



Published in final edited form as:

*Nat Chem.* 2015 September ; 7(9): 752–758. doi:10.1038/nchem.2307.

## Organocatalytic Removal of Formaldehyde Adducts from RNA and DNA Bases

Saswata Karmakar, Emily M. Harcourt, David S. Hewings, Alexander F. Lovejoy<sup>1</sup>, David M. Kurtz<sup>1</sup>, Thomas Ehrenschwender, Luzi J. Barandun, Caroline Roost, Ash A. Alizadeh<sup>1</sup>, and Eric T. Kool<sup>2</sup>

Department of Chemistry, Stanford University, Stanford, CA 94305

<sup>1</sup>Divisions of Oncology and of Hematology, Stanford School of Medicine, Stanford, CA 94305

### Abstract

Formaldehyde is universally employed to fix tissue specimens, where it forms hemiaminal and aминаl adducts with biomolecules, hindering the ability to retrieve molecular information. Common methods for removing these adducts involve extended heating, which can cause extensive degradation of nucleic acids, particularly RNA. Here we show that water-soluble bifunctional catalysts (anthranilates and phosphanilates) speed the reversal of formaldehyde adducts of mononucleotides over standard buffers. Studies with formaldehyde-treated RNA oligonucleotides show that the catalysts enhance adduct removal, restoring unmodified RNA at 37 °C even when extensively modified, and avoiding high temperatures that promote RNA degradation. Experiments with formalin-fixed, paraffin-embedded cell samples show that the catalysis is compatible with common RNA extraction protocols, with detectable RNA yields increased by 1.5–2.4 fold using a catalyst under optimized conditions, and by 7–25 fold compared to a commercial kit. Such catalytic strategies show promise for general use in reversing formaldehyde adducts in clinical specimens.

There is a major trend in medicine toward molecular characterization of disease in order to improve individual patient treatment<sup>1,2</sup>. Virtually all clinical tissue specimens (including biopsies and surgical specimens) are prepared prior to analysis by fixation with formalin (formaldehyde) in formalin-fixed paraffin-embedded (FFPE) tissue block format<sup>3</sup>. Formaldehyde treatment causes extensive molecular crosslinks and adducts to the biomolecules in the sample, greatly diminishing the signals that can be obtained from later molecular analysis via hybridization, PCR, and sequencing<sup>4–6</sup>. These adducts limit the length of PCR amplicons that can be analyzed<sup>7</sup>, hinder or prevent the quantitation of RNA transcripts<sup>8</sup>, and diminish antigen signals in immunohistochemistry<sup>9</sup>.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>2</sup>Corresponding author: kool@stanford.edu.

**Author Contributions.** E.T.K. designed the project, and A.A.A., A.F.L., and D.M.K. designed the cell studies. S.K., E.M.H. and D.H. carried out the experiments, with early exploratory studies carried out by T.E. The catalyst **3** was synthesized by S.K. and L.J.B., and RNA was prepared by C.R. The manuscript was written by E.T.K. with input from E.M.H., S.K., A.A.A. and D.H.

Current methods for treating formalin-fixed tissue prior to analysis involve extended heating in buffers (commonly Tris)<sup>6,10</sup>. Heating removes only a fraction of the adducts, and conditions are harsh enough (60–70 °C for several hours to days of incubation<sup>11</sup>) that nucleic acids in the specimen can be permanently damaged in the process<sup>12,13</sup>. Finding strategies to efficiently remove formaldehyde adducts from RNA and DNA under mild conditions could constitute a major advance in molecular pathology in general, and facilitate the diagnosis and prognosis of a broad range of cancers.

Surprisingly, despite widespread and long-term recognition of this problem among researchers, few if any studies exist that analyze and exploit the chemical mechanisms of formation and breakdown of hemiaminal and aminor adducts to find solutions. Formation of formaldehyde adducts of biological amines involves the intermediacy of tetrahedral adducts and imine structures, based on extensive studies of the mechanism of imine formation by Jencks and others<sup>14,15</sup>. Catalysts that promote imine formation<sup>16–18</sup> and breakdown of tetrahedral intermediates<sup>19,20</sup> have the potential to speed transamination reactions in both the forward and reverse directions. Our recent development of bifunctional transamination catalysts that accelerate the formation of hydrazones and oximes<sup>19,20</sup> led us to consider the possibility that such compounds might also exhibit activity in the reversal of imine-related hemiaminal and aminor structures that are commonly formed on exocyclic amines of nucleic acid bases in the presence of formaldehyde<sup>4</sup>. Here we describe the discovery of water-soluble organocatalysts that speed the reversal of formaldehyde adducts from RNA and DNA bases.

## Results and Discussion

### Preparation and characterization of model formalin-adducted nucleotides

To test whether transamination catalysts speed the reversal of formaldehyde linkages, we prepared water-soluble adducts of monomeric nucleotides AMP and dAMP that could serve as well-behaved, kinetically characterizable models of polymeric nucleic acids. The N6-hydroxymethyl monoadduct of dAMP<sup>21</sup> was prepared in the presence of 10% formaldehyde (see details in Supporting Information). NMR and mass spectrometric analysis show that it exists primarily in the hemiaminal state (Fig. 1) as previously reported, presumably in equilibrium with the dehydrated imine form, which is not observed. We used reverse-phase HPLC to resolve the monoadduct from unmodified dAMP (Figs. S1–S3 in the Supporting file); in pH 7.0 buffer at room temperature, we find that this nucleotide formaldehyde adduct slowly reverts to unmodified dAMP, with a half-life of ca. 6 h (23 °C). For an aminor crosslink model, we prepared a previously unidentified N6-dimer of AMP in buffer by extended incubation of AMP with 10% formaldehyde<sup>22</sup>. Our early experiments revealed that this aminor is much more stable than the hemiaminal model, showing little or no reversal to the uncrosslinked state in neutral buffer over at least a day at room temperature as monitored by HPLC (Fig. S4).

### Bifunctional catalysts promote removal of formaldehyde adducts

Having these model systems in hand, we proceeded to test the effects of potential catalysts on the reversal of the adducts. We screened a range of arylacids and amines, including a

number of bifunctional compounds containing both groups, since such compounds were previously observed to behave as catalysts of transimination<sup>19,20</sup>. Compounds were tested at 5 mM in 30 mM pH 7 Tris buffer, and we measured yields of reversal after 1 and 2 h at 23 °C (Figs. S5,6). Although arylphosphonic acids and arylsulfonic acids were initially found to be active, they were removed from further consideration after it was found that their activity resulted from solution acidification after overwhelming the buffer. Of the remaining compounds, phosphanilate compound **3** was found to promote substantial adduct reversal for both the hemiaminal and aminal models. Bifunctional anthranilate catalyst **2** was also among the more active compounds. Notably, both compounds were previously identified as highly active catalysts for imine formation reactions<sup>19–20</sup>. For formaldehyde adduct reversal, the effects of catalysts added at 5 mM were substantial; for example, the monoadduct was >50% reversed after 2 h with compound **3**, whereas in buffer alone only 11% reversal was seen. Although the aminal dimer was much more stable, showing ~0.5% reversal in buffer at 2 h, the catalyst yielded ca. 14% reversal at this time point (23 °C).

