Supporting Information

Jones et al. 10.1073/pnas.1319658111

SI Methods

Preparation of Nucleic Acids. Transactivation (TAR)/polyadenylation (polyA) A34U, primer binding site (PBS)/tRNA-like element (TLE), extended PBS/TLE, $tRNA^{Lys3}$, and Psi domain RNAs were transcribed in vitro from FokI-linearized plasmids using T7 RNA polymerase as described previously (1) and purified via denaturing PAGE. The 18-mer anti-PBS DNA oligomer (5′-GTCCCTGTTCGGGCGCCA-3′) was purchased from Integrated DNA Technologies. The A34U mutation in TAR/polyA, the dimerization initiation site (DIS) to GAGA tetraloop conversion in Psi, and the TLE extension in PBS/TLE were introduced via QuikChange site-directed mutagenesis (Agilent).

Sample homogeneity was assessed by size-exclusion chromatography (SEC) and native PAGE. Before SEC purification, RNAs (~300–400 μg) were folded in a buffer containing 50 mM Hepes (pH 7.4) and 1 mM $MgCl_2$ in a final volume of 100 μL. RNAs were heated at 80 °C for 2 min and then at 60 °C for 2 min, after which $MgCl_2$ was added, followed by incubation on ice for at least 30 min. For PBS/TLE–anti-PBS complexes, antiPBS DNA was included in 1.5 times molar excess over PBS, and complexes were incubated for an additional 5 min at 37 °C before incubation on ice. Folded RNAs were purified via SEC using a 24-mL Superdex 200 10/300 GL column (GE Healthcare) in running buffer [50 mM Hepes (pH 7.4), 150 mM NaCl, 1 mM glycerol, $MgCl_2$, 3% glycerol (wt/vol)] at a flow rate of 0.30 mL/min. Peak fractions containing the desired RNA species were pooled, concentrated to 60–100 μL (~1 mg/mL) using an Amicon 0.5-mL 10K molecular weight cutoff spin concentrator, and stored at 4 °C. An aliquot of the SEC buffer was used for sample dilutions and small-angle X-ray scattering (SAXS) buffer subtraction. On the day of shipment, RNAs were serial diluted 1:1 with matching SAXS buffer to yield 24-μL samples at three RNA concentrations: ~1 mg/mL, 0.5 mg/mL, and 0.25 mg/mL.

SAXS Data Acquisition and Analysis. Samples were shipped in 96-well plates (Axygen Scientific) at 4 °C to the 12.3.1 SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory (2, 3). Scattering data were collected as described previously (2). Data scaling, merging, and Guinier analysis to estimate the radius of gyration (R_g) and scattering intensity at zero scattering angle (I_0) were performed in PRIMUS (4). The presence of concentration-dependent effects was evaluated by comparing the three RNA concentrations for increasing R_g as a function of concentration or by a nonlinear curve at low scattering angles (3). Samples that displayed significant concentration-dependent effects were not analyzed further.

For well-behaved samples, Kratky analysis, a method of evaluating how well folded or globular a molecule is, was performed by graphing q vs. $q^2 \cdot I(q)$, where $I(q)$ is scattering intensity at a given scattering angle, q ($Å^{-1}$). Scattering curves at multiple concentrations were then scaled and merged into a single scattering curve for further analysis. $P(r)$ functions were calculated using GNOM (5) by varying D_{max} incrementally until the $P(r)$ smoothly decayed to 0 and fit the experimental data well. $P(r)$ function-derived R_g values were then compared with those obtained from Guinier analysis to ensure they were in agreement.

Envelopes were generated using DAMMIN by performing 10–20 runs in fast mode with no symmetry constraints and merging into a final envelope using ATSAS programs DAMSEL, DAMSUP, DAMAVER, DAMFILT, and DAMSTART (6). DAMMIN uses the manually assigned D_{max} value as the maximum dimension of a sphere of "dummy atoms" from which the final envelope is generated. To ensure choosing the correct D_{max} value, values up to 15 Å larger than the originally estimated value were screened at 3-Å intervals. The final D_{max} value was chosen based on the fit to the experimental data (χ^2 value) and reproducibility of the envelopes [normalized spatial discrepancy (NSD) values]. The resulting average and filtered envelopes were visualized in PyMOL (Schrödinger). Theoretical scattering curves were calculated from envelopes and compared with experimental scattering.

