Profiling of RNA modifications by multiplexed stable isotope labelling

Stefanie Kellner[a], Jennifer Neumann[a], David Rosenkranz[b], Svetlana Lebedeva[c], René Ketting[c], Hans Zischler[b], Dirk Schneider[a], and Mark Helm*[a]

SUPPLEMENTARY INFORMATION

1. Materials & Methods
2. Figures

Figure S 1 Fragmentation pathway of modified RNA nucleosides used for neutral loss scan method. (a) Fragmentation of ribonucleosides at the glycosidic bond leads to a neutral loss of ribose (highlighted in blue). (b) Fragmentation of ribose-methylated nucleosides is observed at the glycosidic bond, which leads to the neutral loss of methyl-ribose (highlighted in blue). (c) Fragmentation of pseudouridine is not observed at the glycosidic bond. Instead a neutral loss of 2 mol water can be observed in the mass spectrum.

Figure S 2 Mass spectrum of mcmo^{5}U reveals the presence of the protonated nucleoside and the sodium and potassium adducts.

Figure S 3 MS spectrum of natural 2-methylthioadenosine (ms^{2}A) and synthetic ms^{2}A. Both substances elute at the same retention time (26.0 minutes) and have identical MS spectra and fragmentation behavior. Here, the isotope peak at m/z 316.2 is noteworthy, as it is caused by natural occurring ^{34}S (4.0% abundance in nature), which is a hint for sulfur containing nucleosides.

Figure S 4 DAD chromatogram at 254 nm serves as loading control to verify injection of equal amounts of RNA. The upper lane displays an overlay of all 5 species chromatograms. Below, the separated, unscaled chromatograms for all species can be seen.

3. Tables

Table S 1Neutral Loss Scan method for ribose loss (-132) of unlabeled RNA

Table S 2 List of category 2 identified nucleosides from *E.coli* (identification by comparison with database)

Table S 3 List of category 2 identified nucleosides from yeast (identification by comparison with database)

1. MATERIALS & METHODS

Preparation of unlabelled, ^{13}C and ^{15}N isotopically labelled *Escherichia coli* (*E. coli*) and *Saccharomyces cerevisiae* (*S. c.*). *E. coli* (strain MC4100) was grown in M9 medium (2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4 % glucose, 6.8 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl, 0.2 M $MgSO_4$ and 0.2 M $CaCl_2$). The unlabelled medium contained unlabelled glucose and NH_4Cl, while ^{13}C-glucose (all carbons exchanged for ^{13}C, from Sigma-Aldrich, Munich, Germany) was used for the ^{13}C medium as the only carbon source and $^{15}NH_4Cl$ (Sigma-Aldrich, Munich, Germany) as the only nitrogen source in the ^{15}N medium. The growth of the cultures was monitored and cells harvested at an OD_{600} of 1.8 by centrifugation. Yeast (strain BY 4771) was grown in commercial media from Silantes, Munich, Germany. All media were ready-to-use OD-2 media, either standard, ^{13}C labelled (Prod.Nr. 111201402) or ^{15}N labelled (Prod. Nr. 111301402) and cells were grown using the standard procedure. Again cell growth was monitored and cells were harvested at an OD of 1.6 by centrifugation. RNA of all six cultures was isolated using TRI-reagent and the included RNA isolation manual (Sigma-Aldrich, Munich, Germany). All other chemicals needed for isolation were purchased from Sigma-Aldrich.

Preparation of zebrafish RNA. Approximately 100 1-day old zebrafish embryos (Danio rerio, wild-type TL strain) were homogenized in 1 mL of Trizol LS (Life Technologies, 10296-028) by passing through a syringe with 23G needle until the solution was homogenous. Following homogenization, total RNA was isolated according to manufacturer's protocol

Preparation of small RNA from pig and human. Total RNA was isolated from *Sus scrofa* testis tissue (50 mg) with TRI Reagent according to the manufacturer's instructions. Total RNA from human testis tissue was purchased from Clontech (CATALOG No. 636533, LOT NUMBER 8082011A). The total RNA samples were applied to a denaturing acrylamide gel (12% acrylamid/bisacrylamid 19:1, 8 M urea, TBE-buffered) and run for 30 minutes at 1200 V/50 mA (=60 W) together with GeneRuler Ultra Low Range DNA Ladder (ThermoScientific). Visualization of RNA was achieved by staining with ethidiumbromid/H_2O (1:20000) for 3 minutes. Gel bands corresponding to 75-100bp ladder bands were excised from the gel, eluted and subsequently precipitated with ethanol.

