Title
RNA Structure Analysis of Viruses Using SHAPE

Journal
Current Protocols in Microbiology, 30(1)

ISSN
1934-8525

Authors
Burrill, Cecily P
Andino, Raul

Publication Date
2013-10-01

DOI
10.1002/9780471729259.mc15h03s30

Peer reviewed

HHS Public Access

Author manuscript

Curr Protoc Microbiol. Author manuscript; available in PMC 2016 April 09.

Published in final edited form as:

Curr Protoc Microbiol. ; 30: Unit–15H.3.. doi:10.1002/9780471729259.mc15h03s30.

RNA Structure Analysis of Viruses Using SHAPE

Cecily P. Burrill[1] and **Raul Andino**[1]

[1]Department of Microbiology & Immunology, University of California, San Francisco, California, USA

Abstract

Selective 2'hydroxyl acylation analyzed by primer extension (SHAPE) provides a means to investigate RNA structure with better resolution and higher throughput than has been possible with traditional methods. We present several protocols, which are based on a variety of previously published methods and were adapted and optimized for the analysis of poliovirus RNA in the Andino laboratory. These include methods for non-denaturing RNA extraction, RNA modification and primer extension, and data processing in ShapeFinder.

Keywords

RNA structure; selective 2'hydroxyl acylation; primer extension; SHAPE

INTRODUCTION

Selective 2'hydroxyl acylation analyzed by primer extension (SHAPE) was developed as a method to interrogate RNA structure (specifically backbone flexibility) with single nucleotide resolution. SHAPE reagents are hydroxyl-selective electrophiles that can acylate the ribose 2'-hydroxyl group to form a stable 2'-*O*-adduct, which can be detected as a stop to primer extension by reverse transcriptase (Wilkinson et al., 2006). There are several SHAPE reagents available, including *N*-methylisotoic anhydride (NMIA) and 1-methyl-7-nitroisotoic anhydride (1M7). These protocols use NMIA, which is commercially available and adequate for most purposes, however 1M7 reacts more quickly and can be useful for monitoring rapid RNA folding events. SHAPE reagents are relatively insensitive to base identity (Wilkinson et al., 2009) but very sensitive to conformational dynamics. They react preferentially with flexible RNA nucleotides and poorly with those that are base-paired or otherwise conformationally constrained (Wilkinson et al., 2006). Thus constrained (structured) nucleotides are generally protected from acylation and are said to have low SHAPE reactivity, while flexible (unstructured) nucleotides are unprotected and have high SHAPE reactivity.

Here we present a series of protocols for the analysis of RNA structure by SHAPE. We begin with RNA extraction in a non-denaturing manner and a discussion of the choice of RNA source material (see Basic Protocol 1). We then move to modification of the extracted

phone: 415-502-6357 (lab); 415-502-6358 (RA office), cecily.burrill@ucsf.edu, raul.andino@ucsf.edu.

RNA and analysis of the modification by primer extension and capillary electrophoresis as described in Basic Protocol 2. The Alternate Protocol describes folding and modification of in vitro transcribed RNA. Finally we outline the processing of the raw electropherograms using the freely available ShapeFinder software and the normalization of the output SHAPE reactivities in Basic Protocol 3. It should be noted that these protocols are based substantially on methods published by the research groups lead by Dr. Kevin M. Weeks, Dr. Morgan C. Giddings, and Dr. David H. Mathews (McGinnis et al., 2009; Vasa et al., 2008; Watts et al., 2009; Wilkinson et al., 2006), as well as the kind advice of members of those groups.

BASIC PROTOCOL 1: NON-DENATURING RNA EXTRACTION

Choosing a source for your RNA of interest is a critical decision. Your RNA of interest should ideally be highly abundant and, as far as possible, should exist as a uniform population in a single conformation. The output SHAPE data will reflect a population average, so if there are multiple conformations of similar abundance, the average may not be terribly informative. For example, when examining poliovirus RNA we chose to examine RNA from purified virions rather than from infected cells (Burrill, et al., 2012, under review). While the latter is more abundant, the viral RNA is thought to exist in multiple confirmations during infection. Virion RNA could more reasonable be supposed to represent a single species of uniform confirmation. Similar reasoning presumably led to the choice of HIV-1 virion RNA, the first full-length viral genome analyzed by SHAPE (Watts et al., 2009). Purity of the RNA is less important than abundance and uniform conformation. The primers used (Basic Protocol 2) are specific so the presence of other RNAs is problematic only if they interact with the RNA of interest and therefore alters its SHAPE reactivity by protecting certain nucleotides.

