All domains of life utilize a diverse set of modified ribonucleotides that can impact the sequence, structure, function, stability, and the fate of RNAs, as well as their interactions with other molecules. Today, more than 160 different RNA modifications are known 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 is 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. Here, it is described that 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, still unanswered questions in the field of prokaryotic epitranscriptomics are pointed out, and its future perspectives in the dawn of next-generation sequencing technologies are outlined.

1. 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 more than 160 known RNA modifications today.[3] These nucleotide derivates vary in their complexity, ranging from simple methylations to cyclizations, large group additions, and glycosylations.[2] The versatile and numerous modifications of RNA within a cell are collectively termed as “epitranscriptome.”[3] The complexity of the epitranscriptome is reflected by the various functions of the post- and cotranscriptional modifications, which can affect RNA structure, stability, and cellular fate.[4] 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.[5] 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 as 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.[6] 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.[7] 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 N6-Methyladenosine (m6A) and inosine (I) that are known to decorate prokaryotic mRNA. [8] Moreover, pseudouridine (Ψ), 5-methylcytosine (m5C), 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 noncanonical RNA-caps. Here, nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD), coenzyme A (CoA), uridine diphosphate N-acetyl glucosamine (UPD-GlcNAc), thiamine, dinucleoside polyphosphate (NpnN), 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.

2. Internal RNA Modifications

The noncanonical RNA nucleosides m6A, I, Ψ, m5C, and Nm are widely distributed in all kingdoms of life.[9] 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 m6A and inosine were identified in prokaryotic mRNA to date.[8]

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 m6A, I, Ψ, m5C, 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.

2.1. 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.[10] (Figure 3). This modification is equally distributed in eukaryotic cells throughout various RNA classes.[11,12] In prokaryotes, the appearance of Ψ is restricted to tRNA and rRNA.[13]
2.1.1. 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.[14] In such a way, the specific pseudouridylation sites were identified for the first time within yeast tRNA.[15] 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,[16] Ψ-Seq,[17] PSI-Seq,[18] and CeU-Seq.[19] The derivatized N3-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 N3-CMC-Ψ. Thus, CMCT derivatization of RNA was successfully performed to confirm the presence of Ψ35 in *Escherichia coli* tRNA.[20] However, besides all advantages of CMCT derivatization, low abundant pseudouridylation events may be missed. The presence of N3-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.[21] Khoddami et al. use a Ψ-monobisulfite addition with subsequent heat-induced ribose ring-opening and Mg²⁺-assisted reorientation that

---

**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.
| Modification          | Structure | Detection                        | Writers, readers, erasers                                                                 | Molecular roles                                                                 |
|-----------------------|-----------|-----------------------------------|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Pseudouridine (Ψ)     | ![Structure](image) | Pseudo-Seq, V-Seq, Psi-Seq, CeuU-Seq, RB5-Seq, HydrApSeq | Writers: site-specific PUSs and snoRNA guided box H/ACA, PNP, Eukaryotic reader: MetRS | Structure stabilization of lRNA and rRNA, translation modulation, alteration of codon translation, mRNA stability |
| N'-methyladenosine (mA) | ![Structure](image) | mA-Seq, MeRIP-Seq, mA-UAC-Seq, PA- mA-Seq, mi-CLIP, mA-label-Seq, DART-Seq, NanoCompore | Writers: RmF, RmJ (prokaryotes), METTL3, METTL4, WTAP (eukaryotes), Eukaryotic erasers: FTO, ALKBH5 (eukaryotes), Eukaryotic readers - YTHDF1-3/4 proteins | In eukaryotes involved in mRNA stability, RNA-protein interactions, mRNA processing, splicing, translation initiation and efficiency regulation. |
| Inosine (I)           | ![Structure](image) | ICE-Seq, EndoVIPER-Seq            | Writers: adenosine deaminases that act on RNA (ADAR) or tRNA (ADAT) in eukaryotes, Prokaryotic homologous – TadA, acts on mRNA and mRNA Reader: endonucleaseV | Codon alterations, alternative splicing, regulation of stress response, alternative splicing and 3′-UTR-variation in eukaryotes |
| 5-methylcytidine (m5C) | ![Structure](image) | Bisulfite sequencing, RIP-Seq, Aza-IP, NSUN-2, 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 hm5C, Eukaryotic readers: ALYREF, YBX1, YPS | tRNA structure stabilization, translational fidelity of ribosome |
| 2′-O-methylation (Nm) | ![Structure](image) | RT with limited dNTPs, 2OMe-Seq, Ribometh-Seq, Nm-Seq, RibCox-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.
causes base-skipping during cDNA synthesis. Further RNA modifications, such as m1C and m1A, 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 residues in Ψ. Recently, the so-called HydraPsiSeq was established and verified for mapping residues in Ψ in Saccharomyces cerevisiae and Homo sapiens.[22] 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 toward 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.[23–25] However, the majority of methods were developed for the identification of Ψ in eukaryotes. Studies focusing on prokaryotic pseudouridylation are still missing.

