Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kevin Xiao April 10, 2022

Establishing the optimal conditions for in vivo RNA structural probing of guanine and uracil base-pairing in yeast

by

Kevin Xiao

Homa Ghalei
Adviser

Biology

Homa Ghalei

Adviser

Sohail Khoshnevis

Committee Member

Anita Corbett

Committee Member

2022

Establishing the optimal conditions for in vivo RNA structural probing of guanine and uracil base-pairing in yeast

By

Kevin Xiao

Homa Ghalei

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2022

Abstract

Establishing the optimal conditions for in vivo RNA structural probing of guanine and uracil base-pairing in yeast

By Kevin Xiao

RNAs can have catalytic or regulatory functions in the cell and play critical roles in many steps of gene regulation. RNA structure can play a key role in its function. Therefore, methods to investigate the structure of RNA in vivo are of great importance for understanding the role of cellular RNAs. RNA structural probing is an indirect method to probe the three-dimensional structure of RNA by analyzing the reactivity of different nucleotides to chemical modifications. The chemical modifications can target either the RNA backbone or the Watson-Crick face of nucleotides. The selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) can probe the ribose sugar in RNA nucleotides. Dimethyl sulfate (DMS) alkylates adenine and cytosine, but it is not reactive to guanine or uracil. Recently, new compounds were used to modify Gs and Us in the plant model system rice Oryza sativa. To complement the scope of RNA structural probing by chemical modifications in yeast, as part of my honors thesis dissertation work, I analyzed the effectiveness of guanine modification by a family of aldehyde derivatives, the glyoxal family, in Saccharomyces cerevisiae and Candida albicans in vivo. We also explored the effectiveness of uracil modification by carbodiimide derivatives in these species in vivo. We show that among the glyoxal family, phenylglyoxal (PGO) is the best guanine probe for in vivo structure probing as the guanine modification effectiveness demonstrates a concentration-dependence in S. cerevisiae, and C. albicans. We also demonstrate a concentration-dependent relationship for uracil modification by carbodiimide N-cyclohexyl-N-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate (CMCT) in S. cerevisiae and C. albicans in vivo. Further, we show that PGO treatment does not affect the processing of different RNA species in the cell and is not toxic for the cells under the conditions we have established for RNA structural probing. Our results provide the conditions for probing the reactivity of guanine and uracil in RNA structures in yeast and offer a useful tool for studying RNA structure and function in two widely used yeast model systems.

Establishing the optimal conditions for in vivo RNA structural probing of guanine and uracil base-pairing in yeast

By

Kevin Xiao

Homa Ghalei

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2022

Acknowledgements

I would like to thank Dr. Sohail Khoshnevis and Dr. Homa Ghalei for their guidance and support on my experiments, thesis development and editing. I am also thankful to all members of the Ghalei and Khoshnevis Lab at Emory University School of Medicine. I would like to further thank Dr. Anita Corbett for accepting to be a member of my thesis committee.

Table of Contents

1. Background and Introduction…….....1

1.1 Introduction of RNAs and their three-dimensional structures……..……1

1.2 The importance of understanding RNA three-dimensional structures…….…….2

1.3 An overview of RNA structural probing techniques……..3

1.4 Established in vivo RNA structural probing reagents for modification of guanines and uracils……..…….....3

1.5 Research plan……..……4

2. Materials and Methods…….…….4

2.1 Yeast Cell Culture……..4

2.2 Incubation with Chemical Probes……..5

2.3 Total RNA Extraction and Purification……...……..……5

2.4 Reverse Transcription……6

2.5 Northern blot for snoRNA and tRNA…….…7

2.6 Serial Dilution Spot Test for Chemical Toxicity Assessment……...…..7

3. Results……...…….…7

3.1 In vivo modification of yeast S. cerevisiae 5.8S rRNA by glyoxal, methylglyoxal, and phenylglyoxal characterized by denaturing UREA-PAGE of cDNAs……...……...…8

3.2 In vivo modification of yeast S. cerevisiae 18S rRNA by glyoxal and derivatives characterized by denaturing UREA-PAGE of cDNAs……..9

3.3 In vivo modification of yeast C. albicans 5.8S rRNA by phenylglyoxal by denaturing UREA-PAGE of cDNAs…….…..…10

3.4 Northern blot analysis for U3-3', U3-5' snoRNAs and tRNAs……...…….11

3.5 Serial dilution spot test for toxicity analysis in PGO-treated S. cerevisiae….12

3.6 In vivo modification of yeast S. cerevisiae 5.8S rRNA characterized by CMCT by denaturing UREA-PAGE of cDNAs……...…….13

4. Discussion…….……..…..14

5. **Acknowledgement**……..…..……..…..16

6. **Contributions**……..……..…17

7. **Reference**……...……..…….17

Table of Figures

Figure 1. In vivo modification of 5.8S rRNA by GO derivatives in *S. cerevisiae*.....…8

Figure 2. In vivo modification of 18S rRNA by GO derivatives in *S. cerevisiae* …….....9

Figure 3. In vivo modification of 5.8S rRNA by PGO in *C. albicans* …….……..……..10

Figure 4. Northern blot analysis of the processing of U3 snoRNA and proline tRNA.....…11

Figure 5. Serial dilution growth assay for toxicity analysis of PGO-treated *S. cerevisiae*……...12