Next we carried out more detailed analysis of the reactions catalyzed by bifunctional compounds **2** and **3**, comparing them to the classic transimination catalyst aniline (**1**)<sup>23</sup>. Testing both reactions at pH 4.5 and 7, we found that catalyst **3** was effective in reversal of the hemiaminal adduct at pH 7, while the catalyst was more effective at lower pH with the aminal dimer (most experiments were carried out at pH 4.5; Figs. S6–S9). Measurement of reaction yields at early time points allowed us to assess relative rates (Fig. 2). For the hemiaminal reversal reaction, we found that phosphanilate **3** (10 mM) enhanced rate over buffer alone by 4-fold (pH 7.0, 23 °C), and by a factor of 3 over aniline. For the more stable aminal dimer, we found that compound **3** accelerated the rate of dimer reversal by 37-fold over buffer alone and by 10-fold over aniline (pH 4.5, 23 °C) (Fig. 2). For both reactions, compound **2** fell between **3** and aniline in its ability to promote adduct reversal.

In order to gain more insight into the mechanism by which these catalysts were active, we carried out experiments varying conditions and catalysts. First, by varying buffer pH we investigated the role of acid catalysis on the rate of adduct reversal. We found that the hemiaminal reversal was relatively insensitive to pH, based on the observation that rates of reaction were little different in phosphate buffers at pH 7 or pH 4.5 (Fig. S6). In contrast, the aminal crosslinked dimer was affected by pH, with substantial rate accelerations occurring at lower pH in citrate buffer (Fig. S7). This suggests that aminal reversal is subject to specific acid catalysis, presumably by direct protonation of one of the nitrogens of the aminal<sup>24</sup>. On the other hand, hemiaminal reversal does not appear to be subject to specific acid catalysis over this pH range, which is consistent with its lower basicity. Next we varied catalyst concentrations, using compound **3** to test for evidence of general acid catalysis. The results showed that for both reactions, increasing concentrations of **3** led to approximately proportionate rate enhancements (Figs. S8,9). This establishes the presence of the catalyst at the transition state for these reactions, and is consistent with general acid catalysis as part of its mechanism of action. Notably, the buffers did not show concentration dependence in these two reactions, and thus are likely not interacting with the substrates at the transition states.

Since catalyst **3** was the most efficient promoter of formaldehyde adduct reversal among compounds tested, we investigated the role of its structure by comparing it to analogs having amine or phosphonate functional groups omitted. The results showed (Fig. 3; Figs. S10,11) that both of these functional groups are necessary to achieve most effective catalysis, as controls with amine or phosphonate groups alone yielded slower reaction. Whether bifunctional compound **3** could foster true turnover catalysis was tested in an experiment using high concentration of a minimal dimer, in molar excess over catalyst (Fig. S12); the data showed a catalyst advantage relative to buffer over at least three apparent turnovers. However, some product inhibition may occur at high formaldehyde concentrations, since an adduct between cat. **3** and formaldehyde was observable by NMR (data not shown). We note that the issue of turnover is likely moot in application with formalin-fixed tissues, where catalyst is expected to be in large excess over substrate. Overall, the data show that bifunctional catalysts are most effective in these reactions, and that general acid catalysis in promoting bond breakage is likely to play a role in their mechanism. It is possible that nucleophilic catalysis may also be involved, since transamination likely proceeds via this mechanism<sup>23</sup>. A hypothetical mechanism that explains roles for both the acid and amine functional groups is proposed in Fig. S13, although further mechanistic studies are needed to confirm this hypothesis.

### **An RNA oligonucleotide is extensively modified by formaldehyde**

We developed a full RNA strand model to assess formaldehyde adducts in the biopolymer. This was done to investigate (a) to what extent formaldehyde adducts are formed on a native biopolymer as opposed to a mononucleotide; (b) to what degree standard heating-in-buffer protocols affect biopolymer stability and adduct removal, and (c) whether our observed catalysis of the reversal of formaldehyde adducts from a mononucleotide might extend to polymeric RNA. To this end, we designed a 16mer RNA strand having a central self-complementary sequence flanked by 5mer (A)<sub>5</sub> ends. This is expected to form a short duplex (Fig. 4A) with several overhanging adenosines to promote adduct and crosslink formation in the unpaired bases. To analyze adduct numbers and types qualitatively and quantitatively, we used denaturing polyacrylamide gel electrophoresis and MALDI-mass spectrometry. To generate adducts, the RNA was treated with 10% formaldehyde for 24 h at 23 °C, precipitated, then redissolved in buffer for analysis of reversal.

Analytical data with this formaldehyde-treated RNA showed clear evidence of extensive adduct formation. PAGE analysis revealed evidence of a high degree of crosslinking to form dimers, and also showed shifting of the monomer RNA band, consistent with the base adducts causing slowed mobility during electrophoresis (Fig. S14). These observations suggested that little unmodified RNA remains under these treatment conditions. Mass spectrometric analysis confirmed this, revealing no detectable unmodified RNA after formaldehyde treatment, while untreated RNA gave a clear parent ion peak at 5181 Da (Fig. 4). The observed masses after formaldehyde treatment corresponded to RNA containing ~1–14 adducts, with a median of ca. 9 adducts. The monomer RNA contains 16 exocyclic amine groups on adenine, cytosine and guanine, and thus the results show that a large fraction of these groups on each RNA molecule carry an adduct. It is also possible that more than one adduct can form on a given exocyclic amine. Closer inspection of the mass spectrum of

products around the monomeric RNA mass range (5000–6000 daltons) shows peaks consistent with increasing numbers of adducts primarily in hemiaminal form, adding 30 daltons of mass successively (see Fig. 4B).

### **Catalyst-enhanced removal of formaldehyde from an RNA oligonucleotide at low temperature**

Using this extensively formaldehyde-modified RNA, we proceeded to test the effect of catalysts and conditions on loss of adducts and on recovery of intact RNA. Simple incubation in the presence of 8 mM catalysts at pH 7 (37 °C) showed relatively rapid apparent loss of adducts as judged by gel electrophoresis, with more efficient loss of crosslinked and shifted bands in the presence of catalysts as opposed to Tris buffer alone (Fig. S14). Analysis by mass spectrometry of adduct RNAs having mass near that of the 16mer confirmed this; by analyzing counts of mass peaks relative to a DNA control spiked into the sample we were able to quantitatively measure amounts of hemiaminal adducts and intact RNA. Incubation at 37 °C with the catalyst over 18 h revealed loss of adducts and recovery of intact unmodified RNA as a major peak by mass spectrometry (Fig. 5A). The quantitative data showed that catalyst **3** effectively enhanced the rate of adduct loss, increasing the ratio of intact RNA generated relative to Tris buffer alone at 37 °C. Heating at a literature-standard 60 °C in buffer increased the rate of loss of formaldehyde adducts, but importantly, it also caused rapid degradation of the RNA, with > 40% degraded after 12 h (Figs. 5B and S15). Thus the data suggest strongly that heating is to be avoided in RNA recovery<sup>25</sup>, and that milder incubation at 37 °C in the presence of catalysts is effective at recovering a high yield of unmodified RNA from extensive hemiaminal adducts after 12–18 h incubation, even given an extensive degree of modification. Similar catalyst incubation with a formalin-treated self-complementary DNA duplex also revealed enhanced loss of formaldehyde adducts from this double-stranded biopolymer relative to buffer treatment alone (Fig. S16).

Next we explicitly tested catalyst effect on aminated-crosslinked RNAs, by quantifying dimer RNA bands and their conversion to monomer RNAs via denaturing gel electrophoresis analyzed by phosphorimaging. The data show that catalyst **3** enhances the reversal of crosslinks between two 16mer RNA strands, yielding a consistently greater degree of uncrosslinking relative to buffer alone at the same pH (Fig. 5C).

