Biochemistry

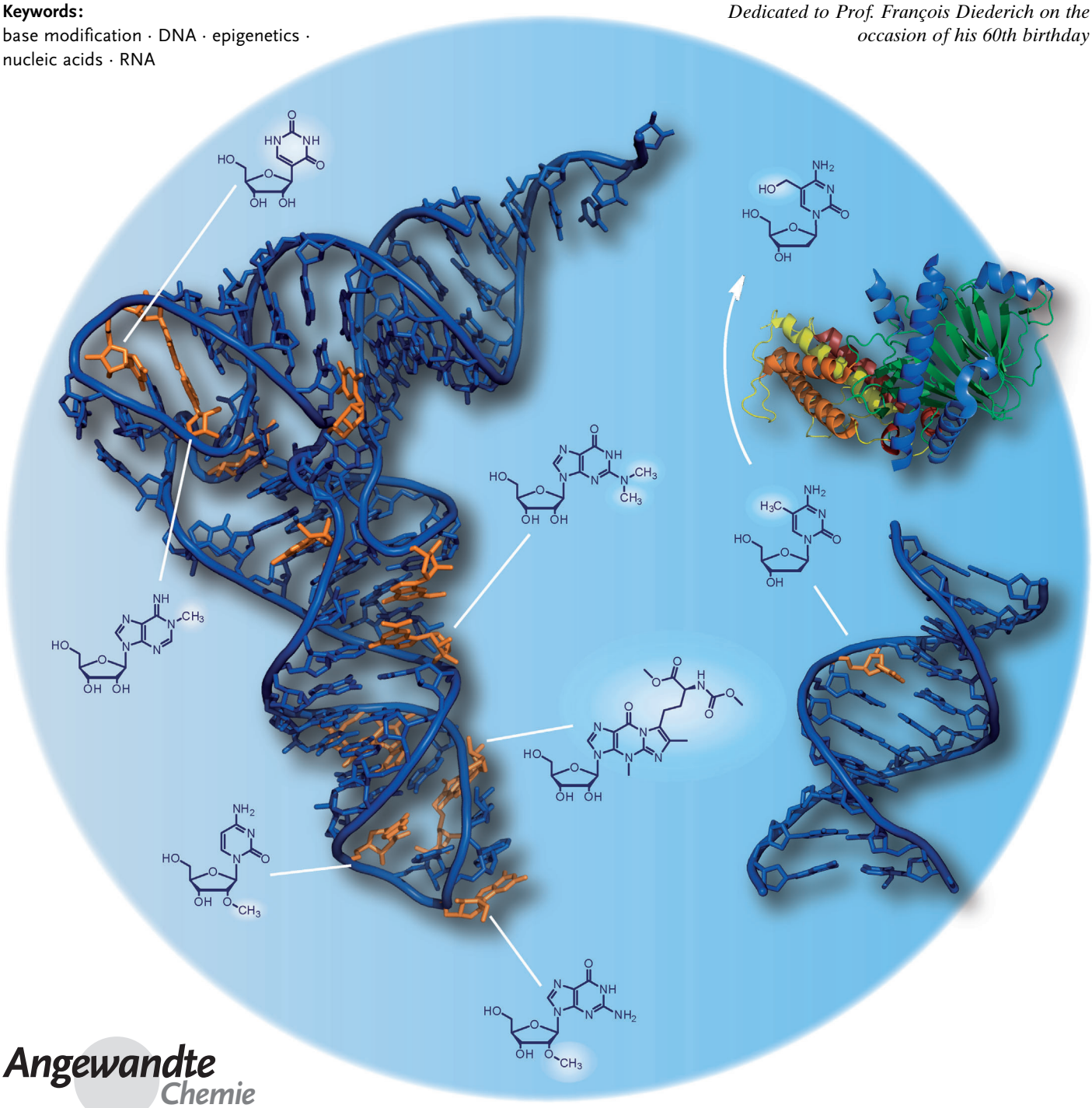
DOI: 10.1002/anie.201201193

Structure and Function of Noncanonical Nucleobases

Thomas Carell, Caterina Brandmayr, Antje Hienzsch, Markus Müller, David Pearson, Veronika Reiter, Ines Thoma, Peter Thumbs, and Mirko Wagner*

Keywords:
base modification · DNA · epigenetics · nucleic acids · RNA

Dedicated to Prof. François Diederich on the occasion of his 60th birthday



Angewandte International Edition Chemie

DNA and RNA contain, next to the four canonical nucleobases, a number of modified nucleosides that extend their chemical information content. RNA is particularly rich in modifications, which is obviously an adaptation to their highly complex and variable functions. In fact, the modified nucleosides and their chemical structures establish a second layer of information which is of central importance to the function of the RNA molecules. Also the chemical diversity of DNA is greater than originally thought. Next to the four canonical bases, the DNA of higher organisms contains a total of four epigenetic bases: m^5dC, hm^5dC, f^5dC und ca^5dC. While all cells of an organism contain the same genetic material, their vastly different function and properties inside complex higher organisms require the controlled silencing and activation of cell-type specific genes. The regulation of the underlying silencing and activation process requires an additional layer of epigenetic information, which is clearly linked to increased chemical diversity. This diversity is provided by the modified noncanonical nucleosides in both DNA and RNA.

From the Contents

1. Introduction

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the basic molecules of life. These molecules encode all genetic information and are the blueprint of every living organism. All organisms use basically the same genetic language, which is constructed from the four canonical nucleosides deoxyadenosine (dA), deoxycytosine (dC), deoxyguanosine (dG), and deoxythymidine (dT) in DNA and the corresponding canonical ribonucleosides A, C, G, and U in RNA (Figure 1). These eight organic molecules are the central building blocks of the genetic system. The DNA molecule itself was originally discovered by Miescher in 1869,[1] and Chargaff later formulated the Chargaff rules,[2] which describe amongst others the relative abundance of the four canonical DNA bases. These bases are connected to deoxyribose to give the four nucleosides, which are linked together through phosphodiester bridges.[3] Groundbreaking work by the research groups of Avery in 1944 and Hershey in 1952 showed finally that DNA acts as a carrier of heritable genetic information.[4] Shortly thereafter, the double-helical structure of the DNA duplex (Figure 1), a true milestone discovery of the last century, was deciphered by Watson and Crick[5] with major scientific contributions from Franklin, Goslin, and Wilkins.[6] The central building blocks of DNA and RNA are the nucleobases, which are attached to the (deoxy)ribose sugar through the N1 nitrogen atom of the pyrimidines (T, C, and U) and through the N9 nitrogen atom of the purine bases (A and G) as shown in Figure 1. We understand today that the basis of the genetic system are the specific hydrogen bonds that can form between the pyrimidine bases (d)C/dT, U, and the purine bases (d)A/(d)G. These specific base-pairing properties establish selective noncovalent interactions that are the basis for the ability of DNA to be replicated and transcribed into RNA. While DNA stores heritable genetic information inside the cell (or inside the cell nucleus), the transcribed RNA strands, which are copies of DNA segments, have been determined to perform a variety of functions. Besides transporting genetic information and regulation of gene activity,[7] RNA molecules can even have catalytic functions, for example in the ribosome, where RNA catalyzes the formation of peptide bonds.[8] In addition to the coding messenger RNA (mRNA) and the well-known ribosomal RNA (rRNA) and transfer

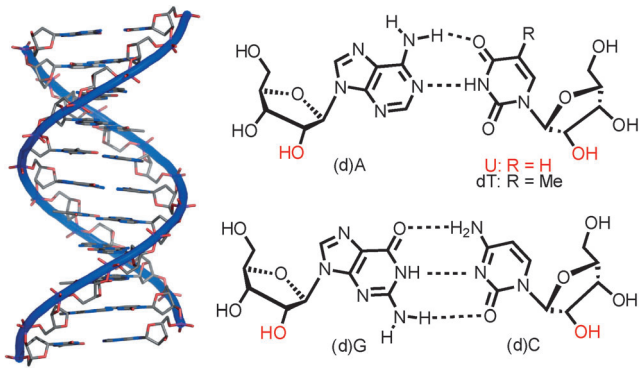


Figure 1. Depiction of a Watson–Crick B-DNA double helix and of the Watson–Crick base pairs. Their sequence (dA, dC, dG, and dT) forms the sequence code, which establishes the first level of genetic information that is encoded. Black: universal components. Red: groups that are present specifically in RNA.

[*] Prof. Dr. T. Carell, MChem. C. Brandmayr, Dipl.-Chem. A. Hienzsch, Dr. M. Müller, Dr. D. Pearson, Dipl.-Chem. V. Reiter, M. Sc. I. Thoma, M. Sc. P. Thumbs, Dipl.-Chem. M. Wagner
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 München (Germany)
Prof. Dr. T. Carell, MChem. C. Brandmayr, Dipl.-Chem. A. Hienzsch, Dr. M. Müller, Dr. D. Pearson, Dipl.-Chem. V. Reiter, M. Sc. I. Thoma, M. Sc. P. Thumbs, Dipl.-Chem. M. Wagner
Department für Chemie, Ludwig-Maximilians Universität München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de

RNA (tRNA) there are various newly discovered forms of noncoding RNA. These species have, over the past two decades, completely changed our view of RNA, making RNA research currently one of the most dynamic and fastest growing fields in the life sciences. The newer species, including small nuclear and nucleolar RNA (snRNA and snoRNA), micro-RNA (miRNA), and long noncoding RNA (lncRNA),[9] are now known to possess important cellular functions that go beyond information transport.[10] Particularly interesting in this respect is the discovery that some lncRNAs and miRNAs are involved in orchestrating the fine details of cellular differentiation and development (Table 1).[11] We now know that only a small amount of the human genome encodes proteins and that a much larger part is cell-specifically transcribed to generate these noncoding RNAs.[12]

Table 1: Abbreviations and explanation of the different RNA species mentioned in this Review.

Abbreviation	Full name	Function
tRNA	transfer RNA	mRNA decoding amino acid carrier
mRNA	messenger RNA	protein coding message
rRNA	ribosomal RNA	nucleic acid component of the ribosome
sncRNA	small noncoding RNA	
miRNA	micro RNA	regulating gene expression
piRNA	piwi-interacting RNA	regulating gene expression
siRNA	short interfering RNA	regulating gene expression
snRNA	small nuclear RNA	involved e.g. in intron splicing
snoRNA	small nucleolar RNA	involved in RNA modification
lncRNA	long noncoding RNA	nonprotein coding transcripts with regulatory or unknown function

We are just beginning to understand that the many different RNA molecules inside the cell create a complex, densely connected network that establishes dedicated regulatory functions. For anyone interested in the field, it is important to note that many RNA molecules contain not only the four canonical bases but a large number of modified structures. Clearly, the chemical diversity established by just four canonical nucleobases is not sufficient for the complex functions that RNA performs beyond transport and translation of DNA sequence information. In addition to the information stored in the base sequence of RNA molecules, RNA species employ structurally distinct posttranscriptional modifications at numerous sites. The position of the modified nucleosides and their chemical structure establishes a second information layer which is of central importance for the function. Many of the modified bases were discovered a long time ago in tRNA, rRNA, and mRNA, but researchers are just starting to look for modified bases in the many noncoding RNA species. The chemical diversity of DNA is less complex and mainly limited to methylated and hydroxymethylated bases such as N4- and C5-methylated cytosine, N6-methylated adenine, and C5-hydroxymethylated cytosine and uracil (nucleobase J).[13] Most recently, however, a few new DNA bases were discovered, which shows that the chemical complexity offered by the four canonical bases is also not sufficient for the encoding of heritable genetic information. We know today that in mammals the modified DNA bases encode epigenetic information needed to orchestrate cellular development processes. The modified bases in RNA, in contrast, have many different functions, which will be discussed in more detail below.

In the first part of this Review we give a general introduction to the field of RNA modifications and describe a number of recent developments and areas we consider to be of view of particular interest. As a consequence of the large size of the field, we are not attempting to provide a comprehensive review. As such, we would like to draw attention to a number of excellent reviews on different aspects of RNA modification, including wobble-base pairing,[14] extended anticodon modification,[15] RNA stability conferred by modificatios,[16] and other aspects.[17] In the second, shorter, part of the manuscript we will discuss the modified DNA bases with a focus on their discovery process. We will show how these bases have started to change our view of how the heritable genetic information is epigenetically programmed.

Thomas Carell studied chemistry at the universities of Münster and Heidelberg, Germany, and received his PhD in 1993 at the Max Planck Institute for Medical Research. After a postdoctoral stay with Julius Rebek at the MIT in Cambridge, USA, he started independent research at the ETH Zurich. In 2000 he accepted a chair for Organic Chemistry at the University of Marburg. In 2004 he moved to his current position at the Ludwig Maximilians University in Munich. He is a member of the German Academy Leopoldina and recipient of the Leibniz award and The Order of Merit of the Federal Republic of Germany.

2. RNA Modifications

To date, more than 100 RNA modifications have been discovered. Chemical alteration of the nucleoside structures is not limited to just particular Watson–Crick bases. In fact, structural variations are known at a variety of positions in all four canonical nucleosides, as depicted in Figure 2. Besides simple modifications, such as methylation or acetylation (N^1-methylguanosine, m^1G; N^2,N^2-dimethylguanosine, m^2_2G; or N^4-acetylcytosine, ac^4C), so-called hypermodified bases are known, some of which even contain a different heterocyclic structure, usually 7-deazapurine (Figure 2). Although all RNA species in the cell are thought to contain modified RNA nucleosides, the largest fraction and diversity is present in tRNA. About one third of the modified bases are present in ribosomal RNA, and roughly 10% of the known modified bases have been found in messenger RNA. So far, 14

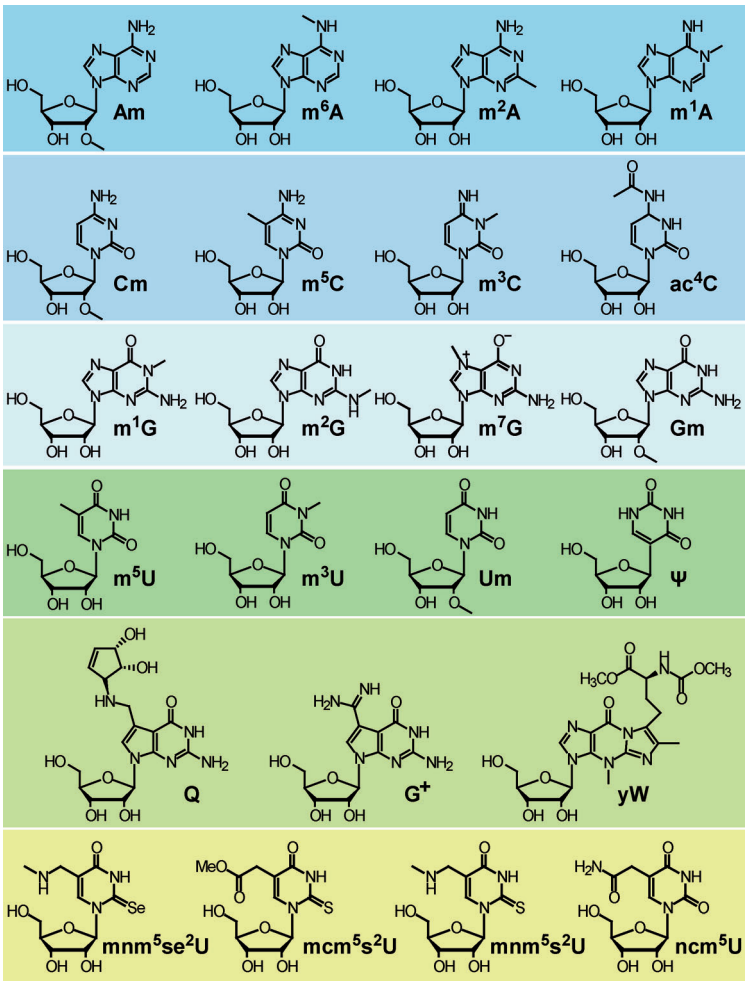
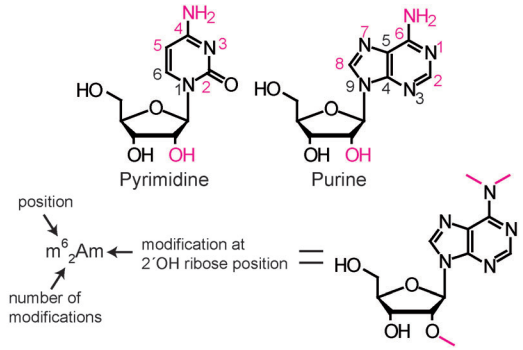


Figure 2. Selected RNA modifications. From the top to the bottom: A-derived modifications (dark blue), C-derived modifications (middle blue), G-derived modifications (light blue), U-derived modifications (dark green), hypermodified purine bases (light green), and hypermodified pyrimidine (uridine) bases (yellow). For the complete set of modifications detected so far see: http://s59.cas.albany.edu/RNAmods/and http://modomics.genesilico.pl/.

modifications have been discovered in other RNA species, mainly in small, noncoding RNA.[18] While some modifications, such as pseudouridine Ψ, are present in almost all types of RNA,[19] others are known to exist only in a single RNA species. Focusing on tRNA, sequencing results show that about 10–15% of the canonical tRNA nucleosides are modified. On average, each tRNA molecule, therefore, contains eight modified nucleosides, many of which are present in, or directly adjacent to, the anticodon loop.[20] The other modified nucleosides are located at various specific positions inside the structural core of the tRNA. In the last few years, the chemical structures of these modifications, their position, and also many of the genes and enzymes that are involved in the biosynthesis of these modifications have been deciphered. The biochemical and biophysical function of some of the modifications in tRNA molecules have been investigated in greater detail.

