LC/MS analysis of cellular RNA reveals NAD-linked RNA

Y. Grace Chen[†], Walter E. Kowtoniuk[†], Isha Agarwal, Yinghua Shen, David R. Liu*

SUPPLEMENTARY INFORMATION

Supplementary Results

A General Method for Detecting Cellular Small Molecule-RNA Conjugates

Recently we developed two methods to detect cellular small molecule-RNA conjugates that both relied on the ability of the conjugates to react with base or nucleophiles in a manner that resulted in a change in the mass of the small molecule or in the mass of the nucleotide to which the small molecule was attached[1]. The molecular complexity of samples derived from cellular RNA requires the use of comparative analysis to distinguish authentic small molecule-RNA conjugates from false positives. Both of these methods therefore provided a matched pair of samples for comparative mass spectrometry analysis in the form of chemically treated samples versus untreated samples.

As the basis for a more general method to detect biological small molecule-nucleic acid conjugates, we envisioned that the samples for comparative analysis could be derived from (*i*) RNA treated with a nuclease enzyme that cleaved the RNA to mononucleotides, and (*ii*) RNA treated with heat-inactivated nuclease that should not generate mononucleotides but that should contain all other small molecules introduced or generated during RNA isolation and sample processing. Non-canonical species more abundant in the active nuclease-treated sample compared with the inactive nuclease-treated sample could then be detected as possible novel small molecule-RNA conjugates, regardless of their chemical lability. Based on the sensitivity of our current liquid chromatography/mass spectrometry (LC/MS) methods[1], this method should be able to detect ~0.5 pmol of a given nucleotide, corresponding to detection of modifications that are present at ~30 copies per bacterial cell. This sensitivity represents a ~200-fold improvement compared with previous efforts to study modified nucleotides, which were typically focused on a specific fraction of the cellular RNA (e.g., tRNAs) and were limited to studying nucleotides at high abundance levels of ≥ ~2%; in *E. coli* this abundance level corresponds to ~4,400 copies per cell[2,3].

Compared with our previous methods, this more general approach places a greater burden on the downstream processing of the resulting comparative LC/MS data. We therefore computationally filtered out (**Supplementary Fig. 1**) canonical mono- and di-ribonucleotides, all known modified nucleotides including aminoacylated AMPs, all known base-modified tRNA and rRNA nucleotides, and the common ionization fragments and isotope peaks from all of these species. This set of known species is significantly larger than the set previously used in the chemical cleavage methods due to the more comprehensive nature of this approach.

Validation Using Aminoacylated tRNAs and Known Nucleoside Modifications

To validate the method, we processed whole *E. coli* cellular RNA and searched for species enriched in the active nuclease sample with masses and retention times consistent with amino acid-linked adenosine monophosphates and nucleobase modifications known to exist in *E. coli*. Of the 20 major 3'-aminoacyl adenosine monophosphates, 16 (80%) were detected as enriched at least 2-fold in the active nuclease-treated samples compared with the heat-inactivated nuclease-treated samples (**Supplementary Figs. 2a and 3a**). In addition, 31 species enriched at least 2-fold were observed with masses consistent with known RNA nucleobase modifications (**Supplementary Figs. 2b and 3b**). We found that an enrichment threshold of 2-fold represented a reasonable balance between detecting the maximum number of known 3'-aminoacyl adenosine monophosphates and nucleobase modifications, and minimizing the number of irreproducible false-positive signals resulting from experimental noise. We note that several of the detected nucleobase modifications such as queuosine, 5-hydroxyuridine, and 2-methylthio-N^6-isopentenyladenosine are not expected to be base- or nucleophile-labile, and were not detected using our previous methods that rely on base cleavage or nucleophile cleavage[1]. These results validate the ability of the nuclease versus heat-inactivated nuclease-based method to detect the presence of known small molecule-RNA conjugates from whole cellular RNA, including conjugates that are not chemically labile.

Detection of Unknown Small Molecule-Nucleotide Conjugates

When applied to *E. coli* RNA the method described above also detected 24 non-canonical, unknown species enriched at least 2-fold (**Supplementary Fig. 4a**). In *S. venezuelae*, this method yielded 28 unknown species that were enriched 2-fold or more (**Supplementary**

Fig. 4b). Independent replicates starting with different cell cultures generate enrichment factors with trial-to-trial correlation coefficients of 0.90 in *E. coli* (**Fig. 4c**) and 0.93 in *S. venezuelae* (**Supplementary Fig. 4d**). None of the observed unknown species were detected from total *E. coli* or *S. venezuelae* RNA if active nuclease P1 was omitted, or if active nuclease P1 treatment was replaced with incubation in formamide and/or 10 mM EDTA at 95°C, conditions expected to abrogate RNA secondary structure. These results suggest that the species in **Supplementary Fig. 5** arise from nuclease-mediated RNA cleavage, and not from the release of small molecules non-covalently associated with RNA.

NADH Oxidation During RNA Isolation

NADH, the reduced form of NAD, is not detected as a small molecule-RNA conjugate by our nuclease-based method. However, the absence of this species (expected $[M-H]^-$ m/z = 664.1175) does not rule out the presence of NADH-linked cellular RNA. Indeed, when authentic NADH was added directly to nuclease digestion reactions, very low NADH signal was observed. Instead, the addition of NADH resulted in an increased NAD signal (**Supplementary Fig. 12**), suggesting that any NADH-RNA present likely oxidized to NAD-RNA during RNA isolation and sample processing.

Characterization of the NAD-RNA Linkage

Unlike nuclease P1, RNase A mediates RNA cleavage by catalyzing the attack of a 2'-hydroxyl group to form a 2', 3'-cyclic phosphonucleotide, which then can be hydrolyzed to generate a nucleotide 3'-monophosphate. Products of RNase A digestion, with the exception of the 3'-terminal base, are therefore monophosphonucleotides[4-6]. When total RNA from *E. coli* and *S. venezuelae* was digested with RNase A, the internal rRNA nucleoside modification $N^6,N^{6'}$-dimethyladenine[7] was detected as a mixture of the cyclic and acyclic monophosphonucleotides, as expected (observed $[M-H]^-$ m/z = 356.0772 and m/z = 374.0873, expected $[M-H]^-$ m/z = 356.0760, and m/z = 374.0866). RNase A digestion also generated monophosphorylated NAD (observed $[M-H]^-$ m/z = 742.0678, expected $[M-H]^-$ m/z = 742.0682). Similar results were observed when authentic NAD-RNA, generated by T7 RNA polymerase[8] was digested with RNase A (observed $[M-H]^-$ m/z = 742.0653). These results are consistent with

the nuclease P1 digestion results and further support a model in which the NAD group is covalently linked not to the 3' terminus but rather to the 5' terminus of RNA.

We note that under base cleavage conditions (pH 8) previously used to discover CoA-linked RNA, the abundance of NAD in the nuclease-treated samples is virtually identical (116 vs. 108 ion counts) to that of the control conditions (pH 4.5) and thus would have been overlooked using our previous methods[1].

NAD radical formation may contribute to the unexpectedly high +1 Da isotope peak in the authentic and cellular NAD spectra

Based on the natural isotope abundances, the expected abundance ratio of the +1 Da isotope peak at $[M-H]^-$ *m/z* = 663.1069 to the isotope peak at $[M-H]^-$ *m/z* = 662.0986 is equal to 26.8% (**Supplementary Fig. 7a**). However, the observed ratio is higher for both authentic and cellular RNA-derived NAD samples (**Supplementary Fig. 7a**). We speculate that the increase in the $[M-H]^-$ *m/z* = 663.1069 signal may be due to the contributions of other species that either coelute with NAD or are produced during ionization. For example, an NAD radical anion (**Supplementary Fig. 7b**) would be observed primarily as a $[M-H]^{\cdot-}$ *m/z* = 663.1091 species. A mixture of NAD anion and NAD radical anion species during ionization therefore would result in a higher-than-expected abundance of the NAD +1 Da peak.