Figure 6. In vivo modification of 5.8S rRNA by CMCT in *S. cerevisiae* ……...…13

Background and Introduction

RNAs are a class of nucleic acids that serve important cellular functions. In eukaryotes, protein-coding messenger RNAs (mRNAs) comprise only 1-2% of the total cellular RNAs (Wu et al, 2014). In contrast, most transcribed eukaryotic RNAs are non-coding and can play key catalytic or regulatory roles in gene expression. These non-coding RNAs make up ~98% of the total cellular RNAs (Wu et al, 2014). As RNAs are single-stranded, they can adopt diverse secondary and tertiary structures in physiological conditions (Bevilacqua and Assmann, 2021). An example of simple RNA three-dimensional structures includes base-paired doubled-stranded areas such as hairpin stems. More complex RNA three-dimensional structures comprise of tertiary and quaternary structures including ribose zippers, kink turns, and pseudoknots (Bevilacqua and Assmann, 2021). The versatility of RNAs in forming simple and intricate three-dimensional structures allows RNAs to perform important cellular functions including catalysis, and promote RNA-protein interactions. As such RNAs are key players in gene expression regulation. For instance, in hairpin ribozymes, RNAs hydrogen bond with each other following the Watson-Crick base-pairing at the stem, whereas the canonical base-pairing is relaxed at the hairpin loop. This structure allows the RNA to perform catalysis with the help of metal cofactors (Scott, W., 1998). Non-coding RNA can also regulate pre-mRNA splicing as part of the posttranscriptional regulation. For instance, the HuR RNA-binding proteins bind to specific nucleotide sequence in pre-mRNA by recognizing specific RNA secondary structures (Buratti and Baralle, 2004). Another instance of an RNA-protein interaction that regulates gene expression is found in the lagging strand of the eukaryotic chromosome. RNA can assume a quadruplex 3D structure, one of the most important of which is found in the 3’ overhang of the telomere. G-quadruplex interacts with telomerase to inhibit DNA replication, stalling tumor cell growths (Want et al, 2011).

Understanding the 3D structures of RNAs is essential for revealing RNA functions in gene expression. However, the diverse range of RNA functions happen in the active forms of RNA in vivo. Researchers, therefore, face the challenge of probing a vast array of RNA structures in vivo to elucidate the functional roles of RNAs in cellular processes (Bevilacqua and Assmann, 2021). RNA structure-function relationships inside the cell are affected by the rate of transcription, the local solution environment, and the presence of small molecules or RNA-protein interactions. The physical state of RNA in vivo can therefore provide insight into the function of the RNA (Spitale et al, 2012).

Structural biology approaches including X-ray crystallography, NMR spectroscopy and single-particle cryo-electron microscopy have shaped our understanding of the three-dimensional structure of many RNAs. While providing us with the atomic resolution structure of the RNA molecules, these techniques have several drawbacks, including the need for highly pure samples and the time it takes to solve the structures. Furthermore, these techniques to study the RNA molecules in vitro may not provide a comprehensive view of the conformations that the RNA molecules adopt in their native environment (Parisien and Major, 2012). Hence, indirect techniques have been developed as an effective alternative to structural approaches to study the RNA structure.

RNA structure probing is the fastest way to indirectly probe RNA structures by using chemical reagents to modify nucleotides at specific positions and analyze the modification frequency of each site. RNA structure probing techniques can target ribose sugars or nitrogen bases. Selective 2'- hydroxyl acylation analyzed by primer extension (SHAPE) uses a carbonyl derivative that is electrophilic enough to be attacked by the strong nucleophile 2'-hydroxyl group on the ribose sugar (Busan et al, 2019). Once the backbone is modified, truncation during the reverse

transcription (RT) is used to identify the modification adduct locations, which are quantified by denaturing UREA polyacrylamide gel electrophoresis (UREA-PAGE) (Busan et al, 2019) or deep-sequencing. Base-specific modification is an alternative method for indirect RNA structural probing. Dimethyl sulfate (DMS) is used as base-specific probing reagent against the N1 of adenines and N3 of cytosines that are not involved in base pairing or hydrogen bonding (Ziehler and Englke, 2013). DMS is cell-permeable because of its small size (Wells et al, 2000). DMS modification reactions therefore take place readily in almost all in vivo conditions without the need for additional permeabilization (Wells et al, 2000). However, DMS is unable to probe uracils or guanines. The lack of reliable chemical probes for guanines and uracils limits the potential of RNA structure probing approaches to assess RNA conformations. Thus, new families of chemicals have been explored to probe guanine and uracil in vivo.

In the glyoxal family, glyoxal (GO), methylglyoxal (MGO), and phenylglyoxal (PGO) are potential candidates for guanine in vivo probing. These chemical probes are carbonyl derivates that are electrophilic towards the nucleophilic amidine group in adenine, cytosine, and guanine. As uracil lacks this functional group, it does not react with glyoxal family compounds. However, because guanines have the amidine group farthest away from the ribose sugar backbone, there is less steric hindrance compared to that in adenines and cytosines, so guanine is most reactive with glyoxal family compounds (Mitchell et al, 2017). Mitchell III et al demonstrated that that glyoxal family compounds are effective guanine probing agents in the eukaryotic plant model rice *Oryza sativa* and in gram-negative bacteria in vivo (Mitchell III et al, 2017). Whether glyoxal and its derivatives can be used for in vivo RNA structure probing in other model systems is unclear.

