

# High-throughput RNA structural analysis by using chemical probing resolved by capillary electrophoresis

Mario A. Pulido<sup>1</sup>, Jane Luo<sup>1</sup>, Steve Questa<sup>1</sup>, Mollie Schubert<sup>2</sup>, Garret Rettig<sup>2</sup>, Manfred Souquet<sup>1</sup>, Handy Yowanto<sup>1</sup>, Susan Darling<sup>1</sup>, Sahana Mollah<sup>1</sup>

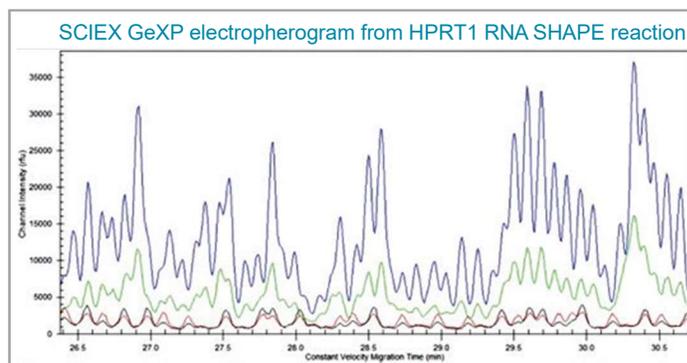
<sup>1</sup>SCIEX, Brea, CA, USA; <sup>2</sup>Integrated DNA Technologies, Coralville, IA, USA

## Introduction

Key studies for RNA structure and function particularly for promising RNA therapeutics such as vaccines, retroviral and lentiviral vectors, aptamers, anti-sense RNA, and small interfering RNA, highly depend on the characterization of the corresponding coding sequence and its structural conformation.<sup>1</sup> The latter gaining momentum for continuous improvements of gene and biological therapies, specifically by expanding the understanding for the RNA mechanism of action, and potential applicability for stability studies or critical quality attributes needed for commercialization. In this technical note, we describe the application of the capillary electrophoresis-based GenomeLab GeXP Genetic Analysis System for RNA structural modeling at the nucleotide level.<sup>2</sup> Thus, providing a solution to solving complex RNA conformation questions to better guide actionable studies in the drug development pipeline, and establish timely, cost-effective quality assays for commercialized products.

Briefly, the visualization of RNA conformations depending on x-ray crystallography, small-angle scattering, cryogenic electron microscopy, or nuclear magnetic resonance provide high atomic resolution; however these biophysical methods are extensively laborious and require costly equipment.<sup>3</sup> In contrast, RNA biochemical probing coupled to automated capillary electrophoresis instruments have been shown to reduce RNA sample processing time, costs, and can be applied to produce evidence-based RNA conformation models for both small and large RNA molecules.

Specifically, we applied the Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) reaction to determine the accessibility of RNA nucleotides by the highly electrophilic probing reagent N-Methylisatoic Anhydride (NMIA) for RNA-2'-O-adduct formation.<sup>4</sup> As a result, a structural map for the human HPRT1 RNA transcript was constructed (fig. 2), after reverse transcription, by applying minimum free-energy computational