A special nomenclature has been developed to describe the multitude of nucleoside modifications. As depicted in Figure 3, the chemical modifications are abbreviated with a one- or two-letter code. The position of the chemical modification on the pyrimidine or purine ring is given as a superscript number after the one- or two-letter code. The number of modified groups that are present on the base at that specific position is indicated by a subscript number. The numbering proceeds according to the nomenclature for pyrimidines and purines. Special hypermodified bases such as queuosine (Q) or wybutosine (yW) are described with an individual letter, similar to the canonical nucleobases.[18,21] Modifications on the sugar moiety are listed after calling the base.

A special modification is pseudouridine (Ψ). This base, depicted in Figure 2, is connected to the sugar backbone not through the pyrimidine N1 atom but through C5, as a result of an isomerization.[19] Hypermodified RNA bases are structures that are more dramatically changed. Base modifications belonging to this class are for example N^6-threonylcarbamoyladenosine (t^6A) and lysidine (k^2C), in which amino acid derived side chains are connected to the purine and pyrimidine skeleton (nomenclature: see Figure 3). Further exam-



Abbr.	full description	Abbr.	full description
m	methyl	se	seleno
ac	acetyl	cmo	5-oxyacetic acid
t	threonylcarbamoyl	mcmo	5-oxyacetic acid methyl ester
тm	taurinomethyl	mo	methoxy
i	isopentenyl	ho	hydroxy
io	*cis*-hydroxyisopentenyl	f	formyl
s	thio	gal	galactosyl
ms	methylthio	glu	glutamyl
inm	isopentenylaminomethyl	man	mannosyl
g	glycinylcarbamoyl	r	2′-*O*-ribosyl
hn	hydroxynorvalylcarbamoyl	I	inosine
hm	hydroxymethyl	Ψ	pseudouridine
nm	aminomethyl	Q	queuosine
acp	(3-amino-3-carboxypropyl)	yW	wybutosine
cm	carboxymethyl	OHyW	hydroxywybutosine
chm	carboxyhydroxymethyl	C^+	agmatidine
cmnm	carboxymethylaminomethyl	k^2C	lysidine
ncm	carbamoylmethyl	D	dihydrouridine
mnm	methylaminomethyl	G^+	archaeosine

Figure 3. Nomenclature for the description of the most common modified nucleosides. Possible modification positions on the purines and pyrimidines are marked in pink (top). The Table (bottom), the first part shows the abbreviations that are used to describe functional groups present on the modified nucleosides. The second part of the Table lists common abbreviations given to certain modified nucleosides.

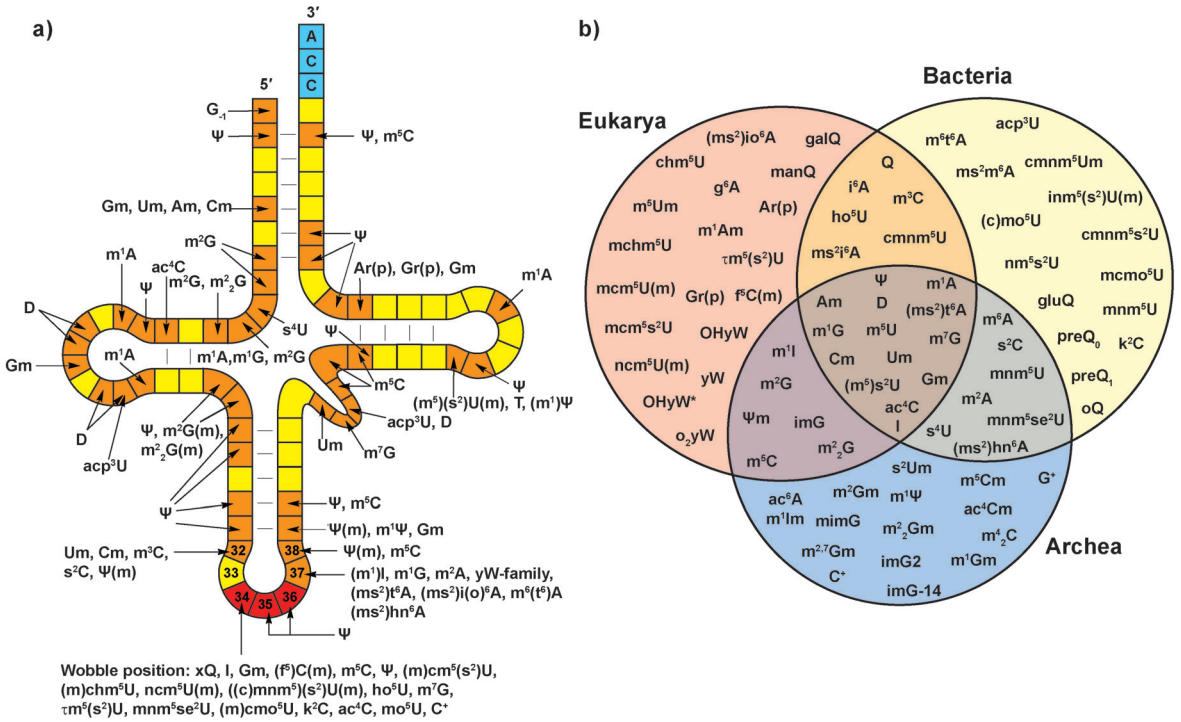


Figure 4. a) 2D representation of a tRNA showing (noncomprehensive) the positions at which most modifications in bacteria, archea, and eukaryotes are found. The anticodon loop is highlighted in red, known modified positions in orange, and the universal CCA-3′ in blue. Position 34 is the wobble position, and position 37 adjacent to the anticodon is the semi-invariable site always occupied by a purine base. These two positions often contain hypermodified bases. Data adapted from Ref. [17c]. b) Modifications sorted by their occurrence in the three domains of life. Figure adapted from Ref. [22].

ples of more complex hypermodified bases are queuosine (Q) and wybutosine (yW), in which a multistep biosynthetic machinery is needed to change the entire heterocyclic skeleton. Figure 4 shows the variety and distribution of many of the known modified nucleosides in a schematic 2D tRNA diagram. The complex structure and diversity of the existing RNA modifications has led to the development of dedicated databases, which are constantly extended. Two of the most important databases are the "RNA modification database" (http://s59.cas.albany.edu/RNAmods/) and "Modomics: a database of RNA Modifications" (http://modomics.genesilico.pl/). These databases contain information about the chemical structure of the various RNA modifications, their trivial names, and the symbols used for their abbreviation. In addition, literature has been deposited about the chemical synthesis and biological function where known. Another important database, tRNAdb (http://trnadb.bioinf.uni-leipzig.de/) places modified nucleosides in context with the tRNA sequences.

2.1. Recently Discovered RNA Modifications

The development of sophisticated liquid chromatography and mass spectrometry (LC-MS) techniques with sensitive mass spectrometers has recently allowed the discovery of new RNA nucleosides (Figure 5). In the last 10 years, around 14 new modifications have been deposited in the databases, including agmatidine (agm^2C or C^+), 8-methyladenosine (m^8A), glutamylqueuosine (gluQ), and 5-taurinomethyluridine (τm^5U).[18] The agmatidine modification was discovered in 2010 by the research groups of Suzuki and RajBhandary. This modification is a cytidine derivative that was found in the wobble position of the AUA-decoding $tRNA_2^{Ile}$ of archaea.[23] Agmatidine is related to the bacterial modification lysidine (k^2C), but differs with respect to the amino acid derived side chain that is linked to cytidine at position C2. Both modifications fulfill the same function. They reprogram the coding potential of C, because the modified C bases k^2C and C^+ base pair preferentially with A instead of G.[23b,24]

Three years ago, researchers discovered 8-methyladenosine (m^8A) as a new base in bacterial 23S rRNA. The base is found at position A2503, where m^2A was also detected a few years earlier.[25] The methylation of A2503 to give m^8A is now known to be performed by the methyltransferase Cfr. This methylation establishes a strong resistance of the corresponding ribosomes to five classes of antibiotics that target peptidyl transferase.[26] These inhibitors can no longer bind to the modified ribosomes. A chemical synthesis of m^8A was already developed in 1993, and it was discovered that the individual nucleoside compound is a selective inhibitor of the vaccinia virus.[27] This result is interesting because it shows that some of the modified bases can have interesting biological properties as individual entities. This is also true for the well-known phytohormone i^6A, which exhibits antiproliferative properties when administered to tumor cells.[28] Two further newly discovered modifications, 5-taurinomethyl-uridine (τm^5U) and 5-taurinomethyl-2-thio-uridine (τm^5s^2U), were detected

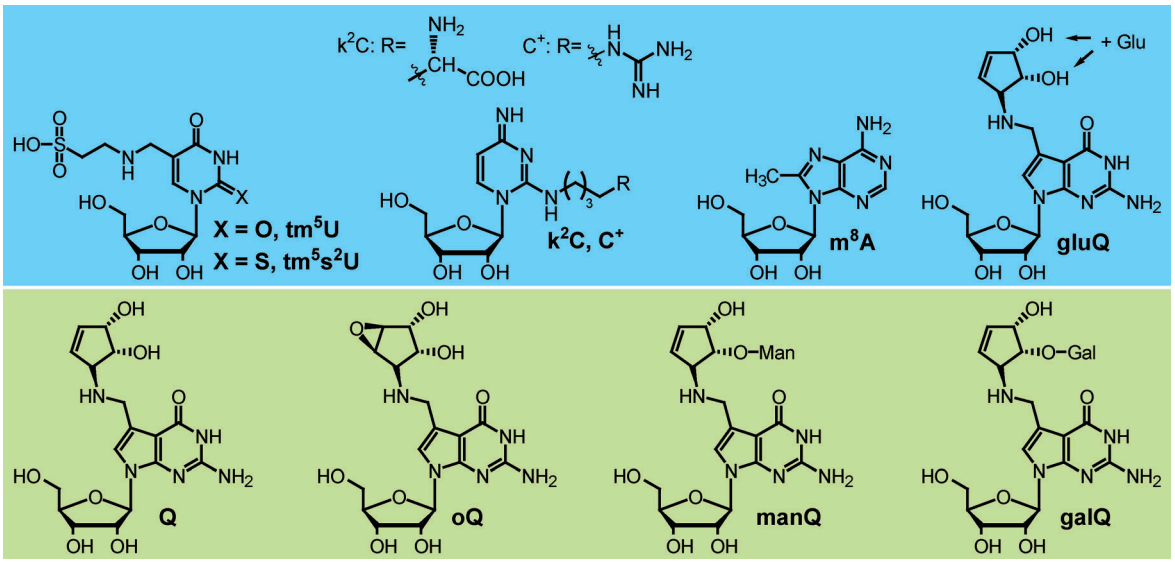


Figure 5. Newly discovered modified RNA bases: k^2C (blue) and derivatives of the hypermodified RNA base Q (green).

in mammalian mitochondrial tRNA. These modifications were identified with high-resolution mass spectrometry in combination with MS^n fragmentation analysis. The structures were further proven by total synthesis and NMR analysis.[29] Lack of these modifications was found to be associated with the development of mitochondrial disease (see Section 2.5.1). The analysis of low-abundance cellular RNA species, enabled by highly sensitive LC-MS techniques, recently revealed that RNA is also modified by the attachment of NAD, CoA, and possibly other molecules. The attachment site of the NAD modification has not yet been identified, despite the relatively high abundance of this modification in *Escherichia coli* and *Streptomyces venezuelae* (about 3000 copies per cell).[30]

Some of the hypermodified RNA bases such as Q or yW exhibit their own chemical diversity. For queuosine, for example, it is known that the base exists in further glycosylated forms with galactose (galQ) and mannose (manQ).[31] The function of the additional sugar units is not clear. Aside from sugar-modified forms of Q, a Q-epoxide (oQ) is also known, which is a biosynthetic intermediate (Figure 5).[32] Most recently, the new Q derivative glutamylqueuosine gluQ was discovered, which contains an amino acid instead of the usual sugar modifications attached to the aminocyclopentendiol substructure.[33] This modification could only be detected in *E. coli* tRNA under special conditions that avoid the facile hydrolysis of the glutamyl–queuosine bond. This result is very important, as it shows that many other sensitive modifications that may be present in cells have not yet been detected by using current methods. At present, neither the exact position nor the function of the glutamyl group is known.[33a] In addition, chemical syntheses of the sugar and amino acid modified Q bases have not yet been reported. In conclusion, it is foreseeable that the discovery process will continue in line with the development of new, more-sensitive mass spectrometers and milder isolation protocols.

2.1.1. *The Q Base Family*

The newly discovered gluQ derivative establishes the Q base as one of the most versatile base modifications known.[33] The basic structural feature of the Q base is the central deazaguanosine heterocycle bearing a carbon atom instead of a nitrogen in position 7 and an Cyclopentenediol moiety attached to the C7-position connected through a methylamine group. The further modified Q species galQ and manQ, in which an aldohexose is attached to the non-allylic alcohol of the cyclopentene, were determined to exist as the β-configured isomers.[31] The queuosine base Q itself was discovered over 40 years ago in $tRNA^{Asn}$, $tRNA^{Asp}$, $tRNA^{His}$, and $tRNA^{Tyr}$. Q derivatives are present in bacteria and eukarya, with the exception of budding yeast and mycoplasma.[34] Despite the structural similarity of the Q derivatives, it is today clear that the further modified Q derivatives are present only in selected domains of life. gluQ, for example, is exclusively found in bacteria,[33b] while galQ and manQ are found only in mammals.[31,35] All Q derivatives are located at position 34 of the tRNA, the wobble position in the anticodon loop.[36] Q derivatives have so far not been observed in archaea. Instead, the structurally similar 7-deazaguanosine archaeosine G^+ is found (Figure 2). G^+ occurs in tRNA at position 15 in the dihydrouridine loop, where it is thought to be important for the stabilization of the tRNA structure.[37]

The partly complex biosynthesis needed to install modified nucleosides in RNA is best exemplified with the Q base. Prokaryotes and eukaryotes introduce this modification by quite different procedures. Eukaryotes use queuine for the biosynthesis, which is simply the Q-base heterocycle.[38] This compound is not de novo synthesized, but instead is taken up through nutrition.[39] The Q-utilizing eukaryotes possess a special tRNA-guanine transglycosylase (TGT), which cleaves the glycosidic bond of G34 to create an abasic site and subsequently attach the queuine heterocyle to the sugar. The biosynthesis, therefore, involves a heterocycle exchange

process that occurs directly at the tRNA.[38,40] Bacteria in contrast are able to synthesize the Q base de novo.[41] To this end they possess a complex biosynthetic machinery that includes several enzymes.[34,42] One of the most interesting steps in the biosynthesis is the formation of the reactive epoxide. Fascinatingly, the cyclopentene ring system is biosynthetically generated from the ribose moiety of the SAM cofactor.[43] The ultimate step of the biosynthesis of Q is performed by a vitamin B12 dependent protein that catalyzes the conversion of oQ into Q.[44] The enzyme performing this reaction was discovered by Bandarian and co-workers,[45] who showed that it is an iron-sulfur reductase, whose structure has not yet been clarified. In addition, the role of vitamin B12 is still elusive.[45]

Despite intensive research, the function of Q and its derivatives is not yet fully known.[46] It is, however, accepted that they influence the recognition of the mRNA codon by the tRNA at the ribosome which, for example, plays a role in the read-through of stop codons by certain viruses.[47] The Q-containing tRNAs recognize codons with the sequence NA*C* and NA*U* (the sequences are always given in the 5′-3′ direction, unless stated differently), where N represents A, C, U, or G. tRNAs which contain G instead of Q favor the interaction with codons that end with a C (see Figure 6B for illustration).[48] However, when a Q base is situated instead of G at position 34, the stop codon UA*G* in the RNA of the tobacco mosaic virus (TMV) is not recognized by the tRNA, thus preventing the translation of viral proteins.[49] Further

