64,75], treating T4 DNA with NgTET presents a promising strategy to also enhance the efficiency of the proposed counterselection process.

It is also important to note that when introducing small mutations, such as point mutations, the small number of mismatches between the wild-type and mutated sequences may not be sufficient for Cas nucleases to discriminate between wild-type and mutant DNA[76]. To achieve efficient counterselection, the introduction of silent mutations into the protospacer region additionally to the intended mutation can be a promising strategy to increase the number of mismatches between wild-type and mutated sequence and thereby enhance the Cas nuclease selectivity. Alternatively, the introduction of silent mutations into the protospacer-adjacent motif (PAM), which is essential for the Cas9 and Cas12 nucleases for target recognition and cleavage, can be another strategy. This approach has already been successfully used to improve the efficiency of phage mutants counterselection using CRISPR-Cas9[77,78].

Apart from phage counterselection on DNA level, RNA-targeting CRISPR-Cas systems have been reported recently to be efficient for phage counterselection[79,80]. Systems like CRISPR-Cas13a induce non-specific RNA degradation upon recognition of their RNA target. This results in the arrested growth of the cell and thereby blocking the progression of phage infection[79,80]. While Cas13a nuclease tolerates up to three mismatches between the guide RNA and its DNA target[79], an engineered Cas13a was recently reported to differentiate even for a single nucleotide polymorphism[81]. This system offers an effective method for counterselecting phage mutants, in which particular proteins are inactivated by a single nucleotide mutation.

Taken together, the mutagenesis strategy developed in this thesis could see enhancements through the integration of a heterologous recombination system to improve recombination efficiency in the future. Additionally, the implementation of effective counterselection strategies might streamline the detection of phage mutants. The development of the presented mutagenesis system was already a big step toward efficient and reliable engineering of bacteriophages in a scarless manner and holds the promising potential to approach fundamental biological questions or generate tailored and therapeutically relevant phages.

7.3. RNAylation – a novel concept of protein-RNA interaction

The high relevance of phage ADP-ribosyltransferases for the efficient reprogramming of *E. coli* by T4 phage is beyond doubt[31,82]. However, the study outlined in Chapter VI revealed that our knowledge of the role of ARTs in the infection cycle is far from being complete, and much remains to be discovered regarding their function.

Our investigation into the mechanisms behind T4 phage control takeover and the involvement of T4 ARTs in these processes led us to the discovery of a post-translational protein modification known as RNAylation. The T4 ADP-ribosyltransferase ModB has been shown to accept NAD-RNA

as a substrate in addition to NAD. In addition to its ADP-ribosylating activity, ModB catalyzes RNAylation by transferring the ADP-ribose-RNA moiety from NAD-RNA to target proteins. This not only provides new meanings to the role of NAD-RNAs but also represents a novel type of protein-RNA interaction. In our study, we have observed several *E. coli* proteins, including ribosomal proteins S1 and L2, being RNAylated upon T4 phage infection. Phages with an inactive ModB variant exhibited an impeded phenotype, underscoring the significance of ModB activity for efficient infection. This study opens new avenues in understanding bacterium-phage interactions, raising several questions about the biological relevance and function of RNAylation.

7.3.1. What is the biological role of RNAylation?

While two other T4 ARTs, Alt and ModA, introduce ADP-ribosylation to *Ec*RNAP, redirecting the transcriptional machinery of the host toward phage gene expression, the third T4 ART ModB RNAylates ribosomal proteins, the key components of the translational apparatus[83]. However, the impact of RNAylation on translation during infection remains to be investigated. Yet, it can be hypothesized that ModB-mediated RNAylation of *E. coli* ribosomal proteins could be another host-hijacking mechanism. This could potentially affect the translational efficiency or integrity of ribosomes, providing an advantage for phage protein biosynthesis.

The RNAylation of ribosomal proteins might regulate the translation of both host and phage transcripts. Particularly, rS1 has been identified as a target of ModB-catalyzed RNAylation[83] and is a crucial component of the 30S ribosomal subunit[84]. rS1 plays a vital role in translation initiation by recruiting mRNAs to the 30S subunit[84] and exhibits RNA helicase activity by unfolding mRNA and mediating its proper positioning on the ribosome[85,86]. It has been shown that the binding of rS1 to the 30S subunit is transient. This was observed in a lower rS1 to 70S ratio when purifying 70S ribosomes compared to the ribosomes *in vivo*[87]. It can be hypothesized that RNAylation of rS1 can increase the strength of non-covalent interactions between rS1 and the 30S ribosomal subunit, thereby increasing the fraction of ribosomes with bound rS1 during infection. This, in turn, could boost protein biosynthesis due to the increased number of fully-assembled ribosomes.

However, the opposite impact of RNAylation can also be hypothesized. RNAylation could potentially weaken the interaction between rS1 and the 30S subunit even further. As this could result in slower overall translation, it could be used by phages as a strategy to minimize the host response to infection, particularly host protein biosynthesis in the early stages of infection.