Sample preparation for LC-MS/MS analysis. The isolated RNAs (final concentration 1 µg/µL) were dissolved in 20 mM NH_4OAc pH 5.3 and digested to nucleosides as described previously.[1]

Development of Neutral-Loss-Scan method.

The samples were analysed on an Agilent 1260 series HPLC equipped with a diode array detector (DAD) and triple quadrupole mass spectrometer Agilent 6460.

HPLC parameters. A Synergy Fusion RP column (4 µm particle size, 80 Å pore size, 250 mm length, 2 mm inner diameter) from Phenomenex (Aschaffenburg, Germany) was used at 35 °C. The solvents consisted of 5 mM ammonium acetate buffer adjusted to pH 5.3 using acetic acid (solvent A) and pure acetonitrile (solvent B). The elution started with 100% solvent A followed by a linear gradient to 10% solvent B at 20 min and 25% solvent B after 30 min. For elution of more hydrophobic compounds solvent B was increased to 80% at 40 minutes. Initial conditions were regenerated by rinsing with 100% solvent A for 14 minutes. The flow rate was 0.35 mL/min.

MS parameters. The effluent from the column was first measured photometrical at 254 nm by the DAD followed by the mass spectrometer equipped with an electrospray ion source (Agilent Jet Stream). ESI parameters were as follows: gas (N_2) temperature 300°C, Gas (N_2) flow 5 L/min, Nebulizer pressure 35 psi, Sheath gas (N_2) temperature 350°C, Sheath gas (N_2) flow 12 L/min, capillary voltage 3500 V and fragmentor voltage 80 V. The MS was operated in positive ion mode using Agilent MassHunter software in neutral loss scan mode. The monitored neutral losses at a collision energy of 15 eV were 132 (ribose), 146 (methyl-ribose) and 36 (loss of 2 water molecules for pseudouridine detection[2]) in a mass range of 230-600 Da. All monitored neutral losses are presented in Figure S1. To exclude major nucleosides from the analysis a segmented method was designed, which does not allow detection of cytidine (m/z 244, R_t = 4.6 min), uridine (m/z 245, R_t = 6.5 min), guanosine (m/z 284, R_t = 11.2 min) and adenosine (m/z 268, R_t = 15.9 min) at their respective retention times. This segmentation allows the injection of high sample amounts without saturating the MS system. The complete method for the neutral loss of 132 Da can be seen in Table S 1. In addition, all segments contain neutral loss of 146 and 36 in the mass range of 230-600 with a dwell time of 555 ms each, which results in 1.5 seconds per cycle.

The NLS methods of ^{15}N and ^{13}C labelled RNA had the same MS parameters as described for the unlabelled RNA. However, for NLS of ^{15}N RNA the excluded m/z of major nucleosides was adjusted to fit the isotopomer m/z (cytidine/247, uridine/247, guanosine/289 and adenosine/273). For the ^{13}C labelled RNA, not only the major nucleoside m/z had to be adjusted (cytidine/253, uridine/254, guanosine/294 and adenosine/278), but in addition, the ribose loss was adjusted to -137 and the methylated-ribose to -152.

NLS method sensitivity determination. For determination of detection sensitivity, a mixture of modified nucleosides (available in weighable quantities) was prepared in the range of 0.01 fmol/µL to 100 fmol/µL and injected to LC-NLS analysis. The peak areas and signal-to-noise (S/N) ratios were determined using MassHunter® Qualitative software. The LOD was defined with a S/N ratio > 3. Nucleosides used: 5-methylcytidine (m^5C), 2'O-methylcytidine (Cm), 4-thiouridine (s^4U), 2'O-methyladenosine (Am) and inosine (I) from Sigma-Aldrich, Munich, Germany; 5-methyluridine (m^5U), pseudouridine (Ψ) and 2'O-methyluridine (Um) from Berry&Associates, Dexter, MI, USA; 6-methyladenosine (m^6A) and 6-dimethyladenosine ($m^6{}_2A$) were a gift from Glen Björk.

