SHAPER: A Web Server for Fast and Accurate SHAPE Reactivity Prediction

Yuanzhe Zhou[1], Jun Li[1], Travis Hurst[1] and Shi-Jie Chen[1,2,3]*

[1]Department of Physics and Astronomy, University of Missouri, Columbia, MO, United States, [2]Department of Biochemistry, University of Missouri, Columbia, MO, United States, [3]Institute of Data Sciences and Informatics, University of Missouri, Columbia, MO, United States

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing serves as a convenient and efficient experiment technique for providing information about RNA local flexibility. The local structural information contained in SHAPE reactivity data can be used as constraints in 2D/3D structure predictions. Here, we present SHAPE predictoR (SHAPER), a web server for fast and accurate SHAPE reactivity prediction. The main purpose of the SHAPER web server is to provide a portal that uses experimental SHAPE data to refine 2D/3D RNA structure selection. Input structures for the SHAPER server can be obtained through experimental or computational modeling. The SHAPER server can accept RNA structures with single or multiple conformations, and the predicted SHAPE profile and correlation with experimental SHAPE data (if provided) for each conformation can be freely downloaded through the web portal. The SHAPER web server is available at http://rna.physics.missouri.edu/shaper/.

Keywords: SHAPE reactivity prediction, structure prediction, structure analysis, web server, coarse-grained simulation

1 INTRODUCTION

With the development of novel ribonucleic acid (RNA) structure determination methods alongside discoveries of new RNA structures and cellular functions, RNA has become increasingly important, contributing new avenues in the development of therapeutic applications for human diseases. Computational modeling of RNA structures could greatly deepen our understanding of RNA folding mechanisms. However, computational prediction of RNA structures from the sequence remains a significant unsolved problem (Shapiro et al., 2007; Laing and Schlick, 2010; Miao and Westhof, 2017).

Although lacking complete structural information, some experimental methods can provide useful details for guiding structure prediction. The selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) method is a convenient and efficient RNA structure probing technology with single nucleotide resolution that can provide information about local nucleotide structural dynamics (Merino et al., 2005; Wilkinson et al., 2006). The SHAPE reactivity of a nucleotide is reflected by the ability to bind SHAPE reagents—small ligands such as 1-methyl-7-nitroisatoic anhydride (1M7)—that preferentially bind to the oxygen of 2′-hydroxyl group of RNA nucleotides (Lee et al., 2017). Previous studies (Gherghe et al., 2008; Weeks, 2010; McGinnis et al., 2012) suggested that SHAPE reactivity is correlated with nucleotide flexibility, where unconstrained nucleotides tend to be more reactive while nucleotides constrained by base pairing, stacking, or other interactions are less reactive. The signals seen in SHAPE experiments intrinsically reflect interactions in the 3D structure, and can therefore be used to place effective constraints on the possible structures in a conformational pool generated by computational modeling software.

