Supporting Information

Deigan *et al.* 10.1073/pnas.0806929106

Materials and Methods

Isolation of *Escherichia coli* Protoplasts. Cell cultures (*E. coli* DH5α) were grown to midlog phase (A600 = 0.8) under shaking in Luria broth, aliquoted (1.5 ml), collected by centrifugation (5 min, 4 °C, 17,000 × *g*), and resuspended in 1 ml of 15 mM Tris (pH 8.0), 450 mM sucrose, and 8 mM EDTA. Lysozyme (400 μg) was added and the solution was incubated at 22 °C for 5 min, and on ice for 10 min. Protoplasts were collected by centrifugation (5 min, 4 °C, 5,000 × *g*).

Recovery of Native RNA. Protoplast pellets were resuspended in 120 μl of 50 mM Hepes (pH 8.0), 200 mM NaCl, 5 mM $MgCl_2$ and 1.5% (wt/vol) SDS and incubated at 22 °C for 5 min and on ice for 5 min. SDS was precipitated by adding 30 μl of 50 mM Hepes (pH 8.0), 1 M potassium acetate, and 5 mM $MgCl_2$. The precipitate was collected and discarded by centrifugation (5 min, 4 °C, 17,000 × *g*) and the buffer of the RNA-containing solution exchanged by gel filtration (G-50, 400-μl column) preequilibrated in 1× folding buffer [50 mM Hepes (pH 8.0), 200 mM potassium acetate (pH 8.0), and 5 mM $MgCl_2$]. RNA was eluted in the same solution. The RNA-containing eluent was extracted 3 times with phenol [(pH 8.0):chloroform:isoamyl alcohol; 25:24:1] and 3 times with chloroform; 1.5 ml of bacterial culture yielded ≈25 μg of cellular RNA.

Recovery of RNA for Sequencing. Protoplast pellets were lysed by resuspension in 250 μl of 10 mM Tris·HCl (pH 8.0), 10 mM NaCl, 1 mM sodium citrate, and 1.5% (wt/vol) SDS, and incubation at 37 °C for 5 min and on ice for 5 min. Saturated NaCl (0.1 ml) was added and the solution was incubated on ice for an additional 10 min. The solution was centrifuged (10 min, 4 °C, 17,000 × *g*) and the precipitate discarded. The resulting solution was extracted as above. RNA was recovered by precipitation with 0.1 vol of sodium acetate (3 M, pH 6.0) and 2.5 vol of ethanol. RNA pellets were washed 3 times with 70% ethanol and resuspended in 50 μl of deionized water (final concentration of total RNA ≈1 mg/ml).

SHAPE on Total *E. coli* RNA. RNA (≈25 μg) in 1× folding buffer (≈500 μl final volume) was divided into 2 equal aliquots, incubated at 37 °C for 30 min, and treated with 1/10 vol of 1-methyl-7-nitro-isatoic anhydride (1M7) (1) in DMSO (1M7, 60 mM) or neat DMSO. Under these conditions, ≈1 in 300 nt are modified. Reactions were incubated for 3 min and RNA was recovered by ethanol precipitation (see above) and resuspended at 1 mg/ml in 1× TE.