The critical criteria for a non-denaturing RNA extraction protocol are that it should avoid both chemical denaturants, such as the chaotropic salts found in many RNA extraction reagents, and also high temperatures. When the RNA is in solution, it is also important to maintain appropriate concentrations of mono- and divalent cations. The following is based on the method described by (Watts et al., 2009) for the extraction of HIV-1 RNA. It has been modified to include a single phenol extraction rather than a succession of phenol:chloroform and chloroform extractions. The single extraction was found to be adequate for the extraction of poliovirus RNA. Poliovirus is a non-enveloped RNA virus. For enveloped viruses, or other RNA species with a higher concentration of lipids in the source material, chloroform extraction may be necessary.

Materials

RNA source material (e.g. purified virions or cell lysate)

20% (w/v) SDS

10 mg/mL Proteinase K

Tris-saturated Phenol (pH 7.4)

3 M NaCl

100% EtOH

RNA Modification Buffer

NanoDrop Spectrophotometer

1. Add 1/20 volume 20% (w/v) SDS and 1/100 volume 10 mg/mL Proteinase K to RNA source material. Incubate 1 h at 37°C.
2. Add an equal volume of Tris-saturated phenol and shake to mix well. Centrifuge 15 min at 12,000× *g*.
3. Transfer aqueous (top) phase to a clean tube. Add 1/10 volume 3 M NaCl and 3 volumes 100% EtOH. Incubate at least 20 min at −70°C.
4. To pellet RNA, centrifuge 15 min at ≥ 12,000× *g*, 4°C.
5. Carefully remove supernatant. Dissolve pellet in RNA modification buffer. Check RNA concentration on NanoDrop spectrophotometer.

   The NanoDrop gives total RNA concentration in ng/μL. In order to accurately adjust the RNA concentration for SHAPE (Basic Protocol 2), you must determine what proportion of total RNA is your RNA of interest. If your source material is relatively pure (e.g. purified virions) it may be sufficient to assume 100%. If not, this can be most accurately determined by qPCR. Store RNA at −70°C.

BASIC PROTOCOL 2: SELECTIVE 2'HYDROXY ACYLATION ANALYZED BY PRIMER EXTENSION (SHAPE)

Having extracted your RNA of interest in a (hopefully) native confirmation, then next step is to probe its structure with the SHAPE reagent (NMIA, in this protocol), perform a reverse transcription reaction, and analyze the cDNA fragments produced. NMIA selectively modifies unconstrained (typically single-stranded) nucleotides. Because the modifications act as stops to the reverse transcriptase, cDNA fragments corresponding to heavily modified nucleotides are more abundant than those corresponding to protected nucleotides. If the reverse transcription reaction is primed with a fluorescently-tagged oligo, the relative abundance of each cDNA fragment can be assessed from a capillary electropherogram, similar to those produced by conventional cDNA sequencing.

Fluorescently-tagged oligos are available from a number of companies including Applied Biosystems, IDT, and TriLink Biotechnologies. Choice of compatible fluorophores depends on the sequencing instrument used for capillary electrophoresis. Some dyes are proprietary and available from only one company. One set of fluorophores to which capillary electrophoresis machines are commonly calibrated is 6-FAM, VIC, NED, and PET (available from Applied Biosystems). Of these, we find that 6-FAM and VIC give the strongest and most consistent signal, and always use VIC as fluorophore 1 and 6-FAM as fluorophore 2.

The following approach uses two fluorophores rather than the three or four required for most published methods (Wilkinson et al., 2006), and is based on the kind advice of Dr. David Mauger. Using only two fluorophores results is significant savings on reagents (fluorescently-tagged oligos are expensive), but requires somewhat more laborious processing of the data in ShapeFinder (see Basic Protocol 3).