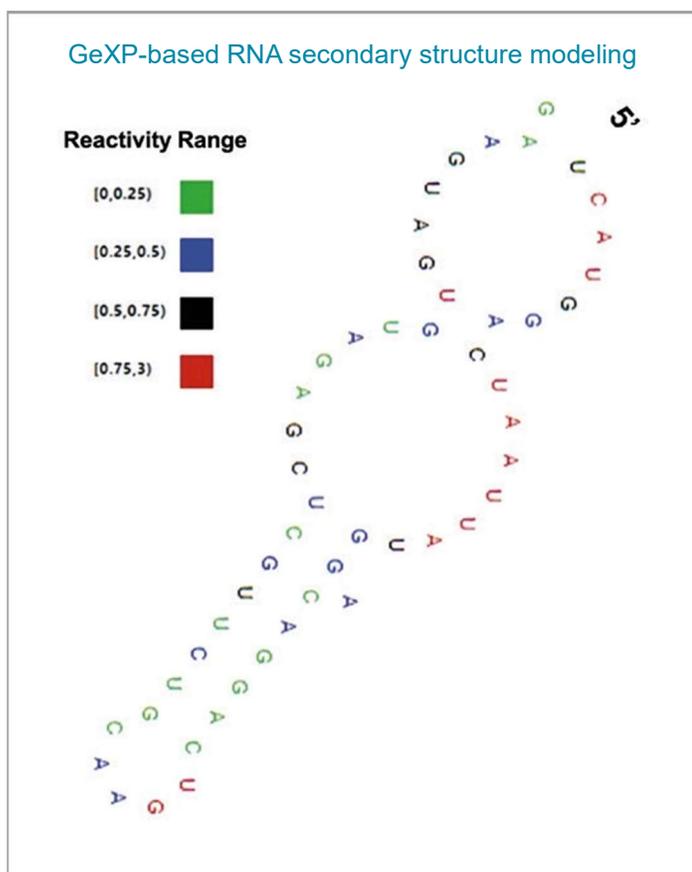


**Figure 1.** Representative section of the GeXP RNA SHAPE HPRT1 reaction profiling. Blue, positive reaction; Green, negative reaction; Black and Red, sequencing ladders.

algorithms. And essential for RNA secondary structure modeling was the high resolution of single modified nucleotides made possible by the separation capabilities and extreme sensitivity of the fluorescent signal found on the SCIEX GenomeLab GeXP.<sup>2</sup>

## Key features

- Evidence based RNA structural modeling reveals regions of nucleotide flexibility, aiding in in-depth understanding of RNA therapeutics actions.
- Built on automated multi-capillary CE platform with ready to use separation reagents, capacity for high-throughput processing, with potential for mapping very large RNA molecules.
- Minimal requirement of RNA samples (6 pmol of RNA per reaction).
- Results within two hours for every eight-capillary run for subsequent computational analysis.
- Fluorescence-based detection, no radioactive waste.
- Single-base resolution, thus every RNA nucleotide can be assessed using optimized GeXP separation conditions.



**Figure 2.** The SHAPE reaction was applied to the full length HPRT1 RNA transcript and resolved by the SCIEX GeXP Genetic Analyzer. This structural section corresponds to the electropherogram shown in figure 1. SHAPE data normalization and RNA conformation modeling was performed by RNAthor software.<sup>5</sup>

## Materials and methods

### Reagents

**Labeled primers.** Fluorescent reverse primers were synthesized by Integrated DNA Technologies (IDT). Oligonucleotides for the positive SHAPE or acylation reaction were labeled with IDT Freedom™ Dye Cy5™ (Ex. 648 nm, Em. 668 nm). The oligonucleotides for the negative SHAPE or control RNA reaction were labeled with IDT Freedom™ Dye Cy5.5™ (Ex. 685nm, Em. 706 nm). The IDT Alexa Fluor Dye 750® (Ex. 649nm, Em. 662nm) was used to label oligonucleotides for DNA ladder preparation. The second reference DNA ladder used oligonucleotides labeled with IDT LI-COR IRDye® 800CW (Ex. 767nm, Em. 791nm).

**HPRT1 RNA.** The human HPRT1 (NCBI, NM\_000194.3, 1395bp) *in vitro* transcribed custom product was provided by the IDT Functional Genomics R&D team, aliquoted into small volumes, and stored at -80° C.

**HPRT1 DNA ladder.** IDT cloning vector (pUCIDT, Amp) carrying the HPRT1 DNA sequence was selected for its capacity to support genomic fragments up to 5000 base pairs long. The resulting plasmid (~4 µg) can be used directly as a template for DNA ladder preparation or expanded using E.coli.

