

Published in final edited form as:

*Biochemistry*. 2015 November 24; 54(46): 6867–6875. doi:10.1021/acs.biochem.5b00977.

# Detection of RNA-protein interactions in living cells with SHAPE

**Matthew J. Smola**[1], **J. Mauro Calabrese**[2], and **Kevin M. Weeks**[1,3]

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

[2]Department of Pharmacology and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

## Abstract

SHAPE-MaP is unique among RNA structure probing strategies in that it both measures flexibility at single-nucleotide resolution and quantifies the uncertainties in these measurements. We report a straightforward analytical framework that incorporates these uncertainties to enable detection of RNA structural differences between any two states, and we use it here to detect RNA-protein interactions in healthy mouse trophoblast stem cells. We validate this approach by analysis of three model cytoplasmic and nuclear ribonucleoprotein complexes, in 2-minute in-cell probing experiments. In contrast, data produced by alternative in-cell SHAPE probing methods correlate poorly ($r = 0.2$) with those generated by SHAPE-MaP and do not yield accurate signals for RNA-protein interactions. We then examine RNA-protein and RNA-substrate interactions in the RNase MRP complex and, by comparing in-cell interaction sites with disease-associated mutations, characterize these non-coding mutations in terms of molecular phenotype. Together, these results reveal that SHAPE-MaP can define true interaction sites and infer RNA functions under native cellular conditions with limited pre-existing knowledge of the proteins or RNAs involved.

## INTRODUCTION

Nearly all RNAs, regardless of function, interact with one or more protein partners in order to function properly[1,2]. Characterizing ribonucleoprotein (RNP) complexes is thus an important step in understanding RNA function. Several well-validated approaches have been developed to explore RNP complexes[3]. These methods provide many valuable insights but often have a limited scope due to affinity purification steps that require prior knowledge about the RNA or protein of interest. As RNA structure studies expand to 'omics scales, direct and accurate approaches for uncovering sites of interaction between the transcriptome and the proteome will become increasingly important.

SHAPE-MaP (selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling) combines well-validated SHAPE RNA structure probing chemistry[4,5] with massively-parallel sequencing to enable high-throughput interrogation of RNA flexibility at single-nucleotide resolution[6,7]. When probed with SHAPE reagents, conformationally

[3]Correspondence, weeks@unc.edu.

Supporting information **available**

Supplemental Figures S1-S6; Primer sequences for RNA-specific SHAPE-MaP.

flexible nucleotides exhibit high reactivity. Conversely, nucleotides constrained by base pairing or by other interactions show low reactivities. The quantitative relationship between SHAPE reactivity and conformational flexibility is maintained even for nucleotides that are not solvent accessible as visualized in static RNPs[5], indicating that SHAPE can be used to probe the interiors of RNA-protein complexes. Previous work has shown that SHAPE reagents readily modify RNAs in living cells[8-13]. Finally, SHAPE-MaP uniquely allows for thorough and quantitative analysis of specific individual RNAs within the contents of an entire transcriptome with the use of targeted primers[6,14]. Thus, SHAPE-MaP offers a broadly useful strategy for probing the structure of the entire transcriptome, or elements thereof, under diverse experimental conditions.

A wide variety of RNA structure probing methods have been proposed[15,16], most of which depend on accurately identifying and quantifying cDNA ends created when reverse transcriptase enzymes encounter a chemical adduct or cleavage site. These methods all involve a critical adapter-ligation step. In principle, these methods make it straightforward to perform RNA structure probing on the entire contents of a given transcriptome; in practice, it is currently almost impossible to perform the adapter-ligation step quantitatively[17,18]. Moreover, transcriptome-wide experiments are strongly subject to the classic multiple and sparse measurement problems such that many measurements are unlikely to be statistically significant[6] and thus do not survive follow-up validation[19]. An important challenge in large-scale and in-cell RNA structure analyses is to robustly detect significant structural changes.

We hypothesized that most RNA-protein interactions would affect the flexibility of nucleotides at the binding site and that by comparing SHAPE reactivities of deproteinized RNA (*ex vivo*) with reactivities obtained by probing RNA in living cells (*in cellulo*), it would be possible to characterize sites of RNP interactions (Fig. 1a). We developed an analysis framework that enables detection of RNP interactions (Supplemental Fig. 1) based upon three principles: (*i*) RNA-protein interactions strongly affect SHAPE reactivity, either positively or negatively; (*ii*) due to measurement errors and the large number of reactivity measurements made, not all apparent reactivity changes are significant; and (*iii*) most RNA-protein interaction sites[20] will span sites of five or more nucleotides in primary sequence.

To identify changes in SHAPE reactivity associated with protein interactions, we used the SHAPE reagent 1-methyl-7-nitroisatoic anhydride (1M7) to generate *in cellulo* and *ex vivo* SHAPE-MaP datasets for U1, 5S, and SRP RNAs (Fig. 1a). These RNAs enable evaluation of RNPs located both in the nucleus and in the cytoplasm and high-resolution structures of their complexes with proteins are available[21-26]. Alternative SHAPE reagents have been proposed for *in cellulo* modification[8,12]. We compared 1M7 SHAPE-MaP with recently published in-cell SHAPE (icSHAPE), which uses a clickable RNA acylation reagent (NAI-N3) to allow enrichment of RNAs modified with this relatively weakly reactive reagent. We found that icSHAPE measurements show very low correlation with those obtained with SHAPE-MaP. Thus, we chose 1M7 for its short half-life, ability to accurately report RNA secondary structure *ex vivo*[4-7,27] and in living cells[9-11], and because in-cell reactivity of 1M7 is sufficiently robust that downstream enrichment is not required.

Differences in SHAPE reactivities (ΔSHAPE) were calculated by subtracting *in cellulo* SHAPE reactivities from *ex vivo* reactivities (Fig. 1b, upper left) and averaging over a three-nucleotide sliding window to reduce local signal fluctuation. By this definition, positive ΔSHAPE values indicate protection from modification in the cellular environment, and negative ΔSHAPE reports enhanced reactivity in cells.

In a SHAPE-MaP experiment, discrete mutation events contribute to the overall reactivity at each nucleotide and are well modeled by a Poisson distribution[6]. The standard error in the SHAPE reactivity measurement can therefore be estimated for every nucleotide[6]. We used these error estimates to perform a modified Z-factor test[6,28] for all positions in a given RNA (Fig. 1b, upper right). This test compares the magnitude of ΔSHAPE with the associated *ex vivo* and *in cellulo* measurement errors, identifying nucleotides for which the magnitudes of the errors are too large for the ΔSHAPE values to be significant. We formulated the Z-factor test such that the underlying *ex vivo* and *in cellulo* SHAPE reactivities must differ by more than 1.96 standard deviations (Z-factor > 0), ensuring that the 95% confidence intervals of each measurement do not overlap.

For many nucleotides, SHAPE-MaP reactivity measurements have very small errors, allowing for the possibility that a trivially small ΔSHAPE could be considered significant according to the Z-factor test. We expected most stable protein-RNA interactions to have a strong effect on the reactivity of nucleotides at the binding site, so we calculated a standard score at each nucleotide to identify the largest ΔSHAPE values (Fig. 1b, lower left). This metric compares ΔSHAPE of a given nucleotide with the ΔSHAPE of all other nucleotides in the RNA, regardless of Z-factor. We required that the absolute value of each standard score be ≥ 1, meaning that individual ΔSHAPE values must be at least one standard deviation away from the mean ΔSHAPE. Thus, only the largest ΔSHAPE values are considered for further analysis. To determine final RNA-protein interaction sites, we filtered by Z-factor and standard score simultaneously (Fig. 1b, lower right). If, in a 5-nucleotide window, at least three nucleotides had a Z-factor > 0 and an absolute standard score ≥ 1, those three (or more) nucleotides were considered to have significant cell-induced changes in SHAPE reactivity.