Molecular Dynamics Simulations. Starting structures were generated as an ensemble of 10 predicted structures using the RNAComposer Web server (7). RNAComposer uses a machine translation system to relate secondary and tertiary structure based on the RNA FRABASE database (8, 9). From the ensemble of 10 structures, initial structures were chosen for molecular dynamics simulations based on the conformation that best fit the SAXS data using CRYSOL (10) in batch mode to back-calculate scattering curves. Standard explicit solvent molecular dynamics (ESMD) simulations were carried out using the AmberTools13 and the GPU accelerated code of the AMBER12 program suite (11, 12). The AMBER ff99SB force field was used for all nucleosides. Constructs were first neutralized with sodium ions and then solvated in an octahedral box of TIP3P water (13), with a cutoff of 8.0 Å for nonbonded interactions.

Minimization and equilibration were each conducted in two steps. First, 2,000 total steps of minimization were performed with 1,000 steps of steepest descent and a particle-mesh Ewald (PME) implementation of constant volume periodic boundaries (14). The RNA was held fixed, with a positional restraint of 500 kcal/mol·$Å^2$. In the second step, 5,000 steps of minimization were performed with 1,000 steps of steepest descent, and the RNA was left unrestrained. The system was heated gradually from 0 to 300 K in the first step of equilibration for a total of 100 ps at a time step of 2 fs, while holding the RNA fixed with a weak positional restraint of 10 kcal/mol·$Å^2$. Although SAXS data collection was carried out at 277 K, the higher temperature used in the simulation allows for more efficient sampling of the conformational space available to the construct.

For the second equilibration and production simulations, the PME implementation was used for a constant pressure periodic boundary. The SHAKE algorithm was used to constrain all bonds involving hydrogen (15), and Langevin dynamics was used at a collision frequency of 1.0 ps^{-1} to control the temperature of the system. A second equilibration was used to relax the system for 900 ps under production conditions. Production simulations were performed for a total of 5 ns using the final equilibration coordinates.

All data analysis and statistical calculations were performed on the final 5 ns of each independent simulation. The use of Langevin dynamics adds the possibility of synchronization artifacts; thus, a random seed was used for all equilibration and production simulations (16, 17). A nonbonded cutoff of 8.0 Å was used for all simulation steps. After completion of production simulations, snapshot structures were generated every 1 ps and fit to the experimental data using CRYSOL in batch mode. The best snapshot was used as the starting structure for a further 5 ns of production simulation, as described above. This process was repeated until no further improvement in the χ^2 value was observed.

Model Refinement and Ensemble Generation. Once a reasonable structure was generated using multiple rounds of ESMD simulations that recapitulated the characteristics of the SAXS

scattering data ($\chi^2 < 2.5$), the structure was refined using a simulated annealing approach in Xplor-NIH (18), similar to that described previously (19). In brief, secondary structure restraints were imposed by introducing distance restraints for atoms located on the Watson–Crick faces of base-paired residues. Guinier R_g values and SAXS scattering data were used as additional potential energy terms (refined against 50 equally spaced data points with a scaling force of 1,000 kcal/mol). A total of 50 structures were calculated, and the 10 lowest-energy structures were chosen for further analysis. Structures were analyzed using modules within Xplor-NIH to calculate the final χ^2 and R_g values. The rmsd values were then calculated using the 10-member ensemble against averaged structural coordinates in VMD (20).

1. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15(21):8783–8798.
2. Hura GL, et al. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* 6(8):606–612.
3. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40(3):191–285.
4. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 36:1277–1282.
5. Semenyuk AV, Svergun DI (1991) GNOM—a program package for small-angle scattering data processing. *J Appl Cryst* 24:537–540.
6. Svergun DI (1999) Restoring low-resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879–2886.
7. Popenda M, et al. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40(14):e112.
8. Popenda M, Blazewicz M, Szachniuk M, Adamiak RW (2008) RNA FRABASE version 1.0: An engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res* 36(Database issue):D386–D391.
9. Popenda M, et al. (2010) RNA FRABASE 2.0: An advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinformatics* 11:231.
10. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 28:768–773.
11. Case DA, et al. (2012) AMBER 12 (University of California, San Francisco).
12. Götz AW, et al. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs, I: Generalized Born. *J Chem Theory Comput* 8(5):1542–1555.
13. Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
14. York DM, Darden TA, Pedersen LG (1993) The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J Chem Phys* 99:8345–8348.
15. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341.
16. Uberuaga BP, Anghel M, Voter AF (2004) Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J Chem Phys* 120(14):6363–6374.
17. Sindhikara DJ, Kim S, Voter AF, Roitberg AE (2009) Bad seeds sprout perilous dynamics: Stochastic thermostat-induced trajectory synchronization in biomolecules. *J Chem Theory Comput* 5:1624–1631.
18. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73.
19. Zuo X, et al. (2008) Global molecular structure and interfaces: Refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* 130(11): 3292–3293.
20. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.

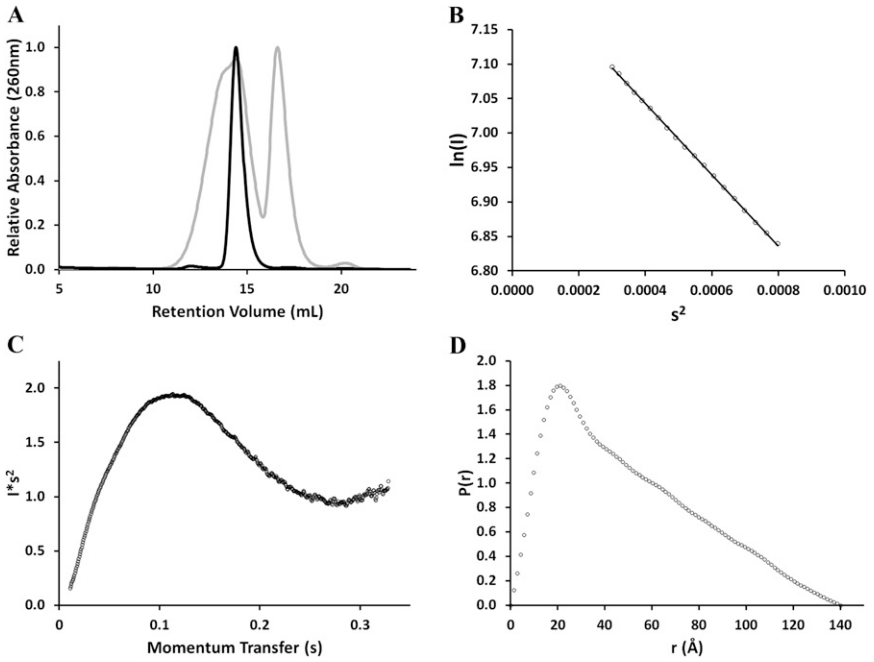


Fig. S1. Purification and analysis of the SAXS scattering curve for TAR/polyA. (*A*) SEC purification of WT (gray) and A34U (black) TAR/polyA constructs reveals that the A34U mutation alleviates sample heterogeneity. (*B–D*) The Guinier (*B*), Kratky (*C*), and *P*(*r*) (*D*) distribution plots calculated from the SAXS scattering curve indicate that the sample is nonaggregated, is well folded, and contains a majority of the electron pair distances expected in a mostly helical RNA, respectively.

