Published in final edited form as:

Methods. 2020 November 01; 183: 68–75. doi:10.1016/j.ymeth.2020.04.001.

Viral RNA structure analysis using DMS-MaPseq

Phillip Tomezsko^{1,2,3}, Harish Swaminathan¹, Silvi Rouskin^{1,*}

¹Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA

²Program in Virology, Harvard Medical School, Boston, Massachusetts, USA

³Brigham and Women's Hospital, Boston, Massachusetts, USA

Abstract

RNA structure is critically important to RNA viruses in every part of the replication cycle. RNA structure is also utilized by DNA viruses in order to regulate gene expression and interact with host factors. Advances in next-generation sequencing have greatly enhanced the utility of chemical probing in order to analyze RNA structure. This review will cover some recent viral RNA structural studies using chemical probing and next-generation sequencing as well as the advantages of dimethyl sulfate (DMS)-mutational profiling and sequencing (MaPseq). DMS-MaPseq is a robust assay that can easily modify RNA *in vitro*, in cell and in virion. A detailed protocol for whole-genome DMS-MaPseq from cells transfected with HIV-1 and the structure of TAR as determined by DMS-MaPseq is presented. DMS-MaPseq has the ability to answer a variety of integral questions about viral RNA, including how they change in different environments and when interacting with different host factors.

Keywords

RNA structure; Chemical probing; DMS-MaPseq; Dimethyl Sulfate; Next-generation sequencing

1. Introduction

Single-stranded RNA is able to form complex structures through both canonical and non-canonical base-pairing interactions and base-stacking [1]. The range of functions ascribed to RNA structure has expanded rapidly [2], [3], precipitated by the discovery that the catalytic element of the ribosome is RNA [4], [5], [6], [7]. All major classes of RNA, including mRNA, form structures that have functional importance [2]. RNA is able to form multiple alternative structures based on thermodynamic properties, but the structure is also influenced by the cellular environment, particularly by RNA binding proteins and RNA helicases [8],

*Corresponding Author: Dr. Silvi Rouskin, Whitehead Institute, 455 Main Street, Cambridge, MA 02142, srouskin@wi.mit.edu, Telephone: 617-324-6482.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Competing Interests

The authors have no competing interests to declare.

[9]. Given these factors, prediction of biologically relevant RNA structures is extremely difficult by thermodynamic modeling alone, although there have been advances in algorithms [10] and methods to ensure biological relevance [11]. Several approaches exist to experimentally study RNA structure, including chemical probing. The basis of chemical probing is to add modifications to open RNA bases but not paired RNA bases, which can be used as constraints to greatly improve the accuracy of prediction programs [12]. Dimethyl sulfate (DMS), the chemical described in the methodology below, is the oldest and one of the most widely used chemicals for RNA structure probing [13]. DMS adds methyl groups to unpaired adenine, cytosine and guanosine at neutral pH. However, methylation of the N1 and N3 positions of unpaired adenosine and cytosine nucleotides are most favored. These modifications occur within the Watson-Crick face and prevent reverse transcriptases from reading the base properly [13], [14]. In slightly basic buffer, DMS has been reported to modify all four bases [15], however this is more efficient *in vitro* than in cells.

The first RNA structure assays converted the modifications into signal by comparing intensity of bands after poly-acrylamide gel electrophoresis of truncated reverse transcription products, in the case of DMS this was termed DMS footprinting [16], [17]. In addition to DMS, there are multiple probing agents, each with a different reactivity profile for footprinting and ability to add modifications to different chemical sites in the nucleotide [14], [18]. One attractive feature of DMS is that it interacts with the Watson-Crick face of both canonical binding pairs [14]. Chemical probing reagents can also be used in combination, such as DMS and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), in order to get reactivity on a wider range of interactions [19]. Another approach to footprinting of RNAs based on structure is to treat with RNases that have specificity for either unpaired or paired bases [14].

In order to increase the throughput of chemical probing techniques, research has been devoted to combining chemical probing with next-generation sequencing. Advances in high-throughput sequencing allowed for DMS sequencing (DMS-Seq), the sequencing of reverse transcription termination products on a transcriptome-wide scale [8]. Discovery of reverse transcriptases that add random mutations in the cDNA when encountering DMS methylations on the RNA, including thermostable group II intron reverse transcriptase (TGIRT-III), has allowed for development of DMS-mutational profiling and sequencing (MaPseq) [20], [21]. The mutations are counted and DMS-MaPseq provides single-molecule RNA structure information which can reveal heterogeneous RNA structures and maximize sequencing depth on low-abundance RNA species [21]. TGIRT-III has an approximately 1:20 deletion to mismatch rate, providing a robust signal [21].

DMS-MaPseq has a number of advantages over other methods to investigate RNA structure, but it provides complementary information to these methods as well. X-ray crystallography and nuclear magnetic resonance (NMR) are regarded as the gold standards of RNA structure analysis; however, both methods are limited to *in vitro* folded RNAs. For X-ray crystallography and to a lesser extent NMR, the RNA must have an extremely rigid and homogenous structure in order to provide interpretable signal. For NMR in particular, the RNA must be small (<155 nt) in order to be resolved [22]. One advantage compared to DMS-MaPseq is that X-ray crystallography and NMR provides 3D structural information.