To investigate the impact of RNAylation on the interaction between 30S ribosomal subunits and rS1, studies on association and dissociation kinetics of ribosomal components could help to determine whether RNAylation of rS1 strengthens or weakens the interaction with 30S. To investigate the impact of RNAylation on ribosomal translational efficiency, non-RNAylated and

RNAylated *E. coli* ribosomes can be isolated from either *E. coli* or T4 phage-infected *E. coli*, respectively. The efficiency of the translation in the presence and absence of RNAylation can be assessed *in vitro* in a cell-free system using mRNA encoding for a fluorescent protein. Translation efficiency can be measured through fluorescence.

Recruitment of specific mRNAs for translation could be another possible biological role of RNAylation of ribosomal proteins. This hypothesis would suggest that RNAylation of rS1 could have the means to deliver specific transcripts to ribosomes or accumulate transcripts within reach of ribosomes to mediate the transcripts' efficient translation. This hypothesis raises the question of whether mRNA, covalently attached to a protein via RNAylation, can be translated into a protein. To answer this question, one could perform RNAylation reactions *in vitro* using NAD-mRNA that encodes a reporter protein, such as a fluorescent protein. Then, the RNAylated rS1 protein can be used to complement 70S ribosomes. Next, *in vitro* translation efficiency could be determined in cell-free systems by measuring the fluorescence signal.

Another ribosomal protein that undergoes RNAylation upon phage infection is rL2. rL2 is an essential component of the 50S ribosomal subunit. The association of the 30S and 50S ribosomal subunits to form 70S ribosomes is entirely dependent on rL2[88,89]. Additionally, rL2 is responsible for maintaining the peptidyltransferase activity and tRNAs binding to both A (aminoacyl-tRNA) and P (peptidyl-tRNA) sites[88]. The RNAylation of rL2 could potentially affect its interaction with tRNAs, especially with respect to the eight tRNAs synthesized by the T4 phage during infection[31]. To test whether RNAylation increases the affinity of *E. coli* ribosomes toward T4 phage tRNAs, additional experiments can be conducted in a cell-free system. The experiment can test the translation efficiency for phage and host transcripts using RNAylated or non-RNAylated ribosomes in the presence or absence of T4 phage tRNAs. Furthermore, the RNAylation of rL2 might again be a negative regulation factor from the phage, aiming to slow down translation as discussed above for rS1. It can be hypothesized that RNAylation of rL2 impedes integrity of 70S ribosomes by negatively affecting 30S and 50S subunit association.

The biological significance of RNAylation was discussed in terms of its impact on the functionality of RNAylated proteins. However, it is important to consider that RNAylation may affect not only the modified protein but also the fate of the RNA used for RNAylation. It is possible that RNAylation could stabilize the RNA by attaching it to a protein, or it could act as an mRNA silencing mechanism by preventing the translation of NAD-RNA through covalent attachment to proteins. The latter hypothesis can be tested by the previously suggested experiment aiming to determine whether RNA that has been covalently attached to proteins by RNAylation can be translated.

Regardless of which hypothesis regarding the putative roles of RNAylation is correct, other questions must also be addressed. Is RNAylation a permanent modification, or is it part of a

regulated process where it can be removed by other proteins after fulfilling its biological function? Furthermore, which proteins are involved in its regulation? These questions highlight the need for further research into the RNAylation phenomenon. The potential roles of RNAylation, whether in directing RNA to ribosomes or inhibiting translation, present exciting challenges and opportunities for future studies.

7.3.2. What is the molecular origin of RNAylation?

In our study we aimed to explore the relevance of RNAylation in the T4 phage infection cycle (Chapter 6)[83]. To address this question, the T4 phage variant referred to as T4 ModB R73A, G74A, harboring an inactive version of ModB, was generated[83]. Our results show that ModB inactivation affects the infection phenotype by delaying lysis, reducing phage burst size, and even impeding phage adsorption to the cell[83]. Nevertheless, the observation is difficult to connect to RNAylation solely. This is due to the fact that the mutations R73A and G74A, which were introduced to ModB, eliminate both its ADP-ribosylation and RNAylation activities. To decouple both functionalities and create ModB variants capable of accepting NAD-RNA or NAD as substrate only, a precise understanding of how ModB achieves substrate specificity is required in the first place. The rationale behind ModB's acceptance of NAD-RNAs, in addition to NAD, as substrates remains unexplored. So what makes ModB different from other ADP-ribosyltansferases that do not accept NAD-RNA as substrates but only NAD?

Understanding how T4 ModB evolved to accept NAD-RNA as a substrate cannot be deduced from protein sequence comparisons between ModB and other T4 ARTs. Although all three T4-encoded ARTs are arginine-specific ARTs with a conserved R-S-EXE active site motif[82,90], ModB shows only 30.9% sequence identity with ModA and 28.2% with Alt (calculated with LALNVIEW[91]). Thus, the substantial differences in sequence do not aid in clarifying the reasons for the partially different reactivity.

