RNA Modifications Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2023,** *62,* e202217128
International Edition: doi.org/10.1002/anie.202217128
German Edition: doi.org/10.1002/ange.202217128

# A New Bacterial Adenosine-Derived Nucleoside as an Example of RNA Modification Damage

*Larissa Bessler, Lea-Marie Vogt, Marc Lander, Christina Dal Magro, Patrick Keller, Jonas Kühlborn, Christopher J. Kampf, Till Opatz, and Mark Helm**

**Abstract:** The fields of RNA modification and RNA damage both exhibit a plethora of non-canonical nucleoside structures. While RNA modifications have evolved to improve RNA function, the term RNA damage implies detrimental effects. Based on stable isotope labelling and mass spectrometry, we report the identification and characterisation of 2-methylthio-1,*N*6-ethenoadenosine ($ms^2\varepsilon A$), which is related to 1,*N*6-ethenoadenine, a lesion resulting from exposure of nucleic acids to alkylating chemicals in vivo. In contrast, a sophisticated isoprene labelling scheme revealed that $ms^2\varepsilon A$ biogenesis involves cleavage of a prenyl moiety in the known transfer RNA (tRNA) modification 2-methylthio-*N*6-isopentenyladenosine ($ms^2i^6A$). The relative abundance of $ms^2\varepsilon A$ in tRNAs from translating ribosomes suggests reduced function in comparison to its parent RNA modification, establishing the nature of the new structure in a newly perceived overlap of the two previously separate fields, namely an RNA modification damage.

The ensemble of post transcriptional modifications of cellular RNA, dubbed the epitranscriptome, features a large chemical variety. About 170 distinct chemical structures are known from various species,[1] and new structures emerge at high frequency.[2–4] Among RNA species, tRNA modifications contribute most to the chemical diversity and newly discovered structures. Increasingly sophisticated instrumentation has driven the discovery of new species occurring at substoichiometric quantities. Yet further non-canonical nucleoside structures are known from the field of nucleic acid damage, although many more of these structures are being intensively scrutinised in DNA than they are in RNA. Most of these structures derive from radiation, oxidative or alkylating agents.[5,6] Rather than investigating damage to canonical nucleosides, a few recent reports focus on irradiation[7] and alkylating damage[3,8] to post-transcriptional modifications.

Given that tRNAs are the most highly modified RNA species, they are also the most probable place for such "RNA modification damage" to occur. Considering the high functional importance of certain RNA modifications, e.g. in the tRNA anticodon, their damage might arguably cause more severe harm than random damage to canonical nucleosides. Consequently, one might suspect the existence of repair mechanisms akin to those described for DNA and RNA.[9] Indeed, 2-methylthiocytidine ($ms^2C$), an alkylating damage to 2-thiocytidine in *Escherichia coli* (*E. coli*) was shown to be reverted by AlkB,[3] an enzyme from a family with reported function in RNA and DNA repair.[6,10] On the other hand, in the case of 4-methylthiouridine ($ms^4U$), repair was AlkB-independent.[8]

While a biological origin of alkylating agents was traced to lipid peroxidation,[11] nucleic acid alkylation from human-made agents such as vinyl halides was extensively studied early on in vitro and in vivo.[12–14] A relevant point in case was investigated upon emergence of unusual liver tumours in the 1970s. These occurred in workers from the polymer industry, with a causal connection to exposure to vinyl chloride.[15] Vinyl chloride was shown to be metabolised to 2-chloroacetaldehyde (CAA),[16] a highly mutagenic reagent that reacts with nucleobases, resulting in the formation of etheno (ε)-adducts such as 1,*N*6-ethenoadenine in DNA[12,17] and in RNA.[13,18] We here report the detection and structural elucidation of a new ε-adenosine, namely $ms^2\varepsilon A$, in bacterial tRNA. Surprisingly, the etheno fragment was not the result of alkylating damage, but oxidatively derived from the isoprene moiety in $ms^2i^6A$. An analysis of the distribution between the cellular tRNA pool and tRNAs on actively translating polysomes indicated decreased functionality, causing us to classify $ms^2\varepsilon A$ as RNA modification damage.