Materials

RNA of interest in RNA modification buffer (Basic Protocol 1)

Dimethyl Sulfoxide (DMSO)

32.5 mM *N*-methylisotoic anhydride (NMIA) in DMSO

20 μg/mL glycogen

50 mM EDTA

3 M NaCl

100% EtOH

0.5× TE

In vitro transcribed RNA of interest in 0.5× TE (for sequencing/alignment reactions)

1 μM oligo 5'-tagged with fluorophore 1

1 μM oligo 5'-tagged with fluorophore 2

SuperScript III (Invitrogen)

5× First-Strand Buffer (supplied with SuperScript III)

0.1 M DTT (supplied with SuperScript III)

dNTP Mix (10 mM each dNTP)

10 mM ddATP

10 mM ddCTP

Nuclease-free water

1 M NaOH

1 M HCl

3 M NaOAc (pH 5.2)

75% (v/v) EtOH

HiDi Formamide (Applied Biosystems)

RNA Modification

1. Adjust RNA of interest concentration to 0.111 pmol/μL.

Once you have determined the concentration in ng/μL of your RNA of interest, you can use its molecular weight to convert to pmol/μL. First

calculate the molecular weight of your RNA product in ng/pmol. [M.W. (ng/pmol) = length (nt) × 0.3205 + 0.1590.] Divide concentration (ng/μL) by molecular weight (ng/pmol) to get pmol/μL.

2. Divide RNA into two equal parts. Incubate both tubes 20 min at 37°C.

3. Keep tubes at 37°C. To one part, add 1/9 volume DMSO. This is the mock-treated control. To the other part, add 1/9 volume 32.5 mM NMIA in DMSO. This is the experimental sample. Pipette up and down to mix each tube well.

4. Incubate both tubes 45 min at 37°C.

5. To each tube, add 1/25 volume 50 mM EDTA, 1/100 volume 20 μg/μL glycogen, 1/10 volume 3 M NaCl and 3 volumes 100% EtOH. Incubate at least 20 min at −70°C.

6. To pellet RNA, centrifuge 15 min at ≥ 12,000× *g*, 4°C.

7. Carefully remove supernatant. Dissolve each pellet in 9 volumes 0.5× TE and divide into 9 μL (1 pmol RNA of interest) aliquots.

   Store RNA at −70°C.

Primer Extension

8. Thaw one 1 pmol aliquot each of treated and mock-treated RNA on ice.

9. Adjust *in vitro* transcribed RNA of interest to 0.111 pmol/μL in 0.5× TE (for sequencing/alignment reactions).

10. Distribute 9 μL RNA to eight PCR tubes as follows:

Tube #	RNA
1	treated (+)
2	mock-treated (−)
3–8	*in vitro* transcribed

11. Incubate 3 min at 95°C, then place tubes on ice.

    This step denatures the template RNA.

12. Add 2 μL 1 μM oligo to tubes as follows:

Tube #	Oligo
1–4	5'-tagged with fluorophore 1
5–8	5'-tagged with fluorophore 2

    All fluorophores are light-sensitive and can become degraded by multiple freeze-thaw cycles. Small aliquots should be stored in the dark. In this method, reactions 1–4 use fluorophore 1 and represent the

four reactions (treated, mock-treated, and 2 different sequencing reactions) of a typical four fluorophore SHAPE experiment. Each of these is combined with an identical sequencing reaction (5–8, using fluorophore 2), which is used only to align the traces during data processing.

13. Incubate 5 min at 65°C, then 10 min at 35°C, then place tubes on ice.

This step anneals the RT primer.

14. Prepare 3 new tubes with RT mixes ("+/−", "A", and "C") as follows:

The volumes below allow for 10% extra to account for pipetting error.

Reagent	+/−	A	C
5× First-Strand Buffer	8.8	22.0	4.4
0.1 M DTT	2.2	5.5	1.1
dNTP Mix	2.2	5.5	1.1
10 mM ddATP	-	5.5	-
10 mM ddCTP	-	-	1.1
Nuclease-free water	4.4	5.5	1.1
SuperScript III	2.2	5.5	1.1

15. Add 9 µL RT mix to tubes as follows:

Tube #	RT mix
1, 2	+/−
3, 5–8	A
4	C

16. Incubate 15 min at 52°C, then place tubes on ice.

This step allows RT primer extension.

