

A Fast-Acting Reagent for Accurate Analysis of RNA Secondary and Tertiary Structure by SHAPE Chemistry

Stefanie A. Mortimer and Kevin M. Weeks*

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received January 18, 2007; E-mail: weeks@unc.edu

RNA sequences fold back on themselves to form structures that are difficult to predict, especially if only a single sequence is known.¹ Current algorithms correctly predict 50–70% of known base pairs on average.² Predicted secondary structure models achieving 50–70% accuracy tend to have regions in which the overall topology differs significantly from the correct one, making it difficult or impossible to develop robust biological hypotheses. Knowledge of which nucleotides are likely to be paired or single-stranded can significantly improve prediction accuracies.³

The ideal technology for chemically assisted RNA structure analysis would (1) be experimentally straightforward, (2) use a single reagent that reacts generically with all four nucleotides, (3) employ a self-quenching reagent, (4) involve short reaction times, and (5) be proven to yield accurate results with complex RNAs of known structure.

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry⁴ takes advantage of the discovery that the nucleophilic reactivity of a ribose 2'-hydroxyl group is gated by local nucleotide flexibility. At nucleotides constrained by base pairing or tertiary interactions, the 3'-phosphodiester anion and other interactions reduce reactivity of the 2'-hydroxyl.^{4a} In contrast, flexible positions preferentially adopt conformations (Figure 1A) that react with *N*-methylisatoic anhydride (**1**, NMIA) to form a 2'-*O*-adduct (Figure 1B). NMIA reacts generically with all four nucleotides, and the reagent undergoes a parallel, self-inactivating, hydrolysis reaction (Figure 1B). Thus SHAPE chemistry meets the first three of the five criteria outlined above.

However, NMIA is relatively unreactive and requires tens of minutes to react to completion. To address the final two criteria for a near-ideal RNA structure interrogation technology, we design a significantly more useful, fast acting, reagent for SHAPE chemistry. We then show that structural constraints obtained using this reagent allow the secondary and tertiary structure of a large RNA to be assessed with high accuracy.

The reactive carbonyl of NMIA is in the benzylic position relative to the aromatic ring system and should be sensitive to substituents in the meta or the para positions, owing to a direct resonance effect. We therefore evaluated the 2'-*O*-adduct-forming and hydrolysis activities of 1-methyl-7-nitroisatoic anhydride (**2**, 1M7) (Figure 1B). The para nitro substituent is strongly electron-withdrawing ($\sigma_p = 0.81$)⁵ and should increase adduct formation and hydrolysis rates in two ways: via a ground-state effect by increasing the electrophilicity of the reactive carbonyl and via a transition-state effect by stabilizing the negative charge in the developing tetrahedral reaction intermediate. We first monitored reagent hydrolysis as the increase in UV absorbance of the aminobenzoate products. 1M7 is significantly more labile toward hydrolysis than NMIA. 1M7 undergoes hydrolysis with a half-life of 14 s and therefore the reaction is complete in ~70 s; in contrast, NMIA requires over 20 min to react to completion (left panel, Figure 2).

We next evaluated the ability of each compound to react with 3'-phosphoethyl-5'-adenosine monophosphate (pAp-ethyl). pAp-

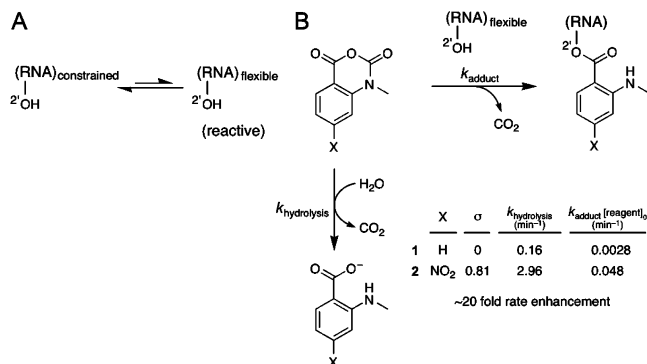


Figure 1. Mechanism of RNA SHAPE chemistry. (A) The nucleophilic reactivity of the 2'-hydroxyl group is selectively enhanced at flexible positions.⁴ (B) Parallel reaction of *N*-methylisatoic anhydride derivatives with RNA 2'-hydroxyl groups and with water. [reagent]₀ was 5 mM.