Another commonly used chemical probing technique is called Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) [23]. Cryogenic electron microscopy (cryoEM) is another extremely powerful tool to determine RNA structure *in vitro*. This method involves freezing purified RNA or RNA-protein complexes in a thin layer of ice and measuring the structure by electron microscopy. The images of the molecule in different orientations need to be sorted based on similarity in order to get a 3D projection of the target. Viral RNA-protein complexes such as the IRES of CrPV bound to the ribosome have unlocked insight into the mechanism of interaction [24]. More recent studies have used cryoEM to study the interaction of viral genomes and viral proteins, such as the nucleocapsid coating of the Hantaan virus and Ebola [25], [26]. Although crystallization is more efficient in the presence of RNA binding proteins, single-stranded RNA and even naked viral genomes have been analyzed with this method [27]. When comparing cryoEM to SHAPE, the secondary structure of naked STMV genome structure was consistent between methods however there was significantly more heterogeneity identified by the cryoEM study [28], [29].

SHAPE and SHAPE-Map use a family of acylating electrophiles to modify predominantly unpaired nucleotides in an RNA chain [30], [31], [32]. DMS-MaPseq and SHAPE-Map share similar theory but one key distinction is that DMS-based techniques are far less sensitive to RNA-binding proteins as compared to SHAPE-based techniques [32]. This difference is due to the fact that DMS is smaller than SHAPE reagents and modifies Watson-Crick face whereas SHAPE modifies the phosphodiester backbone. In order for a protein to provide protection from DMS it must directly interact with N1A or N3C whereas many different types of RNA-protein interactions provide protection from SHAPE [33]. In a comparison of *in vitro* and *in vivo* structure therefore, DMS measures the change in base-pairing induced by protein binding, whereas SHAPE measures both the direct change in accessibility from the protein and the indirect effect of protein binding on secondary structure. Both of these chemical probing strategies then can be useful depending on the proposed mechanism of structural change induced by protein binding. DMS probing in basic conditions by addition of bicine to the buffer followed by mutational profiling is termed PAIR-MaP. This method provides more information on the mutational profile of the RNA structure by modifying all four bases. The added information can be used to infer base-pairing partners based on correlation of DMS modifications. However, in order to use the information on all bases for PAIR-MaP analysis, sequencing needs to be deep enough to provide signal on U and G, which are modified at a much lower rate. Mustoe *et al.* recommend this technique for RNA structure analysis of a targeted RNA, rather than transcriptome-wide analysis [15]. Another method to probe base-pairing partners is called Mutate-and-Map; mutations are introduced into the DNA template before transcription by mutagenesis or an error-prone polymerase [34], [35]. The effects of the mutations on the secondary structure are determined by DMS-MaPseq. The data from all mutations together is used to identify base-pairing partners.

RNA viruses have been an area of particular interest for the RNA field because of an abundance of well-studied functional RNA structures. Embedding information in RNA structure allows the virus to maximize a small genome. Viral RNA structures are utilized in many diverse stages of the replication cycle, including packaging, replication, transcriptional

regulation, translational regulation and anti-host defense [36]. A notable family of RNA structures is the IRES family, that allow the viral RNA to bypass some or all of the normally essential host translational cofactors associated with the ribosome [37]. RNA structure is also relevant to DNA viruses and have similar functions; HHV8 was the first DNA virus to be shown to use an IRES [38] but other IRES elements have been identified in other DNA viral transcripts, such as polyomavirus SV40 [39]. Additionally, several classes of highly structured viral non-coding RNAs in gammaherpesviruses (such as HHV8 and EBV) contribute to transformation of the cell [40]. Even for the most well-characterized of viral RNA structures, many open questions remain that can benefit from a high-throughput, robust *in vivo* chemical probing methods such as DMS-MaPseq: 1) How do different cellular environments, most notably different expression patterns of RNA binding proteins, change the RNA structure? 2) How does the RNA structure change during the course of infection? 3) How do alternative forms of the structure impact regulation of the structure's function? 4) How do mutations or natural viral sequence diversity change RNA structures?

DMS-MaPseq and SHAPE-Map have been used to answer some of these questions for viral RNA structures. DMS-MaPseq was used to confirm the NMR structural prediction of an enterovirus IRES structure [41]. SHAPE-Map was recently used to determine the differential binding of RNA binding proteins in the nucleus, cytoplasm and virion for the PAN RNA of HHV8, which is crucial for suppression of the host anti-viral response during infection [42]. Two intriguing reports on the structure of the influenza virus genome in cells using DMS-MaPseq [43] and in virions using SHAPE-Map [44] had shed insight into how the viral segments form a complex network of structures and intramolecular interactions that allow for proper packaging and re-assortment. Importantly, these structures are stable even in the presence of RNA binding proteins. SHAPE-Map has also been used to study the HIV-1 RRE in different contexts. The RRE is a stable structure that enables export of unspliced and incompletely spliced HIV-1 RNA from the nucleus of infected cells through interaction with the viral protein Rev [45]. Recent studies using SHAPE-Map were able to uncover the binding site of two new inhibitors on the RRE [46], [47]. Another study was able to track slight changes in structure of the RRE over the course of infection and correlate structural changes to activity of the RRE, thus showing that the RRE is under selection pressure to maintain or increase activity as HIV-1 evolves in the host [48]. These studies reflect the power of DMS-MaPseq to answer complex questions about viral RNA structure.