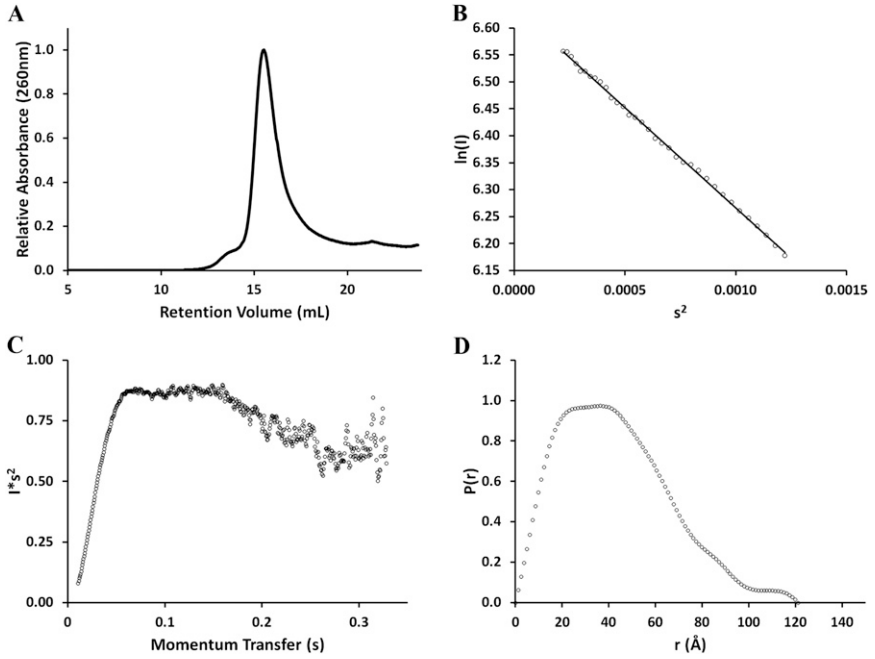
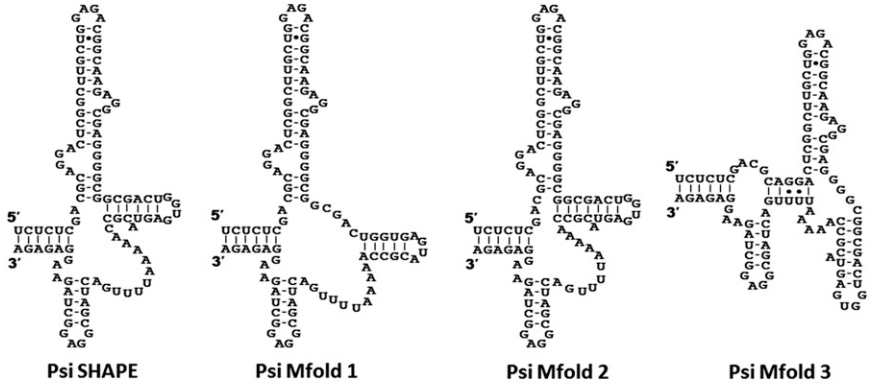


Fig. S2. Purification and analysis of the SAXS scattering curve for Psi. (*A*) SEC purification of the WT (black) Psi construct (DIS to GAGA tetraloop variant) reveals that the sample is homogeneous. (*B–D*) The Guinier (*B*), Kratky (*C*), and *P*(*r*) (*D*) distribution plots calculated from the SAXS scattering curve indicate that the sample is nonaggregated, is nonglobular, and contains a majority of electron pair distances between 20 and 45 Å, respectively.



Secondary Structure	R_G (Å)	χ^2	ΔG (kcal/mol)
Psi SHAPE	26.2 ± 0.8	9.07 ± 1.17	Not determined
Psi Mfold 1	34.8 ± 0.6	3.20 ± 0.61	-38.38
Psi Mfold 2	27.8 ± 0.9	7.04 ± 1.36	-38.75
Psi Mfold 3	28.1 ± 0.6	6.17 ± 0.79	-36.13

Fig. S3. Evaluation of alternative secondary structures for Psi. The secondary structure model predicted based on SHAPE probing (1) was compared with the three structures predicted using Mfold (2). Statistics calculated from 10-member ensembles of homology-modeled tertiary structures using RNAcomposer (3) and folding free energy values calculated by Mfold are given in the table below. All errors were calculated as SD.

1. Wilkinson KA, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6(4):e96.
2. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415.
3. Popenda M, et al. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40(14):e112.