Primer Extension of *E. coli* rRNA. DNA primers were designed to span the entirety of both 16S and 23S RNA; 4 primers were used to analyze 16S RNA and 10 primers for 23S RNA. Sequences of the 4 DNA primers used in the analysis of 16S RNA, named according to the most 5′ nucleotide of the rRNA sequence to which they anneal, were: 559, 5′-CTT TAC GCC CAG TAA TT-3′; 947, 5′-TCG AAT TAA ACC ACA TGC-3′; 1452, 5′-GTA AGC GCC CTC CCG-3′; and 1492, 5′-CCT ACG GTT ACC TTG TTA CGA CTT-3′. Sequences of the ten primers used to analyze 23*S* RNA were: 367, 5′-GTC CCG CCC TAC TCA TC-3′; 728, 5′-CAA CAT TAC TCG GTT CGG TCC-3′; 1109, 5′-CTT CCG CGC AGG CCG ACT CG-3′; 1514, 5′-GCC TCG TCA TCA CGC CTC-3′; 1832, 5′-CCT TCC GGC ACC GGG CAG G-3′; 1909, 5′-CCT TAG GAC CGT TAT AGT TAC G-3′; 2117, 5′-CTA TAG TAA AGG TTC ACG GGG-3′; 2421, 5′-GTA CCT TTT ATC CGT TGA GC-3′; 2581, 5′-ATG TGA TGA GCC GAC ATC G-3′; and 2888, 5′-AAG GTT AAG CCT CAC GG-3′. All primers contained 5′ amino C6 modifiers (H_2N-C_6H_{12}-*p*-DNA). All 23S primers and the 1,492 16S primer were labeled with 5-FAM, 6-JOE, 6-TAMRA, or 5-ROX dyes; other 16S primers were labeled with 6-FAM, VIC, NED and PET dyes. Gel-purified fluorescently labeled DNA primer [2.5 pmol, 5′-labels 6-JOE or 5-FAM (Anaspec); VIC or 6-FAM, (Applied Biosystems)] was added to the appropriate RNA generated above (5.0 μg, in 1× TE). Typically 6-JOE or VIC was used for the (+) reagent channel and 5- or 6-FAM was used for the (−) reagent channel. The RNA-primer solution was diluted to 6.5 μl with water and incubated at 65 °C for 5 min, 45 °C for 5 min, and placed on ice. Primer extension was initiated by addition of enzyme mix [3 μl; 250 mM KCl; 167 mM Tris·HCl (pH 8.3), 1.67 mM each deoxynucleotide, 10 mM $MgCl_2$; 52 °C, 1 min], SuperScript III (0.5 μl, 100 units, Invitrogen), and incubation at 45 °C for 1 min, 52 °C for 7 min, and 65 °C for 5 min. Sequencing reactions were identical, except that they used unmodified nonnative rRNA (4.8 μg), the RNA-primer solution was diluted to 6.0 μl in deionized water, and 0.5 μl of a 10 mM ddNTP solution was added immediately before SuperScript III. Primers used for sequencing were typically labeled with either 6-TAMRA and 5-ROX (Anaspec) or NED and PET (Applied Biosystems) fluorophores. Appropriate reactions [(+) and (−) reagent, 2 sequencing extensions; each extension by using the same primer sequence but labeled with a different fluorophore] were quenched by precipitation with ethanol, washed 3 times with 70% ethanol, dried under vacuum, and redissolved in 10 μl of deionized formamide. The cDNA samples were resolved on an Applied Biosystems 3130 capillary electrophoresis DNA sequencer.

SHAPE Data Processing. Raw electropherograms of fluorescence intensity versus elution time (Fig. 2*B*) were analyzed by using ShapeFinder (2, 3). Data processing steps included baseline and mobility shift corrections and a correction for signal decay as a function of primer extension length. Peaks for the (+) and (−) reagent channels were aligned with each other and also with the RNA sequence. The area under each peak was quantified by whole-channel Gaussian integration. After subtracting background, SHAPE reactivities from each primer read were placed on a normalized scale by dividing by the average intensity of the 10% most highly reactive nucleotides, after first excluding highly reactive outliers. Outliers in each primer dataset were identified by using a model-free box plot analysis as reactivities >1.5× the interquartile range (4). For small RNA datasets (< ≈100 nt), the maximum number of outliners is capped at 5%. Use of model-free statistics is important because SHAPE reactivities do not exhibit a normal distribution. These calculations place all SHAPE reactivities on a scale spanning 0 to ≈2 (Fig. 2*C*). Reactivity data for each primer were processed and normalized independently. Final SHAPE data for different, overlapping, primers for the 16S and 23S rRNAs consistently fell on the same scale, without the need for additional normalization.

SHAPE Analysis of $tRNA^{Asp}$, HCV IRES Domain II, and bI3 Group I Intron P546 Domain. Data for $tRNA^{Asp}$ and the bI3 intron were reported previously (5, 6). Data for the HCV IRES will be reported elsewhere. Accepted target structures for these RNAs were taken from refs. 7–9.

