analytical chemistry

pubs.acs.org/ac Article

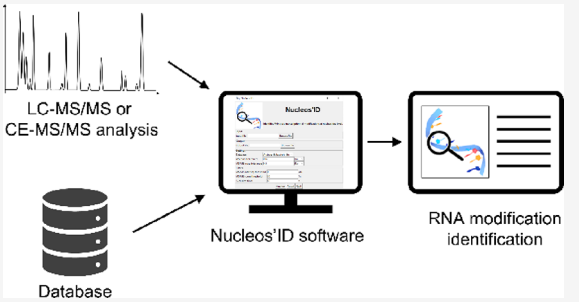
Nucleos'ID: A New Search Engine Enabling the Untargeted Identification of RNA Post-transcriptional Modifications from Tandem Mass Spectrometry Analyses of Nucleosides

Clarisse Gosset-Erard,[#] Mévie Didierjean,[#] Jérome Pansanel,[#] Antony Lechner, Philippe Wolff, Lauriane Kuhn, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, and Yannis-Nicolas François*

Cite This: https://doi.org/10.1021/acs.analchem.2c04722 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: As RNA post-transcriptional modifications are of growing interest, several methods were developed for their characterization. One of them established for their identification, at the nucleosidic level, is the hyphenation of separation methods, such as liquid chromatography or capillary electrophoresis, to tandem mass spectrometry. However, to our knowledge, no software is yet available for the untargeted identification of RNA post-transcriptional modifications from MS/MS data-dependent acquisitions. Thus, very long and tedious manual data interpretations are required. To meet the need of easier and faster data interpretation, a new user-friendly search engine, called Nucleos'ID, was developed for CE-MS/MS and LC−MS/MS users. Performances of this new software were evaluated on CE-MS/MS data from nucleoside analyses of already well-described *Saccharomyces cerevisiae* transfer RNA and *Bos taurus* total tRNA extract. All samples showed great true positive, true negative, and false discovery rates considering the database size containing all modified and unmodified nucleosides referenced in the literature. The true positive and true negative rates obtained were above 0.94, while the false discovery rates were between 0.09 and 0.17. To increase the level of sample complexity, untargeted identification of several RNA modifications from *Pseudomonas aeruginosa* 70S ribosome was achieved by the Nucleos'ID search following CE-MS/MS analysis.



Due to their implication in many biological and structural functions, RNA post-transcriptional modifications are of great interest. Indeed, RNA modifications were reported to play an important role in RNA structural folding and stability,[1] base pairing,[2] translation fidelity,[3] regulation of gene expression,[4] and also cell development.[5] Some modifications are also involved in antibiotic resistance mechanisms[6] or several diseases.[7] Over the years, therapeutic strategies based on post-transcriptional RNA modifications have received growing interest. For instance, modified messenger RNA were reported as potential treatment for infectious disease[8] and some RNA modifications were reported as cancer biomarkers for diagnostic purposes.[9]

Nowadays, more than 150 post-transcriptional modifications are described in the literature and referenced in the Modomics database[10] for all types of RNA. The variety of modifications goes from simple chemical groups such as methylations, which are the most naturally abundant, to more complex modifications such as total rearrangements of the nucleic bases.

One of the methods enabling the characterization of those modifications is mass spectrometry (MS), especially tandem mass spectrometry (MS/MS). Over the years, MS has become a benchmark method for the characterization of biomolecules owing to its high sensitivity, high specificity, and ability to obtain structural information. For RNA characterization, MS is usually hyphenated with a separation method, like liquid chromatography (LC)[11−28] or capillary electrophoresis (CE).[29] Such hyphenations enable lower sample consumption, isomer separation, and higher sensitivity, making them suitable, and essential, techniques for the characterization of RNA post-transcriptional modifications.

Characterization of RNA modifications with CE-MS/MS or LC−MS/MS is based on a two-level strategy. Briefly, the first level consists in identifying all the different post-transcriptional modifications present in the sample after a total hydrolysis of RNA followed by dephosphorylation to obtain nucleosides.[16−25,29] In the second level, the identified modifications

Received: October 26, 2022
Accepted: December 21, 2022

are then precisely mapped on the sequence after specific RNase digestions to obtain oligonucleotides.[15,23−25,29]

To get information on the modified positions on the nucleosides, collision-induced dissociation (CID) is classically used for MS/MS fragmentation. Indeed, CID leads to the fragmentation of the N-glycosidic bond and thus to the neutral loss of the modified or unmodified ribose, enabling the discrimination between a modified base or a modified ribose.[29,30]

Two modes have been reported for MS/MS analyses of nucleotides. The most common is the multiple reaction monitoring (MRM) mode.[11−14,16,18−23] MRM enables fast and easy data interpretation, including for quantification studies.[20−22] However, MRM is a targeted strategy that requires a pre-established list of modified and non-modified nucleosides of interest to be monitored. If one modification is not well predicted and monitored, it will not be detected. Thus, MRM is not really suited for untargeted analyses. The second one, much less common, is the classical MS/MS using data-dependent acquisition mode.[29] This mode enables identification of RNA modification without any preconceptions. However, to our knowledge, no software is yet available for the untargeted identification of RNA post-transcriptional modifications from classical data dependent tandem MS data. Therefore, a long and tedious manual interpretation is required.

To meet the need in bioinformatic tools to fasten and ease data interpretation of nucleosides analyses, a new user-friendly search engine, called Nucleos'ID, was developed. After entering different evaluation parameters, like MS and MS/MS mass tolerances, MS/MS absolute intensity threshold, MS/MS score threshold, and active exclusion time, the software will compare experimental MS/MS data with a database to identify modifications. Each identified modification will then be scored, and results will be saved in different formats (CSV and XLSX) after being sorted to ease the user's reading. All modifications described in the literature and referenced in the Modomics database[10] are considered, thus enabling untargeted searches.

EXPERIMENTAL SECTION

Chemicals. The chemicals used were of analytical grade or high purity. Water was obtained using a Direct-Q3 UV-R water purification system (Merck Millipore, Amsterdam, The Netherlands). Nuclease P1 from *Penicillium citrinum*, ammonium acetate, and zinc chloride ($ZnCl_2$) were purchased from Sigma−Aldrich (Saint Louis, MO, U.S.A.). Glacial acetic acid and bacterial alkaline phosphatase (BAP) (150 U/μL) from *Escherichia coli* C90 were purchased from ThermoFisher Scientific (Marietta, OH, U.S.A.).