We previously reported a list of unknown ribonucleosides consisting of potentially new RNA modifications, that were identified by combination of various stable isotope labelling experiments with a so-called neutral loss scan (NLS) by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a triple quadrupole (QQQ) mass spectrometer.[2] One elucidated structure was $msms^2i^6A$ (2-methylthiomethylenethio-*N*6-isopentenyladenosine), a

[*] L. Bessler, L.-M. Vogt, M. Lander, Dr. C. Dal Magro, Dr. P. Keller, Prof. Dr. M. Helm
Institute of Pharmaceutical and Biomedical Sciences (IPBS), Johannes Gutenberg University Mainz
Staudingerweg 5, 55128 Mainz (Germany)
E-mail: mhelm@uni-mainz.de
Dr. J. Kühlborn, Dr. C. J. Kampf, Prof. Dr. T. Opatz
Department of Chemistry, Johannes Gutenberg University Mainz
Duesbergweg 10–14, 55128 Mainz (Germany)

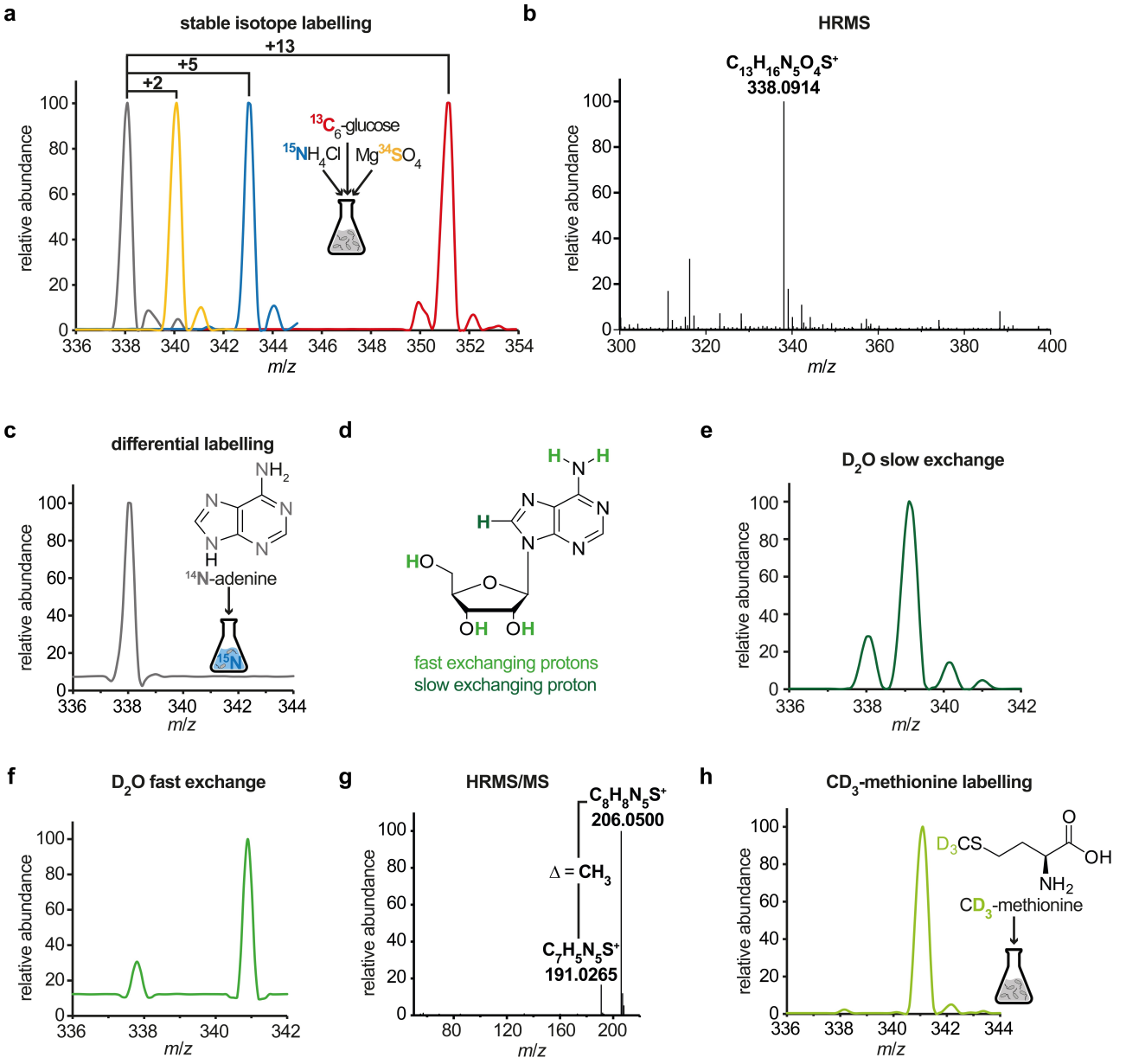


***Figure 1.*** Structural characterisation of candidate 338 by LC-MS/MS. Unless otherwise stated, abundances were set in relation to the highest signal in the respective mass spectrum (relative abundance). a) Overlay of QQQ mass spectra at 27.7 min corresponding to candidate 338 from a NLS of digested labelled ($^{13}C$ = red, $^{15}N$ = blue, $^{34}S$ = yellow) and unlabelled (grey) *E. coli* tRNA. Mass shifts indicating the number of respective atoms within the molecule are assigned. b) Mass spectrum and ion formula of candidate 338, calculated from the exact mass detected by HRMS analysis of digested tRNA enriched for candidate 338. c) QQQ NLS mass spectrum