In this work, we show that biochemical RNA structure probing data generated with the well-validated SHAPE-MaP approach can be used to identify significant, meaningful changes in RNA structure between two states. Here, these states are the RNA in healthy mouse trophoblast stem cells and the same RNAs gently extracted from cells. We validate our approach with the abundant and well-characterized U1 small nuclear RNA (snRNA), 5S rRNA, and signal recognition particle (SRP) RNP complexes, illustrating that the statistical filters implemented in our analysis robustly identify sites of protein interactions. We then examine RNase MRP, an important RNP complex whose in-cell architecture is relatively poorly understood. Our analysis confirms several reported RNA-protein interactions within the complex, and also characterizes the underlying molecular phenotype of many disease-associated mutations.

## EXPERIMENTAL

### *In cellulo* modification

Mouse trophoblast stem cells (TSCs) were cultured as described[29]. Live TSCs were washed once with PBS, and 900 μl of fresh growth media was added. For samples subjected to in-cell SHAPE probing, 100 μl of 100 mM 1M7 in neat DMSO (10 mM final concentration) were added and rapidly mixed by swirling the culture dish. Cells were then incubated at 37 °C for 5 minutes (although the 1M7 reagent is completely quenched by hydrolysis in ~2 minutes). Media was removed and the cells were washed once with PBS before isolation of total RNA (1 mL TRIzol; Ambion). The no-reagent negative control RNA was prepared similarly with the exception that neat DMSO was used instead of 1M7 in DMSO.

### *Ex vivo* RNA extraction and modification

To preserve native secondary structures, RNA for *ex vivo* analysis was extracted using a gentle procedure, avoiding the use of harsh chemical denaturants. Approximately $10^6$ TSCs were washed and pelleted in ice-cold PBS, resuspended in 2.5 ml Lysis Buffer [40 mM Tris, pH 7.9, 25 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 256 mM sucrose, 0.5% Triton X-100, 1,000 U/ml RNasin (Promega), 450 U/ml DNase I (Roche)], and rotated at 4 °C for 5 minutes. Cells were then pelleted at 4 °C for 2 minutes at 2250 *g*, resuspended in 40 mM Tris pH 7.9, 200 mM NaCl, 1.5% SDS, and 500 μg/ml of Proteinase K, and rotated at 20 °C for 45 minutes. RNA was then extracted twice with phenol:chloroform:isoamyl alcohol (24:24:1) pre-equilibrated with 1× Folding Buffer (100 mM HEPES, pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$), followed by one extraction with chloroform. Note that use of TRIzol and similar reagents should be specifically avoided for native-like purification of RNA. RNA was exchanged into 1.1× Folding Buffer using a desalting column (PD-10, GE Life Sciences) and incubated at 37 °C for 20 minutes. Approximately 3 μg RNA was then added to a one-ninth volume of 100 mM 1M7 in neat DMSO (10 mM final concentration) and incubated at 37 °C for 5 minutes. Modified RNA was purified (RNeasy Midi spin column, Qiagen) and eluted in approximately 50 μl $H_2O$. No-reagent negative control RNA was prepared in the same way but substituting neat DMSO for 1M7.

### Denaturing control

TSCs were grown as described[29] and total RNA isolated using TRIzol (Ambion). Approximately 500 ng RNA was then resuspended in 1.1× Denaturing Control Buffer [55 mM HEPES pH 8.0, 4.4 mM EDTA, 55% formamide (v/v)] and incubated at 95 °C for 1 minute. An aliquot of 45 μl of denatured RNA was added to 5 μl of 100 mM 1M7 and allowed to react at 95 °C for 1 minute. After modification, RNA was purified (RNeasy Mini spin column, Qiagen) and eluted in approximately 50 μl $H_2O$.

### U1, SRP, and 5S SHAPE-MaP

Mutational profiling reverse transcription reactions were carried out using RNA-specific primers (Table S1)[6,14], which maximizes efficient use of sequencing reads. cDNA was purified using G-50 spin columns (GE Life Sciences). SHAPE-MaP sequencing libraries were created for each experimental condition (*ex vivo* +1M7, *ex vivo* DMSO, *in cellulo*

+1M7, *in cellulo* DMSO, denaturing control +1M7) and RNA (U1 snRNA, 5S rRNA, SRP RNA) using the targeted specific-RNA approach[6] with minor changes. PCR 1 followed the touchdown format[30] and was performed as follows: 98 °C for 30 s, 20 cycles of [98 °C for 10 s, 72 °C for 30 s (decreasing by 1 °C per cycle until 64 °C), 72 °C for 20 s], 72 °C for 2 min. PCR 2 was performed for 10 cycles using 2 μl unpurified PCR 1 product as template in a 50 μl reaction. Final libraries were then purified (PureLink PCR Micro spin columns; Life Technologies) prior to sequencing.

### Whole-transcriptome SHAPE-MaP

Total RNA was modified as described above and then depleted of ribosomal RNA (mouse RiboZero; Epicentre). Mutational profiling reverse transcription reactions were primed with random DNA nonamers[6,14]. cDNA was purified (Agencourt RNAClean XP beads, Beckman Coulter) and then converted to double-stranded DNA (NEBNext mRNA second-strand synthesis kit, New England Biolabs). The resulting DNA was purified (Ampure XP beads, Beckman Coulter) before construction of whole-transcriptome sequencing libraries (Nextera XT, Illumina).

### Sequencing and SHAPE profile generation

Purified U1, 5S, SRP, or whole-transcriptome sequencing libraries were sequenced on an Illumina MiSeq (U1, 5S, and SRP) or NextSeq (transcriptome) instrument, generating 2 × 150 paired-end datasets. Initial SHAPE reactivity profiles, including error estimates, were created by aligning reads to U1 snRNA, 5S rRNA, SRP or RNase MRP RNA reference sequences (GenBank accession no. FM991912.1, M31319.1, HG323689.1, and NR_001460.1, respectively) using *ShapeMapper* (v1.0, http://chem.unc.edu/rna/software.html)[6,14]. Median per-nucleotide read depth was greater than 8,500, 153,000, 100,000, and 17,000 for U1, 5S, SRP, and RMRP, respectively.

From transcriptome-wide datasets, we identified the 50 most abundant transcripts using *Tophat*[31]. SHAPE reactivity profiles were then generated for each of these RNAs by aligning to respective sequences with *ShapeMapper*. Transcripts with complete sequencing coverage and sufficient depth (median read depth > 5,000) were selected for comparison to icSHAPE profiles.

### SHAPE reactivity normalization

SHAPE-MaP quantifies adduct formation based on the observed mutation rates of modified RNA relative to no-reagent and denaturing controls[6]. We observed higher mutation rates in *ex vivo*-modified U1 snRNA than in the 5S and SRP RNAs (Supplemental Fig. 2). As a result, the SHAPE reactivities of U1 snRNA were generally elevated compared to the other RNAs. Independent normalization of U1 snRNA did not preserve the intrinsically high reactivity of this RNA relative to 5S and SRP. Thus, we normalized SHAPE reactivities with a common normalization factor to preserve the relative distribution of reactivities among the three simultaneously probed RNAs. RNase MRP SHAPE profiles were normalized independently, as they were derived from RNA probed separately from the three model RNAs. Initial SHAPE reactivities for both *in cellulo* and *ex vivo*-modified RNAs were first pooled together into a single distribution from which primer-binding sites were excluded.

The first five nucleotides synthesized during reverse transcription were also excluded to eliminate spurious mutations caused by the suboptimal processivity of initiating retroviral reverse transcriptase[32]. A normalization factor for the entire distribution was calculated by the boxplot method[27]: the interquartile range (IQR) of the distribution was calculated; and reactivity values greater than 1.5 times the IQR were excluded as outliers with the number of outliers capped at 10%. The average of the 10% most reactive remaining nucleotides was then calculated, yielding the common normalization factor. Initial individual SHAPE profiles were then adjusted by dividing each reactivity and standard error by the common normalization factor.