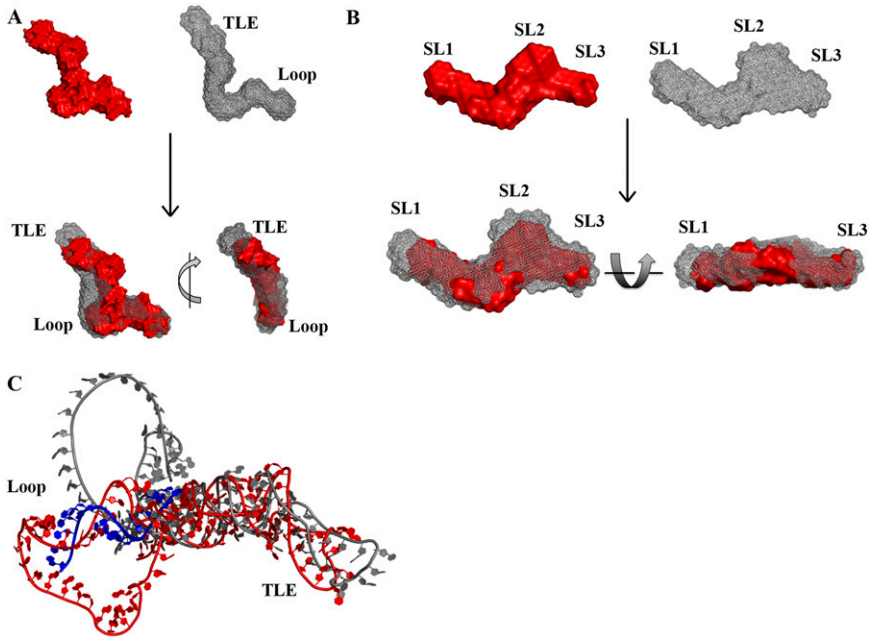


Fig. S4. Assessment of TLE helix elongation constructs and comparison of apo and anti-PBS–annealed PBS/TLE. (*A*) An overlay of the WT (red surface) and helix-extended (gray mesh) PBS/TLE domains allows for unambigious assignment of the TLE stem-loop region of the ab initio envelope. (*B*) Superposition of the WT (red surface) and SL1 helix-extended (gray mesh) Psi domains allows for unambigious assignment of the SL1 stem-loop region of the ab initio envelope. (*C*) Superposition of the apo (gray) and anti-PBS–annealed (red/blue) PBS/TLE models using the TLE as a register reveals the extent of conformational rearrangement within the largely single-stranded loop region.

