Profiling ribonucleotide modifications at full-transcriptome level: a step toward MS-based epitranscriptomics

REBECCA E. ROSE, RYAN QUINN, JACKIE L. SAYRE, and DANIELE FABRIS
The RNA Institute, University at Albany, Albany, New York 12222, USA

ABSTRACT
The elucidation of the biological significance of RNA post-transcriptional modifications is hampered by the dearth of effective high-throughput sequencing approaches for detecting, locating, and tracking their levels as a function of predetermined experimental factors. With the goal of confronting this knowledge gap, we devised a strategy for completing global surveys of all ribonucleotide modifications in a cell, which is based on the analysis of whole cell extracts by direct infusion electrospray ionization mass spectrometry (ESI-MS). Our approach eschews chromatographic separation to promote instead the direct application of MS techniques capable of providing detection, differentiation, and quantification of post-transcriptional modifications (PTMs) in complex ribonucleotide mixtures. Accurate mass analysis was used to carry out database-aided identification of PTMs, whereas multistep tandem mass spectrometry (MSn) and consecutive reaction monitoring (CRM) provided the necessary structural corroboration. We demonstrated that heat-map plots afforded by ion mobility spectrometry mass spectrometry (IMS-MS) can provide comprehensive modification profiles that are unique for different cell types and metabolic states. We showed that isolated tRNA samples can be used as controlled sources of PTMs in standard-additions quantification. Intrinsic internal standards enable direct comparisons of heat-maps obtained under different experimental conditions, thus offering the opportunity to evaluate the global effects of such conditions on the expression levels of all PTMs simultaneously. This type of comparative analysis will be expected to support the investigation of the system biology of RNA modifications, which will be aimed at exploring mutual correlations of their expression levels and providing new valuable insights into their biological significance.

Keywords: RNAomics; global profiling; mass spectrometry; metabolomics; ribonucleotide modifications; epitranscriptomics

INTRODUCTION
A tenet of systems biology is that the behavior of a biological system arises from the complex network of functional interactions between its components. Owing to essential activities in protein synthesis and newly discovered functions in gene regulation, RNA is uniquely positioned in such a network to accurately capture the overall behavior of biological systems, as well as the specific metabolic and epigenetic state of a cell. The RNA building blocks display numerous variations of the four canonical bases, which contribute to defining the breathtaking diversity of structures and functions characteristic of natural RNA (Chang and Varani 1997; Carell et al. 2012). These post-transcriptional modifications (PTMs) are introduced by the activity of specialized enzymes that, in many cases, have been identified and investigated (Ferré-D’Amaré 2003). Over 100 ribonucleotide PTMs and corresponding metabolic pathways are currently described in the RNA Modifications (Limbach et al. 1994; Cantara et al. 2011) and MODOMICS (Dunin-Horkawicz et al. 2006; Machnicka et al. 2012) databases. However, with the exception of a handful of PTMs involved in molecular recognition and stabilization of RNA structure (Kowalak et al. 1994; Ofengand 2002; Helm 2006), their biological function is still largely unknown. The observation that methylation of the 3′ nucleotide protects miRNAs from uridylation, a prelude to exonucleolytic degradation (Li et al. 2005), suggests that many PTMs may act as signals or modulators of vital cellular processes. This type of observation has been made possible by the availability of targeted analytical approaches based on bisulfite chemistry (Frommer et al. 1992; Herman et al. 1996) or specific restriction enzymes (Singer-Sam et al. 1990; Issa et al. 1994), which enable the detection of methylation sites by high-throughput sequencing techniques (Ajay et al. 2011; Koboldt et al. 2013). Unfortunately, there are no high-throughput approaches for the majority of other PTMs. For this reason, their functional elucidation has been severely hampered by the inability...
to detect, locate, and track their levels as a function of predetermined experimental factors.

Mass spectrometry (MS)-based approaches have historically played a determinant role in the discovery and characterization of RNA modifications (McCloskey 1979, 1985; Grain 1990a; Nordhoff et al. 1996). This platform affords the ability to recognize the characteristic mass signatures associated with the different variants, as well as the unique fragmentation patterns necessary to confirm their structures (Banoub and Limbach 2010 and reference therein). High-resolution determinations enable the unambiguous differentiation of mononucleotides with very similar elemental compositions that produce nearly overlapping isotopic distributions (Quinn et al. 2013). At the same time, multistep tandem mass spectrometry (MSn) (Solouki et al. 1996; Collings et al. 2001) and ion mobility spectrometry mass spectrometry (IMS-MS) (von Helden et al. 1995; Clemmer and Jarrold 1997; Verbeck et al. 2002) have been proven capable of tackling mixtures of isobaric mononucleotides that share the same elemental composition, but display different structures (Quinn et al. 2013). These capabilities are exemplified by the analysis of the isomeric species uridine and pseudouridine, which include either an N- or C-glycosidic bond between the pyrimidine ring and ribose unit. This distinctive feature confers different stability to collisional activation, which is substantiated by a greater incidence of base loss from the N- then the C-glycosidic form (Wu and McLuckey 2004). At the same time, the unique conformations associated with the different ring attachments influence their interactions with background gas during IMS-MS analysis, thus enabling unambiguous differentiation even when both isomers are present simultaneously in the same sample (Quinn et al. 2013).

Modified ribonucleotides can be analyzed in complex mixtures obtained by hydrolyzing larger RNA samples into mononucleotide components, which may be further treated with phosphatase to obtain the corresponding nucleosides (Crain 1990b). The ensuing samples are typically resolved by coupling liquid chromatography (Esmans et al. 1998; Chan et al. 2010; Su et al. 2014) or capillary electrophoresis (Apruzzese and Vouros 1998) with MS detection (i.e., LC- and CE-MS, respectively), which are meant to provide separation and reduction of chemical background, while avoiding undesirable analyte bias. In previous work, we investigated the merits of direct infusion electrospray ionization (ESI) (Yamashita and Fenn 1984; Banks et al. 1994) to perform nucleotide analysis in the absence of front-end chromatographic procedures (Quinn et al. 2013). The direct approach was evaluated by using standard samples that contained the canonical ribo- and deoxyribonucleotides with the addition of pseudouridine that constitutes the most abundant variant present in nature (Charette and Gray 2000). We have now extended the investigation to mixtures obtained directly from cell samples that contained the full complement of natural PTMs expressed by the selected organism. We assessed the ability of this approach to unambiguously recognize modified mononucleotides from the remaining background in the absence of chromatography, as well as the possibility of determining their abundance in complex biological samples. The reproducibility of this type of analysis was determined to learn whether this approach could provide comprehensive epitranscriptomic profiles and reveal possible correlations between RNA modifications and specific cellular states.

RESULTS AND DISCUSSION

Direct infusion analysis of cellular nucleotide mixtures

A sensible but deceptively challenging way to reduce sample losses and analyte bias consists of reducing the number of sample-handling steps included in a prospective experimental workflow. We have initially explored the utilization of a classic phenol–chloroform extraction to simultaneously achieve cell lysis and rapid isolation of nucleic acid components (Chomczynski and Sacchi 1987), followed by digestion into separate mononucleotides (Crain 1990b). Owing to the absence of high-resolution separation steps, the final samples were anticipated to contain not only the desired mononucleotide analytes, but also unrelated cellular components carried through the entire workflow (Scheme 1, see Materials and Methods). The complexity of these types of mixtures can be immediately appreciated by examining representative data obtained from a digest of total RNA from a sample of Saccharomyces cerevisiae grown in yeast extract, peptone,
dextrose (YPD) medium, which are shown in Figure 1. A broad distribution of signals with very different intensities covered the entire range between \( m/z \) 300 and 700—the region in which PTMs are typically observed. Abundant signals were readily assigned to the deprotonated molecular ions of canonical ribonucleotides (i.e., \([\text{NMP-H}]^-\), where N indicates any nucleoside) (Quinn et al. 2013). Their experimental masses exhibited an average of \( \sim 100 \) ppb deviation from values calculated from the corresponding elemental compositions, which matched the typical accuracy afforded by these types of LTQ-orbitrap determinations (Quinn et al. 2013).