Incorporation of SHAPE Pseudo-Free Energy Change Terms into a Dynamic Programming Algorithm. All structure calculations were performed by using RNAstructure (10). The maximum allowed base-pairing distance for all structure calculations was 600 nt. $\Delta G_{SHAPE}(i)$ is added to the free energy change for each nucleotide i in a base-pairing stack (11). For the total folding free energy change, nucleotide i therefore contributes ΔG_{SHAPE} if it is involved in a stack at the end of helix, $2\times$ ΔG_{SHAPE} if it is paired and adjacent to 2 stacked pairs or a single bulged nucleotide, or zero if it is single-stranded.

Calculation of Prediction Sensitivity, PPV, and Fraction Correct Helices. When calculating sensitivity and PPV, base pairs between nucleotides i and j are considered correctly predicted pairs if the known structure contains a pair between i and j, $i + 1$ or $i - 1$ and j, or i and $j + 1$ or $j - 1$. We allow these slipped pairs because of the difficulty in conclusively determining the exact pairing and because these predictions are qualitatively consistent with correct pairs (12). We define a helix as a continuous stack of 3 or more canonical base pairs interrupted by no more than a single nucleotide bulge. By this definition, there are 69 helices in *E. coli* 16S rRNA. Helices were considered correct if >50% of the constituent base pairs were predicted correctly. Helices were not included in the calculation if >50% of their nucleotides lie in omit regions or have no SHAPE data. For the calculation of prediction accuracy after allowing local refolding (target 3), we required (*i*) single additional base pairs to stack on an existing helix (for example, the 587–754 base pair; Fig. 5) and (*ii*) multiple base pairs be separated from a phylogenetically supported helix by a bulge of 2 nucleotides or fewer (for example, the 980 loop; Fig. 5).

1. Mortimer SA, Weeks KM (2007) A fast acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129:4144–4145.
2. 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:e96.
3. Vasa SM, *et al.* (2008) ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14:1979–1990.
4. Chernick MR, Friis RH (2003) *Introductory Biostatistics for the Health Sciences* (Wiley, New York), pp 58–60.
5. Wilkinson KA, Merino EJ, Weeks KM (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts. *J Am Chem Soc* 127:4659–4667.
6. Duncan CDS, Weeks KM (2008) SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* 47:8504–8513.
7. Westhof E, Dumas P, Moras D (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J Mol Biol* 184:119–145.
8. Lukavsky PJ, Kim I, Otto GA, Puglisi JD (2003) Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 10:1033–8.
9. Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610.
10. Mathews DH, *et al.* (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 101:7287–7292.
11. Xia T, *et al.* (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719–14735.
12. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288:911–940.
13. Moazed D, Stern S, Noller HF (1986) Rapid chemical probing of conformation in 16*S* ribosomal RNA and 30*S* ribosomal subunits using primer extension. *J Mol Biol* 187:399–416.
14. Cannone JJ, *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.
15. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127:4223–4231.