of tRNA isolated from a $^{15}N$-labelled *E. coli* culture supplemented with $^{14}N$-adenine at 27.7 min. d) Overview of slow (dark green) and fast (light green) exchanging hydrogen atoms in adenosine. e) QQQ NLS mass spectrum after incubation of RNA samples with $D_2O$ (slow exchange). f) QQQ NLS mass spectrum after the use of a $D_2O$ LC-buffer (fast exchange). g) HRMS fragmentation pattern of candidate 338, as well as calculated ion formulas of the generated fragments which differ by a methyl group. h) QQQ NLS mass spectrum of candidate 338 in tRNA isolated from *E. coli* cells supplemented with $CD_3$-methionine.

highly lipophilic derivative of the hypermodification $ms^2i^6A$ found at position 37 of certain bacterial tRNAs. The structural elucidation established "inverse" feeding experiments as highly useful, where bacteria were raised in "heavy" media containing e.g. $^{13}C$-glucose as only carbon source. Incorporation into RNA of any metabolite consisting of "light" $^{12}C$-carbon added to the culture was thereby easily detectable as a mass difference corresponding to the number of carbons incorporated. While application to nucleobases was uncomplicated, compounds that were not naturally taken up from the medium included prenylation reagents and precursors. Thus, while we could indirectly demonstrate the presence of an isopentenyl residue by mass spectrometry and the use of the knockout strain of the prenyltransferase MiaA, there was no direct way to monitor prenylation. This was of interest in the context of another