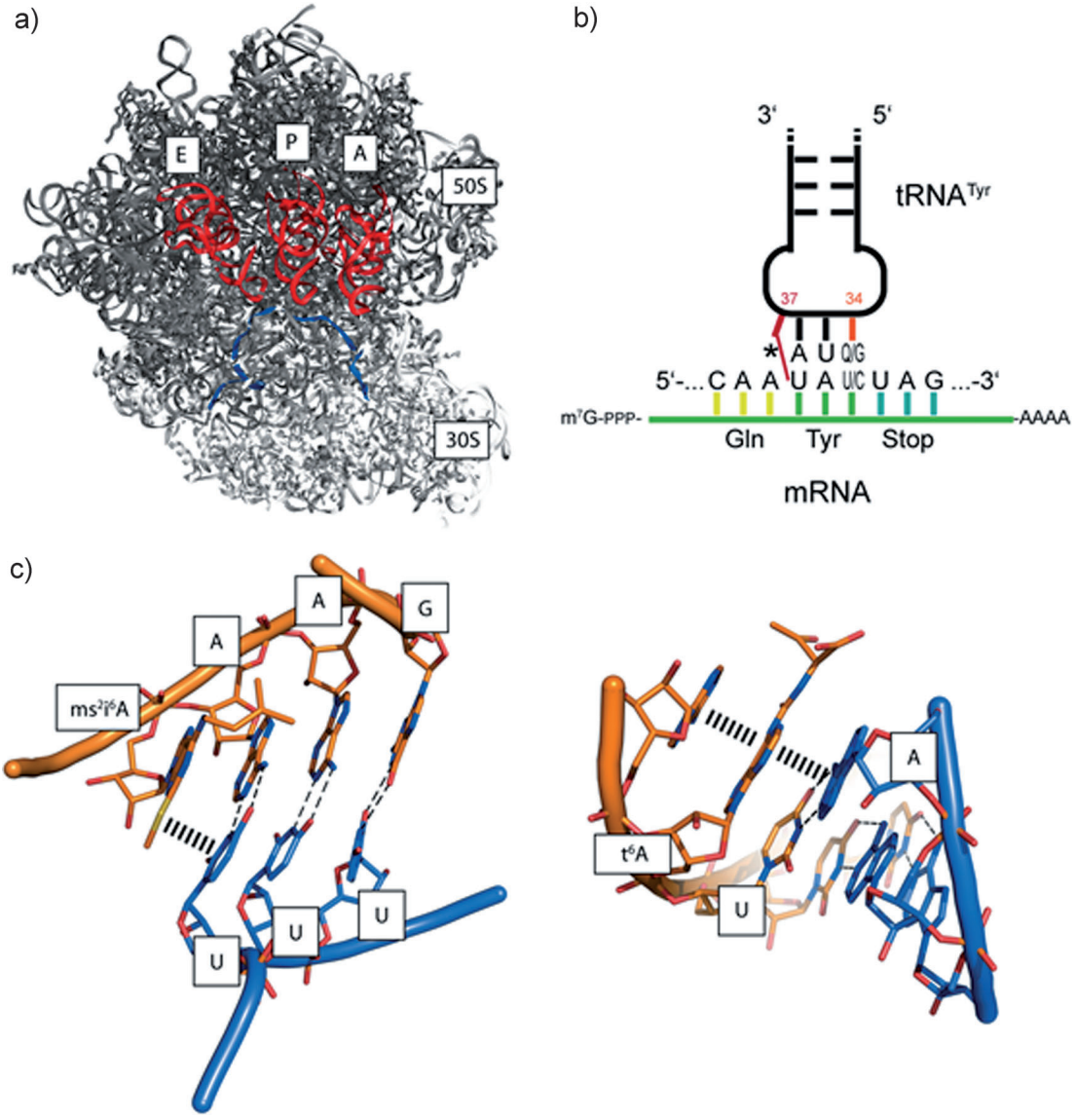


Figure 6. Crystal structures showing codon:anticodon binding at the ribosome. a) Structure of the 70S ribosome bound to mRNA and tRNA in the A, P, and E sites. tRNA is shown in red, and mRNA in blue; the dark gray 50S and light gray 30S subunits of the ribosome are labeled. Parts of the 30S subunit are removed for clarity. b) Illustration of the codon/anticodon interaction and the role of the modified bases at positions 34 and 37. * = ms^2i^6A and Q = queuosine c) Sections of the tRNA/mRNA interaction, showing the stabilizing interactions of bulky modifications at position 37 of tRNA adjacent to the anticodon.[56] tRNA is shown in orange, and mRNA in purple. Stacking interactions between the modified bases and tRNA/mRNA bases are indicated with wide dotted lines. Data were obtained from crystal structures 3I8G (left) and 1XMQ (right).

influences discovered in connection with Q are described in Section 2.5, which covers disease aspects.

2.2. Functions of RNA Modifications

To date, a large amount of information has been assembled in regard to the function of specific base modifications in particular positions. The modifications present in or adjacent to the tRNA anticodon have been most extensively studied. However, the functions of many other modifications are as yet unclear, especially outside of tRNA. A well-studied exception is the base m^7G in mRNA, known to establish the 5′-cap structure of eukaryotes that together with the poly-A-tail serves multiple purposes. m^7G helps, among other things, to define the mRNA reading frame, influences nuclear mRNA export, protects mRNA from degradation, and regulates splicing.[50] Essentially, the cap defines the identity of the RNA strand as a messenger. Interestingly, m^7G is attached through a rather labile triphosphate linkage to the mRNA strand.[51] Another well-described modification in mRNA is inosine, which is generated by deamination of adenosine. As a consequence of the altered base-pairing properties of inosine (pairing with C instead of U), this modification directly influences the genetic code and the process is, therefore, called RNA editing.[52] In tRNA, the large number of base modifications can be roughly divided into three groups in terms of functionality.[17c] The first group of modifications is present directly in or adjacent to the anticodon (e.g. the Q base in position 34). The modifications present here are mostly involved in the fine-tuning of the translational process and hence contribute to the translational fitness of the organism.[14a] A second class of tRNA modifications outside the anticodon loop controls the stability and the proper folding of the tRNA molecule. Some tRNA modifications, which we categorize into a third class, determine the identity of the corresponding tRNA. These modifications are particularly important for the interaction of the tRNA with the corresponding synthetase, which loads the specific amino acids onto the cognate tRNA species.[53] Such modified nucleosides are present at many positions in the tRNA, including the anticodon.

Many of the most heavily modified nucleobases belong to the first group. They are, for example, present in the wobble position 34 or in position 37 (Figure 6) directly to the 3′ side of the anticodon. Position 37 is a semi-invariable position, which is always occupied by a (usually modified) purine base. The most prominent modified bases present in position 37 are N^6-threonylcarbanyladenosine (t^6A) and N^6-isopentenyladenosine (i^6A), as well as the guanine modifications N^1-methylguanosine (m^1G) and the structurally highly complex wybutosine (yW), which is present only in phenylalanine-coding tRNA. These modifications directly influence the decoding process and are not generally required for the interaction of the tRNA with the corresponding amino acyl synthetase. Besides this specific function, they also reduce the conformational flexibility of the anticodon loop. The nucleosides enforce a more rigid conformation and, therefore, optimize the interaction with the codon of the messenger RNA.[15] This stabilization is proposed to enhance the rate and accuracy of tRNA binding to the mRNA codon and to the ribosomal A site. Furthermore, these modifications help to stabilize the positioning in the ribosomal P site, thereby preventing frameshifts.[54] Crystal-structure data[55] show that the "bulky" side chains present in bases that occupy position 37 (e.g. in t^6A, ms^2i^6A, and yW) help to stabilize the weak codon:anticodon interaction by providing an additional base-stacking interaction with base 38 of the tRNA and the first base of the mRNA codon (Figure 6).[56] Without these additional stabilizing interactions, U/A-rich codon:anticodon interactions (e.g. the AAA codon) could not form efficiently. Additionally, the large side chains keep the anticodon loop open and block base pairing between position 32 or 33 with position 38 in the anticodon loop (Figure 4). Such an interloop interaction reduces the size of the loop, which would disturb the anticodon recognition.[57]

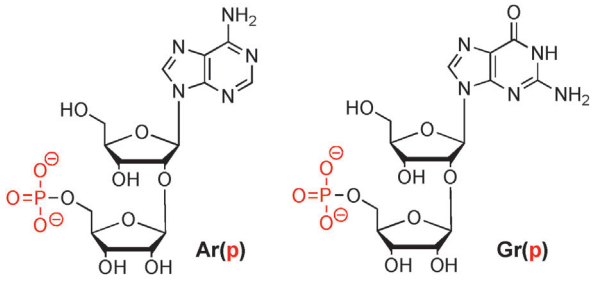
Modifications at the wobble position 34 include Q and a variety of 5′-modified uridines (see also Figure 2). These bases have a strong influence on the nature of wobble-base pairing, in addition to stabilizing the anticodon loop structure. Recent reviews have summarized the possible molecular mechanisms underlying wobble-base pair interactions,[14a] as well as the impact of modified wobble-base pairing on the decoding of the genetic information.[14b] The mechanisms of modified wobble-base pairing appear to be complex and specific to particular tRNAs, as demonstrated for example by recent crystal structures involving the mnm^5U34 and cmo^5U34 modifications.[56a,58] These structures indicate that the modifications stabilize and appropriately orient the modified base to enable optimal stacking and hydrogen-bonding interactions at the ribosome.

Another interesting example of how modified bases influence decoding and define the identity of the corresponding tRNA is provided by the lysidine modification (k^2C, Figure 5). This base is found in position 34 of tRNAs that code for the amino acid isoleucine through the CAU anticodon ($tRNA^{Ile}_{CAU}$).[24b] The CAU anticodon with an unmodified cytidine at position 34 typically recognizes the methionine codon AUG in mRNA. However, if the cytidine is posttranscriptionally modified to lysidine, then the isoleucine messenger RNA codon AUA is recognized. The modification lysidine has a dual function because it is also involved in tuning the tRNA synthetase interaction and thus belongs to both classes 1 and 3. While the isoleucine-carrying amino acyl synthetase (IleRS) recognizes $tRNA^{Ile}_{k^2CAU}$ and loads an isoleucine onto this tRNA, the correspondingly unmodified tRNA ($tRNA^{Ile}_{CAU}$) is recognized much more slowly. This unmodified tRNA, however, is readily aminoacylated by the MetRS, which loads a methionine on the tRNA.[24a]

Modifications of the second group are known to influence the correct folding of the tRNA molecule and increase the stability of the tRNA, as reviewed recently.[16,22] In support of this view, it was determined that the melting temperature of a completely unmodified or undermodified tRNA is lower than that of normally modified tRNA, thus illustrating the stabilizing effect.[59] It has been shown that tRNAs that lack at least two modifications, such as m^7G46 and m^5C or ac^4C and Um44, are temperature sensitive, and in addition these

undermodified tRNAs are quickly degraded.[20,60] It is known that the modifications pseudouridine[61] and 2′-O-methyladenosine[62] stabilize double-helix structures inside the tRNA molecule. The modified base m^1A was found to be important for the correct folding of the human mitochondrial $tRNA^{Lys}$.[63] This base seems to stabilize a particular folding intermediate.

The third class of modifications determines the identity of the corresponding tRNA, such as the previously mentioned case of lysidine, which is a specific recognition signal for the IleRS.[24a] Several other modifications, such as s^2U[64] and m^1G, are similarly required for the recognition of aminoacyl tRNA synthetase.[65] Another interesting example is that of a unique phosphoribosyl modification present at position 64 of the initiator $tRNA_i^{Met}$ in yeast, plants, and fungi. This tRNA initiates peptide synthesis with methionine at an AUG start codon, but is not involved in later elongation of the peptide at other methionine codons. The phosphoribosyl modification prevents $tRNA_i^{Met}$ from binding to elongation factors and thus restricts it to only initiation processes.[66] This modification is observed as either the 2′-O-ribosyladenosine Ar(p) or the 2′-O-ribosylguanosine Gr(p). The large modification



protrudes from the T arm of the tRNA molecule, significantly adding to the steric bulk in this region.[67] While in yeast the initiator tRNA ($tRNA_i^{Met}$) is normally unable to bind to the elongation factor EF-1α, a chemically modified version of the $tRNA_i^{Met}$ lacking the Ar(p) nucleoside is able to bind, thus showing that the modfication directly influences the interaction.[68] Consequently, the initiator tRNA lacking Ar(p) was found to add Met in response to internal AUG (Met) codons in vitro, thus performing an elongator role. A strain lacking both the genes for this modifying enzyme and the elongator $tRNA^{Met}$ was able to grow successfully by using the unmodified initiator tRNA for elongation.[69] These observations show clearly that the phosphoribosylated nucleoside is the determining factor that prevents initiator tRNA from acting during elongation in yeast, plants, and fungi. Other differentiation strategies are observed in bacteria and higher eukaryotes.

In summary, although it is often not exactly known how a nucleoside modification influences the biochemistry, we are starting to obtain general knowledge of their function, for example, how they influence the stability of the anticodon loop[57b] to optimize the accuracy of the decoding process,[70] how they change codon–anticodon preferences, and optimize the codon–anticodon interaction to reduce frameshifting events.[54] These observation led early on to the idea that tRNA decode the mRNA sequence by using an "extended codon" that uses information beyond the sequence of just the three bases in the anticodon.[14a,71]

2.3. Biosynthesis of RNA Modifications

We already discussed in Section 2.1.1 with the example of queuosine that the chemistry needed to create certain base modifications can be complex and that it requires dedicated biosynthetic enzymes, which use partly unknown chemical mechanisms. More information regarding the enzymatic machineries needed to install modified nucleosides in RNA can be found in a database of enzymes under http://modomics.genesilico.pl.

The biosynthetic enzymes often operate in complexes, in which multiple reactions are orchestrated to finally create an individual RNA modification.[17b] Recently identified biosynthetic complexes include the KEOPS complex that generates the universal t^6A37 base in tRNAs recognizing ANN codons,[72] as well as the mammalian ABH8 enzyme (an AlkB homologue) that performs both methylation and hydroxylation modifications and leads to the formation of $mcm^5U34/mchm^5U34$.[73] The ABH8 enzyme is particularly interesting as it is known to be involved in survival after DNA damage, thereby suggesting a role of tRNA modification in the regulation of the response to DNA damage. Moreover, the hydroxylating subunit of ABH8 is an iron(II)/2-oxoglutarate-dependent dioxygenase similar to the TET enzymes that hydroxylate 5-methylcytosine in DNA (see Section 3.2).

Although much is known about the enzymes and pathways involved in RNA modification, it is currently not entirely clear how the positions at which a modification is introduced are recognized. One set of modifying enzymes is certainly structure-specific. They recognize a particular RNA structure for starting the modification reactions. Other modifying protein complexes require small nucleolar RNA guide strands (snoRNAs), which associate with the complementary RNA sequence that is subsequently modified. These guiding RNA strands are part of small nuclear ribonucleoprotein particles (snoRNPs). The box C/D type snoRNAs, for example, localize complementary sequences in the ribosome. Following recognition, the protein fibrillarin of the snoRNP is activated to perform 2′-O-methylation using the cofactor SAM.[74] The snoRNAs of the box H/ACA type recruit the protein dyskerin to the modification site, which converts a uridine base into pseudouridine.[75] Recently, similar modificatied guide RNAs were discovered in the Cajal bodies of cells (scaRNAs), which seem to be involved in the modification of small nuclear RNAs.[76]

tRNAs, which we concentrate on in this part of the Review because of their large chemical diversity, are transcribed as primary transcripts (pri-tRNAs) inside the nucleus. They subsequently undergo 3′- and 5′-end processing and are often spliced before the required base modifications are introduced. In other cases, such as the lysidine tRNA modification, the modification reaction occurs while the tRNA is in a precursor form, presumably to avoid the consequences of having unmodified tRNA in the cell.[77] In all

cases, the cell has to exactly control the kind and position of the modification reaction. For a long time, it was thought that the fully processed tRNA is irreversibly transported into the cytoplasm, where the amino acid is attached to the CCA terminus of the tRNA.[78] However, in the last couple of years it was discovered that this simple view is not correct. The process of aminoacylation[79] and splicing[80] can occur in the nucleus as well as in the cytoplasm. We know today that incompletely processed tRNAs can move from the nucleus into the cytoplasm and that these tRNAs can subsequently move back into the nucleus to complete the posttranscriptional modification process.[81] RNA maturation is consequently a highly complex process that also varies between organisms.[82] To exemplify the complex processes that cells perform to create a mature tRNA we will discuss in more detail the processing of $tRNA^{Phe}$ of *Saccharomyces cerevisiae* and thereafter the biosynthesis of the hypermodified base wybutosine.