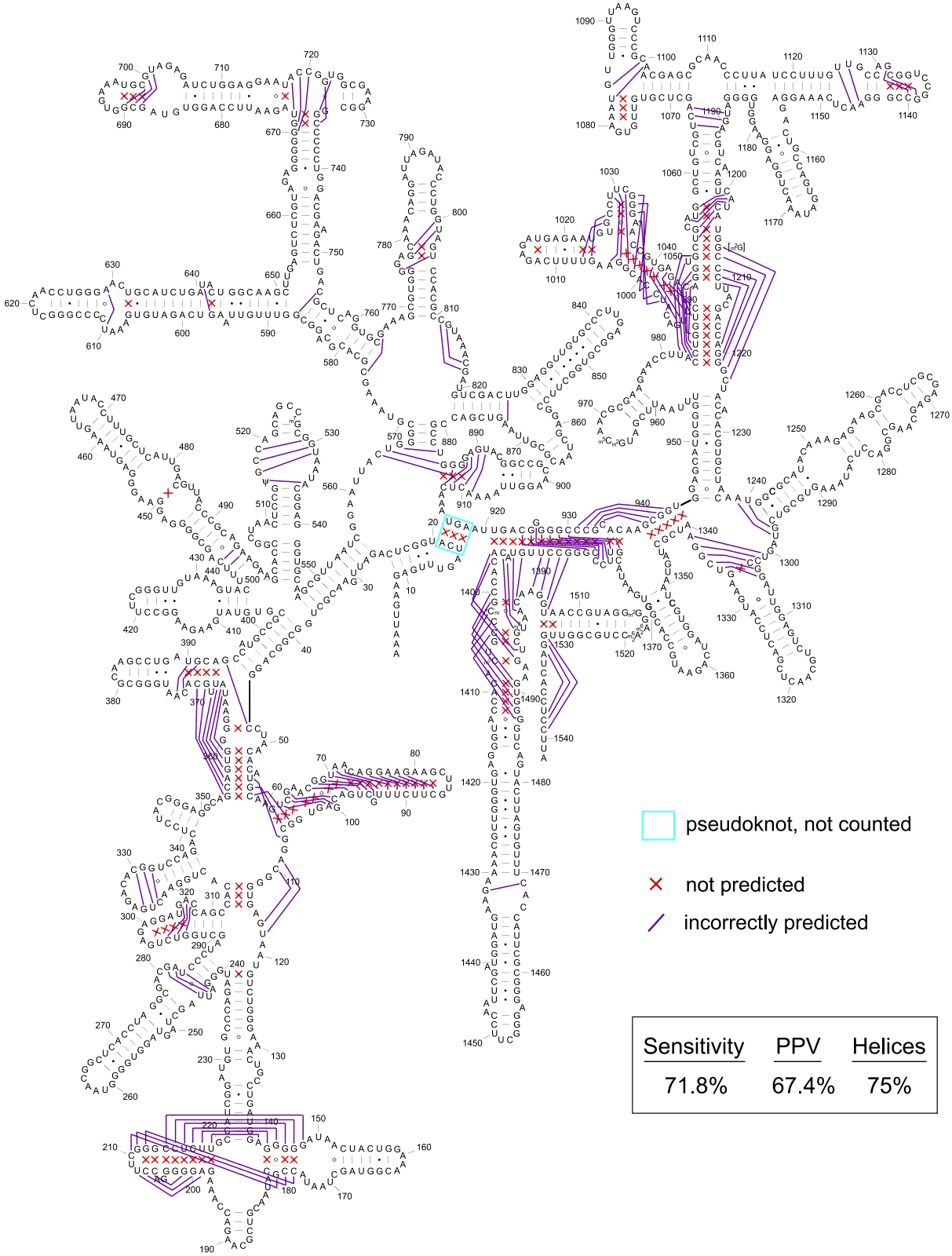


Fig. S1. Accuracy of secondary structure prediction for *E. coli* 16S rRNA using conventional chemical reagents. Reagent data included DMS, kethoxal and CMCT (13). Nucleotides judged to show strong or moderate reactivity toward chemical probes (13) were prohibited from forming Watson–Crick base pairs at internal positions in a helix unless they are adjacent to a GU pair. Missed base pairs are represented by red x's; incorrectly predicted base pairs are represented by purple lines.