2.1.2. Writers, Erasers, and Functions of Ψ

The isomerization of uracil to Ψ is post-transcriptionally catalyzed by so-called pseudouridine synthases in bacteria.[26] For example, the prokaryotic model organism E. coli contains 11 pseudouridine synthases, of which four are tRNA specific, six are rRNA specific, and one is observed to catalyze the pseudouridylation on both tRNA and rRNA.[27]

Independent of the pseudouridine synthase class, generally, two distinct reactions occur during U to Ψ isomerization—the cleavage of N1=C′ glycosidic bond that allows for a base rotation by 180° as well as the subsequent formation of C2=C′ glycosidic bond.[28] The resulting nucleotide possesses partially changed chemical properties and altered biophysical characteristics. As a result, the introduction of Ψ into tRNA and rRNA 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.[29] Second, the RNA duplex stability is improved by enhanced base stacking between the Ψ-A base pair.[30] 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.[31] Moreover, in E. coli, abolished Ψ levels are known to inhibit the translation.[31,32] In eukaryotes, pseudouridylation of mRNA can modulate important cellular processes, in particular translation. Ψ-containing codons affect the translation of cognate codons.[33] Interestingly, the translation efficiency of proteins increases in the presence of Ψ containing mRNA in eukaryotes.[34] 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.[35] 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.[33] However, the first Ψ reader protein, methionine aminocyl tRNA Met synthetase (MetRS), was recently discovered in S. cerevisiae.[35] 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.

3. N6-Methyladenosine (m6A)

N6-Methyladenosine (m6A) RNA methylation was first observed by Perry and Kelley in 1974 and later recognized to be the one of the most abundant post-transcriptional mRNA modifications. m6A is widely distributed in the eukaryotic, prokaryotic, and viral transcriptome[36] (Figure 3). While in eukaryotes, m6A modification is present in most classes of RNA, including mRNA, ncRNA, tRNA, and rRNA,[37] in bacteria, only rRNA, mRNA, and ncRNA were observed carrying m6A modification so far.[38]

2.2.1. Identification, Quantification, and Validation of m6A

The m6A has almost identical chemical properties as its precursor adenosine. Therefore, selective chemical derivatization of either A or m6A, which induce RT aberration, is not possible. For this reason, methods for transcriptome-wide m6A mapping based on principles other than chemical derivatization are required.

The first developed transcriptome-wide adenosine methylome mapping approaches, such as m6A-Seq, MeRIP (m6A methylated RNA immunoprecipitation)-Seq, and m6A-LAIC (m6A-level and isoform-characterization)-Seq, were based on the immunoprecipitation of m6A-containing transcripts with an anti-m6A antibody followed by subsequent high-throughput sequencing.[38–40] 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 m6A mapping. Furthermore, the antibody-based approaches have a resolution of ~100–200 nt, which hampers the m6A detection at single-base level. The low resolution of m6A mapping can be improved to ~23 nt by the application of PA-m6A (photo-crosslinking-assisted m6A)-Seq.[41] 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-m6A-antibody. Afterward, 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 mi-CLIP (m^6^A individual-nucleotide-resolution cross-linking and immunoprecipitation)-Seq approach. 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^6^A-adjacent C to T during RT. Similar to the described methods, mi-CLIP requires an anti-m^6^A-antibody. Thus, the specificity of the latter is crucial for the correct identification of m^6^A sites.

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

In 2020, Shu et al. developed the m^6^A-label-Seq approach to address these issues. m^6^A-label-Seq is a single-base resolution technique. Cells are grown in the presence of Se-allyl-l-selenohomocysteine (SeAM). Eukaryotic m^6^A writers accept SeAM as a methyl donor cofactor, thereby transferring the allyl group to adenosine, generating N^6^-allyladenosine (a^6^A) instead of m^6^A. Chemically induced cyclization of a^6^A results in a nucleotide misincorporation during RT allowing for the detection of m^6^A at single-base level. So far, m^6^A-label-Seq was applied to human cell lines. By feeding the cells with SeAM, only 10% of the naturally occurring m^6^A sites were substituted with a^6^A. In contrast to other m^6^A detection methods, m^6^A-label-Seq is RNA motif-independent and can be used to identify clustered m^6^A sites compared to the methods described before. This approach can potentially be applied to study the role of prokaryotic m^6^A mRNA. However, the acceptance of the SeAM by prokaryotic methyl transferases and the uptake of this artificial cosubstrate by the prokaryotic cell have to be studied first.