### **Catalyst-enhanced recovery of RNA from paraffin-embedded formalin-fixed cells**

Cellular RNAs fold into highly varied structures and thus are expected to react with formaldehyde to yield a greater diversity of hemiaminal and aminated adducts than tested in the above model systems. Moreover, recovery of RNA from formalin-fixed tissues requires additional steps including paraffin removal and proteolysis. Thus we sought to test whether our organocatalytic approach would retain benefits in RNA recovery in the more complex cellular milieu, and whether the presence of catalyst might interfere with the common steps of proteolysis, extraction and isolation, and PCR amplification that are carried out with clinical FFPE specimens. Because primary FFPE specimens from patients are highly variable from sample to sample<sup>25</sup>, we prepared reproducible FFPE specimens via a literature

approach<sup>26</sup>, using a cultured cell line (Raji cells<sup>27</sup>), treating the cell pellet with formalin and embedding it in a paraffin block following standard procedures<sup>26</sup>.

Formalin-fixed samples were deparaffinized with xylenes and treated with proteinase K following standard procedures with a widely used commercial FFPE extraction kit. Catalyst **3** was then added (20 mM), and incubation times and temperatures varied to optimize RNA yield (Fig. S17), with an eye to avoiding high temperatures that rapidly degrade RNA (Fig. S15). RNA was recovered by commercial silica-based spin columns from the kit and detected by qRT-PCR; experiments showed that this method of RNA isolation was not adversely affected by the presence of catalyst. Notably, the addition of catalyst increased yields at all temperatures and times tested. Using lower temperatures (45–65 °C) and prolonged incubation times allowed greater yields of RNA to be recovered, compared to the commercial kit (80 °C). 18 h incubation at 55 °C appeared to be optimal (Fig. S17); higher temperatures or longer times (65 °C, 24 h) offered only marginal benefits or even slightly reduced yields, consistent with heat-induced RNA degradation by hydrolysis.

RNAs from three different genes (*GAPDH*, *AGO1*, *AICDA*) with varied copy number and amplicon lengths (85–514 bp) were then quantified, comparing the optimized catalyst protocol (which employs heating at 55 °C) to the commercial kit protocol (80 °C) and a widely used literature protocol<sup>28</sup> (70 °C). Experiments revealed (Fig. 6) that for all eight amplicons tested, amplifiable RNA yields were enhanced by the catalyst protocol relative to the commercial kit and the literature procedure. The amount of amplifiable RNA was increased by a factor of 7 over the commercial method in the least efficient case (Fig. S18) and by 25-fold in the most efficient case, with an average enhancement of 13-fold. Addition of the catalyst to the optimized incubation conditions enhanced RNA detection by 1.5–2.4 fold. The enhancements in amplicon quantity with catalyst **3** under optimized conditions relative to the commercial kit and the control without catalyst were statistically significant in all cases ( $P < 0.05$ , 1-tailed paired samples t-test) (Fig. 6). While all amplicons benefitted from the catalytic method, longer RNA amplicons showed the greatest increase in quantities retrieved: the three longest (180–514 bp) showed a mean enhancement of 18-fold over the kit.

### Mechanistic considerations

Taken together, our data show that organic water-soluble catalysts containing arylamines and proton donors are effective at speeding the reversal of formaldehyde-derived hemiaminal and aminor adducts from RNA and DNA bases, both for simple model systems and for cellular RNAs. Although many previous studies have attempted to enhance biomolecule recovery from formalin by varying buffers and heating protocols<sup>6</sup>, we are aware of no prior studies that make use of catalysts directed at the chemical mechanism of hemiaminal/aminor breakdown. The present studies suggest that Tris buffer – perhaps the most common buffer used in RNA/DNA recovery from FFPE specimens<sup>11,21</sup> – is not an ideal choice, mechanistically speaking, because its  $pK_a$  (8.1) is too high to be effective as a general acid proton donor at lower pH values. Indeed, while the hemiaminal adduct is eventually reversed in Tris buffer, we find that the aminor crosslink is much more stable, and we observed very little reversal in Tris buffer alone. We find that bifunctional catalysts

containing both amines and proton-donating groups are considerably more active than simple buffers alone.

Our early mechanistic studies reveal that the hemiaminal reversal is not highly pH sensitive between pH 4.5–8, and suggest that the reaction is catalyzed both by general acid and nucleophilic catalysis. Bifunctional catalysts such as **2** and **3** show substantial ability to speed the reaction over phosphate or Tris buffers alone, and the greater activity of catalyst **3** over **2** may reflect a better matching of its pK<sub>a</sub> (7.3) to solution pH. On the other hand, our results show that the aminal crosslink is indeed pH sensitive, and is catalyzed by lowered pH, suggesting a specific acid catalysis mechanism. Most catalysts and buffers tested here appear to have little effect beyond a simple protonation effect; however, very low pH values are to be avoided with nucleic acids because of their likely effect in promoting depurination<sup>29</sup>. We chose pH 4.5 here as a compromise between keeping RNA stable and promoting reaction as much as possible. At this pH we find that phosphanilate compound **3** has an additional catalytic effect over buffer alone, and our studies suggest that general acid catalysis is operative. We hypothesize that this compound is unique among those tested in having a pK<sub>a</sub> near that of the solution pH, being diprotic, and containing an *ortho*-amine group, all of which may enhance complex formation with the aminal at the transition state (see hypothesized transition state in Fig. S12). Additional mechanistic studies and comparison to new catalyst analogues will be useful in testing this hypothesis in the future.

Our experiments with full RNA strands treated with formaldehyde are in accordance with our nucleotide model studies, and show that catalyst **3** is the most active of all tested in reversing hemiaminal adducts and crosslinks. We find that compound **3** allows essentially complete removal of adducts in a shorter time and at lower temperatures than Tris buffer alone, and can do so without extended heating, which we and others<sup>30,31</sup> have shown causes substantial RNA degradation. When applied to formalin-fixed cellular RNAs, catalyst **3** fosters repair and recovery at mild temperatures, and enhances amplifiable RNA yields by more than an order of magnitude under optimized conditions compared to a commercial kit.

The current studies establish the utility of organocatalysts such as **2** and **3** in recovery of intact nucleic acids from formalin-fixed tumor samples. There are hundreds of millions of banked FFPE samples in the US alone<sup>32</sup>, with over a million new samples generated each year from tumor biopsies and cancer surgery. With an increasing clinical focus on molecular analysis of cancers in diagnosis and treatment, the development of strategies for enhanced recovery of biomolecular information from stored tissues is an important goal.

## Methods

### Materials and reagents

Adenosine monophosphate (AMP), deoxyadenosine monophosphate (dAMP), lithium perchlorate, and all organic catalysts (except catalyst **3**) were purchased from Sigma-Aldrich Co. Methanol-free 10% formaldehyde, EM grade was purchased from Polysciences, Inc. Solvents and reagents were purchased from Fisher Scientific, Aldrich or ACROS unless otherwise noted. 1.0 M Tris-HCl (pH 7.0) buffer was purchased from Invitrogen.

### Synthesis of N<sup>6</sup>-hydroxymethyl-dAMP

Synthesis of the hemiaminal monoadduct was performed following a literature procedure<sup>21</sup>. The compound was purified by reverse phase HPLC using a TEAA buffer/acetonitrile gradient; see Supporting Information file for details. The product was characterized by high-resolution ESI-mass spectrometry as C<sub>11</sub>H<sub>17</sub>N<sub>5</sub>O<sub>7</sub>P (M+H).