The complexity of whole-cell extracts was also confronted by utilizing IMS-MS, which enables the differentiation of ions according to their size/conformation. In this technique, ions are dispersed on the time dimension as they travel across a low-pressure region of the instrument (Dwivedi et al. 2006; Lapthorn et al. 2013). The probability of undergoing low-energy collisions with background gas is a function of their conformation, which determines the travel time. The corresponding data are displayed in the form of heat-maps or 3D plots, in which the different dimensions consist of arrival time (\( t_D \)), mass to charge ratio (\( m/z \)), and signal intensity. A representative heat-map obtained from the same yeast extract is shown in Figure 2. The canonical ribonucleotides were immediately recognized on the basis of their characteristic \( m/z \) and \( t_D \) values, as exemplified in the enlarged region of the map. As demonstrated in previous work, the additional dimension facilitated the differentiation of isomeric/isobaric species, such as uridine and pseudouridine, which cannot be discriminated solely by mass analysis (Quinn et al. 2013). Given that the sample was derived directly from a whole-cell extract with no RNA fractionation, the corresponding heat-map provided an immediate and comprehensive representation of all the PTMs present in the cells.

Identification of modified ribonucleotides

The vast majority of the detected signals were readily assigned with the aid of database searching. Initial data reduction followed a conservative approach that eschewed the application of a pre-determined threshold to eliminate background noise on the basis of signal intensity, but relied instead on the detection of recognizable isotopic envelopes to differentiate signal from noise (see Materials and Methods). This task used a deconvolution algorithm included in the instrument’s data system, which was designed to infer the charge state of any given signal from the respective isotopic distribution (Bushberg et al. 2012). When the ESI-MS data shown in Figure 1 were processed, the filtering operation returned only 1206 of the 14,639 entries contained in the initial mass list, which were subsequently used for database searching.

The searches were performed against a database specialized on RNA modifications, as well as a more comprehensive metabolomics registry capable of handling unrelated species present in the workflow carryover. The latter consisted of the METLIN database (hosted and maintained by the Scripps Institute) (Smith et al. 2005; http://metlin.scripps.edu/index.php), which comprises in excess of 75,000 endogenous and exogenous metabolites from a broad selection of living organisms, ranging from bacteria, to plants, to animals and humans. In addition, a non-redundant index of known RNA PTMs was generated in house by combining the information contained in the RNA Modifications Database (hosted by the RNA Institute of University at Albany) (Limbach et al. 1994; Cantara et al. 2011; http://mods.rna.albany.edu/) and MODOMICS (hosted by the International Institute of Molecular and Cell Biology in Warsaw) (Dunin-Horkawicz et al. 2006; Machnicka et al. 2012; http://modomics.genesilico.pl/). After redundant entries were eliminated, we ensured that the mass of each PTM appeared in both the nucleoside and nucleotide form to allow for the recognition of all possible products present in the nuclease digests. For each entry, the mass of the deprotonated and protonated species (i.e.,...
[M−H]− and [M+H]+ were calculated to enable proper matching of data obtained in either polarity. The final custom registry included 254 searchable entries.

A total of 268 database hits were obtained when the reduced experimental data were searched against METLIN, whereas 41 were found in the custom registry (Table 1). The observed experimental masses matched very closely those found in the databases. The majority of hits provided an average deviation between experimental and calculated masses that fell within the accuracy assessed from canonical ribonucleotides, whereas weaker signals displayed slightly higher deviations. The majority of such hits corresponded to modifications typically observed in S. cerevisiae’s ribosomal-RNA (rRNA) or transfer-RNA (tRNA), but 13 of them had not been previously reported for this organism according to the information included in the RNA Modifications and MODOMICS databases (marked with asterisk in Table 1). This observation may be a consequence of the broader scope of these analyses, which was not limited to specific rRNA/tRNA fractions but targeted whole-cell extracts. It is interesting to note that, with only a few exceptions, the majority of the hits afforded by the custom registry were also found in METLIN. This observation provides an indication of the excellent but not absolute overlap between the databases used to search the custom registry. This operation yielded the same database hits produced by the ESI-MS data listed in Table 1. In this case, the experimental values afforded by the canonical ribonucleotides displayed an average of 13 ppm deviation from the theoretical values obtained from their elemental composition, which is consistent with the typical accuracy achieved with this type of instrumentation. In analogy with the ESI-MS data displayed in Figure 1, carryover species contributed significantly to the complexity of the observed heat-map, but did not have any adverse effect on PTM detection.

### Assignment confirmation

A close match between experimental and theoretical masses calculated from known elemental compositions is essential to achieve positive identification. For species of this size, sub-ppm accuracy can greatly reduce the number of possible elemental compositions that could match the experimental data, thus minimizing the risk of erroneous interpretations (Quinn et al. 2013). However, regardless of the accuracy afforded by the available instrumentation, positive identification cannot be based solely on matching mass values, but must receive further corroboration by gas-phase fragmentation data consistent with the putative analyte structure. The facile cleavage of the N-glycosidic bond represents a characteristic dissociation channel that is diagnostic of nucleotide structures (Biemann and McCloskey 1962; Crain 1990a) and can be frequently used to discriminate between isomeric forms present simultaneously in a sample (Quinn et al. 2013). Cleavage products can immediately reveal whether the modifying group may be situated on the phosphoribose or nucleobase moiety of the PTM structure. Further, these first-generation fragments can be submitted to subsequent isolation/activation steps in MS² experiments to obtain additional details on the nature and position of the modification (Quinn et al. 2013).

These points are exemplified by the analysis of the species detected at m/z 376.0684 in Figure 1, which could potentially match different methyl-G isomers (i.e., 377.0762 u neutral mass in Table 1). In anionic mode, characteristic phosphoribose fragments produced by base loss were readily observed upon collisional activation of the deprotonated precursor in the LTQ-orbitrap analyzer (Fig. 3A, see Materials and Methods for details). The fact that such products were detected in both methylated and unmethylated form (the latter

### TABLE 1. Hits obtained by searching the ESI-MS data in Figure 1 against the non-redundant database generated in house

| Exp. mass (u) | Monoisotopic (u) | Hit* |
|---------------|-----------------|------|
| 323.0519      | 323.05185       | C*   |
| 324.0359      | 324.03587       | Ψ*, U* |
| 326.0515      | 326.05152       | D*   |
| 337.0675      | 337.06750       | m1C, m1T, Cm1, m2C* |
| 338.0515      | 338.05152       | mΨ*, Ψm1, m1T, m1Ψ, Ψm2, m1U |
| 347.0631      | 347.06308       | ΨA* |
| 348.0471      | 348.04710       | γ*   |
| 361.0787      | 361.07873       | m1A, m2A, m3A, m4A, Am* |
| 363.0580      | 363.05800       | G*   |
| 365.0623      | 365.06242       | acC, acT, f1Cm1, f1Tm1 |
| 375.0942      | 375.09438       | m1Am*, m1Am1, m2A* |
| 377.0762      | 377.07635       | m1G, m2G, m2G, m3G, m4G |
| 379.0762      | 379.07807       | acCm1, f1 |
| 381.0572      | 381.05733       | ncm1U* |
| 391.0893      | 391.08923       | m1Gm*, m2G, preQ1, m3G, m4G |
| 427.0429      | 427.04505       | cmnm2U |
| 492.1005      | 492.10059       | γA* |
| 588.1581      | 588.15811       | γΨ* |

*See Results and Discussion. The experimental mass of the neutral species is expressed in mass units (u). Monoisotopic mass was calculated from the respective elemental composition.

†Full names available at http://mods.rna.albany.edu/.

‡Assignments corroborated by tandem mass spectrometry (i.e., MS² and CRM determinations).