The constraints coincide with metabolic m^6^A labeling strategies that can be circumvented by transiting to chemical labeling approaches, such as m^6^A-SEAL-Seq. m^6^A-SEAL-Seq utilizes the human originated m^6^A eraser FTO, which oxidizes m^6^A to N^6^-hydroxymethyladenosine (hm^6^A). The latter can be further converted to N^6^-dithiolsitolmethyladenosine (dm^6^A) 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. MazF specifically cleaves RNA at unmethylated sites upstream of the ACA sequence motif. 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). 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 polymerase chain reaction (PCR) amplification prior to high-throughput sequencing, which introduces bias in sample composition. 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. Recently, the DRS method was applied to identify post-transcriptional RNA modifications. Leger et al. developed a NanoCompare approach for m^6^A mapping in the mammalian transcriptome. In this approach, two DRS datasets (nontreated control sample vs 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.

2.2.2. 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. 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. In prokaryotes, the m^6^A modification is prevalently abundant within open reading frames (ORFs) (72%) and rarely present at the termini of mRNA. Both the function and the regulation of m^6^A 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.\[38,48\] Eukaryotic m^6A modifications are described to be involved in mRNA stability, RNA–protein interactions, miRNA processing, splicing, initiation, and regulation of translational efficiency.\[62,8a,49\] The methylation of adenosine in eukaryotes is highly regulated by writers METTL3 and METTL14.\[56\] FTO and ALKBH5 were identified being the eukaryotic erasers of m^6A.\[31\] Eukaryotic YTHDF1-3 proteins are the readers and selectively bind to m^6A mRNA and promote its decay and regulate the translation.\[52\] 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 as DART (deamination adjacent to RNA modification targets)-Seq.\[53\] 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 (∼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.

2.3. 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.\[54\] 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,\[55\] I-to-A modification in RNA is significant and essential for its biological function.

2.3.1. 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.\[56\] 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.\[57\]

Higher accuracy in the mapping of inosine was achieved by ICE (inosine chemical erasing)-Seq.\[38\] 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.\[59\] In this approach, Knutson et al. applied the E. coli inosine reader EndoVIPER to enrich inosine-containing transcripts. EndoVIPER 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.\[60\] Interestingly, EndoVIPER 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 EndoVIPER 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 EndoVIPER, the application of this technique can enable the identification of previously unknown inosine sites.

2.3.2. Writers, Erasers, and Functions of Inosine

The C6 deamination of adenosine to inosine in RNA can occur either by spontaneous hydrolysis or enzymatic conversion.\[60\] Spontaneous deamination can be caused by environmental factors, such as exposure to nitrosative compounds.\[61\] The enzymatic mediated A-to-I editing is specifically catalyzed by adenosine deaminases.\[61\] Within eukaryotes, adenosine deaminases are classified into adenosine deaminases acting on tRNA (ADAT) and adenosine deaminases acting on mRNA (ADAR).\[61,62\] Both deaminase classes possess double-stranded RNA binding domains.\[63\] Prokaryotic TadA protein was initially identified as a bacterial ADAT analog. For a long time, its deamination activity was assumed to be restricted to tRNA,\[64\] precisely tRNA^AGG^ACG and tRNA^AAG^AGG.\[65\] 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.\[86\] 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.\[86\]

In 2020, Nie et al. proceeded with the discovery of inosine in prokaryotes.\[86\] 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 subsequent sequencing (iRIP-Seq).\[86\] 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 toward 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.

### 2.4. 5-Methylcytidine (m\(^5\)C)

5-Methylcytidine (m\(^5\)C) is a modification initially identified in DNA and later observed in RNA.\[^{66}\] 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\[^{67}\] (Figure 3). While the m\(^5\)C 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.\[^{68}\] In eukaryotic mRNA, m\(^5\)C was described to be involved in the modulation of mRNA nuclear export and regulation of protein translation.\[^{69}\]

#### 2.4.1. Identification, Quantification, and Validation of 5-Methylcytidine

The first discovery of m\(^5\)C was triggered by the application of different biophysical technologies, among them chromatography-based approaches, mass spectrometry (MS), and NMR.\[^{70}\] However, all these methods are not applicable for high-throughput sequencing and single-base m\(^5\)C 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\(^5\)C in DNA, was adapted by Schäfer et al. to specifically map m\(^5\)C RNA modifications.\[^{71}\] 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\(^5\)C can be achieved. However, other cytidine modifications such as 3-methylcytidine, N4-methylcytidine, N4,2'-O-dimethylcytidine, and N4-acetylated variants can prevent the C-to-U conversion by bisulfite.\[^{72}\]

Bisulfite-Seq was applied to identify m\(^5\)C modification in archaeal mRNA (Solfolobus solfataricus), whereupon the presence of m\(^5\)C in bacterial mRNA could not be verified with this approach.\[^{68b}\]

To improve the quality of m\(^5\)C detection, alternative m\(^5\)C 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)\[^{72}\] and miCLIP (methylation individual-nucleotide-resolution crosslinking and immunoprecipitation)\[^{73}\] utilize eukaryotic m\(^5\)C RNA writers NSUN2 and DNMT2. The Aza-IP method includes metabolic RNA labeling\[^{74}\] and is based on the ability of m\(^5\)C methyltransferases to form a covalent enzyme-substrate intermediate with their targets.\[^{75}\] In this approach, 5-azacytidine (5-azaC) is incorporated into the RNA in vivo. The applied cytidine analog is a suicide inhibitor of the m\(^5\)C 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\(^5\)C RNA methyltransferases NSUN2 and DNMT2. However, m\(^5\)C 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 5-azaC is low.\[^{73}\]

miCLIP does not require metabolic labeling of RNA and is based on the application of a modified m\(^5\)C writer.\[^{73}\] Here, an irreversible covalent bond between the m\(^5\)C 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\(^5\)C modification in prokaryotes remain unknown.