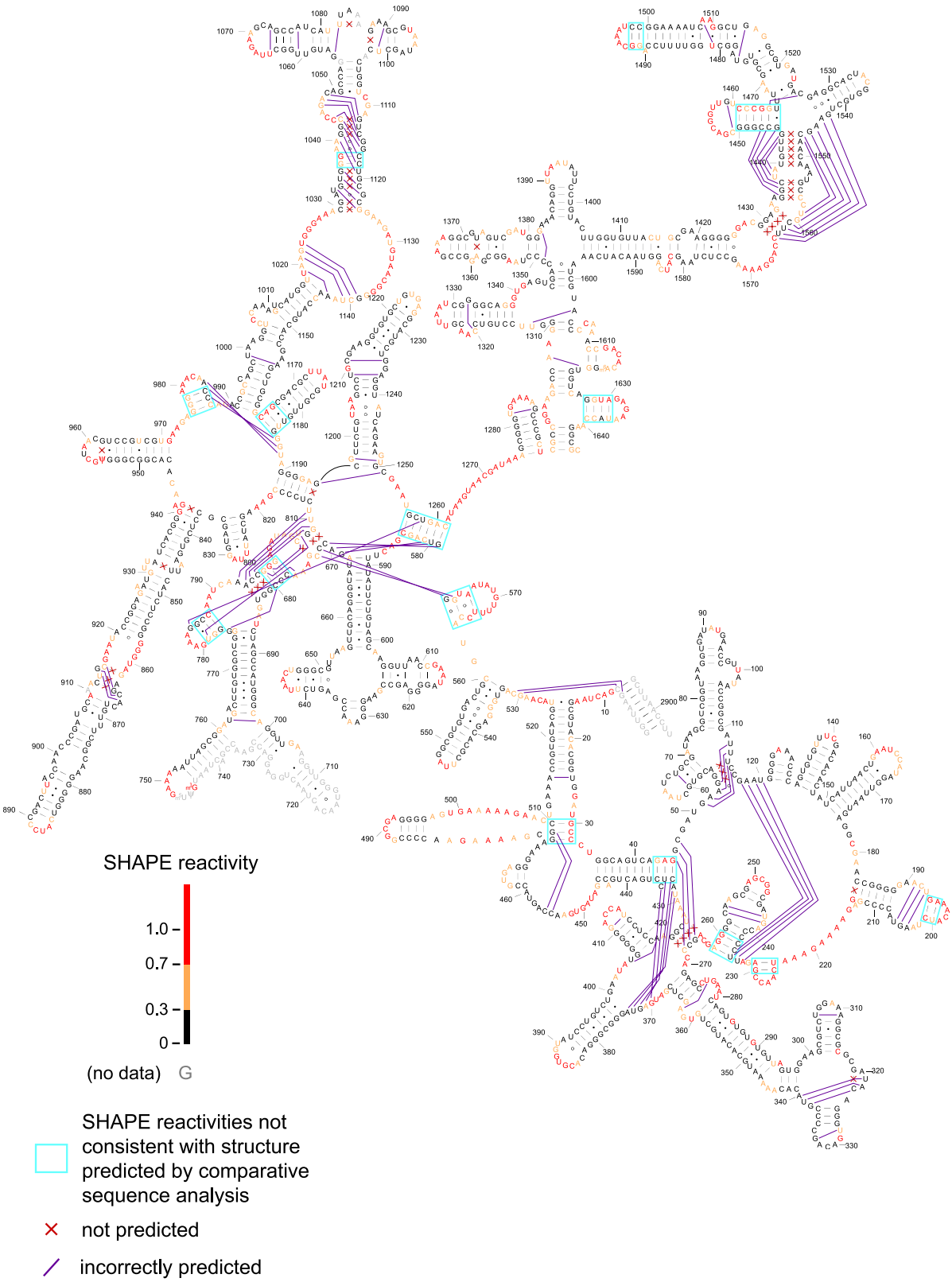