Uracils are probed by carbodiimides family reagents, including 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and carbodiimide N-cyclohexyl-N-(2-

morpholinoethyl) carbodiimide metho-p-toluenesulfonate (CMCT). EDC has been validated for uracil probing in vivo for its rapid cell wall penetration ability (Mitchell III et al, 2019 and Wang et al.). However, EDC reaction in vivo is slower than CMCT due to the presence of hyperconjugation, weakening the electrophilicity of the compound. Therefore, CMCT is supposed to be a much faster and thus better reagent than EDC for uracils probing due to the strengthening of electrophilicity from the inductive effect rendered by toluenesulfonate functional group.

Saccharomyces cerevisiae is a simple eukaryotic organism whose RNA biology shares many features with that of higher eukaryotes, making it a suitable model organism to study different aspects of gene expression. While there are established procedures for in vivo probing of the adenine and cytosine positions within yeast RNAs using DMS (Tijerina et al, 2007), guanines and uracils have so far escaped in vivo probing in yeast cells. In this work, we test the application of glyoxal and its derivatives for RNA structure probing in the widely used budding yeast model system *Saccharomyces cerevisiae* and in the human fungal pathogen model system, *Candida albicans*. We compare three different glyoxal derivatives and show that PGO yields the highest modification rate of guanines in yeast without affecting cell viability or RNA processing. We also present the conditions for in vivo modification of uracils in yeast using CMCT.

Materials and Methods

Yeast Cell Culture

BY4741 strain *Saccharomyces cerevisiae* and BWP17 strain of *Candida albicans* were grown in YPD media to optical density of 0.5-0.6 prior to the application of the desired chemical probes.

Incubation with Chemical Probes

S. cerevisiae and *C. albicans* were incubated with glyoxal family compounds including GO, MGO, or PGO, or with CMCT. For each compound, three different concentrations of the compound (plus a no-compound negative control) and two incubation times were tested in biological duplicates. GO concentrations were 30 mM, 60 mM, and 120 mM. MGO and PGO concentrations were 5 mM, 10 mM, and 20 mM. CMCT concentrations were 50 mM, 100 mM, and 200 mM. The incubation times were 5 and 15 minutes. Upon the completion of incubation, samples were cooled down immediately on ice and cells were harvested. Cells were washed three times with ice-cold water prior to RNA extraction.

Total RNA Extraction and Purification

Cell pellets from 10 mL of culture grown to mid-log phase were resuspended in 400 µL of TES buffer (10 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 0.5% SDS), followed by the addition of 400 µL of acid phenol (Sigma). Samples were vortexed for 10 seconds, before incubating at 65℃ for 45 minutes, with consistent mixing every 5 to 10 minutes. Samples were then placed on ice for 5 minutes before centrifuging at maximum speed for 5 minutes at room temperature. Immediately following the completion of the centrifugation, the aqueous layer on the top was transferred to a new tube. 400 µL of acid phenol was then added to the aqueous phase and mixed. Samples were incubated on ice for 5 minutes and spun down at maximum speed for 5 minutes at room temperature. The aqueous layer was transferred to a new tube, and 400 µL of chloroform was added and mixed. Samples were centrifuged at maximum speed for 5 minutes at room temperature. The aqueous layer was transferred to a new tube and the RNA was precipitated by the addition of

750 µL of cold 100% ethanol and 25 µL of 3M sodium acetate followed by incubation at -20℃ overnight.

The precipitated RNA was pelleted by centrifugation at maximum speed for 20 minutes in the cold room. Supernatant was aspirated and the RNA pellet was washed by addition of 0.5 mL of 80% cold ethanol. Samples were centrifuged at maximum speed for 20 minutes in the cold room, and the pellet was dried at 42℃ for 20 minutes and resuspended in 45 µL of water, before addition of 5 µL of 10X DNase I buffer and 1 µL of DNase I (NEB). DNase treatment was performed for 15-minute at 37 ℃, followed by addition of 1 µL of 0.5M EDTA and 15-minute incubation. The RNA was further purified using a Quick-RNA miniprep kit (Zymo Research) according to the manufacturer's protocol. Final purified RNAs were eluted in 50 µL of DNase/RNase-Free water.

Reverse Transcription (RT)

Approximately 1 µg of purified RNA in 10 µL was mixed with 2 µL of 0.6 µM primers labelled with γ ^{32}P-ATP. Annealing was performed by incubation at 65℃ for 5 minutes followed by a gradual cool down at room temperature over 10 minutes after which the tube was placed on ice. The reverse transcription was performed using SuperScript III (ThermoFisher) at 50℃ for 5 minutes per manufacturer's manual. 1 µL of 4M NaOH was added before heating the RNA at 95℃ for 5 minutes to remove the RNA templates. The cDNA products were mixed with 19.5 µL of formamide stop dye and heated at 95℃ for 5 minutes. The cDNA products were separated on a prewarmed 6% urea/acrylamide sequencing gel at 2200 V for 1.5 h in 0.5X TBE buffer. The gel was dried and exposed to a phosphoscreen.

Northern Blot for snoRNA and tRNA

Total RNA from two biological replicates treated with PGO or CMCT was isolated using the hot phenol method. snoRNAs were separated on 8% acrylamide/urea gels, transferred to Hybond nylon membrane (GE Healthcare), and probed as indicated.