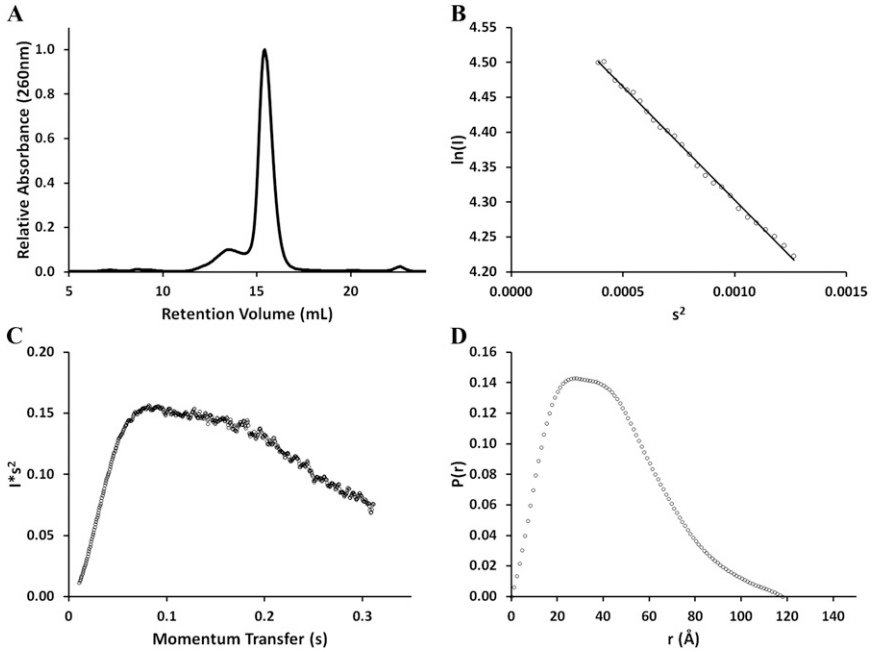
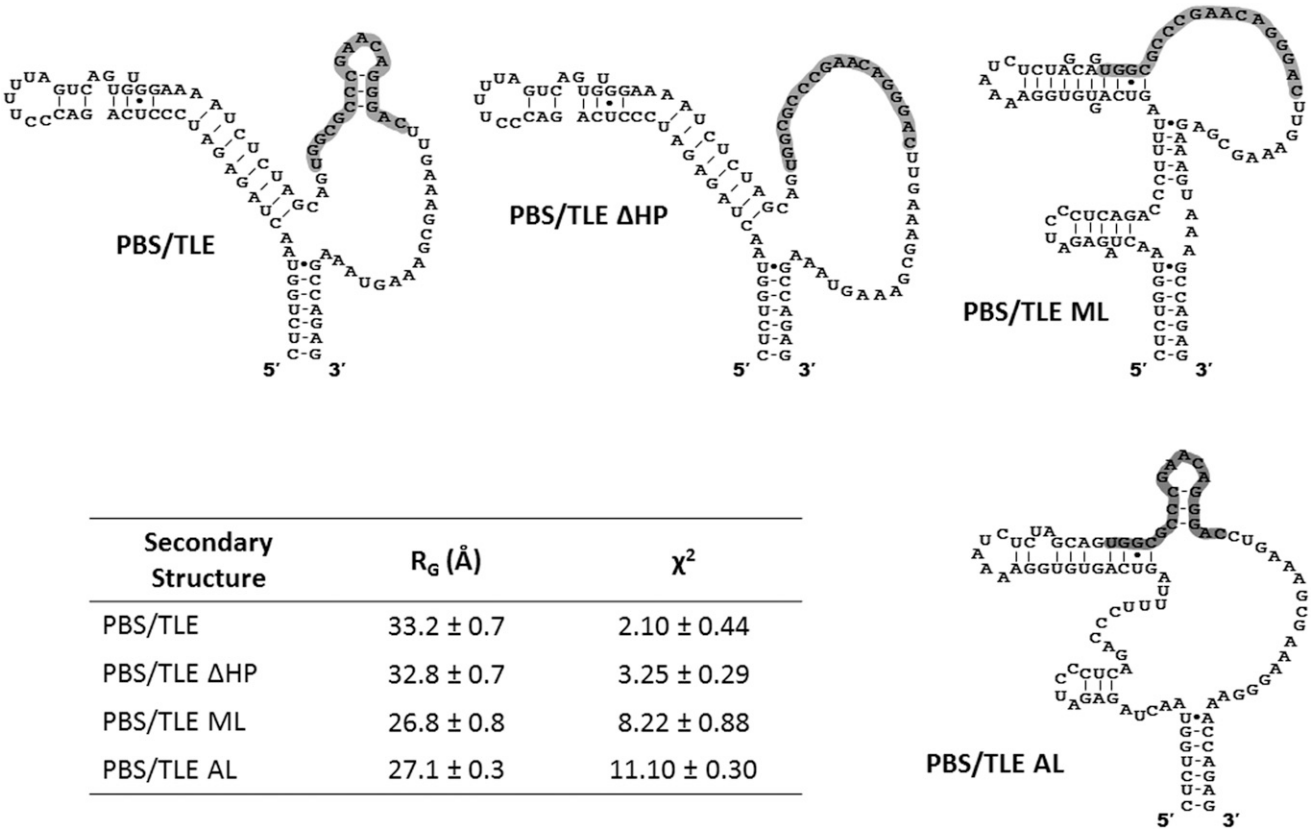


Fig. S5. Purification and analysis of the SAXS scattering curve for PBS/TLE. (*A*) SEC purification of the PBS/TLE construct reveals that the sample is homogeneous with a small aggregation impurity. (*B–D*) The Guinier (*B*), Kratky (*C*), and *P*(*r*) (*D*) distribution plots calculated from the SAXS scattering curve indicate that the sample is nonaggregated, is nonglobular, and contains a majority of electron pair distances between 20 and 40 Å, respectively.



Secondary Structure	R_G (Å)	χ^2
PBS/TLE	33.2 ± 0.7	2.10 ± 0.44
PBS/TLE ΔHP	32.8 ± 0.7	3.25 ± 0.29
PBS/TLE ML	26.8 ± 0.8	8.22 ± 0.88
PBS/TLE AL	27.1 ± 0.3	11.10 ± 0.30

Fig. S6. Evaluation of alternative secondary structures for PBS/TLE. The predicted secondary structures for PBS/TLE (1–4) are shown, with the 18-nt PBS sequence highlighted in gray. The proposed secondary structure models calculated from 10-member ensembles of homology-modeled tertiary structures using RNAcomposer (5) are given in the table below. All errors were calculated as SD.