**DNA purification.** CleanSEQ® Beckman Coulter Life Sciences paramagnetic beads were used for DNA ladder purification (PN A29151). Magnetic plate required (PN A32782). Ethanol precipitation could also be used for sequencing ladder purification.

**GeXP sample loading solution.** (SLS, SCIEX. PN 608082).

**GeXP separation buffer.** (SCIEX. PN 608012). mineral oil (PN 608114). separation gel (20 mL: PN 391438).

**DNA Sep cap array.** 33-75B (SCIEX. PN 608087).

### Other molecular biology grade reagents

N-Methylisatoic anhydride (NMIA) (PN, M25), SuperScript™ III (PN 18080044), dNTP Mix (PN R0192), Sequenase Cycle Sequencing Kit (PN 78500), Magnesium Chloride (PN AM9530G), Potassium Chloride (PN AM9640G), Tris-ETDA buffer (PN 12090015), and Tris- HCl, pH 8.0 (PN 15568025) were purchased from Thermo Fisher Scientific.

Dimethyl Sulfoxide (DMSO) (PN D8418), Sodium Acetate 3M pH 5.2 (PN S7899), Sodium Hydroxide (PN 72068), and Hydrochloric Acid (PN H1758) were purchased from Sigma-Aldrich.

Ethanol (PN BP2818) and EDTA (PN AAJ15694AE) were purchased from Thermo Fisher Scientific.

Nuclease-free water and not DEPC-treated (PN AM9930), was obtained from Invitrogen.



A mixture of (positive and negative SHAPE reactions) cDNA products and sequencing ladders were combined at an optimum ratio, reflecting unsaturated and above background signals, and loaded onto the GenomeLab GeXP Genetic Analyzer.

An electropherogram was generated within two hours after loading the sample for analysis, yielding signals in the blue, green, black, and red channels originating from cDNA products labeled with Cy5, Cy5.5 and sequencing ladders labeled with Alexa Fluor Dye 750 and LI-COR IRDye 800CW, respectively.

#### Section IV.

#### HPRT1 RNA structural analysis and modeling

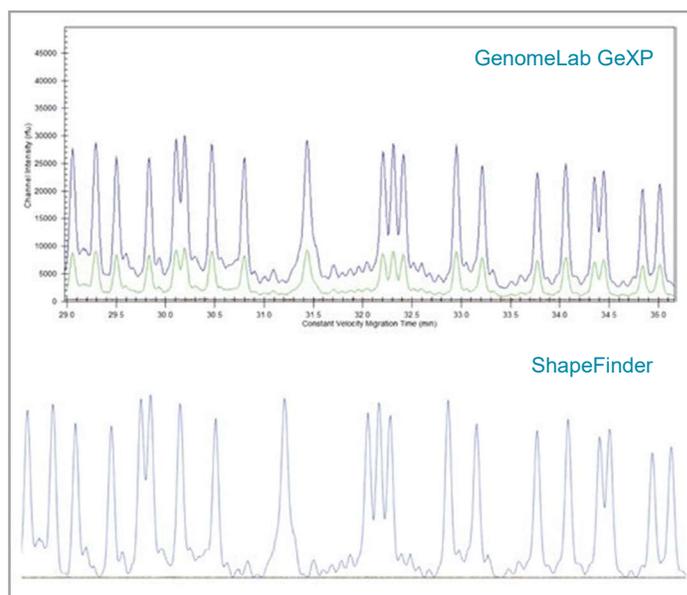
Nucleic acid reactivity following exposure to NMIA was examined through truncated cDNA products, as a result of premature reverse transcription termination, sized to specific nucleotide positions found at flexible regions of the HPRT1 RNA transcript due to RNA-2'-O-adduct formation. For this protocol, the computational set of algorithms found within the ShapeFinder software were used to quantitate nucleic acid reactivity by deciphering the GenomeLab GeXP electropherogram.<sup>8</sup>

These reactivity values were normalized and used for RNA conformational modeling by using RNAThor, ViennaRNA Web Services, or RNAstructure software.