RNA Samples. Two RNA samples were used to evaluate the results obtained with Nucleos'ID. Those samples are a purified $tRNA^{Phe\ GAA}$ from *Saccharomyces cerevisiae* and a total tRNA extract from *Bos taurus*. The purity of both samples was controlled by polyacrylamide gel electrophoresis. In addition to the four canonical nucleosides (A, C, G, and U), the *S. cerevisiae* $tRNA^{Phe\ GAA}$ is composed of 11 post-transcriptional modifications, and the *B. taurus* total tRNA extract is composed of 27 modifications. Those modifications are well known and already described in the literature (Table 1).[10] Thus, both samples are good models of increasing complexity to evaluate the performance of the software. A third sample, whose modifications are not described in the literature yet, was analyzed to apply Nucleos'ID on a more complex case. This sample is a purified 70S ribosome containing 5S, 16S, and 23S ribosomal RNA (rRNA) from *Pseudomonas aeruginosa*. The purity of this sample was controlled by agarose gel electrophoresis. All nucleosides described in this study are annotated

Table 1. List of Modified and Non-modified Nucleosides Reported in the Modomics Database[10] for *S. cerevisiae* $tRNA^{Phe\ GAA}$ and *B. taurus* Total tRNA Extract[a]

Nucleoside	S. cerevisiae $tRNA^{Phe\ GAA}$	*B. taurus* total tRNA extract
A	✓	✓
C	✓	✓
G	✓	✓
U	✓	✓
m^1A	✓	✓
t^6A	-	✓
i^6A	-	✓
I	-	✓
m^3C	-	✓
m^5C	✓	✓
Cm	✓	✓
ac^4C	-	✓
f^5Cm	-	✓
m^1G	-	✓
m^2G	✓	✓
m^7G	✓	✓
Gm	✓	✓
$m^{2,2}G$	✓	✓
Q	-	✓
galQ	-	✓
manQ	-	✓
yW	✓	-
o_2yW	-	✓
Y	✓	✓
D	✓	✓
Um	-	✓
m^5U	✓	✓
Ym	-	✓
m^5Um	-	✓
mcm^5U	-	✓
mcm^5s^2U	-	✓
acp^3U	-	✓

[a]Modified RNA sequences for both tRNA extracts are available in the Supporting Information (Figures S1 and S2).

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

with their short name. All full names are available in the Supporting Information (Table S1).

Sample Preparation. First, three successive precipitations with ammonium acetate and ethanol were performed to desalt RNA samples. Then, nuclease P1 and BAP were used to perform total hydrolysis of RNA, following a previously described protocol.[29] Briefly, P1 10× buffer (2 mM $ZnCl_2$, 250 mM ammonium acetate, pH 5.0) and RNA samples were successively diluted into deionized water before adding 0.5 U of P1 (previously prepared at 0.5 U/μL in 100 mM ammonium acetate) per μg of tRNA. A first incubation was carried out at 37 °C for 2 h. After adding 7.5 U of BAP (previously prepared at 1.5 U/μL in 100 mM ammonium acetate) for each μg of RNA, a second incubation was carried out at 37 °C for 4 h.

Finally, samples were fully evaporated and dissolved in ammonium acetate (200 mM, pH 4) to a final concentration of 0.05 μg/μL for the $tRNA^{Phe\ GAA}$ from *S. cerevisiae*, 0.1 μg/μL for the total tRNA extract from *B. taurus*, and 1.0 μg/μL for the 70S ribosome of *P. aeruginosa*.

Capillary Electrophoresis-Tandem Mass Spectrometry (CE-MS/MS). A CESI8000 capillary electrophoresis system (Sciex Separation, Brea, CA, USA) was hyphenated to a maXis 4G mass spectrometer (Bruker Daltonics, Bremen, Germany) owing to a porous tip sheathless interface. CE-MS/MS analyses were performed following a previously described method.[29] Briefly, 5 nL of sample was injected into the CE capillary and CE separation was carried out by applying +30 kV voltage on a bare-fused silica capillary (100 cm total length, 30 μm i.d.) filled with 10% acetic acid as a background electrolyte. MS detection was performed in ion positive mode over the range 80−800 *m/z*, with an accumulation time of 1 s and a resolution of 49,426 at 410 *m/z*. Data-dependent MS/MS was performed using an accumulation time from 0.14 to 0.56 s depending on the number of precursor ions, CID fragmentation mode, and a collision energy of 20 eV for all precursor ions. From three to seven most intense precursor ions were selected in the quadrupole with a window of 1.5 *m/z*. Dynamic exclusion was activated for 0.7 min after each spectrum. Finally, MS/MS data were converted into MGF (Mascot Generic Format) files for input in Nucleos'ID. The conversion was performed using DataAnalysis software (Bruker Daltonics, Bremen, Germany). However, MSconvert can also be used.

Nucleos'ID Workflow. Nucleos'ID is a software developed in Python 3 and providing an efficient and sober user interface. It is available for Linux, Mac OS X, and Windows operating systems. Both executables and source code can be downloaded from the Github repository (https://github.com/MSARN/NucleosID). The code is distributed under an open-source license (Apache 2) and is registered on Zenodo.[31]

The global workflow for data analysis and data interpretation using Nucleos'ID is presented in Figure 1. For the identification of nucleosides, RNA is reduced into nucleosides before CE-MS/MS or LC−MS/MS analysis. MS/MS data are then converted to an MGF file for interpretation with the software. After suppressing the background noise, Nucleos'ID compares MS/MS data with a database to identify modifications. Each identified modification is then scored, and results are displayed after being sorted to ease the user's reading. Nucleos'ID can be used with low- or high-resolution MS/MS data; however, the best results would be obtained

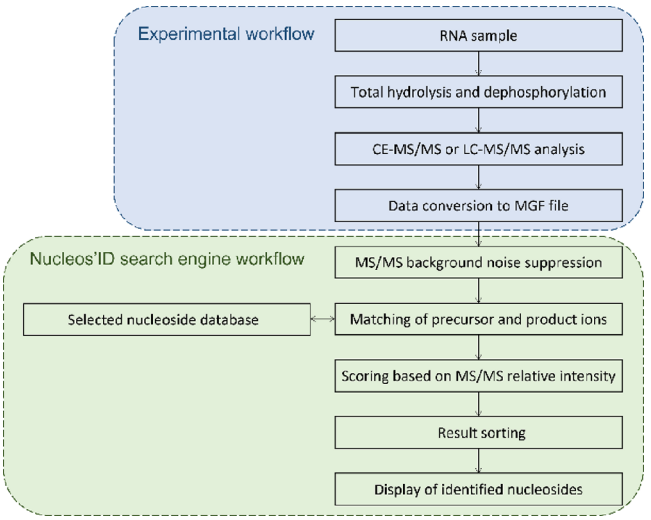


Figure 1. Workflow of identification of RNA modifications by CE-MS/MS or LC−MS/MS.

while using with MS resolution higher than 25,000 and mass errors lower than 30 ppm.