2.3.1. *Distribution of the tRNA Maturation Processes in the Cell: The Example of yW-Modified tRNA*

Transcription[83] and the 3′ processing[84] of the yeast pri-$tRNA^{Phe}$ occur inside the nucleolus of the cell nucleus (Figure 7). The 3′-processing steps and the attachment of the CCA 3′ terminus as well as the introduction of the first 10 nucleoside modifications happen subsequently in the nucleoplasm and at the inner nuclear membrane.[85] During these processes, the 19 nucleotide long intron of this pre-$tRNA^{Phe}$ is still present, which is crucial for the formation of the modification m^5C at position 40.[86] However, prior to the biosynthesis of wybutosine and of three other modifications, the intron between positions 37 and 38 has to be cut out of the tRNA. The tRNA-splicing endonucleases responsible for this reaction are present in the outer membrane of mitochondria.[80] The corresponding intron-containing pre-$tRNA^{Phe}$ is, therefore, transported out of the cell nucleus to perform the splicing event at the mitochondrial membrane. After this splicing process, the $tRNA^{Phe}$ is further modified by cytosolic methyltransferases, which introduce the modifications m^2G10 and Cm32/Gm34, and is subsequently transported back into the cell nucleus,[85] where the first step of the complex biosynthesis of wybutosine, the N1-methylation of guanosine, takes place. This modification is performed by the enzyme TRM5 with the help of the cofactor *S*-adenosylmethionine (SAM). After this modification, the $tRNA^{Phe}$ is transported back into the cytoplasm, where all other steps of the yW biosynthesis occur.[85] The final aminoacylation step is likely to occur in the cytoplasm with the intact yW base present in position 37. However, it has been shown that the amino acid loading can also happen in the nucleus.[87]

It is currently believed that the transport of tRNAs in and out of the nucleus also has regulatory functions and that it is needed for the proofreading of erroneously modified or misfolded tRNAs.[81b] For example, it is known that the eukaryotic carrier protein exportin-t recognizes intact aminoacylated tRNAs much better than pre-tRNAs or tRNAs with structural defects.[79b,87] Hypomodified or misfolded tRNAs can in this way be recognized in the cell nucleus to target them for degradation.[88] It is also known that under specific stress conditions tRNAs are concentrated inside the nucleus, thereby providing important impulses for the biosynthesis of amino acids.[81b,89] If these amino acids are subsequently added to the medium and are loaded onto the tRNAs, these are immediately excreted from the nucleus into the cytoplasm.[89] This observation indicates that the distribution of tRNAs between the cytoplasm and the nucleus is reversible and has indeed some regulatory function. A very interesting observa-

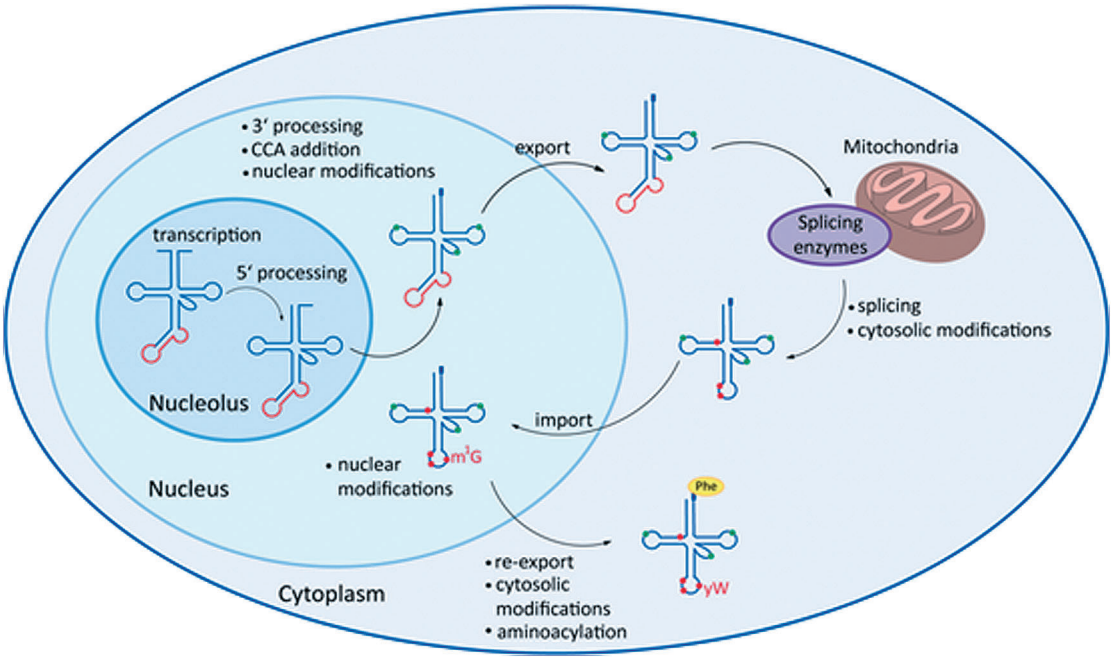


Figure 7. Depiction of the biosynthesis of the yW base and the locations in the cell where the individual tRNA maturation steps occur.

tion is that the import and export of tRNAs is also connected to the recognition of DNA lesions. If many DNA lesions are recognized in the genome of yeast, the export protein Los1 is arrested in the cytosol. This leads to an accumulation of unspliced tRNAs in the cell nucleus, thus resulting in down-regulation of translation and arrest in the G1 state.[90]

2.3.2. *Biosynthesis of the yW Base*

The biosynthesis of the hypermodified bases wybutosine yW and its hydroxylated derivative hydroxywybutosine OHyW has been fully elucidated over the last few years (Figure 8). Of the six enzymes (TRM5, TYW1–5) that are needed for the biosynthesis of wybutosine, TYW1, TYW2, TYW4, and TYW5 have been structurally characterized by the Nureki research group.[91] As mentioned previously, the first step in the biosynthesis of both bases involves methylation of the N1-position of guanosine by the methyltransferase TRM5.[92] This is followed by the formation of the characteristic tricyclic ring structure of yW by using pyruvate as substrate.[93] This reaction is catalyzed by the enzyme TYW1.[91a,94] In the next step the α-amino-α-carboxypropyl (acp) group of SAM is attached to the tricyclic core structure by TYW2.[91d,94b,95] Subsequently, an additional methyl group is introduced at position 4 by TYW3.[94b] The last step of the biosynthesis of yW consists of two reactions, which are both catalyzed by TYW4.[91b] In these steps the carboxy group of the acp group is first methylated and then the amino group is methoxycarbonylated. This process involves fixation of CO_2. The even more complex modification hydroxywybutosine,

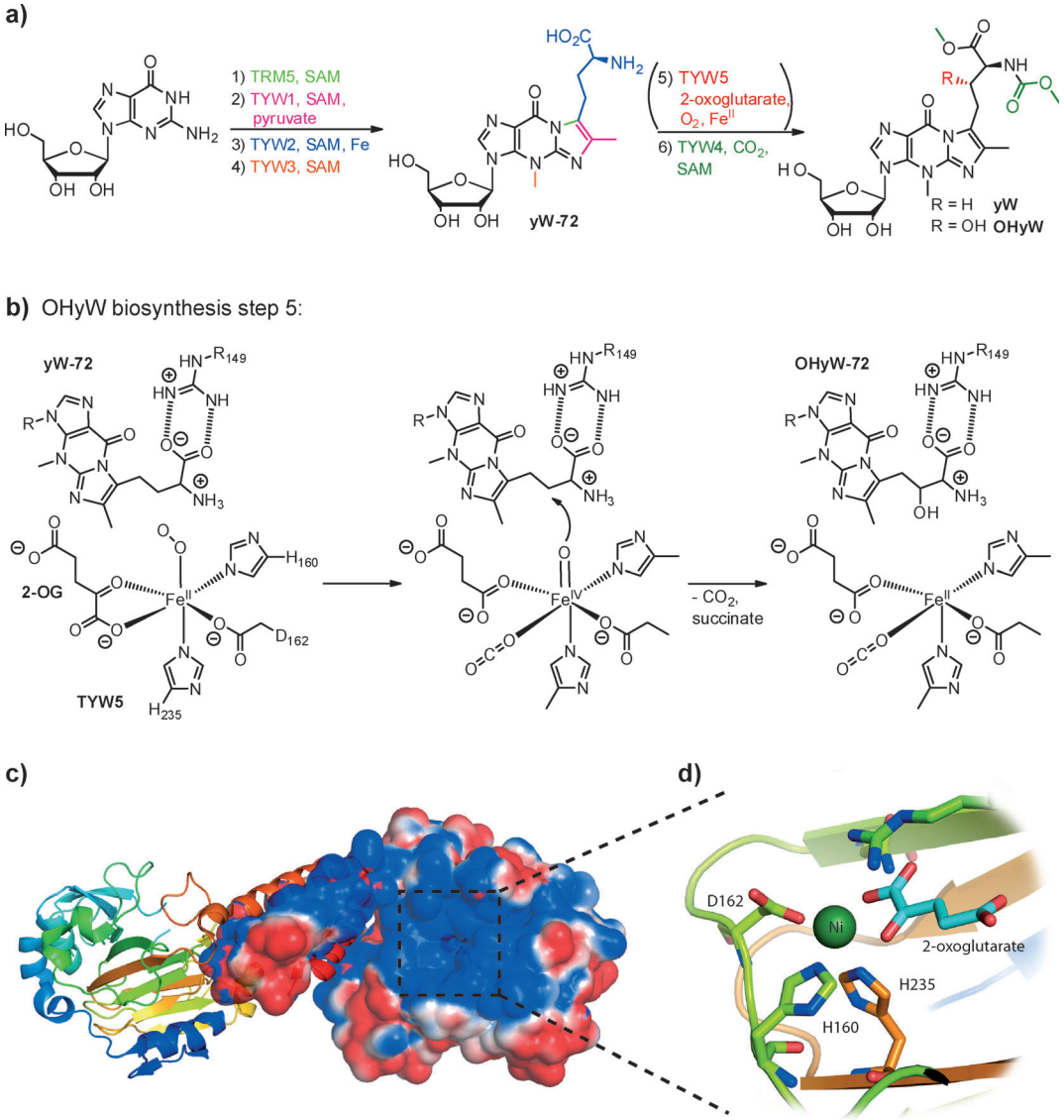


Figure 8. a) Biosynthesis of wybutosine and hydroxywybutosine. b) Proposed mechanism for the introduction of the hydroxy group at the β position of the wybutosine acp side chain. Figure adapted from Ref. [91c]. c) Schematic view of the TYW5 dimer. The left subunit is shown as a secondary structure representation; the right subunit shows the solvent-exposed surface with positive charges in blue and negative charges in red. Positive charges surround the active site, which is likely to be the RNA-binding interface. d) Magnified view of the active site of TYW5, which was crystallized with nickel (instead of iron) and 2-oxoglutarate, but without the RNA substrate.

which is found in higher eukaryotes, carries an additional hydroxy group in the β position to the acp side chain. The synthesis of this modification requires an unusual C–H activation step. In seminal work, the Suzuki and Nureki research groups isolated a human Jumonji C domain containing protein (hTYW5) in 2010 that performs this unusual hydroxylation reaction.[96] Similar to the previously mentioned ABH8, TYW5 is a member of the iron(II)/2-OG dependent dioxygenase family, that will be further discussed later in Section 3. The crystal structure of the human protein (hTYW5) shows a conserved His-X-(Asp/Glu)-X_n-His motif in the active center that chelates an iron(II) ion (Figure 8).[91c] Furthermore, two conserved arginine residues complex the acp side chain. The proposed mechanism for the C–H activation reaction (Figure 8) is similar to that of other dioxygenases.[97] The enzyme first binds an iron(II) ion, molecular oxygen, 2-oxoglutarate, and wybutosine. A reactive oxoferryl species is generated by iron(II), O_2, and 2-oxoglutarate, which is then converted into CO_2 and succinate. The reactive oxoferryl species thereafter attacks the β position of the acp side chain to form the hydroxywybutosine derivative OHyW-72,[91c] which is subsequently transformed to OHyW by TYW4. It is fascinating that nature is able to use complex radical chemistry to perform the needed transformations on a complex DNA/RNA substrate that is normally easily degraded by radical reactions.[98]

2.3.3. Modifications Present in Noncoding RNA

Noncoding RNAs (ncRNA) are RNA molecules that are not translated into a protein.[99] The RNA family includes tRNA, rRNA, snoRNA, miRNA, siRNA, piRNA (slightly longer miRNAs associated with piwi proteins in germ line cells), and also long noncoding RNAs (lncRNA). The small noncoding RNAs are known to be internally and externally modified. In 2005 the research group of Chen showed that the 3′ end of miRNAs are 2′-O-methylated in *Arabidopsis thaliana*.[100] They showed that the methyltransferase protein HEN1 adds a methyl group at the end of miRNAs and siRNAs to stabilizes these molecules. Moreover, the abundance of miRNAs was reduced and the length of the miRNAs was different in *hen1* mutants.[101] It was proven that the methylation reaction protects the miRNA from uridylation with up to five uridines, which, in turn, would lead to degradation of the miRNA.[102] While the 3′ termini of animal miRNAs are generally unmodified, the 3′ end of piRNAs in mouse,[103] zebrafish,[104] and *Drosophila*[105] are 2′-O-methylated. The siRNA in Drosophila is also 2′-O-methylated at the 3′ end. To modify the 3′-end of the *Drosophila* siRNA, the HEN1 orthologue (DmHEN1) has to assemble with the argonaute protein 2 (Ago2), which implies that the modification may be the final step in building the RISC complex.[105] Internal 2′-O-methylations are, for example, important in the small noncoding BC1-RNA, which is highly expressed in neurons and enriched in synapses. The presence of 2′-O-methylation alters the RNA structure and modulates the interaction with important proteins. When the modifications are absent, the protein can bind the BC1-RNA. In this way, the modification status regulates the translation of target mRNAs at synapses.[106] A recent review by Kim et al. covered these important aspects of ncRNA modifications involved in the dynamic regulation of gene silencing.[107] There are also other modifications that are incorporated internally in noncoding RNAs. Adenosine to inosine editing was shown to take place in some pri-miRNAs. An A to I edited pri-miRNA was not able to undergo further processing towards pre-miRNA and subsequently miRNA, thus giving rise to the speculation that this kind of editing changes the structure of the precursor miRNA and prevents the miRNAs from further processing.[108] Another editing process, the isomerization of uridine to pseudouridine (see Section 2.3), takes place in several noncoding RNAs such as rRNA, tRNA, small nuclear RNA,[109] and small nucleolar RNA. Furthermore, the RNA component of the human telomerase complex contains potential pseudouridinylation sites.[110] Pseudouridylation influences the regulation of gene expression, which was shown by a co-activator RNA molecule—a so called steroid receptor RNA activator (SRA). This SRA is pseudouridylated by mouse pseudouridine synthase (mPUS1p) at different sites, which seems to establish an altered structure that can interact differently with proteins that regulate the transcription of target genes.[111]

In summary we can conclude that although tRNA is the RNA species with the most and chemically most complex modified RNA bases (e.g. queuosine, wybutysine) we are learning that modifications are also of central importance in other RNA species. Modified nucleosides are found in ribosomal RNAs (rRNAs) and small noncoding RNAs (sncRNAs), such as miRNA, piRNA, and most recently also in other long (lncRNAs), where they influence the folding, the stability of the molecules, and in particular the interaction with critical proteins. Noncanonical nucleosides are to a great extent responsible for the diverse functions that these noncoding RNAs perform.