### Synthesis of Methylene-bis-adenosine-5'-monophosphate (dimer)

0.06 M solution of dAMP (in deionized water) and 0.3 M solution of 10% formaldehyde (methanol free) in 0.2 M sodium acetate buffer (pH 4.8) were mixed in equal volume. The mixture was stirred for 2–3 days at room temperature. The reaction mixture was then briefly stored by freezing at –20 °C. After thawing, the crude product was precipitated from ice-cold 2% LiClO<sub>4</sub> in acetone and purified by reverse phase HPLC using a TEAA buffer/acetonitrile gradient. The product was characterized by HR-MALDI-MS as C<sub>21</sub>H<sub>29</sub>N<sub>10</sub>O<sub>14</sub>P<sub>2</sub> and further confirmed by <sup>1</sup>H NMR. See the supporting data file for details.

### Catalyst 3

Phosphonate catalyst **3** was synthesized via a new procedure in three steps from *p*-iodoaniline. All spectral properties matched those of the same compound synthesized by our previous route<sup>20</sup>. See Supporting Data file for detailed methods.

### Reverse crosslinking of monoadduct and dimer monitored by HPLC

Reactions were carried out on 1 mL scale. 100 μL of reaction mixture was collected at one-hour intervals and injected into the HPLC (same conditions as for the purification of monoadduct and dimer). The progress of the reaction (reverse cross linking) was monitored at 260 nm. See Supporting Data file for details.

### Oligoribonucleotide synthesis

An RNA oligomer containing a self-complementary region (5'-AAAAACGCGCGAAAAA-3', 5181.31 Da) was designed and synthesized using standard β-cyanoethyl phosphoramidite chemistry and 2'-*O*-TBDMS-protected ribonucleosides. Phosphoramidites were purchased from Glen Research. Deprotection and initial purification of the RNA were carried out using Glen-Pak RNA purification columns. The oligonucleotide was further purified using polyacrylamide gel electrophoresis. The RNA was analyzed by MALDI-MS. See Supporting Data file for details.

### RNA formaldehyde treatment

To 1 equivalent of RNA stock solution (650 μM or 325 μM) in a 200 μL microcentrifuge tube was added 2 equivalents of methanol-free 10% formaldehyde solution with 1 M sodium chloride; the preparation scale ranged from 3 to 60 μL. The RNA-formaldehyde mixture was incubated at room temperature for 24 hours, after which time the RNA was isolated by ethanol precipitation. The pellet was either redissolved immediately or stored at –80 °C until use (within 24 hours). See Supporting Data file for details.



## RNA analysis by MALDI-MS

After formaldehyde treatment and post-treatment as described above, 1  $\mu$ L of 50  $\mu$ M DNA standard (5'-TCGGATCGTGATAT-3', 4293.86 Da) was added to each reaction. The samples were then desalted using C18 ZipTips (EMD Millipore) and eluted directly onto a 100-well plate, on which they were cospotted with 3-HPA matrix containing ammonium citrate. MALDI-TOF mass spectrometry analysis was carried out on an ABI Voyager-DE RP mass spectrometer in linear negative ion mode. Spectra were recorded from 500 to 12000 Da. All experiments were repeated five times, with 100–200 shots per spectrum and at least two spectra taken from each spot. In each case, the spectrum with the highest signal-to-background ratio was used for subsequent analysis.

The program Data Explorer was used to extract data. 5-point Gaussian smoothing was carried out, along with automatic baseline correction, calibration relative to the internal DNA standard and peak detection with a 1% intensity cutoff. Peak data was imported to Microsoft Excel for analysis. The amount of intact RNA was measured by taking the ratio of the RNA peak height at 5181.3 Da to the reference DNA peak height at 4293.86 Da. The recovery of RNA from adducts was calculated by comparing the peak height of the intact RNA (5181.3 Da) to the sum of the peak heights of intact RNA and adducts: peaks between 5181.3 Da and 5676 Da with at least 10% intensity. Finally, degradation was quantified by measuring ion counts for the total non-degraded RNA (intact RNA+adducts) relative to the reference DNA. See Supporting Data file for additional details and data.

## RNA recovery from FFPE specimens

RNA was extracted from a FFPE Raji cell pellet using either the spin-column-based AllPrep<sup>®</sup> DNA/RNA FFPE kit (Qiagen), according to the manufacturer's protocol, or a phenol-chloroform-isoamyl alcohol (PCI) extraction procedure.<sup>26</sup> Catalyst **3** was incorporated into the AllPrep<sup>®</sup> protocol by addition of an aqueous solution of **3** at pH 7 to the RNA-containing lysate. The mixture was heated at the stated temperature and for the stated time, then isolation continued as described in the manufacturer's protocol. RNA was quantified using the Qubit<sup>®</sup> RNA HS Assay Kit (Life Technologies). 200 ng RNA was used to synthesize cDNA using the Invitrogen SuperScript<sup>®</sup> III First-Strand Synthesis System for RT-PCR (Life Technologies). qPCR was performed in 384-well plates using SYBR<sup>®</sup> Green dye. Full experimental details for preparation of the FFPE sample, RNA isolation, quantification, cDNA synthesis, and qPCR are provided in the Supporting Data file.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