Apart from the immunoprecipitation of the writer-RNA complex, the m\(^5\)C-containing RNA can be directly enriched by RIP (RNA immunoprecipitation)-Seq.\[^{68b}\] In this approach, an m\(^5\)C-specific antibody is applied. However, similar to all other antibody-based techniques, RIP-Seq is limited by the antibody specificity toward m\(^5\)C modification.\[^{66d}\] Also, the m\(^5\)C antibody cannot distinguish between m\(^5\)C 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\(^5\)C modification in the prokaryotic transcriptome is still under investigation.

#### 2.4.2. Writers, Erasers, and Functions of 5-Methylcytidine

The introduction of m\(^5\)C modification into eukaryotic RNA is catalyzed by the SAM-dependent methyltransferases NSUN (NSUN1-NSUN7) and DNMT2.\[^{74,76}\] Next to these writers, the proteins involved in the removal of m\(^5\)C are identified in eukaryotes. For instance, ten-eleven translocation (TET) enzymes were reported to erase the m\(^5\)C from RNA by its oxidation to 5-hydroxymethyl cytidine (hm\(^5\)C).\[^{77}\] Recently, eukaryotic m\(^5\)C readers ALYREF, YBX1, and YPS were identified.\[^{78}\] All of them exhibit preferential binding to m\(^5\)C-containing mRNA and promote its cytoplasmic export. Based on these results, m\(^5\)C 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\(^5\)C existence in bacterial mRNA was still not proved. Therefore, writers, readers, and erasers of m\(^5\)C were not identified in bacteria so far.

#### 2.5. 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.\[^{79}\] 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.\cite{79b,80}

2.5.1. 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.\cite{81} 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.\cite{82} 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.\cite{83} In combination with liquid chromatography–mass spectrometry (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.\cite{83}

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.\cite{84a} 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 mRNA.\cite{86b}

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.\cite{84a} 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.\cite{84a} However, this approach cannot be used quantitatively to compare methylation intensity between different sites.

Furthermore, the presence of Nm in RNA affects the reverse transcriptase. Interestingly, under limited dNTP concentrations, Nm-modified residues trigger specific pausing of the reverse transcriptase.\cite{85} In this approach, called 2OMe-Seq,\cite{85} RT reactions are performed in the presence of high and low dNTPs concentrations using the same RNA as a template for the reverse transcriptase. The subsequent 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.

2.5.2. Writers, Erasers, and Functions of 2′-O-Methylation

Generally, the introduction of Nm modification into RNA is performed either by a stand-alone methyltransferase\cite{86a} or a snoRNA-guided 2′-O-methylase.\cite{87} 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.\cite{88} 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.\cite{88b,89} One of these modification sites in bacterial tRNA, particularly Gm18, was involved in suppressing the innate human immune activation.\cite{79b}

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.\cite{34,79d,90} The position of Nm within eukaryotic mRNA affects the extent of translation inhibition.\cite{90} 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.

3. 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.\cite{91}

Apart from these different triphosphate-derived 5′-ends of RNA, viral and eukaryotic mRNAs possess a 5′-cap.\cite{92} This RNA cap comprises an N7-methylguanosine linked to the first nucleotide of the RNA via a 5′-to-5′ triphosphate (5′-m7GpppN). 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.\cite{91} The functions of this eukaryotic cap are associated with pre-mRNA splicing, poly(A)-tailing, nuclear mRNA export, translation, and mRNA stability.\cite{91}

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. These primary transcripts are characterized by a 5′-hydroxyl group (5′-OH-RNAs).

3.1. 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. 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). Afterward, 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 (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. This dRNA-seq approach has already been applied to various bacterial species such as Staphylococcus aureus and Bacillus subtilis or E. coli. Further, it has already been used to identify TSSs in host and phage genomes during infection.