2.4. Quantification and Genome-Wide Mapping of RNA Modifications

The detection and structural characterization of modified RNA nucleosides is a well-developed field and a number of excellent reviews have been published.[112] It is still problematic to localize and quantify modified nucleosides in different RNA species and different tissues. Most sequencing technologies rely on the prior amplification of a reverse-transcribed RNA by the polymerase chain reaction (PCR). Modifications present in the RNA strand that do not affect the base-pairing behavior will be lost during this step. Furthermore, the enzymes needed to perform reverse transcription and PCR may also be blocked by the modification, possibly resulting in premature termination of the cDNA fragment or in localized mutagenesis.[113] However, by smart assay design, these features were utilized for genome-wide mapping of some modifications. A to I editing sites, for example, were detected by chemical modification with acrylonitrile, which converts inosine into a block of reverse transcription.[114] Next-generation sequencing (NGS) of bisulfite-treated RNA enabled the genome-wide identification of m^5C in human RNA.[115]

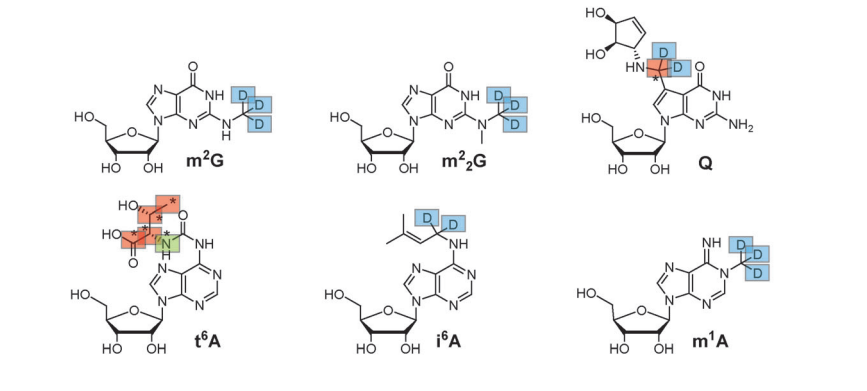
Durairaj and Limbach applied chemical derivatization of pseudouridines in combination with endonuclease digest and MALDI-TOF MS to identify the pseudouridination sites in complex RNA mixtures.[116] Just recently, a genome-wide map of the m^6A modification was published that is based on antibody enrichment of m^6A-containing fragments, followed by next-generation RNA sequencing (RNAseq).[117] These methods enable the detection of certain modifications inside RNA. For comparison, a whole range of modifications could be recently detected on a fine-tuned tiled microarray. By comparing wildtype and modification-deficient yeast stains, enzyme-specific modification sites were detected and even novel sites could be identified.[118] Despite this progress, it is today difficult to characterize the exact RNA species or even position at which a certain modification occurs. What is needed is an efficient method that allows the isolation of individual RNA species from complex tissue samples, for subsequent MS analysis after partial digestion. Such a technique could speed up the positional scanning of tRNAs and maybe other RNA species for base modifications. Towards this goal a method called reciprocal circulating chromatography (RCC) was recently developed to isolate individual RNA species from complex mixtures by using a solid support, coated with DNA capture probes with sequences complementary to segments present in the RNA species under investigation. The advantage of the RCC technology is that it is amenable to high-throughput strategies.[119] A different approach using specific probes for known modification sites was recently developed to quantify the levels of m^1A modifications in tRNA on a genome-wide scale. Here, primer extension in combination with PCR amplification followed by microarray hybridization was employed.[120]

The quantification of the modification levels in tissues and cells has also been the focus of intense investigation, and various methods have been designed for this purpose.[112a,121] To date, some of the most precise techniques rely on mass spectrometry, namely LC-MS combined with isotope dilution protocols[122] or LC-coupled tandem mass spectrometry.[123] Although LC-MS is perfectly suited to detect individual modifications and MS^n techniques can help to decipher the structure of the molecules, MS methods are not quantitative because the signal intensity of a molecule depends on its ionization potential rather than on the amount of the compound in the analyte. However, with the help of isotope-labeled standards, which possess the same retention time and ionization potentials as their natural counterparts, the exact amount of a certain modification can be quantified. The position of the modified base in the RNA species, however, cannot be resolved. By using this method it was recently possible to get broad quantitative insights into the distribution of modifications in different mouse and pork tissues. Sample data obtained for pork tissues are shown in Figure 9.[124]

As shown in Figure 9, the levels of modified nucleosides vary quite significantly in different tissues, thus highlighting that, while the sequence code is the same in all cells of a complex organism, the level of overall tRNA modification content is tissue-specific. This could reflect a tissue-specific modulation of protein translation and metabolism.

Other quantification approaches show that environmental conditions can influence the modification content. A recent study showed that the levels of specific tRNA modifications in *Saccharomyces cerevisiae* dynamically change as a function of various stress conditions,[123b] thus suggesting that modification levels also respond to environmental changes.

Isotope dilution based LC-MS analysis was also used to investigate and quantify modified nucleosides in tRNA of different bacterial and eukaryotic species. The obtained quantification data varied so strongly between the investigated organisms that it was possible to use the data for the construction of a "phylogenetic" tree (Figure 10) that correlates to trees generated using classical genomic sequences data.[125] These data show that while smaller variations in the



Nucleosides / Tissues	m^2G	m^2_2G	Q	t^6A	i^6A	m^1A
Cerebellum	1036,4	566,8	83,5	237,3	41,7	1102,8
Liver	1176,5	562,0	55,2	214,6	39,6	941,6
Lung	879,3	427,2	33,9	184,1	33,6	876,3
Kidney	896,0	451,5	32,3	177,8	30,7	871,6
Spleen	793,6	464,3	42,7	182,1	30,7	833,6
Heart	673,1	372,9	29,6	175,4	29,0	810,9
Tongue	794,0	385,9	21,9	170,6	25,7	846,5

Figure 9. Depiction of synthetic stable isotope labeled modified nucleosides and the approximate positions of their natural analogues within tRNA. The color-coded table shows the amounts of selected modified tRNA nucleosides in different pork tissues, revealing tissue-specific overall modification levels (values are given per 1000 tRNA molecules, colors mirror the modification content, red: high, green: low).

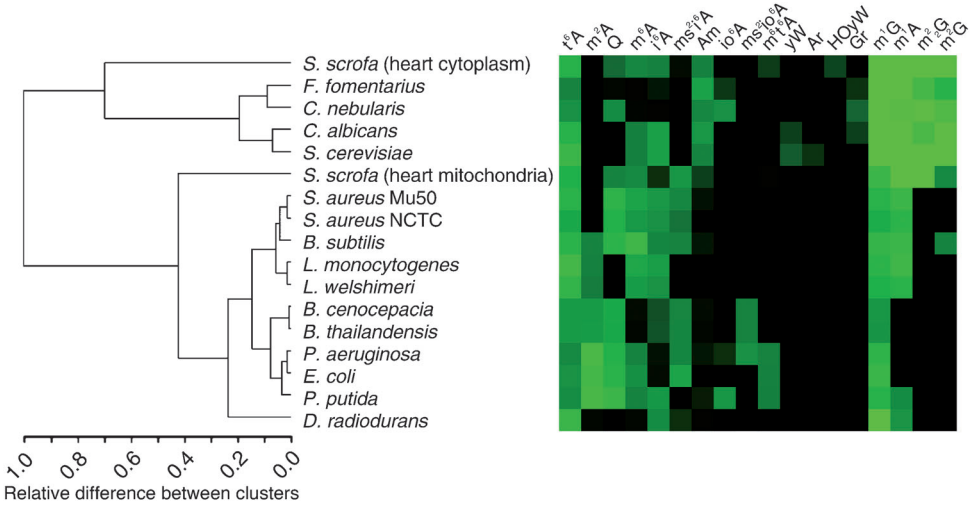


Figure 10. "Phylogenetic" tree constructed from the tRNA modification level data for a range of bacteria and eukaryotes. The intensity of the green color in the plot represents the level of the modifications shown on the *x*-axis for each species shown on the *y*-axis.

content of modified nucleosides occur in response to stress or in adaptation to organ-specific tasks, the overall set and degree of nucleoside modifications is determined by the phylogenetic origin of the species.

2.5. The Role of RNA Modifications in Human Diseases and Diagnosis

Although it is not clear for most modifications to what extent they influence the health of an organism, a few recent results have shown that altered base modifications contribute to the development of diseases. Here it is interesting that most diseases associated with base modifications are connected to energy metabolism (mitochondrial disease, type-2 diabetes, and obesity).

2.5.1. The Role of RNA Modifications in Human Diseases

A particular modification whose absence is closely associated with a disease is the taurine modification of the wobble-base uridine in mitochondrial $tRNA^{Leu}_{UUR}$ and $tRNA^{Lys}$.[126] Undermodified wobble positions in these tRNA have been shown to cause two major classes of mitochondrial diseases, MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes) and MERRF (myoclonus epilepsy associated with ragged-red fibers).[126,127] The lack of modification leads to insufficient decoding of leucine and lysine wobble codons, which then leads to translation deficits in mitochondria.

Similarly, another wobble position modification Q has also been connected to several diseases. A low Q content seems to be a characteristic of various tumor types. For example, it is known that lung cancer tissue contains a lower Q level than healthy tissue.[128] Furthermore, lower Q levels were correlated with higher expression of a tRNA-guanine transglycosylases TGT subunit in leukaemia and colon carcinoma cells.[129] Multiple diseases are associated with defects in the adenosine (A) to inosine (I) editing system by adenine deaminases acting on RNA (ADAR) enzymes. Deamination of adenosine yields inosine, which pairs preferentially with cytosine. Most editing events have been described for proteins of the central nervous system. Important examples are the glutamate receptor family (GluR) and the serotonin receptor 5-HT2C. The diseases associated with defects in ADAR function include neurodegenerative diseases, diabetes, cancer, and AIDS (see the review by Gallo and Locatelli).[130]

Recently, two other modifications have become the center of attention, after their biosynthetic enzymes turned out to be well-known genetic factors of obesity (FTO) and type-2 diabetes (CDKAL1). While alleles of both genes could be identified early on with high statistical significance to cause prevalence for the development of a disease,[131] the function of these proteins remained unknown until recently. The research groups of Fontecave and Tomizawa proved that the type-2 diabetes (T2D) related protein CDKAL1 (CDK5 regulatory subunit associated protein 1-like 1) is an *S*-adenosylmethionine (SAM) dependent methylthiotransferase, which converts t^6A in $tRNA^{Lys}_{UUU}$ into ms^2t^6A.[132] This protein is located in the endoplasmtic reticulum (ER) and its knockout in mice significantly affects insulin secretion and leads to the dysfunction of β cells, most probably by causing misreading of the Lys_{UUU} codon.[132b] Interestingly, CDKAL1-deficient mice are more susceptible to stress from a high-fat diet, while the second reinvestigated protein FTO (fat mass and obesity associated protein) is a well-studied factor in obesity. Both genes are interconnected by numerous T2D-related clinical studies (see http://string-db.org). FTO was first identified by Schofield and co-workers as an iron(II) and 2-oxoglutarate-dependent dioxygenase, which is able to demethylate the alkylated DNA lesions m^3dT, m^1dA, and to a smaller extent m^3dC in single-stranded DNA.[133] Later He and co-workers described an extended substrate spectrum towards m^3U and RNA.[134] Thus, the precise substrate for the FTO was somewhat unclear for a long time, until He and co-workers found that FTO is an RNA-specific m^6A demethylating enzyme.[135] m^6A is a highly abundant modification in mRNA, and the activity of FTO on m^6A containing prolactin mRNA could be shown. The recent sequencing of m^6A found it to be enriched in transcription start sites and near to stop codons. It was further found that m^6A influences the splicing of mRNA.[117] FTO is a nuclear protein that localizes in distinct speckles, which are potential mRNA splicing sites.[135]

It might be interesting to note in this respect that a paralogue of CDKAL1, CDK5RAP1 (CDK5 regulatory subunit associated protein 1), is a known repressor of the highly investigated kinase CDK5.[136] CDK5 is not only involved in the regulation of brain development and neurodegenerative diseases such as Alzheimer's, but also plays a role in insulin secretion.[137] Its inhibitor CDK5RAP1 is the human orthologue of the bacterial MiaB proteins and the methylthiotransferase responsible for the synthesis of ms^2i^6A in mitochondrial tRNA and other RNA species.[138] All diseases mentioned here are interconnected by the function of mitochondria.[139] Taking these results together, they clearly imply important and interconnected functions of RNA modifications in the development of the lifestyle diseases obesity and type-2 diabetes.

2.5.2. RNA Modifications as Diagnostic Markers and Drug Targets

Besides the direct involvement of RNA modifications in disease, methods to quantify the levels of a specific modification can be used as diagnostic tools. As early as 1978, Nishimura and co-workers detected decreased levels of queuosine in tumor tissues.[140] This property was later developed as a biomarker for lung cancer.[141] The presence of this hypermodification is known to be responsible for the virulence of *Shigella flexneri*, a bacterium that causes dysentery (shigellosis). Studies showed that mutations of the TGT gene, which is involved in the exchange of guanine to $preQ_1$ in *S. flexneri* lead to a significant loss of pathogenicity.[142] This makes the TGT enzyme an interesting target to conquer shigellosis.[142a] Medicinal chemical approaches directed at blocking the TGT enzyme are currently being pursued.[143] In particular, the application of sensitive mass spectrometric techniques allowed the development of RNA modification levels as biomarkers. The analysis of urinary nucleosides by LC or CE in combination with mass spectrometry has frequently been used in the diagnosis of different cancer types.[144]

3. Modified Bases in DNA

In comparison to the large diversity of base modifications that are found in RNA and the complex biosynthetic machinery that involves either single-acting modifying enzymes or snRNP protein complexes, the number of base modifications found in DNA is small. This is certainly connected with the important but limited function of DNA as a carrier of genetic information. As such, the molecules persist mainly in a DNA duplex form in which the bases are pointing inwards to the helix core to form Watson–Crick hydrogen bonds with the counterbase. RNA, in contrast, is mostly single stranded and possesses complex, three-dimensional folds in which many but not all the bases form base pairs. Important modifications to the DNA are, however, DNA lesions. Lesions are modified DNA bases, which form in response to damaging agents such as UV irradiation as well as reaction of DNA with reactive oxygen species or molecules that are taken up from the environment, for example, with nutrition, such as aromatic amines. These are activated in the liver to give nitronium intermediates, which react with the nucleophilic sites of the DNA bases to give defined, so-called bulky adduct lesions. All these DNA lesions are mutagenic or cytotoxic. Thus, cells possess in consequence dedicated DNA repair and lesion tolerance systems to counteract the mutagenic and often even lethal effects of these modified DNA bases. DNA lesions are not biosynthetically modified bases, and hence not covered in this Review. For more information regarding this important research field we recommend further reading.[145]

In addition to these DNA lesions, a few enzymatically introduced DNA modifications are known, which are depicted in Figure 11. Adenines and cytosine are methylated

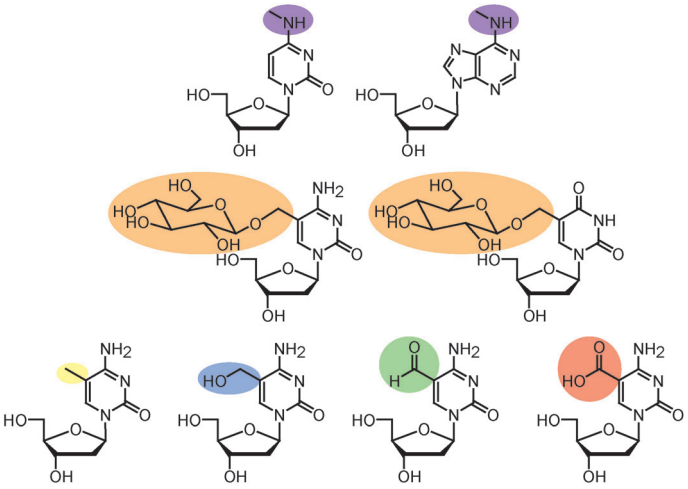


Figure 11. Top: m^4dC and m^6dA. Middle: Left: β-D-glycosylated forms of 5-hmC/hm^5dC and 5-hmU/hm^5dU. Bottom: Depiction of the four epigenetic bases, which establish a second level of genetic information. This level encodes the epigenetic information of gene activity. In accordance with the RNA nomenclature, the 5-modified DNA bases should be abbreviated as m^5dC, hm^5dC, hm^5dU, f^5dC, and ca^5dC. However, the abbreviation 5-mC, 5-hmC, 5-fC, and 5-caC or even shorter mC, hmC, fC, and caC are commonly used in the DNA community and we use this accepted nomenclature in the text as well.

at the N6- and C5- or N4-positions, respectively, in specific sequences of bacterial genomes. Both m^5dC and m^4dC protect the DNA of the bacterium from degradation by its own restriction enzymes. These restriction enzymes are produced to degrade foreign DNA, which is unmethylated, as part of a defense system against infection by bacteriophages. m^6dA isin addition to this function also involved in the regulation of virulence, mismatch repair, timing of DNA replication, and control of gene expression.[146] Since dA, which is introduced during replication, is initially unmethylated, it allows cells to distinguish for the few seconds required for re-methylation, the newly synthesized (nascent) strand from the original parental template strand. This time allows replication errors to be repaired. Newly synthesized origin of replication sites remain longer in the hemimethylated state and are protected from methylation by binding of the SeqA protein. Since these

hemimethylated origins of replication are inactive, any further cell division is blocked until the origin of replication is again fully methylated.[147]

The modified bases 5-hydroxymethyluracil (hmU/hm^{5}dU, nucleoside J) and 5-hydroxymethylcytosine (hmC/hm^{5}dC) were found to occur in even-numbered T bacteriophages (hm^{5}dC) and in the species *Trypanosoma brucei* (hm^{5}dU). In both cases, the hydroxymethyl group is introduced on the corresponding triphosphates prior to incorporation into DNA and subsequently glycosylated. The function of both bases is again to protect the DNA of the organism from degradation by restriction enzymes produced by the host.[13b] hmU is also thought to be involved in repression of transcription.