17. Add 2.5 µL 1 M NaOH to each tube.

18. Incubate 15 min at 98°C, then place tubes on ice.

This step degrades the template RNA by alkaline hydrolysis.

19. Add 2.5 µL 1 M HCl and 25 µL nuclease-free water to each tube.

20. Pool reactions in a pairwise manner: 1 & 5, 2 & 6, 3 & 7, and 4 & 8.

21. Add 1 µL 20 µg/mL glycogen, 10 µL 3 M NaOAc (pH 5.2), and 300 µL 100% EtOH.

22. Incubate at least 20 min at −70°C.

23. To pellet RNA, centrifuge 15 min at ≥ 12,000× *g*, 4°C.

24. Remove supernatants. Wash pellets with 1 mL 75% EtOH.

25. Centrifuge 5 min at ≥ 12,000× *g*, 4°C.

26. Repeat steps 24–25.

27. Remove supernatants. Carefully remove residual EtOH with a pipette. Dissolve each pellet in 10 µL HiDi formamide.

28. Perform capillary electrophoresis according a Fragment Analysis protocol. Analyze each of the four pooled reactions in a separate capillary.

    Fragment analysis can be performed on most conventional capillary electrophoresis-based DNA sequencing machines. If you have access to such a machine, refer to the manufacturer's instructions. Alternatively, many commercial or core DNA sequencing facilities will perform this type of analysis upon request at pricing similar to standard sequencing.

ALTERNATE PROTOCOL 1: FOLDING AND MODIFICATION OF *IN VITRO* TRANSCRIBED RNA

There are often technical barriers to obtaining large amounts of the RNA of interest from an *in vivo* source, while *in vitro* transcript is easily and cheaply obtainable. If RNA structures are robust they are likely to form under non-physiological conditions and can be analyzed in this assay. It may also easier to screen mutants this way. The caveat is always that detected structures may not be biologically significant.

Materials

0.5× TE

In vitro transcribed RNA of interest in 0.5× TE

3.3× RNA Folding Buffer

Dimethyl Sulfoxide (DMSO)

32.5 mM *N*-methylisotoic anhydride (NMIA) in DMSO

20 µg/mL glycogen

50 mM EDTA

3 M NaCl

100% EtOH

1. Adjust RNA of interest concentration to 0.167 pmol/µL.

   Once you have determined the concentration in ng/µL of your RNA of interest, you can use its molecular weight to convert to pmol/µL. First calculate the molecular weight of your RNA product in ng/pmol. [M.W. (ng/pmol) = length (nt) × 0.3205 + 0.1590.] Divide concentration (ng/µL) by molecular weight (ng/pmol) to get pmol/µL.

2. Divide RNA into two equal parts. Incubate both tubes 3 min at 95°C, then place on ice.
3. On ice, add 0.5 volume 3.3× RNA Folding Buffer.
4. Incubate both tubes 20 min at 37°C.
5. Keep tubes at 37°C. To one part, add 1/9 volume DMSO. This is the mock-treated control. To the other part, add 1/9 volume 32.5 mM NMIA in DMSO. This is the experimental sample. Mix each tube well.
6. Incubate both tubes 45 min at 37°C.
7. To each tube, add 1/25 volume 50 mM EDTA, 1/100 volume 20 μg/μL glycogen, 1/10 volume 3 M NaCl and 3 volumes 100% EtOH. Incubate at least 20 min at −70°C.
8. To pellet RNA, centrifuge 15 min at ≥ 12,000× *g*, 4°C.
9. Carefully remove supernatant. Dissolve each pellet in 9 volumes 0.5× TE and divide into 9 μL (1 pmol RNA of interest) aliquots.

   Store RNA at −70°C. To analyze by primer extension, follow steps 8–28 of Basic Protocol 2.