Serial Dilution Spot Test for Chemical Toxicity Assessment

S. cerevisiae was incubated with PGO or CMCT as described above. The density of the culture was adjusted to a final concentration of 10^7 cells/mL, followed by four successive cascade dilutions in a 1:10 ratio. Dilutions were spotted onto YPD plates and grown at 25℃, 30℃, and 37℃ for 48 hours.

Results

To define the best glyoxal derivative to probe guanines in yeast, we tested different concentrations of glyoxal (GO), methylglyoxal (MGO), and phenylglyoxal (PGO) and as well as two incubation times. The tested concentrations of GO and MGO were chosen based on their effect on yeast cell growth (Hoon et al., 2011) as a proxy for their cell penetration. In that study, 60 mM GO and 10 mM MGO resulted in ~ 50% growth rate reduction in *S. cerevisiae* cells. Therefore, we tested those conditions as well as 0.5X and 2X of each of these (30, 60, and 120 mM GO, and 5, 10, and 20 mM MGO). PGO concentrations were chosen based on the effect on yeast mitochondrial ATP synthase (Mueller 1988 and Guo et al., 2018) where 10 mM PGO greatly destabilized the F_1-ATPase in *S. cerevisiae*. We therefore tested 5, 10 and 20 mM PGO.

The 5.8S rRNA is a part of the large ribosomal subunit. Several positions on 5.8S are subject to glyoxal modification in rice 5.8S rRNA (Mitchell et al., 2017). We therefore studied the

modification of the 5.8 rRNA of *S. cerevisiae* by glyoxal (GO), methylglyoxal (MGO), and phenylglyoxal (PGO) (Fig. 1).

GO weakly modifies guanines in the 5.8S rRNA of *S. cerevisiae* in vivo, as evident from the weak modification of nucleotides at positions G78 and G85 (Fig. 1A). The 0 mM concentration of GO serves as the negative control to ensure that modifications displayed in experimental groups (30 mM, 60 mM, and 120 mM) are the results of GO treatment.

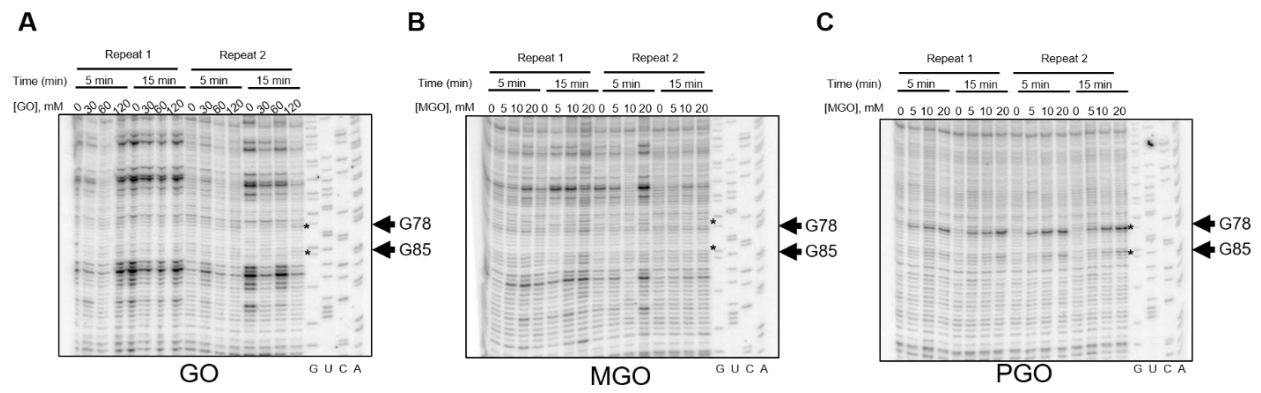

FIG 1. In vivo modification of yeast *S. cerevisiae* 5.8S rRNA by glyoxal and derivatives characterized by denaturing UREA-PAGE of cDNAs after reverse transcription. RNA structure probing with A) glyoxal B) methylglyoxal, and C) phenylglyoxal. Reaction conditions are specified at the top of each panel. The data show two biological replicates. Arrows and the asterisks indicate RNA nucleotide positions modified by glyoxal and glyoxal derivatives.

At the 15 min time point, the band intensity for G78 and G85 increases as the GO concentration increases (Fig.1A). At the 5 min time point, GO modifications do not show an appreciable concentration-dependent relationship, as band intensity across all GO concentrations appears similar (Fig. 1A). The reactivity of MGO towards guanines in the 5.8S rRNA of *S. cerevisiae* in vivo is weaker than that of GO (Fig.1B). G78 and G85 modifications are comparable

across all MGO concentrations, as band intensity remains the same at each time point and at each MGO concentration (Fig. 1B). PGO demonstrates the highest reactivity towards guanine in the 5.8S rRNA of *S. cerevisiae* in vivo (Fig. 1C). At G78 and G85, modifications by PGO appear to be concentration-dependent, as band intensity at both positions increases as the PGO concentration rises (Fig. 1C).