The best studied noncanonical nucleobase however is 5-methyldeoxycytosine (m^{5}dC). It is considered to be the fifth major component of the DNA molecule. The existence of the DNA base, which contains just one methyl group at position C5 of cytosine, was initially postulated in 1925 by Johnson and Cokel, who discovered the base in the tuberculosis bacterium.[148] The result, which was initially treated as pure speculation, was confirmed 24 years later by Wyatt, who proved that m^{5}dC is indeed a component of the genetic material.[149] It is today well established that this base modification regulates gene transcription. While the DNA sequence remains unchanged, the transcriptional activity of the gene can be changed by the methylation of cytosine bases, which defines the research field of epigenetics. m^{5}dC is an epigenetic modification in plants, fungi, and animals that controls gene expression, while in bacteria it serves again as a protective base against infection by bacteriophages (see above).

3.1. Discovery of the Modified DNA Bases 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in Higher Organisms

The research field of m^{5}dC is so vast that it cannot be covered here. Instead, we concentrate on the discovery process that led recently to the observation of other modified cytosine-derived bases in DNA.

In 2009, two research groups showed independently that 5-hydroxymethylcytosine (hmC/hm^{5}dC) is present in mouse stem cell DNA and in neurons, in addition to the known 5-methylcytosine (mC/m^{5}dC).[150] Actually, this was a re-discovery, because in 1972 hmC was already detected as a component of animal DNA by Yura and co-workers.[151] In subsequent work, it was shown that hydroxymethylcytosine is particularly prevalent in the DNA from the central nervous system (CNS).[152] This discovery was possible after the synthesis of isotope-labeled mC and hmC and the use of quantitative LC-MS techniques, already described for the quantification of RNA modifications.[122a] It was shown with this technology that while the mC content is rather constant in all tissues at around 4.5%, the hmC level varies strongly. We know today that this base is present in almost all tissues, although in largely different amounts, and that hmC is strongly accumulated in brain tissues and in particular in brain tissues associated with higher cognitive function. The hmC levels were found to be strongly reduced in brain tumor tissues.[153] Besides neuronal tissue, significant levels of hmC are present in embryonic stem cells. This observation fueled the idea that the hmC base is linked to the epigenetic programming of the genome and in particular to the enigmatic process of active demethylation.[154]

An initial attempt to find the most likely occurring oxidation products of hmC, namely 5-formylcytosine (fC/f^{5}dC) and 5-carboxylcytosine (caC/ca^{5}dC), in brain tissues failed.[155] It was shown that even though hmC is highly abundant in the central nervous system (CNS), neither fC nor caC were detectable at significant levels. However, in a continuation of this study, fC was recently discovered in stem cell DNA.[156] Here, caC was also discovered shortly thereafter.[157] It was reported that caC is detectable at larger levels in special stem cells that lack a DNA repair enzyme (TDG) known to cleave the glycosylic bond between the deoxyribose and caC to repair this modification.[157b] This result led to the conclusion that caC is removed from DNA in a base excision repair process. A crystal structure of TDG bound to caC-containing DNA further supported this hypothesis.[158]

3.2. Role of the Modified DNA Bases 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine

Even though the exact levels and the distribution of fC and caC in stem cells and somatic tissues are not yet clear, the question is what their function might be. mC is an important base that is involved in many biological processes such as gene expression, genome imprinting, X chromosome inactivation, and the suppression of transposable elements.[159] If a promoter features significant levels of mC, the corresponding gene is frequently found to be silenced. It is currently believed that the base hmC has a similar epigenetic function. Proteins that are known to bind to mC do not interact with hmC, which might change the transcriptional activity of hydroxymethylated versus methylated genes. As such, hydroxymethylated promoters could already represent a reversal of silencing caused by cytosine methylation.[160] In addition, the bases hmC, fC, and caC are potential intermediates in a process that aims at replacing the epigenetic mC base with an unmodified dC.[161] This would allow cells to flexibly reprogram the epigenetic state of a gene. The mechanism is so far unclear, but it can in principle occur in two different ways (Figure 12). The first mechanism relies on base excision repair.[162] Here it is thought that the modified bases fC and caC are recognized by specific glycosylases such as the thymine-DNA glycosylase (TDG) enzyme, which can cleave the glycosylic bonds between the heterocycles and the sugar to create an abasic site (Figure 12), which is subsequently repaired.[163] In the course of this process the modified bases fC and caC are replaced by an unmodified dC, which leads finally to the exchange of an epigenetic mC mark by an unmodified canonical dC. A similar mechanism may enable cells to exchange hmC by a dC. It was shown, however, that hmC is not a good substrate for TDG. Enzymes such as AID/APOBEC were shown to deaminate hmC to hmU, and because hmU is a good substrate for the glycosylases SMUG1

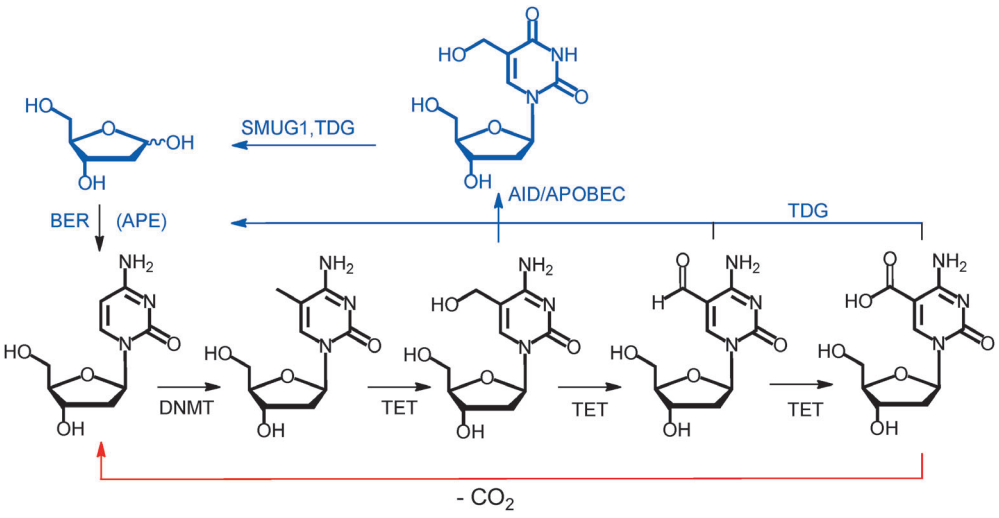


Figure 12. Possible demethylation mechanisms based on the oxidation of mC. In blue: the glycosylase/BER mechanism that creates first an abasic site by the action of the thymine DNA glycosylase (TDG) or SMUG 1 and then a strand nick during the removal process by BER as a result of the action of the apurinic/apyrimidinic endonuclease (APE) followed by insertion of a new dC. hmC is not a substrate for TDG. Here, a first deamination of hydroxymethyluracil (hmU) by the enzyme AID/APOBEC is presumed to occur. The putative decarboxylative pathway is given in red.

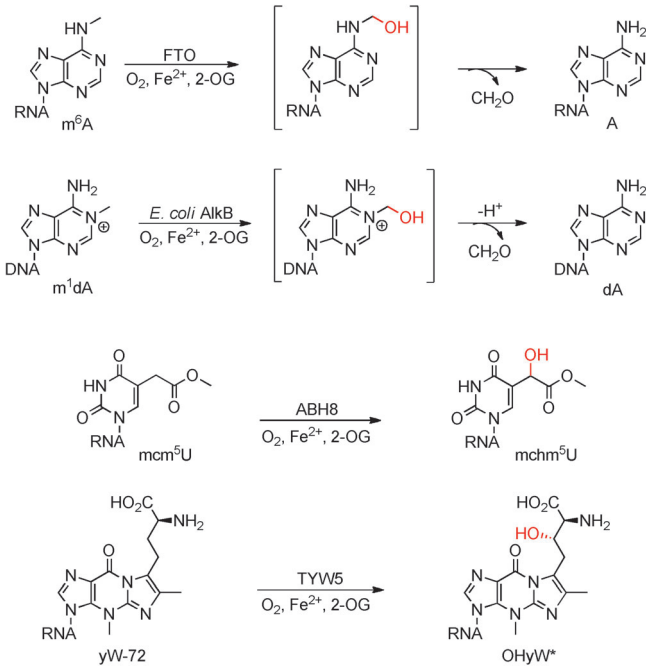
and TDG a replacement of hmC by dC is possible by the deamination "detour".[164]

The second chemically very attractive possibility to achieve the demethylation would be the final decarboxylation of caC or alternatively dehydroxymethylation of hmC or deformylation of fC. In this scenario, enzymes would specifically oxidize mC bases that are targeted for removal in a stepwise process to caC so as to perform one of these processes. Decarboxylation of caC seems to be chemically the most attractive mechanism because similar decarboxylation reactions are known with orotate and isoorotate enzymes, which decarboxylate orotate and isoorotate to uracil.[165]

Although a specific decarboxylase acting on caC has not been discovered, it was recently shown that stem cell extracts can decarboxylate caC to dC, and it was shown that the decarboxylation of caC is chemically possible after saturation of the C5–C6 double bond,[166] as recently suggested.[167]

A major discovery in the field is the enzymes that perform the successive oxidation of mC to hmC, fC, and caC. These are the ten-eleven-translocation enzymes (TET 1, 2, and 3). These proteins are 2-oxoglutarate dioxygenases that use oxygen and an iron(II)-containing reactive site to perform a radical-based oxidation of the methyl group to initially give hmC. These enzymes are also able to convert hmC into fC and caC.[157] The later property was developed recently into a method for hmC sequencing with single-base precision.[168] A recombinant TET enzyme oxidizes hmC to caC, which reacts differently in bisulfite sequencing. An alternative approach uses a perruthenate-based chemical oxidation of hmC to fC for the subsequent bisulfite sequencing, thus achieving the same goal.[169]

In this Review, the interesting class of iron(II)/2-oxoglutarate-dependent dioxygenases was mentioned several times in different contexts. These enzymes are currently of tremendous interest because they are able to functionalize and to remove methyl groups that are attached to nucleobases. The removal reaction reverses the effect that these special base modifications impose. The protein FTO and the DNA repair enzymes of the ALKB family remove, for example, the methyl groups of m^6A or m^1dA, respectively. They oxidize the methyl group to give a hydroxymethyl functionality (Scheme 1). Since the methyl group is situated at a heteroatom, an "acetal" structure is thereby generated that is readily hydrolyzed, which leads to the complete loss of the methyl group. In contrast, the proteins ABH8 and TYW5 oxidize methyl groups to give hydroxymethyl groups in $mchm^5U$ and OHyW that are not eliminated because here the methyl groups are connected to another carbon atom. In this situation, the hydroxymethyl group is the endpoint of the biosynthesis. Finally, the recently discovered TET enzymes have the formidable task of fully removing the C5-methyl groups present in mC. This is achieved by oxidation of the methyl group to give fC and caC. These bases are known to be removed by TDG-based cleavage of the glycosidic bond, which circumvents the difficult C–C bond cleavage reaction. Further studies are needed to find the activity, either enzymatic or otherwise-catalyzed, that is responsible for the decarboxylating effect.



Scheme 1. Nucleoside oxidation reactions performed by different enzymes of the iron(II)/2-OG dioxygenase family.

4. Summary

It is well-known that DNA and RNA are composed of the four canonical nucleobases. Besides these, however, both polymers of life contain a variety of highly modified bases, which are able to increase the chemical information content. RNA is particularly rich in modifications, which reflects the highly diverse functions that RNA molecules perform inside cells. RNA not only encodes information, but translates the information into an amino acid sequence. In addition, RNAs possess regulatory and occasionally catalytic functions and they are critically involved as guide RNA strands in the modification processes themselves. The broad diversity and large number of base modifications present in RNA is clearly the result of an adaptation to their complex and varied functions. Most recently we learned that the chemical diversity in DNA is also larger than originally anticipated. In addition to the four canonical bases and the long-known DNA base modifications m^4dC, m^5dC, m^6dA, hm^5dU/hmU, and hm^5dC/hmC we are now aware that DNA contains in total four epigenetic bases (mC, hmC, fC, caC) of which fC and caC were just recently discovered. Again the modifications are needed for regulatory, epigenetic purposes. While all cells, as different as they are, contain the same genetic material, their vastly different function and properties inside complex higher organisms require the specific silencing and activation of cell-type-specific genes. The regulation of the underlying silencing and activation process requires an additional layer of epigenetic information, which is clearly linked to increased chemical diversity. This diversity is provided by the four epigenetic bases mC, hmC, fC, and caC. While DNA and RNA contain, in addition to the first coding level provided by the sequence of the canonical Watson–Crick bases, a second level of chemical complexity that is established by the introduction of specific DNA and RNA modifications. Research in the coming years will surely be dedicated to deciphering the biochemistry of the compounds, their spatial (sequence) arrangement in DNA and RNA, and the orchestration of the modification events inside cells. The research area comprising the investigation of DNA and RNA modifications is situated between the areas of chemical synthesis, biochemistry, and cell biology. This gives synthetically orientated researchers in the area of chemical biology a great chance to contribute to our further understanding of how nucleoside modifications are introduced and how they influence life.

We would like to thank the Excellence Cluster: Center for Integrated Protein Science CiPS^M, SFB 749, SFB 646, and the DFG-Normalverfahrensprojekt (CA 275/8-4) for financial support. D.P. thanks the Alexander von Humboldt Foundation for a postdoctoral fellowship.

Received: February 13, 2012
Revised: May 7, 2012
Published online: June 28, 2012

[1] a) R. Dahm, *Hum. Genet.* **2008**, *122*, 565 – 581; b) R. Olby, *Nature* **2003**, *421*, 402 – 405.
[2] E. Chargaff, *Experientia* **1950**, *6*, 201 – 209.
[3] P. A. Levene, *J. Biol. Chem.* **1919**, *40*, 415 – 424.
[4] a) A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39 – 56; b) M. McCarty, *Nature* **2003**, *421*, 406; c) O. T. Avery, C. M. MacLeod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137 – 158.
[5] a) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737 – 738; b) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964 – 967.
[6] a) R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740 – 741; b) M. H. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738 – 740.
[7] a) M. Guttman, J. L. Rinn, *Nature* **2012**, *482*, 339 – 346; b) L. He, G. J. Hannon, *Nat. Rev. Genet.* **2004**, *5*, 522 – 531; c) E. D. Wiklund, J. Kjems, S. J. Clark, *Epigenomics* **2010**, *2*, 823 – 840.
[8] a) T. R. Cech, *Gene* **1993**, *135*, 33 – 36; b) S. Altman, *Biosci. Rep.* **1990**, *10*, 317 – 337.
[9] J. S. Mattick, *EMBO J.* **2012**, *31*, 515 – 516.
[10] S. Y. Ng, R. Johnson, L. W. Stanton, *EMBO J.* **2012**, *31*, 522 – 533.
[11] a) M. E. Dinger, P. P. Amaral, T. R. Mercer, K. C. Pang, S. J. Bruce, B. B. Gardiner, M. E. Askarian-Amiri, K. Ru, G. Soldá, C. Simons, S. M. Sunkin, M. L. Crowe, S. M. Grimmond, A. C. Perkins, J. S. Mattick, *Genome Res.* **2008**, *18*, 1433 – 1445; b) M. Guttman, J. Donaghey, B. W. Carey, M. Garber, J. K. Grenier, G. Munson, G. Young, A. B. Lucas, R. Ach, L. Bruhn, X. Yang, I. Amit, A. Meissner, A. Regev, J. L. Rinn, D. E. Root, E. S. Lander, *Nature* **2011**, *477*, 295 – 300.
[12] T. R. Mercer, D. J. Gerhardt, M. E. Dinger, J. Crawford, C. Trapnell, J. A. Jeddeloh, J. S. Mattick, J. L. Rinn, *Nat. Biotechnol.* **2011**, *30*, 99 – 104.
[13] a) M. Ehrlich, G. G. Wilson, K. C. Kuo, C. W. Gehrke, *J. Bacteriol.* **1987**, *169*, 939 – 943; b) J. H. Gommers-Ampt, F. Van Leeuwen, A. L. de Beer, J. F. Vliegenthart, M. Dizdaroglu, J. A. Kowalak, P. F. Crain, P. Borst, *Cell* **1993**, *75*, 1129 – 1136.
[14] a) P. F. Agris, F. A. P. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1 – 13; b) H. Grosjean, V. de Crécy-Lagard, C. Marck, *FEBS Lett.* **2010**, *584*, 252 – 264.
[15] P. F. Agris, *EMBO Rep.* **2008**, *9*, 629 – 635.
[16] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934 – 4944.
[17] a) K. Phelps, A. Morris, P. A. Beal, *ACS Chem. Biol.* **2012**, *7*, 100 – 109; b) D. Iwata-Reuyl, *Curr. Opin. Chem. Biol.* **2008**, *12*, 126 – 133; c) E. M. Phizicky, A. K. Hopper, *Genes Dev.* **2010**, *24*, 1832 – 1860.
[18] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195 – D201.
[19] M. Charette, M. W. Gray, *IUBMB Life* **2000**, *49*, 341 – 351.
[20] E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, *584*, 265 – 271.
[21] A. Czerwoniec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118 – 121.
[22] Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611 – 631.
[23] a) Y. Ikeuchi, S. Kimura, T. Numata, D. Nakamura, T. Yokogawa, T. Ogata, T. Wada, T. Suzuki, T. Suzuki, *Nat. Chem. Biol.* **2010**, *6*, 277 – 282; b) D. Mandal, C. Köhrer, D. Su, S. P. Russell, K. Krivos, C. M. Castleberry, P. Blum, P. A. Limbach, D. Söll, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2872 – 2877.
[24] a) T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, *Nature* **1988**, *336*, 179 – 181; b) T. Muramatsu, S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, T. Miyazawa, *J. Biol. Chem.* **1988**, *263*, 9261 – 9267.