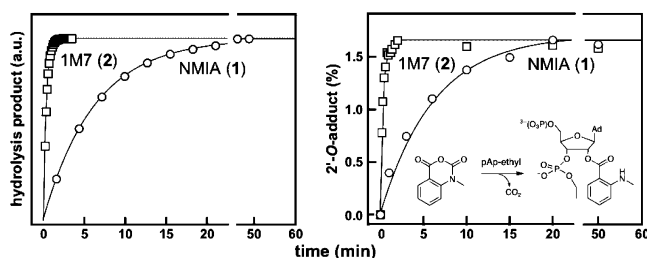


Figure 2. Comparative reactivity of 1M7 and NMIA via hydrolysis (left) and 2'-*O*-adduct formation with pAp-ethyl (right).

ethyl contains a 2'-hydroxyl and 3'-phosphodiester monoanion and is a good analogue for an unstructured RNA nucleotide.^{4a} 1M7 reacts significantly more rapidly with pAp-ethyl than does NMIA; however, the final extent of 2'-*O*-adduct formation for the two compounds is identical, within error (right panel, Figure 2).

Identical extents of reaction for NMIA and 1M7, despite the much faster reactivity of 1M7, indicate that the rates of hydrolysis and of 2'-hydroxyl acylation have increased by precisely the same 20-fold increment. These experiments indicate that 1M7 has the ideal chemical characteristics for a fast-acting and self-quenching reagent for RNA SHAPE chemistry.

We next evaluated the extent to which 1M7 provides accurate and quantitative information regarding RNA structure using the specificity domain of the *Bacillus subtilis* RNase P enzyme. This domain was chosen because it is a large (154 nt) RNA with a known structure⁶ that does not contain pseudoknots, which are not well predicted by current algorithms. This RNA spans numerous typical base-pairing and stacking interactions, a tetraloop-receptor tertiary interaction (involving L12 and P10.1) common to many large RNAs, and two large internal loops (J11/12 and J12/11) stabilized by an extensive series of noncanonical interactions.⁶

A SHAPE experiment was performed on the RNase P domain under conditions that stabilize the native tertiary fold (6 mM MgCl₂, 100 mM NaCl, pH 8.0) by treating the RNA with 6.5 mM 1M7.

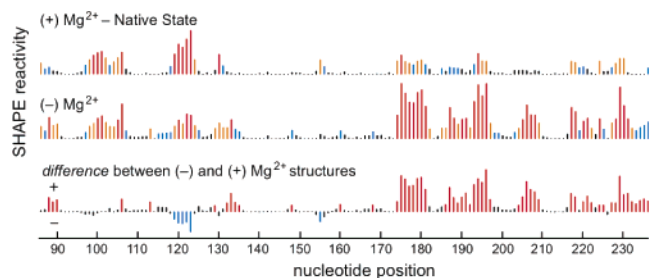


Figure 3. Histograms and difference plot of absolute nucleotide reactivities for SHAPE experiments performed with 1M7 in the presence and absence of 6 mM Mg^{2+} .

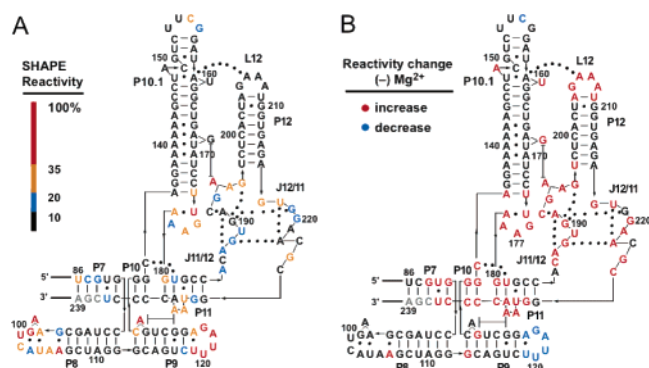


Figure 4. Base pairing and tertiary interactions for the specificity domain of *Bacillus subtilis* RNase P.

Sites of 2'-*O*-adduct formation were identified as stops to primer extension, using fluorescently labeled DNA primers, resolved by capillary electrophoresis⁷ (Figure S1). Absolute SHAPE reactivities were calculated by subtracting the background observed in no-reagent control experiments that omitted 1M7. Reactivity at each nucleotide was classified as high, medium, low, or near-zero (red, orange, blue, and black columns, respectively, in Figure 3).

Superposition of the quantitative reactivity information on a secondary structure diagram⁶ for the RNase P specificity domain shows that a 70 s reaction with 1M7 accurately reports the known secondary and tertiary structure for this RNA (Figure 4A). Essentially all nucleotides involved in Watson–Crick base-pairs are unreactive; moreover, many noncanonical, but stable, U·G, A·A, and A·G pairs are unreactive. Nucleotides in P10.1 and in L12 that form the tetraloop-receptor tertiary structure motif are also unreactive. In contrast, nucleotides in loops or adjacent to bulges or other irregularities are reactive. Nucleotides in the structurally idiosyncratic module involving J11/12 and J12/11 show a wide range of reactivities. Strikingly, the most highly conserved nucleotides in this module (A187, A191, G219–G220, A222), which participate in stabilizing tertiary interactions,⁶ also show the lowest SHAPE reactivities using 1M7.