BASIC PROTOCOL 3: PROCESSING AND ANALYSIS OF SHAPE DATA

Capillary electrophoresis data are typically returned as .fsa electropherogram files. Processing these data requires both the ShapeFinder software and a spreadsheet program capable of reading text files, such as Microsoft Excel. This protocol covers data processing in ShapeFinder and the normalization of a single experiment. This will enable the reader to determine which nucleotides are flexible or constrained, which can be extremely helpful in confirming or refuting a predicted structure, or in assessing the effect of mutations. Downstream applications, such as the incorporation of normalized SHAPE data as a parameter in the RNA folding algorithm RNAStructure for the *de novo* prediction of RNA structures, are well described elsewhere (for example, see (Low and Weeks, 2010)) and are not covered here.

ShapeFinder is available at: http://bioinfo.boisestate.edu/Downloads/index.html

RNAStructure is available at: http://rna.urmc.rochester.edu/RNAstructure.html

Both programs are Linux-based and available for Mac OS X. RNAStructure is also available for Windows. The different steps involved in processing an electropherogram in ShapeFinder, and the rationales for each step, are described in (Vasa et al., 2008). The ShapeFinder Help Documentation included with the software also contains helpful information about each step.

1. Open each of your four experimental .fsa files in ShapeFinder.

   Each of your four experimental samples from Basic Protocol 2 corresponds to one .fsa file/electropherogram:

File	Reactions (Basic Protocol 2)
A	1 & 5
B	2 & 6
C	3 & 7
D	4 & 8

2. Note how many channels are displayed and which channels correspond to the two fluorophores used in your experiment.

   This information is available by selecting "Channel Attributes" from the "Display Options" pulldown menu. For example, if the electropherogram is from a machine calibrated for the 6-FAM/VIC/NED/PET dye set, Channels 1 and 2 typically correspond to 6-FAM and VIC respectively.

3. Save each .fsa file as a .txt file.

   Each channel becomes a column of values in a tab-delimited text file.

4. Open each .txt file in Microsoft Excel or a similar spreadsheet program.

   Column 1 contains data from Channel 1, Column 2 from Channel 2, etc. If you used VIC as fluorophore 1 and 6-FAM as fluorophore 2 as described in Basic Protocol 2, your experimental data are in column 2 and your alignment data in column 1. If you used different fluorophores, be sure to confirm which channels/columns contain which data.

5. Select the two columns containing your channels of interest from each of your four files. Copy and paste these columns into a single spreadsheet. This sheet should contain eight columns of data.

   Make sure to copy and paste in a consistent order. Our preferred order is:

Source File & Column	Paste Column	Reaction (Basic Protocol 2)	Fluorophore
File A, Column 1	1	5	2
File A, Column 2	2	1	1
File B, Column 1	3	6	2
File B, Column 2	4	2	1
File C, Column 1	5	7	2
File C, Column 2	6	3	1
File D, Column 1	7	8	2
File D, Column 2	8	4	1

   Thus, the even columns contain the two different experimental and two different sequencing reactions that were primed with fluorophore 1-tagged oligo (VIC) and the odd columns contain the four identical sequencing reactions that were primed with fluorophore-2 tagged oligo (6-FAM).

These reactions will be used to align the experimental and sequencing reactions in Step 8–10.

6. Save the file containing all eight columns as a .txt file and open in ShapeFinder. Save the file as a .shape file.

   If you need to pause during Steps 8–10, simply save the .shape file. You can re-open and resume where you left off. Most of the actions described in the following steps (Fitted Baseline Adjust, Cubic Mobility Shift, Scale Factor, Signal Decay Correction, and Align and Integrate) are standard processing steps available from the Tool Inspector menu. The details of each step and the rationales are explained in detail in (Vasa et al., 2008) and the ShapeFinder Help documentation.

7. Perform a Fitted Baseline Adjust on all eight channels using a Window Width of 200.

8. Perform a Cubic Mobility Shift by sliding the traces on Channel 3 to align with Channel 1. "Sliding traces" is a drag and drop process to manually align the electropherograms. After you save and apply the shift, confirm visually that the alignment is good.

   Steps 8–12 represent the additional processing steps necessary for a two (rather than three or four) fluorophore experiment. Because the four experimental samples use the same fluorophore, they were necessarily analyzed in separate capillaries and must aligned to each other before further processing can occur. This alignment depends on the fact that the four sequencing/alignment reactions contain identical peaks and can be easily, if tediously, aligned by hand. The more manual "slides" you perform, the better your alignment will be.