Cappable-seq is another sequencing technique to detect TSSs. Here, the vaccinia virus capping enzyme uses 3′-biotinylated GTP as a substrate to specifically cap 5′-PPP- and 5′-PP-RNA, while 5′-P-RNA remains unmodified. Capped and biotinylated RNA is subsequently enriched by streptavidin pull-down. Comparing enriched and nonenriched samples bioinformatically allows for the identification of primary transcripts and their TSSs. Thereby, 41% of novel TSSs in E. coli could be annotated. Also, Cappable-seq has proven suitable for meta-transcriptome analysis, as it enabled the mapping of TSSs in the mouse gut microbiome. SMRT (single molecule, real time)-Cappable-seq is adjusted to sequence the intact, full-length primary transcripts. The classical Cappable-seq relies on the fragmentation of cDNA before Illumina sequencing. To determine TSSs of entire

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′-PPP-RNAs with the PACO assay relies on the specific protection of 5′-PPP-RNAs and the ligation of adapters. 5′-PPP-RNAs are m7G-capped by the Guanylyltransferase (1). Afterward uncapped RNAs are dephosphorylated to 5′-OH-RNA by alkaline phosphatase (2). Pyrophosphatase converts capped RNAs into 5′-P-RNAs (3). 5′-P-RNAs are ligated to an adapter (PABLO-assay) (4). Shifted bands in the Northern blot indicate 5′-PPP-RNAs (5).
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. 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 (TTTs). These findings suggest the existence of more operon variants than previously annotated, which may be determined by properties of the respective TTTs and growth conditions. The latter could resemble a gene regulatory mechanism that defines the number of cistrons per operon to fine-tune metabolic activities. 

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. They are used by eukaryotic, prokaryotic as well as archaeal RNA polymerases (RNAPs) as substrates in vitro. The first demonstration of nanoRNA-driven transcription initiation in vivo has been provided by studies of *Pseudomonas aeruginosa*. Here, nanoRNAs are frequently used for transcription initiation upon their stabilization. Further studies have demonstrated their presence in *E. coli* and *Vibrio cholerae*. Transcripts derived from nanoRNA-mediated transcription initiation typically carry a 5′-OH-group. nanoRNA-derivable transcripts can be identified by the ectopic expression of an oligoRNase coupled with 5′-RNA-seq. 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.

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

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

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 (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," an ssDNA complementary to the 5′-end of the RNA and the 3′-end of the DNA oligo, enables the specific ligation of the adapter. Afterward, the ligation product and the nonligated product can be analyzed by Northern blotting to evaluate the levels of 5′-PP-RNA. The fraction of 5′-PP-RNA for a particular transcript can be quantified using the phosphorylation assay of capping outcome (PACO). In the first step, 5′-PP-RNA is capped by the eukaryotic guanylyltransferase while other RNA 5′-termini remain unmodified (Figure 4C). Afterward, 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. 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. 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).

### 3.2. 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. Several studies have indicated that these cofactors are used as so-called noncanonical initiating nucleotides (NCINs) by the bacterial RNAP. Therefore, the cofactors’ adenosine moiety base pairs with the complementary strand at the +1 position if the transcript initiates with an adenosine. 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).

#### 3.2.1. 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. 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. The abundance of these 5′-NAD-capped RNAs (NAD-RNAs) was estimated to 3000 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. This method, termed as NAD captureSeq (Figure 6A), initially labels NAD-capped RNA with an alkyn 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 alkyn-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 was subjected to NGS (Illumina). 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.
| Modification | Structure | Detection | Writers, readers, erasers | Molecular roles |
|--------------|-----------|-----------|---------------------------|-----------------|
| NAD          | ![NAD Structure](image) | NAD captureSeq, NAD-tagSeq, Mass spectrometry, NAD-capQ | Incorporation by RNAP during transcription initiation, Decapping by nuclease hydrases NudC in *E. coli* and RppH in *B. subtilis*, DeNADing via DXO/Ral 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          | ![FAD Structure](image) | Mass spectrometry, FAD-capQ | Incorporation by RNAP during transcription initiation in vitro, DeFADing in eukaryotes by DXO/Ral family proteins | Decoration of 5′-termini of predominantly sRNAs; RNA sequences remain unidentified |
| CoA          | ![CoA Structure](image) | Mass spectrometry | Incorporation by RNAP during transcription initiation in vitro, Decapping by nuclease hydrase NudC, DeCoAping via eukaryotic DXO/Ral family proteins | Decoration of 5′-termini of predominantly sRNAs; RNA sequences remain unidentified, RNA protection against 5′-end dependent RNA processing and nuclease hydrase RppH |
| UDP-GlcNac/ UPD-Glc | ![UDP-Glc Structure](image) | Mass spectrometry | Incorporation by RNAP during transcription initiation in vitro | RNA sequences remain unidentified |
| ATH (Thiamine) | ![ATH Structure](image) | None | Incorporation by T7 RNAP during transcription initiation in vitro | Existence in vivo and RNA sequences remain unidentified |
| ApoA (Dinucleoside tetraphosphates) | ![ApoA Structure](image) | Mass spectrometry, Northern blotting | Incorporation by RNAP during transcription initiation; capping by mRNA 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 | ![5′-phospho-ADP-ribose Structure](image) | None | Deposition post-transcriptionally by Tpt1 from 5′-P-RNA and NAD | Existence in vivo and RNA sequences remain unidentified |