2.1 Method Overview

The goal of this method is to produce *in vivo* and *in virion* DMS-modified RNA with a strong signal-to-noise ratio that can be used for library generation. This allows secondary structure analysis for the whole viral genome and all transcripts, as well as host genes. This protocol is written for HEK293t cells transfected with a plasmid containing full-length HIV-1, however this method is adaptable and notes will be included on how best to modify the approach for other cell systems and viruses. The general steps are cell culture and infection, DMS-modification and RNA extraction, and library generation. The libraries are then sequenced and the reads are aligned to the viral genome or any viral genomic regions of interest, in this case using Bowtie2. The DMS-induced mutations are then counted and ratiometric calculations are performed in order to determine the most highly reactive bases.

2.2 Detailed Method

2.2.1 HEK293t culture and transfection

Culture HEK293t (ATCC) in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 50 U/mL Penicillin/Streptomycin (ThermoFisher Scientific) at a concentration of ~1–2 million/mL during maintenance. For experimental setup, seed 0.9 million HEK293t cells per well of a 6-well plate and allow them to attach overnight. Transfect plasmid containing HIV-1_{NHG} using X-tremegene9 (MilliporeSigma) according to manufacturer's instructions. Incubate cells for 48 hours in order to achieve peak virion production. We treat a single plate as a single sample, however it is flexible in that a half plate or individual wells can be treated if the either cells or virus are limited. If suspension cells are used for DMS-modification, grow them at a density of ~1–2million/mL before DMS-modification.

2.2.2 DMS-modification of HIV-1 virions

First, isolate virions from supernatant. Remove all supernatant from the plate and add fresh media. Filter the supernatant through a syringe driven 0.22 µm PVDF filter (MilliporeSigma). Centrifuge the supernatant for 1 hour at 28,000xg, 4°C. Remove supernatant and resuspend pellet in 100 µL ice-cold PBS. Add 100 µL 2x modification buffer (400mM NaCl and 6mM MgCl₂²⁺) to virions and incubate at 37°C for 10 minutes on a Thermomixer. Add 20 µL DMS (MilliporeSigma) for a final percentage of ~9.1% DMS and incubate 10 minutes at 37°C while shaking at 1000 rpm on a Thermomixer. Add 440 µL β-mercaptoethanol (BME; MilliporeSigma) to neutralize the DMS. Purify RNA with the Clean and Concentrator –5 kit (Zymo Research) according to manufacturer's specifications for all RNA. We typically elute in 30 µL nuclease-free water and have a concentration of 20–50 ng/µL.

We have tried this protocol with and without detergent (1% SDS) in the modification buffer to lyse the virion, and found there was no difference. We speculate the acidification of the media from DMS disrupts the viral membrane and capsid. However, it is possible that other viruses are more resistant and need a lysing reagent in the modification buffer.

2.2.3 DMS-modification of HIV-1+ HEK293t cells

Pre-warm media to 37°C before DMS-modification. Add 200 µL DMS to 15 mL warm media (~1.33% final percentage of DMS). Remove supernatant from the wells of the plate and wash with PBS. Add 2 mL of the DMS-media to each well and place immediately in the incubator, incubate for 4 minutes. Remove media (save all DMS-containing waste for proper disposal) and add 2 mL of ice-cold 1:1 PBS:BME to each well. Scrape the cells off with a cell scraper and transfer to 50 mL conical. Centrifuge for 5 min at 1000xg, 4°C. Remove supernatant, resuspend pellet in 15 mL ice-cold PBS and centrifuge for 5 min at 1000xg, 4°C. Resuspend pellet in 1 mL Trizol reagent (ThermoFisher) and extract RNA according to manufacturer's specifications. We resuspend the purified RNA in 50 µL nuclease-free water, and have a typical concentration of ~2µg/µL.

2.2.4 rRNA subtraction

As DMS-modified RNA is often fragmented, and the poly-A tail is susceptible to methylation that could interfere with Oligo(dt) beads from getting quality mRNA isolation, rRNA subtraction is necessary. It is also possible that the viral RNA of interest doesn't have a poly-A tail, in which case rRNA subtraction is also preferred.

We have had success with commercially available rRNA subtraction kits including RiboZero (Illumina; discontinued), FastSelect (Qiagen) and RiboMinus (ThermoFisher Scientific), used according to manufacturer's specifications. However, for cost considerations, we also use an in-house oligo cocktail designed based on work by Adiconis *et al* [49]. For this method, start with 1–3 µg in a volume of 7 µL nuclease-free water (it is possible to use multiple reaction for library generation and combine after rRNA subtraction). Add 2 µL 5x hyb buffer (1M NaCl and 500 mM Tris-HCl pH 7.5) and 1 µL rRNA subtraction mix. Run on PCR machine –1°C/minute starting at 68°C and ending at 45°C. Once the reaction is at 45°C, add 33 µL nuclease-free water, 5 µL Hybridase buffer and 2 µL Hybridase Thermostable RNase H (Lucigen). Incubate reaction at 45°C for 30 minutes. Purify RNA using the RNA Clean and Concentrator –5 kit, eluting in 44 µL nuclease-free water. If there were multiple reactions, elute such that they are a total final volume of 44 µL and combine. Add 5 µL 10x Turbo DNase buffer and 1 µL Turbo DNase (ThermoFisher Scientific), and incubate for 30 minutes at 37°C. Add 5.5 µL Turbo DNase Inactivation reagent and incubate for 5 minutes at room temp. Centrifuge briefly and transfer supernatant to new tube. Purify RNA using the RNA Clean and Concentrator –5 kit and elute in 9 µL nuclease-free water.