Analysis of unlabelled, ^{13}C labelled, and ^{15}N labelled total RNA from *E. coli* and *S. cerevisiae*. 10 µg of each nucleoside mixture of the RNA digest was injected and analysed as described in method development with the respective NLS method. The resulting TIC chromatograms of the unlabelled RNA samples were searched for prominent peaks with a minimal abundance of at least 10^3 (chosen on the basis that a signal with abundance of 10^3 is 10 fold higher than the average background noise) and mass spectra of these peaks were extracted. The remaining peaks were intensively scrutinized to eliminate artefacts. A major artefact is caused by the high amount of injected RNA digests (> 10 µg), which leads to the detection of natural isotopes of the highly abundant major nucleosides (*e.g.* 3 to 4 ^{13}C atoms per nucleoside gave rise to signals above 10^3). Another observed artefact were nucleoside adducts of sodium, and potassium ions, acetonitrile and acetate clusters, which gave rise to false positive results. In a next step, the respective ^{13}C and ^{15}N RNA chromatograms were overlaid with the unlabelled chromatogram and coinciding peaks and mass spectra were extracted. The resulting total number of peaks (58) was then subdivided in three categories. Category 1 included peaks of modified nucleosides identified with synthetic standards. Category 2 included peaks of modified nucleosides with fitting m/z, carbon and nitrogen content compared to Modomics database (see **Table S 2** and **Table S 3**). In addition, retention times compared to the major nucleosides were used for further identification (*e.g.* ct^6A). Category 3 included peaks with undescribed m/z values. During this stage several peaks from the preliminary peak-list were found to be sodium or potassium adducts of known modified nucleosides and were removed from the list.

Identification of the peaks with m/z 255 and m/z 314. Analysis of unlabelled and $^{13}C/^{15}N$ labelled RNA from *E. coli* and yeast revealed masses and elemental composition of several peaks. However, two of these peaks with a neutral loss of ribose (-132) were found to fit to the nucleosides N-ribosylnicotinamid (r-NA) and 2-methylthioadenosine (ms^2A) in all analysed criteria.

r-NA	M_w = 255 g/mol	elemental composition: $C_{11}N_2O_5H_{15}$	Positively charged N1
Found Peak	m/z = 255	Isotope result: $C_{11}N_2$	R_t = 3.1 min
ms^2A	M_w = 313 g/mol	elemental composition: $C_{11}N_5SO_4H_{16}$	
Found Peak	m/z = 314	Isotope result: $C_{11}N_5$	R_t = 26.0 min

To further prove the identity of these peaks synthetic standards were used. For identification of r-NA, commercially available NADP (Sigma-Aldrich, Munich, Germany) was digested in the presence of 3% ammoniumhydroxide (Sigma-Aldrich, Munich, Germany) and 1u FastAlkaline Phosphatase (Fisher scientific, Schwerte, Germany) per 100 ng for 2 hours at 37 °C. 100 ng ms^2A-diphosphate (Sigma-Aldrich, Munich, Germany) was digested in the presence of 1 u FastAP to yield the free nucleoside. The synthetic standards were injected and the resulting mass spectra and retention times compared to the peaks found in the samples of *E. coli* and yeast (see Figure 2b in main manuscript). The synthetic standards were found to have the same mass spectra and retention times as the found peaks, which can be therefore assigned as r-NA and ms^2A. The mass spectra of natural and synthetic ms^2A can be seen in Figure S 3. Here, the m/z 316.2 indicates the presence of a sulfur atom in the modified nucleoside (^{34}S isotope peak), which can be used for identification of thiolated and other sulfur containing modified nucleosides.

RNA modification profiling of tissue. 10 µg of each RNA (total RNA *E. coli, S. c.* and zebrafish, small RNAs pig and human) was injected and analysed using the NLS method as described. The UV chromatogram recorded at 254 nm served as loading control to confirm the injection of equal amounts of RNA; the peak heights of the major nucleosides are identical in all five samples (see Figure S 4). In a next step, the chromatograms of all categorized peaks were extracted (EIC function of Agilent MassHunter Qualitative software) and integrated. For each nucleoside, the sample with the highest peak area was defined as 100 % and all other species peak areas were related to the respective peak. The heatmap was generated using tab-stop separated .txt-files and the online platform[2]. While blue indicates a 100% modification status, white indicates an absence of the modification in the species.