G913 in *S. cerevisiae* 18S rRNA changes conformation at different stages of the translation cycle (Sulima et al., 2013). We therefore tested the in vivo reactivity of this nucleotide towards different derivatives of glyoxal (Fig. 2). GO weakly modifies G913 in the 18S rRNA because bands across all GO concentrations at two timepoints (5 and 15 min) are nearly clear, indicating weak to no modifications by GO at G913 (Fig. 2A). MGO shows a slightly better result for G815

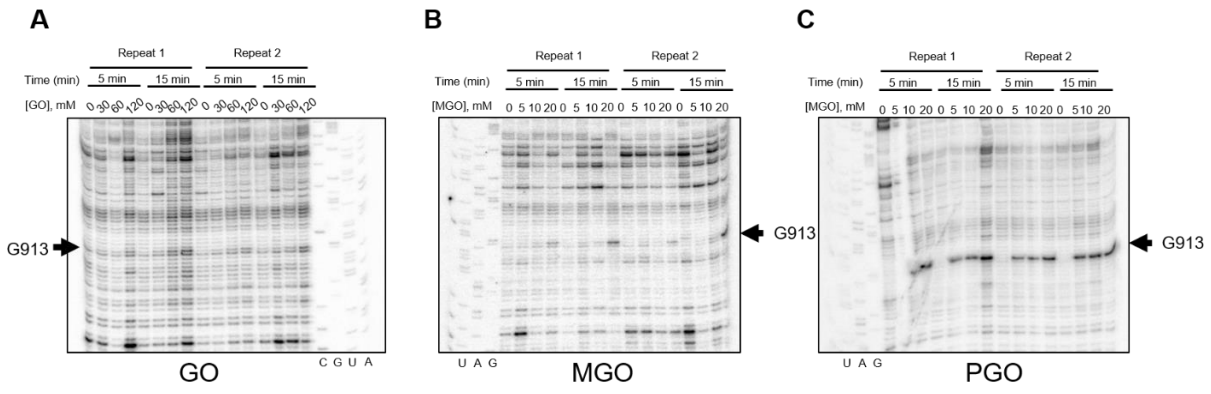

FIG 2. In vivo modification of yeast *S. cerevisiae* 18S rRNA by glyoxal and derivatives characterized by denaturing UREA-PAGE of cDNAs after reverse transcription. Figs from left to right represent glyoxal reactions, methylglyoxal reactions, and phenylglyoxal reactions, respectively. Reaction conditions at two timepoints are specified at the top of the Fig, and concentrations of glyoxal and its derivatives used are listed below reaction time conditions. One repeat is done. Arrows represent modified RNA nucleotide positions by glyoxal and its derivatives.

and G913 modifications. A weak concentration-dependent relationship is found at G913 as band intensity increases when MGO concentration is increased (Fig. 2B). Interestingly, there is a sharp increase in the band intensity when MGO concentration increases from 10 mM to 20 mM (Fig. 2B). PGO demonstrates the best guanine modification among the three glyoxal derivatives. As the PGO concentration increases, the band intensity increases, resulting in higher reactivity and better modifications at G913 (Fig. 2).

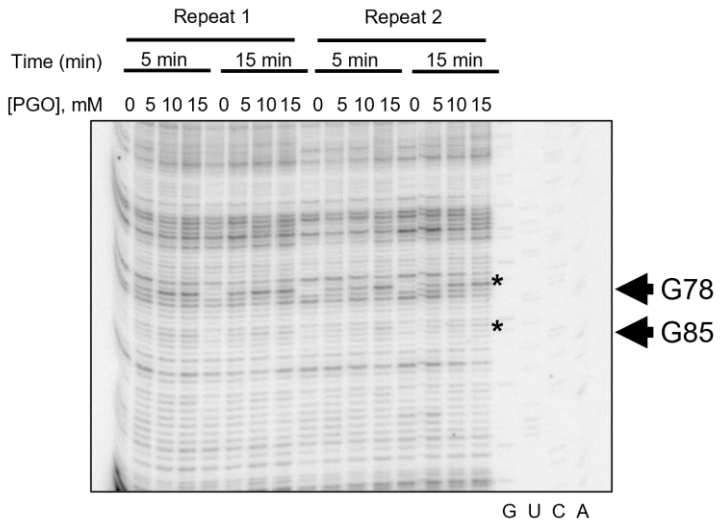

FIG 3. In vivo modification of yeast *C. albicans* 5.8S rRNA by phenylglyoxal by denaturing UREA-PAGE of cDNAs after reverse transcription. Reaction conditions at two timepoints are specified at the top of the Fig, and concentrations of glyoxal and its derivatives used are listed below reaction time conditions. One repeat is done. Arrows represent modified RNA nucleotide positions by PGO.

Next, we sought to determine whether the conditions for modifying guanines in *S. cerevisiae* are applicable to other fungi. To this end, we tested the modification of 5.8S rRNA of *Candida albicans*, a human fungal pathogen. Post-transcriptional regulation of gene expression is important for the pathogenicity of *C. albicans* (Verma-Gaur and Tavran, 2016). Therefore, establishing the effective condition for guanine probing in *C. albicans* is of great importance for future pharmaceutical and biochemical applications. Having already established that PGO is the best guanine probe in *S. cerevisiae,* we used PGO to probe the 5.8S rRNA in *C. albicans*. PGO shows

effective modification reactivity towards G78 and G85 in *C. albicans* (Fig. 3). Specifically, PGO modifies these nucleotides in a concentration dependent manner, similar to that observed in *S. cerevisiae* (Fig. 3). PGO modification of guanine demonstrates weak band intensities overall, compared to that in *S. cerevisiae*, representing lower frequency of guanine modifications in *C. albicans* than that in *S. cerevisiae*.