## icSHAPE profile generation

icSHAPE reads[12] were downloaded from the gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo, accession GSE60034). Reads corresponding to U1 snRNA, 5S rRNA, SRP and RNase MRP RNA were extracted by alignment to the respective sequence. Relevant reads were then converted to fastq format and analyzed using the published icSHAPE pipeline (https://github.com/qczhang/icSHAPE)[12]. Limited reads for U1 snRNA, 5S rRNA, and RNase MRP RNA resulted in icSHAPE profiles with unanalyzably sparse data, so we restricted our comparison of RNP complexes to the SRP RNA.

## Calculating standard error, ΔSHAPE, Z-factors, and standard scores to determine binding sites

The derivation of nucleotide-resolution standard error values associated with SHAPE reactivity measurements has been described fully[6], and is reviewed briefly here. Mutation rates for each experimental measurement (+1M7, no-reagent, denaturing control) are modeled as a Poisson distribution because discrete mutation events contribute to the overall reactivity at each nucleotide. The variance of a Poisson distribution equals the number of observations, and the standard error of a mutation rate ($SE_{rate}$) can be estimated as

$$SE_{rate}=\frac{\sqrt{\lambda}}{\text{reads}}=\frac{\sqrt{\text{rate}}}{\sqrt{\text{reads}}} \quad (1)$$

where λ is the number of mutations observed, *reads* is the read depth at a given nucleotide, and *rate* is the number of mutation events per read. The standard errors from each experimental measurement are then combined to yield SHAPE reactivity standard errors[6].

The change in SHAPE reactivity (ΔSHAPE) for each nucleotide *i* was calculated as

$$\Delta\text{SHAPE}_i=\frac{1}{3}\left[\left(\sum_{n=i-1}^{i+1}X_n\right)-\left(\sum_{n=i-1}^{i+1}C_n\right)\right] \quad (2)$$

where *X* and *C* are the *ex vivo* and *in cellulo* SHAPE reactivities, respectively. This produces ΔSHAPE values that reflect the difference in reactivity between *ex vivo* and *in*

*cellulo* conditions averaged over a three-nucleotide sliding window. To account for smoothing, standard error values were averaged as

$$SE_i = \frac{1}{3}\sqrt{\sigma_{i-1}^2 + \sigma_i^2 + \sigma_{i+1}^2} \quad (3)$$

where $\sigma_i$ and $SE_i$ refer to the original error and smoothed error at nucleotide *i*, respectively. Z-factors (*Z*)[28] for each nucleotide *i* were calculated according to Eqn. 4, where the subscripts *X* and *C* indicate *ex vivo* and *in cellulo* conditions, respectively. Nucleotides for which $Z > 0$ were considered to undergo significant changes in SHAPE reactivity.

$$Z_i = 1 - \frac{1.96\left(SE_{X,i} + SE_{C,i}\right)}{|\Delta\mathrm{SHAPE}_i|} \quad (4)$$

Standard scores (*S*) were calculated for each nucleotide *i* according to Eqn. 5, where $\mu_{\Delta SHAPE}$ and $\sigma_{\Delta SHAPE}$ represent the mean and standard deviation of the distribution of ΔSHAPE values, respectively.

$$S_i = \frac{\Delta\mathrm{SHAPE}_i - \mu_{\Delta\mathrm{SHAPE}}}{\sigma_{\Delta\mathrm{SHAPE}}} \quad (5)$$

Putative binding sites were identified as regions within five-nucleotide sliding windows for which at least three nucleotides had $Z > 0$ and $|S| \geq 1$. Nucleotides that met these requirements were denoted as undergoing changes in SHAPE reactivity due to the influence of the cellular environment.

### Modeling

The model of the complete U1 snRNP complex used in this study was generated from three individual models. Phosphorus atoms in a U1 snRNP model (omitting stem-loop 2 and the U1A protein and kindly provided by Kiyoshi Nagai) were first aligned to the phosphorus atoms in a 5.5-Å model of the complete complex (PDB: 3PGW)[23]. To incorporate the U1A/stem-loop 2 interaction, we aligned the Cα atoms of U1A in a high-resolution model (PDB: 4PKD)[33] to the 5.5-Å model. The model of the SRP S domain bound to SRP68/72, SRP19, and SRP54 was generated by overlaying the SRP68- and SRP19-bound structure (PDB: 4P3E)[24] with the SRP19- and SRP54-bound structure (PDB: 1MFQ)[22] via alignment of the SRP19 atoms.

## RESULTS

### Comparison of SHAPE-MaP and icSHAPE

We compared the similarity of RNA structure probing data for SHAPE-MaP and icSHAPE[12] experiments using the SRP RNP complex and six mRNAs. When probing SRP *ex vivo*, we found that strong icSHAPE signals are generally indicative of flexible nucleotides

(Supplemental Fig. S3), as expected for SHAPE reagents. However, in comparison to SHAPE-MaP, the icSHAPE results appear roughly binary, with relatively few intermediate reactivity values. In comparing *in cellulo* data, SHAPE-MaP and icSHAPE data show very poor correlations. The differences between *ex vivo* and *in cellulo* icSHAPE values exhibit strong, punctate positive values throughout the RNA and dramatically strong negative values near the 5′ end (Supplemental Fig. S3). The icSHAPE data would suggest that the SRP RNA undergoes extreme and widespread conformational changes in cells, which is not consistent with prior work on this RNA[24,25]. Further *in cellulo* comparison of six mRNAs produced similar results; the correlation between icSHAPE and SHAPE-MaP was consistently poor, with correlation coefficients ranging from 0.1-0.3 (Supplemental Fig. S3). This analysis suggests that icSHAPE does not measure the same features of RNA as does SHAPE-MaP and that icSHAPE does not accurately measure in-cell RNA structure.

## Validation of the ΔSHAPE approach

We used SHAPE-MaP to analyze three model RNAs *ex vivo* and *in cellulo*. The U1 snRNA is localized in the nucleus and forms the U1 snRNP complex upon binding several proteins: U1A, U1C, U1-70K, and the heteroheptameric Sm ring. Comparison of U1 snRNA *ex vivo* and *in cellulo* SHAPE reactivities revealed distinct qualitative reactivity differences throughout the RNA (Fig. 2a). Due to differences in the number of individual mutation events observed relative to the times a given nucleotide was sequenced, the estimated errors vary as a function of nucleotide position and are greater for some reactivity measurements than others. This is a feature shared by all RNA structure probing experiments read out by massively parallel sequencing but is explicitly and uniquely measured using the MaP strategy. If a naïve approach had been taken that ignored these errors, multiple regions would have been (incorrectly) identified as having significant SHAPE reactivity differences (Fig. 2b, grey and green shading). Only a subset of these regions are involved in true RNA-protein interactions; the remainder are analysis artifacts caused by the measurement uncertainties that occur in any experiment, especially those read out by massively parallel sequencing. When we applied the complete analysis framework in which the Z-factor test is used to account for these errors, only three regions of significant ΔSHAPE were identified (Fig. 2b, green shading only). The locations of these positive ΔSHAPE values correspond precisely to known interactions sites of U1-70K, U1A, and the Sm ring proteins (Fig. 2c and Supplemental Fig. 4).

We next examined the differences in reactivities of the SRP RNA *ex vivo* versus *in cellulo* (Fig. 3a). The SRP RNA associates with six proteins and is comprised of an Alu domain at the 5′ end connected by a long central helix to the S domain. The Alu domain is bound by the SRP9-SRP14 (SRP9/14) heterodimer, and the larger S domain interacts with SRP19, SRP54, and the SRP68-SRP72 (SRP68/72) heterodimer. The SHAPE reactivity changes identified by our analysis were largely localized to these two domains (Supplemental Fig. 5b), consistent with a lack of protein binding in the central helix.