Identification of less abundant modified nucleosides by extracted ion chromatogram (EIC) analysis. The m/z values and elemental composition of all modified nucleosides were extracted from the Modomics database. The datasets from unlabelled *E. coli* and yeast RNA analysis was searched for the literature-based m/z values that were not covered by the nucleosides in category 1 or 2. 7 additional modified nucleosides were thus identified (m^{28}A, mcm^5U, ncm^5s^2U and t^6A, mnm^5s^2U, acp^3D and g^6A). In small RNA samples from pig and human, the modified nucleosides m^5Um, $m^4{}_2$Cm and m^6Am were identified by TIC analysis. EIC analysis of these samples revealed minor presence of 5 modified wybutosine derivatives (yW-86, yW-72, yW-58, OHyW and o2yW) and ms^2t^6A.

2. FIGURES

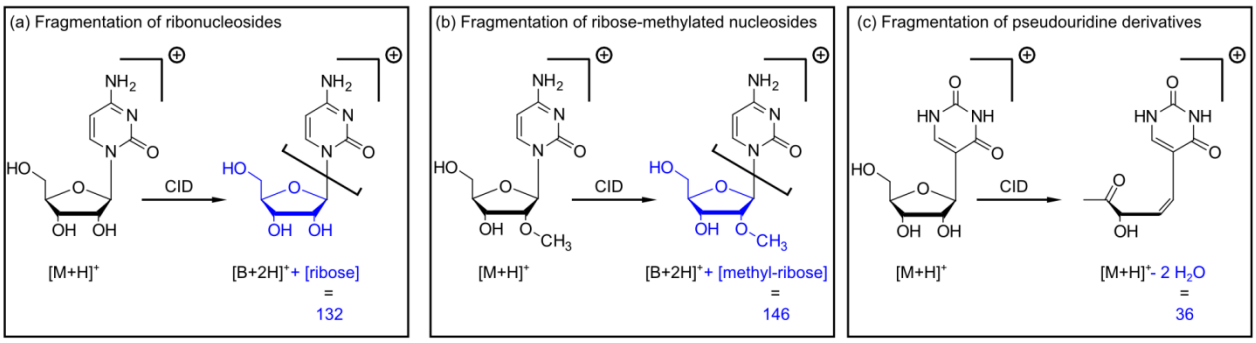

Figure S 1 Fragmentation pathway of modified RNA nucleosides used for neutral loss scan method. (a) Fragmentation of ribonucleosides at the glycosidic bond leads to a neutral loss of ribose (highlighted in blue). (b) Fragmentation of ribose-methylated nucleosides is observed at the glycosidic bond, which leads to the neutral loss of methyl-ribose (highlighted in blue). (c) Fragmentation of pseudouridine is not observed at the glycosidic bond. Instead a neutral loss of 2 mol water can be observed in the mass spectrum.