An important premise in RNA structural probing is that the RNA metabolic pathways in the cell should be unaltered upon treatment with the probing agent. Therefore, we investigated different RNA processing pathways of some important RNAs, including the small nucleolar RNA U3 (U3 snoRNA) and a transfer RNA (proline tRNA) using Northern blot analysis (Fig. 4).

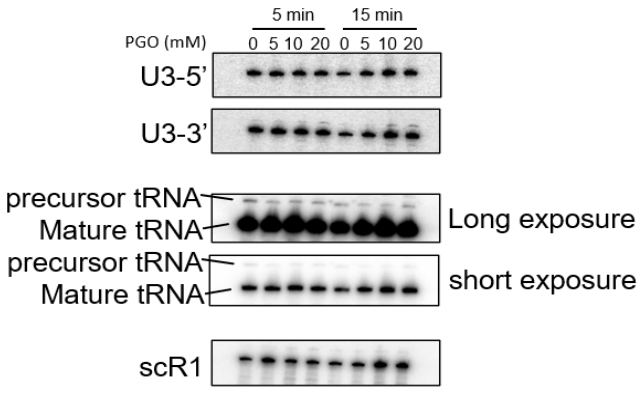

FIG 4. Northern blot analysis for U3-3', U3-5' snoRNAs and tRNAs. Reaction conditions at two timepoints are specified at the top of the Fig, and concentrations of PGO incubation are listed below reaction time conditions. scR1 RNA is the loading control.

We first analyzed the processing of U3 snoRNA, required for proper ribosome biogenesis. U3 is transcribed as a precursor with 5'- and 3'-extensions which need to be removed. We used radioactive probes against sequences within U3-5' and U3-3'. The band intensities for U3-3' and U3-5' species remain the same relative to the loading control (scR1 RNA) even after the concentration of PGO increases, indicating that in the presence of PGO probing reagent, the U3 snoRNA processing pathway remains intact (Fig. 4). We next analyzed the processing pathway of the proline tRNA ($tRNA^{Pro}$) which is transcribed as a precursor and undergoes processing before entering the translation pool. Our analysis revealed that the $tRNA^{Pro}$ processing pathway is

unaltered as well, as both mature and precursor tRNAs have similar band intensities in the presence of PGO, once again indicating comparable tRNA concentrations in PGO-treated yeast cells (Fig. 4).

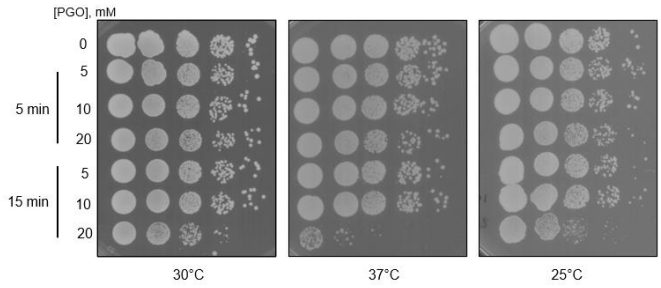

FIG 5. Serial dilution test for toxicity analysis in PGO-treated *S. cerevisiae*. Reaction conditions at two timepoints and concentrations of PGO are specified at the left side of the figure. Plates are incubated at three temperature temperatures- at 30℃, 37℃, and 25℃ as indicated.

Having established that the U3 and $tRNA^{Pro}$ processing pathways remains intact in PGO-treated *S. cerevisiae*, it is essential to establish the condition in which PGO treatment does not cause toxicity in this yeast strain. A serial dilution spot test was conducted on PGO-treated *S. cerevisiae* (Fig. 5). At concentrations below 20 mM, PGO does not cause toxicity for *S. cerevisiae* as judged by the similar size of individual colonies (Fig. 5). However, at the concentration of 20 mM and at 15-minute incubation period, *S. cerevisiae* no longer withstands the toxicity of PGO (Fig. 5).

Although we provide evidence that PGO is a viable guanine probe in *S. cerevisiae* and *C. albicans* in vivo, uracil probing chemical reagent and its reaction conditions have yet to be established in yeast in vivo. Carbodiimide N-cyclohexyl-N-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate (CMCT), a carbodiimide derivative, has been used to probe Us (Mitchell et al., 2019). We sought to investigate CMCT reactivity to probe uracil in *S. cerevisiae* in vivo. The 5.8S rRNA in *S. cerevisiae* is probed for uracil modification by CMCT in vivo (Fig. 6). CMCT

demonstrates effective uracil modification in *S. cerevisiae* in vivo (Fig. 6). A concentration-dependent relationship for U81 and U82 modification is observed (Fig.6). These data establish the conditions for probing uracils in RNA using CMCT.

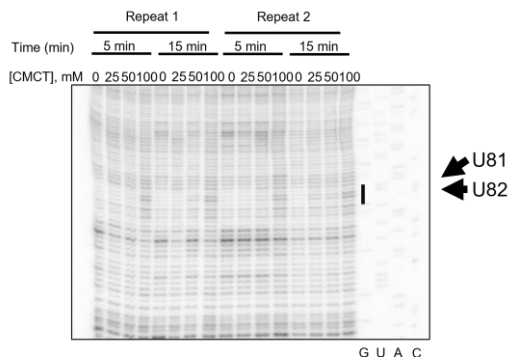

FIG 6. In vivo modification of yeast *S. cerevisiae* 5.8S rRNA by CMCT characterized by denaturing UREA-PAGE of cDNAs after reverse transcription. Reaction conditions at two timepoints and concentrations of glyoxal and its derivatives used are listed. Arrows represent modified RNA nucleotide positions by CMCT.