The purified RNA is ready for library generation. However, if library generation is not necessary for your workflow, the RNA is suitable for RT-PCR and sequencing at this point.

2.2.5 Library Generation

Start with fragmentation of the RNA. Incubate RNA at 95°C for 1 minute to denature. Add 1 µL 10x RNA fragmentation reagent (ThermoFisher Scientific) and incubate at 70°C for 45 seconds. Place on ice and add 1 µL 10x Stop solution. Purify RNA using RNA Clean and Concentrator –5 following instructions for all fragment size collection and elute in 6.5 µL nuclease-free water.

The next step is to dephosphorylate the RNA fragment ends and add a linker to the 3' end. To 6.5 µL sample, add 1 µL CutSmart buffer (New England Biolabs), 1 µL Shrimp Alkaline Phosphatase (New England Biolabs) and 1 µL RNaseOUT (ThermoFisher Scientific). Incubate at 37°C for 1 hour. Add 6 µL 50% PEG 8000, 2.1 µL 10x T4 RNA Ligase Buffer, 2 µL T4 RNA ligase 2, truncKO (ThermoFisher Scientific) and 1 µL 20 M N12 linker (see Appendix A: Oligos). Incubate for 18 hours at 22°C. Purify RNA using RNA Clean and Concentrator –5 kit for all RNA fragments and elute in 15 µL nuclease-free water. Add 2 µL RecJ buffer, 1 µL RecJ exonuclease (Lucigen), 1 µL 5' Deadenylase (New England Biolabs) and 1 µL RNaseOUT to degrade excess linker. Incubate for 1 hour at 30°C. Purify using the RNA Clean and Concentrator –5 kit, following directions for RNA >200 nt as long as your RNA fragmentation conditions allow for this. By only purifying large fragments, more of the

linker is removed. However, if your target insert size is short, it is acceptable to purify using directions for all RNA. Elute in 11 μ L nuclease-free water.

For reverse transcription, add the following reagents to the sample:

- 4 μ L M-MLV reverse transcriptase buffer (ThermoFisher Scientific)
- 1 μ L dNTP mix
- 1 μ L 0.1M DTT
- 1 μ L 10 μ M Library RT Primer
- 1 μ L RNaseOUT
- 1 μ L TGIRT-III (Ingex)

Incubate at 65°C for 1.5 hours. Add 1 μ L 4N NaOH and incubate at 95°C for 3 minutes to degrade RNA. Add 20 μ L 2x TBE-Urea sample loading buffer (ThermoFisher Scientific) and load onto a pre-cast 10% TBE-Urea 10 mm Novex gel (ThermoFisher Scientific). Run for ~2 hours at 180V. Stain the gel with SybrGold (ThermoFisher Scientific). Excise the band of the expected size on a blue light box. Place the excised gel piece into a 0.65 mL Eppendorf tube that has been punctured through the bottom. Place the 0.65 mL Eppendorf into a 1.5 mL Eppendorf and centrifuge for 30 seconds at max speed to extrude the gel through the punctured tube. Add 400 μ L 300nM NaCl and incubate while shaking at 70°C to extract the DNA from the gel. Place sample in a 0.22m Costar Spin-X column (MilliporeSigma) and centrifuge at max speed for 30 seconds, then discard the column. Add 500 μ L 2-propanol and 3 μ L glycoblue (ThermoFisher Scientific) and freeze on dry ice. Centrifuge for 45 minutes at 18,000xg, 4°C. Remove supernatant and wash pellet with 250 μ L ice-cold 70% ethanol. Resuspend pellet in 15 μ L nuclease-free water. To sample add:

- 2 μ L 10x CircLigase reaction buffer
- 1 μ L 1mM ATP
- 1 μ L 50mM MnCl₂
- 1 μ L CircLigase (Lucigen)

Incubate for 2 hours at 60°C and then 10 minutes at 80°C. In a fresh PCR strip, add the following reagents:

- 11 μ L Nuclease-free water
- 4 μ L 5x HF Phusion Buffer
- 0.5 μ L Phusion (New England Biolabs)
- 1 μ L 10 M library reverse primer
- 1 μ L 10 M library forward primer
- 0.5 μ L dNTP
- 2 μ L circularized cDNA

Run the following PCR program:

- i. 30 seconds at 95°C
- ii. 15 seconds at 95°C
- iii. 5 seconds at 55°C
- iv. 10 seconds at 65°C
- v. Go to ii for 8–14 cycles
- vi. Hold at 4°C

We typically run the PCR for 10 and 12 cycles on the first attempt, and adjust according cycles higher or lower if the PCR needs to be rerun. After PCR, add 2 µL 6x loading dye (ThermoFisher Scientific) to sample and run on a pre-cast, 10 mm 8% TBE Novex gel (Thermo Fisher Scientific) for ~50 minutes at 180V. Stain gel with SybrGold. Repeat the gel extraction process from gel extrusion to 2-propanol precipitation and centrifugation. After 70% EtOH wash, resuspend in 11 µL nuclease-free water. Dilute 1 µL of the sample 1:5 and submit to Bioanalyzer in order to check concentration and size of product.