*Modifications previously unreported in S. cerevisiae.
The first-generation fragment obtained in positive ion mode, activation of the protonated species detected at m/z 378.0840 led to complementary products corresponding to the purine system (i.e., 1-methyl-GMP [m1G], N2-methyl-GMP [m2G], or 7-methyl-GMP [m7G]). In positive ion mode, activation of the protonated species detected at m/z 378.0840 led to complementary products corresponding to the purine system. The observed fragments were consistent with N1-, N2-, and N7-methylation, whereas additional fragmentation of the latter identified the position of methylation sites on the purine system. A close examination of these data revealed that characteristic fragments, such as the methylated/unmethylated purine moieties, could be detected with distinctive intensity patterns in different sections of the time domain (e.g., Supplemental Fig. S1, cf. a and b), thus suggesting possible overlap between isomers with very similar mobility properties. The possible ambiguity was resolved by extracting the mobility profiles with much greater abundance) was consistent with the presence of alternative isomers with the methyl group on the ribose (i.e., 2′-O-methyl-GMP [Gm] only possible match) or the purine system (i.e., 1-methyl-GMP [m1G], N2-methyl-GMP [m2G], or 7-methyl-GMP [m7G]). In positive ion mode, activation of the protonated species detected at m/z 378.0840 led to complementary products corresponding to the purine system. The observed fragments were consistent with N1-, N2-, and N7-methylation, thus supporting the simultaneous presence of all these isomers in the S. cerevisiae sample.

The information obtained from these types of determinations corroborated the vast majority of the hits returned by database searching (marked with § in Table 1). The exception consisted of a few species for which limited signal intensity hindered the direct application of multiple activation steps. In this case, an alternative approach was implemented, which involved the application of consecutive reaction monitoring (CRM) (Tomer et al. 1988) to detect diagnostic precursor–product relationships characteristic of target nucleotides (Quinn et al. 2013). This technique affords excellent noise suppression and high duty cycle, which enhance the ability to detect low-abundance analytes in the presence of elevated background. However, its implementation requires prior knowledge of characteristic transitions to be monitored during analysis. In this case, this approach was facilitated by the predictable nature of ribonucleotide dissociation pathways (i.e., the above described base loss followed by ribose and nucleobase fragmentation), which provided series of specific precursor → product transitions for each entry of our non-redundant registry. The actual analyses were performed in automated mode with no operator intervention (see Materials and Methods). In this way, the results of multiple CRM experiments were combined to corroborate the identity of low-abundance analytes and, in most cases, confirm the simultaneous presence of isomeric species (Table 1).

In the case of IMS-MS determinations, gas-phase dissociation was used in analogous fashion to corroborate the initial assignments obtained by searching the observed m/z values in the custom registry. As described in previous work, the assignments were confirmed by activating in parallel all species dispersed on the time dimension by the ion mobility element, before transfer to the mass analyzer (Quinn et al. 2013). Called time aligned parallel (TAP) dissociation (Damen et al. 2009; Castro-Perez et al. 2011), this technique provided extensive fragmentation data that matched those observed for the yeast extract in the LTQ-orbitrap determinations. In addition, we explored an alternative characterization strategy that mimicked more traditional tandem MS spectrometry by fragmenting only those species that were recognized as potential PTMs. As illustrated for the methyl-G species described above, this type of determination was completed by isolating the precursor ion at m/z 378 in the mass-selective quadrupole, by allowing the various isomers to disperse on the time domain in the ion mobility element, and by then activating their gas-phase dissociation before final mass analysis. In what could be defined as mass-selected time-resolved (MaSTeR) dissociation experiment, the data obtained at different intervals displayed fragmentation patterns that were characteristic of the various methyl-G isomers (Supplemental Fig. S1) and matched those observed in LTQ-orbitrap experiments (Fig. 3). As discussed above, dissociation of the N-glycosidic bond differentiated isomers with the methyl group located either on the phosphoribose or nucleobase moiety, whereas additional fragmentation of the latter identified the position of methylation sites on the purine system. A close examination of these data revealed that characteristic fragments, such as the methylated/unmethylated purine moieties, could be detected with distinctive intensity patterns in different sections of the time domain (e.g., Supplemental Fig. S1, cf. a and b), thus suggesting possible overlap between isomers with very similar mobility properties. The possible ambiguity was resolved by extracting the mobility profiles
of diagnostic fragments unique for each isomer (Fig. 4). The resulting reconstructed ion chromatograms (RICs) clearly differentiated the individual species on the time scale, which helped explain the observed spectral overlap. When the mobility profile of \( m/z 378 \) was submitted to Gaussian fitting (Fig. 4A), the various components displayed a distribution that matched that of the individual RIC traces (Fig. 4B–E), thus providing the weight of the different isomers in the sample mixture. The benefits and perils of MaSTeR dissociation will be expounded in a dedicated report.

It should be noted that none of the approaches described in this report use specific \( t_D \) values to achieve positive identification of the various species, which is instead based on corroborating fragmentation information. The time domain was used to achieve separation between isomers and to enable the observation of their specific fragmentation patterns. In our hands, \( t_D \) determinations showed an average reproducibility of \( \pm 0.006 \) msec over repeated analyses on different days, which could potentially support the utilization of \( t_D \) as a unique identifying characteristic. However, in light of the number of experimental variables that may affect such quantity, it will be important to verify the reproducibility of this type of experiment across different instruments/platforms before \( t_D \) values can be used directly for identification purposes. In the meantime, the ability to achieve positive corroboration was determined not only by the comprehensive nature of the structural information afforded by gas-phase dissociation, but also by the rather conservative principles used for initial data reduction. Filtering signals that did not possess recognizable isotopic patterns increased the efficiency of database searches and the effectiveness of subsequent analyses. While it cannot be excluded that this criterion might have caused the occasional rejection of potentially valid information, conservative data reduction minimized the incidence of false positives by removing questionable signals from the initial mass lists. In this way, subsequent analyses targeted only the species that had a legitimate probability of yielding viable fragmentation data for assignment confirmation.

**Absolute versus relative assessment of modification levels**

The implementation of these MS approaches, either individually or in concert, can provide a comprehensive inventory of all PTMs detectable in a lysate. However, the ability to merely identify their presence is not sufficient to support functional studies based on the evaluation of their expression levels as a function of experimental variables. The selected platform must be able to provide valid information on the respective abundances to appreciate possible up- or down-regulation and explore functional hypotheses. Classic quantitative approaches require the availability of target analyte in neat form to generate a calibration curve through serial dilutions, or to perform incremental additions to the original sample according to the standard additions method. Unfortunately, the broad implementation of such strategies has been severely limited by the inadequate availability of standards for the majority of known modifications. The workflow proposed here presents the opportunity to overcome this challenge by utilizing purified tRNA samples as intrinsic sources of PTMs in standard-addition determinations. According to this strategy, tRNA was added to the sample as an internal standard capable of releasing all its PTMs at once during the RNase digestion step (Scheme 1). Samples containing incremental amounts of tRNA were then used to generate the signal/concentration curves necessary to determine unknown concentrations.
We first established the basis for this strategy by exploring the utilization of *S. cerevisiae* tRNA<sup>Phc</sup> (commercially available in isolated form) as a controlled source of selected PTMs. In preliminary experiments, predetermined amounts were submitted to the entire workflow to replicate actual application conditions (see Materials and Methods). The concentration of intact tRNA<sup>Phc</sup> was monitored by UV absorption determinations through the steps preceding RNase digestion (Scheme 1), which revealed an average ~25% sample recovery. Upon hydrolysis, ESI-MS analysis on the LTQ-orbitrap displayed signals for the entire complement of PTMs represented in this type of tRNA, with no trace of undigested substrate (Supplemental Fig. S2), thus indicating that the concomitant modifications did not hamper nuclease activity. The analysis was repeated on samples that contained decreasing concentrations of tRNA<sup>Phc</sup> to evaluate the limit of detection of the proposed workflow. The results revealed that, on average, the various PTMs required a sample consumption in the amol range to produce a 3:1 or better signal-to-noise ratio (Supplemental Table S1). In addition to putative detection limits, this exercise enabled us to obtain individual signal/concentration curves that provided the signal response for each ribonucleotide in tRNA<sup>Phc</sup> (Supplemental Table S1).