9. Record the values of the constants a, b, c, and d.

10. Perform a Cubic Mobility Shift on Channel 4, by entering the equations by hand, using the constants obtained from shifting Channel 3 to Channel 1. After you save and apply the shift, it should align the peaks in Channel 4 with the peaks in Channel 2, although the heights of the peaks will be different. Confirm visually that the alignment is good.

11. Repeat steps 8–10 twice to align Channel 5, then 7, to Channel 1 by sliding traces, and use the constant values to align Channel 6, then 8, to Channel 2.

12. Save the fully-aligned file both .shape and .txt files.

13. Open the aligned .txt file and delete Columns 1, 3, 5, and 7. Save as a .txt file.

    This removes the data from the four identical sequencing reactions that were primed with fluorophore-2 tagged oligo. Once the four channels of interest have been aligned, these are no longer needed.

14. Open the new .txt file in ShapeFinder and save as a .shape file.

15. Perform a Signal Decay Correction on each of the four channels. Use the same Region of Interest for all four channels. Determine the region of interest based on the channel corresponding to the treated RNA (typically Channel 1) and apply the same region to all four channels (typically from between 1000 and 2000 to between 5500 and 7000). The rescale factor and equation parameters, A, q, and C, should be kept at their default values of 10000, 1000000, 0.999, and 10000, respectively.

    You may need to try several different regions on the first trace before finding one that works well.

16. Apply a Scale Factor to the channel corresponding to the mock-treated RNA (typically Channel 2) such that most peaks are of equal or lesser height than the corresponding treated RNA (Channel 1) peak. This factor is typically between 0.3 and 0.7.

    It is not necessary to scale the sequencing channels (typically Channels 3 and 4) but if the peaks are dramatically larger or smaller than the experimental peaks, it may make the Alignment and Integration process easier if you scale them to a similar size.

17. Align and Integrate:

    a. In the Setup stage, indicate which Channels correspond to which data (typically +Reagent, Channel 1; -Reagent, Channel 2; ddATP, Channel 3, ddCTP, Channel 4). Set a trace range equal to or narrower than that used for signal decay correction and identify a text file containing the RNA sequence.

    b. In the Modify stage, manually inspect and correct all peaks.

        At this stage, if there are missing or extra peaks, parts of the trace will be misaligned. It is easiest to find a region (usually in the middle) where the alignment is correct and work outwards in both directions from there. This process is a little fiddly and frustrating at first, but becomes easier with practice.

    c. In the Fit stage, the ShapeFinder software performs a whole-channel Gaussian integration to quantify all individual peak areas, and the mock-treated (Channel 2) peak areas are subtracted from the treated (Channel 1).

18. There are two output files after Integration: filename_myfit.txt and filename_mypeaks.txt. The filename_mypeaks.txt file contains your SHAPE reactivities in the (RX.area-BG.area) column (furthest to the right).

19. Normalize the raw reactivity data using either a model-free box plot analysis or the "2–8% rule" as described in (Deigan et al., 2009; McGinnis et al., 2009). This normalization is typically performed in Microsoft Excel or by using scripts written for the purpose. Set normalized reactivity values between −0.5 and 0 to zero. Discard as extreme outliers normalized reactivity values of less than −0.5 or greater than 3.0.

This normalization assumes that at least 10% of nucleotides in a given dataset are single-stranded and therefore reactive. If this is not a valid assumption for your data, you may wish to explore other methods of normalization. The extreme outliers often represent nucleotides with strong signal in the mock-treated channel, for which signal decay correction and scaling are imperfect, or else nucleotides that are constrained in a highly reactive conformation. It can be helpful to take note of them when examining an individual structure, however they tend to distort further analysis and typically represent < 1 % of all data, so discarding them for some purposes can be more helpful than harmful.

20. Your output after normalization will be a normalized SHAPE reactivity value for each nucleotide analyzed. A SHAPE reactivity less than 0.3 typically represents a base-paired or otherwise structured/constrained nucleotide. Reactivities greater than 0.7 typically represent flexible nucleotides. Intermediate values can represent moderately constrained nucleotides (e.g. the first or last base-pair of a stem), but should be interpreted with caution. In addition to examining the reactivity of individual nucleotides, SHAPE data sets can be used to bias RNA folding algorithms and predict structures *de novo*, a process which is described elsewhere (Deigan et al., 2009; Low and Weeks, 2010).