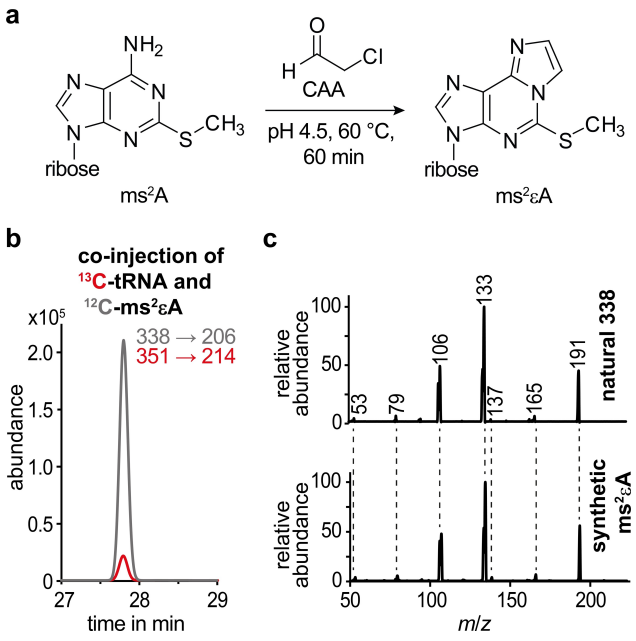


***Figure 2.*** Structural confirmation of ms$^2$εA by LC-MS/MS analysis of the synthesised compound. a) Reaction scheme displaying the synthesis of ms$^2$εA, starting from 2-methylthioadenosine (ms$^2$A). b) QQQ LC-MS/MS analysis of a digested $^{13}$C-labelled tRNA sample (red) spiked with the synthetic $^{12}$C-compound (grey). An overlay of the extracted ion chromatograms at a mass transition of *m/z* 338→206 and *m/z* 351→214 is shown. c) Comparison of fragmentation patterns of candidate 338, recorded from digested native tRNA (upper panel), and the synthesised reference ms$^2$εA (lower panel) in a QQQ pseudo-MS$^3$ scan.

candidate nucleoside species from the list, which displayed a mass-to-charge ratio of 338 (herein referred to as candidate 338). It eluted at 27.7 min, significantly after *N*6,*N*6-dimethyladenosine (m$^{6,6}$A) but before *N*6-isopentenyladenosine (i$^6$A), from the reversed-phase column. Stable isotope labelling revealed that this compound was elementally constituted of 13 carbon atoms, 5 nitrogen atoms and one sulphur atom which, in combination with the results from high-resolution mass spectrometry (HRMS), suggested a sum formula of $C_{13}H_{15}N_5O_4S$ (Figure 1a+b). Inverse feeding identified adenosine as parent nucleoside prior to modification (Figure 1c).

Comparison of adenosine to the constitution of candidate 338 showed a difference of $C_3H_2S$ and two double-bond equivalents. Given that the initial NLS had provided the fragmentation pattern of an unsubstituted ribose, the above formula implied that at least some of the nitrogens of the parental nucleoside carried substituents for the original hydrogens. To locate the modification sites, we performed slow and fast deuterium exchange experiments, which provide insights into the substitution of the C8-position and the substitution of heteroatoms, respectively (Figure 1d).[2,19] Interestingly, while the 8-position was suggested to be unmodified (Figure 1e), all nitrogens bound to fast exchanging hydrogen atoms in the parent adenosine were suggested to carry substituents in candidate 338 (Figure 1f). HRMS fragmentation experiments indicated the loss of a methyl group after loss of the ribose (Figure 1g). The presence of a methyl group was further substantiated by a corresponding +3 *m/z* shift in a $CD_3$-methionine labelling experiment (Figure 1h).

Given the presence of a sulphur atom in conjunction with a methyl group in a modified adenosine structure, we suspected a thiomethylation in position 2 (as for example in ms$^2$i$^6$A), which would typically be catalysed by the iron-sulphur enzyme MiaB. The absence of candidate 338 in the MiaB knockout established a thiomethyl moiety in the adenine-C2 position (Figure 1g and S1). However, this only left two further carbon atoms to be assigned to the structure of candidate 338. Considering the double-bond equivalents and the fact that none of the nitrogen atoms was bound to fast exchanging protons, the most plausible structure proposal contained an additional ring structure, i.e. an etheno bridge linking the exocyclic *N*6 with *N*1 of the nucleobase. A theoretically conceivable triple bond at the exocyclic *N*6-position would have been incompatible with the instability of *N*-ethynylamines[20] and the $D_2O$ fast exchange results (Figure 1f). With only one plausible structure left, we synthesised ms$^2$εA, adapting a protocol from Bhatt et al. (Figure 2a).[21] LC-MS/MS analysis of the $^{13}$C-labelled *E. coli* tRNA digest, spiked with the synthetic compound, showed co-elution of the synthetic compound and its isotopologue, confirming the hypothetical structure (Figure 2b). Ultimately, the perfectly matching pseudo-MS$^3$ spectra of candidate 338, enriched from digested *E. coli* tRNA, and the synthesised ms$^2$εA demonstrated identity of the molecular structures of the naturally occurring candidate 338 and the synthesised compound ms$^2$εA (Figure 2c).