In the Alu domain, we observed *in cellulo* protection at the SRP9/14 binding site (nts 24-26). We also detected enhanced *in cellulo* reactivity at nucleotides 35-37 and 46-48, consistent with protein-induced tertiary structure changes (Fig. 3b). In the S domain, we

observed extensive *in cellulo* protection where SRP19 and SRP54 bind (Fig. 3c). Binding by SRP68/72 involves insertion of an α–helix into the major groove of the central helix, causing an adjacent asymmetric internal loop to open[24]. Consistent with this observation, we detect enhanced *in cellulo* reactivity on the opened side of this loop at positions 230-232 (Fig. 3c). The interaction between SRP RNA and the complete SRP68/72 heterodimer has not been characterized at high resolution; however, cryo-electron microscopy data provide evidence that a portion of SRP68/72 interacts with the central helix at an internal "hinge" loop comprised of nucleotides 97-104 and 249-253[25]. In-cell SHAPE supports this observation, as enhanced *in cellulo* reactivity was noted on both sides of the loop at nucleotides 99-101 and 251-253, and suggests a local conformational change also occurs at nucleotides 230-232 (Supplementary Fig. 5b). Overall, every region of significant *in cellulo* protection in the SRP RNA identified by our analysis framework corresponds to sites of direct protein binding.

In examining the 5S rRNA, which forms a complex with ribosomal protein L5, we detected several regions of *in cellulo* protection (Fig. 4a). These sites correspond to previously identified contacts between 5S rRNA and L5 (Fig. 4b)[8,34]. There were no other sites with significant ΔSHAPE values, although many ribosomal proteins are known to be located near the 5S particle in fully assembled ribosomes. These results are consistent with the observations that a significant fraction of cellular 5S RNPs are not ribosome-associated[35] and that 5S rRNA adopts multiple conformations even when associated with the ribosome[36]. We infer that ΔSHAPE analysis primarily detects only the stable protein-RNA interactions in the 5S rRNA, and that these involve L5.

## Application of ΔSHAPE to RNase MRP

We next applied the ΔSHAPE analysis framework to in-cell analysis of the RNA component of mouse RNase MRP (RMRP). This RNA forms a complex with 10 proteins in eukaryotes that functions in rRNA processing and mitochondrial replication[37]. In humans, numerous mutations within RMRP RNA cause a spectrum of autosomal recessive skeletal diseases ranging from cartilage-hair hypoplasia (CHH) to anauxetic dysplasia (AD)[38]. The structure of and protein interactions with the RNA component of RMRP have been investigated *in vitro* using affinity selection, chemical probing, and crosslinking experiments[37,39-41]. A recent cryo-EM study has revealed the overall three-dimensional architecture of the complex in yeast[42]. However, the precise binding sites of proteins and interactions with substrate have not been examined natively in cells.

Multiple regions of the RMRP RNA have statistically significant enhanced reactivity or protection *in cellulo* (Fig. 5a) and many of these can be attributed to interaction with protein components. These include in the P3 domain, a functionally critical element (Fig. 5b)[43], as well as nucleotides near the junction of helices P8, P9, and P12. Cryo-EM data suggest this latter region interacts with protein Pop4 and perhaps additional proteins (Fig. 5c). We also observed enhanced reactivity at internal loops in helix P12. Although the complete P12 helix is not present in the cryo-EM model, its proximity to the Pop3 protein suggests that the reactivity enhancements located in the P12 helix may be due to conformational changes induced by Pop3 (Supplemental Fig. 6).

We also observed protections involving helices P2 and P19 that are not attributable to RNA-protein interactions. In the cryo-EM model of RMRP, these two regions are adjacent to the active site and are oriented such that they may stabilize or direct RMRP substrates to the catalytic center (Fig. 5d). Additional density in the cryo-EM map adjacent to these sites of protection may reflect RMRP substrates co-purified with the complex, and supports the hypothesis that P2 and P19 play roles in substrate recognition. There is notable overlap between ΔSHAPE-detected protection in P2 and P19 and sites of disease-associated mutations in RMRP (Supplemental Fig. 6)[44]. The substantial level of in-cell protection in the RMRP active site cleft suggests that this RNP enzyme is saturated with its RNA substrates[37] in the cellular steady state.

## DISCUSSION

Our experiments with the well-characterized U1, SRP, and 5S RNPs validate the ability of the ΔSHAPE analytical framework (Fig. 1), enabled by SHAPE-MaP, to correctly and specifically identify regions of RNA protected by stably-associated proteins *in cellulo*, even in the context of a large number of individual measurements and variable level of confidence in each. In addition, this work illustrates the robust ability of the well-validated 1M7 reagent to react with RNP complexes located in both cytoplasmic and nuclear compartments in cells.

In comparing SHAPE-MaP with icSHAPE, we found poor agreement between the two approaches. SHAPE-MaP has previously been extensively validated against a large set of RNAs with complex structures[6], suggesting that icSHAPE does not provide a robust view of RNA structure *ex vivo* or *in cellulo* (Supplemental Fig. 2). icSHAPE also reports that the SRP RNA undergoes extensive internal conformational changes in cells, which is not consistent with prior studies of this RNA[24,25]. icSHAPE differs from SHAPE-MaP in important ways. First, NAI-N3 reacts more slowly than 1M7 ($t_{1/2}$ = ~30 min vs. ~17 sec, respectively), which has important consequences. These include, first, that slow (but not faster) reagents are highly sensitive to specific ion and buffer choices[45] making it very difficult to compare in-cell and *ex vivo* experiments and, second, that long reaction times will reflect RNP assembly and disassembly, cellular turnover, and other events unrelated to the steady-state structure of an RNA. icSHAPE is also one of the many proposed strategies that require a complex purification procedure followed by multi-step adapter-ligation sequencing library construction, steps that are difficult to perform quantitatively[17,18].

In addition to defining in-cell RNA-protein and RNA-substrate interactions, we investigated whether ΔSHAPE analysis might enable categorization of disease-associated mutations in terms of their likely phenotypic effects (Fig. 5a). Our analysis supports the interpretation that most mutations leading to CHH/AD spectrum diseases in the RNase MRP complex result from misfolding of the RNA secondary or tertiary structure, as they are not located near protein or substrate interaction sites. These structural changes occur in helices P1, P3, P4, P9, and P12 (Fig. 5a, blue nucleotides). We also identified a subset of CHH/AD-related mutations located near protein interaction sites (Fig. 5a, in red). In individuals with these mutations, which are most concentrated within helix P8 and the P8-P9-P12 junction, improper assembly of the RNase MRP RNA-protein complex may be the root cause of disease. Finally, the remaining disease-related mutations are most consistent with

compromising RNA-substrate interactions. These involve nucleotides that comprise the active site along with portions of P2 and P19 that are protected *in cellulo* due to putative substrate interactions (Fig. 5a, yellow).

The ΔSHAPE analysis framework is clearly a broadly useful tool for defining RNA-protein interactions. ΔSHAPE is also subject to limitations. Because ΔSHAPE requires a change in SHAPE reactivity between conditions, proteins that interact primarily with double-stranded RNA may be difficult to detect. For the RNAs studied here, in-cell protections almost always corresponded to direct protein-RNA interactions, while enhancements generally reported RNA conformational changes. In other cases, protein-induced conformational changes may lead to apparent protections in regions unrelated to protein binding. While the ΔSHAPE framework correctly identified sites of stable RNA-protein interaction, the stringency implemented here may lead to missing weaker protein binding sites. For example, nucleotides stably bound by Sm ring proteins are detected by ΔSHAPE (Fig. 2) but other nucleotides inside the Sm ring do not display protection. Finally, as with any chemical probing experiment, ΔSHAPE requires sufficient sequencing coverage of the RNA of interest in both tested conditions.

In sum, SHAPE-MaP efficiently and accurately detects RNA-protein interaction sites and occupancy in living cells. Using simple and intuitive statistical filtering, significant differences between *ex vivo* and *in cellulo* SHAPE reactivities were identified while avoiding false positive detection. The analysis framework developed here identified RNA binding sites for all stably bound protein factors for three model RNPs, found in both cytoplasmic and nuclear compartments, under native growth conditions without the need for specialized affinity purification. Application to the RNase MRP ribonucleoprotein enzyme complex both identified sites of RNA-protein interaction and extensive substrate recognition in the active site cleft, and also facilitated categorization of CHH/AD-related mutations by molecular phenotype.