REAGENTS AND SOLUTIONS

3.3× RNA Folding Buffer

333 mM HEPES (pH 8.0)

20 mM $MgCl_2$

333 mM NaCl

RNA Modification Buffer

50 mM HEPES (pH 8.0)

200 mM KOAc (pH 8.0)

3 mM $MgCl_2$

0.5× TE

5 mM Tris-HCl (pH 8.0)

0.5 mM EDTA (pH 8.0)

COMMENTARY

Background Information

The fundamental methods of chemical and/or enzymatic probing to detect RNA structure were developed in the 1980s (Ehresmann et al., 1987; Peattie and Gilbert, 1980; Stern et al., 1988) (reviewed in (Weeks, 2010)). Essentially, a chemical or enzymatic reagent modifies the RNA. To be useful, this modification must have two characteristics. First, it must

selectively modify nucleotides based on their structural context. For example, DMS methylates adenosine, cytidine, and guanosine residues, but only when they are not "protected" from modification by a base-pairing interaction. Second, the modification must be detectable as a stop to primer extension by reverse transcriptase. The DMS methylations of adenosine and cytidine (but not guanosine) meet this criterion (Tijerina et al., 2007). Traditionally, such foot-printing techniques have involved gel electrophoresis of radioactively labeled cDNA fragments.

Newer methods, such as inline probing (Regulski and Breaker, 2008) and selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) (Merino et al., 2005; Wilkinson et al., 2006), are advantageous in that they interrogate RNA nucleotide flexibility at the sugar-phosphate backbone. Structured nucleotides are still protected, but all four bases can be interrogated in a single experiment. SHAPE further improves on traditional methods by eliminating radioactive sequencing gels in favor of fluorescently tagged oligos and capillary electrophoresis. When the electropherograms are processed computationally using the ShapeFinder software (Vasa et al., 2008), the result is a quantitative measure of structure at single-nucleotide resolution. The process is sufficiently high-throughput to be applied to large RNA molecules. For example, the full-length HIV-1 genome was analyzed by SHAPE (Watts et al., 2009).

While capillary electrophoresis represents a substantial improvement in throughput, the combination of chemical or enzymatic probing with next generation sequencing technologies has the potential to increase throughput exponentially, potentially allowing the analysis of whole populations of RNA molecules rather than a single species (Aviran et al., 2011; Kertesz et al., 2010; Lucks et al., 2011). However these methods still present considerable computational challenges. When a single RNA species, (such as a viral genome) is of interest, such methods are probably excessive.

Computational methods for predicting RNA structures have also improved. Tertiary structures have always been much more difficult to predict than secondary structures for sequences of any length, due to the processing time and computer memory required (Rivas and Eddy, 2000), but improvements are continually forthcoming (Wiebe and Meyer, 2010). These approaches have traditionally been base either on thermodynamic folding algorithms or on comparative sequence analysis, but increasingly these computational approaches can be combined with each other and with experimental probing data to improve structure prediction (reviewed in (Laing and Schlick, 2011; Mathews et al., 2010)). For example, SHAPE reactivity data can be incorporated as a pseudo-free energy change term into the RNA folding algorithm RNAstructure (Low and Weeks, 2010; Reuter and Mathews, 2010).

Critical Parameters and Troubleshooting

These protocols suggest 1 pmol of RNA per primer extension reaction, which usually works well. However, given the variability of capillary electrophoresis equipment, you may wish to try several sequencing reactions with your oligos and *in vitro* transcribed RNA, to determine the minimum amount of RNA needed to give consistent signal. This is especially true if your experimental RNA is scarce and you wish to use as little as possible in each experiment.

Before embarking on a large-scale project, it is critical to ensure that experiments are highly reproducible, giving similar normalized SHAPE reactivity values for the same nucleotide. You should examine both replicate experiments using the same primer and experiments using an up- or downstream primer giving some overlapping data. If the data are not reproducible, examine the raw reactivity values and the unprocessed electropherograms, to determine whether the inconsistency is in the data processing or in the actual experimental results.