Experimental Methods

General. Unless otherwise noted, all starting materials were obtained from commercial suppliers and were used without further purification.

Bacterial Growth and Crude Nucleic Acid Isolation. *E. coli* TOP10 (Invitrogen) was cultured to OD_{600} = 0.7-0.8 at 37ºC in 2 L LB broth Miller (EMD Bioscience). *S. venezuelae* ATCC #10595 was cultured to OD_{600} = 0.7-0.8 at 30 ºC in 2 L MYME media (100 g/L sucrose, 10 g/L maltose, 5 g/L peptone, 3 g/L yeast extract, 3 g/L malt extract)[9]. All subsequent manipulations of cells were carried out on ice or at 4 °C. The bacteria were centrifuged (10 min at 6,750 x g), resuspended in lysis buffer (1% SDS, 2 mM EDTA, 32 mM NaOAc, pH 4.5), and vortexed vigorously. After incubation on ice for 15 min, the lysate was cleared by centrifugation (10 min at 10,000 x g), extracted with acid-phenol chloroform (Ambion) until the organic-aqueous

interface was clear, and the aqueous layer was washed once with chloroform. An equal volume of isopropanol was added to the resulting aqueous extract and the mixture was incubated on crushed dry ice 20-30 min prior to centrifugation (20 min, 15,000 x g). The resulting pellet was dissolved in 50 mM NH_4OAc, pH 4.5, and subjected to size-exclusion chromatography using NAP5 columns (GE Healthcare). The macromolecular fraction was treated with 0.04 U/μL TURBO DNase (Ambion) at 25 ºC for 30 min and then with 0.06 U/μL proteinase K (New England Biolabs) at 25 °C for 30 min. The resulting solution was extracted with acid-phenol chloroform (Ambion) twice, washed with chloroform, and again subjected to size-exclusion chromatography using NAP5 columns. The macromolecular fraction was divided into aliquots, lyophilized, and stored as a dry powder at -80 ºC.

Nuclease Digestion. The cleavage condition sample was prepared by incubating 500 μg of *E. coli* RNA or 500 μg of *S. venezuelae* RNA with 10 U nuclease P1 (Sigma-Aldrich) in 200 μL of 50 mM NH_4OAc, pH 4.5 at 37 ºC for 40 min. The control condition samples was prepared by incubating 500 μg aliquot of *E. coli* RNA or *S. venezuelae* RNA with 10 U heat-inactivated nuclease P1 (95 ºC for one hour) in 200 μL of 50 mM NH_4OAc, pH 4.5 at 37 ºC for 40 min. The digestion products were purified by size-exclusion chromatography (NAP5) and the small-molecule fraction was retained. The nuclease P1 digestion with $H_2{}^{18}O$ (Cambridge Isotope Laboratories) was performed as described above except in buffer with a final composition containing 86% $H_2{}^{18}O$ and 14% $H_2{}^{16}O$.

LC/MS Data Collection and Analysis. LC/MS was performed using a Waters Aquity UPLC Q-TOF Premier instrument with an Aquity UPLC BEH C18 column (1.7 μm, 2.1 mm x 100 mm, Waters). Mobile phase A was 0.1% aqueous ammonium formate, and mobile phase B was 100% methanol. The flow rate was a constant 0.300 mL/min and the mobile phase composition was as follows: 0% B for 3 min; linear increase over 17 min to 100% B; maintain at 100% B for 2 min before returning linearly to 0 % B over 1 min. Electrospray ionization (ESI) was used, with a capillary voltage of 3.5 kV, sampling cone voltage of 40.0 kV, and collision voltage of 1.0 V. The drying gas temperature was 300 ºC, the drying gas flow rate was 800 L/hour, the source temperature was 150 ºC, and the detector was operated in negative ion mode. To observe the NAD parent ion ($[M-H]^-$ m/z = 662.1018), the capillary voltage was 2.75 kV and the sampling

cone voltage was 20.0 kV. For each sample, 15 µL of the redissolved lyophilized material was injected. Ions with cleavage condition average integrated ion intensities below 50 ion counts were not considered for further analysis. The XCMS program[10] quantified the area under detected ion abundance peaks as it stepped through each ion chromatogram; the step size was set to 0.050 Da. Integrated ion abundances were averaged among replicates, and the enrichment values reported were the ratios of these average ion intensities between active nuclease conditions and heat-inactivated nuclease (control) conditions.

MS/MS Fragmentation Analysis. MS/MS experiments were performed using the same instrument, LC gradient, and MS parameters described above. The collision voltage was varied empirically from 20.0-30.0 V.

Isotope Labeling of *S. venezuelae* RNA. *S. venezuelae* ATCC #10595 was cultured to OD_{600} = 0.7-0.8 at 30 °C in minimal media (4.3 g/L NaH_2PO_4, 6.1 g/L K_2HPO_4, 2.0 g/L NaCl, 2.0 mg/L $FeSO_4 \cdot 7H_2O$, 2.0 mg/L $MnCl_2 \cdot 4H_2O$, 2.0 mg/L $ZnSO_4 \cdot 7H_2O$, 2.0 mg/L $CaCl_2$, 0.6 g/L $MgSO_4 \cdot 7H_2O$, 10.0 g/L glucose, and 2.0 g/L $(NH_4)_2SO_4$)[9]. For ^{13}C labeling, ^{13}C-glucose (99% ^{13}C, Cambridge Isotope Laboratories) was used as the sole carbon source. For ^{15}N labeling, ^{15}N-ammonium sulfate (99% ^{15}N, Cambridge Isotope Laboratories) was used as the sole nitrogen source. RNA isolation, nuclease P1 digestion, and LC/MS analysis was performed as described above.

NAD Spiking Into Cell Lysate. *E. coli* TOP10 cell lysate was prepared as described above and divided into three equal aliquots. NAD and NADH (Sigma-Aldrich) were added to the cell lysate as follows: aliquot 1: no added NAD or NADH; aliquot 2: 500 nmol of NAD; aliquot 3: 500 nmol of NADH. Based on our standard curve analysis 500 nmol of NADH or 500 nmol of NAD theoretically represents 4,000,000 ion counts, however in practice this amount of NAD or NADH saturates the detector. Cell lysates were processed and analyzed as described above.

RNase A Digestion. Total RNA prepared as described above was subjected to digestion with 1 µg RNase A (Ambion) in 200 µL of 50 mM NH_4OAc, pH 4.5, at 37 ºC for 20 min). After digestion, the sample was subjected to size-exclusion chromatography (NAP5, GE Healthcare)

and the small-molecule fraction was lyophilized. The lyophilized product was redissolved in 20 µL of 0.1 % aqueous ammonium formate and analyzed by LC/MS as described above.

Number of NAD-RNA(s) per *E. coli* cell. To establish a standard curve of MS ion counts per pmol NAD, 5, 25, 50 and 100 pmols of authentic NAD (Sigma-Aldrich) were analyzed by initial and mild ionization conditions as described above. The NAD signal from 500 ug of total *E. coli* RNA was compared to the standard curves to determine the number of pmols of cellular NAD per µg RNA. The amount molecules of NAD-RNA per cell was then calculated based on $59x10^{-15}$ g RNA per *E. coli* cell[11].