Based on these results, we tested the utilization of tRNA<sup>Phc</sup> in a modified standard-addition procedure that involved mixing weighed amounts of *S. cerevisiae* pellet with accurately known increments of standard. This *modus operandi* ensured that standard and endogenous RNA underwent together the entire workflow. Also in this case, UV determination was used to evaluate RNA recovery, which matched the ~25% average observed earlier for isolated standard. The excellent match between recoveries observed in the absence/presence of cell material indicated that lysis debris did not significantly interfere with phenol–chloroform extraction and subsequent workflow operations. The data obtained from the standard-addition series were used to generate the curves necessary to complete the quantitative determination of the PTMs in the extract (Table 2). The absence of accurate estimates of cellular volumes precluded a correct translation of extract concentrations into actual cellular concentrations. For this reason, the total amounts of PTMs in the sample were more conveniently expressed in terms of mol per gram of wet pellet (mol/g, Table 2), which were based on the weight of initial cell material used in the determination. For conversion purposes, we estimated that 1 g of wet pellet corresponded to ~86 mL of a culture suspension with 0.3 OD<sub>600</sub> (see Materials and Methods). The results clearly displayed the typical gulf between abundant canonic ribonucleotides and low-abundance modifications representing the bulk of the observed analytes, which showcased the excellent dynamic range afforded by this approach. In the context of the detection limits obtained from isolated tRNA<sup>Phc</sup> (Supplemental Table S1), the observed values indicated that valid determinations could be comfortably accomplished for even the least abundant modifications (i.e., ac<sup>Cm</sup> and cmnm<sup>5</sup>s<sup>2</sup>U) with as little as 800 μg of wet pellet (~69 μL of a culture suspension with 0.3 OD<sub>600</sub>). It should be noted that, although this determination covered only a subset of the entire complement of cellular PTMs—those represented in tRNA<sup>Phc</sup>—the utilization of different tRNAs or other controlled sources of natural PTMs could extend the coverage to virtually any type of modified ribonucleotides, thus making this strategy viable for a wide range of possible applications.

### Table 2. Quantitative determination of ribonucleotides present in total RNA extract of *S. cerevisiae*

| Hit<sup>a</sup> | Exp. mass (u) | Conc. (M)<sup>b</sup> | Amount (mol/g)<sup>c</sup> | Exp. AvP<sup>d</sup> |
|----------------|---------------|------------------------|---------------------------|---------------------|
| C              | 323.0519      | 2.60 × 10<sup>−5</sup> | 1.25 × 10<sup>−7</sup>   | 28.7               |
| UΨ             | 324.0359      | 2.26 × 10<sup>−5</sup> | 1.09 × 10<sup>−7</sup>   | 22.0               |
| D              | 326.0515      | 2.40 × 10<sup>−6</sup> | 1.16 × 10<sup>−9</sup>   | 2.95               |
| m<sup>1</sup>C, m<sup>3</sup>C, Cm, m<sup>6</sup>C | 337.0675      | 2.48 × 10<sup>−7</sup> | 1.19 × 10<sup>−9</sup>   | 2.45 × 10<sup>−3</sup> |
| m<sup>1</sup>Ψ, m<sup>6</sup>U, m<sup>8</sup>U, m<sup>3</sup>Ψ, Ψn, m<sup>3</sup>U | 338.0151      | NA                      | NA               | 0.30               |
| A              | 347.0631      | 2.41 × 10<sup>−5</sup> | 1.16 × 10<sup>−7</sup>   | 21.6               |
| I              | 348.0471      | NA                      | NA               | 1.44 × 10<sup>−4</sup> |
| m<sup>1</sup>A, m<sup>8</sup>A, m<sup>6</sup>A, Am, m<sup>8</sup>A | 361.0787      | 4.40 × 10<sup>−7</sup> | 4.40 × 10<sup>−9</sup>   | 1.02 × 10<sup>−4</sup> |
| G              | 363.0380      | 2.42 × 10<sup>−6</sup> | 1.17 × 10<sup>−7</sup>   | 27.7               |
| ac<sup>C</sup>, f<sup>Cm</sup> | 365.0623      | NA                      | NA               | 4.75 × 10<sup>−3</sup> |
| m<sup>2</sup>Am, m<sup>2</sup>Am, m<sup>6</sup>A | 375.0942      | NA                      | NA               | 4.75 × 10<sup>−3</sup> |
| m<sup>1</sup>G, m<sup>6</sup>G, m<sup>6</sup>G, Gm | 377.0762      | 1.32 × 10<sup>−7</sup> | 6.34 × 10<sup>−10</sup> | 1.02               |
| ac<sup>Cm</sup> | 379.0762      | NA                      | NA               | 6.85 × 10<sup>−3</sup> |
| ncm<sup>U</sup> | 381.0572      | NA                      | NA               | 3.37 × 10<sup>−2</sup> |
| m<sup>2</sup>Gm, m<sup>2</sup>G, preQ<sup>1</sup>, m<sup>2</sup>G, m<sup>2</sup>Gm | 391.0893      | 5.64 × 10<sup>−4</sup> | 2.72 × 10<sup>−9</sup>   | 6.26 × 10<sup>−3</sup> |
| cmnm<sup>5</sup>s<sup>2</sup>U | 427.0429      | NA                      | NA               | 1.02 × 10<sup>−2</sup> |
| f<sup>A</sup> | 492.1005      | NA                      | NA               | 1.34 × 10<sup>−4</sup> |
| yW | 588.1580 | 2.50 × 10<sup>−7</sup> | 1.20 × 10<sup>−9</sup> | 1.84 × 10<sup>−3</sup> |

This standard-additions determination used tRNA<sup>Phc</sup> purified from *S. cerevisiae* to achieve in situ release of PTM standards (see Materials and Methods). Name abbreviations and neutral experimental mass in units (u) are provided for each ribonucleotide.

<sup>a</sup>Full names available at [http://mods.rnajournal.org/](http://mods.rnajournal.org/).

<sup>b</sup>The concentration of each PTM in the extract was calculated from the respective curve afforded by the standard-additions determination. This figure accounts also for the ~25% recovery estimated from the standard tRNAPhe added to each sample. NA indicates PTMs that could not be determined due to their absence in the standard tRNA<sup>Phc</sup>.

<sup>c</sup>The amount of each PTM per gram of wet pellet was calculated from the respective curve afforded by the standard-additions determination. This figure accounts also for the ~25% recovery estimated from the standard tRNAPhe added to each sample. NA indicates PTMs that could not be determined due to their absence in the standard tRNA<sup>Phc</sup>.

<sup>d</sup>For each PTM, the value of abundance versus proxy (AvP) was calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonic ribonucleotides (see Materials and Methods). Each value was the average of five repeat analyses. This figure represents a relative measure of the abundance of each PTM in the sample, which can be always calculated across the board in the absence of PTM standards (see Results and Discussion).
The proposed strategy can circumvent but not eliminate the hurdles associated with the dearth of suitable standards for rigorous quantitative determinations. In many cases, however, obtaining the absolute amount of a given PTM is not as crucial as monitoring its relative abundance versus others in the sample. In proteomics studies, for example, the ability to appreciate mutual variations of post-translational modifications—reliable indicators of up- and down-regulation—is at least as valuable as the ability to determine their absolute levels. For this reason, we explored the possibility of utilizing the four canonical ribonucleotides, whose overall amounts and distributions are dependent on the cell’s genetic makeup, as a convenient internal reference to observe relative variations within a given organism. More specifically, we combined their signal intensities to establish a multicomponent reference, which could fit the MS definition of a proxy base-peak, for quantifying the various PTMs in terms of abundance versus proxy (i.e., AvP, see Materials and Methods). As shown in Table 2, this figure was readily attainable for all PTMs in the extract, regardless of their representation in a putative standard, thus providing a self-consistent and comprehensive measure of their relative abundances in the sample. We evaluated the effectiveness of this approach by comparing experimental AvP values obtained from isolated tRNA\textsuperscript{Phe} with putative figures calculated for a fixed concentration by using the respective signal/concentration curves (Supplemental Table S1). The excellent match between corresponding values provided the justification for a broader application of this treatment to monitor the relative variations of PTM levels in actual cell material.