Beyond protein sequence, protein structural analysis can offer insights into the ModB residues or its structural elements that confer RNAylation activity to ModB. However, the structures for ModB and other T4 ARTs have not yet been experimentally determined. The application of AlphaFold for computational modeling allows for the sequence-based prediction of ART structures (Figure 2)[92,93]. The superposition of predicted ModA and ModB protein structures by structural alignment results in an RMSD value (root-mean-square deviation of atomic positions) of 1.09 Å. This value indicates a high degree of structural similarity between the two ART models. Thus, detailed structural analysis to identify the specific differences – such as the arrangement of amino acid residues within the catalytic site of the enzymes – could reveal the molecular reason for differences in the substrate scope of both ARTs. The AlphaFold-predicted ModB model can be used in molecular

docking simulations with NAD-RNA to pinpoint sites that facilitate NAD-RNA interaction, potentially uncovering the evolutionary development of ModB's RNAylation activity.

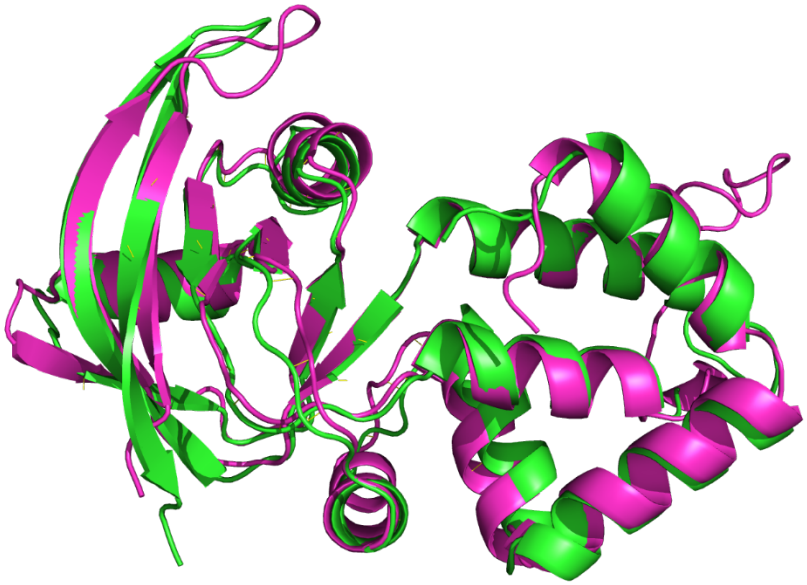


Figure 21: Alignment of ModA (green) and ModB (magenta) structural models predicted with Colab AlphaFold[93]. The alignment results in an RMSD of 1.09 Å, reflecting a high similarity between the two predicted structures, with 30.9% sequence identity for both proteins.

Another way to approach clarification of the molecular reason for ModB's ability to perform RNAylation alongside ADP-ribosylation might be the prediction of ancestor proteins from the ModB lineage via ancestor sequence reconstruction (ASR). ASR resurrects ancestral proteins using the sequences of extant proteins and their evolutionary relationships, summarized in a phylogenetic tree[94,95]. This approximation, paired with experimental validation, could elucidate the substitutions and changes that enabled ModB the functional shift for gaining RNAylation activity. However, the utility of ASR for phage proteins may be constrained by their rapid evolution and diversity, which lead to significant divergence and low sequence similarity and poses a challenge for the alignment process[95].

Furthermore, it would be helpful to compare ModB not only with other ARTs performing ADP-ribosylation but also with ARTs showing RNAylation activity. Yet, this is challenging since ModB is the first and only ART identified with RNAylation activity. This raises the question: is it specific to the T4 phage-host interaction, or does it have broader biological significance across different species and interactions? The strongest argument supporting the hypothesis of the presence of RNAylation beyond the *E. coli*-T4 phage pair is the ubiquity of NAD-RNAs and ARTs across all

domains of life[7,96-103] along with ARTs' key roles in host-pathogen interactions[103,104]. This suggests that RNAylation might be a common regulatory mechanism in various inter-species interactions. In summary, understanding the molecular and evolutionary background of RNAylation activity and experimentally verifying the presence of RNAylation in other organisms could significantly expand our understanding of the biological role of this post-translation protein modification.

7.4. Conclusion

This thesis was committed to study the infection cycle of bacteriophage T4 and its mechanisms of hijacking its bacterial host *E. coli*. In order to re-assess the mutual arms race with a long scientific history between T4 phage and *E. coli*, a comprehensive approach to gain novel insights and perspectives on the fundamental mechanisms of infection was chosen. We demonstrated that the T4 phage still harbors numerous unexpected and unique molecular mechanisms, previously undiscovered post-translational modifications, and disregarded aspects to invade and control its host. The development of a precise and efficient T4 phage mutagenesis system within this thesis paves the way for the identification and investigation of unknown infection modulators utilized by T4 phage. In conclusion, this thesis illuminates previously hidden facets of the underlying infection mechanisms and identifies future perspectives for studies on T4 phage infection. The highlighted results represent a step forward in leveraging T4 phage and its molecular principles in medical and biotechnological applications.