2.2.6 Sequencing

Longer-sequencing reads are preferred, however due to the fragmentation induced by DMS, there is a need to balance optimize read-length. Paired-end is also highly recommended. We have had success running 75x75nt, 150x150nt and 300x300nt on Illumina iSeq, MiSeq and HiSeq. The right sequencing length also relies on the target insert size. For a library generation, we recommend a target insert size of ~100–250 nt depending on the sequencing length.

2.2.7 Quality Control

The following steps are contained within the following code for DMS-MaPseq. First, FastQC is run on all samples in order to summarize basic quality of the reads. Next Trimalore is used to remove bases of reads with a Phred score <20 from the reads. The '--fastqc' option is included to run FastQC on the post-trimming reads.

2.2.8 Mapping

We typically use Bowtie2 [50] as the aligner and provide a transcriptome or genome for the reference as needed. It is also possible to use a splice-aware aligner such as HISAT2 [51] however. For Bowtie2, we align with the following options '--local --no-unal --no-discordant --no-mixed -X 1000 -L 12'. This set of options functions to allow for the maximum number of mismatches by reducing the seed length and running with the local option. Allowing for short seed length is important since there will be mismatches throughout the read because of the DMS-induced mutations. The remaining options work to reduce run time and prevent other technical artefacts stemming from discordant and mixed pairs. The output of Bowtie2 is a .sam file that is used downstream to count mutations and compare to sequencing depth.

2.2.8 Bitvector generation, RNA and visualization

After mapping, a bitvector file is generated that counts the mutations. Each read is converted into a vector of '0' for matches, '1' for deletion, or 'A', 'T', 'C' or 'G' to indicate the identity of the mutation. Mates are compared, and a '?' or '.' replaces any ambiguous or missing bases. The bitvectors are then used to count mutations and normalized by sequencing depth and mutation rate in order to provide normalized DMS reactivity. We typically aim for a minimum of 1000-fold coverage per base for population average DMS signal analysis for high-quality data.

The RNA structure prediction program RNAstructure is run using the RSample function in order to use the normalized DMS reactivities per base as constraints for folding [52]. This produces both a visualization of the RNA secondary structure and bracket notation folding constraints. For final visualization, we recommend using VARNA with the bracket notation constraints and color code by reactivity [53]. A downloadable version of the DMS-MaPseq pipeline can be found at rundmc.wi.mit.edu/cluster/dreem.

3.1 Results

HEK293t cells were transfected with HIV-1_{NHG}. The libraries were generated using 10 g total RNA, rRNA was subtracted by using both the RNase H hybridization methods and RiboMinus. The bioanalyzer showed that the final PCR constructs had an average size of 379 nt with a range of 253–496 for one particular library. This corresponded with an average insert size of 259 nt. The libraries were sequenced and filtered for quality, then mapped to the HIV-1 genome in windows of 50–100 nt. The number of mutations per read for libraries was compared to number of mutations per read for a site-specific PCR based approach for a region of the HIV-1 genome from the same starting total RNA.

The number of mutations per read from the PCR amplified read is approximately a normal distribution around an average of 4.42 mutations per alignment with a length of 243 nt. The number of mutations per read for the library more closely approximates a Poisson's distribution with an average of 0.55 mutations per alignment with a length of 100 nt. Since the DMS-modified RNA came from the same sample, part of the difference is likely due to the shorter read length in the library. If we align the PCR sample to a 100 nt window, we find that the average mutations per alignment is 1.98. We therefore speculate that the fragmentation step present in library generation enriches for RNA that has a low modification rate since modified RNA is susceptible to over-fragmentation. Next, we viewed the first window of the HIV-1 genome, which contains the well-described TAR structure, which is a positive regulator of HIV-1 transcription [54]. We plotted the raw DMS-MaPseq data in a bar graph of mutational frequency per position of the genome. The only reactive bases are A's and C's, as is expected for quality DMS-MaPseq data. The baseline mutational frequency (ie U and G) is ~10-fold less than the mutational frequency for reactive bases. Next, we made a structural model with RNAstructure, using the DMS-MaPseq data on A and C bases as constraints. The model was visualized with VARNA. The model of TAR recapitulates what has been shown before in literature and represents an extremely stable RNA structure, with a free energy of -23.6 kcal/mol.