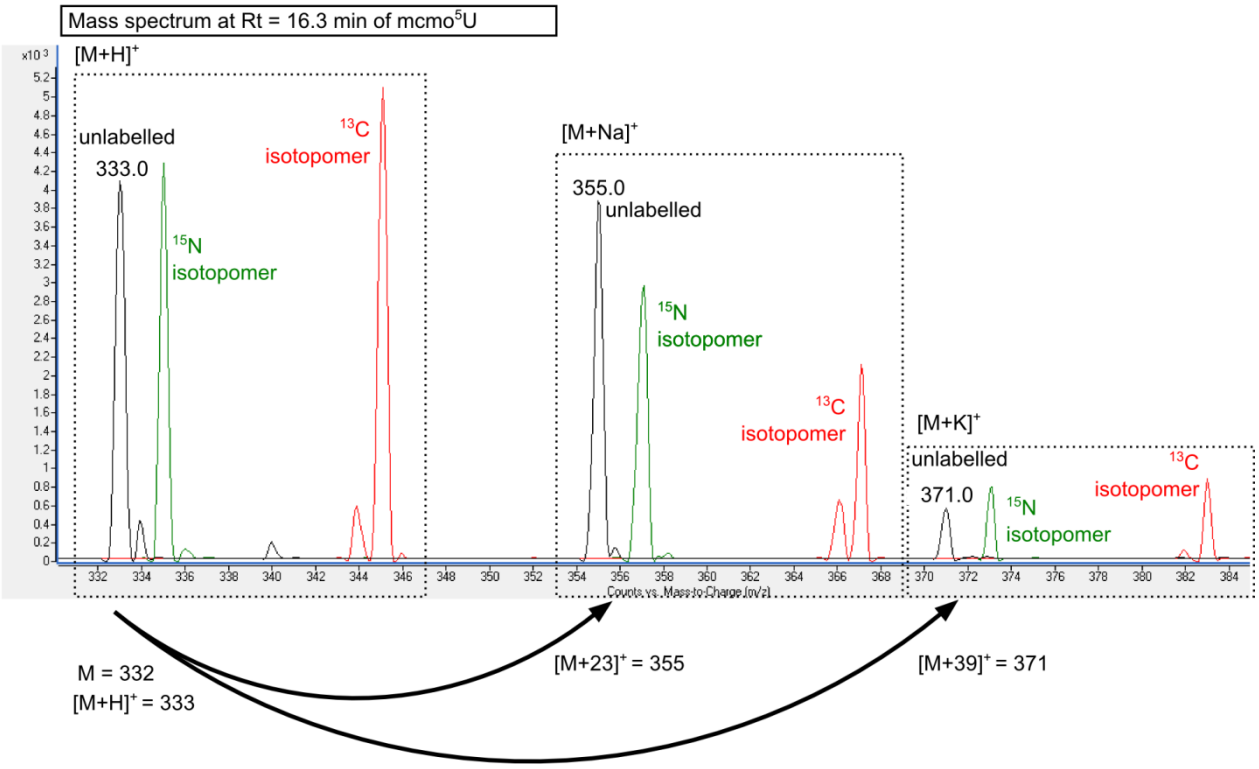

Figure S 2 Mass spectrum of mcmo^{5}U reveals the presence of the protonated nucleoside and the sodium and potassium adducts.

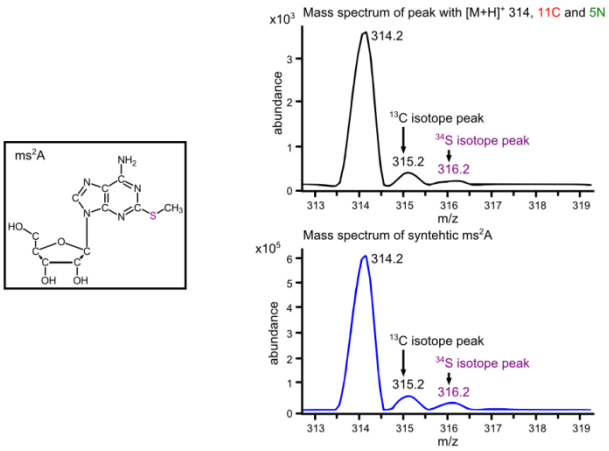

Figure S 3 MS spectrum of natural 2-methylthioadenosine (ms^2A) and synthetic ms^2A. Both substances elute at the same retention time (26.0 minutes) and have identical MS spectra and fragmentation behaviour. Here, the isotope peak at m/z 316.2 is noteworthy, as it is caused by natural occurring ^{34}S (4.0% abundance in nature), which is a hint for sulfur containing nucleosides.