The unusual tricyclic structure of ms$^2$εA evoked the question on the metabolic origin of the two carbon atoms of the etheno bridge. Based on the involvement of MiaB and the metabolic as well as structural proximity to ms$^2$i$^6$A, we speculated that this bridge might be the vestigial relic of an originally present *N*6-isopentenyl moiety. We hence tested RNA isolated from a knockout of the MiaA enzyme. The latter catalyses RNA prenylation of the *N*6-position of an adenosine, resulting in i$^6$A, using dimethylallyl pyrophosphate (DMAPP), an isoprene cofactor that is provided via the methylerythritol-phosphate (MEP) pathway. We did not detect any ms$^2$εA in the MiaA knockout (Figure S1), suggesting that the carbon atoms of the etheno bridge might indeed originate from the MEP pathway. Formation of ms$^2$εA would involve ms$^2$i$^6$A as a hypothetical precursor, with its isoprene moiety being the carbon source of the etheno bridge (Figure 3a). To verify this hypothesis, we isolated 30 tRNA species and found ms$^2$εA only in those 9 tRNAs known to carry ms$^2$i$^6$A (Figure S2).

For direct clarification, we needed to label the isoprene moiety in i$^6$A/ms$^2$i$^6$A by supplying a "light" C5 compound (i.e. composed of $^{12}$C) to the isoprene pathway in an inverse feeding experiment (see above). To circumvent uptake problems of the negatively charged isoprene cofactors DMAPP and isopentenyl pyrophosphate by *E. coli*, we turned to a metabolic engineering approach that required inhibiting the native hemiterpene biosynthesis pathway by addition of fosmidomycin. To supply a "light" C5 surrogate,