4.1 Hints and Tips for Troubleshooting

1. For suspension cells, follow the same steps as adherent cells, except the step where the DMS-media is removed from the adherent cells before detaching them. Since the media cannot be removed, simply add the PBS:BME wash directly and proceed with washing.
2. A range of DMS concentrations should be tried in an initial experiment for each particular virus and cell type. 2% DMS for five minutes is a good starting point. The optimal DMS concentration allows for the most mutations per sequencing read possible without fragmenting the RNA so much that library generation is compromised. Over-modification can be seen by Bioanalyzer or by sequencing reads that are low-quality with high sequence duplication and overrepresented reads. In some scenarios, the reads in an over-modified sample will have very few mutations, because what is sequenced are the RNAs with a low modification rate. Under modification yields a low signal-to-noise ratio after sequencing.
3. Troubleshooting the fragmentation conditions is critically important for library generation. First, you should choose a desired insert size. Then, you should use a portion of your sample to try different fragmentation conditions (or DMS-modified RNA using the same treatment condition if the sample is too precious to spare any). We have found that > 1 min at 70°C is a good starting place, but changing duration of fragmentation and temperature may be necessary. If the RNA of interest is extremely structured and there is concern for fragmentation bias, higher temperature (80–95°C) would be better. It is important to keep in mind that DMS treatment itself has fragmented the RNA.
4. It is important to include untreated controls of the same sample. This is helpful in determining the baseline mutation rate of your enzyme and PCR setup, as well as for identifying any mutations present in your sample compared to the reference genome that are not DMS-induced.
5. It is helpful to include an RNA of known size. We typically order RNA oligos from IDT and use 1–10 M in addition to experimental conditions to help guide troubleshooting and gel excision. We use this control in parallel in all steps except rRNA subtraction, for cost consideration.
6. When performing this library generation for the first time or when troubleshooting problems, it can be helpful to submit the sample for bioanalyzer after multiple steps. For instance, we submit after rRNA subtraction, after linker ligation, after excess linker degradation and after the final PCR gel extraction if we have problems with a particular sample.

Funding

This work was supported by the Center of HIV-1 RNA Studies (CRNA), the Smith Family Foundation, the Burroughs Wellcome fund and the National Institutes of Health [grant number R21AI134365], awarded to S.R.

Appendix A:: Oligos

We ordered all oligos from IDT, Coralville, IA, USA.

N12 Linker-/5rApp/

TCNNNNNNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGA/3ddC/

Library RT primer-/5Phos/AGATCGGAAGAGCACACGTCTGAACTCCAG/iSp18/

TCTTTCCCTACACGAC GCTCTTCCGATCT

Library PCR forward primer-

CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCT

C XXXXXX=Index

Library PCR reverse primer-

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC

Appendix B:: Software used

FastQC v 0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)

TrimGalore v 0.4.1 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)

Bowtie2 v 2.3.4.1 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>)

RNAstructure v 6.0.1 (<https://rna.urmc.rochester.edu/RNAstructure.html>)

VARNA v 3.93 (<http://varna.lri.fr/>)

Plotly v 3.2.1 (<https://plot.ly/#/>)

Adobe Illustrator CC 23.0.1

Biorender (<https://biorender.com>)

DMS-MaPseq rundmc.wi.mit.edu/cluster/dreem

Abbreviations:

DMS	Dimethyl Sulfate
DMS-MaPseq	DMS Mutational Profiling and Sequencing
DMS-Seq	DMS Sequencing
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
SHAPE-Map	SHAPE and Mutational Profiling
TGIRT-III	thermostable group II intron reverse transcriptase

rRNA	ribosomal RNA
IRES	internal ribosome entry sites
RRE	Rev Response Element

References

- [1]. Lemieux S, Major F, RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire, *Nucleic Acids Res* 30(19) (2002) 4250–63. [PubMed: 12364604]
- [2]. Mortimer SA, Kidwell MA, Doudna JA, Insights into RNA structure and function from genome-wide studies, *Nat Rev Genet* 15(7) (2014) 469–79. [PubMed: 24821474]
- [3]. Strobel EJ, Yu AM, Lucks JB, High-throughput determination of RNA structures, *Nat Rev Genet* 19(10) (2018) 615–634. [PubMed: 30054568]
- [4]. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* 289(5481) (2000) 905–20. [PubMed: 10937989]
- [5]. Schlutzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A, Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution, *Cell* 102(5) (2000) 615–23. [PubMed: 11007480]
- [6]. Wimberly BT, Brodersen DE, Clemons WM Jr., Morgan-Warren RJ, Carter AP, Vonnrhein C, Hartsch T, Ramakrishnan V, Structure of the 30S ribosomal subunit, *Nature* 407(6802) (2000) 327–39. [PubMed: 11014182]
- [7]. Ramakrishnan V, The ribosome emerges from a black box, *Cell* 159(5) (2014) 979–984. [PubMed: 25416939]
- [8]. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS, Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo, *Nature* 505(7485) (2014) 701–5. [PubMed: 24336214]
- [9]. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM, Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq, *Nat Protoc* 10(7) (2015) 1050–66. [PubMed: 26086407]
- [10]. Fallmann J, Will S, Engelhardt J, Gruning B, Backofen R, Stadler PF, Recent advances in RNA folding, *J Biotechnol* 261 (2017) 97–104. [PubMed: 28690134]
- [11]. Mathews DH, How to benchmark RNA secondary structure prediction accuracy, *Methods* 162–163 (2019) 60–67.
- [12]. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH, Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc Natl Acad Sci U S A* 101(19) (2004) 7287–92. [PubMed: 15123812]
- [13]. Peattie DA, Gilbert W, Chemical probes for higher-order structure in RNA, *Proc Natl Acad Sci U S A* 77(8) (1980) 4679–82. [PubMed: 6159633]
- [14]. Ehresmann C, Baudin F, Mougél M, Romby P, Ebel JP, Ehresmann B, Probing the structure of RNAs in solution, *Nucleic Acids Res* 15(22) (1987) 9109–28. [PubMed: 2446263]
- [15]. Mustoe AM, Lama NN, Irving PS, Olson SW, Weeks KM, RNA base-pairing complexity in living cells visualized by correlated chemical probing, *Proc Natl Acad Sci U S A* 116(49) (2019) 24574–24582. [PubMed: 31744869]
- [16]. Inoue T, Cech TR, Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase, *Proc Natl Acad Sci U S A* 82(3) (1985) 648–52. [PubMed: 2579378]
- [17]. Tijerina P, Mohr S, Russell R, DMS footprinting of structured RNAs and RNA-protein complexes, *Nat Protoc* 2(10) (2007) 2608–23. [PubMed: 17948004]
- [18]. Chang C, Lee CG, Chemical modification of ribonucleic acid. A direct study by carbon-13 nuclear magnetic resonance spectroscopy, *Biochemistry* 20(9) (1981) 2657–61. [PubMed: 7236629]