Figure 5. An overview about 5′-terminal RNA modifications in prokaryotes.
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.
Since the development of the NAD captureSeq technology, this method has enabled the discovery of NAD-RNA in other organisms from the prokaryotic world.[109] In B. subtilis, mostly mRNAs are NAD-capped. The overall abundance of these 5′-modified RNAs amounts to ≈6.3 fmol µg⁻¹ RNA, which corresponds to 220 copies per cell.[110a] In Staphylococcus aureus, mRNAs and sRNAs are NAD-capped that account for 25 fmol µg⁻¹ RNA, ≈900 NAD-RNAs per cell.[110b]

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,[111] Arabidopsis thaliana,[112] and human embryonic kidney (HEK) cell lines.[113]

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.[114] Further, the combination of SPAAC-NAD-Seq with mΓ-G-RNA depletion tackles the issue of the NAD captureSeq reactivity toward these canonically capped eukaryotic mRNAs.[114]

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 NAD-RNA with pentyn-1-ol but uses a 3′-azide-RNA instead of NAD-RNA, which competitively binds to the full-length RNA.[115] Former NAD-RNA is thereby 5′-labeled 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 nonenriched samples.[115a,115b]

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.[116]

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′-PP-P-RNA, initiating with adenosine, is coupled to nicotinamide mononucleotide (NMN) via a nucleophilic substitution reaction.[117] 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).[106a,110] This fact is supported by structural studies of the E. coli RNAP, which binds NAD similar to ATP.[106b] Thereby, NAD-capping by the RNAP occurs ab initio during transcription. As the adenine 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.[106c] 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.[108] 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. While NAD capping is not affected by sigma factors,[110] preferentially A and T are most frequently located.
at the −1 position of respective NAD-capped transcripts in B. subtilis and S. aureus.\[^{110}\] It has been shown that the −1 position influences the incorporation efficiency of NAD by the respective RNAP.\[^{110}\] This is supported by structural studies of the NADP during transcription initiation.\[^{106a}\]

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.\[^{122}\] 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—\[^{123}\] 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.\[^{111}\]

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.\[^{108,110a}\] 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.\[^{124}\] This RNA-cleavage product may be degraded more easily by nucleases such as RNase E.\[^{108,114}\] Homologs of this decapping enzyme exist in yeast (NPY1)\[^{125}\] 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.\[^{108,110a}\] It has been shown that the RNA polymerase can use NAD instead of ATP in order to start transcription of +1A transcripts.\[^{110}\] 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.\[^{111}\] 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.

![Figure 7](image-url)

**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.

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. While 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. FADH\(_2\) is an important electron carrier within the electron transport chain (ETC) and plays a central role in energy metabolism. In addition to its role as an electron donor, FAD also functions as a redox cofactor in a variety of enzymes, including those involved in biosynthesis, degradation, and signaling pathways. The redox potential of FADH\(_2\) is lower than that of NADH, allowing it to donate electrons to a variety of acceptors.

In the context of transcription, FAD has been shown to play a role in transcription initiation and elongation. The T7 RNA polymerase (RNAP) is known to accept NAD as a cap analog, and FADH\(_2\) has been shown to enhance the processivity of T7 RNAP. This effect is thought to be due to the ability of FADH\(_2\) to donate electrons to the enzyme, thereby regulating its activity. Moreover, FADH\(_2\) has been shown to influence the incorporation efficiency of NAD and other NCINs by the RNAP during transcription initiation.\[^{110}\] It has been shown that the −1 position influences the incorporation efficiency of NAD by the respective RNAP.\[^{110}\] This is supported by structural studies of the NADP during transcription initiation.\[^{106a}\]

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 tRNAs are NAD-capped in E. coli, mRNAs are NAD-capped in B. subtilis and S. aureus.\[^{108,110}\] 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.\[^{110a}\] 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.\[^{112a}\] Nevertheless, NAD-RNA functions beyond stabilization remain to be elucidated in prokaryotes.\[^{112a}\]

### 3.2.2. 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. While 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 noncanonical cap of RNA in 2003 (Figure 5). Huang showed that the T7 RNAP accepts FAD as an NCIN.\[^{107}\] 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. Consequently, FAD-capped RNA (FAD-RNA) may also exist in E. coli. The existence of FAD-RNA in E. coli total RNA has already been verified by CapQuant. Thereby, the cellular FAD-RNA concentration was quantified to 0.17 fmol μg⁻¹ E. coli RNA—more than tenfold lower than the amount of NAD-RNA.

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. This technique quantifies FAD-RNA in total RNA using Schizosaccharomyces pombe SpRai1 (SpRai1) which removes intact FAD from FAD-RNA in vitro (Figure 8A). Thereby, FAD is released from RNA and serves as a substrate for an oxidase reaction of a commercially available fluorometric assay. In HEK 293T cells, FAD-capQ identified 0.33 fmol μg⁻¹ short RNA (<200 nt). In the absence of DXO, the mammalian decapping enzyme of FAD-RNA, FAD-RNA levels were doubled.