7.5. References

1 Wolfram-Schauerte, M., Pozhydaieva, N., Viering, M., Glatter, T. & Höfer, K. Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of E. coli on Proteome and Transcriptome Levels. *Viruses* **14** (2022). https://doi.org:10.3390/v14112502

2 Parson, K. A. & Snustad, D. P. Host DNA degradation after infection of Escherichia coli with bacteriophage T4: dependence of the alternate pathway of degradation which occurs in the absence of both T4 endonuclease II and nuclear disruption on T4 endonuclease IV. *J Virol* **15**, 221-224 (1975). https://doi.org:10.1128/JVI.15.1.221-224.1975

3 Yang, J. Y. *et al.* Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Syst* **12**, 771-779 e775 (2021). https://doi.org:10.1016/j.cels.2021.05.019

4 Warner, H. R., Drong, R. F. & Berget, S. M. Early events after infection of Escherichia coli by bacteriophage T5. Induction of a 5'-nucleotidase activity and excretion of free bases. *J Virol* **15**, 273-280 (1975). https://doi.org:10.1128/JVI.15.2.273-280.1975

5 Panayotatos, N. & Fontaine, A. An endonuclease specific for single-stranded DNA selectively damages the genomic DNA and induces the SOS response. *J Biol Chem* **260**, 3173-3177 (1985).

6 Laalami, S., Zig, L. & Putzer, H. Initiation of mRNA decay in bacteria. *Cell Mol Life Sci* **71**, 1799-1828 (2014). https://doi.org:10.1007/s00018-013-1472-4

7 Cahova, H., Winz, M. L., Höfer, K., Nubel, G. & Jäschke, A. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**, 374-377 (2015). https://doi.org:10.1038/nature14020

8 Ueno, H. & Yonesaki, T. Phage-induced change in the stability of mRNAs. *Virology* **329**, 134-141 (2004). https://doi.org:10.1016/j.virol.2004.08.001

9 Chevallereau, A. *et al.* Next-Generation "-omics" Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of Pseudomonas aeruginosa. *PLoS Genet* **12**, e1006134 (2016). https://doi.org:10.1371/journal.pgen.1006134

10 Blasdel, B. G., Chevallereau, A., Monot, M., Lavigne, R. & Debarbieux, L. Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera. *ISME J* **11**, 1988-1996 (2017). https://doi.org:10.1038/ismej.2017.63

11 Marchand, I., Nicholson, A. W. & Dreyfus, M. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Mol Microbiol* **42**, 767-776 (2001). https://doi.org:10.1046/j.1365-2958.2001.02668.x

12 Stazic, D., Lindell, D. & Steglich, C. Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Res* **39**, 4890-4899 (2011). https://doi.org:10.1093/nar/gkr037

13 Stazic, D., Pekarski, I., Kopf, M., Lindell, D. & Steglich, C. A Novel Strategy for Exploitation of Host RNase E Activity by a Marine Cyanophage. *Genetics* **203**, 1149-1159 (2016). https://doi.org:10.1534/genetics.115.183475

14 Qi, D., Alawneh, A. M., Yonesaki, T. & Otsuka, Y. Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. *Genetics* **201**, 977-987 (2015). https://doi.org:10.1534/genetics.115.180364

15 Ueno, H. & Yonesaki, T. Recognition and specific degradation of bacteriophage T4 mRNAs. *Genetics* **158**, 7-17 (2001). https://doi.org:10.1093/genetics/158.1.7

16 Kai, T., Selick, H. E. & Yonesaki, T. Destabilization of bacteriophage T4 mRNAs by a mutation of gene 61.5. *Genetics* **144**, 7-14 (1996). https://doi.org:10.1093/genetics/144.1.7

17 Uzan, M. & Miller, E. S. Post-transcriptional control by bacteriophage T4: mRNA decay and inhibition of translation initiation. *Virol J* **7**, 360 (2010). https://doi.org:10.1186/1743-422X-7-360

18 Ruckman, J., Ringquist, S., Brody, E. & Gold, L. The bacteriophage T4 regB ribonuclease. Stimulation of the purified enzyme by ribosomal protein S1. *J Biol Chem* **269**, 26655-26662 (1994).

19 Juskauskas, A. *et al.* Interaction between Phage T4 Protein RIII and Host Ribosomal Protein S1 Inhibits Endoribonuclease RegB Activation. *Int J Mol Sci* **23** (2022). https://doi.org:10.3390/ijms23169483

20 Li, X. & Wang, C. Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* **13**, 36 (2021). https://doi.org:10.1038/s41368-021-00146-0

21 Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat Methods* **14**, 267-270 (2017). https://doi.org:10.1038/nmeth.4145

22 Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol* **5**, 1192-1201 (2020). https://doi.org:10.1038/s41564-020-0729-6

23 Kuchina, A. *et al.* Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371** (2021). https://doi.org:10.1126/science.aba5257

24 Wang, B. *et al.* Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nat Microbiol* **8**, 1846-1862 (2023). https://doi.org:10.1038/s41564-023-01462-3

25 Klovins, J., van Duin, J. & Olsthoorn, R. C. Rescue of the RNA phage genome from RNase III cleavage. *Nucleic Acids Res* **25**, 4201-4208 (1997). https://doi.org:10.1093/nar/25.21.4201