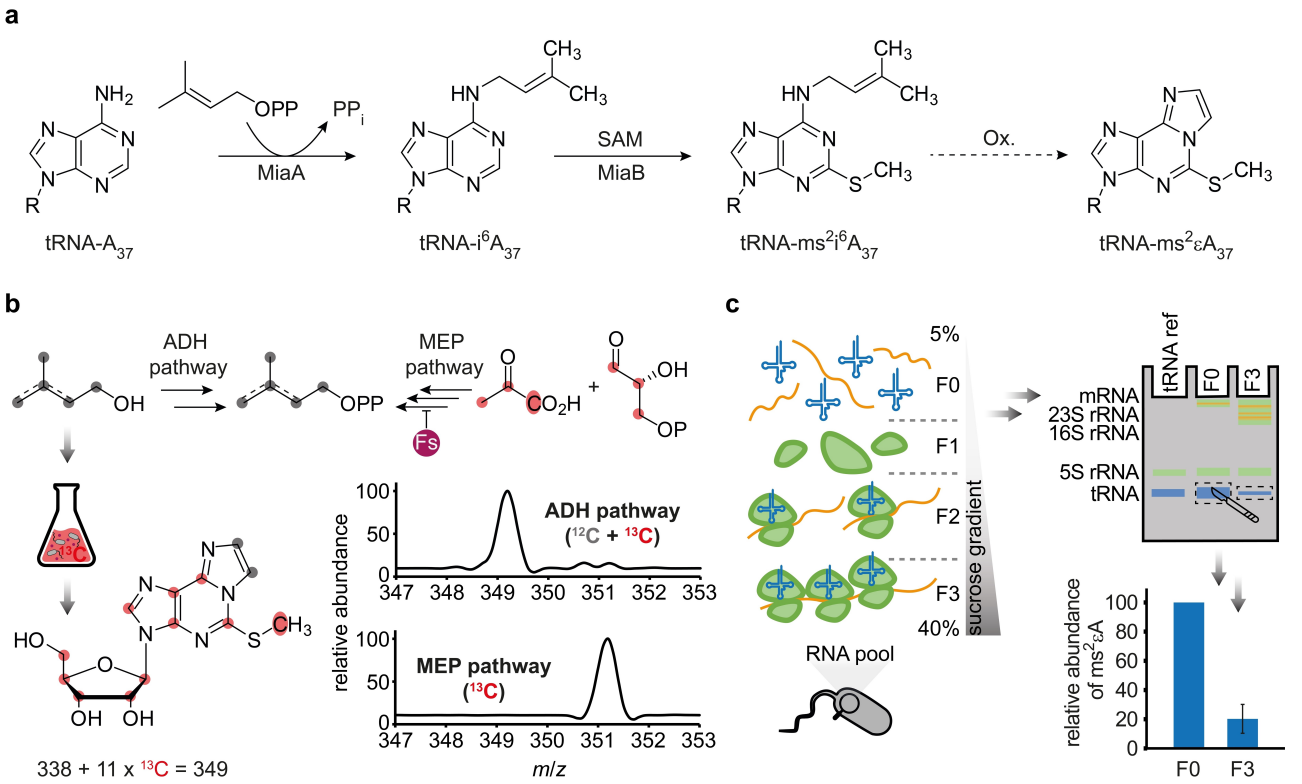


***Figure 3.*** Biological origin and impact of ms$^2$εA. a) Biosynthesis of ms$^2$i$^6$A and proposed generation of ms$^2$εA. b) Inverse isoprene labelling experiment and mass spectra of subsequent QQQ NLS analysis: red ($^{13}$C) and grey ($^{12}$C) markers in the structure formula illustrate the mass-to-charge ratio upon inverse feeding (ADH pathway). In the untreated control sample, the native MEP pathway leads to the completely $^{13}$C-labelled structure. Fs: fosmidomycin, PP$_i$: pyrophosphate, SAM: *S*-adenosylmethionine. c) Constitution of the cellular RNA pool with its most abundant representatives tRNA, messenger RNA (mRNA) and ribosomal RNA (rRNA) and their distribution upon sucrose gradient fractionation. Relative LC-MS/MS quantification of ms$^2$εA in tRNA purified from F0 corresponding to the cellular tRNA pool and F3 corresponding to actively translating tRNAs via gel elution. Normalised peak areas were related pairwise to the respective F0 fraction (set to 100%) and the average of three independent biological replicates was calculated.

we adapted a protocol to introduce an artificial alcohol-dependent hemiterpene (ADH) pathway for isoprenoid biosynthesis into *E. coli* and combined it with stable isotope labelling.[22] Bacteria cultured in $^{13}$C-labelled medium were supplemented with the bioavailable $^{12}$C-alcohol precursors dimethylallyl alcohol and isopentenol (detailed description in Supporting Information 1.1.3). LC-MS/MS analysis of prenylated nucleosides from isolated tRNA was then conducted comparing a "heavy" $^{13}$C-culture with a similar culture supplemented with the $^{12}$C-alcohols. A mass difference in the signals of i$^6$A and ms$^2$i$^6$A indicated incorporation of the C5 alcohols into the hemiterpene pathway (Figure S3). Importantly, a mass shift of −2 units (*m/z* 349 versus *m/z* 351) was observed for the ms$^2$εA signal (Figure 3b). This shows that the etheno bridge originates from the MEP pathway, and thereby that ms$^2$εA is derived from ms$^2$i$^6$A.

Synthetic ms$^2$εA (Figure 2) as calibration standard permitted absolute quantification e.g. in purified tRNA$^{Leu}_{(BAA)}$, showing that only 0.06% of these tRNA molecules contained ms$^2$εA. This low abundance might signify that the presence of ms$^2$εA in tRNAs is a spurious side product that might be non-functional or even detrimental to the cell. We investigated this hypothesis by comparing the ms$^2$εA content of the cellular tRNA pool with that of tRNAs from actively translating polysomes, which are considered functional "by definition". The observed significant reduction of ms$^2$εA in polysomal tRNA relative to free tRNA (Figure 3c), was interpreted to mean that tRNAs carrying this derivative of ms$^2$i$^6$A cannot fully perform their designated function and are excluded from translation by yet unknown mechanisms. This makes ms$^2$εA an RNA modification damage of ms$^2$i$^6$A.