This analysis framework works well for *de novo* identification of functionally essential regions in non-coding RNAs, and is complementary to RNA-protein crosslinking and immunoprecipitation (CLIP)[3] experiments. Critically, ΔSHAPE specifically detects the occupancy of a given site. As RNA structure studies increasingly shift towards in-cell and transcriptome-wide analyses, the robust analytical approach presented here will become an essential tool for rapid discovery and analysis of true RNA-protein interactions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We are indebted to Chris Oubridge and Kiyoshi Nagai for sharing coordinates of their refined U1 snRNP model, and to Bettina Böttcher for providing the model of the RNase MRP complex. We are also grateful to members of the Weeks laboratory for thoughtful discussions regarding the analyses presented here.

**Funding sources**

This work was supported by the National Science Foundation (MCB-1121024 to K.M.W.). M.J.S. is a Graduate Research Fellow of the National Science Foundation (DGE-1144081) and was supported in part by an NIH training grant in molecular and cellular biophysics (T32 GM08570).

## References

(1). Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. 2010; 11:75–87. [PubMed: 20019688]

(2). Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012; 482:339–346. [PubMed: 22337053]

(3). McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol. 2014; 15:203. [PubMed: 24467948]

(4). 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]

(5). Mcginnis JL, Dunkle JA, Cate JHD, Weeks KM. The mechanisms of RNA SHAPE chemistry. J. Am. Chem. Soc. 2012; 134:6617–6624. [PubMed: 22475022]

(6). Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat. Methods. 2014; 11:959–965. [PubMed: 25028896]

(7). Mauger DM, Golden M, Yamane D, Williford S, Lemon SM, Martin DP, Weeks KM. Functionally conserved architecture of hepatitis C virus RNA genomes. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:3692–3697. [PubMed: 25775547]

(8). Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. RNA SHAPE analysis in living cells. Nat. Chem. Biol. 2012; 9:18–20. [PubMed: 23178934]

(9). McGinnis JL, Weeks KM. Ribosome RNA assembly intermediates visualized in living cells. Biochemistry. 2014; 53:3237–3247. [PubMed: 24818530]

(10). Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ. The cellular environment stabilizes adenine riboswitch RNA structure. Biochemistry. 2013; 52:8777–8785. [PubMed: 24215455]

(11). McGinnis JL, Liu Q, Lavender CA, Devaraj A, McClory SP, Fredrick K, Weeks KM. In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. Proc. Natl. Acad. Sci. U.S.A. 2015; 112:2425–2430. [PubMed: 25675474]

(12). Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung J-W, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Chang HY. Structural imprints in vivo decode RNA regulatory mechanisms. Nature. 2015; 519:486–490. [PubMed: 25799993]

(13). Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. 2015 [Epub ahead of print] doi: 10.1093/nar/gkv879.

(14). 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. 2015; 10:1643–1669. [PubMed: 26426499]

(15). Mortimer SA, Kidwell MA, Doudna JA. Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genet. 2014; 15:469–479. [PubMed: 24821474]

(16). Kwok CK, Tang Y, Assmann SM, Bevilacqua PC. The RNA structurome:transcriptome-wide structure probingwith next-generation sequencing. Trends Biochem. Sci. 2015; 40:221–232. [PubMed: 25797096]

(17). Raabe CA, Tang T-H, Brosius J, Rozhdestvensky TS. Biases in small RNA deep sequencing data. Nucleic Acids Res. 2014; 42:1414–1426. [PubMed: 24198247]

(18). Weeks KM. Review toward all RNA structures, concisely. Biopolymers. 2015; 103:438–448. [PubMed: 25546503]

(19). Corley M, Solem A, Qu K, Chang HY, Laederach A. Detecting riboSNitches with RNA folding algorithms: a genome-wide benchmark. Nucleic Acids Res. 2015; 43:1859–1868. [PubMed: 25618847]

(20). Draper DE. Themes in RNA-protein recognition. J. Mol. Biol. 1999; 293:255–270. [PubMed: 10550207]

(21). Weichenrieder O, Wild K, Strub K, Cusack S. Structure and assembly of the Alu domain of the mammalian signal recognition particle. Nature. 2000; 408:167–173. [PubMed: 11089964]

(22). Kuglstatter A, Oubridge C, Nagai K. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. Nat. Struct. Biol. 2002; 9:740–744. [PubMed: 12244299]

(23). Krummel DAP, Oubridge C, Leung AKW, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. 2009; 458:475–480. [PubMed: 19325628]

(24). Grotwinkel JT, Wild K, Segnitz B, Sinning I. SRP RNA remodeling by SRP68 explains its role in protein translocation. Science. 2014; 344:101–104. [PubMed: 24700861]

(25). Halic M, Becker T, Pool MR, Spahn C, Grassucci RA, Frank J, Beckmann R. Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature. 2004; 427:808–814. [PubMed: 14985753]

(26). Voorhees RM, Fernández IS, Scheres SHW, Hegde RS. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. Cell. 2014; 157:1632–1643. [PubMed: 24930395]

(27). Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc. Natl. Acad. Sci. U.S.A. 2013; 110:5498–5503. [PubMed: 23503844]

(28). Zhang J, Chung T, Oldenburg K. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J. Biomol. Screen. 1999; 4:67–73. [PubMed: 10838414]

(29). Quinn, J.; Kunath, T.; Rossant, J. Placenta and Trophoblast. Humana Press; New Jersey: 2005. Mouse Trophoblast Stem Cells; p. 123-146.

(30). Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. “Touchdown” PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 1991; 19:4008. [PubMed: 1861999]

(31). Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L, Trapnell C. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 2012; 7:562–578. [PubMed: 22383036]

(32). Majumdar C, Abbotts J, Broder S, Wilson SH. Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition. J. Biol. Chem. 1988; 263:15657–15665. [PubMed: 2459125]

(33). Kondo Y, Oubridge C, van Roon A-MM, Nagai K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. Elife. 2015; 4

(34). Scripture JB, Huber PW. Binding site for Xenopus ribosomal protein L5 and accompanying structural changes in 5S rRNA. Biochemistry. 2011; 50:3827–3839. [PubMed: 21446704]

(35). Steitz JA, Berg C, Hendrick JP, La Branche-Chabot H, Metspalu A, Rinke J, Yario T. A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. J. Cell Biol. 1988; 106:545–556. [PubMed: 3279045]

(36). Leidig C, Thoms M, Holdermann I, Bradatsch B, Berninghausen O, Bange G, Sinning I, Hurt E, Beckmann R. 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. Nat. Commun. 2014; 5:1–8.

(37). Esakova O, Krasilnikov AS. Of proteins and RNA: the RNase P/MRP family. RNA. 2010; 16:1725–1747. [PubMed: 20627997]

(38). Mattijssen S, Welting TJM, Pruijn GJM. RNase MRP and disease. WIREs RNA. 2010; 1:102–116. [PubMed: 21956908]

(39). Pluk H, van Eenennaam H, Rutjes SA, Pruijn GJ, van Venrooij WJ. RNA-protein interactions in the human RNase MRP ribonucleoprotein complex. RNA. 1999; 5:512–524. [PubMed: 10199568]

(40). Welting TJM, van Venrooij WJ, Pruijn GJM. Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex. Nucleic Acids Res. 2004; 32:2138–2146. [PubMed: 15096576]

(41). Khanova E, Esakova O, Perederina A, Berezin I, Krasilnikov AS. Structural organizations of yeast RNase P and RNase MRP holoenzymes as revealed by UV-crosslinking studies of RNA-protein interactions. RNA. 2012; 18:720–728. [PubMed: 22332141]

(42). Hipp K, Galani K, Batisse C, Prinz S, Bottcher B. Modular architecture of eukaryotic RNase P and RNase MRP revealed by electron microscopy. Nucleic Acids Res. 2012; 40:3275–3288. [PubMed: 22167472]