26 Babitzke, P., Baker, C. S. & Romeo, T. Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol* **63**, 27-44 (2009). https://doi.org:10.1146/annurev.micro.091208.073514

27 McPheeters, D. S., Christensen, A., Young, E. T., Stormo, G. & Gold, L. Translational regulation of expression of the bacteriophage T4 lysozyme gene. *Nucleic Acids Res* **14**, 5813-5826 (1986). https://doi.org:10.1093/nar/14.14.5813

28 Schauerte, M., Pozhydaieva, N. & Höfer, K. Shaping the Bacterial Epitranscriptome-5'-Terminal and Internal RNA Modifications. *Adv Biol (Weinh)* **5**, e2100834 (2021). https://doi.org:10.1002/adbi.202100834

29 Pozhydaieva, N., Wolfram-Schauerte, M., Keuthen, H. & Höfer, K. The enigmatic epitranscriptome of bacteriophages: putative RNA modifications in viral infections. *Curr Opin Microbiol* **77**, 102417 (2024). https://doi.org:10.1016/j.mib.2023.102417

30 Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* **6**, e00648 (2015). https://doi.org:10.1128/mBio.00648-15

31 Miller, E. S. *et al.* Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents (2003). https://doi.org:10.1128/MMBR.67.1.86-156.2003

32 Höfer, K. & Jäschke, A. Epitranscriptomics: RNA Modifications in Bacteria and Archaea. *Microbiol Spectr* **6** (2018). https://doi.org:10.1128/microbiolspec.RWR-0015-2017

33 Cappannini, A. *et al.* MODOMICS: a database of RNA modifications and related information. 2023 update. *Nucleic Acids Res* **52**, D239-D244 (2024). https://doi.org:10.1093/nar/gkad1083

34 Leger, A. *et al.* RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun* **12**, 7198 (2021). https://doi.org:10.1038/s41467-021-27393-3

35 Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat Biotechnol* **39**, 1394-1402 (2021). https://doi.org:10.1038/s41587-021-00949-w

36 Fleming, A. M., Bommisetti, P., Xiao, S., Bandarian, V. & Burrows, C. J. Direct Nanopore Sequencing for the 17 RNA Modification Types in 36 Locations in the E. coli Ribosome Enables Monitoring of Stress-Dependent Changes. *ACS Chem Biol* **18**, 2211-2223 (2023). https://doi.org:10.1021/acschembio.3c00166

37 Bakshi, S., Choi, H. & Weisshaar, J. C. The spatial biology of transcription and translation in rapidly growing Escherichia coli. *Front Microbiol* **6**, 636 (2015). https://doi.org:10.3389/fmicb.2015.00636

38 Fan, H. *et al.* Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res* **45**, 11043-11055 (2017). https://doi.org:10.1093/nar/gkx719

39 Miller, E., Karam, J. & Spicer, E. Control of translation initiation: mRNA structure and protein repressors. *Molecular biology of bacteriophage* **4**, 193-205 (1994).

40 Luke, K. *et al.* Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* **299**, 182-191 (2002). https://doi.org:10.1006/viro.2002.1409

41 Nivinskas, R., Malys, N., Klausa, V., Vaiskunaite, R. & Gineikiene, E. Post-transcriptional control of bacteriophage T4 gene 25 expression: mRNA secondary structure that enhances translational initiation. *J Mol Biol* **288**, 291-304 (1999). https://doi.org:10.1006/jmbi.1999.2695

42 Eyler, D. E. *et al.* Pseudouridinylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* **116**, 23068-23074 (2019). https://doi.org:10.1073/pnas.1821754116

43 Gligorijevic, V. *et al.* Structure-based protein function prediction using graph convolutional networks. *Nat Commun* **12**, 3168 (2021). https://doi.org:10.1038/s41467-021-23303-9

44 Srikant, S., Guegler, C. K. & Laub, M. T. The evolution of a counter-defense mechanism in a virus constrains its host range. *Elife* **11** (2022). https://doi.org:10.7554/eLife.79549

45 Antonova, D. *et al.* The Dynamics of Synthesis and Localization of Jumbo Phage RNA Polymerases inside Infected Cells. *Viruses* **15** (2023). https://doi.org:10.3390/v15102096

46 Tabib-Salazar, A., Mulvenna, N., Severinov, K., Matthews, S. J. & Wigneshweraraj, S. Xenogeneic Regulation of the Bacterial Transcription Machinery. *J Mol Biol* **431**, 4078-4092 (2019). https://doi.org:10.1016/j.jmb.2019.02.008

47 Mayo-Munoz, D., Pinilla-Redondo, R., Camara-Wilpert, S., Birkholz, N. & Fineran, P. C. Inhibitors of bacterial immune systems: discovery, mechanisms and applications. *Nat Rev Genet* **25**, 237-254 (2024). https://doi.org:10.1038/s41576-023-00676-9

48 Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat Rev Microbiol* **21**, 686-700 (2023). https://doi.org:10.1038/s41579-023-00934-x