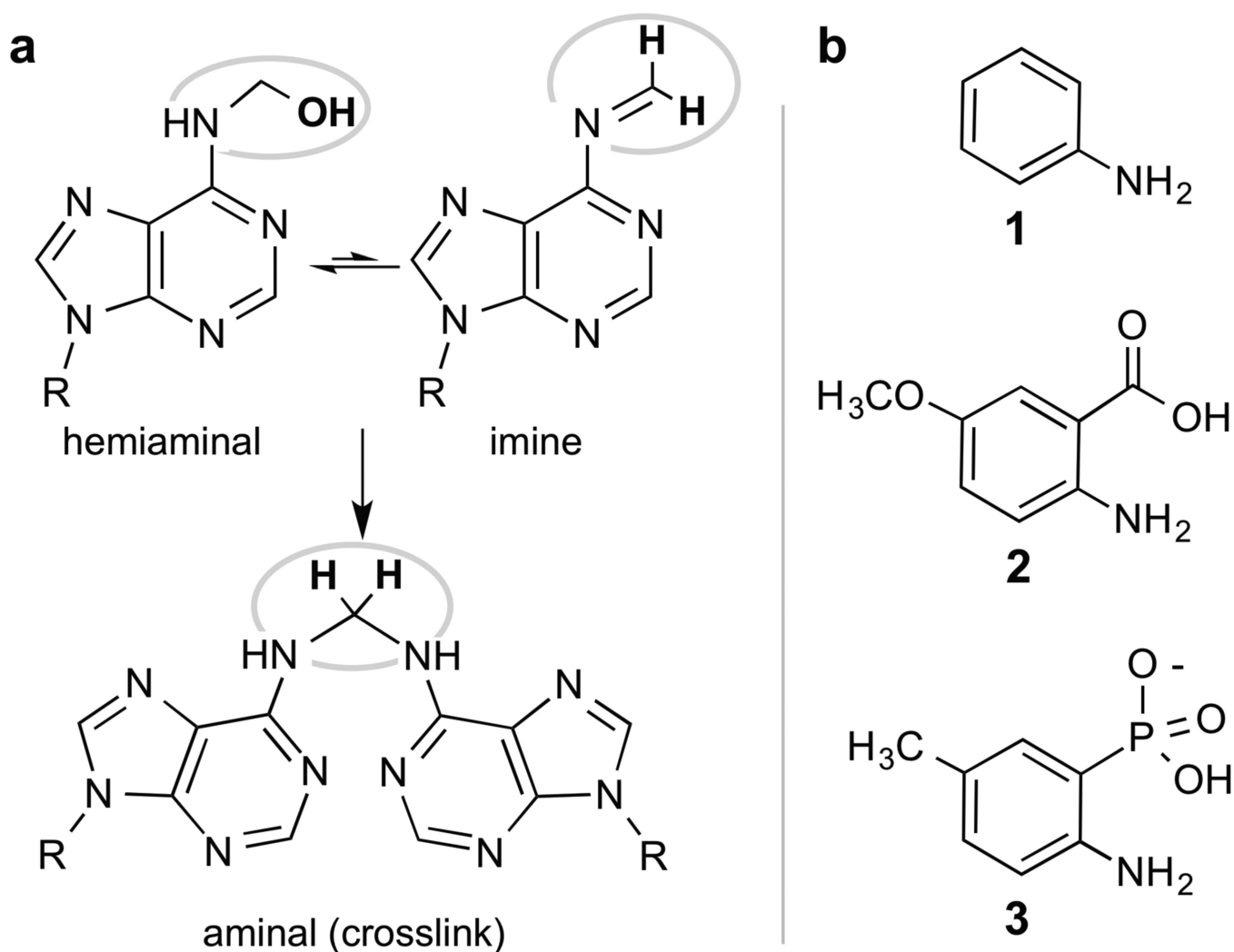
## Acknowledgements

We thank the U.S. National Institutes of Health (GM068122 and GM110050) to E.T.K., and Lymphoma Research Foundation (LRF#245480) to A.A.A., for support. E.H. acknowledges funding from an NSF graduate fellowship (DGE-1147470), and L.B. and C.R. acknowledge support from Swiss National Science Foundation. D.H. received funding from the Stanford ChEM-H Institute, and T.E. acknowledges support from the DAAD. We thank Dick Winant for assistance with MALDI analysis of RNAs, and Rashi Krishnan and Florian Scherer for assistance with the cell samples and RNA isolation.