[25] a) J. A. Kowalak, E. Bruenger, J. A. McCloskey, *J. Biol. Chem.* **1995**, *270*, 17758 – 17764; b) A. M. Giessing, S. S. Jensen, A. Rasmussen, L. H. Hansen, A. Gondela, K. Long, B. Vester, F. Kirpekar, *RNA* **2009**, *15*, 327 – 336.
[26] a) K. S. Long, J. Poehlsgaard, C. Kehrenberg, S. Schwarz, B. Vester, *Antimicrob. Agents Chemother.* **2006**, *50*, 2500 – 2505; b) D. N. Wilson, *Ann. N. Y. Acad. Sci.* **2011**, *1241*, 1 – 16.
[27] A. A. Van Aerschot, P. Mamos, N. J. Weyns, S. Ikeda, E. De Clercq, P. A. Herdewijn, *J. Med. Chem.* **1993**, *36*, 2938 – 2942.
[28] a) C. Laezza, M. Notarnicola, M. G. Caruso, C. Messa, M. Macchia, S. Bertini, F. Minutolo, G. Portella, L. Fiorentino, S. Stingo, M. Bifulco, *FASEB J.* **2006**, *20*, 412 – 418; b) H. Sakakibara, *Ann. Rev. Plant Biol.* **2006**, *57*, 431 – 449.
[29] T. Suzuki, T. Wada, K. Saigo, K. Watanabe, *Nucleic Acids Res. Suppl.* **2001**, 257 – 258.
[30] a) Y. G. Chen, W. E. Kowtoniuk, I. Agarwal, Y. Shen, D. R. Liu, *Nat. Chem. Biol.* **2009**, *5*, 879 – 881; b) W. E. Kowtoniuk, Y. Shen, J. M. Heemstra, I. Agarwal, D. R. Liu, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7768 – 7773.
[31] H. Kasai, K. Nakanishi, R. D. Macfarlane, D. F. Torgerson, Z. Ohashi, J. A. McCloskey, H. J. Gross, S. Nishimura, *J. Am. Chem. Soc.* **1976**, *98*, 5044 – 5046.
[32] D. W. Phillipson, C. G. Edmonds, P. F. Crain, D. L. Smith, D. R. Davis, J. A. McCloskey, *J. Biol. Chem.* **1987**, *262*, 3462 – 3471.
[33] a) J. C. Salazar, A. Ambrogelly, P. F. Crain, J. A. McCloskey, D. Soll, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7536 – 7541; b) M. Blaise, H. D. Becker, J. Lapointe, C. Cambillau, R. Giege, D. Kern, *Biochimie* **2005**, *87*, 847 – 861.
[34] D. Iwata-Reuyl, *Bioorg. Chem.* **2003**, *31*, 24 – 43.
[35] N. Okada, S. Nishimura, *Nucleic Acids Res.* **1977**, *4*, 2931 – 2938.
[36] F. Harada, S. Nishimura, *Biochemistry* **1972**, *11*, 301 – 308.
[37] a) J. M. Gregson, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, D. W. Phillipson, J. A. McCloskey, *J. Biol. Chem.* **1993**, *268*, 10076 – 10086; b) M. Sprinzl, T. Hartmann, J. Weber, J. Blank, R. Zeidler, *Nucleic Acids Res.* **1989**, *17*, R1 – 172; c) T. Ohgi, T. Kondo, T. Goto, *J. Am. Chem. Soc.* **1979**, *101*, 3629 – 3633; d) F. Klepper, E. M. Jahn, V. Hickmann, T. Carell, *Angew. Chem.* **2007**, *119*, 2377 – 2379; *Angew. Chem. Int. Ed.* **2007**, *46*, 2325 – 2327; e) T. Brückl, F. Klepper, K. Gutsmiedl, T. Carell, *Org. Biomol. Chem.* **2007**, *5*, 3821 – 3825.
[38] N. Shindo-Okada, N. Okada, T. Ohgi, T. Goto, S. Nishimura, *Biochemistry* **1980**, *19*, 395 – 400.
[39] a) U. Günduz, J. R. Katze, *Biochem. Biophys. Res. Commun.* **1982**, *109*, 159 – 167; b) U. Gunduz, J. R. Katze, *J. Biol. Chem.* **1984**, *259*, 1110 – 1113.
[40] a) G. A. Garcia, J. D. Kittendorf, *Bioorg. Chem.* **2005**, *33*, 229 – 251; b) N. Okada, S. Noguchi, H. Kasai, N. Shindo-Okada, T. Ohgi, T. Goto, S. Nishimura, *J. Biol. Chem.* **1979**, *254*, 3067 – 3073; c) B. Stengl, K. Reuter, G. Klebe, *ChemBioChem* **2005**, *6*, 1926 – 1939.
[41] Y. Kuchino, H. Kasai, K. Nihei, S. Nishimura, *Nucleic Acids Res.* **1976**, *3*, 393 – 398.
[42] a) B. W. Lee, S. G. Van Lanen, D. Iwata-Reuyl, *Biochemistry* **2007**, *46*, 12844 – 12854; b) R. M. McCarty, A. Somogyi, G. Lin, N. E. Jacobsen, V. Bandarian, *Biochemistry* **2009**, *48*, 3847 – 3852.
[43] R. K. Slany, M. Bosl, P. F. Crain, H. Kersten, *Biochemistry* **1993**, *32*, 7811 – 7817.
[44] B. Frey, J. McCloskey, W. Kersten, H. Kersten, *J. Bacteriol.* **1988**, *170*, 2078 – 2082.
[45] Z. D. Miles, R. M. McCarty, G. Molnar, V. Bandarian, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7368 – 7372.
[46] M. Vinayak, C. Pathak, *Biosci. Rep.* **2010**, *30*, 135 – 148.
[47] a) J. M. Skuzeski, L. M. Nichols, R. F. Gesteland, J. F. Atkins, *J. Mol. Biol.* **1991**, *218*, 365 – 373; b) K. Zerfass, H. Beier, *Nucleic Acids Res.* **1992**, *20*, 5911 – 5918.
[48] F. Meier, B. Suter, H. Grosjean, G. Keith, E. Kubli, *EMBO J.* **1985**, *4*, 823 – 827.
[49] M. Bienz, E. Kubli, *Nature* **1981**, *294*, 188 – 190.
[50] V. H. Cowling, *Biochem. J.* **2010**, *425*, 295 – 302.
[51] a) R. Reddy, R. Singh, S. Shimba, *Pharmacol. Ther.* **1992**, *54*, 249 – 267; b) F. Rottman, A. J. Shatkin, R. P. Perry, *Cell* **1974**, *3*, 197 – 199.
[52] S. Farajollahi, S. Maas, *Trends Genet.* **2010**, *26*, 221 – 230.
[53] R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res.* **1998**, *26*, 5017 – 5035.
[54] J. Urbonavicius, Q. Qian, J. M. B. Durand, T. G. Hagervall, G. R. Bjork, *EMBO J.* **2001**, *20*, 4863 – 4873.
[55] a) N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905 – 920; b) F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, *Cell* **2000**, *102*, 615 – 623; c) V. Ramakrishnan, *Angew. Chem.* **2010**, *122*, 4454 – 4481; *Angew. Chem. Int. Ed.* **2010**, *49*, 4355 – 4380.
[56] a) F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186 – 1191; b) L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, *17*, 555 – 560.
[57] a) J. W. Stuart, Z. Gdaniec, R. Guenther, M. Marszalek, E. Sochacka, A. Malkiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13396 – 13404; b) J. W. Stuart, K. M. Koshlap, R. Guenther, P. F. Agris, *J. Mol. Biol.* **2003**, *334*, 901 – 918; c) V. Dao, R. Guenther, A. Malkiewicz, B. Nawrot, E. Sochacka, A. Kraszewski, J. Jankowska, K. Everett, P. F. Agris, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2125 – 2129; d) P. C. Durant, A. C. Bajji, M. Sundaram, R. K. Kumar, D. R. Davis, *Biochemistry* **2005**, *44*, 8078 – 8089.
[58] A. Weixlbaumer, F. V. Murphy, A. Dziergowska, A. Malkiewicz, F. A. P. Vendeix, P. F. Agris, V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498 – 502.
[59] a) J. R. Sampson, O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1033 – 1037; b) K. B. Hall, J. R. Sampson, O. C. Uhlenbeck, A. G. Redfield, *Biochemistry* **1989**, *28*, 5794 – 5801; c) V. Perret, A. Garcia, J. Puglisi, H. Grosjean, J. P. Ebel, C. Florentz, R. Giegé, *Biochimie* **1990**, *72*, 735 – 743; d) P. Davanloo, M. Sprinzl, K. Watanabe, M. Albani, H. Kersten, *Nucleic Acids Res.* **1979**, *6*, 1571 – 1581.
[60] a) A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, E. M. Phizicky, *Mol. Cell* **2006**, *21*, 87 – 96; b) L. Kotelawala, E. J. Grayhack, E. M. Phizicky, *RNA* **2008**, *14*, 158 – 169.
[61] P. C. Durant, D. R. Davis, *J. Mol. Biol.* **1999**, *285*, 115 – 131.
[62] M. L. Wilkinson, S. M. Crary, J. E. Jackman, E. J. Grayhack, E. M. Phizicky, *RNA* **2007**, *13*, 404 – 413.
[63] M. Helm, R. Giegé, C. Florentz, *Biochemistry* **1999**, *38*, 13338 – 13346.
[64] a) E. Madore, C. Florentz, R. Giege, S.-i. Sekine, S. Yokoyama, J. Lapointe, *Eur. J. Biochem.* **1999**, *266*, 1128 – 1135; b) L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Soll, *Biochemistry* **1993**, *32*, 3836 – 3841.
[65] J. Pütz, C. Florentz, F. Benseler, R. Giege, *Nat. Struct. Biol.* **1994**, *1*, 580 – 582.
[66] S. Kiesewetter, G. Ott, M. Sprinzl, *Nucleic Acids Res.* **1990**, *18*, 4677 – 4681.
[67] R. Basavappa, P. B. Sigler, *EMBO J.* **1991**, *10*, 3105 – 3111.
[68] C. Forster, K. Chakraburtty, M. Sprinzl, *Nucleic Acids Res.* **1993**, *21*, 5679 – 5683.
[69] S. U. Åström, A. S. Byström, *Cell* **1994**, *79*, 535 – 546.
[70] V. I. Lim, *J. Mol. Biol.* **1994**, *240*, 8 – 19.
[71] a) E. M. Gustilo, F. A. P. Vendeix, P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134 – 140; b) M. Yarus, *Science* **1982**, *218*, 646 – 652.

[72] a) M. Srinivasan, P. Mehta, Y. Yu, E. Prugar, E. V. Koonin, A. W. Karzai, R. Sternglanz, *EMBO J.* **2011**, *30*, 873 – 881; b) B. El Yacoubi, B. Lyons, Y. Cruz, R. Reddy, B. Nordin, F. Agnelli, J. R. Williamson, P. Schimmel, M. A. Swairjo, V. de Crécy-Lagard, *Nucleic Acids Res.* **2009**, *37*, 2894 – 2909.

[73] a) D. Fu, J. A. N. Brophy, C. T. Y. Chan, K. A. Atmore, U. Begley, R. S. Paules, P. C. Dedon, T. J. Begley, L. D. Samson, *Mol. Cell. Biol.* **2010**, *30*, 2449 – 2459; b) Y. Fu, Q. Dai, W. Zhang, J. Ren, T. Pan, C. He, *Angew. Chem.* **2010**, *122*, 9069 – 9072; *Angew. Chem. Int. Ed.* **2010**, *49*, 8885 – 8888; c) E. van den Born, C. B. Vågbø, L. Songe-Møller, V. Leihne, G. F. Lien, G. Leszczynska, A. Malkiewicz, H. E. Krokan, F. Kirpekar, A. Klungland, P. Ø. Falnes, *Nat. Commun.* **2011**, *2*, 172.

[74] a) J. P. Aris, G. Blobel, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 931 – 935; b) R. P. Jansen, E. C. Hurt, H. Kern, H. Lehtonen, M. Carmo-Fonseca, B. Lapeyre, D. Tollervey, *J. Cell Biol.* **1991**, *113*, 715 – 729.

[75] N. S. Heiss, S. W. Knight, T. J. Vulliamy, S. M. Klauck, S. Wiemann, P. J. Mason, A. Poustka, I. Dokal, *Nat. Genet.* **1998**, *19*, 32 – 38.

[76] X. Darzacq, B. E. Jady, C. Verheggen, A. M. Kiss, E. Bertrand, T. Kiss, *EMBO J.* **2002**, *21*, 2746 – 2756.

[77] K. Nakanishi, L. Bonnefond, S. Kimura, T. Suzuki, R. Ishitani, O. Nureki, *Nature* **2009**, *461*, 1144 – 1148.

[78] D. A. Melton, E. M. De Robertis, R. Cortese, *Nature* **1980**, *284*, 143 – 148.

[79] a) M. Steiner-Mosonyi, D. Mangroo, *Biochem. J.* **2004**, *378*, 809 – 816; b) E. Lund, J. E. Dahlberg, *Science* **1998**, *282*, 2082 – 2085.

[80] a) W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, E. K. O'Shea, *Nature* **2003**, *425*, 686 – 691; b) T. Yoshihisa, K. Yunoki-Esaki, C. Ohshima, N. Tanaka, T. Endo, *Mol. Biol. Cell* **2003**, *14*, 3266 – 3279.

[81] a) A. Takano, T. Endo, T. Yoshihisa, *Science* **2005**, *309*, 140 – 142; b) H. H. Shaheen, A. K. Hopper, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11290 – 11295.

[82] A. K. Hopper, H. H. Shaheen, *Trends Cell Biol.* **2008**, *18*, 98 – 104.

[83] M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, *Science* **2003**, *302*, 1399 – 1401.

[84] E. Bertrand, F. Houser-Scott, A. Kendall, R. H. Singer, D. R. Engelke, *Genes Dev.* **1998**, *12*, 2463 – 2468.

[85] T. Ohira, T. Suzuki, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10502 – 10507.

[86] H.-Q. Jiang, Y. Motorin, Y.-X. Jin, H. Grosjean, *Nucleic Acids Res.* **1997**, *25*, 2694 – 2701.

[87] G.-J. Arts, S. Kuersten, P. Romby, B. Ehresmann, I. W. Mattaj, *EMBO J.* **1998**, *17*, 7430 – 7441.

[88] S. Kadaba, A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch, J. Anderson, *Genes Dev.* **2004**, *18*, 1227 – 1240.

