AlkAniline-Seq: Profiling of m7G and m3C RNA Modifications at Single Nucleotide Resolution

Virginie Marchand, Lilia Ayadi, Felix G. M. Ernst, Jasmin Hertler, Valérie Bourguignon-Igel, Adeline Galvanin, Annika Kotter, Mark Helm, Denis L. J. Lafontaine, and Yuri Motorin*

Abstract: RNA modifications play essential roles in gene expression regulation. Only seven out of > 150 known RNA modifications are detectable transcriptome-wide by deep sequencing. Here we describe a new principle of RNAsseq library preparation, which relies on a chemistry based positive enrichment of reads in the resulting libraries, and therefore leads to unprecedented signal-to-noise ratios. The proposed approach eschews conventional RNA sequencing chemistry and rather exploits the generation of abasic sites and subsequent aniline cleavage. The newly generated 5'-phosphates are used as unique entry for ligation of an adapter in library preparation. This positive selection, embodied in the AlkAniline-Seq, enables a deep sequencing-based technology for the simultaneous detection of 7-methylguanosine (m7G) and 3'-methylcytidine (m3C) in RNA at single nucleotide resolution. As a proof-of-concept, we used AlkAniline-Seq to comprehensively validate known m7G and m3C sites in bacterial, yeast, and human cytoplasmic and mitochondrial tRNAs and rRNAs, as well as for identifying previously unmapped positions.

RNA modifications are playing essential roles in gene expression regulation.[1–3] Only seven (m6A, m6Am, m1A, m5C, hm5C, Nm and y) out of > 150 RNA modifications are detectable transcriptome-wide by deep sequencing. Modified nucleotides in tRNA, rRNA and mRNA do not only affect RNA processing, transport and stability, but these residues also impact mRNA translation.[11–15] Despite these important findings, details are scarce and disputed on the distribution and functions of many modified nucleotides in different transcriptomes. Only recently, high-throughput sequencing (NGS) methods coupled to antibody-directed enrichment provided comprehensive maps of m6A and m1A residues in different species.[16–18] For other RNA modifications, the use of specific chemical reagents allowed their high-throughput mapping.[19–22] While these approaches brought important discoveries in the field, several challenges remain open. For one, the list of NGS-detectable RNA modifications is still extremely restricted. Considering that the natural RNA modification repertoire comprises > 150 modifications, there is an urgent need to develop novel chemistry of detection, ideally directly amenable to deep sequencing.

Here we report a fundamentally new detection principle that hinges upon a chemistry-based enrichment of RNAsseq library with RNA fragments containing certain modifications. The first clues in the discovery of this highly effective method emerged from the application of traditional chemical treatments to induce cleavage in RNAs containing m7G, for which this specific detection in RNA was described in the 70's.[23,24] We initially intended to apply sequential treatments with sodium borohydride (NaBH4) and aniline, for high throughput detection of cleavage sites. We found that NaBH4 reduction combined with subsequent aniline cleavage of the resulting abasic site[25] produced high background, as is typical for detection of abortive cDNA.[26] Coupling to deep sequencing, this technique produced signals at some known m7G sites in tRNAs and rRNA, but also led to numerous false discovery hits (Figure S1 in the Supporting Information) thus making it unsuitable for a search of m7G in low abundant RNAs.

We therefore reasoned, that the 5'-phosphate generated during the aniline cleavage could be exploited as a selective entry into a library preparation including a limiting ligation step with exactly such phosphate as prerequisite (Figure 1b and Figure S2a). This approach did not only yield very clear signals, but, in addition, revealed that the NaBH4 treatment ultimately leads to ring opening, base elimination and nucleophilic attack by hydroxide anions, a reaction that produces signals, but, in addition, revealed that the NaBH4 treatment could not be omitted in favor of a limited alkaline hydrolysis. The resulting novel approach (named AlkAniline-Seq) can map abasic sites and modified nucleotides that generate the latter upon treatment. AlkAniline-Seq combines three successive treatments: i) RNA alkaline hydrolysis, ii) extensive 5'-phosphate dephosphorylation and iii) aniline cleavage (Figure 1a). According to the literature,[27,28] m7G residues are subject of nucleophilic attack by hydroxide anions, a reaction that ultimately leads to ring opening, base elimination and creation of an abasic site.[29]

Upon exposing m7G nucleosides to the alkaline conditions of the AlkAniline-Seq protocol, known intermediates of this