(43). Shadel GS, Buckenmeyer GA, Clayton DA, Schmitt ME. Mutational analysis of the RNA component of Saccharomyces cerevisiae RNase MRP reveals distinct nuclear phenotypes. Gene. 2000; 245:175–184. [PubMed: 10713458]

(44). Thiel CT, Mortier G, Kaitila I, Reis A, Rauch A. Type and Level of RMRP Functional Impairment Predicts Phenotype in the Cartilage Hair Hypoplasia–Anauxetic Dysplasia Spectrum. Am. J. Hum. Genet. 2007; 81:519–529. [PubMed: 17701897]

(45). Mortimer SA, Weeks KM. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J. Am. Chem. Soc. 2007; 129:4144–4145. [PubMed: 17367143]

(46). Schmitt ME, Bennett JL, Dairaghi DJ, Clayton DA. Secondary structure of RNase MRP RNA as predicted by phylogenetic comparison. FASEB J. 1993; 7:208–213. [PubMed: 7678563]

(47). Perederina A, Esakova O, Quan C, Khanova E, Krasilnikov AS. Eukaryotic ribonucleases P/MRP: the crystal structure of the P3 domain. EMBO J. 2010; 29:761–769. [PubMed: 20075859]

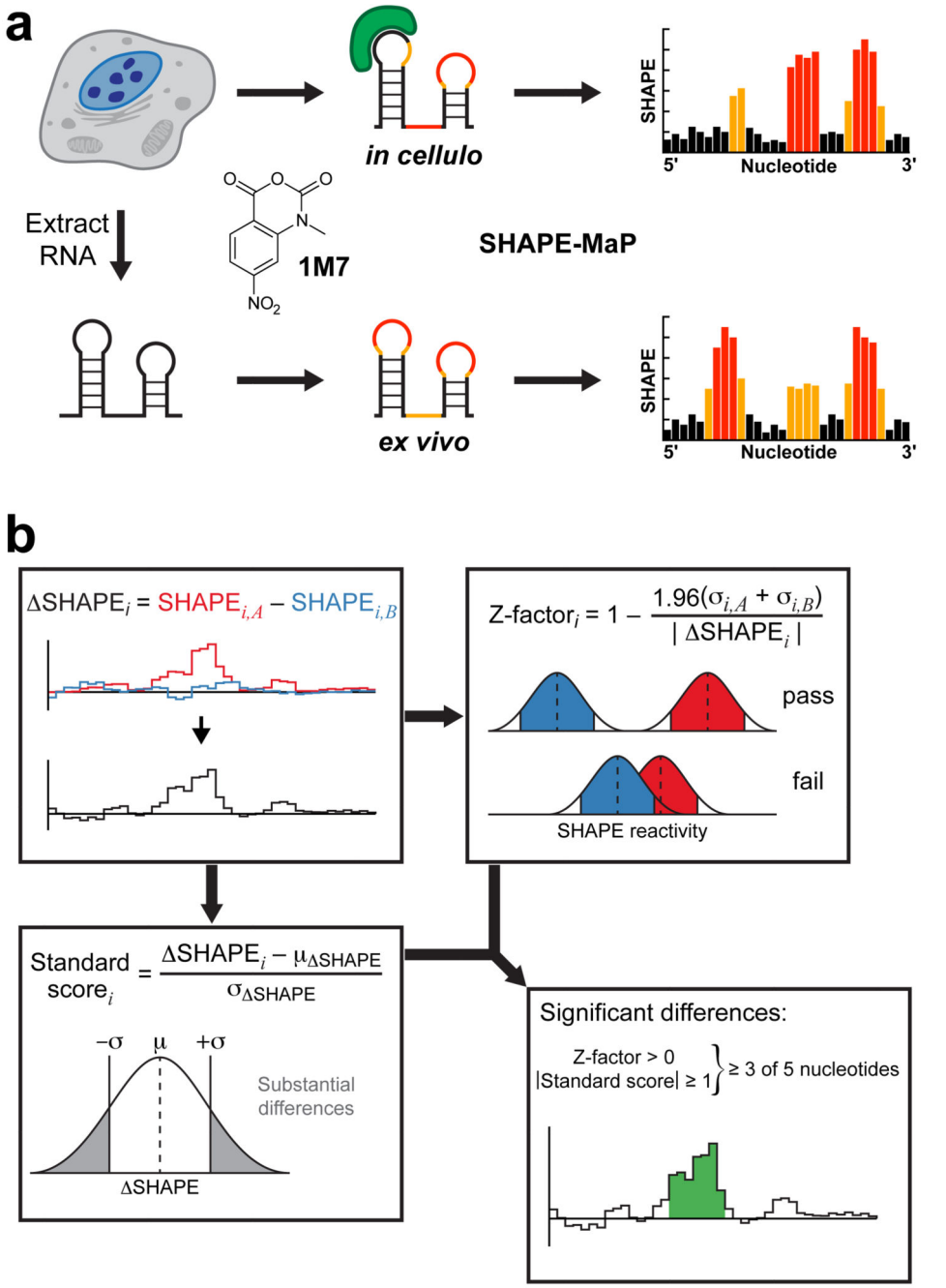


**Figure 1.**

Experimental and analytical framework for detecting SHAPE-MaP reactivity differences. (a) Total cellular RNA is treated with 1M7 under native conditions in living cells (top) or following non-denaturing extraction into folding buffer (bottom). RNAs that interact stably with cellular proteins (green) exhibit different SHAPE reactivities under *in cellulo* versus *ex vivo* conditions. Black, orange, and red illustrate low, moderate, and high reactivities, respectively on the secondary structure diagram and in SHAPE-MaP profiles. (b) Calculation of differences in SHAPE reactivities (ΔSHAPE) between experimental

conditions *A* and *B* (upper left). If (*i*) the Z-factor for a nucleotide is greater than zero, indicating that the 95% confidence intervals of measurements in the two conditions do not overlap, (*ii*) the standard score is greater than one standard deviation from the mean ΔSHAPE (lower left), and (*iii*) three of five nucleotides in a sliding window meet both Z-factor and standard score criteria (lower right), the (ΔSHAPE) reactivity difference is accepted as significant.