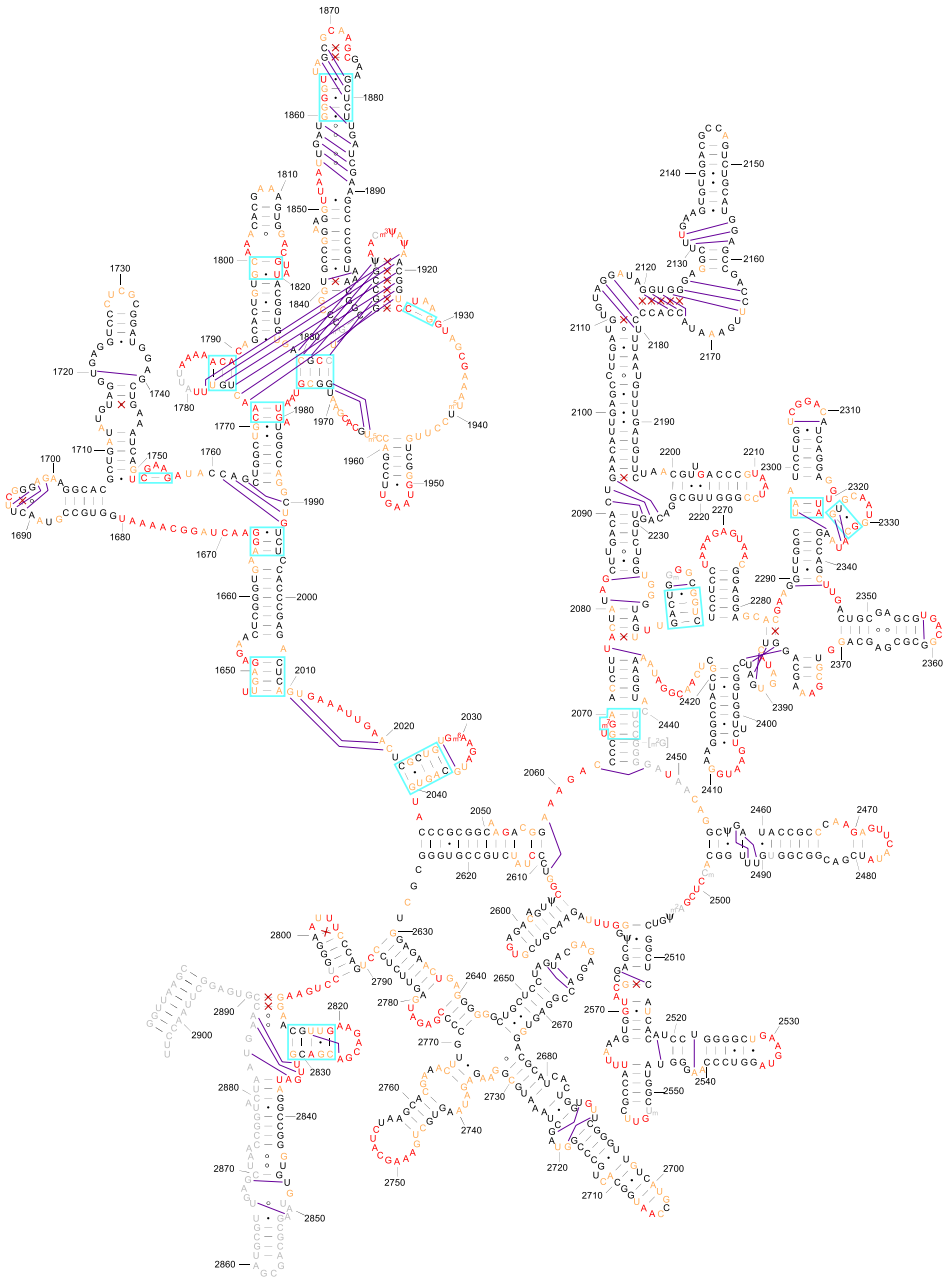