Discussion

RNA structural probing is an important indirect technique to study RNA conformations in the physiological environment. Specifically, it uses chemical reagents, mostly reactive electrophiles, to modify either the ribose backbone or nitrogenous bases in RNA. In this study, we demonstrated that PGO is a reactive and effective chemical probe for guanine modification in *S. cerevisiae* and is more potent than other glyoxal derivatives we studied here (Fig. 1). This conclusion is supported by prior research, which investigated the effectiveness and the condition to probe RNAs in vivo in rice *Oryza sativa* and in a gram-negative bacteria *E. coli* (Mitchell et al., 2017). This study demonstrated that PGO is the most effective probe for guanines among glyoxal family derivatives (Mitchell et al., 2017). The high reactivity of PGO comes from its increased hydrophobicity conferred by the phenyl function group in the molecule which allows it to penetrate through the phospholipid membrane bilayer (Mitchell et al., 2017). The hydrophobic phenyl group on PGO can strengthen interactions between PGO and hydrophobic

protein residues, orienting the electrophilic PGO carbonyl group in place for nucleophilic attack by the amidine of guanine (Mitchell et al., 2017).

Expanding upon this finding, we then tested the effectiveness of PGO to probe guanines in the 18S rRNA in *S. cerevisiae* in order to generalize PGO probing to a more diverse group of RNAs in *S. cerevisiae*. Our data show that PGO can modify the G913 residue in the 18S rRNA, (Fig. 2). Similar to PGO-probed 5.8S rRNA, a positive concentration-dependent relationship is found for the 18S probing (Fig. 2).

We expanded our findings to other fungi. Given the emergence of the human fungal pathogen *Candida albicans* as a public health threat, it is important to develop tools to study gene expression in *C. albicans*. We therefore applied the same experimental conditions to test the effectiveness and the reactivity of PGO towards *C. albicans* RNAs. The 5.8S rRNA is modified by PGO effectively in vivo, as G78 and G85 are glyoxalated in a positive concentration-dependent manner at both nucleotide positions (Fig. 3). These findings establish PGO as a suitable chemical reagent for probing guanines in *C. albicans*.

Our findings add PGO to the current arsenal of chemical probes used to probe the Watson-Crick face of the RNA. In addition to PGO which targets guanines, DMS modifies adenine and cytosine, leaving behind uracil. Mitchell III et al. have established the viability of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) for uracil probing in *O. sativa* and in a gram-negative bacteria *E. coli*, but carbodiimide N-cyclohexyl-N-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate (CMCT) is supposed to be 20 times more reactive than EDC due to the inductive effect that makes it a better electrophile for nucleophilic attack (Mitchell et al., 2019 and Wang et al., 2019). We could not obtain highly pure RNA from EDC-treated yeast cells. Therefore, CMCT was used to probe uracil in the 5.8S rRNA of *S. cerevisiae* to establish the chemical condition for

which it can perform uracil probing most effectively. CMCT is demonstrated to probe U81 and U82 effectively (Fig. 6). CMCT is too large to effectively penetrate cell wall because of the presence of a quaternary ammonium ion that constitutes a positive charge in CMCT (Want et al., 2019). The weak band intensities displayed may be due to a lack of entry of CMCT into the cytoplasm (Fig. 6).

The optimal condition for RNA structural probing should not impact the RNA processing pathways. To test this in PGO-treated cells, we analyzed the processing pathways of examples from two important small RNAs- the U3 snoRNA and $tRNA^{Pro}$ (Fig. 4). Our results indicate that important small RNA processing pathways in PGO-treated *S. cerevisiae* remain unaltered. As there are other pathways which can affect different RNA molecules in the cell and are not studied here, we studied the overall fitness of yeast cells upon PGO treatment. We conducted a serial dilution toxicity assay to analyze the effect of PGO concentration and incubation period on the growth of *S. cerevisiae*. We find that at the PGO concentration at 20 mM and for 15 minutes, cells are no longer thriving, as individual colony size is reduced (Fig. 5). The detrimental effect of treatment is further exacerbated by growing cells in temperature conditions other than 30℃ (Fig. 5). Therefore, at less than 20 mM PGO and less than 15-minute incubation period, *S. cerevisiae* is tolerant of PGO toxic effect (Fig. 5),

In summary, we established the optimal chemical conditions in which PGO and CMCT can effectively probe guanine and uracil, respectively. We found that PGO is a potent probe within the glyoxal family derivatives to probe guanine in yeast in vivo. PGO incubation with yeast does not affect its RNA processing pathways, and at the PGO concentrations less than 20 mM and at less than 15-minute incubation period, yeast cells can withstand the toxic effect of PGO. CMCT can effectively probe uracil in yeast in vivo, with a positive concentration-dependent relationship.