[89] M. L. Whitney, R. L. Hurto, H. H. Shaheen, A. K. Hopper, *Mol. Biol. Cell* **2007**, *18*, 2678 – 2686.

[90] A. Ghavidel, T. Kislinger, O. Pogoutse, R. Sopko, I. Jurisica, A. Emili, *Cell* **2007**, *131*, 915 – 926.

[91] a) Y. Suzuki, A. Noma, T. Suzuki, M. Senda, T. Senda, R. Ishitani, O. Nureki, *J. Mol. Biol.* **2007**, *372*, 1204 – 1214; b) Y. Suzuki, A. Noma, T. Suzuki, R. Ishitani, O. Nureki, *Nucleic Acids Res.* **2009**, *37*, 2910 – 2925; c) M. Kato, Y. Araiso, A. Noma, A. Nagao, T. Suzuki, R. Ishitani, O. Nureki, *Nucleic Acids Res.* **2011**, *39*, 1576 – 1585; d) M. Umitsu, H. Nishimasu, A. Noma, T. Suzuki, R. Ishitani, O. Nureki, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15616 – 15621.

[92] a) L. Droogmans, H. Grosjean, *EMBO J.* **1987**, *6*, 477 – 483; b) G. R. Bjork, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Bystrom, O. P. Persson, *EMBO J.* **2001**, *20*, 231 – 239.

[93] A. P. Young, V. Bandarian, *Biochemistry* **2011**, *50*, 10573 – 10575.

[94] a) W. F. Waas, V. de Crécy-Lagard, P. Schimmel, *J. Biol. Chem.* **2005**, *280*, 37616 – 37622; b) A. Noma, Y. Kirino, Y. Ikeuchi, T. Suzuki, *EMBO J.* **2006**, *25*, 2142 – 2154.

[95] H. R. Kalhor, M. Penjwini, S. Clarke, *Biochem. Biophys. Res. Commun.* **2005**, *334*, 433 – 440.

[96] A. Noma, R. Ishitani, M. Kato, A. Nagao, O. Nureki, T. Suzuki, *J. Biol. Chem.* **2010**, *285*, 34503 – 34507.

[97] N. R. Rose, M. A. McDonough, O. N. F. King, A. Kawamura, C. J. Schofield, *Chem. Soc. Rev.* **2011**, *40*, 4364 – 4397.

[98] J. R. Wagner, J. Cadet, *Acc. Chem. Res.* **2010**, *43*, 564 – 571.

[99] J. S. Mattick, I. V. Makunin, *Hum. Mol. Genet.* **2006**, *15*, R17 – 29.

[100] B. Yu, Z. Yang, J. Li, S. Minakhina, M. Yang, R. W. Padgett, R. Steward, *Science* **2005**, *307*, 932 – 935.

[101] J. Li, Z. Yang, B. Yu, J. Liu, X. Chen, *Curr. Biol.* **2005**, *15*, 1501 – 1507.

[102] a) Z. Yang, Y. W. Ebright, B. Yu, X. Chen, *Nucleic Acids Res.* **2006**, *34*, 667 – 675; b) Y. Zhao, Y. Yu, J. Zhai, V. Ramachandran, T. T. Dinh, et al., *Curr. Biol.* **2012**, *22*, 689 – 694.

[103] a) T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, *Nat. Struct. Mol. Biol.* **2007**, *14*, 349 – 350; b) Y. Kirino, Z. Mourelatos, *RNA* **2007**, *13*, 1397 – 1401.

[104] a) L. M. Kamminga, M. J. Luteijn, M. J. den Broeder, S. Redl, L. J. T. Kaaij, E. F. Roovers, P. Ladurner, E. Berezikov, R. F. Ketting, *EMBO J.* **2010**, *29*, 3688 – 3700; b) S. Houwing, L. M. Kamminga, E. Berezikov, D. Cronembold, A. Girard, H. van den Elst, D. V. Flippov, H. Blaser, E. Raz, C. B. Moens, R. H. A. Plasterk, G. J. Hannon, B. W. Draper, R. F. Ketting, *Cell* **2007**, *129*, 69 – 82.

[105] M. D. Horwich, C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, P. D. Zamore, *Curr. Biol.* **2007**, *17*, 1265 – 1272.

[106] C. Lacoux, D. Di Marino, P. Pilo Boyl, F. Zalfa, B. Yan, M. T. Ciotti, M. Falconi, H. Urlaub, T. Achsel, A. Mougin, M. Caizergues-Ferrer, C. Bagni, *Nucleic Acids Res.* **2012**, *40*, 4086 – 4096.

[107] Y.-K. Kim, I. Heo, V. N. Kim, *Cell* **2010**, *143*, 703 – 709.

[108] a) W. Yang, T. P. Chendrimada, Q. Wang, M. Higuchi, P. H. Seeburg, R. Shiekhattar, K. Nishikura, *Nat. Struct. Mol. Biol.* **2006**, *13*, 13 – 21; b) Y. Kawahara, B. Zinshteyn, T. P. Chendrimada, R. Shiekhattar, K. Nishikura, *EMBO Rep.* **2007**, *8*, 763 – 768.

[109] J. Karijolich, Y.-T. Yu, *RNA Biol.* **2010**, *7*, 192 – 204.

[110] N. K. Kim, C. A. Theimer, J. R. Mitchell, K. Collins, J. Feigon, *Nucleic Acids Res.* **2010**, *38*, 6746 – 6756.

[111] X. Zhao, J. R. Patton, S. L. Davis, B. Florence, S. J. Ames, R. A. Spanjaard, *Mol. Cell* **2004**, *15*, 549 – 558.

[112] a) M. Buck, M. Connick, B. N. Ames, *Anal. Biochem.* **1983**, *129*, 1 – 13; b) P. F. Crain, *Mass Spectrom. Rev.* **1990**, *9*, 505 – 554; c) J. A. McCloskey, *Acc. Chem. Res.* **1991**, *24*, 81 – 88; d) J. Ofengand, M. Del Campo, Y. Kaya, *Methods* **2001**, *25*, 365 – 373; e) I. Behm-Ansmant, M. Helm, Y. Motorin, *J. Nucleic Acids* **2011**, *2011*, 408053; f) S. Kellner, J. Burhenne, M. Helm, *RNA Biol.* **2010**, *7*, 237 – 247.

[113] Y. Motorin, S. Muller, I. Behm-Ansmant, C. Branlant, *Methods Enzymol.* **2007**, *425*, 21 – 53.

[114] M. Sakurai, T. Yano, H. Kawabata, H. Ueda, T. Suzuki, *Nat. Chem. Biol.* **2010**, *6*, 733 – 740.

[115] J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, *Nucleic Acids Res.* **2012**, *40*, 5023 – 5033.

[116] A. Durairaj, P. A. Limbach, *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3727 – 3734.

[117] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, et al., *Nature* **2012**, *485*, 201 – 206.

[118] S. L. Hiley, J. Jackman, T. Babak, M. Trochesset, Q. D. Morris, E. Phizicky, T. R. Hughes, *Nucleic Acids Res.* **2005**, *33*, e2.
[119] K. Miyauchi, T. Ohara, T. Suzuki, *Nucleic Acids Res.* **2007**, *35*, e24.
[120] M. Saikia, Y. Fu, M. Pavon-Eternod, C. He, T. Pan, *RNA* **2010**, *16*, 1317 – 1327.
[121] a) C. W. Gehrke, K. C. Kuo, *J. Chromatogr.* **1989**, *471*, 3 – 36; b) J. D. Johnson, J. Horowitz, *Biochim. Biophys. Acta* **1971**, *247*, 262 – 279; c) X. L. Zhao, Y. T. Yu, *RNA* **2004**, *10*, 996 – 1002.
[122] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074 – 8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932 – 7934; b) J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 3242 – 3245.
[123] a) B. Addepalli, P. A. Limbach, *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1363 – 1372; b) C. T. Y. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, *PLoS Genet.* **2010**, *6*, e1001247.
[124] T. Brückl, PhD thesis, Ludwig-Maximilians-Universität, München, Germany, **2010**.
[125] D. Globisch, D. Pearson, A. Hienzsch, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem.* **2011**, *123*, 9913 – 9916; *Angew. Chem. Int. Ed.* **2011**, *50*, 9739 – 9742.
[126] T. Suzuki, T. Wada, K. Saigo, K. Watanabe, *EMBO J.* **2002**, *21*, 6581 – 6589.
[127] Y. Kirino, T. Yasukawa, S. Ohta, S. Akira, K. Ishihara, K. Watanabe, T. Suzuki, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15070 – 15075.
[128] a) G. Dirheimer, W. Baranowski, G. Keith, *Biochimie* **1995**, *77*, 99 – 103; b) C. Pathak, Y. K. Jaiswal, M. Vinayak, *RNA Biol.* **2005**, *2*, 143 – 148.
[129] a) S. Ishiwata, J. Katayama, H. Shindo, Y. Ozawa, K. Itoh, M. Mizugaki, *J. Biochem.* **2001**, *129*, 13 – 17; b) S. Ishiwata, Y. Ozawa, J. Katayama, S. Kaneko, H. Shindo, Y. Tomioka, T. Ishiwata, G. Asano, S. Ikegawa, M. Mizugaki, *Cancer Lett.* **2004**, *212*, 113 – 119.
[130] A. Gallo, F. Locatelli, *Biol. Rev. Cambridge Philos. Soc.* **2012**, *87*, 95 – 110.
[131] a) K. A. Fawcett, I. Barroso, *Trends Genet.* **2010**, *26*, 266 – 274; b) J. R. Perry, T. M. Frayling, *Curr. Opin. Clin. Nutr. Metab. Care* **2008**, *11*, 371 – 377.
[132] a) S. Arragain, S. K. Handelman, F. Forouhar, F.-Y. Wei, K. Tomizawa, J. F. Hunt, T. Douki, M. Fontecave, E. Mulliez, M. Atta, *J. Biol. Chem.* **2010**, *285*, 28425 – 28433; b) F.-Y. Wei, T. Suzuki, S. Watanabe, S. Kimura, T. Kaitsuka, A. Fujimura, H. Matsui, M. Atta, H. Michiue, M. Fontecave, K. Yamagata, T. Suzuki, K. Tomizawa, *J. Clin. Invest.* **2011**, *121*, 3598 – 3608.
[133] T. Gerken, C. A. Girard, Y. C. Tung, C. J. Webby, V. Saudek, K. S. Hewitson, G. S. H. Yeo, M. A. McDonough, S. Cunliffe, L. A. McNeill, J. Galvanovskis, P. Rorsman, P. Robins, X. Prieur, A. P. Coll, M. Ma, Z. Jovanovic, I. S. Farooqi, B. Sedgwick, I. Barroso, T. Lindahl, C. P. Ponting, F. M. Ashcroft, S. O'Rahilly, C. J. Schofield, *Science* **2007**, *318*, 1469 – 1472.
[134] G. Jia, C. G. Yang, S. Yang, X. Jian, C. Yi, Z. Zhou, C. He, *FEBS Lett.* **2008**, *582*, 3313 – 3319.
[135] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885 – 887.
[136] Y. P. Ching, A. S. Pang, W. H. Lam, R. Z. Qi, J. H. Wang, *J. Biol. Chem.* **2002**, *277*, 15237 – 15240.
[137] F.-Y. Wei, K. Nagashima, T. Ohshima, Y. Saheki, Y.-F. Lu, M. Matsushita, Y. Yamada, K. Mikoshiba, Y. Seino, H. Matsui, K. Tomizawa, *Nat. Med.* **2005**, *11*, 1104 – 1108.
[138] a) K. H. Kaminska, U. Baraniak, M. Boniecki, K. Nowaczyk, A. Czerwoniec, J. M. Bujnicki, *Proteins* **2008**, *70*, 1 – 18; b) V. Reiter, D. M. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, DOI: 10.1093/nar/gks240.
[139] a) P. I. Moreira, S. M. Cardoso, C. M. Pereira, M. S. Santos, C. R. Oliveira, *CNS Neurol. Disord. Drug Targets* **2009**, *8*, 492 – 511; b) P. I. Moreira, A. I. Duarte, M. S. Santos, A. C. Rego, C. R. Oliveira, *J. Alzheimer's Dis.* **2009**, *16*, 741 – 761.
[140] N. Okada, S. Noguchi, S. Nishimura, T. Ohgi, T. Goto, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1978**, *5*, 2289 – 2296.
[141] B. S. Huang, R. T. Wu, K. Y. Chien, *Cancer Res.* **1992**, *52*, 4696 – 4700.
[142] a) U. Grädler, H. D. Gerber, D. M. Goodenough-Lashua, G. A. Garcia, R. Ficner, K. Reuter, M. T. Stubbs, G. Klebe, *J. Mol. Biol.* **2001**, *306*, 455 – 467; b) J. K. Hurt, S. Olgen, G. A. Garcia, *Nucleic Acids Res.* **2007**, *35*, 4905 – 4913.
[143] a) S. R. Hörtner, T. Ritschel, B. Stengl, C. Kramer, W. B. Schweizer, B. Wagner, M. Kansy, G. Klebe, F. Diederich, *Angew. Chem.* **2007**, *119*, 8414 – 8417; *Angew. Chem. Int. Ed.* **2007**, *46*, 8266 – 8269; b) B. Stengl, E. A. Meyer, A. Heine, R. Brenk, F. Diederich, G. Klebe, *J. Mol. Biol.* **2007**, *370*, 492 – 511.
[144] a) S. Gamagedara, Y. Ma, *Bioanalysis* **2011**, *3*, 2129 – 2142; b) W.-Y. Hsu, W.-D. Lin, Y. Tsai, C.-T. Lin, H.-C. Wang, L.-B. Jeng, C.-C. Lee, Y.-C. Lin, C.-C. Lai, F.-J. Tsai, *Clin. Chim. Acta* **2011**, *412*, 1861 – 1866; c) A. Seidel, S. Brunner, P. Seidel, G. I. Fritz, O. Herbarth, *Br. J. Cancer* **2006**, *94*, 1726 – 1733.
[145] a) E. Friedberg, G. Walker, W. Siede, R. Wood, R. Schultz, T. Ellenberger, *DNA Repair and Mutagenesis*, *2nd ed.*, ASM, Washington, **2006**; b) M. P. Stone, H. Huang, K. L. Brown, G. Shanmugam, *Chem. Biodiversity* **2011**, *8*, 1571 – 1615.
[146] D. Ratel, J. L. Ravanat, F. Berger, D. Wion, *Bioessays* **2006**, *28*, 309 – 315.
[147] B. R. Palmer, M. G. Marinus, *Gene* **1994**, *143*, 1 – 12.
[148] T. B. Johnson, R. D. Coghill, *J. Am. Chem. Soc.* **1925**, *47*, 2838 – 2844.
[149] G. R. Wyatt, *Nature* **1950**, *166*, 237 – 238.
[150] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935.
[151] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781 – 790.
[152] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503 – 5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375 – 5377.
[153] a) T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, DOI: 10.1002/ijc.27429; b) H. Yang, Y. Liu, F. Bai, J. Y. Zhang, S. H. Ma, J. Liu, Z. D. Xu, H. G. Zhu, Z. Q. Ling, D. Ye, K. L. Guan, Y. Xiong, *Oncogene* **2012**, DOI: 10.1038/onc.2012.67.
[154] K. I. Ladwein, M. Jung, *Angew. Chem.* **2011**, *123*, 12347 – 12349; *Angew. Chem. Int. Ed.* **2011**, *50*, 12143 – 12145.
[155] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367.
[156] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146 – 7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012.
[157] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303; b) Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, *333*, 1303 – 1307.
[158] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, *8*, 328 – 330.

[159] a) H. Cedar, Y. Bergman, *Nat. Rev. Genet.* **2009**, *10*, 295 – 304; b) M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465 – 476; c) H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129 – 140.
[160] M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **2012**, *13*, 7 – 13.
[161] S. K. Ooi, T. H. Bestor, *Cell* **2008**, *133*, 1145 – 1148.
[162] a) C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101 – 1126; b) M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Schöler, J. Walter, *EMBO J.* **2010**, *29*, 1877 – 1888; c) P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78 – 82; d) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Sorprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67 – 79.
[163] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334 – 35338.
[164] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, *145*, 423 – 434.
[165] a) A. Radzicka, R. Wolfenden, *Science* **1995**, *267*, 90 – 93; b) J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, Sir, A. A. Axhemi, *Biochim. Biophys. Acta Gen. Subj.* **2005**, *1723*, 256 – 264.
[166] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem.* **2012**, DOI: 10.1002/ange.201202583; *Angew. Chem. Int. Ed.* **2012**, DOI: 10.1002/anie.201202583.
[167] T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 2543 – 2545.
[168] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368 – 1380.
[169] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934 – 937.