49 Kever, L. *et al.* Aminoglycoside Antibiotics Inhibit Phage Infection by Blocking an Early Step of the Infection Cycle. *mBio* **13**, e0078322 (2022). https://doi.org:10.1128/mbio.00783-22

50 Kronheim, S. *et al.* A chemical defence against phage infection. *Nature* **564**, 283-286 (2018). https://doi.org:10.1038/s41586-018-0767-x

51 Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077-1084 (2020). https://doi.org:10.1126/science.aba0372

52 Patel, P. H. *et al.* Anti-phage defence through inhibition of virion assembly. *Nat Commun* **15**, 1644 (2024). https://doi.org:10.1038/s41467-024-45892-x

53 Hobbs, S. J. *et al.* Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522-526 (2022). https://doi.org:10.1038/s41586-022-04716-y

54 Ma, W. *et al.* Enhancing Protein Function Prediction Performance by Utilizing AlphaFold-Predicted Protein Structures. *J Chem Inf Model* **62**, 4008-4017 (2022). https://doi.org:10.1021/acs.jcim.2c00885

55 Malone, L. M., Birkholz, N. & Fineran, P. C. Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Curr Opin Biotechnol* **68**, 30-36 (2021). https://doi.org:10.1016/j.copbio.2020.09.008

56 Wang, Y., Fan, H. & Tong, Y. Unveil the Secret of the Bacteria and Phage Arms Race. *Int J Mol Sci* **24** (2023). https://doi.org:10.3390/ijms24054363

57 Stanley, S. Y. *et al.* Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. *Cell* **178**, 1452-1464 e1413 (2019). https://doi.org:10.1016/j.cell.2019.07.046

58 Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol* **8**, 995-1005 (2007). https://doi.org:10.1038/nrm2281

59 Hamamsy, T. *et al.* Protein remote homology detection and structural alignment using deep learning. *Nat Biotechnol* (2023). https://doi.org:10.1038/s41587-023-01917-2

60 Vanni, C. *et al.* Unifying the known and unknown microbial coding sequence space. *Elife* **11** (2022). https://doi.org:10.7554/eLife.67667

61 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

62 Duan, N., Hand, E., Pheko, M., Sharma, S. & Emiola, A. Structure-guided discovery of anti-CRISPR and anti-phage defense proteins. *Nat Commun* **15**, 649 (2024). https://doi.org:10.1038/s41467-024-45068-7

63 Hinton, D. M. Transcriptional control in the prereplicative phase of T4 development. *Virol J* **7**, 289 (2010). https://doi.org:10.1186/1743-422X-7-289

64 Tao, P., Wu, X., Tang, W. C., Zhu, J. & Rao, V. Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synth Biol* **6**, 1952-1961 (2017). https://doi.org:10.1021/acssynbio.7b00179

65 Dong, J. *et al.* Engineering T4 Bacteriophage for In Vivo Display by Type V CRISPR-Cas Genome Editing. *ACS Synth Biol* **10**, 2639-2648 (2021). https://doi.org:10.1021/acssynbio.1c00251

66 Pozhydaieva, N. *et al.* Temporal epigenome modulation enables efficient bacteriophage engineering and functional analysis of phage DNA modifications. *bioRxiv*, 2024.2001. 2028.577628 (2024).

67 Kiro, R., Shitrit, D. & Qimron, U. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biol* **11**, 42-44 (2014). https://doi.org:10.4161/rna.27766

68 Liu, J. & Morrical, S. W. Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virol J* **7**, 357 (2010). https://doi.org:10.1186/1743-422X-7-357

69 Bleuit, J. S. *et al.* Mediator proteins orchestrate enzyme-ssDNA assembly during T4 recombination-dependent DNA replication and repair. *Proc Natl Acad Sci U S A* **98**, 8298-8305 (2001). https://doi.org:10.1073/pnas.131007498

70 Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front Biosci* **3**, D570-603 (1998). https://doi.org:10.2741/a304

71 Nelson, S. W. & Benkovic, S. J. The T4 phage UvsW protein contains both DNA unwinding and strand annealing activities. *J Biol Chem* **282**, 407-416 (2007). https://doi.org:10.1074/jbc.M608153200

72 Lagerback, P. & Carlson, K. Amino acid residues in the GIY-YIG endonuclease II of phage T4 affecting sequence recognition and binding as well as catalysis. *J Bacteriol* **190**, 5533-5544 (2008). https://doi.org:10.1128/JB.00094-08

73 Carlson, K., Lagerback, P. & Nystrom, A. C. Bacteriophage T4 endonuclease II: concerted single-strand nicks yield double-strand cleavage. *Mol Microbiol* **52**, 1403-1411 (2004). https://doi.org:10.1111/j.1365-2958.2004.04062.x

74 Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186**, 791-799 (2010). https://doi.org:10.1534/genetics.110.120782

75 Duong, M. M., Carmody, C. M., Ma, Q., Peters, J. E. & Nugen, S. R. Optimization of T4 phage engineering via CRISPR/Cas9. *Sci Rep* **10**, 18229 (2020). https://doi.org:10.1038/s41598-020-75426-6