Authentic NAD and NADH Carried Through Mock RNA Isolation. To determine if cellular NADH-RNA could be detected by our nuclease based method, 50 pmol of NAD and NADH (Sigma-Aldrich) in 500 µL of 50 mM NH_4OAc, pH 4.5 were incubated on ice for 15 min, and then at 4 °C for 10 min. The samples were moved back to ice for 5 min, and then placed at 4 °C for 10 min. This procedure was repeated three times, corresponding to the number of acid-phenol chloroform and chloroform extractions during the crude nucleic acid isolation. The NAD and NADH samples were then placed on crushed dry ice for 20-30 min prior to incubation at 4 °C for 20 min, mimicking the isopropanol precipitation of cellular RNA. While the crude nucleic acids were subjected to size-exclusion chromatography, the NAD and NADH samples were exposed to air and incubated at 25 °C for 10-15 min. The samples were then incubated at 25 °C for 1 h, and lyophilized. The dry powder was redissolved in 200 µL of 50 mM NH_4OAc, pH 4.5 and incubated at 37 ºC for 40 min. The NAD and NADH samples were exposed to air and incubated at 25 ºC for 15-20 min. The samples were lyophilized, and stored as a dry powder at -20 ºC.

Preparation of Authentic 5'-NAD-RNA. 5'-NAD-RNA was prepared as previously described[8]. A DNA template containing the T7 class II promoter (Φ2.5) and encoding a 223-base RNA was prepared by PCR. Transcription was carried out in 1x New England Biolabs (NEB) RNA pol buffer supplemented with 1 mM each NTP, 1 mM NADH (Aldrich), 0.01% Triton X-100, 5 mM DTT, 0.2 U/µl RNase inhibitor (NEB), 0.2 µM DNA template, and 5 U/µL T7 RNA polymerase (NEB). The reactions were incubated at 37 °C for 2 hrs before the product

was isolated by precipitation with ethanol, dissolved in TURBO-DNase buffer (Ambion), and treated with TURBO-DNase (0.04 U/μL final concentration, 37 °C, 30 min). The DNase-treated RNA was precipitated with ethanol, redissolved in water, and purified by size-exclusion chromatography (NAP5, GE Healthcare). The precipitated NAD-RNA product following DNase treatment was dissolved in water, purified by silica column (Qiagen RNeasy), quantified by A_{260}, digested with nuclease P1, and analyzed by LC/MS as described above in order to estimate the fraction of the resulting RNA strands that contained the NAD modification. We observed that ~3% of the transcripts generated by this procedure were linked to NAD.

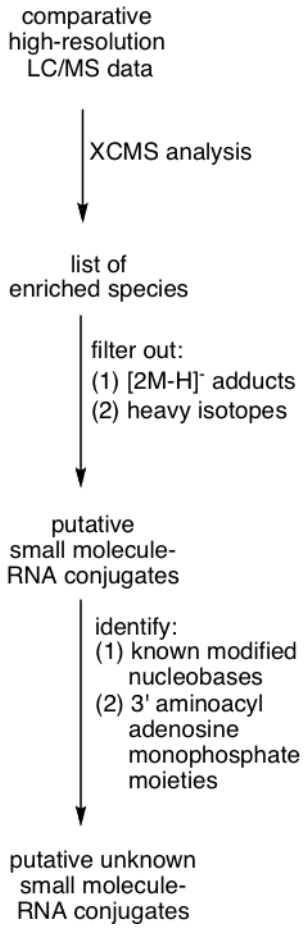
***In Vitro* Transcription.** *In vitro* transcription reactions contained *E. coli* RNA polymerase (Epicentre Biotechnologies) in 50 mM Tris-HCl, pH 7.5; 150 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100 supplemented with NTPs (0.5 mM each), DTT (10 mM), and NADH (0.5 mM or 5.0 mM). The plasmid or genomic DNA template was added to a final concentration of 0.02 μg/μL and *E. coli* RNA polymerase was added to a final concentration of 0.04 U/μL. The reactions were incubated at 37 °C for 22 hrs before the product was isolated by precipitation with ethanol, dissolved in turbo-DNase buffer (Ambion), and treated with TURBO-DNase (Ambion) at a final concentration of 0.04 U/μL for 30 minutes at 37 °C. The RNA was purified by RNeasy spin-column (Qiagen) to remove free NTPs and NADH. The resulting RNA was digested with nuclease P1 and analyzed by LC/MS as described above.

The first template was a modified pUC19 plasmid, which encoded an adenosine at the +1 position of each of its four predicted transcripts. An *in vitro* transcription reaction containing 0.5 mM of each NTP and either 0.5 mM or 5 mM of NADH yielded 222 μg or 174 μg of RNA, respectively. The second template used was *E. coli* genomic DNA. *In vitro* transcription in the presence of either 0.5 mM or 5 mM of NADH yielded 78 μg or 70 μg of RNA. Once again, this material contained no detectable NAD or NADH after nuclease P1 digestion (**Supplementary Fig. 11**).

RNA Size Fractionation by Silica Spin-Column. Total *E. coli* or *S. venezuelae* RNA was dissolved in 100 mM aqueous DTT and purified with Qiagen RNeasy kit (Qiagen) following the manufacturer's instructions. RNA from either the flow-through (< ~200 nucleotides) or the

eluted fraction (> ~200 nucleotides) was recovered by two consecutive precipitations with ethanol followed by desalting using NAP5 columns (GE Healthcare).

a



b

3'-aminoacyl adenosine monophosphate	expected $[M-H]^-$ *m/z*
alanine	417.0929
arginine	502.1569
asparagine	460.0987
aspartic acid	461.0827
cysteine	449.0650
glutamic acid	475.0984
glutamine	474.1144
glycine	403.0772
histidine	483.1147
isoleucine	459.1399
leucine	459.1399
lysine	474.1507
methionine	477.0963
phenylalanine	493.1242
proline	443.1085
serine	433.0878
threonine	447.1035
tryptophan	532.1351
tyrosine	509.1191
valine	445.1242