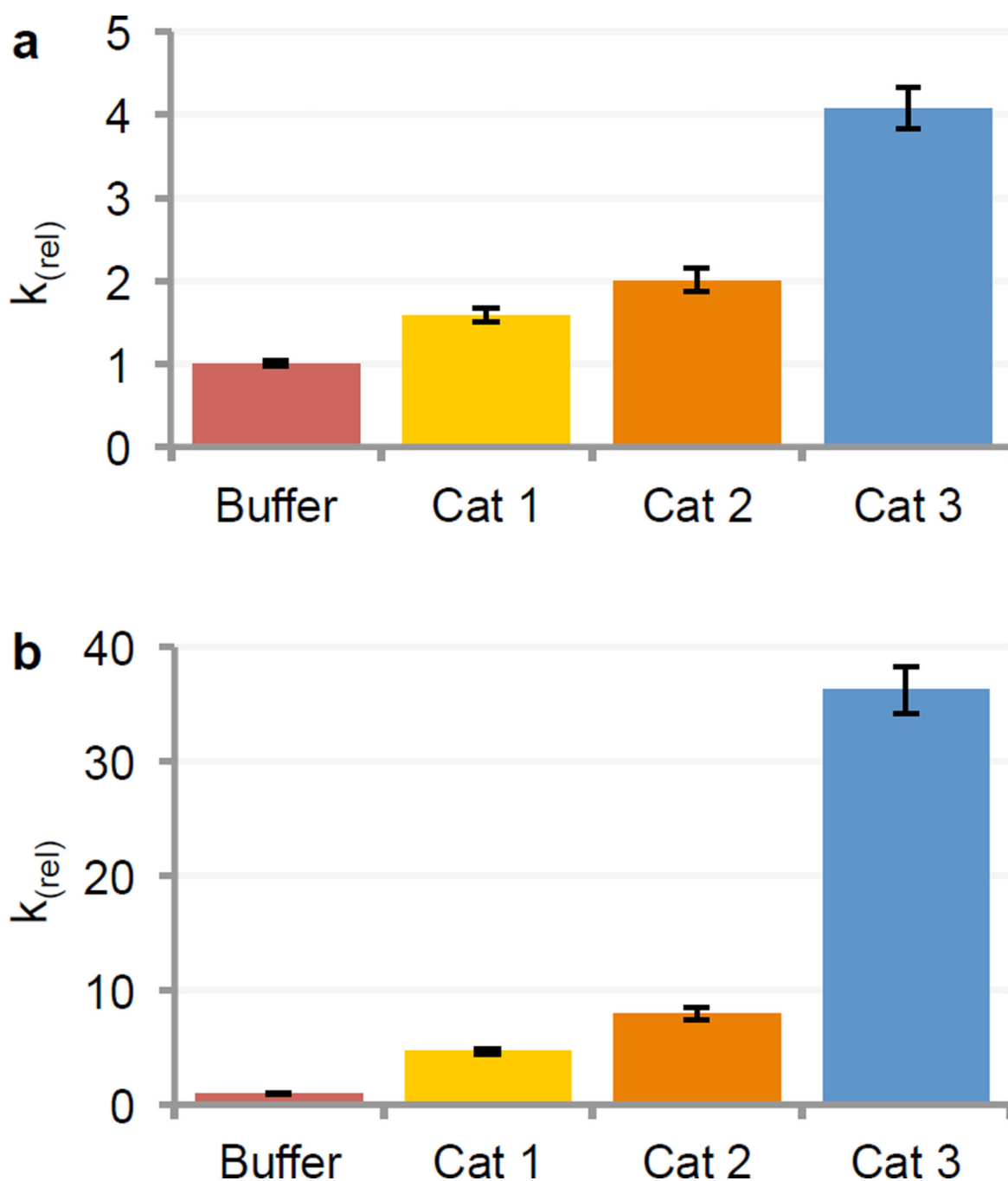
## References

1. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nature Rev. Cancer*. 2007; 7:545–553. [PubMed: 17585334]
2. Chin L, Gray JW. Translating insights from the cancer genome into clinical practice. *Nature*. 2014; 452:553–563. [PubMed: 18385729]
3. Fairley JA, Gilmour K, Walsh K. Making the most of pathological specimens: molecular diagnosis in formalin-fixed, paraffin embedded tissue. *Curr. Drug Targets*. 2012; 13:1475–1487. [PubMed: 22974391]
4. McGhee JD, von Hippel PH. Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases. *Biochemistry*. 1975; 14:1281–1296. [PubMed: 235285]
5. Krafft AE, Duncan BW, Bijwaard KE, Taubenberger JK, Lichy JH. Optimization of the isolation and amplification of RNA from formalin-fixed, paraffin-embedded tissue: The armed forces institute of pathology experience and literature review. *Mol. Diagn*. 1997; 2:217–230. [PubMed: 10462613]
6. Okello JB, Zurek J, Devault AM, Kuch M, Okwi AL, Sewankambo NK, Bimenya GS, Poinar D, Poinar HN. Comparison of methods in the recovery of nucleic acids from archival formalin-fixed paraffin-embedded autopsy tissues. *Anal. Biochem*. 2010; 400:110–117. [PubMed: 20079706]
7. Páska C, Bögi K, Szilák L, Tóth A, Szabó E, Sziller I, Rigó J, Sobel G, Szabó I, Kaposi-Novák P, Kiss A, Schaff Z. Effect of formalin, acetone, and RNAlater fixatives on tissue preservation and different size amplicons by real-time PCR from paraffin-embedded tissue. *Appl. Immunohistochem. Mol. Morphol*. 2004; 13:234–240.
8. Linton KM, Hey Y, Saunders E, Jeziorska M, Denton J, Wilson CL, Swindell R, Dibben S, Miller CJ, Pepper SD, Radford JA, Freemont AJ. Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours. *Br. J. Cancer*. 2008; 98:1403–1414. [PubMed: 18382428]
9. Shi SR, Shi Y, Taylor CR. Antigen retrieval immunohistochemistry review and future prospects in research and diagnosis over two decades. *J. Histochem. Cytochem*. 2011; 59:13–32. [PubMed: 21339172]
10. Chen J, Byrne GE Jr, Lossos IS. Optimization of RNA extraction from formalin-fixed, paraffin-embedded lymphoid tissues. *Diagn. Mol. Pathol*. 2007; 16:61–72. [PubMed: 17525674]
11. Gilbert MT, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, Jewell LD, Van Marck E, Worobey M. The isolation of nucleic acids from fixed, paraffin-embedded tissues—which methods are useful when? *PLoS One*. 2007; 2:e537. [PubMed: 17579711]
12. Lindahl T. Instability and decay of the primary structure of DNA. *Nature*. 1993; 362:709–715. [PubMed: 8469282]
13. Kubista M, Björkman J, Svec D, Sjöback R. RNA quality matters. *Eur Pharm. Rev*. 2012; 17:63–67.
14. Jencks WP. Mechanism and catalysis of simple carbonyl group reactions. *Prog. Phys. Org. Chem*. 1964; 2:63.
15. Jencks WP. Studies on the mechanism of oxime and semicarbazone formation. *J. Am. Chem. Soc*. 1959; 81:475–481.
16. Cordes EH, Jencks WP. Nucleophilic catalysis of semicarbazone formation by anilines. *J. Am. Chem. Soc*. 1962; 84:826–831.
17. Wendeler M, Grinberg L, Wang X, Dawson PE, Baca M. Enhanced catalysis of oxime-based bioconjugations by substituted anilines. *Bioconjugate Chem*. 2014; 25:93–101.
18. Rashidian M, Mahmoodi MM, Shah R, Dozier JK, Wagner CR, Distefano MD. A highly efficient catalyst for oxime ligation and hydrazone-oxime exchange suitable for bioconjugation. *Bioconjugate Chem*. 2013; 24:333–342.
19. Crisalli P, Kool ET. Water-soluble organocatalysts for hydrazone and oxime formation. *J. Org. Chem*. 2013; 78:1184–1189. [PubMed: 23289546]
20. Crisalli P, Kool ET. Importance of ortho proton donors in catalysis of hydrazone formation. *Org. Lett*. 2013; 15:1646–1649. [PubMed: 23477719]

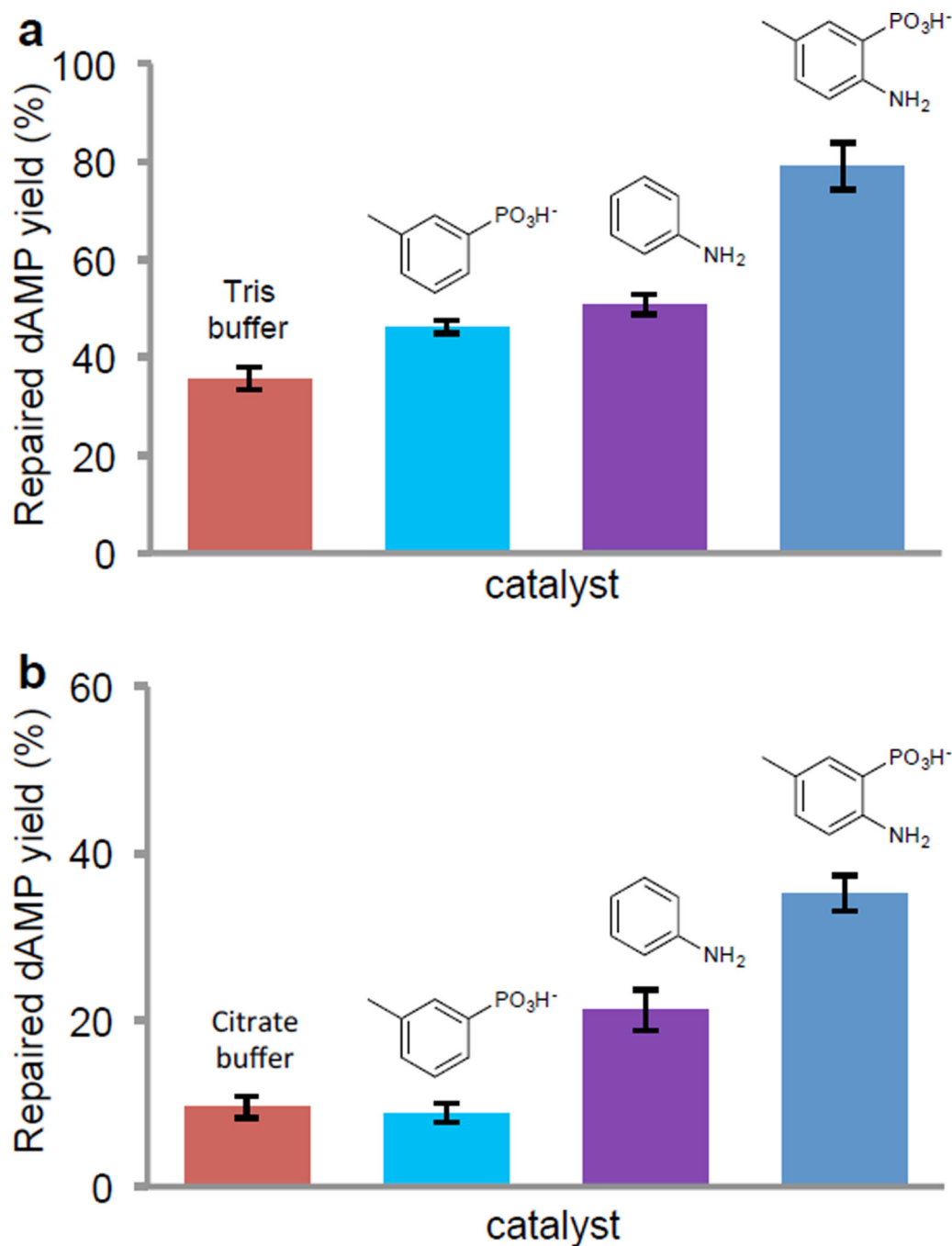
21. Rait VK, Zhang Q, Fabris D, Mason JT, O'Leary TJ. Conversions of formaldehyde-modified 2'-deoxyadenosine 5'-monophosphate in conditions modeling formalin-fixed tissue dehydration. *J. Histochem. Cytochem.* 2006; 54:301–310. [PubMed: 16116034]
22. Chaw YF, Crane LE, Lange P, Shapiro R. Isolation and identification of cross-links from formaldehyde-treated nucleic acids. *Biochemistry.* 1980; 19:5525–5531. [PubMed: 7459328]
23. Dirksen A, Dirksen S, Hackeng TM, Dawson PE. Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *J. Am. Chem. Soc.* 2006; 128:15602–15603. [PubMed: 17147365]
24. Rowland GB, Zhang H, Rowland EB, Chennamadhavuni S, Wang Y, Antilla JC. Brønsted acid-catalyzed imine amidation. *J. Am. Chem. Soc.* 2005; 127:15696–15697. [PubMed: 16277499]
25. Kerick M, Isau M, Timmermann B, Sültmann H, Herwig R, Krobitch S, Schaefer G, Verdorfer I, Bartsch G, Klockner H, Lehrach H, Schweiger MR. Targeted high-throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med. Genomics.* 2011; 4:68. [PubMed: 21958464]
26. Montgomery K, Zhao S, van de Rijn M, Natkunam Y. A novel method for making 'tissue' microarrays from small numbers of suspension cells. *Appl. Immunohistochem. Mol. Morphol.* 2005; 13:80–84. [PubMed: 15722798]
27. Karpova MB, Schoumans J, Ernberg I, Henter JI, Nordenskjöld M, Fadeel B. Raji revisited: cytogenetics of the original Burkitt's lymphoma cell line. *Leukemia.* 2005; 19:159–161. [PubMed: 15457187]
28. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res.* 1999; 27:4436–4443. [PubMed: 10536153]
29. Lindahl T, Nyberg B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry.* 1972; 11:3610–3618. [PubMed: 4626532]
30. Brisco MJ, Morley AA. Quantification of RNA integrity and its use for measurement of transcript number. *Nucleic Acids Res.* 2012; 40:e144. [PubMed: 22735698]
31. Gingrich J, Rubio T, Karlak C. Effect of RNA degradation on data quality in quantitative PCR and microarray experiments. *Bio-Rad. Bull.* 2006:5452A.
32. Asslaber M, Abuja PM, Stark K, Eder J, Gottweis H, Trauner M, Samonigg H, Mischinger HJ, Schippinger W, Berghold A, Zatloukal K. The genome Austria tissue bank (GATiB). *Pathobiology.* 2007; 74:251–258. [PubMed: 17709968]