- [19]. Mitchell D 3rd, Renda AJ, Douds CA, Babitzke P, Assmann SM, Bevilacqua PC, In vivo RNA structural probing of uracil and guanine base-pairing by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC), *RNA* 25(1) (2019) 147–157. [PubMed: 30341176]
- [20]. Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunicke-Smith S, Swamy S, Kuersten S, Lambowitz AM, Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing, *RNA* 19(7) (2013) 958–70. [PubMed: 23697550]
- [21]. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S, DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo, *Nat Methods* 14(1) (2017) 75–82. [PubMed: 27819661]
- [22]. Barnwal RP, Yang F, Varani G, Applications of NMR to structure determination of RNAs large and small, *Arch Biochem Biophys* 628 (2017) 42–56. [PubMed: 28600200]
- [23]. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM, RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE), *J Am Chem Soc* 127(12) (2005) 4223–31. [PubMed: 15783204]
- [24]. Spahn CM, Jan E, Mulder A, Grassucci RA, Sarnow P, Frank J, Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor, *Cell* 118(4) (2004) 465–75. [PubMed: 15315759]
- [25]. Arragain B, Reguera J, Desfosses A, Gutsche I, Schoehn G, Malet H, High resolution cryo-EM structure of the helical RNA-bound Hantaan virus nucleocapsid reveals its assembly mechanisms, *Elife* 8 (2019).
- [26]. Kirchdoerfer RN, Saphire EO, Ward AB, Cryo-EM structure of the Ebola virus nucleoprotein-RNA complex, *Acta Crystallogr F Struct Biol Commun* 75(Pt 5) (2019) 340–347. [PubMed: 31045563]
- [27]. Gopal A, Zhou ZH, Knobler CM, Gelbart WM, Visualizing large RNA molecules in solution, *RNA* 18(2) (2012) 284–99. [PubMed: 22190747]
- [28]. Athavale SS, Gossett JJ, Bowman JC, Hud NV, Williams LD, Harvey SC, In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus, *PLoS One* 8(1) (2013) e54384. [PubMed: 23349871]
- [29]. Garmann RF, Gopal A, Athavale SS, Knobler CM, Gelbart WM, Harvey SC, Visualizing the global secondary structure of a viral RNA genome with cryo-electron microscopy, *RNA* 21(5) (2015) 877–86. [PubMed: 25752599]
- [30]. Siegfried NA, Busan S, Rice GM, Nelson JA, Weeks KM, RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP), *Nat Methods* 11(9) (2014) 959–65. [PubMed: 25028896]
- [31]. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM, Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis, *Nat Protoc* 10(11) (2015) 1643–69. [PubMed: 26426499]
- [32]. Smola MJ, Weeks KM, In-cell RNA structure probing with SHAPE-MaP, *Nat Protoc* 13(6) (2018) 1181–1195. [PubMed: 29725122]
- [33]. Smola MJ, Calabrese JM, Weeks KM, Detection of RNA-Protein Interactions in Living Cells with SHAPE, *Biochemistry* 54(46) (2015) 6867–75. [PubMed: 26544910]
- [34]. Kladwang W, Das R, A mutate-and-map strategy for inferring base pairs in structured nucleic acids: proof of concept on a DNA/RNA helix, *Biochemistry* 49(35) (2010) 7414–6. [PubMed: 20677780]
- [35]. Cheng CY, Kladwang W, Yesselman JD, Das R, RNA structure inference through chemical mapping after accidental or intentional mutations, *Proc Natl Acad Sci U S A* 114(37) (2017) 9876–9881. [PubMed: 28851837]
- [36]. Rausch JW, Sztuba-Solinska J, Le Grice SFJ, Probing the Structures of Viral RNA Regulatory Elements with SHAPE and Related Methodologies, *Front Microbiol* 8 (2017) 2634. [PubMed: 29375504]
- [37]. Mailliot J, Martin F, Viral internal ribosomal entry sites: four classes for one goal, *Wiley Interdiscip Rev RNA* 9(2) (2018).
- [38]. Bielecki L, Talbot SJ, Kaposi's sarcoma-associated herpesvirus vCyclin open reading frame contains an internal ribosome entry site, *J Virol* 75(4) (2001) 1864–9. [PubMed: 11160685]