Databases. Three databases, recalculated from the Modomics database,[10] are available depending on the primary phylogenetic domain of the sample (Eubacteria, Archaea, and Eukarya).[32,33] They can be reused as long as the attribution to Modomics is cited. The databases will be kept up to date in the future owing to the release of new versions. Those databases contain *m/z* values of precursor ions (+1 charge state) and product ions (+1 charge state) generated by CID for all non-modified and modified nucleosides (thus, after dephosphorylation). Most nucleosides only have one product ions in the database because the N-glycosidic linkage is the one that is preferentially fragmented by CID. Moreover, several modifications, such as m^1G, m^2G, and m^7G, are grouped into sets as they share the same precursor and product ions. Therefore, it is important to know that the discrimination of isomers grouped into a set is not possible with Nucleos'ID. The user can select one of those databases, or a combination of two or three of them, for the search. By default, a combination of the three databases is considered.

MS/MS Background Noise Suppression. The first step of the Nucleos'ID workflow is a cleaning procedure of the MS/MS background noise. This process removes, in the MGF file, all product ions below an absolute intensity threshold defined by the user. By default, the absolute intensity threshold is set to 0, such as no cleaning is performed. Indeed, noise is instrument-dependent, and a wrong threshold setting can lead to many false negatives. To estimate this threshold, the intensity of MS/MS spectra with only low intensity background noise must be manually checked. For instance, with our setup, a threshold around 500 AU could have been set (one MS/MS spectrum used for this estimation is available in Figure S3).

When used, this step also enables the reduction of the number of comparisons. Therefore, the execution time and the needed computer resources are also reduced.

Modification Matching. To identify post-transcriptional modifications, MS/MS data are compared to a database containing the *m/z* of precursor ions and products ions. Once a precursor ion is matching experimentally and theoretically within MS mass tolerance, all the product ions from this

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

precursor are searched in the MS/MS spectra within MS/MS mass tolerance. A modification is identified when the precursor ion and at least one of the product ions are matching. MS and MS/MS mass tolerances used for the comparison are defined by the user, either in Da or in ppm.[34] As the result file is well detailed and includes accurate measured m/z of all matching precursor and product ions, it is possible to reprocess the results afterward with lower mass tolerances.

Scoring System. As with any data interpretation software, it is essential to associate a confidence score to each result. This score is used to differentiate which result can be a false positive among all the true positives. To our knowledge, scoring systems in proteomics or oligonucleotides sequencing are always probability-based. Indeed, the number of matched fragments among all theoretical fragments is used to calculate a probability that the match is only due to chance.[35−38] However, in the case of nucleosides, a similar scoring function would not be representative since most of them have only one specific product ion after CID fragmentation. Moreover, for some modifications, all MS/MS fragments in the database are not always detected. These are the reasons why our scoring system is solely based on the MS/MS relative intensity of the most intensive matched product ion. For instance, if a matched product ion is the most intense peak of its MS/MS spectrum, the score associated with this modification will be 100%.

Result Sorting. Before saving the output, results are filtered according to two user-defined criteria. The first sort is based on the score. If a result has a score below the user-defined threshold, it will not be displayed to reduce the number of false positives. The second sorting is done to reduce the redundancy of the modifications identified in a short period of time. Indeed, if the same modification is selected several times to be fragmented in MS/MS, it will be identified as many times by the software. For this, the user can define an exclusion time. If the same modification is identified several times during this exclusion time, then only the result with the best MS/MS intensity is selected. Both filters make it convenient to select the most significant matches.

Result Saving. Final results are then saved using one of the two available file formats (CSV or XLSX). By default, the CSV format is proposed as it enables easier additional post-processing of the data and can also be open by any spreadsheet software.

Data Analysis. Data interpretation was performed with Nucleos'ID software on a Dell Inc. laptop using Windows 10 operating system with an Intel Core i5 processor and 8 GB RAM. Nucleos'ID search was performed with default parameters for the purified *S. cerevisiae* $tRNA^{Phe\ GAA}$ and *B. taurus* total tRNA extract. The default settings are presented on the user-friendly interface in Figure 2. For the *P. aeruginosa* 70S ribosome, the score threshold was set to 0% to ensure that no modification was missed as none are described in the literature for this sample, and only the Eubacteria database was used. Manual search of all nucleosides in the database was also performed, and results were compared.

Evaluation of True Positive, True Negative, and False Discovery Rates. To assess the performances of the new search engine, three common rates based on the confusion matrix containing the numbers of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results (Table 2) were chosen. Those rates are the true positive rate, the true negative rate, and the false discovery rate.[39] They

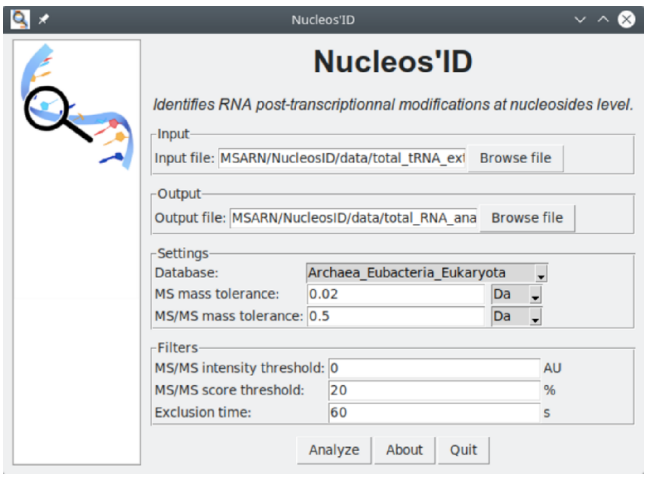


Figure 2. User-friendly interface of Nucleos'ID with its default settings.

Table 2. Confusion Matrix for the Classification of True Positives, True Negatives, False Positives, and False Negatives Depending on the Nucleos'ID Results and the Actual Nucleosides Content of the Sample

		actual nucleoside content of the sample	
		present nucleoside	non-present nucleoside
Nucleos'ID search	identified nucleoside	true positive (TP)	false positive (FP)
	nonidentified nucleoside	false negative (FN)	true negative (TN)

are commonly used as statistical metrics to assess software performances.[38,40,41]

To identify the actual nucleoside content of the sample, manual interpretation was performed using the same databases (Eubacteria, Archaea, and Eukarya) as the Nucleos'ID search. Each set of nucleosides from the Nucleos'ID search database was counted as 1 true positive, true negative, false positive, or false negative. When using all three databases for the search, a total of 119 sets were considered. When using only the Eubacteria database for the search, a total of 69 sets were considered. As only the manual interpretation enables the discrimination of some isomers with similar fragmentation pattern (based on electrophoretic mobility comparison), only the identifications of the sets are compared with the Nucleos'ID search.