known nucleotide monophosphate	expected $[M-H]^-$ *m/z*	known nucleotide monophosphate	expected $[M-H]^-$ *m/z*
galQ	650.1716	$preQ_0$	386.0507
manQ		mnm^5s^2U	382.0480
o_2yW	619.1407	cm^5U	381.0341
OHyW	603.1458	ncm^5U	380.0501
yW	587.1508	ncm^5Um	
ms^2hn^6A	551.0967	ac^4Cm	378.0708
ms^2t^6A	537.0810	Gm	376.0664
OHyW*	531.1246	m^1G	
hn^6A	505.1090	m^2G	
m^6t^6A		m^7G	
oQ	504.1137	m^1Im	375.0711
Gr(p)	494.0930	m^1Am	374.0871
t^6A	491.0933	$m^6{}_2A$	
Q	488.1188	m^6Am	
Ar(p)	478.0981	nm^5s^2U	368.0323
ms^2io^6A	476.1010	mnm^5U	366.0708
tm^5s^2U	476.0204	$m^4{}_2Cm$	364.0915
ms^2i^6A	460.1061	ac^4C	364.0551
tm^5U	460.0433	f^5Cm	
k^2C	450.1395	GMP	362.0507
g^6A	447.0671	Im	361.0555
m^1acp^3Y	438.0919	m^1I	
inm^5s^2U	436.0949	Am	360.0715
inm^5Um	434.1334	m^1A	
io^6A	430.1133	m^2A	
mnm^5se^2U	429.9924	m^6A	
mimG	428.0977	mo^5U	353.0392
$cmnm^5s^2U$	426.0378	s^2Um	
acp^3U	424.0763	m^5s^2U	353.0214
$cmnm^5Um$		hm^5C	352.0551
inm^5U	420.1177	m^3Um	351.0599
i^6A	414.1184	m^5Um	
imG	414.0820	m^4Cm	350.0759
imG2		m^5Cm	
$mcmo^5U$	411.0446	f^5C	350.0395
$mcmo^5U$		I	347.0398
mcm^5s^2U	411.0269	AMP	346.0558
$cmnm^5U$	410.0606	m^5D	339.0599
mcm^5Um	409.0654	ho^5U	339.0235
ms^2m^6A	406.0592	s^2U	339.0058
$m^2{}_2Gm$	404.0977	s^4U	
$m^2{}_7Gm$		s^2C	338.0217
G^+	403.0773	m^3U	337.0442
imG-14	400.0664	m^3Y	
chm^5U	397.0290	m^5U	
cmo^5U		Um	
mcm^5U	395.0497	Cm	336.0602
m^1Gm	390.0820	m^3C	
$m^2{}_2G$		m^5C	
$m^2{}_7G$		D	325.0442
m^2Gm		UMP	323.0286
$preQ_1$		Y	
$m^6{}_2Am$	388.1028	CMP	322.0446
ac^6A	388.0664		

Supplementary Figure 1. Computational filtering to identify putative unknown small molecule-RNA conjugates. (**a**) Following XCMS analysis[10], we filtered out MS adducts and isotopic species resulting in a list of putative small molecule-RNA conjugates. $[2M-H]^-$ adducts were filtered out by removing the heavier of any two peaks separated in *m/z* value by M Da with the same retention time. Heavy isotope species were filtered out by removing any peak with a corresponding peak 1.00627 Da (the mass of a neutron) lower in *m/z* with identical retention times. From the resulting filtered list we identify known modified nucleobases, shown in (**b**). Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

a

amino acid	expected [M-H]⁻ m/z	observed [M-H]⁻ m/z	nuclease/control enrichment
alanine	417.0929	417.0934	61
asparagine	460.0987	460.0977	6.7
glutamine	474.1144	474.1135	34
glycine	403.0772	403.0769	111
histidine	483.1147	483.1142	43
isoleuicine	459.1399	459.1392	746
leucine	459.1399	459.1387	537
lysine	474.1507	474.1499	53
methionine	477.0963	477.0959	123
phenylalanine	493.1242	493.1243	244
proline	443.1085	443.1086	17
serine	433.0878	433.0866	20
threonine	447.1035	447.1027	107
tryptophan	532.1351	532.1346	45
tyrosine	509.1191	509.1190	36
valine	445.1242	445.1229	728

b

RNA modification	expected [M-H]⁻ *m/z*	observed [M-H]⁻ *m/z*	retention time (min)	nuclease/control enrichment
m^6t^6A	505.1090	505.0983	8.1	150
hn^6A				
oQ	504.1137	504.1142	6.3	2.7
t^6A	491.0933	491.0971	2.0	13
Q	488.1333	488.1330	2.6	22
ms^2i^6A	460.1061	460.1089	13.7	142
$cmnm^5s^2U$	426.0378	426.0381	2.0	11
i^6A	414.1184	414.1173	10.3	67
$mcmp^5U$	411.0446	411.0434	3.9	206
cmo^5U	397.0290	397.0314	0.8	9.0
$m^2{}_2G$	390.0820	390.0802	6.2	109
$preQ_0$	386.0507	386.0493	0.8	1.7
mnm^5s^2U	382.0480	382.0498	1.6	1.5
m^1G	376.0664	376.0678	1.6	88
m^2G			1.8	3.9
m^7G			2.0	5.1
Gm			2.6	7.9
$m^6{}_2A$	374.0871	374.0878	7.0	14
ac^4C	364.0551	364.0543	1.2	64
m^1A	360.0715	360.0710	4.3	36
m^2A			5.3	54
m^6A			5.8	7
m^5s^2U			1.0	5.2
ho^5U	339.0235	339.0249	1.9	19
s^2U			1.6	5.6
s^2C	338.0217	338.0271	0.8	6.3
m^3U	337.0442	337.0442	1.3	19
Um			1.7	37
m^5U			1.9	13
m^3Y			2.4	5.6
m^3C	336.0602	336.0602	1.1	4.1
m^5C		336.0605	1.4	2.2
Cm		336.0611	1.7	3.5
D	325.0442	325.0461	0.8	12

Supplementary Figure 2. (**a**) 3'-aminoacyl adenosine monophosphates from total *E. coli* RNA detected with the method developed in this work. (**b**) Species detected with the nuclease digestion method that are consistent with known rRNA and tRNA nucleotide modifications from total *E. coli*. Although plausible identifications are shown based on high-resolution LC/MS data, we note that definitive identifications require comparison with authentic samples. Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

a

amino acid	expected $[M-H]^-$ m/z	observed $[M-H]^-$ m/z	nuclease/control enrichment
alanine	417.0929	417.0936	11
aspartic acid	461.0827	461.0816	7.8
glutamic acid	475.0984	475.0978	69
glutamine	474.1144	474.1149	20
glycine	403.07724	403.0770	41
isoleuicine	459.1399	459.1396	148
leucine	459.1399	474.1400	97
lysine	474.1507	474.1503	25
methionine	477.0963	477.0960	12
phenylalanine	493.1242	493.1238	62
proline	443.10854	443.1061	10
serine	433.0878	433.0887	10
threonine	447.1035	447.1035	33
tryptophan	532.1351	532.1348	28
tyrosine	509.1191	509.1196	16
valine	445.1242	445.1240	163

b

RNA modification	expected $[M-H]^-$ *m/z*	observed $[M-H]^-$ *m/z*	retention time (min)	nuclease/control enrichment
m^6t^6A	505.1090	505.0952	8.2	145
hn^6A				
Q	488.1333	491.0922	2.1	26
ms^2i^6A	460.1061	460.1085	13.7	29
$cmnm^5s^2U$	426.0378	426.0342	1.6	53
$cmnm^5Um$	424.0763	424.0680	1.0	3.0
inm^5U	420.1177	420.1159	6.0	11
i^6A	414.1184	414.1191	10.4	123
$m^2{}_2G$	390.0820	390.0852	6.2	99
m^7G			2.0	58
Gm			2.6	6.7
$m^6{}_2A$	374.0871	374.0892	7.0	83
nm^5s^2U	368.0323	368.0359	1.3	7.3
ac^4C	364.0551	364.0563	1.0	95
m^1A	360.0715	360.0739	5.0	3.5
m^2A			5.3	22
m^6A			5.7	26
m^5s^2U			1.0	5.2
m^3U	337.0442	337.0448	1.6	74
Um			2.0	6.7
m^5U			2.3	18
m^3Y			3.3	4.9
m^3C	336.0602	336.0572	1.4	49
m^5C			2.0	33
Cm			2.3	7.3
D	325.0442	325.0442	1.0	64

Supplementary Figure 3. (**a**) 3'-aminoacyl adenosine monophosphates from total *S. venezuelae* RNA detected with the nuclease digestion method. (**b**) Species detected with the nuclease digestion method that are consistent with known rRNA and tRNA nucleotide modifications from total *S. venezuelae*. Although plausible identifications are reported based on high-resolution LC/MS data, definitive identifications require additional comparisons with authentic samples of each modified nucleotide. Isomeric nucleotides are listed sequentially, with expected *m/z* values shown once.