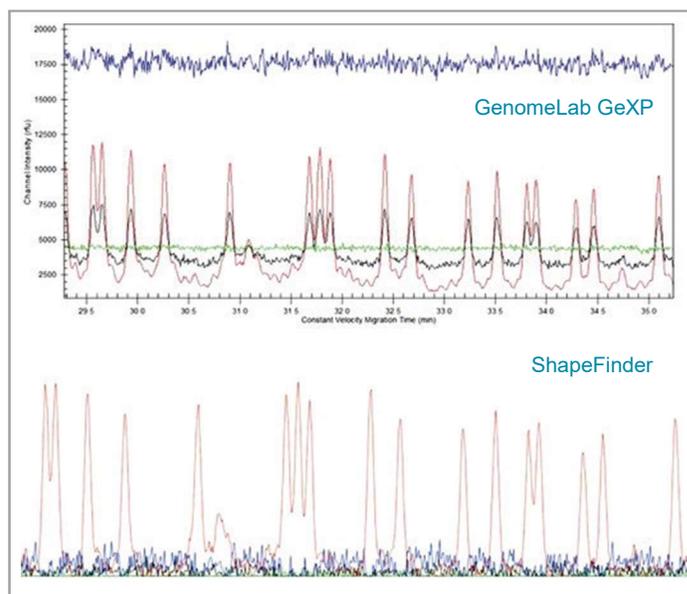
**Preliminary analysis: fluorescent dye cross-talk and mobility shift correction.** Prior to analyzing GenomeLab GeXP electropherograms for nucleotide reactivity values deriving from NMIA exposure, fluorescent dye cross-talk and mobility shift correction reactions were set up for proper analysis. As shown in figures 4a and 4b; using CleanSEQ paramagnetic beads for DNA purification, following HPRT1-pUCIDT plasmid amplification by using the Sequenase Cycle Sequencing kit and keeping the same ddTTP termination mix in the four separate reactions, with each reaction utilizing the same labeled reverse primer sequence with Cy5 (fig 4a), Cy5.5, Fluor Dye 750, or IRDye 800CW (fig. 4b), corresponding samples were analyzed followed by signal cross-talk correction. Figure 4c, shows a mixture of the reactions mentioned above, needed for mobility shift correction due to the difference in molecular weight as a result of the reverse primer labeled with different fluorescent dyes.

#### Preliminary analysis for SHAPE value calculations.

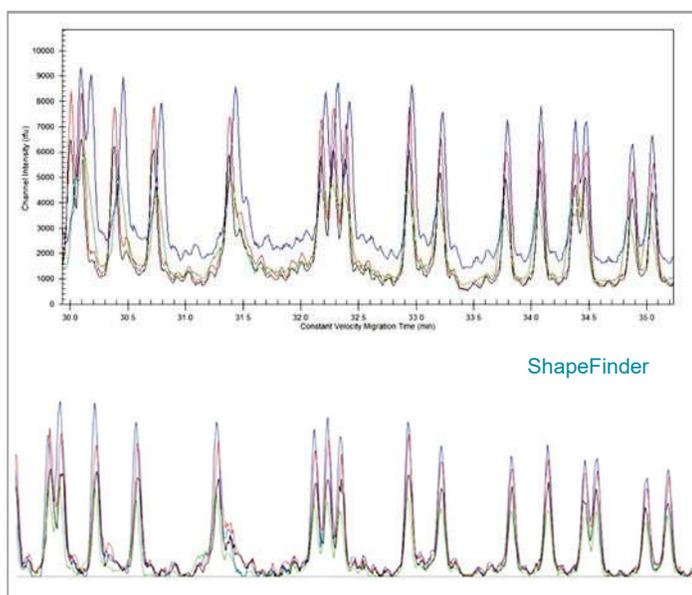
Figures 4a and 4b demonstrate the HPRT1 fluorescent dye cross-talk and mobility shift correction for nucleotide alignment and reactivity value calculations for the first ~400 bases of the full length transcript. This step was also applied to correct Cy5.5 and Alexa-750 signal cross-talk. Figures not shown.