Supporting information (methods, including statements of data availability and any associated accession codes and references) and the ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/anie.201810946. The datasets generated and analyzed in the current study are available in the European Nucleotide Archive, accession number PRJEB26005.
Figure 1. Overview of the AlkAniline-Seq technology. a) Schematic representation of the RNA chain cleavage and primer ligation to unique 5'-phosphates resulting from abasic site decomposition. Enrichment of sequencing reads starting at the nucleotide N + 1 is shown. Modified residue (m7G, m6C or D) is shown as a blue dot. Broken blue dot corresponds to RNA abasic site. b) Chemical reactions leading to the formation of an abasic site at m7G/m6C/D residues in RNA. The first step consists of an alkaline hydrolysis (RNA fragmentation), and the second step is an aniline cleavage followed by β-elimination of ribose-aniline adduct. c) Principles of scoring of AlkAniline-Seq signals by counting of 5'-reads' extremities and calculation of normalized cleavage (left) or Stop-ratio (right). Normalized cleavage is expressed in units (reads starting at a given position in RNA / total number of reads aligned to this RNA, ranging from 5–25 units for background to a maximum of 1000 units for a single positive hit in RNA). "Stop-ratio" is calculated as the proportion of reads starting at a position out of all the reads overlapping it. "Stop-ratio" was extensively used so far in transcriptome-wide mapping of RNA modifications (e.g. in Ψ-seq).
**Figure 2.** Detection and quantification of the unique m$^7$G1575 residue in *S. cerevisiae* 18S rRNA. a) Influence of aniline treatment on the level of observed m$^7$G1575 signal. Normalized cleavage at m$^7$G1575 in WT and bud23.Δ 18S rRNA in mock experiment (no aniline) and in aniline-treated samples (treated) by RNA position. The m$^7$G1575 is only detected in the wild-type treated sample. The inset gives LC-MS quantification of m$^7$G molar amount in WT and bud23.Δ 18S rRNA. Molar amount is a mean of n = 3 independent measurements, which are shown as yellow dots. (b) Sensitivity of m$^7$G detection in mixes of unmodified (bud23.Δ) and modified (WT) 18S rRNA. Proportion of modified 18S rRNA varied from 0 to 100% (indicated to the right). Inset gives a trace of normalized cleavage (in black) and Stop-ratio (in grey) as a function of molar ratio of m$^7$G1575 modification.

**bud23.Δ (KO strain for m$^7$G1575),** with the proportion of m$^7$G1575-containing WT rRNA varying from 1 to 100%. Remarkably, a saturation curve of the signal was obtained (Figure 2c), where the signal is well visible over background.

**bud23.Δ (KO strain for m$^7$G1575),** with the proportion of m$^7$G1575-containing WT rRNA varying from 1 to 100%. Remarkably, a saturation curve of the signal was obtained (Figure 2c), where the signal is well visible over background.

**Figure 3.** Detection of m$^7$G46 and m$^7$C32 residues in *S. cerevisiae* tRNAs. a) Normalized cleavage signals for *S. cerevisiae* tRNA$^\text{IAC}$ in *trm8*, *trm82*, WT and bud23.Δ yeast strains. Inset on the right shows the cloverleaf structure of tRNA with the modified position and the corresponding enzymatic activity (heterocomplex Trm8/Trm82). Inset on the left gives quantification of normalized cleavage in the four strains used. b) Heat map shows normalized cleavage values for all *S. cerevisiae* RNAs containing m$^7$G residues in the same strains. LC-MS quantification of the residual m$^7$G content in tRNA fraction from *trm8* strain is also shown. Molar amount is a mean of n = 3 independent measurements shown on the Figure as yellow dots. The residual signals for tRNA$^{\text{NGG}}$ and tRNA$^{\text{CUU}}$ in *trm8* and *trm82* strains are not due to residual m$^7$G level, but correspond to incomplete dephosphorylation of some positions in the compact 3D tRNA structure. c,d) Normalized cleavage signals for *S. cerevisiae* tRNA$^{\text{NGA}}$ and tRNA$^{\text{IGU}}$ in *trm140*, WT, bud23.Δ, *trm8*, and *trm82* yeast strains. Inset on the left shows the cloverleaf structure of tRNA with the modified position and corresponding enzymatic activity (Trm140). Inset on the right gives quantification of normalized cleavage in the different strains used.
value already at \(\approx 2\%\) of modified RNA, which is unprecedented in any RNAseq approach for RNA detection. This demonstrates a high sensitivity of AlkAniline-Seq even for the detection of substoichiometrically modified m\(^7\)G sites.