Reproducibility of PTM analysis

Separate aliquots of the same S. cerevisiae pellet were submitted in parallel to the entire workflow to assess the technical reproducibility (precision) of the approach. The analysis of five individual samples consistently produced 40 hits. Their relative abundances were expressed in AvP units to enable direct comparisons of their distributions (Table 3). Overall, the values displayed an average of ±4.4% relative standard deviations (RSD\%) for all the PTMs, which offered a measure of the reproducibility of these determinations. Not surprisingly, the species at the higher end of the AvP scale displayed better reproducibility (i.e., smaller RSD\% values) than those at the lower end, owing to the greater susceptibility of the latter to possible fluctuations of experimental conditions throughout the workflow. For comparison purposes, the reproducibility of the ESI-MS analysis itself was evaluated separately by repeating the determination of the same digestion mixture for a total of five times. The results provided an average RSD\% of ±1.6% calculated from all the PTMs in the sample, thus suggesting that workup operations, such as extraction/lysis, digestion, etc., contributed the lion’s share of the overall ±4.4% uncertainty intrinsic in these determinations. It should be noted that in general the observed reproducibility benefited significantly from the utilization of relative rather than absolute notations. Indeed, any experimental inconsistency affecting detection is typically expected to influence analyte and reference in the same direction, as they both undergo simultaneously the same procedure. When abundances are expressed in relation to the reference, these effects tend to cancel out, thus minimizing the impact of analytical fluctuations. This explains the observation that RSD\% obtained directly from ion counts (absolute notation) were distinctively larger than those calculated from the corresponding AvPs (relative notation, Table 3).

| Hit\(^a\) | Exp. mass (u) | Technical reproducibility\(^b\) | Biological reproducibility\(^c\) |
|-----------|---------------|-------------------------------|-------------------------------|
|           | Ave. AvP | AvP RSD\% | Ave. AvP | AvP RSD\% |
| C         | 323.0519  | 28.7 ±5.0 | 29.1 ±4.9 |
| Ψ, U      | 324.0359  | 22.0 ±4.1 | 22.1 ±9.1 |
| D         | 326.0515  | 2.95 ±3.7 | 1.18 ±14 |
| m\(^6\)C, m\(^7\)C, C, m\(^6\)G | 337.0675  | 2.45 × 10\(^{-1}\) | 7.04 × 10\(^{-1}\) | 7.5 |
| m\(^8\)Ψ, U, m\(^6\)Ψ, Ψm, m\(^6\)U | 338.0515  | 1.03 ±3.6 | 7.28 × 10\(^{-1}\) | 5.5 |
| A         | 347.0631  | 21.6 ±4.7 | 21.3 ±6.6 |
| m\(^6\)A, m\(^7\)A, m\(^6\)A, m\(^3\)A, Am | 348.0471  | 1.44 × 10\(^{-1}\) | 7.3 ±10\(^{-2}\) | 6.1 |
| G         | 361.0787  | 1.02 ±6.9 | 4.53 ±10\(^{-1}\) | 11 |
| m\(^5\)G, m\(^7\)G, Cm | 363.0580  | 27.7 ±2.6 | 27 ±4.2 |
| ac\(^6\)C, f\(^5\)Cm | 365.0623  | 4.75 × 10\(^{-1}\) | 4.87 ±10\(^{-1}\) | 2.9 |
| m\(^8\)Am, m\(^7\)Am, m\(^6\)A | 375.0942  | 5.01 × 10\(^{-1}\) | 5.15 ±10\(^{-2}\) | 11 |
| m\(^6\)G, m\(^7\)G, Cm | 377.0762  | 1.02 ±2.5 | 7.52 ±10\(^{-1}\) | 4.8 |
| ac\(^6\)Cm | 379.0762  | 6.85 × 10\(^{-3}\) | 7.34 ±10\(^{-1}\) | 5.4 |
| mcm\(^3\)U | 381.0572  | 3.37 × 10\(^{-2}\) | 2.76 ±10\(^{-2}\) | 13 |
| m\(^7\)Gm, m\(^5\)Gm, preQ1, m\(^5\)Gm, m\(^7\)Gm | 391.0893  | 6.26 × 10\(^{-1}\) | 2.25 ±10\(^{-1}\) | 9.9 |
| cmm3\(^3\)sU | 427.0429  | 1.02 × 10\(^{-1}\) | 4.62 ±10\(^{-1}\) | 6.5 |
| r\(^5\)A | 492.1005  | 1.34 × 10\(^{-1}\) | 5.89 ±10\(^{-2}\) | 8.9 |
| yW | 588.1580  | 1.84 × 10\(^{-1}\) | 1.86 ±10\(^{-2}\) | 9.7 |

\(^a\)Full names available at http://mods.rna.albany.edu/.
\(^b\)Assessed by applying the proposed workflow to five separate aliquots of the same S. cerevisiae pellet. For each PTM, abundance versus proxy (AvP) was calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonical ribonucleotides (see Materials and Methods). Average and relative standard deviation (RSD \%) are reported.
\(^c\)Assessed from five different samples of S. cerevisiae grown under identical conditions in separateYPD cultures.
To weigh sample-to-sample variability (i.e., biological reproducibility) against the observed technical reproducibility, we performed parallel analyses of individual samples grown in separate cultures under otherwise identical conditions (see Materials and Methods). These experiments produced consistently the same hits obtained from the technical repeats, but their relative abundances displayed an average RSD% of ±7.8 (Table 3). At least at first sight, sample-to-sample fluctuations are typically ascribable to variations of total RNA in each sample. However, great effort was placed into growing the cultures in parallel under identical conditions, harvesting them at the same growth phase, and diluting the culture before aliquoting to approximate the same number of cells per sample (see Materials and Methods). In addition, a closer look at the results revealed that the PTMs manifested widely different fluctuation levels from one another (e.g., compare ±14% for D with ±2.9% for ac4C/f5Cm). Any variation of overall RNA content would be expected to affect all PTMs in the same direction, leading to comparable swings. Therefore, these considerations ruled out possible variations of total RNA as a source of uncertainty and suggested the influence of uncontrolled experimental variables that will warrant further investigation. When evaluating the uncertainty intrinsic in these determinations, it is important to recognize that the observed biological reproducibility of ±7.8% included also the ±4.4% contribution of the underlying technical reproducibility present in every determination. Taken together, these figures provided a measure of the typical range within which the incidence of PTMs may vary sample-to-sample under strictly controlled conditions in these yeast experiments. Determining such range for the specific system under investigation will be indispensable to recognize with confidence whether a certain variation is significant and may be unambiguously attributed to actual biological factors rather than mere sample variability.

Epitranscriptomic profiling

We finally evaluated the merits of heat-maps generated by IMS-MS analysis as direct representations of global PTM profiles. In this type of plot, the independent variables describing molecular mass and ion mobility behavior are dispersed onto orthogonal dimensions. Their intersection is unique for each analyte and enables their accurate differentiation. For this reason, a heat-map can provide a comprehensive view of the distribution of all species in the sample. As exemplified in Figure 2, the full complement of cellular PTMs and carryovers from the original lysate was observed in a single experiment that placed the abundances on a common scale. The visual nature of these plots lends itself to immediate comparisons between related samples. For example, inspection of the map obtained from a sample grown in synthetic complete medium (SC, Fig. 5A) revealed numerous signals in common with those observed for the YPD sample (Fig. 2), which corresponded to legitimate PTMs. These findings were readily confirmed by performing database searching of their respective m/z values, which led to the positive identification of 48 hits corroborated by gas-phase activation experiments (Table 4). Of this total, 40 hits matched PTMs observed in the YPD analysis (Table 1): Eight were unique for the SC and one for the YPD sample. The vast majority of the discrepancies between the SC and YPD samples corresponded instead to different carryover components, a direct consequence of the widely different compositions of these growth media and their putative effects on cell metabolism.