The true positive rate (TPR), also called sensitivity, indicates the probability that an actual nucleoside present in the sample will be correctly identified during the Nucleos'ID search. To avoid missing a positive result, TPR must be as high as possible. TPR was calculated with eq 1:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

The true negative rate (TNR), also called specificity, indicates the probability that a nucleoside that is not present in the sample will correctly be undetected with the Nucleos'ID search. TNR was calculated with Eq 2:

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{2}$$

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

Table 3. Output CSV File Imported in Excel for the Nucleos'ID Search of *S. cerevisiae* tRNA$^{Phe\ GAA}$

modification	observed MS (Da)	theoretical MS (Da)	observed MS/MS (Da)	theoretical MS/MS (Da)	score (%)	detection time (s)
C	244.0933	244.0928	112.05079	112.050538	100	959.703
Cm	258.1079	258.1084	112.05057	112.050538	22.59	975.819
m3C/m4C/m5C	258.1079	258.1084	126.06634	126.066188	33.6	978.505
m1G/m2G/m7G	298.1136	298.1146	166.0726	166.072336	100	981.19
m1A/m2A/m6A/m8A	282.1189	282.1197	150.0776	150.077422	100	981.469
A	268.1053	268.104	136.06332	136.061772	100	991.932
m2,2g/m2,7G/preQ1	312.1292	312.1302	180.08824	180.087986	100	1236.39
yW	509.1937	509.1991	377.15438	377.156794	100	1239.07
m1G/m2G/m7G	298.1137	298.1146	166.07272	166.072336	100	1265.93
Cnm5U	284.0981	284.0877	152.05712	152.045453	100	1327.72
G	284.0981	284.0989	152.05712	152.056686	100	1327.72
Gm	298.1136	298.1146	152.05691	152.056686	100	1359.95

Both TPR and TNR take on values between 0 and 1. A TPR equal to 1 means that all the nucleosides present in the sample have been identified by the software. Similarly, a TNR equal to 1 means that all the nucleosides that are not present in the sample have not been identified by the Nucleos'ID search.

The false discovery rate (FDR) indicates the proportion of false positives among all significant results identified by Nucleos'ID. Usually, in proteomics and transcriptomics, the FDR is calculated using a decoy sequence database.[41] However, for the identification at the nucleosidic level, no decoy database can be constructed. Therefore, we decided, as mentioned before, to use only the numbers of false positives and true positives from the selected database. The FDR was calculated with eq 3:

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TP}} \tag{3}$$

The FDR also takes on values between 0 and 1. An FDR equal to 0 means that all the nucleosides identified by the software are present in the sample. If, for instance, an FDR is equal to 0.10, 10% of the nucleosides identified by Nucleos'ID are false positives and thus not present in the sample.

To summarize, high TPR and TNR, and low FDR, indicate a high level of confidence in the results obtained with the Nucleos'ID search.

RESULTS AND DISCUSSION

Evaluation of the Nucleos'ID Software on a Non-complex Sample: Identification of Post-transcriptional Modifications from *S. cerevisiae* tRNA$^{Phe\ GAA}$. The output file of the Nucleos'ID search is presented in Table 3. To evaluate the results obtained from Nucleos'ID search, all expected and unexpected nucleosides were checked with manual interpretation. The electropherogram obtained with manual interpretation is presented in Figure 3.

First, both interpretation methods, Nucleos'ID search and manual interpretation, identified the four canonical nucleosides and seven expected sets of post-transcriptional modifications (Table 4). Those results are considered as true positives. Second, the expected modifications identified manually but not found by the software are modifications that were not selected for fragmentation due to low MS relative intensity. As no MS/MS spectrum is available in the input MGF file, it is therefore normal that they do not appear in the Nucleos'ID search results. Thus, those four modifications are not false negatives and can be considered as true negatives.

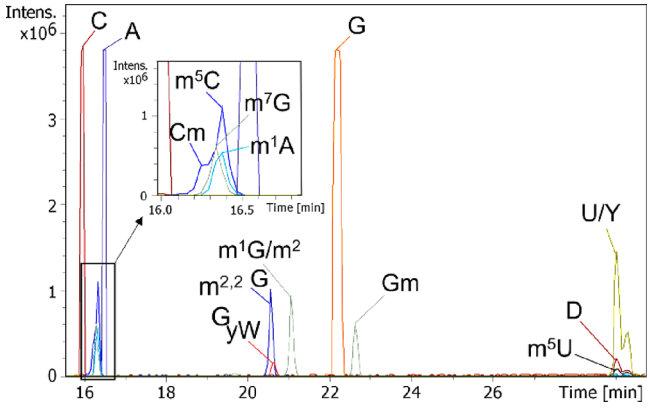


Figure 3. Multiple extracted ion electropherograms for the nucleosides analysis of tRNA$^{Phe\ GAA}$ from *S. cerevisiae* obtained after manual interpretation. Nucleosides annotated with asterisks (*) are expected nucleosides that were only detected by MS as they were not selected for fragmentation.

Finally, only one additional modification, cnm^{5}U, was found by Nucleos'ID when using 0.02 Da of mass tolerance (Table 3) but was not identified during the manual interpretation (Figure 3). cnm^{5}U was not expected as it is not annotated on the sequence of *S. cerevisiae* tRNA$^{Phe\ GAA}$ (Figure S1) referenced in the Modomics database.[10] The modification cnm^{5}U was identified at a detection time of 22.1 min (1327.72 s), which is the same detection time of G. The mass differences between the theoretical precursor ions, and between the theoretical product ions, of cnm^{5}U and G are, respectively, below 0.02 and 0.5 Da (0.011233 Da for both to be exact). Thus, with MS and MS/MS mass tolerances of, respectively, 0.02 and 0.5 Da, Nucleos'ID software does not differentiate the two nucleosides. With MS and MS/MS mass tolerances set at 0.02 and 0.5 Da, cnm^{5}U is then considered as a false positive. For instance, setting the MS mass tolerance below 0.01 Da could have been useful to enable differentiation of both nucleosides.