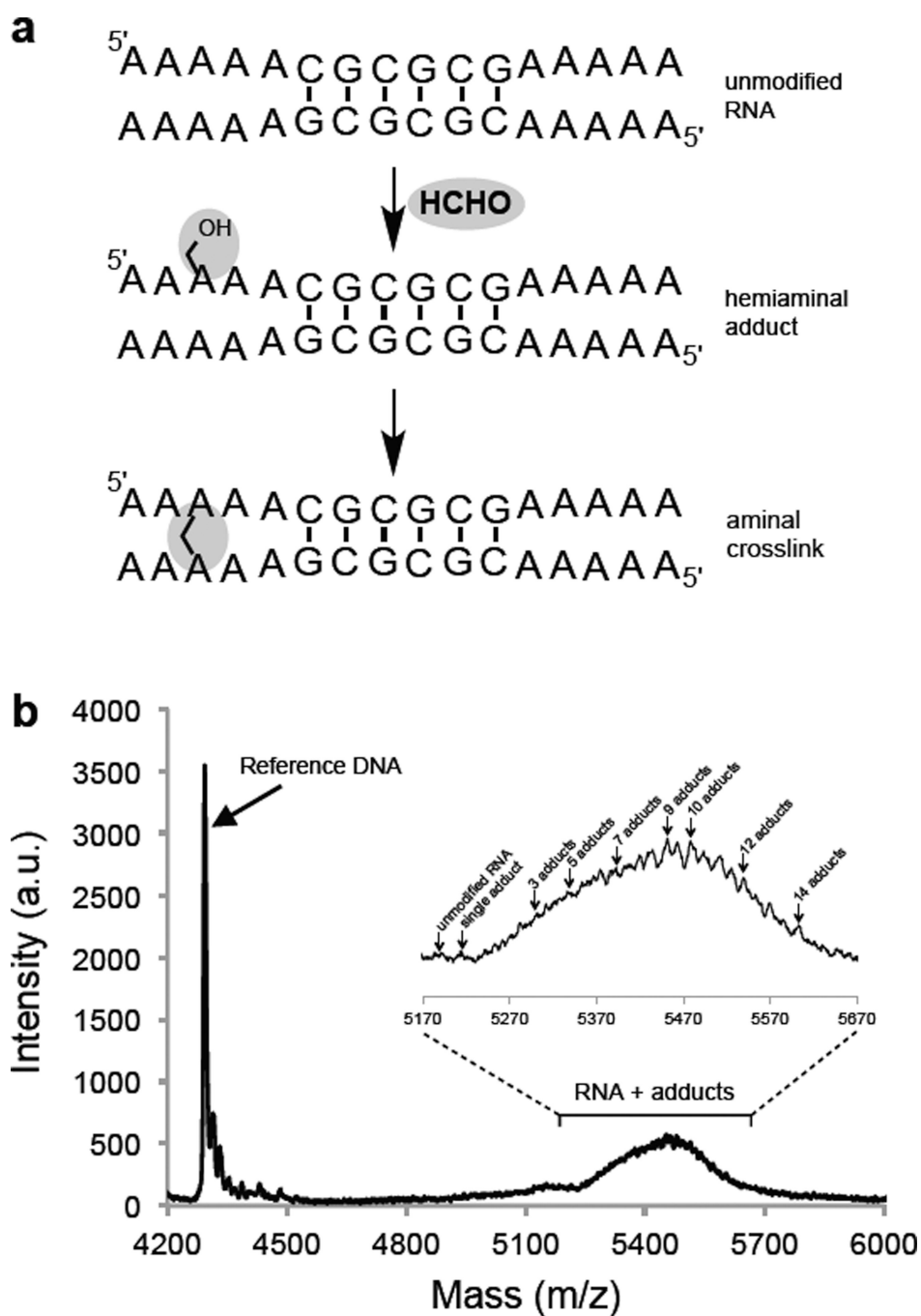
**Figure 1.** Formaldehyde adducts and catalysts in this study. **(a)** Adducts on N6 of adenine (R= ribose or deoxyribose); formaldehyde residues are shown in red. Similar adducts are formed on exocyclic amines of cytosine and guanine. **(b)** Transimination catalyst structures studied here.