1. Wilkinson KA, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6(4):e96.
2. Berkhout B (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog Nucleic Acid Res Mol Biol* 54:1–34.
3. Damgaard CK, Dyhr-Mikkelsen H, Kjems J (1998) Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res* 26(16):3667–3676.
4. Stephenson JD, et al. (2013) Three-dimensional RNA structure of the major HIV-1 packaging signal region. *Structure* 21(6):951–962.
5. Popenda M, et al. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40(14):e112.

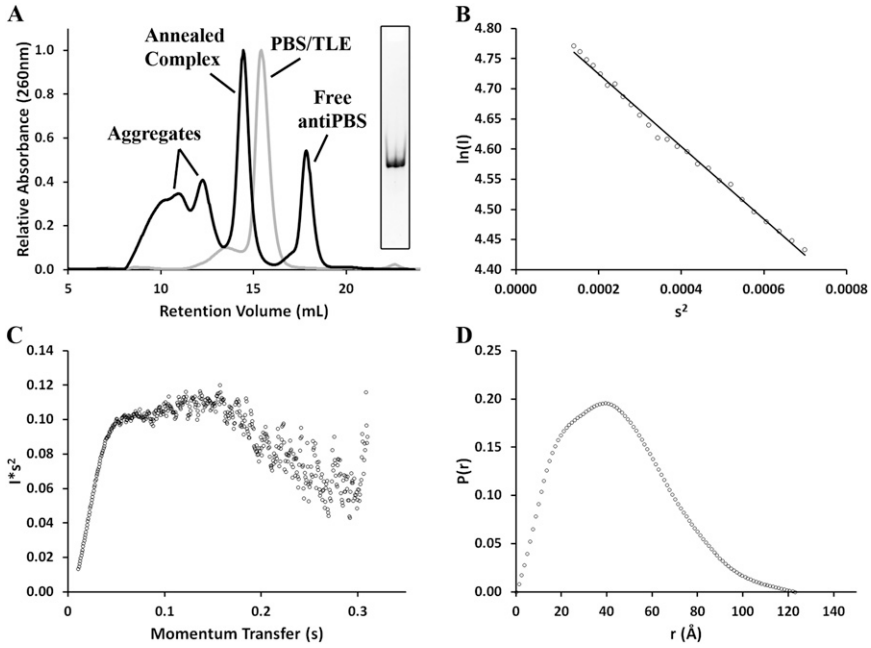


Fig. S7. Purification and analysis of the SAXS scattering curve for the PBS/TLE:anti-PBS annealed complex. (*A*) SEC purification of apo (gray) and anti-PBS–annealed (black) PBS/TLE constructs reveals that the excess anti-PBS used completely eliminates free PBS/TLE in the annealed complex sample. (*Inset*) A native polyacrylamide gel confirming sample homogeneity before SAXS analysis. (*B–D*) The Guinier (*B*), Kratky (*C*), and *P*(*r*) (*D*) distribution plots calculated from the SAXS scattering curve indicate that the sample is nonaggregated, is nonglobular, and contains a majority of electron pair distances between 20 and 60 Å, respectively.