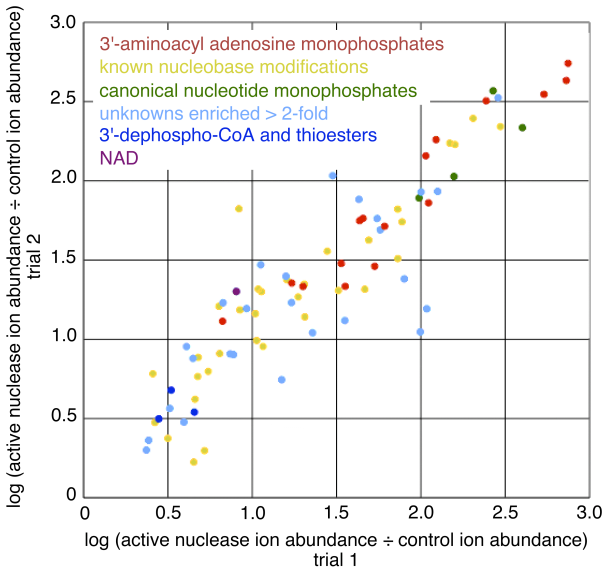
a

E. coli		
observed [M-H]⁻ m/z	retention time (min)	nuclease/control enrichment
729.1087	1.7	117
721.1160	5.8	102
668.1103	2.3	100
642.0883	3.7	72
725.1068	1.6	57
643.1063	2.4	55
473.0827	4.0	36
655.1179	4.0	33
820.2175	7.2	23
771.0726	1.4	23
490.1119	5.7	21
789.1588	12.6	20
565.1339	6.5	20
726.0653	4.9	19
760.0740	1.7	17
822.2748	2.7	17
536.0481	1.0	16
824.1958	13.8	14
540.0533	**5.3**	**8**
602.1188	6.9	8
965.2397	8.3	7
786.1589	**6.8**	**5**
801.1505	8.4	4
654.3013	12.1	4
686.1414	**7.1**	**3**
728.1541	**7.9**	**3**
811.3491	2.4	2

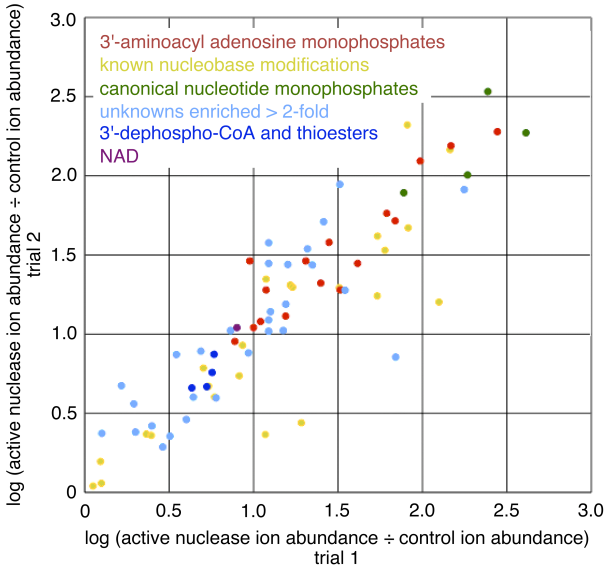
b

S. venezuelae		
observed [M-H]⁻ m/z	retention time (min)	nuclease/control enrichment
659.0447	6.8	189
668.1072	2.4	97
602.1252	6.9	51
725.0881	1.1	38
822.2739	2.7	35
565.1754	5.6	28
536.0485	1.0	28
771.0714	1.3	23
616.1401	7.3	22
651.0952	2.0	21
652.0778	4.7	18
372.0320	2.5	14
820.2169	7.2	12
424.0033	1.0	11
750.0549	1.1	11
789.1590	12.6	10
540.0607	**5.3**	**8**
859.0962	1.0	8
595.1417	7.0	8
549.0873	5.8	7
572.1117	6.4	6
965.2407	8.2	4
801.1506	8.4	4
686.1436	**7.1**	**4**
786.1563	**6.6**	**3**
691.1004	2.1	3
654.2997	12.0	3
728.1506	**7.9**	**3**
786.1567	**6.8**	**3**
816.1701	6.6	2
742.1639	6.8	2
811.3444	2.1	2

c



d

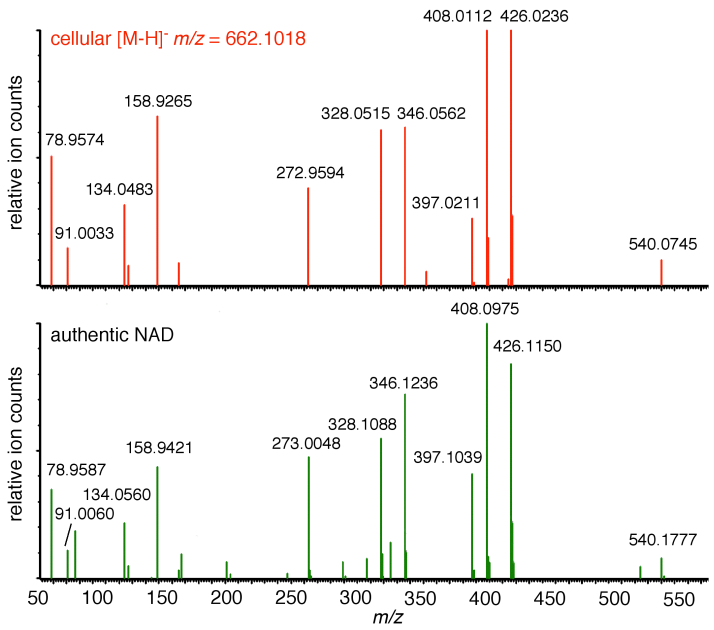


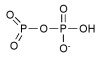
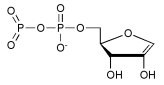
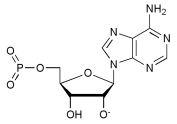
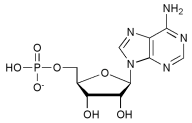
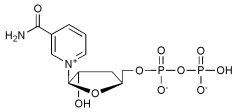
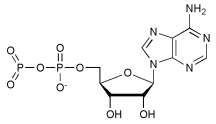
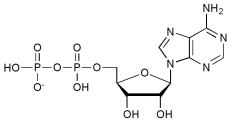
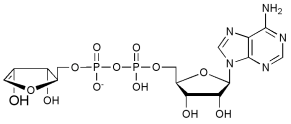
e

				free amino acid ion counts			
				E. coli		*S. venezuelae*	
amino acid	expected [M-H]⁻ *m/z*	observed [M-H]⁻ m/z	retention time (min)	cleavage conditions	control conditions	cleavage conditions	control conditions
phenyalanine	164.0717	164.0722	3.8	29	32	42	38
tyrosine	180.0666	180.0675	1.9	113	118	126	125
tryptophan	203.0826	203.0830	6.5	36	35	67	71