**Figure 4a.** GenomeLab GeXP raw electropherogram (zoomed image) showing IDT Cy5 Dye (blue trace) cross-talk to another channel (green trace). Bottom image shows ShapeFinder cross-talk correction by applying Matrixing-2.1.



**Figure 4b.** GenomeLab GeXP raw electropherogram (zoomed image) showing IDT LI-COR IRDye 800CW Dye (red trace) cross-talk to another channel (black trace). Bottom image shows ShapeFinder cross-talk correction by applying Matrixing-2.1.

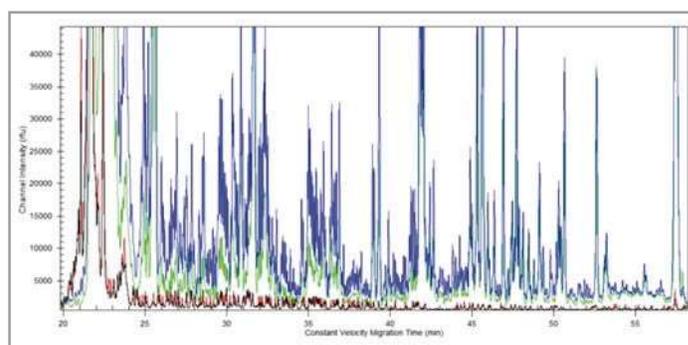


**Figure 4c.** Genomelab GeXP raw electropherogram (zoomed image) showing a product mixture of amplified HPRT1-pUC1DT plasmid using a common ddNTP and differently labeled reverse primers with Cy5 (blue trace), Cy5.5 (green trace), Fluor Dye 750 (black trace), or IRDye (red trace) 800CW. Bottom image shows ShapeFinder mobility-shift correction by applying the Cubic function.

## Results and Discussion

### HPRT1 RNA SHAPE reaction resolved by SCIEX GeXP

To empirically determine an RNA conformation model of the full length HPRT1 transcript conjugated with NMIA (see experimental part), and to maximize reverse transcription efficiency, reverse primers labeled with respective fluorescent dyes were designed at ~250-400 nucleotides apart depending on coding sequence properties like GC-rich regions or long repeats. Figure 5 shows the entire Genomelab GeXP raw electropherogram for the first 408 nucleotides of the HPRT1 RNA molecule assessed for SHAPE cDNA products along with the complementary sequencing ladders. Using a previously determined NMIA concentration obtained by performing a titration assay with varying NMIA concentrations and including a control reaction by adding DMSO, varying Cy5 peaks were observed following optimized separation conditions. This peak variation (blue trace) suggests the quantitation of Cy5 cDNA products as a result of truncated reverse transcription due to RNA-2'-O-adduct formation. The green trace (Cy5.5) shows the DMSO treated HPRT RNA molecule, suggesting the baseline cDNA products due to random or possibly rigid structures. The mostly uniform peak height or signal intensity for the black (Alexa Fluor Dye 750) and red (IRDye 800CW) traces correspond to the sequencing ladders used for alignment and sequencing functions.



**Figure 5.** Raw superimposed electropherograms show SHAPE positive (Cy5, blue trace) and negative (Cy5.5, green trace) reactions, and sequencing ladder reactions: ddATP (Alexa Fluor Dye 750, black trace) and ddTTP (IRDye 800CW, red trace), for the HPRT1 RNA transcript fragment up to the first 408 bases.