We performed a similar SHAPE experiment in the absence of magnesium ion (middle histogram, Figure 3). Control experiments show that both reaction with the model nucleotide, pAp-ethyl, and 1M7 hydrolysis are independent of Mg^{2+} concentration (Figure S2). This Mg^{2+} -independence represents an additional significant improvement over the parent compound, NMIA, whose reactivity is strongly dependent on ionic strength (Figure S2). Thus, observed changes in SHAPE reactivity with 1M7 reflect changes in RNA secondary and tertiary structure and not Mg^{2+} -induced differences in reagent properties.

We quantified the effect of Mg^{2+} on the structure of the RNA using a difference plot in which nucleotide reactivities in the (+) Mg^{2+} experiment were subtracted from the (-) Mg^{2+} experiment.

Positive and negative peaks thus indicate an increase or decrease in local nucleotide flexibility in the absence of Mg^{2+} , respectively (lower panel, Figure 3). Many sites in the (-) Mg^{2+} experiment show increased SHAPE reactivity. Strikingly, increased reactivity occurs precisely at nucleotides that participate in tertiary interactions in the RNase P domain (in red, Figure 4B). SHAPE reactivity also shows that the irregularly stacked P7–P10–P11 helical domain unfolds when Mg^{2+} is removed (Figure 4B).

We then evaluated how well SHAPE information can be used to constrain the output of an RNA secondary structure prediction algorithm. We calculated prediction accuracies both using the native secondary structure (Figure 4A) as the target and using a modified structure that excluded the Mg^{2+} -dependent base pairs in the P7–P10–P11 domain. When the specificity domain of *Bacillus subtilis* RNase P is folded in RNAstructure,^{3d} the lowest free energy structure contains 52% of the correct pairs and features an overall topology that is radically different from the correct structure (Figure S3A). When SHAPE reactivity information is added to constrain single-stranded and non-internal base pairs, the lowest free energy structure is 76% correct using the native secondary structure as the target and 91% correct when base pairs in the P7–10–P11 domain (which do not form in the absence of native tertiary interactions) are excluded (Figure S3B). Using either target structure, the SHAPE-constrained prediction features an overall topology that closely resembles the correct structure.

SHAPE chemistry performed with 1M7 accurately reports the known structure of the RNase P specificity domain under native conditions. 1M7 reactivity detects nucleotides constrained both by base pairing and by idiosyncratic, noncanonical tertiary interactions (Figure 4). SHAPE chemistry enables very precise analysis of the differences between two structures, such as Mg^{2+} -dependent tertiary interactions. 1M7 is easily handled in the laboratory and enables analysis of large RNA structures at single nucleotide resolution in less than 70 s.

Acknowledgment. We are indebted to T. Pan and A. Mondragon for a *B. subtilis* RNase P specificity domain plasmid, to J. Chattopadhyaya for adenosine 3'-ethyl phosphate, and to J. Morken for helpful discussions. This work was supported by a grant from the NSF (Grant MCB-0416941 to K.M.W.).

Supporting Information Available: Experimental procedures and three figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) (a) Tinoco, I.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271–281. (b) Eddy, S. R. *Nat. Biotechnol.* **2004**, *22*, 1457–1458. (c) Doshi, K. J.; Cannone, J. J.; Cobaugh, C. W.; Gutell, R. R. *BMC Bioinformatics* **2004**, *5*, 105.
- (2) (a) Dowell, R. D.; Eddy, S. R. *BMC Bioinformatics* **2004**, *5*, 71. (b) Mathews, D. H.; Turner, D. H. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270–278.
- (3) (a) Peattie, D. A.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 4679–4682. (b) Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J.-P.; Ehresmann, B. *Nucleic Acids Res.* **1987**, *15*, 9109–9128. (c) Knapp, G. *Methods Enzymol.* **1989**, *180*, 192–212. (d) Mathews, D.; Disney, M.; Childs, J.; Schroeder, S.; Zuker, M.; Turner, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7287–7292. (e) Badorrek, C. S.; Weeks, K. M. *Nature Chem. Biol.* **2005**, *1*, 104–111.
- (4) (a) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231. (b) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4659–4667. (c) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *Nature Protocols* **2006**, *1*, 1610–1616.
- (5) Exner, O. *Correlation Analysis in Chemistry*; Plenum Press: New York, 1978.
- (6) Krasilnikov, A. S.; Yang, X.; Pan, T.; Mondragon, A. *Nature* **2003**, *421*, 760–764.
- (7) Wilkinson, K. A.; Gorelick, R. J.; Vasa, S. M.; Guex, N.; Rein, A.; Mathews, D. H.; Giddings, M. C.; Weeks, K. M., submitted for publication, 2007.

JA0704028