Fig. S2. Secondary structure prediction for the 5′ (*A*) and 3′ (*B*) regions of *E. coli* 23S RNA using SHAPE constraints. SHAPE data are superimposed on the structure determined by comparative sequence analysis (14); nucleotides are colored by their SHAPE reactivities. Nucleotides with no data are gray. Missed base pairs are represented by red x's; incorrectly predicted base pairs are represented by purple lines. ΔG_{SHAPE} parameters were an intercept of −0.8 kcal/mol and a slope of 2.6 kcal/mol; the maximum base pairing distance was 600 nt.



Prediction accuracies for 23*S* rRNA as a function of SHAPE constraints

Constraints	Target	Base Pairs	
		Sensitivity	PPV
none		65.4	60.6
SHAPE	comparative sequence model	78.7	76.7
SHAPE	with omit regions	89.8	77.6

Fig. S2. continued.

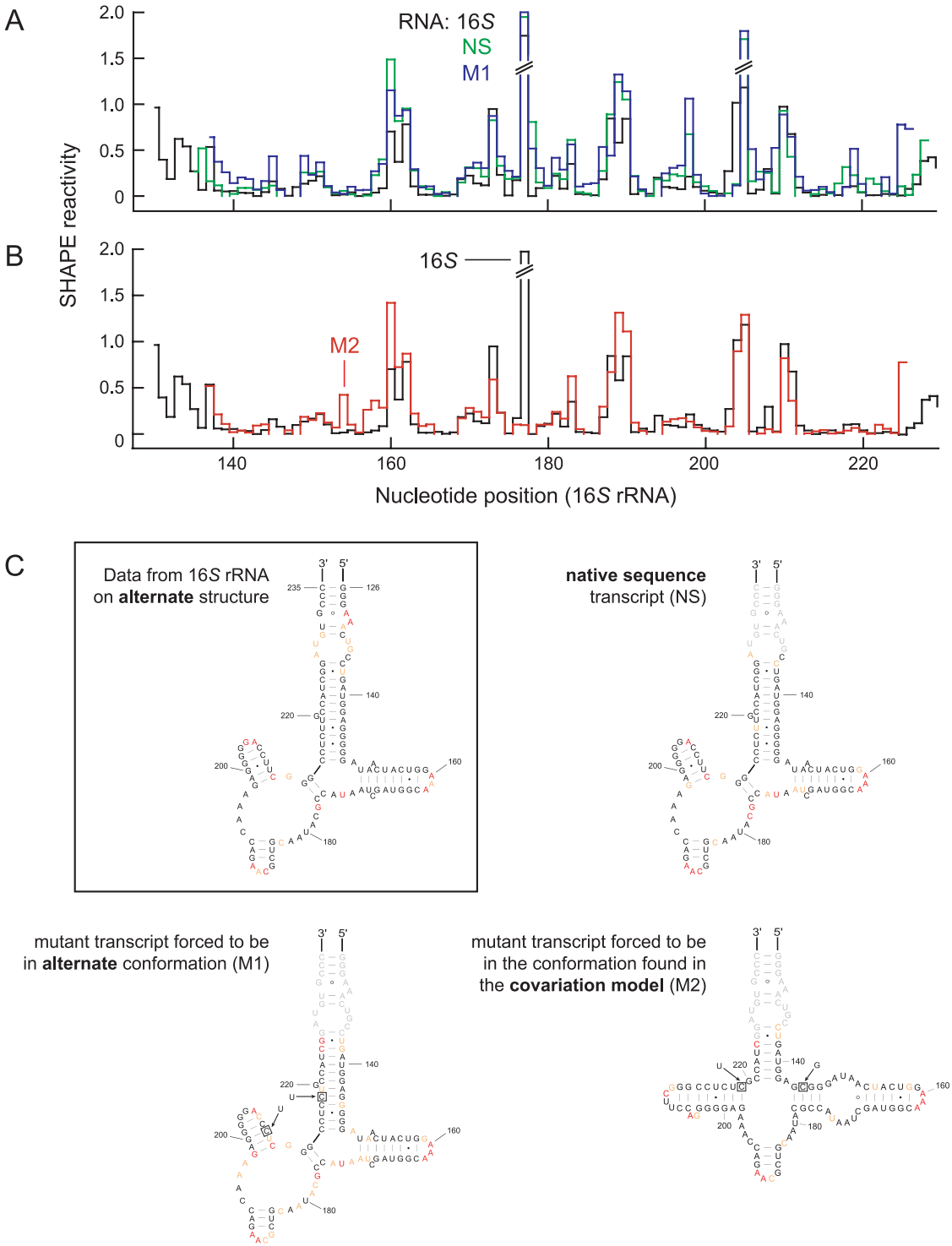


Fig. S3. Comparative SHAPE analysis of 16S RNA at the four-helix junction at nucleotides 139–224. Two possible conformations for this region were tested using in vitro transcripts and comparison with the structure in the intact 16S rRNA. The in vitro transcripts (spanning nucleotides 126–235) were imbedded within 5′ and 3′ structure cassette sequences (15) to facilitate analysis by SHAPE. Three RNA transcripts were tested: (*i*) a native sequence RNA (NS), (*ii*) a mutant that strengthens the proposed alternate structure (M1), and (*iii*) a mutant that strengthens the conventional structure proposed based on sequence covariation (M2). (*A*) Integrated SHAPE data for 16S RNA, the NS RNA, and M1. Overall SHAPE reactivities for these three RNAs are similar. (*B*) Integrated SHAPE data for 16S RNA and M2. There are clear differences in SHAPE reactivities between the two RNAs, most notably at positions 158 and 177. (*C*) SHAPE data superimposed on proposed alternative and covariation-based secondary structures. Boxes indicate mutations; arrows point from the native sequence nucleotide to the mutated nucleotide. In sum, these experiments strongly support the interpretation that the 3-helix junction in deproteinized 16S rRNA does not fold to the structure derived from covariation analysis but, instead, folds to the alternate structure emphasized by the box in *C*.

