Supporting Information for:

Strong Correlation between SHAPE Chemistry and the Generalized NMR Order Parameter (S^2) in RNA

C.M. Gherghe, Z. Shajani, K.A. Wilkinson, G. Varani and K.M. Weeks

Experimental Procedures

RNA constructs. RNAs (TAR: 5'-GGCAG AUCUG AGCCU GGGAG CUCUC UGCC-3'; U1A: 5'-GGCAG AGUCC UUCGG GACAU UGCAC CUGCC-3'; T-SL4: 5'-GAGAC UAUCG ACAUU UGAUA CACUA UUUAU CAAUG GAUGU CUC-3') were synthesized by T7 RNA polymerase-mediated *in vitro* transcription using a PCR-generated template (SHAPE experiments) or a single stranded template with a double stranded promoter region (NMR measurements). SHAPE experiments for the TAR RNA were performed using a 976 nt RNA transcript.[1] For SHAPE experiments, U1A and T-SL4 RNAs were embedded within 5' and 3' structure cassette[2] sequences. RNAs were purified by denaturing polyacrylamide gel electrophoresis, excised from the gel, and recovered by electroelution and precipitation with ethanol. Purified RNAs were resuspended in either TE [10 mM Tris (pH 8.0), 1 mM EDTA] or sterile water and stored at –20 °C.

SHAPE analysis. RNAs were heated to 95 °C for 2 min, cooled on ice, and incubated at 37 °C for 10 min in SHAPE buffer [50 mM Hepes (pH 8.0), 50 mM KCl]. TAR and U1A RNAs were allowed to slowly cool to room temperature over 15 min; the T-SL4 RNA was incubated at 40 °C for 15 min. These conditions mimic those used for the NMR experiments, with the exception that the pH was 8.0. RNAs were then treated with 1-methyl-7-nitro-isatoic anhydride (1M7)[3] (1 μL; 100 mM; in anhydrous DMSO) and allowed to react for 5 min. No-reagent controls contained 1 μL neat DMSO. Modified RNAs were recovered by ethanol precipitation [80 μL water, 10 μL NaCl (5 M), 1 μL glycogen (20 mg/mL), 400 μL ethanol; 30 min at -80 °C; centrifugation at 10,000 *g* for 10 min] and resuspended in 5 μL TE. For the U1A and T-SL4 RNA, sites of 2'-*O*-adduct formation were identified by primer extension using a 5'-[^{32}P]-label primer as described,[4] with the exception that the extension reaction was incubated at 52 °C for 7 min. Dideoxy sequencing markers were generated using unmodified RNA. cDNA extension products were separated by gel electrophoresis and visualized using phosphorimaging. cDNA band intensities for the (+) and (–) reagent reactions were integrated using SAFA.[5] Primer extension products for the TAR RNA were analyzed by capillary electrophoresis, as described.[1] SHAPE reactivity profiles were obtained by subtracting the (–) reagent control from the (+) reaction intensities. SHAPE reactivities were normalized to a scale in which 1.0 is defined as the average intensity of highly reactive positions, after excluding outliers using a box plot analysis.[6] Two independent SHAPE experiments were performed for each RNA. Errors in reactivities are reported as the absolute differences between experiments, divided by the maximum expected reactivity value, 1.

NMR relaxation experiments. NMR measurements for the U1A target RNA have been reported previously.[7,8] For this work, we additionally measured relaxation parameters for the HIV-1 TAR and for most C and A nucleotides in the T-SL4 RNA. RNAs used for the NMR experiments were dialyzed into 10 mM potassium phosphate buffer containing 0.1 mM EDTA at pH 6.0; the final concentration of potassium ion (as the obligate RNA counterion) was ~50 mM. Data collection was executed on a Bruker Avance-500 spectrometer using a TXI triple resonance HCN probe in 99.9% D_2O at 25 °C for the U1A target and HIV-1 TAR RNAs. Relaxation data was collected at 40 °C for the T-SL4 RNA. T_1, $T_{1\rho}$ and Het-NOE experiments were recorded as a series of 2D NMR spectra, in which the relaxation delay (τ) was parametrically increased. Experiments were performed in constant-time mode, essentially as described.[7]