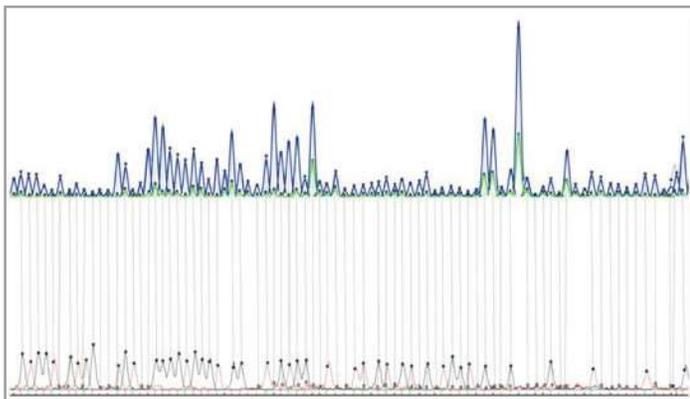
### SCIEX GeXP separation technology and dye quality: Critical factors for SHAPE reactivity value calculation

For this SCIEX technical note, ShapeFinder was applied for the quantitation of SHAPE based nucleic acid reactivity by taking the difference of the positive reaction (Cy5, blue trace) over the negative reaction (Cy5.5, green trace). In addition, to their alignment to the reference HPRT1 RNA sequence by integrating the ddATP and ddTTP based sequencing ladders: black and red traces (Alexa Fluor Dye 750, and IRDye 800CW, respectively). Figure 6, highlights the single nucleotide resolution made possible by the SCIEX GeXP proprietary separation technology as shown by the sequencing ladders, and critical for the alignment and integration of the SHAPE reactions and the sequencing ladders was the quality of the fluorescently labeled reverse primers obtained from Integrated DNA Technologies. As a result ShapeFinder peak calling occurred mostly automatically, thus providing an advantage for minimizing the manual alignment of SHAPE reactions and sequencing ladders to the reference HPRT1 RNA sequence.

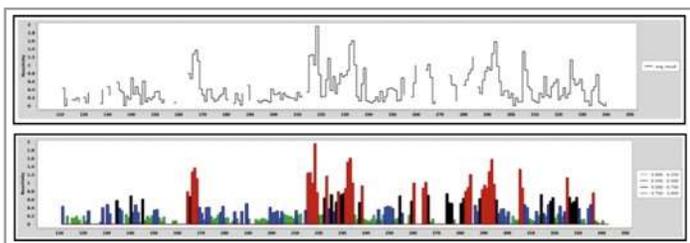
### SCIEX GeXP SHAPE quality data and normalization

RNAthor, a new software tool, was recently introduced for the fully automated normalization of SHAPE data resolved by capillary electrophoresis instruments like the SCIEX GeXP.<sup>5</sup> This cloud-based computational tool takes ShapeFinder files following alignment and integration and performs a series of steps to filter out outliers while retaining quality data. Also, a normalization step occurs resulting in a comprehensive view or reactivity map profiling indicating the respective Shape value or pseudo-energy

constraint information for each RNA nucleotide due to NMIA exposure. As shown in figure 7, gaps in the reactivity profiling plot for the first 408 nucleotides of the HPRT1-RNA molecule, suggest exclusion of data points as a result of hard-stops or outliers in the data set. Applying RNAtor or functionally-similar computational tools may become highly beneficial for assay assessment, quality of RNA sampling and processing, or dominant structural properties of RNA molecules.



**Figure 6.** Representative section of the GeXP SHAPE HPRT1 reaction profile. Blue, positive reaction; Green, negative reaction; Black and Red, sequencing ladders (ShapeFinder, zoom 200X).



**Figure 7.** (top figure). Normalized SHAPE reactivity values for the first 408 bases of the HPRT RNA transcript treated with NMIA reagent. Gaps in the “step-plot” suggest hard-stops or outliers in the data set. The bar-plot (bottom figure) illustrates a relative SHAPE reactivity breakdown. This color-coded reactivity profile provides a deeper understanding of RNA structural conformation.