A first application consisted in the faithful mapping of all 10 previously known m\(^7\)G sites in yeast tRNAs, all of which depend on the Trm8/Trm82 complex\[^{33}\] (Figure 3a and Figure S6a). A heat map summarizing normalized cleavage efficiency for yeast m\(^7\)G-containing tRNAs\[^{34}\] for WT, bud23\(^D\), trm8\(^D\) and trm82\(^D\) strains is shown in Figure 3b. In addition to known sites, a readily detectable signal for m\(^G\)46 residue was detected in tRNA\(^{Ala}\)(IGC) in WT and bud23\(^D\) cells, but not in trm8\(^D\)/trm82\(^D\) strains, thus identifying a new substrate of the Trm8/Trm82 complex (Figure 3b and Figure S7). A further application to E. coli tRNAs successfully detected all known m\(^7\)G positions, without any false-positive hits (representative examples are shown in Figure S5c).

An inspection of the above results at residues other than guanosines identified further strong signals at selected cytosines and uridines, strongly suggesting that the specificity of AlkAniline-Seq, was not limited to m\(^7\)G, but reacted to additional of the altogether 46\[^{34}\] modifications present in yeast and E. coli. In particular, the AlkAniline-Seq profiles for yeast tRNA\(^{Ser}\) and tRNA\(^{Thr}\) showed a consistent and prominent signal in the anticodon loop, which corresponds to m\(^3\)C32 (known to be modified by Trm140).\[^{35}\]

Accordingly, in both tRNAs, the signal specifically disappeared in trm140\(^D\) yeast strain (Figure 3c,d, Figure S6c). However, exposure of m\(^3\)C nucleoside to alkaline conditions showed only deamination to m\(^3\)U (Figure S8), which is not reactive in AlkAniline-Seq.

Dihydrouridine (D) in yeast and bacterial tRNAs also provided detectable signal in AlkAniline-Seq. Signals were observed for a limited subset of known D residues in tRNAs and the relative signal strength was comparably lower (100–300 units) (Figure S9). To validate consistency of D detection, we performed AlkAniline-Seq analysis on yeast mutants carrying a deletion of one of the four known tRNA:dihydrouridine synthases (Dus1, Dus2, Dus3, and Dus4), responsible for formation of D16/D17, D20, D47 and D20ab, respectively.\[^{36,37}\] Analysis of D signals for tRNAs and their absence in respective knockout strains followed previous assignments of DUS genes to D sites in tRNAs. Thus, AlkAniline-Seq responds to the presence of D, although the signal strength is considerably lower than for m\(^7\)G and m\(^3\)C, likely due to incomplete formation of abasic sites from D. While the chemistry of m\(^7\)G detection could be readily reenacted (Figure S2), deeper investigations must clarify the basis for the detection of m\(^3\)C and D.

AlkAniline-Seq was further applied to human cytoplasmic and mitochondrial rRNAs and tRNAs. Using total RNA as a starting material, we established a complete profile of m\(^7\)G, m\(^3\)C and D sites in these RNA molecules (Table S3). As anticipated, analysis of rRNAs revealed a single 18S-m\(^7\)G1639 residue\[^{38}\] without additional signals in other human cytoplasmic and mitochondrial rRNA species (Figure S10). Profiling of human cytoplasmic and mitochondrial tRNAs revealed seventeen m\(^7\)G, twelve m\(^3\)C and thirty-three D sites (Figure 4, Table S4). This comprehensive listing once more
illustrates the exceptional sensitivity of the method, given that some of the low-abundant RNA species were only present in pg amounts.

The presence of mG, mC, and D in eukaryotic mRNA is still questionable, even if some of these nucleotides were reported.\[39\] To perform our analysis transcriptome-wide, AlkAniline-Seq, converted to library and sequenced. The resulting reads were aligned to the yeast SacCer3 genome as well as to the RefSeq mRNA database. Stop-ratios between reads terminating at a given position and reads overlapping the same were calculated (Figure 1 c).

Sufficient amount of sequencing information (> 100 reads coverage per mRNA) was obtained for \( \approx 1500 \) yeast mRNAs. However, only very few (< 20) hits satisfied the criteria validated for rRNA and tRNA (Stop-ratio > 0.75 and > 100 reads coverage). The majority of them were either cryptic tRNA sequences or transcription start sites of highly expressed yeast genes, most probably due to incomplete de-phosphorylation of mG/Capped mRNAs during step 2 of the protocol (Figure S11). Moreover, no mG/mC/(D) signals were detected for non-polyadenylated yeast ncRNA (snRNAs, snoRNAs, etc).