In the near future, we will complete the investigation of uracil modification by CMCT in *S. cerevisiae* in vivo by conducting the Northern blot to analyze RNA processing pathways and by performing the spot test to evaluate the condition at which *S. cerevisiae* cannot survive the toxic effect of CMCT. Furthermore, we will apply the same experimental conditions for uracil probing by CMCT in *C. albicans* in vivo in order to establish versatility of this probing agent in different yeast strains. Finally, we will investigate rRNA ribosomal subunit structures and their intermolecular interactions using mutational profiling (MaP) after chemical probe treatments, as this method enables sequencing of larger rRNAs and transcriptome-wide systems as a whole by creating mutations at the modification sites instead of conventional truncations by the RT, followed by sequencing (Zubradt et al. 2016). In conclusion, in vivo RNA structural probing is critical for understanding the stability and function of RNA. As the majority of the RNAs are involved in metabolic processes and the regulation of gene expression, elucidating the structural-function relationship in RNAs in vivo paves the way for pharmaceutical and biochemical applications to combat genetic disease.

Acknowledgement

I would like to thank Dr. Sohail Khoshnevis and Dr. Homa Ghalei for their guidance and support on my experiments, thesis development and editing. I am also thankful to all members of the Ghalei Lab at Emory University School of Medicine. I would like to further thank Dr. Anita Corbett for accepting to be a member of my thesis committee.

Contributions

I performed yeast growth assays with chemical probes, harvested cells and isolated total cell RNA. I also measured the RNA concentrations and shadowed reverse transcription and northern blot assays. Analysis of the RNA samples by reverse transcription, sequencing gels and northern blots was performed by Dr. Sohail Khoshnevis. Dr. Ghalei helped with the design of the experiments, interpretation of the data and provided funding (1R35GM138123).

Reference

1. Mitchell III, David.; S. Assmann.; P. Bevilacqua. “Probing RNA structure in vivo.” Current Opinion in Biology, 2019.
2. Bevilacqua, P.; S. Assmann. “Technique Development for Probing RNA Structure In Vivo and Genome-Wide.” Cold Spring Harbor Perspectives in Biology, 2018.
3. Mitchell III, D.; L. Ritchey.; H. Park et al. “Glyoxals as in vivo RNA structural probes of guanine base-pairing.” Cold Spring Harbor Laboratory Press, 2017.
4. Kubota, M.; C. Tran.; and R. Spitale. “Progress and challenges for chemical probing of RNA structure inside living cells.” Nature Chemical Biology, 2015.
5. Mitchell III, D.; A. Renda.; C. Douds et al. “In vivo RNA structural probing of uracil and guanine base-pairing by 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC).” Cold Spring Harbor Laboratory Press, 2022.
6. Wang, P.; A. Sexton.; W. Culligan et al. “Carbodiimide reagents for the chemical probing of RNA structure in cells.” Cold Spring Harbor Laboratory Press, 2019.

7. Wang, Q.; J. Liu.; Z. Chen et al. “G-quadruplex formation at the 30 end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase.” *Nucleic Acids Research,* 2011.
8. Wells, S.; J. Hughes, A. Igel, and M. Ares. “Use of dimethylsulfate to probe RNA structure in vivo.” *Methods in Enzymology,* 2000.
9. Ziehler, W.; and D. Engelke. “Chemical and Enzymatic Probes for Nucleic Acid Structure.” *Current Protocols in Nucleic Acid Chemistry,* 2000.
10. Scott, G. “RNA catalysis.” *Current Opinions in Structural Biology,* 1998.
11. Wu, J et al. “Ribogenomics: the Science and Knowledge of RNA.” *Genomics Proteomics Bioinformatics,* 2014.
12. Wang, Q et al. “G-quadruplex formation at the 3’ end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase.” *Nucleic Acids Research,* 2011.
13. Spitale, R et al. “RNA SHAPE analysis in living cells.” *Nature,* 2012.
14. Wilkinson, K.; E, Merino, and K. Weeks. “Selective 2’-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide solution.” *Nature,* 2006.
15. Parisien, M.; and F. Major. “Determining RNA three-dimensional structures using low-resolution data.” *Journal of Structural Biology,* 2012.
16. Buratti, E.; and F. Baralle. “Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process.” *Molecular and Cellular Biology,* 2004.
17. Zubradt, M et al. “DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo.” *Nature,* 2017.

18. Busan, S.; and K. Weeks. “Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2.” *Bioinformatics*, 2018.
19. Busan, S.; C. Weidmann.; A. Sengupta.; and K. Weeks. “Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies.” *Biochemistry,* 2019.
20. Verma-Gaur, J and A. Traven. “Post-transcriptional gene regulation in the biology and virulence of Candida albicans.” *Cellular Microbiology,* 2016.
21. Tijerina, P.; S. Mohr, and R. Russell. “DMS footprinting of structured RNAs and RNA–protein complexes.” *Nature*, 2007.
22. Mueller, D. “Arginine 328 of the &Subunit of the Mitochondrial ATPase in Yeast Is Essential for Protein Stability.” *The Journal of Biological Chemistry,* 1988.
23. Sulima, S et al. “Eukaryotic rpL10 drives ribosomal rotation.” *Nucleic Acid Research,* 2014.
24. Guo, L et al. “Arginine 107 of yeast ATP synthase subunit g mediates sensitivity of the mitochondrial permeability transition to phenylglyoxal.” *J. Biol. Chem*., 2018.
25. Hoon, S et al. “A Global Perspective of the Genetic Basis for Carbonyl Stress Resistance.” *G3 Genes*|*Genomes*|*Genetics,* 2011.