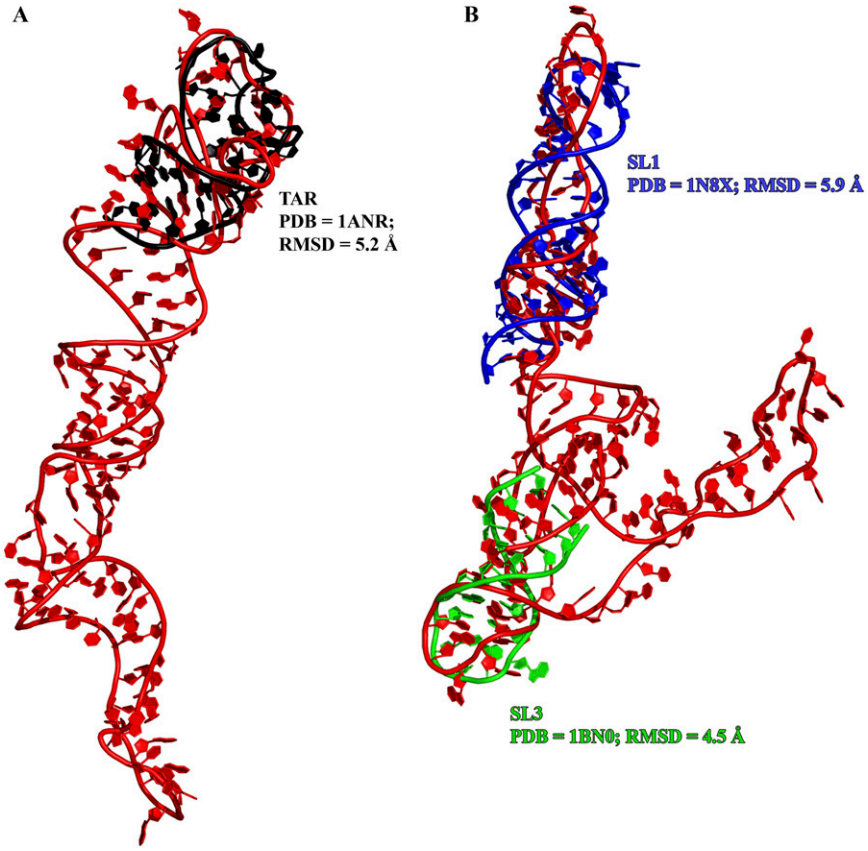


Fig. S8. Comparison of TAR/polyA and Psi models with NMR structures. (*A*) Superimposition of the previously determined NMR structure of the apical stem and loop of TAR (black) (1) onto the corresponding region of our SAXS-derived structural model (red) revealed only minor differences. (*B*) The NMR structures of SL1 (blue) (2) and SL3 (green) (3) were aligned with our SAXS-derived model of Psi, indicating a high degree of conformational similarity.

1. Aboul-ela F, Karn J, Varani G (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res* 24(20):3974–3981.
2. Greatorex J, Gallego J, Varani G, Lever A (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J Mol Biol* 322(3):543–557.
3. Pappalardo L, Kerwood DJ, Pelczer I, Borer PN (1998) Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J Mol Biol* 282(4):801–818.

Table S1. Statistics calculated from analysis of the SAXS scattering curve, construction of ab intio envelopes using DAMMIN, and the final molecular modeling ensemble

RNA	TAR/polyA (1-104; A34U)	Psi (228-334; DIS GAGA)	PBS/TLE (125-223; WT)	PBS/TLE:anti-PBS
Scattering curve				
R_g, Å, Guinier	39.1 ± 0.5	33.8 ± 0.4	34.1 ± 0.4	33.5 ± 0.3
R_g, Å, $P(r)$	41.1	34.5	33.9	35.3
D_{max}, Å	140	121	118	123
DAMMIN (n = 20)				
χ^2	1.15 ± 0.16	1.14 ± 0.02	0.90 ± 0.02	1.30 ± 0.01
NSD	0.67 ± 0.03	0.86 ± 0.05	0.76 ± 0.05	0.84 ± 0.08
Modeling (n = 10)				
R_g, Å	40.8 ± 0.1	34.0 ± 0.1	33.7 ± 0.1	34.6 ± 0.1
χ^2	1.22 ± 0.13	0.99 ± 0.03	0.61 ± 0.01	1.14 ± 0.01
rmsd, backbone	2.60 ± 0.34	0.84 ± 0.11	1.61 ± 0.20	1.53 ± 0.24
rmsd, total	2.68 ± 0.32	0.94 ± 0.09	1.81 ± 0.19	1.68 ± 0.24
MolProbity core	0.30 (99th)	0.00 (100th)	0.00 (100th)	0.00 (100th)

All DAMMIN and modeling statistics are calculated as the average of 20 and 10 models, respectively. MolProbity scores were calculated using the online server (1) with the associated percentile in parenthesis. All errors are reported as SD.

1. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.