FIGURE 1 | A schematic view of the organization and function of the SHAPER web server.
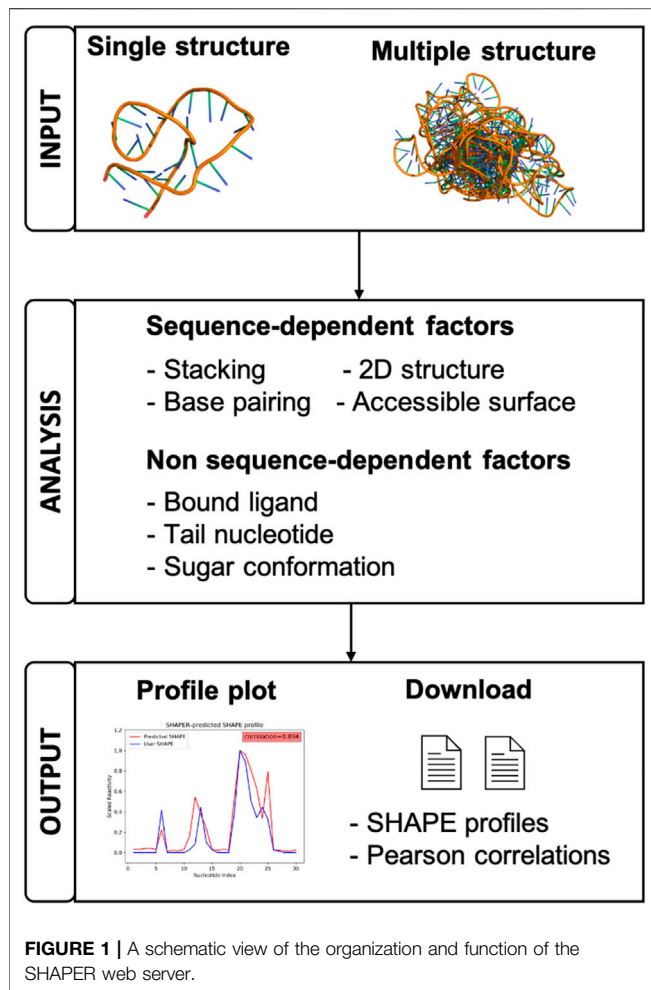
Since many RNA structure prediction studies would benefit from utilizing experimental SHAPE data, having a freely available, dedicated web server for rapidly predicting SHAPE profiles and filtering structural ensembles is essential. In this paper, we present our SHAPE predictoR (SHAPER) web server for predicting the SHAPE profile of any given RNA structure. The organization of the server is shown in **Figure 1**. The SHAPER server only requires the 3D coordinates of the target RNA (in PDB format). These structures can come from experimental structures, simulation snapshots, or computational structure-prediction models, etc. The SHAPER server can accept either individual structures or a structural ensemble, and the output contains predicted SHAPE profiles with the correlations between predicted profiles and a provided experimental SHAPE profile (if available). The engine powering the SHAPER web server is the new SHAPE prediction model (Hurst and Chen, 2021), which is an updated version of the original 3D Structure-SHAPE Relationship (3DSSR) model (Hurst et al., 2018). The SHAPER model incorporates RNA sequence-dependent bias into the prediction and is able to provide higher correlations between SHAPE data and the native RNA structure, which improves our ability to discern between SHAPE-compatible and -incompatible structures on decoys than the previous 3DSSR model.

2 MATERIALS AND METHODS

2.1 Workflow

The following shows both the workflow and theoretical background of the SHAPER server. Detailed description and analysis of the SHAPER model can be found in the original paper (Hurst and Chen, 2021).

2.1.1 Step 1: Uploading Input Data

As shown in **Figure 2**, Step 1, the input parameters are the following: (1) the input RNA structure file in PDB format, (2) user provided SHAPE profile, (3) user provided MASK file for the target RNA (for masking nucleotides that interact with ligands), (4) an email address for delivery of the calculation results, and (5) a simple text verification to prevent robotic usage. Required parameters are labeled by red asterisks. After submitting the job, the user will be redirected to a waiting page (**Figure 2**, Step 2), where they can view information about the running job. The information shown in the table in **Figure 2**, Step 2 includes: JobID—an identification code used to look up the results—and the file names of the RNA, SHAPE, and MASK file uploaded by user, respectively.

2.1.2 Step 2: Calculating SHAPE on Server Side

After submitting the job, SHAPER will put the job in a queue and will run the job once the computational resources are available. Usually, it takes less than a minute for a single structure with around 100 nucleotides. The procedures taken by the SHAPER server are listed in the order of execution.

- *Validating input.*

The input RNA file (in PDB format) is checked before any further processing. Entities other than RNA will be removed from the PDB file, only the backbone of modified residues and the first occurrence of atoms with multiple alternative locations will be kept for SHAPE reactivity calculations.