76 Anderson, E. M. *et al.* Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity. *J Biotechnol* **211**, 56-65 (2015). https://doi.org:10.1016/j.jbiotec.2015.06.427

77 Du, J. *et al.* Enhancing bacteriophage therapeutics through in situ production and release of heterologous antimicrobial effectors. *Nat Commun* **14**, 4337 (2023). https://doi.org:10.1038/s41467-023-39612-0

78 Meile, S. *et al.* Engineered reporter phages for detection of Escherichia coli, Enterococcus, and Klebsiella in urine. *Nat Commun* **14**, 4336 (2023). https://doi.org:10.1038/s41467-023-39863-x

79 Adler, B. A. *et al.* Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nat Microbiol* **7**, 1967-1979 (2022). https://doi.org:10.1038/s41564-022-01258-x

80 Guan, J. *et al.* Bacteriophage genome engineering with CRISPR-Cas13a. *Nat Microbiol* **7**, 1956-1966 (2022). https://doi.org:10.1038/s41564-022-01243-4

81 Molina Vargas, A. M. *et al.* New design strategies for ultra-specific CRISPR-Cas13a-based RNA detection with single-nucleotide mismatch sensitivity. *Nucleic Acids Res* **52**, 921-939 (2024). https://doi.org:10.1093/nar/gkad1132

82 Tiemann, B. *et al.* ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* **186**, 7262-7272 (2004). https://doi.org:10.1128/JB.186.21.7262-7272.2004

83 Wolfram-Schauerte, M. *et al.* A viral ADP-ribosyltransferase attaches RNA chains to host proteins. *Nature* **620**, 1054-1062 (2023). https://doi.org:10.1038/s41586-023-06429-2

84 Sorensen, M. A., Fricke, J. & Pedersen, S. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo. *J Mol Biol* **280**, 561-569 (1998). https://doi.org:10.1006/jmbi.1998.1909

85 Duval, M. *et al.* Escherichia coli ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol* **11**, e1001731 (2013). https://doi.org:10.1371/journal.pbio.1001731

86 Qu, X., Lancaster, L., Noller, H. F., Bustamante, C. & Tinoco, I., Jr. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc Natl Acad Sci U S A* **109**, 14458-14463 (2012). https://doi.org:10.1073/pnas.1208950109

87 Voynow, P. & Kurland, C. G. Stoichiometry of the 30S ribosomal proteins of Escherichia coli. *Biochemistry* **10**, 517-524 (1971). https://doi.org:10.1021/bi00779a026

88 Diedrich, G. *et al.* Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding to A and P sites and peptidyl transfer. *EMBO J* **19**, 5241-5250 (2000). https://doi.org:10.1093/emboj/19.19.5241

89 Lambert, J. M. & Traut, R. R. The subunit interface of the Escherichia coli ribosome. Identification of proteins at the interface between the 30 S and 50 S subunits by crosslinking with 2-iminothiolane. *J Mol Biol* **149**, 451-476 (1981). https://doi.org:10.1016/0022-2836(81)90481-2

90 Domenighini, M. & Rappuoli, R. Three conserved consensus sequences identify the NAD-binding site of ADP-ribosylating enzymes, expressed by eukaryotes, bacteria and T-even bacteriophages. *Mol Microbiol* **21**, 667-674 (1996). https://doi.org:10.1046/j.1365-2958.1996.321396.x

91 Duret, L., Gasteiger, E. & Perriere, G. LALNVIEW: a graphical viewer for pairwise sequence alignments. *Comput Appl Biosci* **12**, 507-510 (1996). https://doi.org:10.1093/bioinformatics/12.6.507

92 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-+ (2022). https://doi.org:10.1038/s41592-022-01488-1

93 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022). https://doi.org:10.1038/s41592-022-01488-1

94 Foley, G. *et al.* Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLoS Comput Biol* **18**, e1010633 (2022). https://doi.org:10.1371/journal.pcbi.1010633

95 Hochberg, G. K. A. & Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu Rev Biophys* **46**, 247-269 (2017). https://doi.org:10.1146/annurev-biophys-070816-033631

96 Yu, X. *et al.* Messenger RNA 5' NAD(+) Capping Is a Dynamic Regulatory Epitranscriptome Mark That Is Required for Proper Response to Abscisic Acid in Arabidopsis. *Dev Cell* **56**, 125-140 e126 (2021). https://doi.org:10.1016/j.devcel.2020.11.009

97 Bird, J. G. *et al.* Highly efficient 5' capping of mitochondrial RNA with NAD(+) and NADH by yeast and human mitochondrial RNA polymerase. *Elife* **7** (2018). https://doi.org:10.7554/eLife.42179

98 Walters, R. W. *et al.* Identification of NAD+ capped mRNAs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **114**, 480-485 (2017). https://doi.org:10.1073/pnas.1619369114

99 Gomes-Filho, J. V. *et al.* Identification of NAD-RNA species and ADPR-RNA decapping in Archaea. *Nat Commun* **14**, 7597 (2023). https://doi.org:10.1038/s41467-023-43377-x