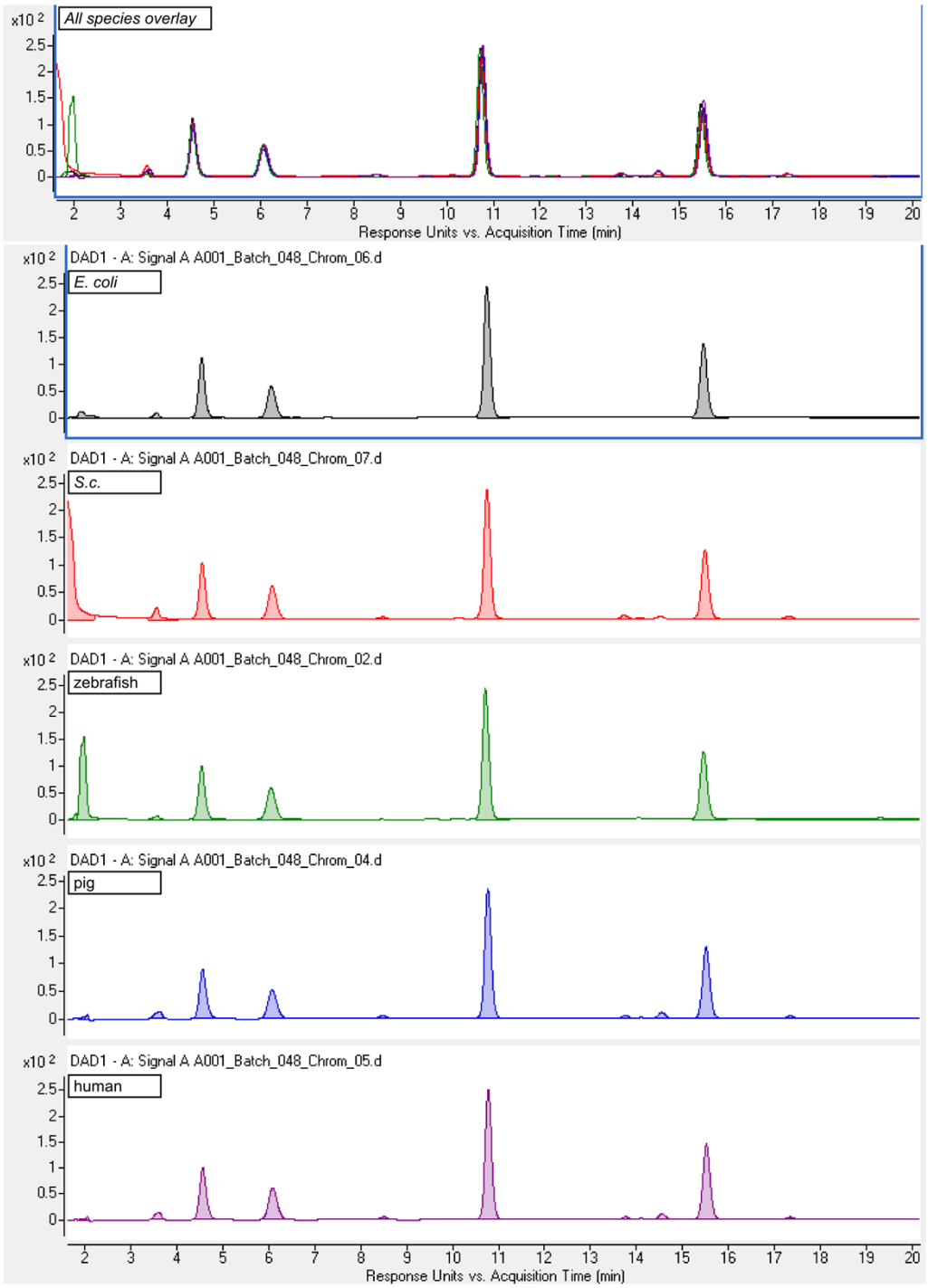

Figure S 4 DAD chromatogram at 254 nm serves as loading control to verify injection of equal amounts of RNA. The upper lane displays an overlay of all 5 species chromatograms. Below, the separated, unscaled chromatograms for all species can be seen.

3. TABLES

Table S 1Neutral Loss Scan method for ribose loss (-132) of unlabeled RNA

Segment	**Time frame**	**Mass range 1**	**Mass range 2**	**Mass excluded**	**Cycles per second**
1	0-2	waste			
2	2-4	230-600 Dwell: 555 ms		none	0.6
3	4-5	230-243 Dwell: 20 ms	246-600 Dwell: 535 ms	Cytidine (Rt =4.6‘)	
4	5-6	230-600 Dwell: 555 ms		none	
5	6-7.5	230-244 Dwell: 21 ms	247-600 Dwell: 534 ms	Uridine (Rt =6.5‘)	
6	7.5-10	230-600 Dwell: 555 ms		none	
7	10-12	230-283 Dwell: 80 ms	286-600 Dwell: 475 ms	Guanosine (Rt =11.2‘)	
8	12-15	230-600 Dwell: 555 ms		none	
9	15-17	230-267 Dwell: 56 ms	270-600 Dwell: 500 ms	Adenosine (Rt =15.9‘)	
10	17-45	230-600 Dwell: 555 ms		none	
11	45-50	waste			

Table S 2 List of category 2 identified nucleosides from *E coli* (identification by comparison with database)