- *Identifying base pairing and stacking interactions.*

Base pairs are identified by RNAView (Yang et al., 2003), while stacking nucleotides are identified by our in-house Perl script. Then, pairing and stacking energies are combined into the interaction energy score (IE) for a given nucleotide i as

$$E_{\text{IE}}(i) = \sum_m [A \cdot E_{\text{bp}}^{(t)}(i, m) + B] + \sum_k E_{\text{st}}^{(i)}(i, k) \qquad (1)$$

where all the type-t base pairing energies $E_{bp}^{(t)}(i, m)$ and all the stacking energies $E_{st}^{(i)}(i, k)$ of nucleotide i are summed together. A and B are two extra parameters trained for the SHAPER model. The base pairing interactions $E_{bp}^{(t)}(i, m)$ were derived through a quasi-chemical statistical potential approach based on the statistical frequencies of the base pairing interactions extracted from the non-redundant RNA Basepair Catalog (Narayanan et al., 2013), and the stacking energies introduce $5' \rightarrow 3'$ polarity-dependence by using different weights and energy parameters for upstream $5'$ and downstream $3'$ nucleotides, respectively.
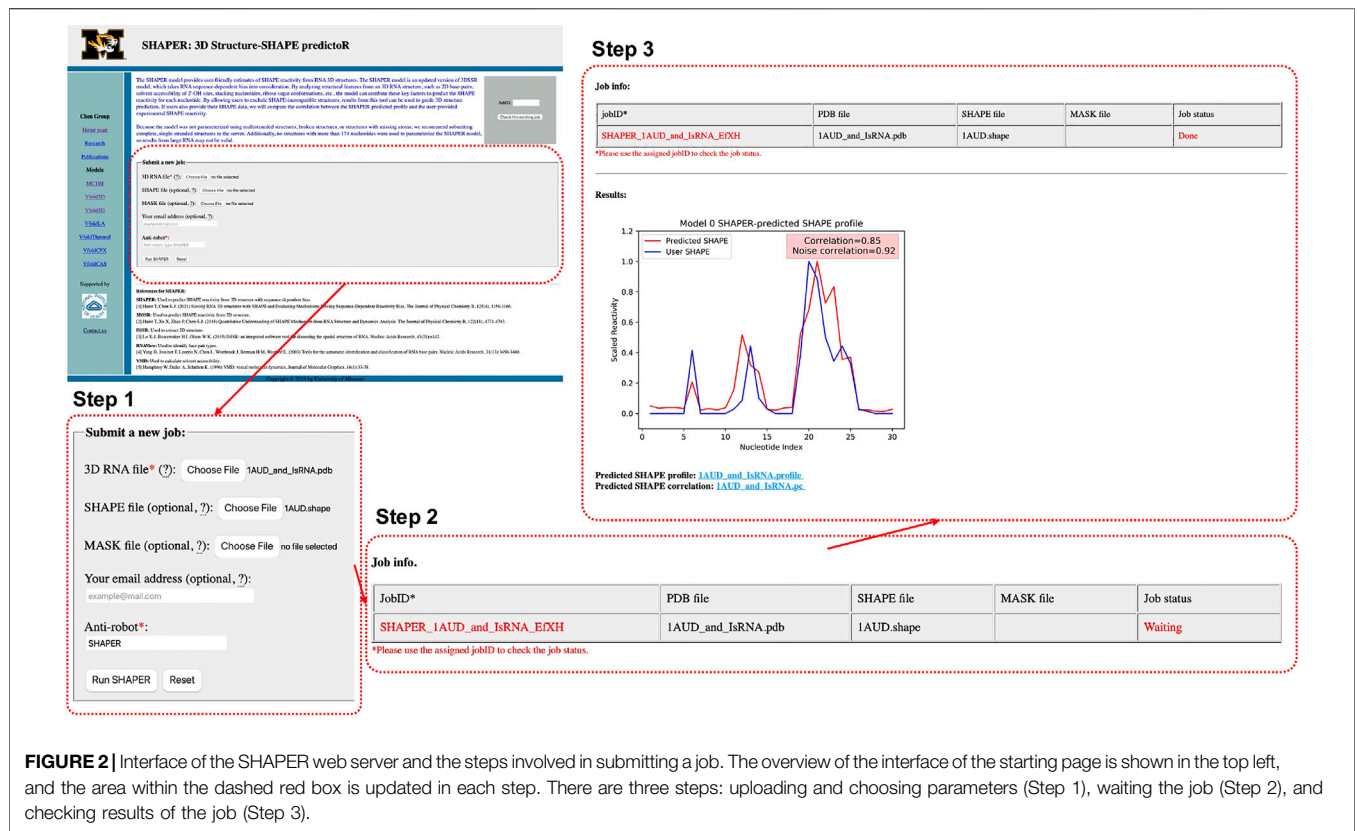
FIGURE 2 | Interface of the SHAPER web server and the steps involved in submitting a job. The overview of the interface of the starting page is shown in the top left, and the area within the dashed red box is updated in each step. There are three steps: uploading and choosing parameters (Step 1), waiting the job (Step 2), and checking results of the job (Step 3).