The new nucleoside structure elucidated here is already the second one derived from ms$^2$i$^6$A.[2] Previous identification of msms$^2$i$^6$A demonstrated a presumed side reaction of the iron-sulphur cluster enzyme MiaB, involving radical chemistry. In the formation of ms$^2$εA from ms$^2$i$^6$A, the type of chemistry catalysed is truly remarkable, involving oxidative cleavage of the isoprene double bond in analogy to an ozonolysis. We are currently pursuing identification of the responsible enzyme, an endeavour that is hampered by the low abundance of the structures involved and by the experimental challenge of isotopic labelling of the isoprene moieties in RNA. In the present case, this was achieved after a significant effort devoted to funnelling exogenous C5 alcohols into bacterial hemiterpene synthesis. This approach, adopted from the work of Lund et al.,[22] allowed to show that the etheno bridge within the third heterocycle derives

from the isoprene moiety in $ms^2i^6A$. While successful in the case at hand, its application is sophisticated in practice and in need of detailed optimisation before becoming a tool for routine investigations. Most biological implications of our findings remain to be discovered, but the nucleoside quantification data of polysome-bound tRNA are strongly suggestive of damage to tRNA functionality and a detrimental character of $ms^2\varepsilon A$ and warrant its designation as RNA damage.

While older examples of damaging reactions on tRNA modifications have received the low attention of isolated findings,[23] we conclude that such events are indeed more frequent than hitherto assumed. The proposed term "RNA modification damage" would not only cover structural alteration of intact RNA modifications as a consequence of exposure to chemical or physical stress,[3,7,8] but also the enzymatic side reactions leading to dysfunctional modified nucleosides. Whether such damage is primarily caused by promiscuity of RNA modification enzymes, as it is the case for $msms^2i^6A$, is not clear for $ms^2\varepsilon A$.

## Acknowledgements

The authors would like to thank Dr. Annika Kotter for helpful discussions and Felix Green for technical assistance. This work was funded by the DFG (HE 3397/14-2 in SPP1784 and TP C03 in TRR 319; Project-ID 439669440) to M.H. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

Mark Helm is a consultant for Moderna Inc.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Keywords:** Epitranscriptome · Isotopic Labeling · Mass Spectrometry · Nucleoside Analysis · RNA Modification Damage

[1] P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, E. Purta, M. Kurkowska, N. Shirvanizadeh, E. Destefanis, P. Groza, G. Avşar, A. Romitelli, P. Pir, E. Dassi, S. G. Conticello, F. Aguilo, J. M. Bujnicki, *Nucleic Acids Res.* **2022**, *50*, D231–D235.

[2] C. Dal Magro, P. Keller, A. Kotter, S. Werner, V. Duarte, V. Marchand, M. Ignarski, A. Freiwald, R.-U. Müller, C. Dieterich, Y. Motorin, F. Butter, M. Atta, M. Helm, *Angew. Chem. Int. Ed.* **2018**, *57*, 7893; *Angew. Chem.* **2018**, *130*, 8019.

[3] V. F. Reichle, D. P. Petrov, V. Weber, K. Jung, S. Kellner, *Nat. Commun.* **2019**, *10*, 5600.

[4] T. Ohira, K. Minowa, K. Sugiyama, S. Yamashita, Y. Sakaguchi, K. Miyauchi, R. Noguchi, A. Kaneko, I. Orita, T. Fukui, K. Tomita, T. Suzuki, *Nature* **2022**, *605*, 372.

[5] a) J. H. J. Hoeijmakers, *N. Engl. J. Med.* **2009**, *361*, 1475; b) J. Cadet, J. R. Wagner, *Cold Spring Harbor Perspect. Biol.* **2013**, *5*, a012559; c) Q. Kong, C.-L. G. Lin, *Cell. Mol. Life Sci.* **2010**, *67*, 1817; d) B. Nawrot, E. Sochacka, M. Düchler, *Cell. Mol. Life Sci.* **2011**, *68*, 4023.

[6] F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, B. Kavli, M. S. Bratlie, J. Peña-Diaz, M. Otterlei, G. Slupphaug, H. E. Krokan, *DNA Repair* **2004**, *3*, 1389.

[7] C. Sun, M. Jora, B. Solivio, P. A. Limbach, B. Addepalli, *ACS Chem. Biol.* **2018**, *13*, 567.