Generalized order parameter calculations. Quantitative analysis of the relaxation data, based on ^{13}C relaxation at the C1' ribose position, was conducted using the model-free approximation[9,10] using ModelFree 4.15.[11] In the analysis, we assumed that the chemical shift tensors are axially symmetric (η=0) and that the

symmetry axis of this tensor is collinear with the C–H bond. Model selection was as described[11] with some modifications.[12] Specifically, when neither model 2 nor 3 could be applied satisfactorily, models 4 and 5 were introduced. ModelFree parameters were fit to one of five models, in which the following parameters are varied: (1) S^2; (2) S^2 and an effective internal correlation time for fast motions (τ_e); (3) S^2 and the transverse relaxation exchange parameter (R_{ex}); (4) S^2, τ_e and R_{ex}; (5) the order parameters for shorter and longer time scale motion (S_f^2 and S_s^2, respectively). For all analyses, a CSA of 30 ppm was used. Errors in S^2 calculations were estimated to be ±5%.

Solvent accessibility calculations were performed using a 1.4 Å radius spherical probe.[13]

Additional information regarding NMR experiments with T-SL4. Relaxation studies with ^{13}C are complicated by the fact that homonuclear dipolar couplings between adjacent carbons contribute significantly to relaxation behavior in uniformly labeled samples. This interference increases with the square of the correlation time, becomes substantial for molecules that tumble slowly, and can introduce large errors in the model-free analysis. To obtain accurate data for the larger T-SL4 RNA, NMR data were therefore collected at 40 °C to increase the rate of rotational diffusion. $1–S^2$ values were obtained for about one-third of the positions in this RNA because (*i*) only A and C residues were isotopically labeled to reduce spectral overlap and (*ii*) spectral overlap nevertheless did obscure some positions in this RNA.

The correlation between SHAPE chemistry and S^2 for the T-SL4 RNA, although good (R = 0.73), is lower than for the TAR and U1A RNAs. Motions slower than overall rigid body molecular tumbling (6-20 ns) are invisible to NMR relaxation measurements. By raising the temperature, as required to reduce homonuclear effects, the correlation time was reduced from 15 to 10 ns and thus a more limited set of motions (only those below 10 ns) are probed for the T-SL4 RNA. An additional contributor to the poorer correlation reflects the increased noise in the NMR data due to the larger molecular mass of the T-SL4 RNA. The overall conclusion is that, while a clear correlation between SHAPE and S^2 is detected for the T-SL4 RNA, this RNA is close to the upper limit in size that can be routinely analyzed by NMR using uniformly labeled ^{13}C nucleotides.

References

(1) Wilkinson, K. A.; Gorelick, R. J.; Vasa, S. M.; Guex, N.; Rein, A.; Mathews, D. H.; Giddings, M. C.; Weeks, K. M. *PLoS Biology* **2008**, *6*, e96.

(2) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4659-4667.

(3) Mortimer, S. A.; Weeks, K. M. *J. Am. Chem. Soc.* **2007**, *129*, 4144-4145.

(4) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *Nature Protocols* **2006**, *1*, 1610-1616.

(5) Das, R.; Laederach, A.; Pearlman, S. M.; Herschlag, D.; Altman, R. B. *RNA* **2005**, *11*, 344-354.

(6) Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M. *submitted* **2008**.

(7) Shajani, Z.; Varani, G. *J. Mol. Biol.* **2005**, *349*, 699-715.

(8) Shajani, Z.; Drobny, G.; Varani, G. *Biochemistry* **2007**, *46*, 5875-5883.

(9) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.

(10) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559-4570.

(11) Mandel, A. M.; Akke, M.; Palmer, A. G. *J. Mol. Biol.* **1995**, *246*, 144-163.

(12) Chen, J.; Brooks, C. L.; Wright, P. E. *J. Biomol. NMR* **2004**, *29*, 243-257.

(13) Brunger, A. T.; Adams, P. D.; Clore, G. M.; Delano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta. Cryst.* **1998**, *D54*, 905-921.