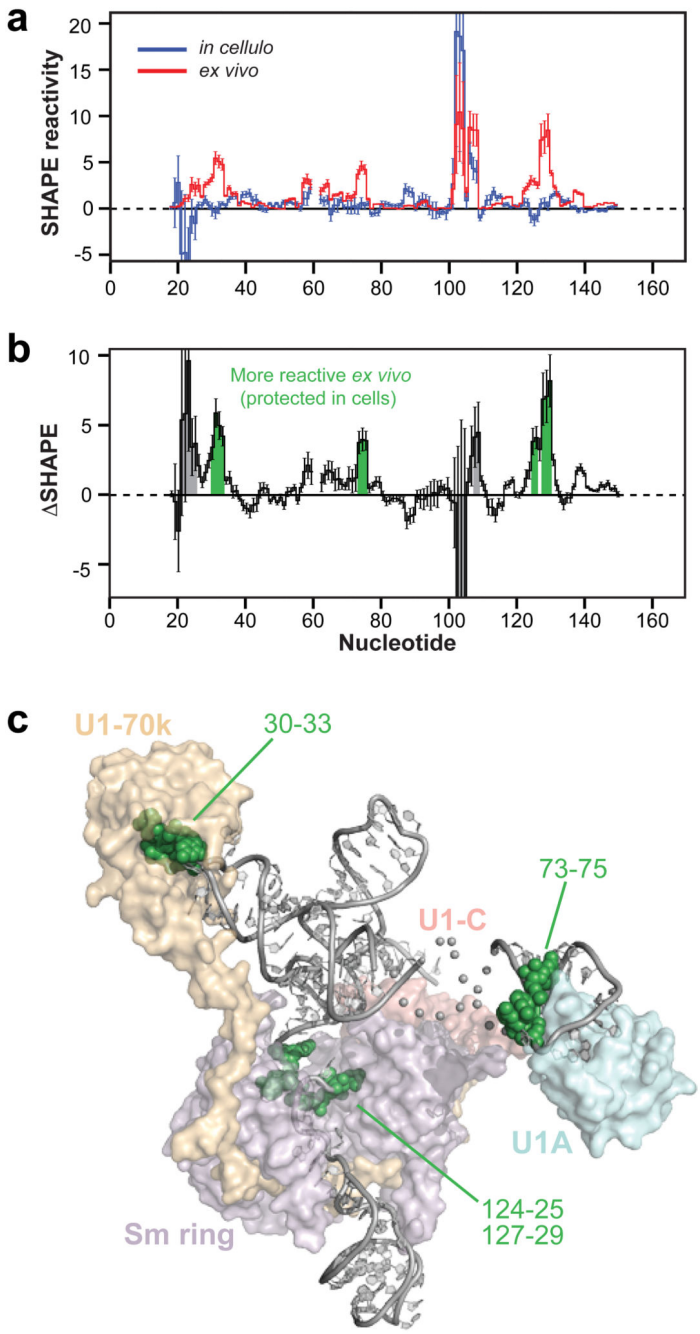


**Figure 2.**

Identification of protein binding sites by ΔSHAPE analysis. (a) Smoothed SHAPE reactivities for U1 snRNA *in cellulo* (blue) and *ex vivo* (red). (b) ΔSHAPE values for the U1 snRNA. Significant reactivity changes as established by the ΔSHAPE analysis framework are shaded green. If measurement errors were not taken into account, several off-target interaction sites would have been incorrectly identified as significant (grey shading). Primer-binding regions for which no data are available are shown with dashed lines. (c) Model of the human U1 snRNA complex including U1-70K (orange), U1-C (red), U1A (blue), and

Sm ring proteins (purple; subunit D1 excluded for clarity). RNA is shown as a ribbon. Nucleotides that exhibit significant ΔSHAPE values are emphasized as spheres.

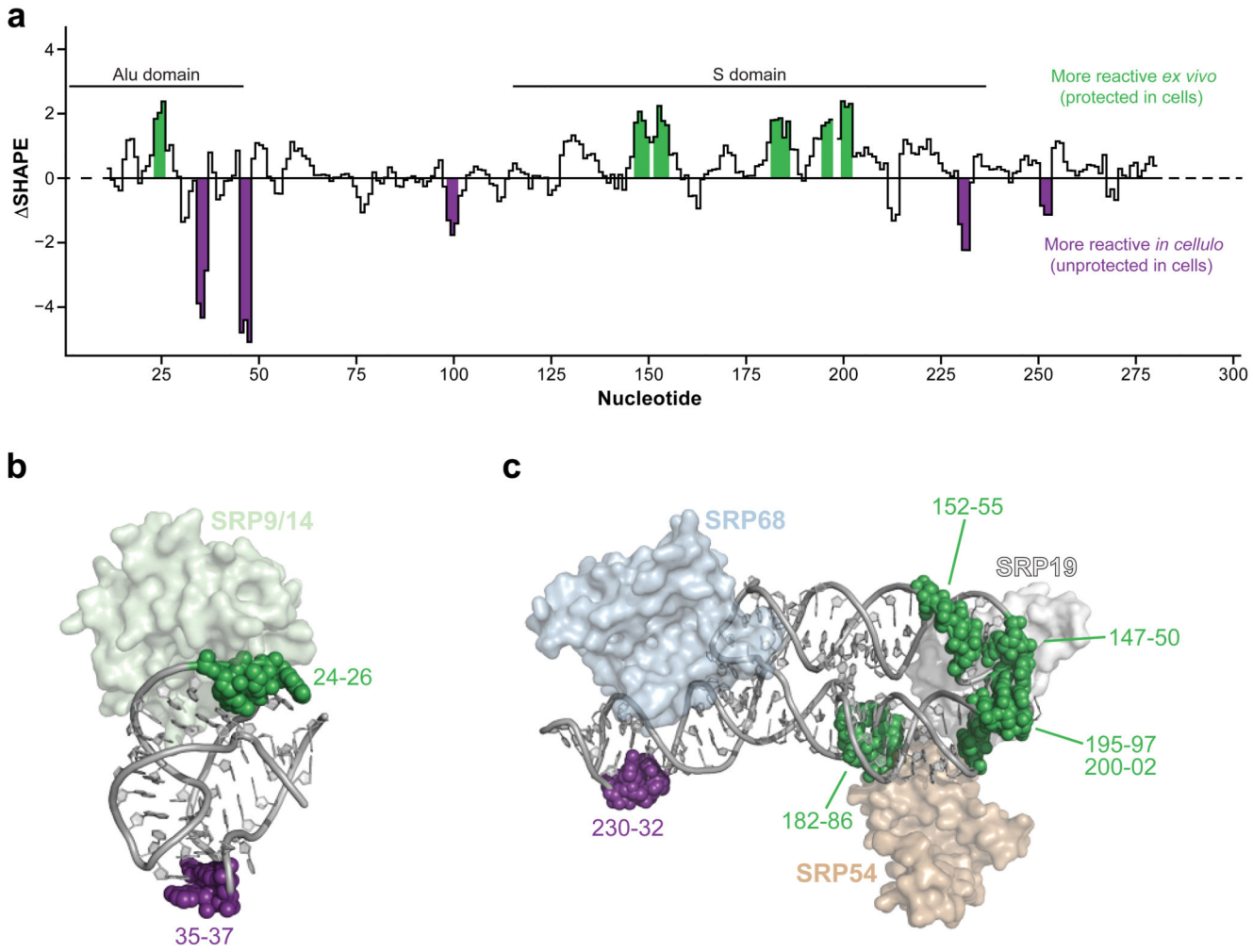


**Figure 3.**

Summary of results obtained for the SRP RNA. (a) ΔSHAPE profile for the entire SRP RNA. *In cellulo* protections are shaded green, and *in cellulo* reactivity enhancements are purple. Locations of the Alu and S domains are indicated. (b) Crystal structure of the Alu domain bound to SRP9/14. Nucleotides with significant reactivity differences are labeled. (c) Model of the S domain bound to SRP68, SRP19 and SRP54 with significant reactivity differences indicated.