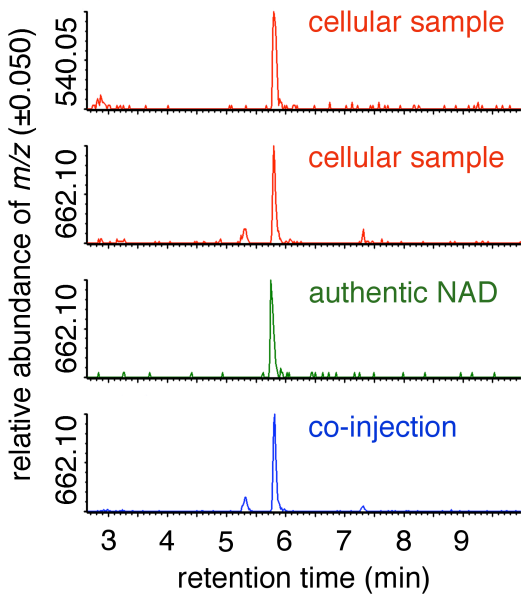
Supplementary Figure 4. (**a**) Unknown species detected by the nuclease digestion method applied to *E. coli* RNA that satisfy the criteria described in the text. (**b**) Unknown species detected by the nuclease digestion method applied to *S. venezuelae* RNA that satisfy the criteria described in the text. The NAD fragment is listed in red, and the CoA derivatives are listed in blue. (**c**) Result of two independent trials (r = 0.90) of the method described in the main text applied to total *E. coli* RNA. The observed species include 16 3'-aminoacyl adenosine monophosphates, 33 known nucleotide modifications, the four canonical RNA nucleotides, 3'-dephospho-CoA and two thioester derivatives, and 24 additional unknown species with a control:base ratio ≥ 2-fold. (**d**) Results of two independent trials (r = 0.93) of the nuclease method applied to total *S. venezuelae* RNA. The observed species include 16 aminoacyl adenosine monophosphates (red), 25 known nucleobase modifications (yellow), four canonical RNA nucleotides (green), 3'-dephospho-CoA and 3'-dephospho-CoA thioesters (dark blue)[1],

NAD (purple), and 24 additional unknown species with a nuclease:control ratio ≥ 2-fold (light blue), 20 of which were not discovered with our previously reported nucleotide cleavage method[1]. In both **(c)** and (**d**) each point represents the average of three experiments. (**e**) After macromolecule purification (main text Figure 1a), the RNA is exposed to the same pH, temperature, and time under both cleavage and control conditions. Therefore the rate of amino acid hydrolysis from tRNA or 3'-aminoacyl adenosine monophosphates should be equivalent under both conditions. Following size-exclusion purification of the small molecules (after nuclease or non-nuclease conditions), we expect to retain the same amount of free amino acid in each case. We observe the same quantity of free amino acids in both cleavage and control conditions, suggesting that there is minimal systematic variation between the treatment of cleavage and control samples prior to LC/MS analysis. The ion counts shown represent the average of three experiments.

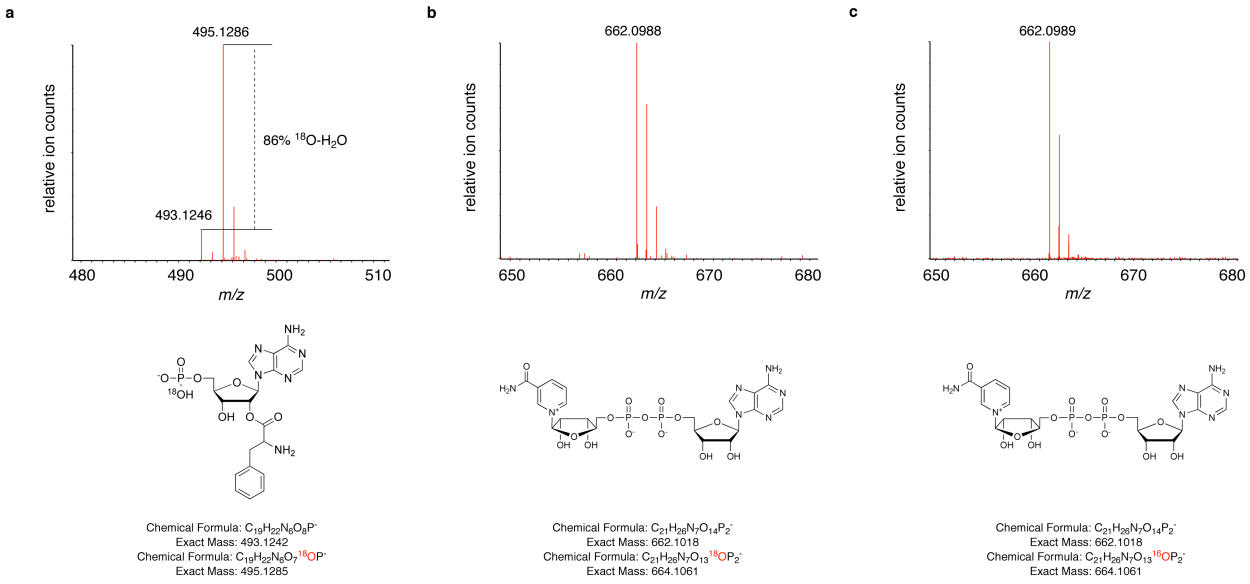


observed [M-H]- *m/z*	proposed structure
78.9574	
134.0483	
158.9265	
272.9594	
328.0515	
346.0562	
397.0211	
408.0115	
426.0236	
540.0745	

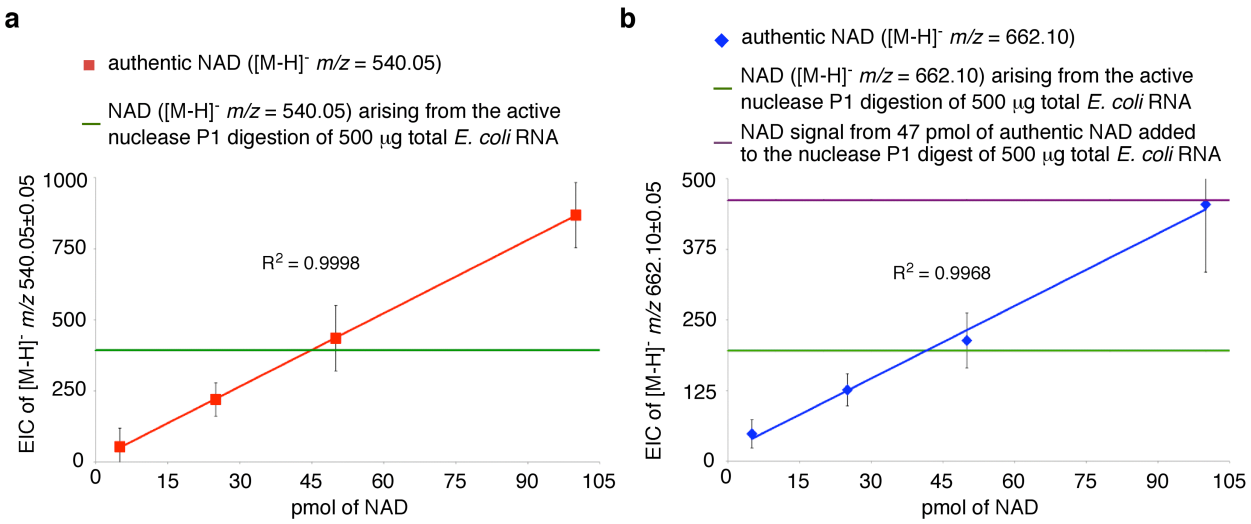
Supplementary Figure 5. MS/MS fragmentation of *E. coli* cellular species $[M-H]^-$ *m/z* = 662.1018 and comparison with authentic NAD confirms the assignment of the $[M-H]^-$ *m/z* = 662.1018 as NAD. A table of plausible fragment ion structures is provided.