With a total of 10 true positives, 108 true negatives, 1 false positive, and 0 false negative, the TPR of this Nucleos'ID search is 1.00, the TNR is 0.99, and the FDR is 0.09 with the default parameters. These excellent rates indicate great confidence in the results obtained by the software and show that a high level of confidence can be given to these results. To reduce the number of false positives, the search parameters can be adjusted depending on the performances of the mass spectrometer. For instance, lowering the mass tolerance to less

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

Table 4. Identification after Manual Interpretation and Nucleos'ID Search of Expected Nucleosides *S. cerevisiae* tRNA$^{Phe\ GAA}$[a]

	A	C	G	U	m^1A	m^5C	Cm	m^2G / m^7G	Gm	$m^{2,2}G$	yW	Y	D	m^5U
Manual	✓	✓	✓	≈	✓	✓	✓	✓	✓	✓	✓	≈	≈	≈
Nucleos'ID	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	×	×	×

[a]The check symbol shows that the identification was based on MS and MS/MS data. The almost equal to symbol shows that the manual identification was only based on MS due to no selection of those nucleosides for MS/MS fragmentation. Nucleosides identified solely on MS are not considered as valid as they cannot be differentiated from their isomers. The multiplication symbol shows that the nucleoside was not identified.

Table 5. Identification after Manual Interpretation and Nucleos'ID Search of Expected Nucleosides in *B. taurus* Total tRNA Extract[a]

	A	C	G	U	m^1A	t^6A	i^6A	I	m^3C / m^5C	Cm	ac^4C	f^5Cm	m^1G / m^2G / m^7G	Gm	$m^{2,2}G$	Q	galQ/manQ	o_2yW	Y	D	m^5U	Um	Ym	m^5Um	mcm^5U	mcm^5s^2U	acp^3U
Manual	✓	✓	≈	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓	≈	✓	✓	✓	≈	≈	≈	✓
Nucleos'ID	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓	×	✓	✓	×	×	×	×	✓

[a]The check symbol shows that the identification was based on MS and MS/MS data. The almost equal to symbol shows that the manual identification was only based on MS due to no selection for MS/MS fragmentation. Nucleosides identified solely on MS are not considered as valid as they cannot be differentiated from their isomers. The multiplication symbol shows that the nucleoside was not identified.

than 0.01 Da could withdraw the false positive cnm^5U and thus a true negative rate of 1.00 could be achieved.

Increase in Sample Complexity: Identification of Post-transcriptional Modifications from *B. taurus* Total tRNA Extract. As in the previous example, to evaluate the results obtained from Nucleos'ID search (output file from Nucleos'ID search available in the Supporting Information in Table S2), all expected and unexpected nucleosides were checked with manual interpretation (the electropherogram obtained with manual interpretation is available in the Supporting Information in Figure S4). The comparison from both interpretation methods is summarized in Table 5 for expected modifications and in Table 6 for unexpected modifications found by Nucleos'ID.

Among the expected unmodified and modified nucleosides for *B. taurus* total tRNA extract, 19 sets of nucleosides have been identified manually and with the Nucleos'ID search. Those results are considered as true positives. One of those modifications(I) was detected at two different times (Table S2) by Nucleos'ID despite having no expected isomer. The second one is a true positive and the first one is due to the first isotope of A. The mass differences between the theoretical precursor ions, and between the theoretical product ions, of I and the first isotope of A are, respectively, below 0.02 and 0.5 Da. As I was truly detected manually at the second detection time, this modification is not considered as a false positive.

In addition, two expected modifications (f^5Cm and o_2yW) were not identified either manually or by the software. This can be explained either by a degradation of RNA during storage, or by their low concentration in the sample, which could be below the detection limit. Indeed, on the one hand, f^5Cm is only present in $tRNA^{Leu\ CAA}$, which is already in low abundance, and this modification can be partial, meaning that a portion of this specific position could also be modified with Cm. On the other hand, the 37th position of $tRNA^{Phe\ GAA}$ must be modified with yW or one of its derivatives, such as o_2yW.

Table 6. Identification after Manual Interpretation and Nucleos'ID Search of Unexpected Nucleosides in *B. taurus* Total tRNA Extract[a]

Nucleoside	Nucleos'ID search	Manual search	Number of detected isomers[b]
hm^5C / nm^5U	✓	×	×
hm^5Cm	✓	✓	1
$m^{2,8}A$ / $m^{6,6}A$	✓	✓	1
ms^2m^6A	✓	✓	1
ac^6A	✓	✓	1
hn^6A / m^6t^6A	✓	✓	1
hm^6A	✓	✓	1
$cmnm^5s^2U$	✓	×	×
m^1Y / m^3Y	✓	✓	1
acp^3Y	✓	?	?
acp^3D	✓	✓	1

[a]The check symbol shows that the identification was based on MS and MS/MS data. The multiplication symbol shows that the nucleoside was not identified manually. The question mark symbol indicates an ambiguous case that is considered as not valid. [b]The number of detected isomers is based on the manual interpretation.

During manual interpretation, yW was found with very low MS signal intensity and thus was not selected for fragmentation. As those modifications were not detected at all with Nucleos'ID and manual searches, they are considered as true negatives.

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

Only one expected modification (Ym) was identified manually but not identified with the Nucleos'ID search. This is due to the low relative intensity of its MS/MS fragment due to the co-migration with many nucleosides, including isomeric ones. Indeed, when several isomeric nucleosides are co-migrating, all isomers are selected for fragmentation at the same time and combined MS/MS spectra are obtained. Matching of several isomers to one combined MS/MS spectrum is possible; however, it potentially leads to a low MS/MS relative intensity, thus a low score, for some of the isomers. In the case of Ym, as its MS/MS relative intensity (3.2%) is lower than the default score threshold (20%), this result was withdrawn from the output file during the result sorting. This modification is considered as a false negative.

The remaining five expected nucleosides were not identified by the software since they were not selected for MS/MS fragmentation. As mentioned in the previous example, these modifications are also considered as true negatives.

Table 6 shows that 11 sets of unexpected modifications have been identified by the Nucleos'ID search. Among those unexpected modifications, eight sets were also found manually and are considered as true positives. The identification of unexpected modifications can be explained, either by the low amount of rRNA in the tRNA extract despite purification, or by potentially missing tRNA sequences for *B. taurus* in the Modomics database.[10] Two other sets of unexpected modifications (hm^5C/nm^5U and $cmnm^5s^2U$) have not been identified manually as their MS/MS intensities were too low to be considered as a real signal, compared to the unusual intensity of the background noise in those MS/MS spectra.

Finally, one modification (acp^3Y) was identified by the Nucleos'ID search. However, manual interpretation indicated that this is an ambiguous case. Indeed, according to the Modomics database,[10] acp^3Y and acp^3U share the same *m/z* for their precursor ions as well as for the theoretical product ion. The modification acp^3Y has only one theoretical product ion, the one in common, while acp^3U has additional theoretical product ions. Those additional product ions have been detected during the CE-MS/MS analysis. Therefore, there is no confident information leading to an unambiguous identification of acp^3Y. Due to this ambiguity, acp^3Y was considered as a false positive.