• *Extracting 2D structure.*

Using the Dissecting the Spatial Structure of RNA (DSSR) tool (Lu et al., 2015), the 2D structure is extracted from the input 3D structure. A parameter $E_{2D}(i)$ is introduced to represent the energy contributed by the base pairing nucleotide i in the 2D structure.

• *Accounting for other structural features.*

(1) Ligand Accessible Surface (A_{SAS}). The accessibility of the SHAPE reagent (1M7) to the 2′-hydroxyl of each nucleotide is calculated using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996) with a bead radius of 2.0 Å.

(2) Ribose sugar conformations. Previous studies (Vicens et al., 2007; Frezza et al., 2019) suggest that the conformation of the ribose sugar is important for SHAPE-reactivity. A correction F_{sug} determined by the pseudorotation angle of the ribose is employed to account for this effect.

(3) Tail nucleotides. Simple parameter F_{term} on terminal nucleotide is used to account for the effect of the short nucleotide sequence added at the terminal regions during SHAPE experiments.

(4) Bound ligands. Nucleotides interacting with a bound ligand need different treatment. To account for these effects, a ligand binding energy penalty E_{lig} is introduced for the nucleotides that are interacting with bound ligands. This is

achieved by masking the nucleotides that interact with the ligand. Users can supply their own mask file when submitting jobs on the web server, supplying 0 and 1 for non-interacting and interacting nucleotides, respectively. By default, the SHAPER server will treat all nucleotides as not interacting with ligand.

• *Accounting for the effects of neighboring nucleotides.*

Due to observations that a free nucleotide next to rigid nucleotides will be less reactive than a free nucleotide that has flexible neighbors, we introduce a weighted averaging scheme to account for this type of correlative effect for E_{IE}, E_{2D}, and A_{SAS} terms as

$$\bar{E}_{IE}(i) = \frac{\sum_{j=0}^{3} w_j \times E_{IE}(i+j-1)}{\sum_{j=0}^{3} w_j} \tag{2}$$

$$\bar{E}_{2D}(i) = \frac{\sum_{j=0}^{3} d_j \times E_{2D}(i+j-1)}{\sum_{j=0}^{3} d_j} \tag{3}$$

$$\bar{A}_{SAS}(i) = \frac{\sum_{j=0}^{3} a_j \times A_{SAS}(i+j-1)}{\sum_{j=0}^{3} a_j} \tag{4}$$

where $w_0 - w_3$, $d_0 - d_3$, and $a_0 - a_3$ are weights accounting for the influence of interactions involving the nucleotide of interest (NOI) and/or neighboring nucleotides.