Supplementary Figure 6. EIC comparison of *E. coli* cellular RNA nuclease P1 digestion products at both the original ($[M-H]^-$ m/z = 540.0533) and mild ($[M-H]^-$ m/z = 662.1018) ionization conditions with authentic NAD.



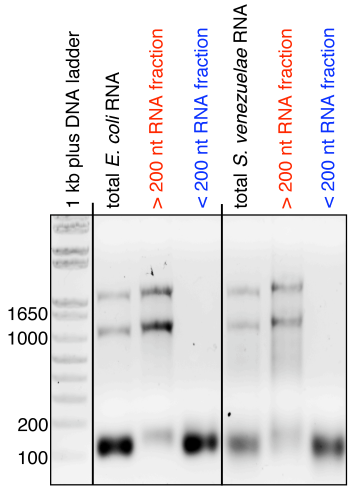
Supplementary Figure 7. In the presence of ^{18}O water, all nuclease P1 digestion products other than the nucleotides at the 5' ends of substrates will exhibit a +2 Da mass shift compared with the products arising from digestion in ^{16}O water. (**a**) Phenylalanyl AMP, a known small molecule-RNA conjugate that exists at the 3' end of tRNA, exhibits the expected +2 Da shift. (**b**) No mass shift is observed in the case of NAD, indicating that this modification is located at the 5' terminus of the RNA(s). (**c**) Authentic NAD has a similar isotope pattern as NAD from nuclease P1 digestion of total RNA.



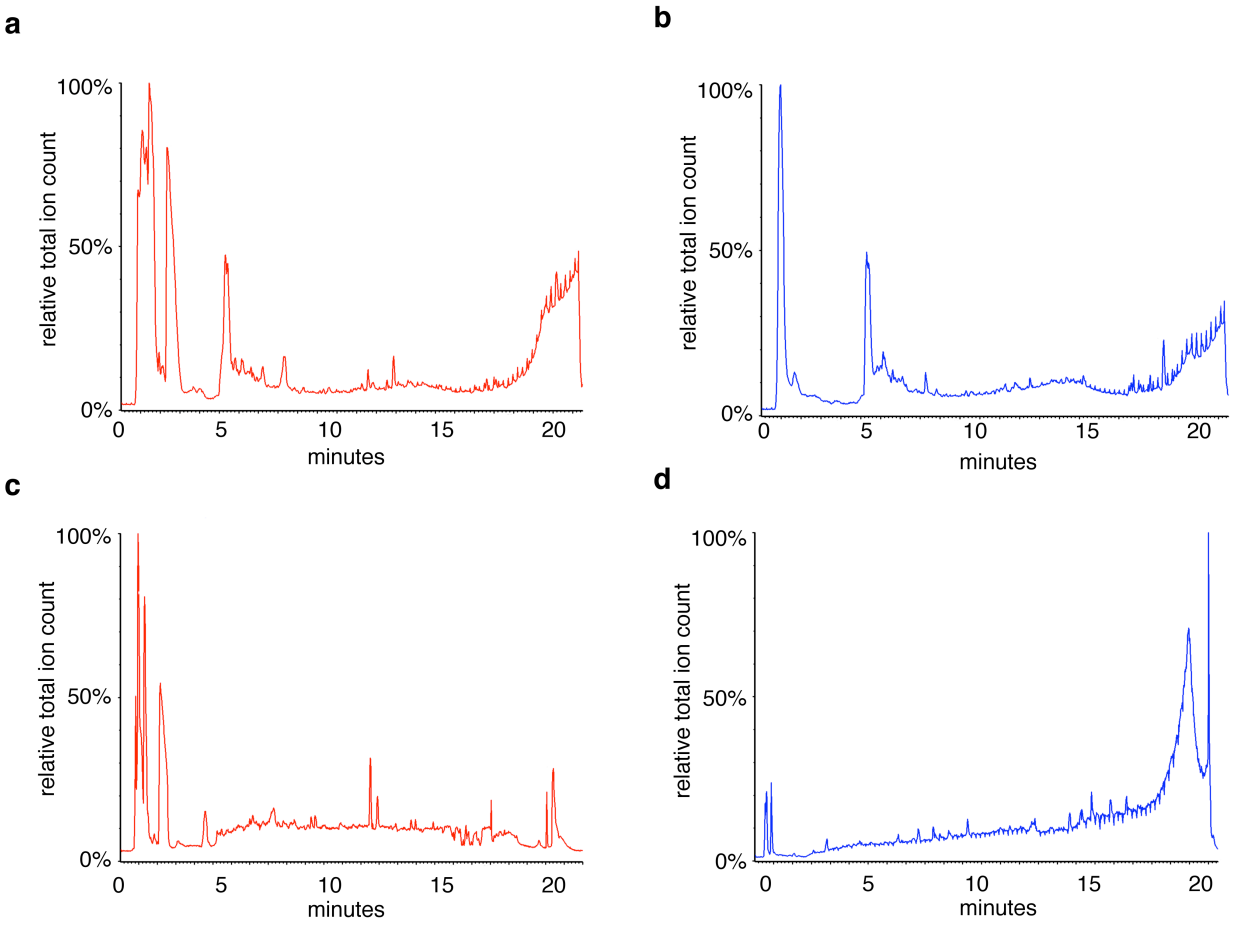
Supplementary Figure 8. Standard curves relating known NAD quantities to the abundance of the major daughter ion ([M-H]$^-$ *m/z* = 540.0538) or of the parent ion ([M-H]$^-$ *m/z* = 662.1018). These curves were used to determine the amount of NAD-RNA per *E. coli* cell. (**a** and **b**) The amount of NAD signal arising from nuclease P1 digestion of total *E. coli* RNA is indicated with the green lines. (**b**) The purple line indicates the observed NAD signal when 47 pmols of authentic NAD is added to nuclease P1-digested total *E. coli* RNA. All three observations suggest that there are ~43-47 pmol of NAD-RNA per 500 μg of E. coli RNA, which corresponds to ~3,000-3,300 copies of NAD-RNA per *E. coli* cell. Error bars represent the standard deviation of three independent trials.

DNA template	spiked NAD-RNA (pmol)	*E. coli* RNA pol.	NADH (mM)	RNA yield (µg)	expected ion counts of NAD or NADH from transcription	expected ion counts of NAD or NADH from spiked NAD-RNA	observed ion counts of NAD or NADH
pUC19+1A	0	-	0.5	1.5	0	0	8
pUC19+1A	0	+	0.5	222.4	≥ 636	0	6
pUC19+1A	0	+	5.0	173.8	≥ 495	0	11
E. coli genomic DNA	0	-	0.5	1.7	0	0	8
E. coli genomic DNA	0	+	0.5	78.4	≥ 217	0	6
E. coli genomic DNA	0	+	5.0	69.8	≥ 192	0	3
E. coli genomic DNA	1.0	-	0.5	2.2	0	9	8
E. coli genomic DNA	1.0	+	0.5	86.4	≥ 240	9	10
E. coli genomic DNA	2.3	-	0.5	2.0	0	20	16
E. coli genomic DNA	2.3	+	0.5	72.9	≥ 201	20	22
E. coli genomic DNA	4.5	-	0.5	1.6	0	38	45
E. coli genomic DNA	4.5	+	0.5	91.0	≥ 254	38	48