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. This electrophoretic method retards in vitro transcribed FAD-RNA relative to PPP-RNA. 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. Northern blotting of the APB-gel with a probe specific toward the RNA of interest allows for the validation and quantification of the FAD-RNA.

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. SpRai1 does not harbor exonuclease activity and only removes FAD, which maintains 5'-P-RNA integrity.

3.2.3. Coenzyme 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. 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.

Proof for the existence of CoA-RNA in vivo has been provided by LC-MS measurements using digested total RNA. Thereby, 3'-dephospho-CoA and its thioester derivatives (e.g., acetyl-thioester) were identified to decorate the 5'-end of bacterial RNA in S. venezuelae (13 fmol μg⁻¹ total RNA) and E. coli (8 fmol μg⁻¹ total RNA). Using size fractionation of total RNA, the approximate maximum size of CoA-RNAs was estimated to 200 nucleotides per RNA. 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.

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 removes intact 3'-dephospho-CoA from CoA-RNA. Then, exonucleolytic activity leads to 5'-P-RNA decay.
Writers, Erasers, and Functions of CoA-RNAs: Apart from T7 RNAP initiating transcription with 3′-dephospho-CoA in vitro,[107] E. coli RNAP can also start transcription of adenosine-initiating transcripts with 3′-dephospho-CoA in vitro.[106a]

In the eukaryotic world, the mitochondrial RNAP accepts 3′-dephospho-CoA as substrate in vitro as well.[122]

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.[106a]

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[106a].

The same is observed for the Nudix hydrolase NudL from E. coli which has recently been reported to hydrolyze CoA while being inactive on CoA-RNA.[110]

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).[126] DeCoAping refers to the 5′-terminal processing of CoA-RNA that releases 3′-dephospho-CoA and 5′-P-RNA (Figure 8C).[126] KIDxo1 combines deCoAping and exonuclease activity which leads to immediate RNA degradation after decapping of CoA-RNA in vitro.[130] 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.

3.2.4. 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.[106b] Thus, UDP-GlcNAc/UDP-Glc-RNAs are capped co-translationally—allogously to RNAs capped with adenosine-derived cofactors.[106b] The concentrations of UDP-Glc (2.5 × 10^{-3} m) and especially UDP-GlcNAc (9.2 × 10^{-3} m) amount to the same order of magnitude as concentrations of UTP (8.3 × 10^{-3} m).[131] 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.[132] 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 (2.3 fmol μg^{-1} RNA) exceeds the concentration of UDP-Glc-RNA (0.22 fmol μg^{-1} RNA) by tenfold. Thus, UDP-GlcNAc-RNAs and NAD-RNAs (2.2 fmol μg^{-1} RNA) are present in similar concentrations.[118] Interestingly, these noncanonical caps also exist in the eukaryotic world (humans, mice, and yeast).[118]

More complex 5′-UDP-derived caps, such as UDP-GlcNActri pentapeptide, do not serve as NCINs for E. coli RNAP in vitro.[106b] 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.[106a] Apart from these speculations, the sequences and functions of UDP-GlcNAc- and UDP-Glc-RNAs remain elusive.

3.2.5. 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).[113] 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.[133] Interestingly, ThTP synthesis is elevated under certain stress conditions such as hypoxia in E. coli.[113] Further, thiamine exists in conjugation with adenine—adenosine thiamine triphosphate (AThTP) and adenosine thiamine diphosphate (AThDP)—in E. coli and eukaryotes.[114] Interestingly, upon carbon starvation of E. coli, AThTP can accumulate (79–120 pmol mg^{-1} protein).[115]

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.[136] 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.[136] 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.[136] The functionality of these potential thiamine-RNAs is as obscure as the sequence of the transcripts subjected to capping.