If visual comparison can provide qualitative information on the presence/absence of a certain PTM, actual data subtraction can reveal more subtle differences between related heat-maps and provide an assessment of the different levels of common PTMs. This type of analysis was accomplished by expressing the intensity scales in AvP units to enable axes alignment and point-by-point subtraction (see Materials and Methods). The resulting differential plot highlighted the subtle changes experienced by low-abundance species, such as ac4C and f5Cm (Fig. 5B). Overall, 31 of the 40 common PTMs were found to be more abundant in the SC than in the YPD sample, whereas the remaining nine were less abundant. A closer look at the relative deviations featured in the differential plot showed that several of them exceeded the average RSD% of ±7.8% that expresses the biological reproducibility for these types of yeast samples (Table 4). A more accurate assessment of the individual variations was obtained by comparing such deviations with the corresponding individual uncertainties provided in Table 3. This analysis indicated that the differences between SC and YPD profiles were confidently ascribable to the effects of the distinct growth media on S. cerevisiae metabolism, which represented the controlled variable between these data sets.

Environmental conditions and metabolic states can influence the expression of PTMs through the different metabolic
pathways responsible for their biogenesis. At the same time, the enzymatic infrastructure that constitutes such pathways is coded by the genome of the organism under consideration. Therefore, global PTM profiles reflect the intersection of the very specific genetic and metabolic makeup of the respective cells. We explored the ability of the proposed approach to tackle this source of diversity by analyzing different microorganisms and comparing their PTM profiles. To this end, *Escherichia coli* cultures were grown in the same SC medium utilized for *S. cerevisiae*, in such a way as to eliminate the type of available nutrients as an environmental variable (see Materials and Methods). As expected, the recorded heat-maps (Fig. 6A) differed significantly from those afforded by the corresponding *S. cerevisiae* sample (Fig. 5A). The plot obtained by subtracting the former from the latter served to accurately assess such differences and to guide subsequent analysis (Fig. 6B). The enlargement helps illustrate the type of variations afforded by low-abundance modifications. Overall, the *E. coli* sample provided a total of 30 hits, of which only 23 were in common with *S. cerevisiae* (Supplemental Table S2). The common hits displayed relative deviations ranging from $1.68 \times 10^{-2}$% to 173%, many of which exceeded the RSD% values obtained from biological repeats of either organism. This observation provided excellent indications that these deviations were statistically significant, consistent with the considerable evolutionary distance between *E. coli* (a prokaryote) and *S. cerevisiae* (a eukaryote) in the phylogenetic tree. Therefore, the outcome of these preliminary experiments bodes well for the utilization of transcriptome-wide PTM profiles as effective tools for taxonomic determinations.

### CONCLUSIONS

With the rare exceptions of 2′-O-methylation, adenosine-N6-methylation, and pseudouridylation, the high-throughput sequencing approaches (e.g., RNA-seq and similar next-generation techniques) that are the pillars of genomics research are incapable of detecting PTMs, owing to the fact that analysis takes place on DNA copies, rather than genuine RNA samples bearing the PTMs. The lack of sufficient data on PTM expression and distribution has significantly hampered the elucidation of their biological functions. MS-based approaches have the potential to fill this gap by enabling PTM recognition and quantification on the basis of their unique mass and fragmentation signatures. These types of approaches have traditionally relied on liquid chromatography and capillary electrophoresis to reduce chemical background and provide separation before analysis. This report provides proof of principle for their possible implementation without any high-resolution separation, which will greatly simplify their incorporation in large scale, high-throughput applications. The results reported here demonstrated that this strategy is capable of providing comprehensive surveys of ribonucleotide modifications at the full-transcriptome level.

**FIGURE 6.** (A) Anionic IMS-MS heat map obtained from *E. coli* grown in SC medium. (B) Difference plot obtained by subtracting the plot in A from the one provided by *S. cerevisiae* grown in the same medium (Fig. 4A).
Direct infusion analysis with either high-resolution MS or IMS-MS detection enabled the positive identification of PTMs in complex cellular extracts based on their individual molecular masses, unique fragmentation patterns, and characteristic conformational features. Eliminating typical front-end chromatographic steps streamlined the operations without affecting detection sensitivity and characterization capabilities. Combining lysozyme and nucleic acid extraction in a single step led to minimal carryover of cellular components, which did not have any appreciable consequence on the ability to detect modified ribonucleotides. The proposed workflow provided comprehensive PTM information by using as little as ~800 μg of wet cell pellet or ~69 μL of culture at 0.3 OD600 (corresponding to ~1.2 × 10^7 cells).

The proposed approaches rely on database searching and gas-phase activation techniques to positively identify the observed PTMs. For this reason, it is not necessarily geared toward the identification of new PTMs that are absent from the available databases. Indeed, it is likely that many of the observed signals that did not return hits in our experiments may correspond to yet undiscovered PTMs. Their presence in total cell extracts should not be considered a surprise in light of the almost exclusive emphasis placed by earlier studies on tRNA/rRNA analysis (McCloskey 1979, 1985; Crain 1990a; Chan et al. 2010; Su et al. 2014). This study has clearly demonstrated that the LTQ-orbitrap and IMS-MS platforms, in concert or individually, are capable of providing the information necessary to support full-fledged structural characterization. The identification of undiscovered PTMs is a possibility that will be pursued in future work.

The utilization of isolated tRNA standard provided an excellent avenue for accomplishing quantification in the absence of pure stocks of ribonucleotide variants. Interpreting a classic standard-additions strategy, purified tRNA from commercial sources was added to total RNA extracts immediately before ribonuclease digestion, which enabled the release in situ of accurately known amounts of specific PTMs. In this way, proper signal-concentration curves were obtained in parallel for all the PTMs in the standard, thus enabling their multiplexed determination in the total ribonucleotide mixture. Furthermore, we evaluated also the possibility of utilizing the endogenous canonical ribonucleotides as a proxy internal reference. This approach allowed us to determine the relative abundance of all PTMs with no addition of individual standards. The fact that the results matched the quantitative data from standard-additions determinations provided validation and enabled us to use AvPs to accurately monitor changes of expression levels across multiple samples. The results demonstrated that the typical technical reproducibility (i.e., sample to sample of the same culture) was significantly better than the biological one (i.e., culture to culture), thus substantiating the robustness of the proposed workflow. The observed reproducibility levels (expressed as average RSD% for all detected PTMs) were obtained without the utilization of stable-isotope standards, which is particularly challenging when multiple PTMs are targeted at the same time. As a natural development of any strategy based on MS platforms, future work will explore the possibility of incorporating different isotope labeling techniques in our approach. It is expected that their implementation will further improve the technical reproducibility, but it is not clear whether they will have any beneficial effect on the biological one. In the meantime, the reproducibility observed for our label-free approach allowed us to define the boundaries for deciding whether any fluctuation observed in yeast samples might be simply ascribable to experimental inconsistencies, or assumed legitimate biological significance.

The heat-maps afforded by IMS-MS analysis clearly substantiated the possibility of visualizing in a very direct and compact format the full complement of PTMs produced by a cell, which will be expected to promote large-scale comparative studies of complete epitranscriptomes. The unique features identified by dispersing the signals on the tID and m/z dimensions can lead to an immediate appreciation of qualitative variations between the types of PTMs in different samples. The ability to complete direct data subtraction offers the opportunity to detect and quantify more subtle variations of expression levels manifested by common PTMs. The possibility to observe concomitant variations of all modifications in comprehensive and self-consistent fashion will enable the investigation of their functional relationships at the system biology level. In particular, it will be possible to investigate the complex network of metabolic interactions that may lead to up- or down-regulation of specific PTMs as a function of environmental conditions, cellular states, or cell types. Taking advantage of these capabilities, future work will focus on testing whether RNA modification profiles may faithfully reflect the difference between healthy and diseased cells, or between distinctive epigenetic states. The results will provide new insights into the biological significance of ribonucleotide modifications and their roles in the cell lifecycle.