With a total of 27 true positives, 88 true negatives, 3 false positive, and 1 false negative, the TPR of this Nucleos'ID search is 0.96, the TNR is 0.97, and the FDR is 0.10. Although the sample is more complex than the previous *S. cerevisiae* $tRNA^{Phe}_{GAA}$ (more numerous and diverse modifications), the TPR, TNR, and FDR are still very good. Once again, search parameters can be optimized to minimize the probability of obtaining false positives. For instance, in this case, the absolute intensity threshold, which is set to 0 by default, can be a powerful tool to eliminate some false positives. However, if the absolute intensity threshold is too high, it could lead to more false negatives.

To assess the suitability of the default score threshold (set at 20%), the same research was performed using thresholds from 0 to 100%. The number of true positives, false positives, true negatives, and false negatives as well as the TPR, TNR, and FDR obtained with these different thresholds are presented in Figure 4.

With a score threshold at 0%, there is 0 false negative (Figure 4D) and the number of true positives is the highest (Figure 4A), meaning that all nucleosides that should be found

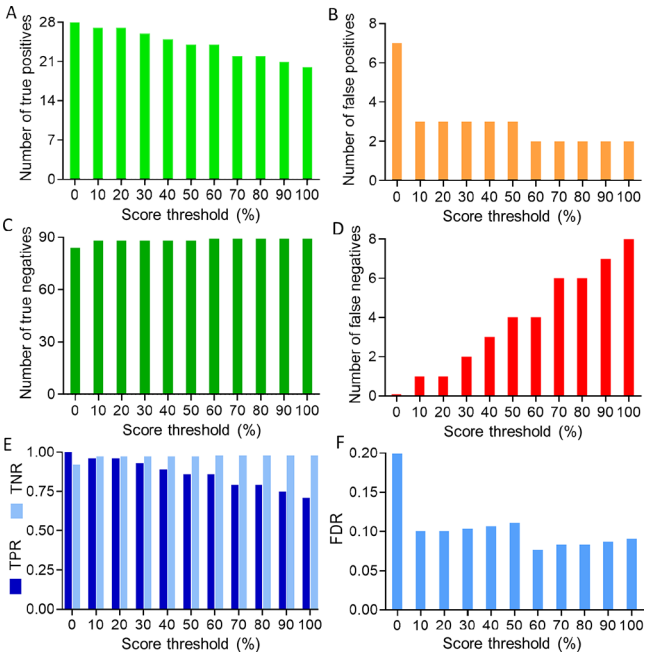


Figure 4. Evolution of the numbers of (A) true positives, (B) false positives, (C) true negatives, and (D) false negatives as well as (E) the TPR and the TNR and (F) the FDR depending on the score threshold used for the Nucleos'ID search for *B. taurus* total tRNA extract.

are identified. However, the number of false positives, and thus the FDR, are quite high (Figure 4B,F). On the contrary, with a score threshold at 100%, the number of true negatives is the highest (Figure 4C), there are fewer false positives (2 instead of 7), and the FDR is way lower (0.09 instead of 0.20). However, the number of true positives is much lower (20 instead of 28), and there are many more false negatives (8 instead of 0). This means that the results found during the Nucleos'ID search are unlikely to be wrong, but some nucleosides will be missed by the software. Thus, finding the right score threshold is all about compromises.

To identify the best compromise, the score threshold should provide the best TPR and TNR. Score thresholds of 10 and 20% gave the highest TPR and TNR (Figure 4E). Thus, thresholds of 10 or 20% are the best compromise. As it appeared to be a good choice for many samples, the default score threshold was set to 20%.

Untargeted Identification of RNA Post-transcriptional Modifications of *P. aeruginosa* 70S Ribosome. To increase sample complexity after the first result evaluation, the Nucleos'ID search engine was used to identify RNA post-transcriptional modifications of *P. aeruginosa* 70S ribosome. Indeed, only well described samples were previously used to evaluate the software, whereas RNA modifications from *P. aeruginosa* 70S ribosome are not described in the literature yet.

As *P. aeruginosa* is part of the Eubacteria domain, an untargeted search was performed on the Eubacteria database. To avoid missing any modification, the score threshold was set to 0%, and all results obtained in the output file (available in the Supporting Information in Table S3) were compared to manual search (the electropherogram obtained with manual interpretation is available in the Supporting Information in Figure S5). No additional nucleosides were found manually, and thus no false negative was obtained. All modified and non-

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

modified nucleosides found by the Nucleos'ID search are summarized in Table 7.

Table 7. Manual Verification of All Nucleosides Found by Nucleos'ID Search in *P. aeruginosa* 70S Ribosome

Abbreviations	Nucleos'ID score (%)	Manual search	Number of detected isomers[b]
A	100	✓	1
C	100	✓	1
G	100	✓	1
U	100	✓	1
m^1A / m^2A / m^6A/ m^8A	100	✓	2
$m^{2,8}A$ / $m^{6,6}A$	100	✓	1
m^3C / m^4C / m^5C	100	✓	1
m^4Cm	100	✓	1
m^1G / m^2G / m^7G	100	✓	2
Gm	100	✓	1
Q	100	✓	1
m^3U / m^5U	100	✓	1
Y	43	✓	1
ncm^5U	90	✓	1
acp^3U	100	✓	1
m^3Y	14	×	×
$cmnm^5Um$	55	×	×
cmo^5U	6	×	×

[a]The check symbol shows that the identification was based on MS and MS/MS data. The multiplication symbol shows that the nucleoside was not identified manually. [b]The number of detected isomers is based on the manual interpretation.

In addition to the 4 canonical nucleosides, 11 sets of modifications detected by the software were confirmed by manual interpretation. All of them were considered as true positives. The last three modifications (m^3Y, $cmnm^5Um$ and cmo^5U) were considered as false positives as their product ions absolute intensities were too low to be considered as a real signal and not as background noise.

Even though, RNA modifications in *P. aeruginosa* 70S ribosome have not been described yet, those of others bacterial species, such as *E. coli*, have already been reported in the literature.[42,43] As species from the same domain, here the Eubacteria, usually show similarities in their RNA post-transcriptional modifications, it is interesting to compare our results to those from *E. coli* ribosome.

All the identified sets of modifications except Q, ncm^5U, and acp^3U are similar to those described in *E. coli* 16S or 23S rRNA.[43] Only three additional modifications (D, Um, and Cm) are reported in *E. coli*. D was not identified with MS/MS due the very intense signal of U. Um and Cm were not detected at all in our analysis of *P. aeruginosa* 70S ribosome.

Finally, it is important to acknowledge that Q, ncm^5U, and acp^3U were detected with very low signal intensities in *P. aeruginosa* 70S ribosome and are usually found in bacterial tRNA, but not in bacterial rRNA. As ribosome is where the translation of proteins takes place, some tRNA can be trapped between both 30S and 50S ribosome subunits during the purification process, leading to a very small proportion of tRNA in the sample.