3.2.6. NpₙNs (Dinucleoside Polyphosphates)

Dinucleoside polyphosphates (NpₙNs) 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.[137] 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.[138] 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,As (Ap,As, Cp,As, Gp,As, Up,As)—were identified as noncanonical 5′-caps of transcripts in E. coli.[139] 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.[139] 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.[139] Moreover, calculations based on 14 randomly investigated transcripts with median Np,n-capping levels over 40% suggest toward a higher abundance of Np,n-capped RNA than NAD-RNA.[139] Interestingly, disulfide stress induced by cadmium or diamide dramatically increases the Np,n-capping levels.[139] It has to be noted that boronate gel electrophoresis is directed toward 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.[140] Consequently, RNA is degraded to single nucleotide level, whereas Np,n-caps remain intact and can be detected by LC-MS/MS analysis.[140] Thereby, Ap,3A, Ap,5A, and Ap,G as well as methylated forms of Np,nNs were detected as 5′-caps of RNA in E. coli.[140] 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,n. 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,3A, Ap,5A, Gp,G). Independent studies have shown that these alarmones are more efficiently used as NCINs by these RNAPs in vitro than NAD and CoA.[140,141] 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.[141] Increasing linker length coincides with lower incorporation efficiency.[141a] Moreover, the promoter influences the efficiency of transcription initiation with Np,nNs. A purine at −1 position increases transcription initiation with Np,A 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.[141] In vitro, dinucleoside polyphosphates such as Ap,A are synthesized by tRNA synthetases such as E. coli lysyl-tRNA synthetase LysU.[142] Interestingly, this enzyme also generates Ap,4-capped RNA in vitro using ATP and 5′-P-RNA as substrates (Figure 9A).[139] 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,n-caps. A) Disulfide stress increases cellular Np,N concentrations in E. coli. The RNA polymerase either incorporates Ap,N ab initio during transcription or the lysyl-tRNA synthetase post-transcriptionally creates this cap from ATP and 5′-P-RNA. B) Ap,A-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.](image-url)
As elaborated above, 5′-cofactor-capped RNAs such as NAD-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.[144] 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 3). As typical for ADP-ribosylation, nicotinamide serves as a leaving group. Thereby, a covalent bond between oxygen of the phosphate group and the Cl 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.

4. 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 harbors 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. While the RNA modifications m3C, m3A, I, and Ψ are merely studied in tRNA and rRNA, 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.[145] 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.[144] Even though the biosynthesis pathways of queuosine-modified RNA in bacteria are described, the existence of this modification in mRNA remains unexplored.[145]

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 noncanonical 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 noncanonical 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, NpN, 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 toward noncanonical 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 m3A in total RNA isolates.[74,146] 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.[145] Which roles might RNA modifications play?
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.[20] So far, it has not been combined with studies of the epitranscriptome, which may fulfill functions in the context of pathogen–host interactions. Further, it appears that 5′-terminal RNA modifications could correlate with the cellular concentrations of the respective cofactors or nucleoside analogs. 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 with the distinct physiological conditions. Also, these RNA modifications could be linked to various metabolic processes and stress conditions in bacteria.[132,141b]

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.[176]

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

M.S. and N.P. contributed equally to this work. K.H. acknowledges funding from the Eliteprogramm of the Baden-Württemberg Stiftung and the Max Planck Society.

Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

capping, cofactor, epitranscriptome, next-generation sequencing, prokaryotes, RNA modifications

Received: April 9, 2021
Revised: May 7, 2021
Published online:
a) I. A. Roundtree, M. E. Evans, T. Pan, C. He, D. Incarnato, F. Anselmi, E. Morandi, F. Neri, M. Maldotti, J. Somme, B. Van Laer, M. Roovers, J. Steyaert, W. Versees, T. P. Hoernes, N. Clementi, K. Faserl, H. Glasner, K. Breuker, H. Hori, T. Suzuki, K. Sugawara, Y. Inoue, T. Shibata, Y. Zhu, S. P. Pirnie, G. G. Carmichael, V. Khoddami, B. R. Cairns, E. Flatau, F. A. Gonzales, L. A. Michalowsky, P. A. Jones, R. Reid, P. J. Greene, D. V. Santi, S. Gehrig, M.-E. Eberle, F. Botschen, K. Rimbach, F. Eberle, J. Cavaille, M. Nicoloso, J. P. Bachellerie, N. Krogh, O. Birkedal, H. Nielsen, in *Adv. Biology*, 2013, 31, 458.

b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090; b) P. P. Dennis, V. Tripp, L. Lui, T. Lowe, L. Randau, *BMC Genomics*, 2015, 16, 2090.
Maik Schauerte obtained his M.Sc. Degree in Biochemistry from the Ruprecht-Karls University of Heidelberg in 2021. He continues his studies of the epitranscriptome as a Ph.D. candidate in Bacterial Epitranscriptomics at the Max-Planck-Institute for Terrestrial Microbiology in Marburg, Germany. His research aims to identify the biological functions of NAD-RNAs in prokaryotes.

Nadiia Pozhydaieva received her M.Sc. Degree in Biochemistry from the Heinrich-Heine University of Düsseldorf in 2020. Now, she is a Ph.D. candidate in Bacterial Epitranscriptomics at the Max-Planck-Institute for Terrestrial Microbiology in Marburg, Germany. Her research focuses on epitranscriptional regulatory mechanisms in *E. coli*.

Katharina Höfer received her Ph.D. degree at the Institute of Pharmacy and Molecular Biotechnology (Heidelberg University) in 2016 for her pioneering discoveries of NAD-RNAs in prokaryotes and their decapping enzymes. Afterward, she continued her work on NAD-RNAs as a postdoctoral researcher at Heidelberg University. In 2020, Katharina Höfer started her group focusing on Bacterial Epitranscriptomics at Max-Planck-Institute for Terrestrial Microbiology in Marburg. Her group is interested in the biological roles of NAD-RNAs in prokaryotes by deciphering the writers, readers, and erasers of these RNAs.