### The path from a SCIEX GeXP Cy5 intensity peak to a pseudo-energy constraint or SHAPE value

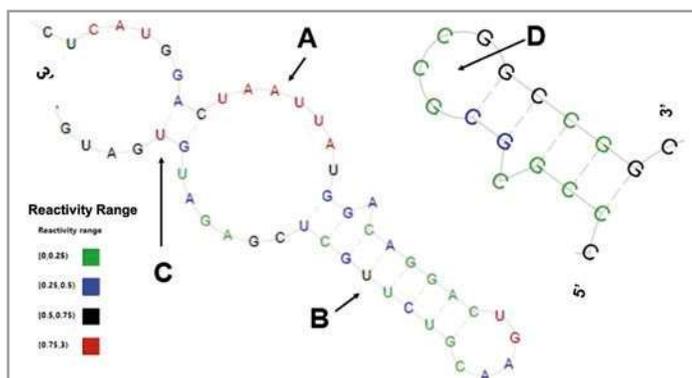
RNA structural changes due to nucleotide variation could also be studied using the SCIEX GeXP Genetic Analyzer by characterizing Cy5 (blue) peaks of interest matching the RNA reference sequence. For example, RNA SHAPE-treated variants following data normalization could be compared for Shape reactivity values or structural conformation. In detail, and as shown in Table 1, RNAtor software converted GeXP electropherogram-based data into relative SHAPE values meaningful for RNA conformation modeling. As a result, iterations of this computational process may be applied to compare structural differences among RNA variants.

seq num	seq	RX.area-BG.area	SHAPE value
249	U	5467.8	0.435
250	A	5251	0.418
251	U	4709.6	0.375
252	G	1507.8	0.12
253	C	1735.3	0.138
254	U	8657.3	0.688
255	G	3305.2	0.263
256	A	1094.7	-999
257	G	1667.2	-999
258	G	2736.6	0.218
259	A	7060.7	0.561
260	U	12593.2	1.001
261	U	11377.2	-999
262	U	15135.4	-999
263	G	17653.2	-999
264	G	11000.4	0.875
265	A	12813.4	1.019
266	A	8861	0.705

**Table 1.** SCIEX GeXP SHAPE values representing the constrained flexibility for the respective nucleotide along the RNA molecule. All together, these SHAPE values provide an empirical pseudo-free energy map, suggesting structure-function for RNA molecules. These results were derived from a representative section of the first 408 bases of the NMIA-treated HPRT1 RNA molecule.

## Solving HPRT RNA secondary structure using the GeXP

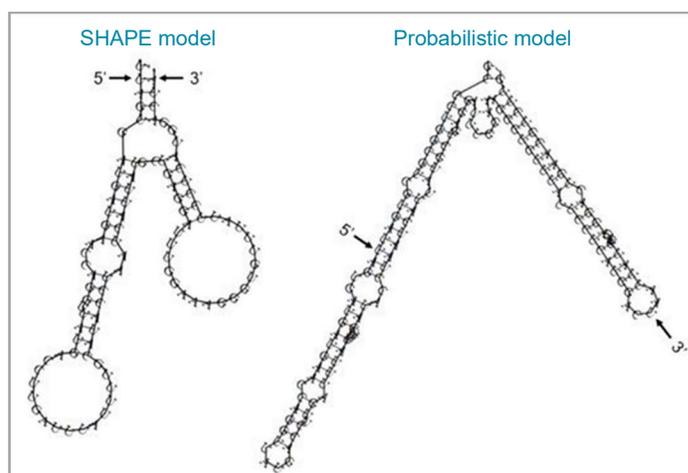
SCIEX GeXP evidence-based RNA nucleotide flexibility has the potential to reveal unexpected structures or to confirm structural information of RNA molecules alone or in combination with ligands such as small molecules, other RNAs, or proteins both *in vitro* and *in vivo*. Figure 8, shows a portion of a color-coded HPRT1 structural modeling using RNAtor. Green colored bases suggest a lower SHAPE value compared to the red colored nucleotides. Interestingly, arrow A shows a cluster of bases with high SHAPE values concentrating as part of a large pocket. Arrow B on the other hand, suggests that bases with a low SHAPE values are involved with rigid structure formation. Arrow C, suggests a breaking point as a nucleotide (U) with a high SHAPE value is found at the edge of a structural pocket. Substitution of this nucleotide or neighboring bases may be of interest to better understand the mechanisms behind RNA structure formation. Lastly, arrow D suggests the possibility of RNA folding involving GC-rich domains.