**MATERIALS AND METHODS**

**Preparation of cellular extracts**

*Saccharomyces cerevisiae* strain BY4741 was grown in yeast extract, peptone, dextrose (YPD), and synthetic complete (SC) media. Cell suspensions were streaked onto YPD agar plates and incubated at 30°C overnight. Five individual colonies were selected from each plate and placed into individual tubes containing 6 mL of either YPD or SC medium. All growth tubes were incubated at 30°C with 200 rpm gyration. Optical density at 600 nm (OD600) was monitored on a ThermoFisher Scientific Nanodrop 2000c spectrophotometer until a value slightly >0.3 units was achieved. Each liquid culture was diluted to a final 0.3 OD600 to ensure that all tubes contained comparable “cell concentrations” (i.e., number of cell per volume unit). A 3 mL aliquot of each culture was centrifuged at 6000g for 5 min to obtain pellets that contained approximately the same number of cells. In this respect, we determined that each
3 mL culture with 0.3 OD_{600} provided a wet pellet weighing on average 34.9 × 10^{-3} g. Escherichia coli K-12 strain MG1655 was grown in synthetic complete (SC) medium according to established procedures (Fritsch and Maniatis 1989; Miller 1972). Harvesting was carried out in analogous way.

Each pellet was disrupted by using Denaturation Solution (Life Technologies) in the presence of 0.5 mm diameter glass beads (BioSpec Products). When required by the standard-additions protocol, accurately known aliquots of S. creusisai trRNA^{305} (Sigma-Aldrich) were introduced at this point into the lysate to serve as internal standard. Total RNA was extracted by using the ToTALLY RNA Extraction Kit (Life Technologies), which is based on a typical phenol/chloroform procedure. The RNA was precipitated by using cold isopropanol and then treated with DNase 1 (New England Biolabs) in 1× DNase buffer to remove any remaining DNA. The recovered RNA was subsequently desalted by ethanol precipitation overnight and reconstituted in 50 μL of RNase-free water (Sigma-Aldrich). The concentration of intact total RNA from each sample was measured by UV absorbance at 260 nm. Nuclease P1 and phosphodiesterase 1 from snake venom (Sigma-Aldrich) were used to complete the digestion of RNA into individual mononucleotides, as previously described (Crain 1990b). Immediately before analysis, samples were diluted 1:10 in 150 mM ammonium acetate and 10% isopropanol.

Mass spectrometry

Samples were analyzed by direct infusion electrospray ionization (ESI) on either a ThermoFisher Scientific LTQ-orbitrap Velos mass spectrometer or a Waters Synapt G2 HDMS IMS MS spectrometer. All analyses were performed in nanoflow ESI mode by using quartz emitters produced in house by a Sutter Instruments Co. P2000 laser pipette puller. Up to 5 μL samples were typically loaded into each emitter by using a gel-loader pipette tip. A stainless steel wire was inserted in the back-end of the emitter to supply an ionizing voltage that ranged between 0.9 and 1.2 kV. Source temperature and desolvation conditions were adjusted by closely monitoring the incidence of ammonium adducts and water clusters (Fabris et al. 2010). For high-resolution determinations, the LTQ-orbitrap instrument was calibrated by using an anion mixture that contained sodium dodecyl-sulfate, sodium taurocholate, and Ultramark. These standards enabled calibration over a range of 150–2000 m/z with up to 100 ppb mass accuracy. Tandem mass spectrometry (MS/MS) was accomplished by isolating the precursor ions of interest in the linear trap quadrupole (LTQ) element, which were then collided with N\textsubscript{2} to activate fragmentation. Multistep activation experiments (MS\textsuperscript{δ}) (Solouki et al. 1996; Collings et al. 2001) were completed by properly isolating first-generation and subsequent fragments prior to activation. The fragmentation of mass-selected ions was activated by using a typical 25 V collision voltage. Ensuring products were mass analyzed either in the LTQ or the Orbitrap region of the instrument. Consecutive reaction monitoring (CRM) (Tomer et al. 1988) was performed by dialing the selected precursor \rightarrow fragment transitions in the instrument data system. Series of diagnostic CRM experiments were performed in systematic fashion by inputting lists of precursor \rightarrow fragment transitions specific for the different modifications, which were completed by the instrument with no further user intervention.

In IMS-MS experiments, apparent drift time (τ\textsubscript{d}) was determined by allowing ions to move through the traveling wave (Tri-WAVE) element of the instrument (Giles et al. 2004), before transferring them for mass analysis into the time-of-flight (TOF) stage operated in single reflectron mode. The instrument was calibrated by using a 2 mg/mL solution of cesium iodide in 50:50 water/methanol, which afforded up to 10 ppm mass accuracy. For comprehensive mixture analysis, the Tri-WAVE region was held at a pressure of \sim 4.40 mbar (uncalibrated gauge reading) by a 90 mL/min flow of N\textsubscript{2} and 180 mL/min of He. It was operated with \sim 650 m/sec IMS wave velocity, a 40 V wave height, a 109 m/sec transfer wave velocity, and a 2.0 V transfer wave height. Time aligned parallel (TAP) dissociation and MaSiTe fragmentation of isolobes were performed by raising the transfer voltage to 17 V and the cell pressure to \sim 4.60 mbar (uncalibrated gauge reading) with a flow of 140 mL/min N\textsubscript{2} and 180 mL/min He. At the same time, IMS wave velocity was raised to \sim 700 m/sec, transfer wave velocity to \sim 600 m/sec, and transfer wave height to 4.0 V.

Data analysis

High-resolution and fragmentation data obtained on the LTQ-orbitrap instrument were processed by Xcalibur 2.1 software (ThermoFisher Scientific). Mass calculations and predictions of elemental composition were performed by using the Molecular Weight Calculator software made available by the Pacific Northwest National Laboratory (v. 6.49, http://ncrr.pnl.gov/software/). A data reduction step was implemented prior to database searching to simplify the operations and minimize the incidence of false positives. Instead of relying on a predefined intensity threshold to discriminate signal from noise, the experimental masses to be used in the searches were selected according to a deconvolution algorithm included in the Xcalibur 2.1 software (Bushberg et al. 2012). This algorithm requires the detection of full-fledged 12C and 13C signals to correctly assign the charge state of observed species. If the 13C peak of a low-abundance component was not recognized from the background (and a plausible charge was not assigned), then the mass of the corresponding 12C was filtered out regardless of whether its intensity afforded an acceptable signal-to-noise ratio. The resulting mass list was then searched against the METLIN database (http://metlin.scripps.edu/index.php) and a non-redundant registry obtained by combining the entries present in the RNA Modifications (http://mods.rna.albany.edu/) and MODOMICS databases (http://modomics.genesilico.pl/). Matching between experimental data and database information was carried out by using software developed in house.

IMS-MS data were displayed in the form of heat-map plots with arrival time (τ\textsubscript{d}) and mass to charge ratio (m/z) placed on the x- and y-axis, respectively, by using OriginPro 9.1 (Origin Lab). A color gradient provided in each plot was used to communicate the signal intensity expressed in arbitrary ion counts. For data subtraction analysis, appropriate scaling factors were utilized to align the intensity scales of the selected plots. Such factors were calculated to match the combined intensities of the four canonical ribonucleotides (i.e., \sum_{i=1}^{4} c_i \text{ with } c_i \text{ corresponding to the respective absolute intensity in arbitrary counts}) observed in each plot. Taking advantage of this proxy, de facto internal reference, it was possible also to express the abundance of each species in relation to that of the canonical ribonucleotides according to

\[ \text{Predicted psi} = \text{baseline} + \sum_{i=1}^{6} P_i \times (\text{psi}_i - \text{baseline}), \]
in which $\text{AvP}$ is the abundance versus proxy of a certain species obtained from its absolute intensity ($a_i$) normalized to the combination of the abundances of the canonical ribonucleotides ($\sum_i^L a_i$). Home-built software was used to process the experimental data, calculate AvP values and, when necessary, apply appropriate scaling factors for data alignment. The same application was used also to perform point-by-point subtraction of aligned data. The software application developed in house is the object of a manuscript in preparation. The results were visualized in heat-map format by utilizing OriginPro 9.1 (Origin Lab).