- *Predicting the SHAPE profile.*

The final SHAPE prediction is a combination of the interaction factors, written as

$$p_i = SF_i \times e^{SE_i} \tag{5}$$

where structural factors SF_i and energy-like scores SE_i are determined by

$$SF_i = (\bar{A}_{SAS}(i) + A^0_{SAS}) \times F_{sug}(i) \times F_{term}(i) \tag{6}$$

$$SE_i = \bar{E}_{2D}(i) + \bar{E}_{IE}(i) + E_{lig}(i) \tag{7}$$

and A^0_{SAS} is a parameter that accounts for the breathing of the RNA structure that may allow an apparently inaccessible nucleotide to become accessible to the SHAPE reagent. The model implies an effective ambient temperature when modeling SHAPE reactivity. Indeed, solution conditions including temperature can influence RNA conformational fluctuation and the reaction for SHAPE reagents (such as 1-methyl-7-nitroisatoic anhydride) to form $2'$-O-adducts with RNA nucleotides. Because SHAPE experimental data were collected under the folding conditions for the respective (folded) RNAs, the parameters in the model may be appropriate for the selection of folded RNA structures for the experimental conditions involved in the training data set. Considering that different SHAPE experimental data for different RNAs were often collected at different solution (such as temperature) conditions, the parameters in the model reflect an average effect of the different experimental conditions.

- *Calculating regular and noise-adjusted Pearson correlations.*

In the original 3DSSR model, the relationship between the predicted SHAPE profile and experimental SHAPE data (if provided) was characterized by the Pearson correlation (PC). However, this regular PC does not account for the log-normality of SHAPE data and noise found by multiple previous studies (Deng et al., 2016; Vaziri et al., 2018). The newer SHAPER model uses a noise-adjusted normalization method to calculate the noise-adjusted PC between the predicted SHAPE profile and reweighted experimental SHAPE reactivities (Hurst and Chen, 2021).

2.1.3 Step 3: Showing Output

After submitting a job, the user will be directed to a result page which shows the job status and information about the input files. This page will be refreshed every few seconds. Once the job is done, The result page will be updated and the Job status will change from "Waiting" to "Done". A plot of the predicted and user-provided (if any) SHAPE profile along with normal and noise-adjusted Pearson correlation coefficients will appear below the status table. Links to download corresponding SHAPE prediction and correlation files will appear at the bottom of the page (**Figure 2** Step 3). Existing results can be accessed by