**Figure 8.** SCIEX GeXP SHAPE-based HPRT1 RNA structural modeling using RNAtor. Color coded bases indicate the respective SHAPE reactivity value range as a result of NMIA treatment. Arrows A, B, C, and D point to key structural regions.

## Surveying SHAPE-less HPRT1 RNA structural modeling

Because RNA structure is known to substantially contribute to biological processes and very likely as well to be involved in the mechanism of action of RNA therapeutics, we modeled the first 408 nucleotides of the HPRT1 RNA transcript, first by using constrained values derived from the SHAPE reaction, and second by using a probabilistic approach. In summary, an analysis comparing a secondary structure model for this HPRT1 RNA fragment constrained by SHAPE values demonstrated substantial divergence from a theoretical model by using the ViennaRNA Web Services.<sup>9</sup> Figure 9, provides a visual for secondary structure conformation calculated from minimum free energy and partition functions. The black arrows illustrate a sequence chosen for comparison within the 408 HPRT1 RNA fragment for visualization purposes, highlighting structural differences between the experimental and the theoretical model. In addition to base-pairing differences, there are various pocket formations inconsistent from the theoretical model. In conclusion, application of the SCIEX GeXP Genetic Analyzer for RNA structural studies may be instrumental for the characterization of promising RNA therapeutics or for quality assessment of commercialized products.



**Figure 9.** A comparison for HPRT1 RNA conformation modeling with and without SHAPE reactivity values. Black arrows indicate the edges for the coding sequence of interest and also suggest RNA directionality: 5' or 3' ends.

## Conclusions

- The SCIEX GeXP can be used to establish a high-through processing and systematic approach for interrogating both small and large RNA molecules.
- IDT Freedom quality dyes contributed to single base resolution and sequencing alignment.
- SCIEX GeXP capillary electrophoresis proprietary technology allows for single base separation, thus instrumental for supporting evidence base RNA conformational modeling.

## References

1. Kenyon, J., et.al. *Biochem. Soc. Trans.* 2014. 42. 1251-1255.
2. SCIEX GenomeLab Genetic Analysis System User's Guide.
3. Li, B. et al. *Front Genet.* 2020 Oct 26;11:574485.
4. Wilkinson, K.A., et. al. *Nat Protoc.* 2006;1(3):1610-6 5. Gumna, J., et. al. *PLOS ONE* 15(10): e0239287.
5. <https://rnathor.cs.put.poznan.pl/>
6. Gumna J., et.al. *RNA Biol.* 2019 Dec;16(12):1749-1763.
7. Smola, M., et. al. 2015. *Nat Protoc* 10, 1643–1669.
8. 8. Vasa, S. M., et. al. *RNA.* 2008 Oct;14(10):1979-90.
9. <http://rna.tbi.univie.ac.at/>

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to [www.sciex.com/diagnostics](http://www.sciex.com/diagnostics). All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries (see [www.sciex.com/trademarks](http://www.sciex.com/trademarks)).

© 2021 DH Tech. Dev. Pte. Ltd. RUO-MKT-02-13866-A



**Headquarters**  
500 Old Connecticut Path | Framingham, MA 01701 USA  
Phone 508-383-7700  
[sciex.com](http://sciex.com)

**International Sales**  
For our office locations please call the division headquarters or refer to our website at [sciex.com/offices](http://sciex.com/offices)