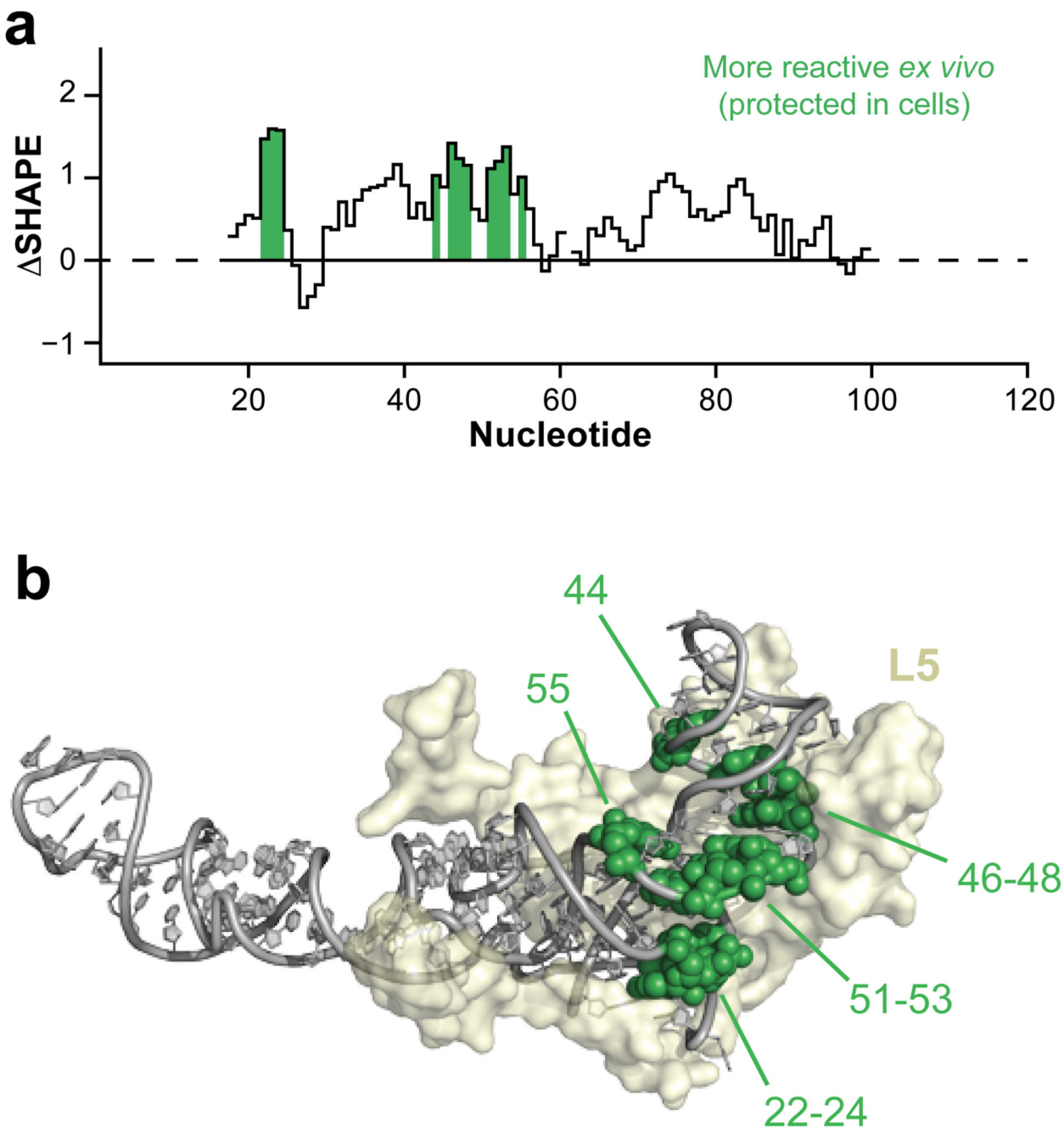


**Figure 4.**

Summary of results obtained for the 5S rRNA. (a) ΔSHAPE profile of 5S rRNA with nucleotides protected *in cellulo* indicated in green. (b) Cryo-EM structure[26] of the 5S rRNA bound to ribosomal protein L5. Sites of significant ΔSHAPE are labeled.

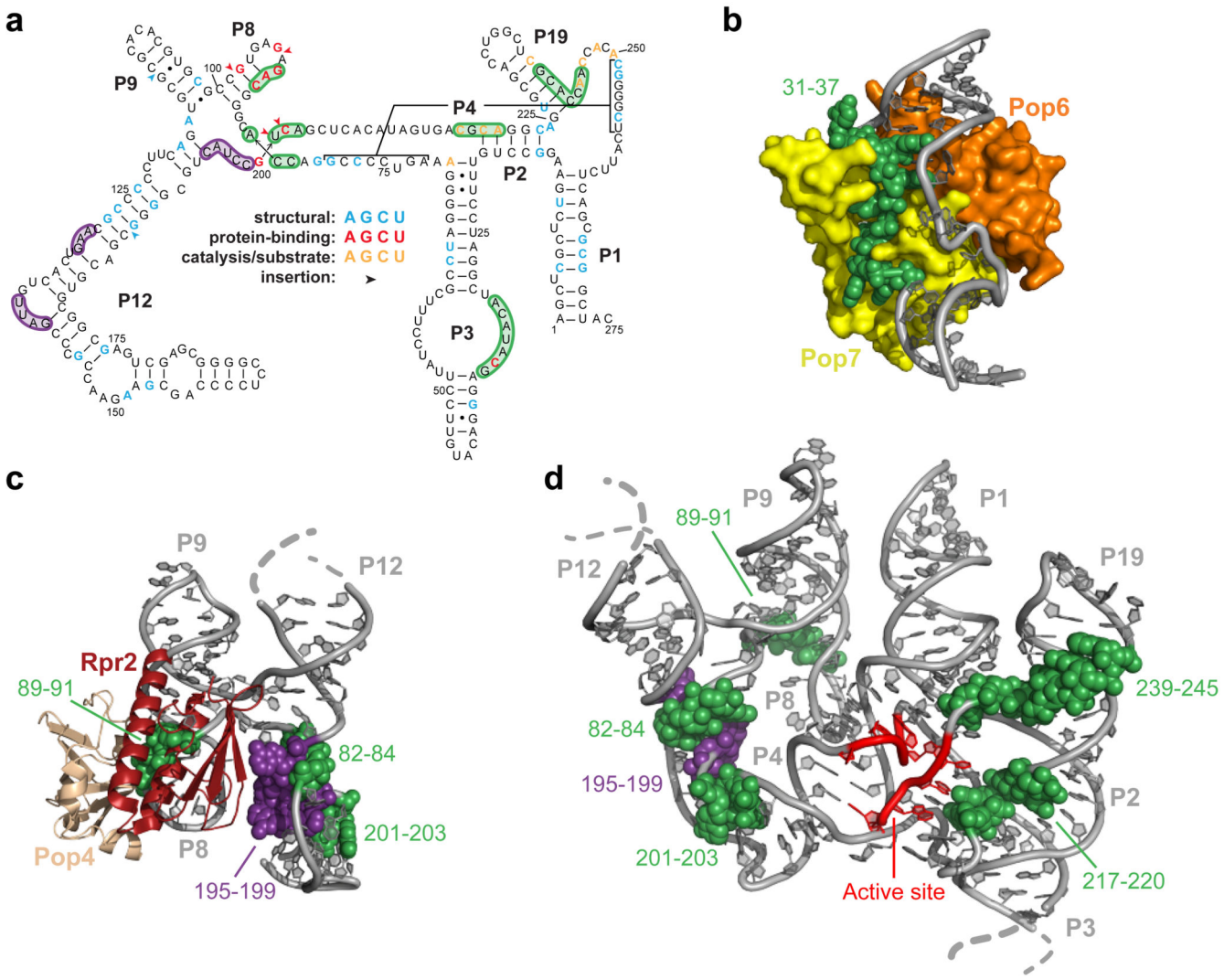


**Figure 5.**

In-cell analysis of RNase MRP RNA interactions. (a) Secondary structure of the RNA component of RMRP[46], showing RNA-protein interactions detected by ΔSHAPE analysis. Nucleotides protected *in cellulo* are shaded green, and those with enhanced reactivity are purple. Nucleotide positions corresponding to disease-associated mutations that affect function due to inferred (based on ΔSHAPE analysis) RNA structure, protein interactions, or catalysis and substrate recognition are shown in blue, red, and yellow, respectively. (b) Crystal structure of eukaryotic Pop6 (orange) and Pop7 (yellow) proteins interacting with the P3 domain of RMRP (3iab)[47]. Nucleotides 31-37 show ΔSHAPE protection *in cellulo* (green spheres) and interact tightly with Pop7. Nucleotides on the opposite side of the P3 internal loop are not tightly associated with Pop6/Pop7 and, correspondingly, do not exhibit strong interactions as assessed by ΔSHAPE. (c) Model of the junction between RMRP RNA helices P8, P9, and P12, showing interactions with Pop4 (tan). Nucleotides exhibiting significant ΔSHAPE values are shown as spheres and colored as in panel (a). In the cryo-EM model, yeast Rpr2 (a potential homolog of Snm1) also binds in this region[42] and this protein

or an alternative mouse protein may interact with nucleotides 82-84, 195-199, and 201-203. (d) Model of core regions in the eukaryotic RNase P RNP, showing regions of protection and enhanced reactivity as in (a-c). Conserved active site nucleotides are colored red. Nucleotides 217-220 and 239-245 are protected *in cellulo* and form a path to the active site, supporting a role in substrate recognition.