From Modomics database			Expected m/z values				Experimental m/z		
Name	Abbreviation	Elemental composition	unlabelled M+H	^{13}C M+H	^{15}N M+H	R_t [min]	unlabelled M+H	^{13}C M+H	^{15}N M+H
N-ribosylnicotinamid*	r-NA	$C_{11}N_2O_5H_{15}$	255	266	257	3.2	255	266	257
dihydrouridine	D	$C_9O_6N_2H_{14}$	247	256	249	chrom.prob.	247	256	249
3-(3-amino-3-carboxypropyl)uridine	acp^3U	$C_{13}O_8N_3H_{19}$	346	359	349	5	346.1	n.d.	349.1
uridine 5-oxyacetic acid	cmo^5U	$C_{11}O_9N_2H_{14}$	319	330	321	5.9	319	n.d.	320.8
2-thiocytidine	s^2C	$C_9O_4N_3H_{13}S_1$	260	269	263	7.4	260	269	262.9
2-lysidine	k^2C	$C_{15}O_6N_5H_{25}$	372	387	377	12.4	372	387.2	377
N4,2′-O-dimethylcytidine	m^4Cm	$C_{11}O_5N_3H_{17}$	272	283	275	13.1	272	283	274.9
cyclic N6-threonylcarbamoyladenosine	ct^6A	$C_{15}O_7N_6H_{18}$	395	410	401	13,9/16.0/20	395	409.1	401.1
1-methylguanosine	m^1G	$C_{11}O_5N_5H_{15}$	298	309	303	14	298	309	303
N4-acetylcytidine	ac^4C	$C_{11}O_6N_3H_{15}$	286	297	289	14.7	286	297	289
uridine 5-oxyacetic acid methyl ester	$mcmo^5U$	$C_{12}O_9N_2H_{16}$	333	345	335	16	333	345	335
N6-threonylcarbamoyladenosine	t^6A	$C_{15}O_8N_6H_{20}$	413	425	418	17.7	413	425	418
2-methyladenosine	m^2A	$C_{11}O_4N_5H_{15}$	282	293	287	20.8	282	293	287
N6-hydroxynorvalylcarbamoyladenosine	hn^6A	$C_{16}O_8N_6H_{22}$	427	443	433	21	427	443	433
N6-methyl-N6-threonylcarbamoyladenosine	m^6t^6A	$C_{16}O_8N_6H_{22}$	427	443	433	21	427	443	433
2-methylthioadenosine*	ms^2A	$C_{11}N_5SO_4H_{16}$	314	325	319	26.0	314.2	325.2	319.2
N6-(cis-hydroxyisopentenyl)adenosine	io^6A	$C_{15}O_5N_5H_{21}$	352	367	357	28.2	352	367	357
2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine	ms^2io^6A	$C_{16}O_5N_5H_{23}S_1$	398	414	403	34.8	398.2	414.2	403.2

Table S 3 List of category 2 identified nucleosides from yeast (identification by comparison with database)

From Modomics database			Expected m/z values				Experimental m/z		
Name	Abbreviation	Elemental composition	unlabelled	^{13}C	^{15}N	R_t	unlabelled	^{13}C	^{15}N
			M+H	M+H	M+H	[min]	M+H	M+H	M+H
2-methylthio-N6-isopentenyladenosine	ms^2i^6A	$C_{16}O_4N_5H_{23}S_1$	382	398	387	38.6	382.1	398.1	387.1
5-carbamoylmethyluridine	ncm^5U	$C_{11}O_7N_3H_{15}$	302	313	305	5.5	302	n.d.	304.9
1-methylinosine	m^1I	$C_{11}O_5N_4H_{14}$	283	294	287	13.5	283	293.9	287
N2-methylguanosine	m^2G	$C_{11}O_5N_5H_{15}$	298	309	303	14.9	298.1	309.1	303.1
N2,N2-dimethylguanosine	$m^{2,2}G$	$C_{12}O_5N_5H_{17}$	312	324	317	17.7	312.1	324.1	317.1
5-methoxycarbonylmethyl-2-thiouridine	mcm^5s^2U	$C_{12}O_7N_2H_{16}S_1$	333	345	335	19.4	333	n.d.	335
wybutosine	yW	$C_{21}O_9N_6H_{28}$	509	530	515	31.9	509	n.d.	515

REFERENCES

1. S. Kellner, S. Seidu-Larry, J. Burhenne, Y. Motorin and M. Helm, *Nucleic acids research*, 2011, **39**, 7348-7360.
2. E. Dudley, R. Tuytten, A. Bond, F. Lemiere, A. G. Brenton, E. L. Esmans and R. P. Newton, *Rapid communications in mass spectrometry : RCM*, 2005, **19**, 3075-3085.