With a total of 15 true positives, 51 true negatives, 3 false positives, and 0 false negative, the TPR of this Nucleos'ID search is 1.00, the TNR is 0.94, and the FDR is 0.17. Once again, the TPR and TNR are very good, and the FDR is also quite good considering the relatively small size of the database.

As mentioned before, search parameters could be optimized to reduce the number of false positives or to increase the number of true negatives. Nevertheless, it is important to remember that, while parameter optimization can be useful in reducing the number of false positives, it can also lead to a loss of information and an increase in false negative. For instance, for the *P. aeruginosa* 70S ribosome, as none of the true positives has a score below 20%, a lower FDR could have been achieved by using the default score threshold. However, using a very high score threshold of 75% would lead to three less false positives (m^3Y, $cmnm^5Um$, and cmo^5U), but one true positive (Y) would be lost. The right compromise must be found by the user depending not only on the performance of the mass spectrometer and on nature and knowledge of the sample but also on the aim of the Nucleos'ID search. For this sample, the aim of the Nucleos'ID search was to save time in the data interpretation before manual checking the results to avoid any missed modifications.

The great results obtained for all three samples demonstrate the strong potential of the Nucleos'ID software for the untargeted identification of post-transcriptional modifications after MS/MS analyses. Indeed, from several hours for manual interpretation if all modifications from the three databases are searched to less than 1 s for the Nucleos'ID search engine, the gain in time is impressive, as for instance, results showed in this article are based on MGF files containing from around 500 spectra to more than 15,600 spectra. In case the user wants to manually check the Nucleos'ID results, only a few more minutes are necessary as the number of potential modifications was significantly reduced compared to the full database.

CONCLUSIONS

To conclude, a new software enabling untargeted identification of RNA modifications from MS/MS data was developed and is freely available for CE-MS/MS or LC−MS/MS users. Nucleos'ID was evaluated on two well-known samples and good TPR, TNR, and FDR were achieved. The evaluation was performed using default settings; however, search settings can be tuned, with common sense, to match the mass spectrometer performances and the aim of the study. Nucleos'ID was then used to characterize *P. aeruginosa* 70S ribosome modifications that were not described in the literature yet. In addition to the 4 canonical nucleosides, 11 sets of modifications have been successfully identified. As the results are displayed instantly when the Nucleos'ID search is performed, it is a timesaving tool compared to several hours of untargeted manual interpretation. In addition to providing structural information

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX

on RNA, the use of this new software is undeniably a time saver for the identification of RNA modifications, especially when the sample is not yet described in the literature, even with a manual verification following the Nucleos'ID search.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c04722.

Sequences of studied tRNA, example of MS/MS spectra for MS/MS absolute intensity threshold estimation, additional Nucleos'ID output files, and additional multiple extracted ion electropherograms (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yannis-Nicolas François − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 (Unistra-CNRS), Université de Strasbourg, Strasbourg 67000, France;* orcid.org/0000-0002-8776-071X; Email: yfrancois@unistra.fr

Authors

Clarisse Gosset-Erard − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 (Unistra-CNRS), Université de Strasbourg, Strasbourg 67000, France; Université de Lorraine, LCP-A2MC, F-57000 Metz, France*

Mévie Didierjean − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 (Unistra-CNRS), Université de Strasbourg, Strasbourg 67000, France*

Jérome Pansanel − *Université de Strasbourg, Institut Pluridisciplinaire Hubert Curien (IPHC), CNRS, UMR7178, Strasbourg 67037, France;* orcid.org/0000-0002-7067-5009

Antony Lechner − *Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire, CNRS UPR9002, Université de Strasbourg, Strasbourg 67084, France*

Philippe Wolff − *Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire, CNRS UPR9002, Université de Strasbourg, Strasbourg 67084, France*

Lauriane Kuhn − *Plateforme Protéomique Strasbourg-Esplanade, Institut de Biologie Moléculaire et Cellulaire, FR1589 CNRS, CEDEX, Strasbourg 67084, France*

Frédéric Aubriet − *Université de Lorraine, LCP-A2MC, F-57000 Metz, France;* orcid.org/0000-0003-2457-7974

Emmanuelle Leize-Wagner − *Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS), UMR 7140 (Unistra-CNRS), Université de Strasbourg, Strasbourg 67000, France*

Patrick Chaimbault − *Université de Lorraine, LCP-A2MC, F-57000 Metz, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c04722

Author Contributions

#C.G.-E., M.D., and J.P. contributed equally.

Notes

The authors declare no competing financial interest.

Source code for Nucleos'ID is available on Github (https://github.com/MSARN/NucleosID) under Apache 2 license.

■ ACKNOWLEDGMENTS

This work was supported by a PhD grant from the RESEX project funded by the European Regional Development Fund (FEDER), the general council of Moselle, Region Grand Est, Metz Metropole and the University of Lorraine.

■ REFERENCES

(1) Helm, M. *Nucleic Acids Res.* **2006**, *34*, 721−733.

(2) Chawla, M.; Oliva, R.; Bujnicki, J. M.; Cavallo, L. *Nucleic Acids Res.* **2015**, *43*, 6714−6729.

(3) Agris, P. F.The Importance of Being Modified: Roles of Modified Nucleosides and Mg^{2+} in RNA Structure and Function. In *Progress in Nucleic Acid Research and Molecular Biology*; Cohn, W. E.; Moldave, K., Eds.; Academic Press, 1996; Vol. *53*, pp. 79−129, DOI: 10.1016/S0079-6603(08)60143-9.

(4) Roundtree, I. A.; Evans, M. E.; Pan, T.; He, C. *Cell* **2017**, *169*, 1187−1200.

(5) Wilkinson, E.; Cui, Y.-H.; He, Y.-Y. *Front. Cell Dev. Biol.* **2022**, *10*, No. 828683.

(6) Vester, B.; Long, K. S.*Antibiotic Resistance in Bacteria Caused by Modified Nucleosides in 23S Ribosomal RNA*; Landes Bioscience, 2009.

(7) Jonkhout, N.; Tran, J.; Smith, M. A.; Schonrock, N.; Mattick, J. S.; Novoa, E. M. *RNA* **2017**, *23*, 1754−1769.

(8) Gao, M.; Zhang, Q.; Feng, X.-H.; Liu, J. *Acta Biomater.* **2021**, *131*, 1−15.

(9) Amalric, A.; Bastide, A.; Attina, A.; Choquet, A.; Vialaret, J.; Lehmann, S.; David, A.; Hirtz, C. *Crit. Rev. Clin. Lab. Sci.* **2022**, *59*, 1−18.