[8] C. Borek, V. F. Reichle, S. Kellner, *ChemBioChem* **2020**, *21*, 2768.

[9] a) N. Chatterjee, G. C. Walker, *Environ. Mol. Mutagen.* **2017**, *58*, 235; b) B. Sedgwick, P. A. Bates, J. Paik, S. C. Jacobs, T. Lindahl, *DNA Repair* **2007**, *6*, 429; c) Z. Li, J. Wu, C. J. Deleo, *IUBMB Life* **2006**, *58*, 581; d) C. L. Simms, H. S. Zaher, *Cell. Mol. Life Sci.* **2016**, *73*, 3639; e) E. J. Wurtmann, S. L. Wolin, *Crit. Rev. Biochem. Mol. Biol.* **2009**, *44*, 34.

[10] a) B. I. Fedeles, V. Singh, J. C. Delaney, D. Li, J. M. Essigmann, *J. Biol. Chem.* **2015**, *290*, 20734; b) G. Zheng, Y. Fu, C. He, *Chem. Rev.* **2014**, *114*, 4602.

[11] a) F. L. Chung, H. J. Chen, R. G. Nath, *Carcinogenesis* **1996**, *17*, 2105; b) F. el Ghissassi, A. Barbin, J. Nair, H. Bartsch, *Chem. Res. Toxicol.* **1995**, *8*, 278; c) C. E. Vaca, J. Wilhelm, M. Harms-Ringdahl, *Mutat. Res. Rev. Genet. Toxicol.* **1988**, *195*, 137.

[12] K. Bergman, *Arch. Toxicol.* **1982**, *49*, 117.

[13] R. J. Laib, H. M. Bolt, *Toxicology* **1977**, *8*, 185.

[14] H. Ottenwälder, R. J. Laib, H. M. Bolt, *Arch. Toxicol.* **1979**, *41*, 279.

[15] J. L. Creech, M. N. Johnson, *J. Occup. Med.* **1974**, *16*, 150.

[16] a) A. Barbin, H. Brésil, A. Croisy, P. Jacquignon, C. Malaveille, R. Montesano, H. Bartsch, *Biochem. Biophys. Res. Commun.* **1975**, *67*, 596; b) C. Malaveille, H. Bartsch, A. Barbin, A. M. Camus, R. Montesano, A. Croisy, P. Jacquignon, *Biochem. Biophys. Res. Commun.* **1975**, *63*, 363.

[17] R. J. Laib, L. M. Gwinner, H. M. Bolt, *Chem.-Biol. Interact.* **1981**, *37*, 219.

[18] W. J. Krzyzosiak, J. Biernat, J. Ciesiołka, K. Gulewicz, M. Wiewiórowski, *Nucleic Acids Res.* **1981**, *9*, 2841.

[19] a) F. J. Bullock, O. Jardetzky, *J. Org. Chem.* **1964**, *29*, 1988; b) J. Tropp, A. G. Redfield, *J. Am. Chem. Soc.* **1980**, *102*, 534; c) M. P. Schweizer, S. I. Chan, G. K. Helmkamp, P. O. P. Ts'o, *J. Am. Chem. Soc.* **1964**, *86*, 696.

[20] H.-W. Winter, C. Wentrup, *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 720; *Angew. Chem.* **1980**, *92*, 743.

[21] D. P. Bhatt, X. Chen, J. D. Geiger, T. A. Rosenberger, *J. Chromatogr. B* **2012**, *889–890*, 110.

[22] S. Lund, R. Hall, G. J. Williams, *ACS Synth. Biol.* **2019**, *8*, 232.

[23] a) A. Favre, E. Hajnsdorf, K. Thiam, A. Caldeira de Araujo, *Biochimie* **1985**, *67*, 335; b) J. Biernat, J. Ciesiołka, P. Górnicki, R. W. Adamiak, W. J. Kryzosiak, M. Wiewiórowski, *Nucleic Acids Res.* **1978**, *5*, 789.

Manuscript received: November 21, 2022
Accepted manuscript online: January 11, 2023
Version of record online: February 7, 2023