In conclusion, AlkAniline-Seq illustrates a new concept for the high-throughput detection of modified RNA nucleotides by deep sequencing: a specific sequence of chemical reactions leading to the production of RNA fragments cleaved at the N + 1 nucleotide to the modification site, and those fragments are selectively converted into sequencing libraries. This dramatically increases the specificity and the sensitivity of the approach, since the only other sources of accessible 5’-phosphate residues are seldom.

Another interesting property of AlkAniline-Seq is that it simultaneously detects at least two different RNA modifications, namely: mG and mC, and, at lower extent, D. Since these modified nucleotides are derived from different parental nucleotides they are readily distinguishable when the reads are aligned.

We foresee multiple applications for AlkAniline-Seq mG/mC/(D) mapping in organisms where RNA modification has not been systematically addressed, which includes many model organisms in developmental biology. Further perspectives include putative stress-inducible and disease-specific modifications in mRNA and ncRNAs other than tRNA/rRNAs in various biological and pathological contexts.\[40,41\] AlkAniline-Seq could also be used for profiling of 16S-mG1405-dependent aminoglycoside resistance in gram-negative bacteria\[42\] and in other biomedical projects aimed at studying the dynamics and regulation of mG, mC, and D in diverse biological context.

Acknowledgements

Y.M. was supported by ANR-DFG grant HTRNA-Mod ANR-13-ISV8-0001 and Research funds from Lorraine Région (France). M.H. was supported by the DFG (HE3397B1-1, HE3397/13-2, SPP1784). The Lab of D.L.J.L. is supported by the Université Libre de Bruxelles (ULB), the Fonds National de la Recherche (F.R.S./FNRS), the Walloon Region (DG06), and the Fédération Wallonie-Bruxelles. Work performed in the frame of the COST EPITRAN network.

Conflict of interest

The authors declare no conflict of interest.

Keywords: abasic site · deep sequencing · epitranscriptomics · methylation · RNA modification

References


Manuscript received: September 26, 2018
Accepted manuscript online: October 29, 2018
Version of record online: 2018

These are not the final page numbers!
Deep sequencing was applied in a new concept of library preparation for detection of RNA modifications. Modified RNA is treated by alkaline hydrolysis, dephosphorylated and subjected to aniline cleavage of abasic sites. The resulting 5'-phosphates in RNA are used for specific ligation of the sequencing adapter. The method can be applied for specific and sensitive detection of m7G, m3C and (D)-residues in RNAs.


Share your work on social media! Angewandte Chemie has added Twitter as a means to promote your article. Twitter is an online microblogging service that enables its users to send and read short messages and media, known as tweets. Please check the pre-written tweet in the galley proofs for accuracy. If you, your team, or institution have a Twitter account, please include its handle @username. Please use hashtags only for the most important keywords, such as #catalysis, #nanoparticles, or #proteindesign. The ToC picture and a link to your article will be added automatically, so the tweet text must not exceed 250 characters. This tweet will be posted on the journal’s Twitter account (follow us @angew_chem) upon publication of your article in its final (possibly unpaginated) form. We recommend you to re-tweet it to alert more researchers about your publication, or to point it out to your institution’s social media team.
Please check that the ORCID identifiers listed below are correct. We encourage all authors to provide an ORCID identifier for each coauthor. ORCID is a registry that provides researchers with a unique digital identifier. Some funding agencies recommend or even require the inclusion of ORCID IDs in all published articles, and authors should consult their funding agency guidelines for details. Registration is easy and free; for further information, see http://orcid.org/.

Virginie Marchand
Lilia Ayadi
Felix G. M. Ernst
Jasmin Hertler
Valérie Bourguignon-Igel
Adeline Galvanin
Annika Kotter
Mark Helm
Denis L. J. Lafontaine
Yuri Motorin http://orcid.org/0000-0002-8018-334X

Author Contributions
VM. and L.A. performed AlkAniline-Seq, F.E., A.G. and V.B.I. prepared strains/cell lines and extracted RNA, J.H. and A.K. performed LC-MS analysis, M.H., D.L.J.L and Y.M. analysed the data and wrote the manuscript.