using the JobID, by bookmarking the address of the result page, or by checking email results (if provided).
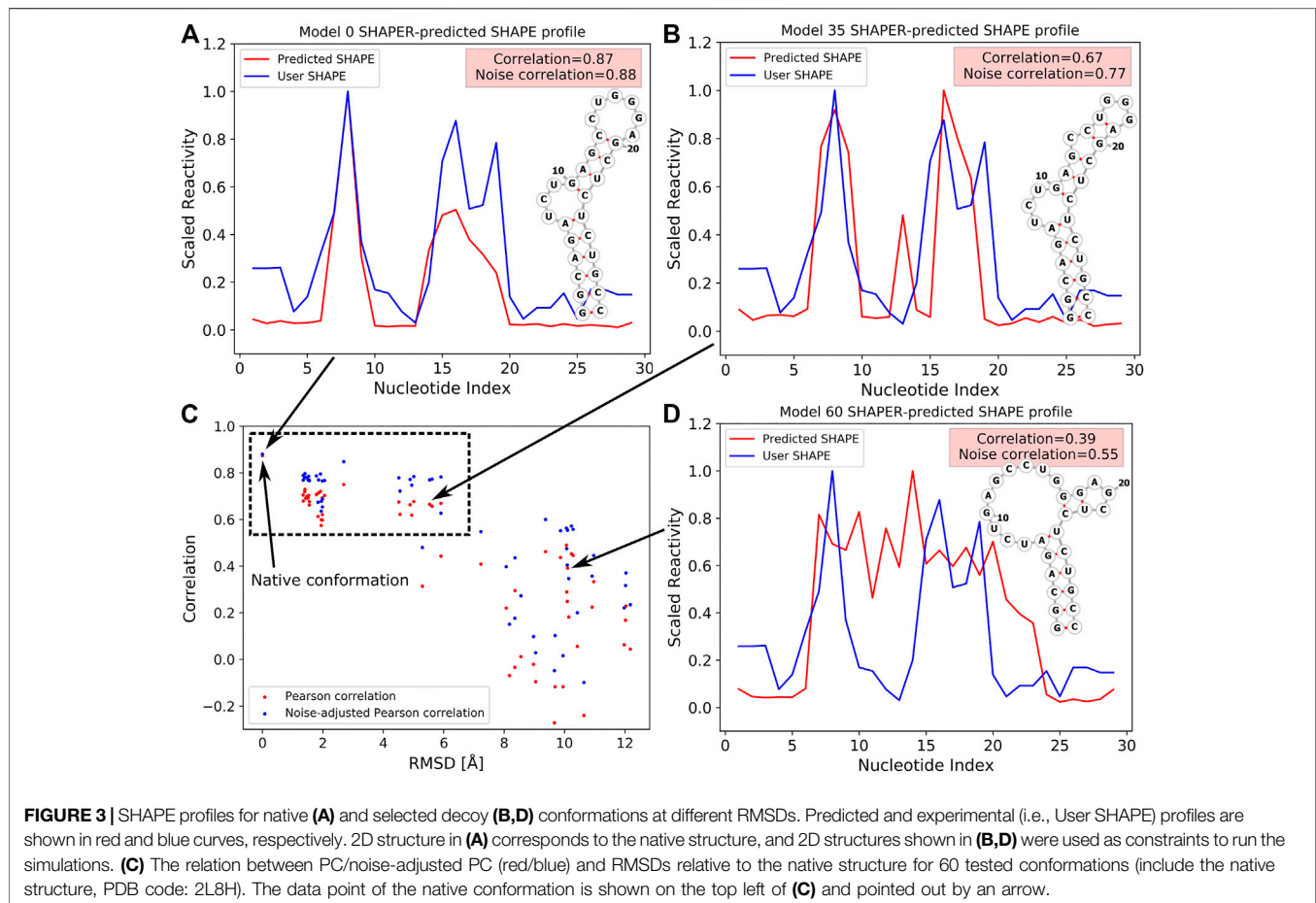
2.2 Server Implementation

Several programming and scripting languages are used in the SHAPER server, including Bash, C++, Python, Perl, and Tcl. The SHAPE prediction module is implemented in C++ for performance. Third party software packages are used in other modules for preparing the necessary input files. Dissecting the Spatial Structure of RNA (DSSR) (Lu et al., 2015) is used to extract the 2D structure and torsional information of the ribose sugars from a 3D structure. RNAView (Yang et al., 2003) is used to identify base pair types shown in 3D structure, and the identification of stacking interactions is carried out by our in-house program written in Perl. The ligand accessible surface of the $2'$-hydroxyl for each nucleotide is calculated using Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). The above tools help automate the preparation process and greatly reduce the potential for human error. All modules were combined by Python and the web server is based on Apache 2.2.15.