100 Aravind, L., Zhang, D., de Souza, R. F., Anand, S. & Iyer, L. M. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. *Curr Top Microbiol Immunol* **384**, 3-32 (2015). https://doi.org:10.1007/82_2014_414

101 Koch-Nolte, F., Kernstock, S., Mueller-Dieckmann, C., Weiss, M. S. & Haag, F. Mammalian ADP-ribosyltransferases and ADP-ribosylhydrolases. *Front Biosci* **13**, 6716-6729 (2008). https://doi.org:10.2741/3184

102 Feng, B., Liu, C., Shan, L. & He, P. Protein ADP-Ribosylation Takes Control in Plant-Bacterium Interactions. *PLoS Pathog* **12**, e1005941 (2016). https://doi.org:10.1371/journal.ppat.1005941

103 Mikolcevic, P., Hlousek-Kasun, A., Ahel, I. & Mikoc, A. ADP-ribosylation systems in bacteria and viruses. *Comput Struct Biotechnol J* **19**, 2366-2383 (2021). https://doi.org:10.1016/j.csbj.2021.04.023

104 Fehr, A. R. *et al.* The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. *Genes Dev* **34**, 341-359 (2020). https://doi.org:10.1101/gad.334425.119

Acknowledgements

This work would not have been possible without the support of many people who have accompanied me on my PhD journey.

I would like to express my deepest gratitude to my PhD supervisor, Katharina Höfer, for the honor of being her first PhD student and for all the knowledge and insights she has given me along the way. Thank you for constantly inspiring me, believing in me, and for all the scientific discussions and cake breaks at the weekend. I am grateful that you convinced me, within seconds of our first meeting, that phages are a fascinating research topic – I could not agree more with you now. Also, I could not wish for a better scientific advisor than you have been these last years. Thank you. I will always remember to keep my wings wide open on my back.

I would also like to thank the members of my thesis advisory committee, Lennart Randau and Timo Glatter, for their continuous advice and support. I want to thank all the collaborators who supported my work scientifically. Special thanks to Nicole Paczia and Daniel Schindler for being supportive and reliable. I also want to thank Timo Glatter again for generously sharing his valuable knowledge on proteomics and his patience with the infinite number of samples I provided, which (luckily) rarely upset the column, and for all the enjoyable conversations that followed the handover of samples. Additionally, I want to thank Jörg Kahnt for his scientific support and advice.

I want to thank Dušica Radoš for her support and being the best IMPRS coordinator ever.

Thanks to my ever-best colleagues, who made these three and a half years absolutely awesome and always supported me. I want to thank to Maik for being so reliable, for all our collaborations, his trust, and the hours of discussion and fun we shared; Nurseda – a friend that I couldn't even dream of having before, for her constant support and incredible food. David – for being a wonderful friend, proven by tolerating all my jokes. Petra – for being an angel in person and magically making everything work that did not work for me in the lab. Helene – for the inspiring calmness and delightful conversations. Moritz – for being such an amazing person, surprising us all daily with some random extraordinary knowledge. Elyés – for sharing wonderfully optimistic view of the world. You are all just absolutely wonderful and supportive people; I could not have wished to be surrounded by better colleagues.

This work would not have been possible without the support and contributions of my students: Madita Viering, who has been with me since the very first day of my PhD, always making me laugh and enjoying all our experiments together; Franziska Billau, a 154 cm “pure Lebensfreude”; Tristan Krause, incredibly curious and critical; Lena Schuck, always energizing; Rusica Sedic, honest and empathetic person. I am extremely thankful to all of you, proud and inspired by all of you.

I would like to thank Franziska Sendker, my first friend in Marburg, an absolutely wonderful, supportive person and a great scientist. Thank you for always being there for me. I also want to

thank Rebecca Hinrichs. It was always enjoyable to work with you and discuss science. I am constantly inspired by your strength and incredibly grateful for your constant encouragement.

I want to thank Julia Frunzke for her support, advice, and inspiration. Thank you for maintaining such a wonderful phage family in Germany.

I also want to thank Claudia, Christina, Inka, Ali, Sarah, and Sebastian from the department for their support.

Thank you to my friends Bianca, Helena, Irma, and Rebecca for their support and being there all these years.

Thanks to my parents, who are, even being far away from me and going through the absolutely worst times of their lives, always had the energy and power to care about me, support me, motivate me and distract me. Thank you for shaping me into the person I am. I also want to thank my brother Pavlo for inspiring me constantly with your example; I am incredibly proud of you. Thank you to my entire wonderful family for always being there for me and caring about me.

I want to thank to Horst and Elke, who believed in me ten years ago and made it possible for me to study. Thank you for supporting me so incredibly along the way and for becoming my second parents. Thank you for making me feel at home and welcome. You are some of the most wonderful people I have ever met.

I want to express my deep gratitude to my husband Moritz for his unwavering support, engaging scientific discussions, and constant belief in me. Thank you for supporting all my decisions and being there for me. I also thank the entire Weber family for their support and belief in me.