- [39]. Yu Y, Alwine JC, 19S late mRNAs of simian virus 40 have an internal ribosome entry site upstream of the virion structural protein 3 coding sequence, *J Virol* 80(13) (2006) 6553–8. [PubMed: 16775341]
- [40]. Chavez-Calvillo G, Martin S, Hamm C, Sztuba-Solinska J, The Structure-To-Function Relationships of Gammaherpesvirus-Encoded Long Non-Coding RNAs and Their Contributions to Viral Pathogenesis, *Noncoding RNA* 4(4) (2018).
- [41]. Tolbert M, Morgan CE, Pollum M, Crespo-Hernandez CE, Li ML, Brewer G, Tolbert BS, HnRNP A1 Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation, *J Mol Biol* 429(19) (2017) 2841–2858. [PubMed: 28625847]
- [42]. Sztuba-Solinska J, Rausch JW, Smith R, Miller JT, Whitby D, Le Grice SFJ, Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins, *Nucleic Acids Res* 45(11) (2017) 6805–6821 [PubMed: 28383682]
- [43]. Simon LM, Morandi E, Luganini A, Griboaldo G, Martinez-Sobrido L, Turner DH, Oliviero S, Incarnato D, In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs, *Nucleic Acids Res* 47(13) (2019) 7003–7017. [PubMed: 31053845]
- [44]. Dadonaite B, Gilbertson B, Knight ML, Trifkovic S, Rockman S, Laederach A, Brown LE, Fodor E, Bauer DLV, The structure of the influenza A virus genome, *Nat Microbiol* (2019).
- [45]. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR, The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA, *Nature* 338(6212) (1989) 254–7. [PubMed: 2784194]
- [46]. Dai Y, Wynn JE, Peralta AN, Sherpa C, Jayaraman B, Li H, Verma A, Frankel AD, Le Grice SF, Santos WL, Discovery of a Branched Peptide That Recognizes the Rev Response Element (RRE) RNA and Blocks HIV-1 Replication, *J Med Chem* 61(21) (2018) 9611–9620. [PubMed: 30289719]
- [47]. Dai Y, Peralta AN, Wynn JE, Sherpa C, Li H, Verma A, Le Grice SFJ, Santos WL, Molecular recognition of a branched peptide with HIV-1 Rev Response Element (RRE) RNA, *Bioorg Med Chem* 27(8) (2019) 1759–1765. [PubMed: 30879859]
- [48]. Sherpa C, Jackson PEH, Gray LR, Anastos K, Le Grice SFJ, Hammarskjold ML, Rekosh D, Evolution of the HIV-1 Rev Response Element during Natural Infection Reveals Nucleotide Changes That Correlate with Altered Structure and Increased Activity over Time, *J Virol* 93(11) (2019).
- [49]. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ, Comparative analysis of RNA sequencing methods for degraded or low-input samples, *Nat Methods* 10(7) (2013) 623–9. [PubMed: 23685885]
- [50]. Langmead B, Salzberg SL, Fast gapped-read alignment with Bowtie 2, *Nat Methods* 9(4) (2012) 357–9. [PubMed: 22388286]
- [51]. Kim D, Langmead B, Salzberg SL, HISAT: a fast spliced aligner with low memory requirements, *Nat Methods* 12(4) (2015) 357–60. [PubMed: 25751142]
- [52]. Reuter JS, Mathews DH, RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinformatics* 11 (2010) 129. [PubMed: 20230624]
- [53]. Darty K, Denise A, Ponty Y, VARNA: Interactive drawing and editing of the RNA secondary structure, *Bioinformatics* 25(15) (2009) 1974–5. [PubMed: 19398448]
- [54]. Muesing MA, Smith DH, Capon DJ, Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein, *Cell* 48(4) (1987) 691–701. [PubMed: 3643816]

Highlights

- DMS-MaPseq is a robust assay that can be used to probe the secondary structure of viral RNA in many environments, including intracellular RNA structure.
- DMS-MaPseq protocol is presented for virally infected or transfected cells as well as virions.
- DMS-modified RNA can be used for RT-PCR or whole-genome library generation.
- Library generation quality control and DMS-MaPseq data for HIV-1 TAR is presented.

Figure 1-

Overview of DMS-MaPseq method with infected/transfected cells and virions. The first column depicted DMS modification of RNA intracellularly or in virion. The second column shows the general library generation protocol. The third column shows sequencing and conceptual analysis.

Figure 2-

Bioanalyzer trace of library generation and distribution of mutations per read. A) Bioanalyzer raw electrophoresis image and quantification for a fully constructed DMS-MaPseq library. B) On the left is a histogram of mutations per read for a site-specific PCR sample directed towards a region of the HIV-1 genome in HEK293t cells transfected with HIV-1_{NHG} with an alignment length of 243 nt. In the center is a histogram of mutations per alignment for the same exact PCR sample aligned to a shorter window in the region of interest. On the right, the same starting DMS-modified RNA was used for library generation and the histogram of mutations per read is shown for the same region of the HIV-1 genome.

Figure 3-

DMS-MaPseq derived structural model for HIV-1 TAR. A) Raw data from a whole-genome library DMS-MaPseq for HIV-1 TAR from HEK293t cells transfected with HIV-1_{NHG}. The bar graph shows the mutational fraction, as a result of DMS-induced methylation, for all reads at each nucleotide position of the region of interest. Different nucleotides are color-coded. B) structural model for HIV-1 TAR based on the DMS-MaPseq constraints. The structural model was made using RNAstructure and visualized with VARNA. C) Scatterplot for replicates of TAR and immediate downstream sequence from two HEK293t and HIV-1_{NHG} libraries.