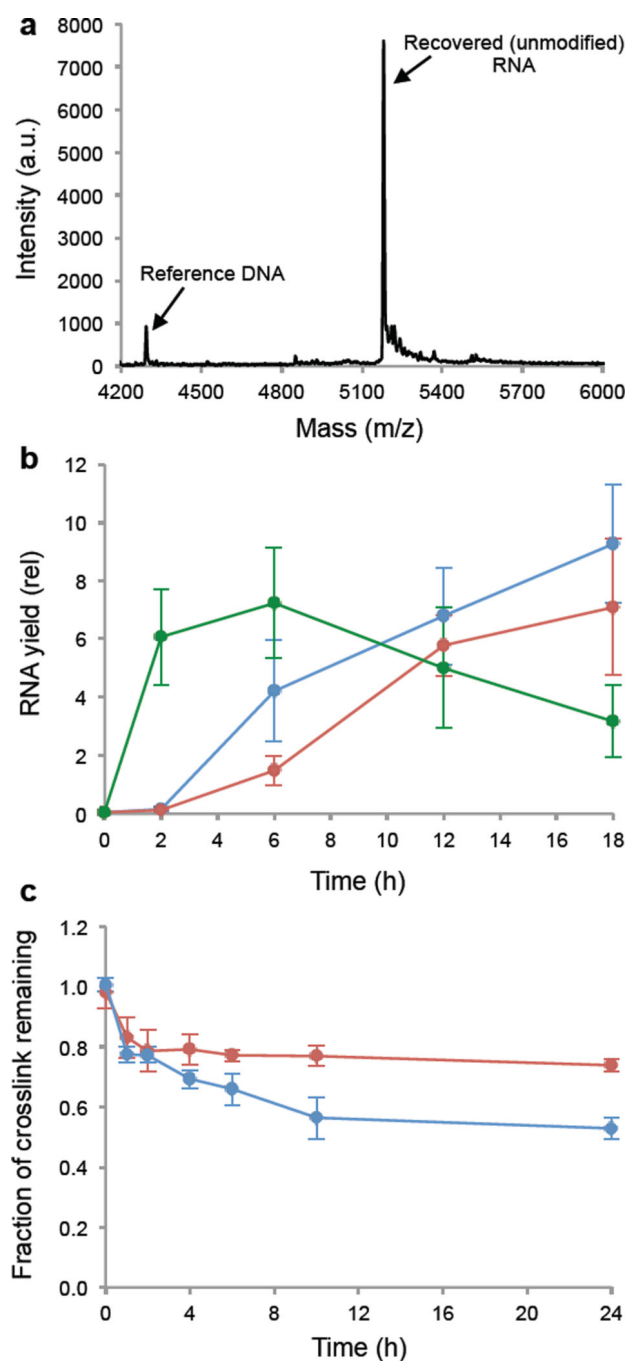
**Figure 2.** Relative rates of formaldehyde adduct reversal. (a) Relative rates of reversal of the hemiaminal adduct of dAMP in Tris buffer alone (pH 7) or with 10 mM added catalysts shown. Catalysts provide 1.6–4.1 fold rate enhancement over buffer, with the largest change in rate due to bifunctional catalyst **3**. (b) Rates of reversal of the aminal crosslink of AMP in phosphate buffer alone (pH 4.5) or with 10 mM added catalysts. Catalysts **1** and **2** show moderate rate enhancements, while catalyst **3** provides a superior 37-fold increase in rate. Data are from three replicates each (error bars are standard deviations).



**Figure 3.** Relative yields of adduct reversal from nucleotides after 1 h with catalyst **3** and with analogs having functional groups omitted (structures shown). (a) Reversal of the hemiaminal adduct of dAMP (pH 7, 16 mM catalyst, 37 °C); (b) Reversal of the aminor dimer of AMP (pH 4.5, 16 mM catalyst, 37 °C). In both cases the bifunctional catalyst **3** performs better than either of the closely-related monofunctional catalysts. Data are from three replicates each (error bars are standard deviations).



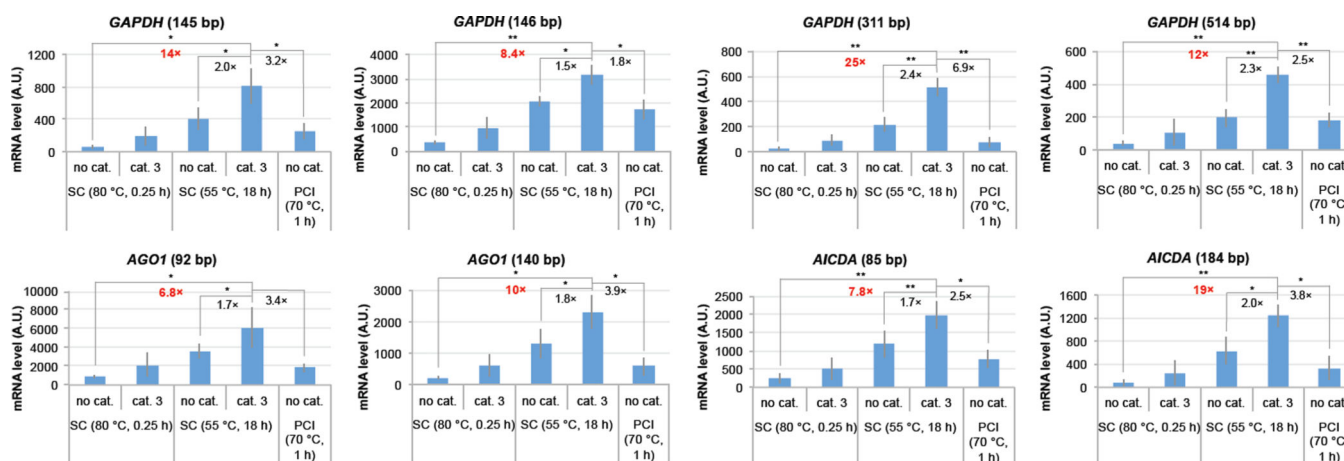
**Figure 4.** Assessing formaldehyde adducts on an RNA strand by mass spectrometry. **(a)** Sequence of the self-complementary 16mer RNA, which was designed to promote adducts and crosslinks on unpaired bases. **(b)** MALDI mass spectrum of formaldehyde-treated RNA, showing extensive adducts after 24 h treatment (up to 14 per strand, see inset) and little or no unmodified RNA (5181 Da) remaining. Unmodified DNA (4294 Da) is spiked in for reference.



**Figure 5.** Improvement in reversal of RNA formaldehyde adducts after low-temperature incubation in the presence of catalyst **3**. RNA pretreated with 10% formaldehyde (see Fig. 4) was used as starting material. **(a)** MALDI mass spectrum of 16mer RNA oligonucleotide after 18 h treatment with catalyst **3** (8 mM, pH 7, 37 °C), showing major recovered RNA peak. **(b)** Time course of RNA recovery at pH 7, comparing 60 °C heating in 8 mM Tris buffer (●) to 37 °C in 8 mM Tris buffer (●) and 37 °C in 8 mM catalyst **3** (no Tris) (●). Higher temperature provides a short-term increase in RNA recovery, but catalyst **3** gives an



advantage at the lower temperature where RNA is more stable. Error bars show standard deviation from 5 experiments. (c) Time course of crosslink reversal in dimerized RNA oligonucleotide, following uncrosslinking to monomer RNA by denaturing PAGE. Treatment at pH 4.5, 37 °C with 16 mM catalyst **3** (●) yields more uncrosslinking than incubation at pH 4.5, 37 °C in 16 mM citrate buffer (●), although full reversal of crosslinks is not observed with this substrate. Error bars show standard deviation from 3 or 4 experiments.

**Figure 6.**

Enhancement in recovery of RNAs from formalin-fixed, paraffin-embedded cell specimens using catalyst **3** (20 mM) as compared with different incubation and isolation conditions. Amplifiable RNA yield is plotted for eight amplicons, and quantity is determined with a standard curve. Lane 1 employs incubation conditions from a commercial kit (Qiagen AllPrep<sup>®</sup> DNA/RNA FFPE kit), which uses an 80 °C, 0.25 h incubation step without catalyst (“no cat”) and a spin column for isolation, with the addition of catalyst to these conditions shown in lane 2 (“cat. 3”). Optimized incubation conditions (55 °C, 18 h) followed by a spin column RNA isolation are shown in lanes 3–4. A common literature procedure is shown in lane 5 (“PCI”)<sup>28</sup>. Addition of the catalyst to the optimized incubation conditions results in a ~2 fold increase in detectable RNA, and more substantial increases relative to the catalyst-free commercial kit protocol (see enhancements in red) or the literature procedure. The means of three independent experiments are shown, error bars indicating the standard deviation of variation in the qRT-PCR yield. SC: spin column isolation. PCI: Masuda protocol of phenol-chloroform-isoamyl alcohol extraction followed by heating in buffer.<sup>28</sup> A.U.: arbitrary units. Significance for pairwise comparisons shown was tested using a 1-tailed paired samples t-test. \*: P < 0.05; \*\*: P < 0.01. Fold enhancements are shown below brackets; enhancement relative to the commercial kit is shown in red.