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

(11) Raćkowska, E.; Bobrowska-Korczak, B.; Giebułtowicz, J. *J. Chromatogr., B* **2019**, *1128*, No. 121775.

(12) Lo, W.-Y.; Jeng, L.-B.; Lai, C.-C.; Tsai, F.-J.; Lin, C.-T.; Chen, W. T.-L. *Clin. Chim. Acta* **2014**, *428*, 57−62.

(13) Chen, F.; Zhang, F.; Yang, N.; Liu, X. *J. Chromatogr. Sci.* **2014**, *52*, 852−861.

(14) Feng, Y.-J.; You, X.-J.; Ding, J.-H.; Zhang, Y.-F.; Yuan, B.-F.; Feng, Y.-Q. *Anal. Chem.* **2022**, *94*, 4747−4755.

(15) Paulines, M. J.; Wetzel, C.; Limbach, P. A. *J. Mass Spectrom.* **2019**, *54*, 906−914.

(16) Sakaguchi, Y.; Miyauchi, K.; Kang, B.; Suzuki, T.Chapter Two - Nucleoside Analysis by Hydrophilic Interaction Liquid Chromatography Coupled with Mass Spectrometry. In *Methods in Enzymology*; He, C., Ed.; RNA Modification; Academic Press, 2015; Vol. *560*, pp. 19−28, DOI: 10.1016/bs.mie.2015.03.015.

(17) Grobe, S.; Doberenz, S.; Ferreira, K.; Krueger, J.; Brönstrup, M.; Kaever, V.; Häussler, S. *ChemBioChem* **2019**, *20*, 1430−1437.

(18) Liu, R.; Zhao, F.; Wei, J.; Yu, P.; Zhang, J.; Wang, Y.; Li, Z.; Zhang, J.; Zhang, X.; Tian, X. *J. Sep. Sci.* **2019**, *42*, 2668−2678.

(19) Thüring, K.; Schmid, K.; Keller, P.; Helm, M. *Methods* **2016**, *107*, 48−56.

(20) Yan, M.; Wang, Y.; Hu, Y.; Feng, Y.; Dai, C.; Wu, J.; Wu, D.; Zhang, F.; Zhai, Q. *Anal. Chem.* **2013**, *85*, 12173−12181.

(21) Zhu, Y.; Zhou, G.; Yu, X.; Xu, Q.; Wang, K.; Xie, D.; Yang, Q.; Wang, L. *PLoS One* **2017**, *12*, No. e0175849.

(22) Su, D.; Chan, C. T. Y.; Gu, C.; Lim, K. S.; Chionh, Y. H.; McBee, M. E.; Russell, B. S.; Babu, I. R.; Begley, T. J.; Dedon, P. C. *Nat. Protoc.* **2014**, *9*, 828−841.

(23) Wolff, P.; Villette, C.; Zumsteg, J.; Heintz, D.; Antoine, L.; Chane-Woon-Ming, B.; Droogmans, L.; Grosjean, H.; Westhof, E. *RNA* **2020**, 1957.

(24) Cao, X.; Limbach, P. A. *Anal. Chem.* **2015**, *87*, 8433−8440.

(25) Ross, R.; Cao, X.; Yu, N.; Limbach, P. A. *Methods* **2016**, *107*, 73−78.

(26) Dai, Y.; Qi, C.-B.; Feng, Y.; Cheng, Q.-Y.; Liu, F.-L.; Cheng, M.-Y.; Yuan, B.-F.; Feng, Y.-Q. *Anal. Chem.* **2021**, *93*, 6938−6946.
(27) Chen, M.-Y.; Qi, C.-B.; Tang, X.-M.; Ding, J.-H.; Yuan, B.-F.; Feng, Y.-Q. *Chin. Chem. Lett.* **2022**, *33*, 3772−3776.
(28) Chen, M.-Y.; Gui, Z.; Chen, K.-K.; Ding, J.-H.; He, J.-G.; Xiong, J.; Li, J.-L.; Wang, J.; Yuan, B.-F.; Feng, Y.-Q. *Chin. Chem. Lett.* **2022**, *33*, 2086−2090.
(29) Lechner, A.; Wolff, P.; Leize-Wagner, E.; François, Y.-N. *Anal. Chem.* **2020**, *92*, 7363−7370.
(30) Jensen, S. S.; Ariza, X.; Nielsen, P.; Vilarrasa, J.; Kirpekar, F. *J. Mass Spectrom.* **2007**, *42*, 49−57.
(31) Didierjean, M.; Gosset-Erard, C.; François, Y.; Pansanel, J.*MSARN/NucleosID: NucleosID v0.8.0*; 2022, DOI: 10.5281/zenodo.7223374.
(32) McCown, P. J.; Ruszkowska, A.; Kunkler, C. N.; Breger, K.; Hulewicz, J. P.; Wang, M. C.; Springer, N. A.; Brown, J. A. *WIREs RNA* **2020**, *11*, No. e1595.
(33) Grosjean, H.*Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour*; Landes Bioscience, 2013.
(34) Brenton, A. G.; Godfrey, A. R. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1821−1835.
(35) Risk, B. A.; Spitzer, W. J.; Giddings, M. C. *J. Proteome Res.* **2013**, *12*, 3019−3025.
(36) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(37) Yu, N.; Lobue, P. A.; Cao, X.; Limbach, P. A. *Anal. Chem.* **2017**, *89*, 10744−10752.
(38) Nakayama, H.; Akiyama, M.; Taoka, M.; Yamauchi, Y.; Nobe, Y.; Ishikawa, H.; Takahashi, N.; Isobe, T. *Nucleic Acids Res.* **2009**, *37*, No. e47.
(39) Brereton, R. G. *J. Chemom.* **2021**, *35*, No. e3241.
(40) Kapp, E. A.; Schütz, F.; Connolly, L. M.; Chakel, J. A.; Meza, J. E.; Miller, C. A.; Fenyo, D.; Eng, J. K.; Adkins, J. N.; Omenn, G. S.; Simpson, R. J. *Proteomics* **2005**, *5*, 3475−3490.
(41) Aggarwal, S.; Yadav, A. K.False Discovery Rate Estimation in Proteomics. In *Statistical Analysis in Proteomics*; Jung, K., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp. 119−128, DOI: 10.1007/978-1-4939-3106-4_7.
(42) Kowalak, J. A.; Bruenger, E.; McCloskey, J. A. *J. Biol. Chem.* **1995**, *270*, 17758−17764.
(43) Wilson, D. N.; Nierhaus, K. H. *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42*, 187−219.

https://doi.org/10.1021/acs.analchem.2c04722
Anal. Chem. XXXX, XXX, XXX−XXX