Supplementary Figure 9. Aberrant transcription initiation by *E. coli* RNA polymerase does not account for NAD-RNA *in vivo*. *In vitro* transcription with *E. coli* RNA polymerase in the presence of NADH does not generate significant quantities of NAD-linked RNA. *E. coli* RNA polymerase was used to generate RNA *in vitro* from plasmid or genomic DNA templates in the presence of 0.5 mM or 5.0 mM of NADH. When the resulting RNA was digested with nuclease P1 and analyzed by LC/MS, the observed (NAD+NADH) signal (column 8) was much lower than the expected ion counts if the observed quantities of NAD-RNA in *E. coli* arise from aberrant transcriptional initiation with NAD or NADH (column 6). Spiking 1.0, 2.3, or 4.5 pmol of authentic NAD-linked RNA (generated with T7 RNA polymerase; see the main text) into the *in vitro* transcription reaction resulted in observed ion counts that were consistent with, but not significantly greater than, the expected levels of NAD ions from the spiked material (column 7). Column 6 lists the expected ion counts of (NAD+NADH) assuming that the fraction of RNA generated from *in vitro* transcription under these conditions is at least as great as that observed from *E. coli* cellular RNA



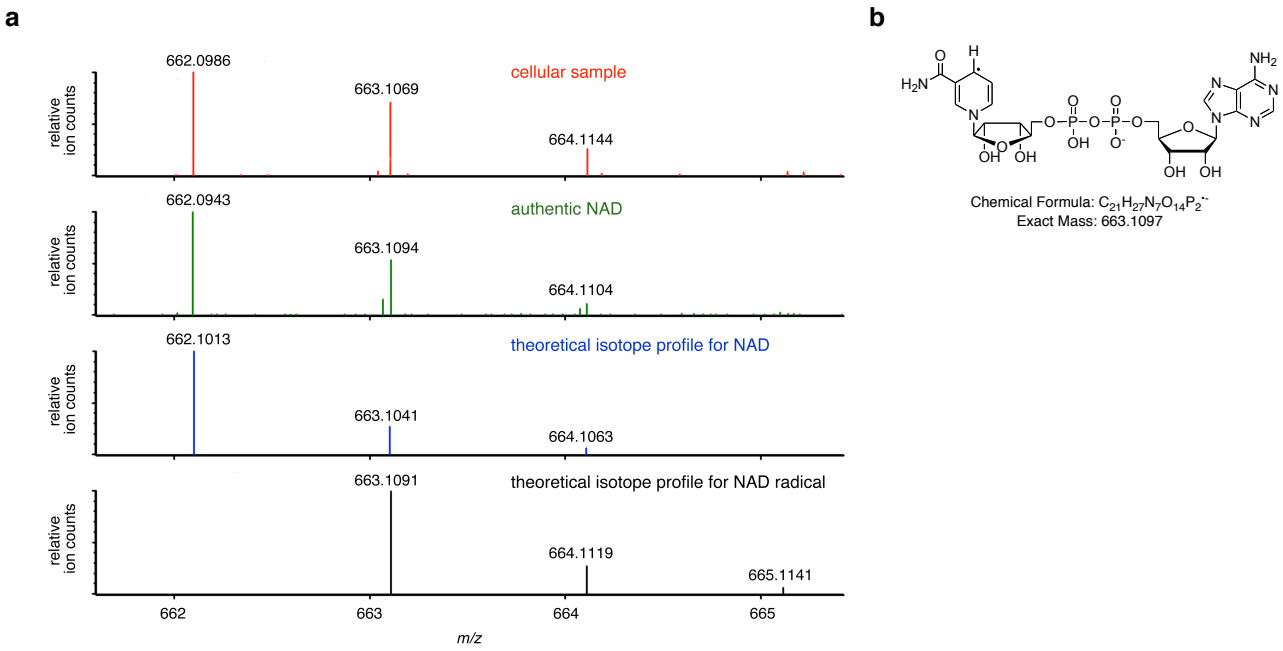
Supplementary Figure 10. Silica column-based size fractionation of cellular RNA. Total *E. coli* or *S. venezuelae* RNA was separated into fraction I (RNAs of length > ~200 nucleotides) and fraction II (RNAs of length < ~200 nucleotides) using a Qiagen RNeasy silica column. The starting total RNA and each fraction (1 µg per lane) was analyzed by 1% TAE-agarose gel electrophoresis and stained with ethidium bromide.



Supplementary Figure 11. Total ion chromatograms of total RNA from *E. coli* (**a**) nuclease P1 digestion, original ionization conditions. (**b**) heat-inactivated nuclease P1 digestion, original ionization conditions. (**c**) nuclease P1 digestion, mild ionization conditions. (**d**) heat-inactivated nuclease P1 digestion, mild ionization conditions. Each chromatogram is scaled relative to the MS scan with the greatest aggregate ion count in that chromatogram.

	observed NAD (pmol) [M-H]$^-$ *m/z* 662.1018	observed NADH (pmol) [M-H]$^-$ *m/z* 664.1175
authentic NAD (50 pmol) carried through mock RNA isolation	49	0
authentic NADH (50 pmol) carried through mock RNA isolation	45	4
authentic NAD added to nuclease P1-treated total *E. coli* RNA	49	0
authentic NADH added to nuclease P1-treated total *E. coli* RNA	47	2
no authentic NAD or NADH added to nuclease P1-treated total *E. coli* RNA	43	0

Supplementary Figure 12. NADH oxidizes to NAD during RNA isolation, sample processing and analysis. The mock RNA isolation mimicked the total RNA isolation from bacteria protocol, but omitted the isopropanol precipitations and size-exclusion chromatography steps (see the Supplementary Experimental Methods above). Observed ion counts are converted into pmol based on NAD and NADH standard curves (**Supplementary Figs. 8a** and **8b**).



Supplementary Figure 13. NAD radical may contribute to the $[M-H]^-$ *m/z* 663.1069 signal in the authentic and cellular NAD spectra. (**a**) Mass spectra of cellular and authentic NAD, and the theoretical isotope profiles for NAD and NAD radical. (**b**) Structure of NAD radical.

Supplementary References

1. Kowtoniuk, W.E., Shen, Y., Heemstra, J.M., Agarwal, I. & Liu, D.R. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci U S A* **106**, 7768-73 (2009).
2. Crain, P.F. & James, A.M. [42] Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. in *Methods in Enzymology*, Vol. Volume 193 782-790 (Academic Press, 1990).
3. Crain, P.F. Mass spectrometric techniques in nucleic acid research. *Mass Spectrometry Reviews* **9**, 505-554 (1990).
4. Roberts, G.C., Dennis, E.A., Meadows, D.H., Cohen, J.S. & Jardetzky, O. The mechanism of action of ribonuclease. *Proc Natl Acad Sci U S A* **62**, 1151-8 (1969).
5. Usher, D.A. On the mechanism of ribonuclease action. *Proc Natl Acad Sci U S A* **62**, 661-7 (1969).
6. Usher, D.A., Richardson, D.I., Jr. & Eckstein, F. Absolute stereochemistry of the second step of ribonuclease action. *Nature* **228**, 663-5 (1970).
7. Grosjean, H.B.R. *Modification and editing of RNA*, xiii, 596 p. (ASM Press, Washington, DC, 1998).
8. Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res* **31**, e8 (2003).
9. Hopwood, D.A. *Genetic manipulation of Streptomyces: a laboratory manual*, xii, 356 p. (John Innes Foundation, Norwich, 1985).
10. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-87 (2006).
11. Neidhardt, F.C.I.J.L.S.M. *Physiology of the bacterial cell : a molecular approach*, xii, 506 p. (Sinauer Associates, Sunderland, Mass., 1990).