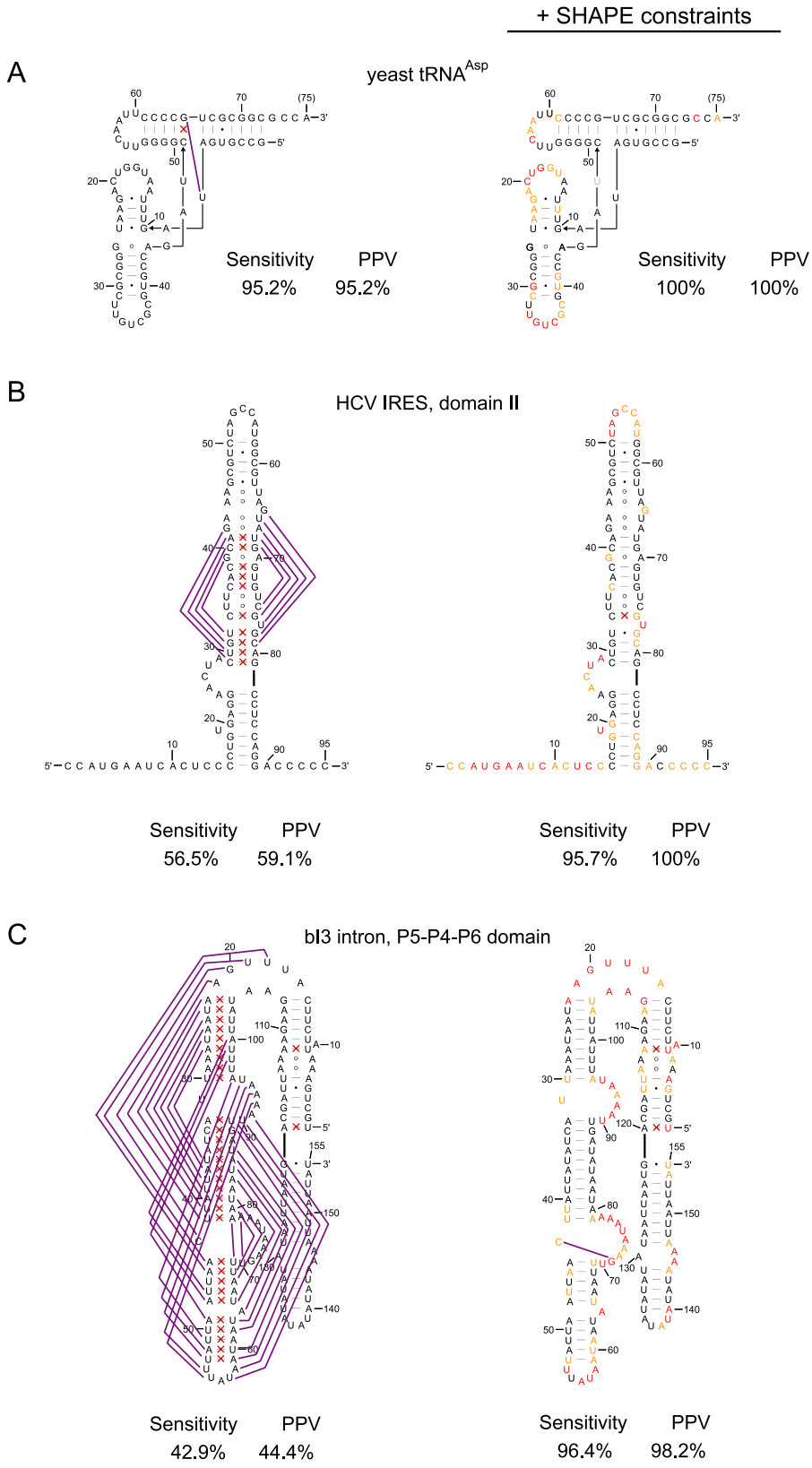


Fig. S4. Accuracy of SHAPE-directed prediction for 3 small, nonribosomal, RNAs using ΔG_{SHAPE} parameters (of −0.8, 2.6 kcal/mol). *Left* and *Right* illustrate prediction accuracies in the absence and presence of SHAPE constraints, respectively.