**SUPPLEMENTAL MATERIAL**

Supplemental material is available for this article.

**ACKNOWLEDGMENTS**

This work was supported by the University at Albany and by the National Institutes of Health (GM064328-14).

Received December 22, 2014; accepted March 25, 2015.

**REFERENCES**

Ajay SS, Parker SCJ, Abaan HO, Fuentes Fajardo KV, Margulies EH. 2011. Accurate and comprehensive sequencing of personal genomes. Genome Res 21: 1498–1505.

Apzuwese WA, Vouros P. 1998. Analysis of DNA adducts by capillary methods coupled to mass spectrometry: a perspective. J Chromatogr A 794: 97–108.

Banks JP, Shen S, Whitehouse CM, Fenn JB. 1994. Ultraconsistently assisted electrospray ionization for LC/MS determination of nucleosides from a transfer RNA digest. Anal Chem 66: 406–414.

Banoub JH, Limbach PA. 2010. Mass spectrometry of nucleosides and nucleic acids. CRC Press Inc, Boca Raton, FL.

Biemann K, McCloskey JA. 1962. Application of mass spectrometry to biopolymers and why. J Am Soc Mass Spectrom 794: 162:

Bushberg JT, Seibert JA, Leidholdt EM Jr, Boone JM. 2012. The essential physics of medical imaging. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA.

Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF. 2011. The RNA modification database, RNAMDB: 2011 update. Nucleic Acids Res 39: D195–D201.

Collings BA, Campbell JM, Mao D, Douglas DJ. 2001. A combined linear ion trap time-of-flight system with improved performance and MSn capabilities. Rapid Commun Mass Spectrom 15: 1777–1795.

Crain PF. 1990a. Mass spectrometric techniques in nucleic acid research. Mass Spectrom Rev 9: 505–554.

Crain PF. 1990b. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. Methods Enzymol 193: 782–790.

Crem BN, Chen W, Chakraborty AB, van Oosterhout M, Mazzeo JR, Gebler JC, Schellens JHM, Rosing H, Beijnen JH. 2009. Electrospray ionization quadrupole ion-mobility time-of-flight mass spectrometry as a tool to distinguish the lot-to-lot heterogeneity in N-glycosylation profile of the therapeutic monoclonal antibody trastuzumab. J Am Soc Mass Spectrom 20: 2021–2033.

Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. 2006. MODOMICS: a database of RNA modification pathways. Nucleic Acids Res 34: D145–D149.

Dwivedi P, Wu C, Matz ML, Clowers BH, Siems WF, Hill HH Jr. 2006. Gas-phase chiral separations by ion mobility spectrometry. Anal Chem 78: 8200–8206.

Esman E, Broes D, Hoes I, Lemière F, Vanhoucke K. 1998. Liquid chromatography–mass spectrometry in nucleoside, nucleotide and modified nucleotide characterization. J Chromatogr A 794: 109–127.

Fabris D, Turner KB, Hagan NA, 2010. Electrospray ionization—mass spectrometry for the investigation of protein–nucleic acids interactions. In Mass spectrometry of nucleosides and nucleic acids (ed. Banoub J, Limbach P), pp. 303–327. CRC Press, Taylor and Francis Group, LLC, London, UK.

Ferrié-D’Amare AR. 2003. RNA-modifying enzymes. Curr Opin Struct Biol 13: 49–55.

Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Frommer M, McDonald LE, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89: 1827–1831.

Giles K, Pringle SD, Worthington KR, Little D, Wildgoose JL, Bateman RH. 2004. Applications of a travelling wave-based radiofrequency-only stacked ring ion guide. Rapid Commun. Mass Spectrom 18: 2401–2414.

Helm M. 2006. Post-transcriptional nucleotide modification and alternative folding of RNA. Nucleic Acids Res 34: 721–733.

Herman JG, Graff JR, Myohansen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA 93: 9821–9826.

Horkawicz S, Rother KM, et al. 2012. MODOMICS: a database of RNA modification pathways. Nucleic Acids Res 39: D201.

Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. 1994. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7: 536–540.

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. 2013. The next-generation sequencing revolution and its impact on genomics. Cell 155: 27–38.

Kowalak JA, Dalluge JI, McCloskey JA, Stetter KO. 1994. The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. Biochemistry 33: 7869–7876.

Lapthorn C, Pullen F, Chowdhry BZ. 2013. Ion mobility spectrometry–mass spectrometry (IMS–MS) of small molecules: separating and assigning structures to ions. Mass Spectrom Rev 32: 43–71.

Li J, Yang Z, Yu B, Liu J, Chen X. 2005. Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Curr Biol 15: 1501–1507.

Limbach PA, Crain PF, McCloskey JA. 1994. Summary: the modified nucleosides of RNA. Nucleic Acids Res 22: 2183–2196.

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowski M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, et al. 2012. MODOMICS: a database of RNA modification pathways—2013 update. Nucleic Acids Res 41: D262–D267.

McCloskey JA. 1979. Characterization of nucleosides by mass spectrometry. Nucleic Acids Symp Ser s109–s113.

www.rnajournal.org 1373
McCloskey JA. 1985. Mass spectrometry of nucleic acid constituents and related compounds. In Mass spectrometry in the health and life sciences (ed. Burlingame AL, Castagnoli N), pp. 521–546. Elsevier, Amsterdam.

Miller JH. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Nordhoff E, Kirpekar F, Roepstorff P. 1996. Mass spectrometry of nucleic acids. Mass Spectrom Rev 15: 67–138.

Ofengand J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. FEBS Lett 514: 17–25.

Quinn R, Basanta-Sanchez M, Rose RE, Fabris D. 2013. Direct infusion analysis of nucleotide mixtures of very similar or identical elemental composition. J Mass Spectrom 48: 703–712.

Singer-Sam J, Grant M, LeBon JM, Okuyama K, Chapman V, Monk M, Riggs AD. 1990. Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgl-1 CpG island of mouse embryos at the time of X-chromosome inactivation. Mol Cell Biol 10: 4987–4989.

Smith CA, O’Maille G, Want E, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. 2005. METLIN: a metabolite mass spectral database. Ther Drug Monit 27: 747–751.

Solouki T, Pasa-Tolić L, Jackson GS, Guan S, Marshall AG. 1996. High-resolution multistage MS, MS2, and MS3 matrix-assisted laser desorption/ionization FT-ICR mass spectra of peptides from a single laser shot. Anal Chem 68: 3718–3725.

Su D, Chan CTY, Gu C, Lim KS, Chionh YH, McBee ME, Russell BS, Babu IR, Begley TJ, Dedon PC. 2014. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. Nat Protoc 9: 828–841.

Tomer KB, Guenat CR, Deterding LJ. 1988. Consecutive reaction monitoring in a four-sector mass spectrometer: MS4 and one step beyond. Anal Chem 60: 2232–2236.

Verbeck GF, Ruotolo BT, Sawyer HA, Gillig KJ, Russell DH. 2002. A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. J Biomol Tech 13: 56–61.

von Helden G, Wyttenbach T, Bowers MT. 1995. Conformation of macromolecules in the gas phase: use of matrix-assisted laser desorption methods in ion chromatography. Science 267: 1483–1485.

Wu J, McLuckey SA. 2004. Gas-phase fragmentation of oligonucleotide ions. Int J Mass Spectrometry 237: 197–241.

Yamashita M, Fenn JB. 1984. Electrospray ion source. Another variation on the free-jet theme. J Phys Chem 88: 4671–4675.