3 CASE STUDY

3.1 Sieving RNA 3D Structures Generated by 3D Structure Prediction Software

To better illustrate the function of the SHAPER web server, we ran an example case with known experimental SHAPE data to show the ability of SHAPER to distinguish near-native conformations from a pool of decoys. The test RNA structure (PDB code: 2L8H) contains 29 nucleotides. We used our coarse grained (CG) simulation software (IsRNA) (Zhang and Chen, 2018; Zhang et al., 2021) and an all-atom molecular dynamics (MD) simulation to generate 59 decoy conformations for the target RNA. We selected 20 near-native conformations and 39 non-native conformations generated with native and non-native 2D structures (Hurst and Chen, 2021). These decoys along with the native structure allow us to show the ability of the SHAPER server to distinguish native conformation from conformational pools. Then we put these 60 structures into the SHAPER web server, and the SHAPE correlation coefficients (PC and noise-adjusted PC) between predicted SHAPE profiles and experimental SHAPE data were calculated. The root mean square deviations (RMSDs) between the native and decoy conformations were calculated for heavy atoms. As shown in **Figure 3C** for the relationship between RMSD and SHAPE correlation coefficients, the native structure shows the highest correlation, and the near-native conformations around 2 Å of RMSD also have high correlations. However, similar correlations were also found for non-native conformations around 4 to 6 Å. This is because the 2D structural constraints (see, **Figure 3B**) used to generate these decoys are similar to the native 2D structural constraints (see, **Figure 3A**). As for the non-native conformations generated by using different 2D structural constraints (see, **Figure 3B**), both correlation coefficients (PC and noise-adjusted PC) drop

FIGURE 3 | SHAPE profiles for native **(A)** and selected decoy **(B,D)** conformations at different RMSDs. Predicted and experimental (i.e., User SHAPE) profiles are shown in red and blue curves, respectively. 2D structure in **(A)** corresponds to the native structure, and 2D structures shown in **(B,D)** were used as constraints to run the simulations. **(C)** The relation between PC/noise-adjusted PC (red/blue) and RMSDs relative to the native structure for 60 tested conformations (include the native structure, PDB code: 2L8H). The data point of the native conformation is shown on the top left of **(C)** and pointed out by an arrow.

significantly relative to the values of the native conformation. The above results suggest that SHAPE correlation may serve as a useful measure to sieve structures and find the native and near-native 2D and 3D structures.

4 CONCLUSION

SHAPER is a fast and accurate web server to predict SHAPE profile for any given RNA structure. Compared to the original 3DSSR model, SHAPER greatly improves performance (Hurst et al., 2018) by accounting for sequence-dependent bias, tail effects, and ligand binding. In addition, the SHAPER model better reflects that SHAPE reactivities are a direct reflection of the underlying system energetics and incorporates effects related to the log-normality of SHAPE data and noise. The server provides functionalities for predicting SHAPE profiles for RNA with either a single structure or a structural ensemble. Combined with the available experimental SHAPE data, SHAPER can provide a reliable measure of the nativeness of the target conformation and serves as a convenient tool to help

researchers select the most probable RNA 3D structures from a pool of decoys.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

S-JC and JL conceived and designed the web server. YZ, JL, and TH developed the server. YZ, TH, and S-JC wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Deng, F., Ledda, M., Vaziri, S., and Aviran, S. (2016). Data-directed RNA Secondary Structure Prediction Using Probabilistic Modeling. *RNA* 22, 1109–1119. doi:10.1261/rna.055756.115

Frezza, E., Courban, A., Allouche, D., Sargueil, B., and Pasquali, S. (2019). The Interplay between Molecular Flexibility and RNA Chemical Probing Reactivities Analyzed at the Nucleotide Level *via* an Extensive Molecular Dynamics Study. *Methods* 162-163, 108–127. Experimental and Computational Techniques for Studying Structural Dynamics and Function of RNA. doi:10.1016/j.ymeth.2019.05.021

Gherghe, C. M., Shajani, Z., Wilkinson, K. A., Varani, G., and Weeks, K. M. (2008). Strong Correlation between SHAPE Chemistry and the Generalized NMR Order Parameter (S2) in RNA. *J. Am. Chem. Soc.* 130, 12244–12245. doi:10.1021/ja804541s

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graphics* 14, 33–38. doi:10.1016/0263-7855(96)00018-5

Hurst, T., and Chen, S.-J. (2021). Sieving RNA 3D Structures with SHAPE and Evaluating Mechanisms Driving Sequence-dependent Reactivity Bias. *The J. Phys. Chem. B* 125, 1156–1166. doi:10.1021/acs.jpcb.0c11365