If the experimental results are inconsistent, one option is to troubleshoot with *in vitro* transcribed RNA. This will determine whether the SHAPE and primer extension reactions are working well. If this is the case, the problem likely lies in your source RNA, which can be more difficult to correct and is specific to the RNA of interest.

Anticipated Results

If your RNA of interest contains one or more known structure, it is always informative to compare the average SHAPE reactivity of predicted paired versus predicted unpaired nucleotides - the unpaired nucleotides should be significantly more reactive. Reactivity changes between different conditions can also be informative regarding structural changes. If you are examining an RNA of unknown structure, SHAPE data can be used to bias the folding algorithm RNAstructure, as described by (Low and Weeks, 2010).

Time Considerations

The experiments described here can all be completed relatively quickly. Generating the source RNA may be time-consuming, but non-denaturing extraction (Basic Protocol 1) takes only about 3 hours, including several spins and incubations. SHAPE (Basic Protocol 2 and/or Alternate Protocol 1) similarly takes several hours. Capillary electrophoresis may take longer, but turn around from commercial sequencing facilities is typically less than 24 hours. Data processing in ShapeFinder and normalization for a single experiment takes 1–2 hours for an experienced user, depending on data quality. For novice users, it will take longer at first.

Acknowledgments

We wish to thank members of the Andino laboratory for comments on the manuscripts. This work was supported by NIH (R01, AI36178, AI40085, P01 AI091575, 5T32AI060537) and the University of California (CCADD).

Literature Cited

Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP, Pachter L. Modeling and automation of sequencing-based characterization of RNA structure. Proc Natl Acad Sci U S A. 2011; 108:11069–11074. [PubMed: 21642536]

Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci U S A. 2009; 106:97–102. [PubMed: 19109441]

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. Probing the structure of RNAs in solution. Nucleic Acids Res. 1987; 15:9109–9128. [PubMed: 2446263]

Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. Nature. 2010; 467:103–107. [PubMed: 20811459]

Laing C, Schlick T. Computational approaches to RNA structure prediction, analysis, and design. Curr Opin Struct Biol. 2011; 21:306–318. [PubMed: 21514143]

Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. Methods. 2010; 52:150–158. [PubMed: 20554050]

Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A. 2011; 108:11063–11068. [PubMed: 21642531]

Mathews DH, Moss WN, Turner DH. Folding and finding RNA secondary structure. Cold Spring Harb Perspect Biol. 2010; 2:a003665. [PubMed: 20685845]

McGinnis JL, Duncan CD, Weeks KM. High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. Methods Enzymol. 2009; 468:67–89. [PubMed: 20946765]

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. 2005; 127:4223–4231. [PubMed: 15783204]

Peattie DA, Gilbert W. Chemical probes for higher-order structure in RNA. Proc Natl Acad Sci U S A. 1980; 77:4679–4682. [PubMed: 6159633]

Regulski EE, Breaker RR. In-line probing analysis of riboswitches. Methods Mol Biol. 2008; 419:53–67. [PubMed: 18369975]

Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 2010; 11:129. [PubMed: 20230624]

Rivas E, Eddy SR. The language of RNA: a formal grammar that includes pseudoknots. Bioinformatics. 2000; 16:334–340. [PubMed: 10869031]

Stern S, Moazed D, Noller HF. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. Methods Enzymol. 1988; 164:481–489. [PubMed: 2468070]

Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. Nature protocols. 2007; 2:2608–2623. [PubMed: 17948004]

Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA (New York, N.Y. 2008; 14:1979–1990.

Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009; 460:711–716. [PubMed: 19661910]

Weeks KM. Advances in RNA structure analysis by chemical probing. Curr Opin Struct Biol. 2010; 20:295–304. [PubMed: 20447823]

Wiebe NJ, Meyer IM. TRANSAT-- method for detecting the conserved helices of functional RNA structures, including transient, pseudo-knotted and alternative structures. PLoS Comput Biol. 2010; 6:e1000823. [PubMed: 20589081]

Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature protocols. 2006; 1:1610–1616. [PubMed: 17406453]

Wilkinson KA, Vasa SM, Deigan KE, Mortimer SA, Giddings MC, Weeks KM. Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. RNA (New York, N.Y. 2009; 15:1314–1321.