Hurst, T., Xu, X., Zhao, P., and Chen, S.-J. (2018). Quantitative Understanding of SHAPE Mechanism from RNA Structure and Dynamics Analysis. *J. Phys. Chem. B* 122, 4771–4783. doi:10.1021/acs.jpcb.8b00575

Laing, C., and Schlick, T. (2010). Computational Approaches to 3D Modeling of RNA. *J. Phys. Condensed Matter* 22, 283101. doi:10.1088/0953-8984/22/28/283101

Lee, B., Flynn, R. A., Kadina, A., Guo, J. K., Kool, E. T., and Chang, H. Y. (2017). Comparison of SHAPE Reagents for Mapping RNA Structures inside Living Cells. *RNA* 23, 169–174. doi:10.1261/rna.058784.116

Lu, X.-J., Bussemaker, H. J., and Olson, W. K. (2015). DSSR: an Integrated Software Tool for Dissecting the Spatial Structure of RNA. *Nucleic Acids Res.* 43, e142. doi:10.1093/nar/gkv716

McGinnis, J. L., Dunkle, J. A., Cate, J. H. D., and Weeks, K. M. (2012). The Mechanisms of RNA SHAPE Chemistry. *J. Am. Chem. Soc.* 134, 6617–6624. doi:10.1021/ja2104075

Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005). RNA Structure Analysis at Single Nucleotide Resolution by Selective 2-hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223–4231. doi:10.1021/ja043822v

Miao, Z., and Westhof, E. (2017). RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annu. Rev. Biophys.* 46, 483–503. doi:10.1146/annurev-biophys-070816-034125

Narayanan, B. C., Westbrook, J., Ghosh, S., Petrov, A. I., Sweeney, B., Zirbel, C. L., et al. (2013). The Nucleic Acid Database: New Features and Capabilities. *Nucleic Acids Res.* 42, D114–D122. doi:10.1093/nar/gkt980

Shapiro, B. A., Yingling, Y. G., Kasprzak, W., and Bindewald, E. (2007). Bridging the gap in RNA Structure predictionTheory and Simulation/Macromolecular Assemblages. *Curr. Opin. Struct. Biol.* 17, 157–165. doi:10.1016/j.sbi.2007.03.001

Vaziri, S., Koehl, P., and Aviran, S. (2018). Extracting Information from RNA SHAPE Data: Kalman Filtering Approach. *PLOS ONE* 13, 1–29. doi:10.1371/journal.pone.0207029

Vicens, Q., Gooding, A. R., Laederach, A., and Cech, T. R. (2007). Local RNA Structural Changes Induced by Crystallization Are Revealed by SHAPE. *RNA* 13, 536–548. doi:10.1261/rna.400207

Weeks, K. M. (2010). Advances in RNA Structure Analysis by Chemical Probing. *Curr. Opin. Struct. Biol.* 20, 295–304. Nucleic acids/Sequences and topology. doi:10.1016/j.sbi.2010.04.001

Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2006). Selective 2-hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative RNA Structure Analysis at Single Nucleotide Resolution. *Nat. Protoc.* 1, 1610–1616. doi:10.1038/nprot.2006.249

Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H., et al. (2003). Tools for the Automatic Identification and Classification of RNA Base Pairs. *Nucleic Acids Res.* 31, 3450–3460. doi:10.1093/nar/gkg529

Zhang, D., and Chen, S.-J. (2018). IsRNA: an Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *J. Chem. Theor. Comput.* 14, 2230–2239. doi:10.1021/acs.jctc.7b01228

Zhang, D., Li, J., and Chen, S.-J. (2021). IsRNA1: De Novo Prediction and Blind Screening of RNA 3D Structures. *J. Chem. Theor. Comput.* 17, 1842–1857. doi:10.1021/acs.jctc.0c01148