Structure, *19*

Supplemental Information

Crystal Structure of the HCV IRES Central Domain

Reveals Strategy for Start-Codon Positioning

Katherine E. Berry, Shruti Waghray, Stefanie A. Mortimer, Yun Bai, and Jennifer A. Doudna

Inventory of Supplemental Information

Figure S1, relates to Figure 1. This supplemental figure presents additional supporting details to the overall crystal structure presented in Figure 1, including experimental and 2Fo-Fc electron density maps, crystal contacts between molecules, atomic displacement parameters, a surface representation of the structure, and a comparison to a previously determined computational model.

Figure S2, relates to Figure 4. This supplemental figure presents the details of the SHAPE experiments summarized in Figure 4. The figure provides a quantitative perspective of the results throughout the entire IRES and gives the reader a sense of experimental noise.

Figure S3, relates to Figure 5. This supplemental figure compares the cryo-EMguided model of the HCV IRES bound to 40S subunit to the positions of ribosomal proteins known to crosslink to the central domain of the IRES.

Supplemental Experimental Procedures. This presents additional details regarding construct design, RNA purification, modeling, and SHAPE experiments. All of these experiments are described in some detail in the main Experimental Procedures section, but this is far more detailed.

SUPPLEMENTAL FIGURES

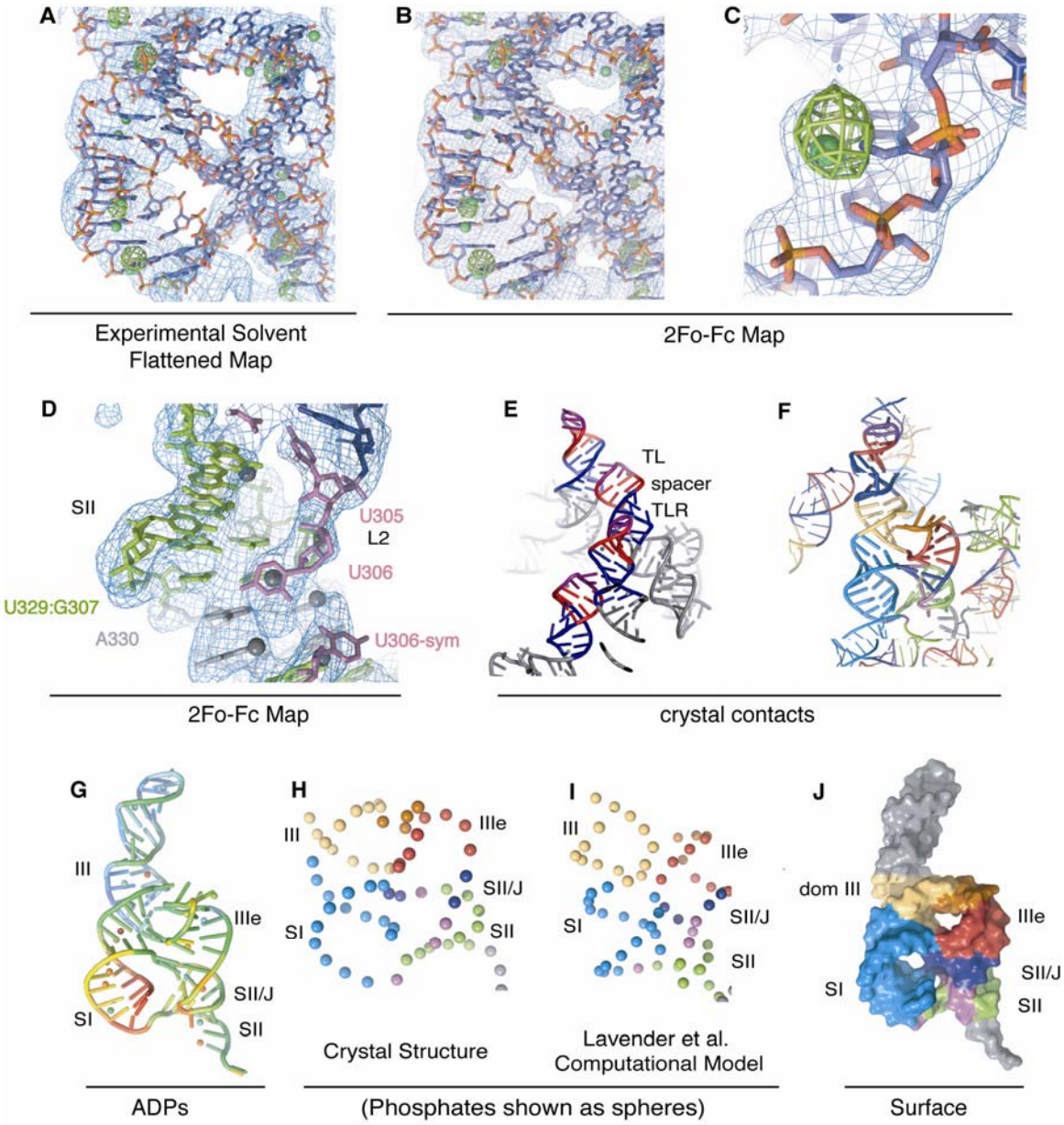

Figure S1, related to Figure 1. (A) Experimental electron density map (blue) after density modification, determined by MAD using cocrystallized nickel (II) ions. An anomalous difference map (green) shows strong sites of nickel (II) occupancy, along the major groove of helical regions. (B) $2F_o$-F_c map after refinement. (C) Close-up view of $2F_o$-F_c map after refinement and original nickel anomalous difference map (green), showing that nickel density is within the inner-sphere coordination distance to N7 positions of guanines, one of the most Lewis basic functional groups within RNA. (D) $2F_o$-F_c electron density map surrounding L2

and the terminus of SII. Nucleotides in the main model and U306 in the symmetry-related model are indicated. Blue maps are contoured at 1 σ. Green nickel anomalous density maps are contoured at 5 σ. (E) The interaction between TLs and TLRs of neighboring molecules is shown with the crystallization module colored as follows: tetraloop (TL) in purple, spacer in red and tetraloop receptor (TLR) in blue. These interactions occur in a head-to-tail fashion and propagate throughout the crystal, providing the four-fold screw axis present in the $P4_12_12$ space group. (F) Additional crystal contacts between SI and SI, SII and SII and the 3′ overhang and IIIe loop are shown between a central molecule (thick cartoon) and surrounding molecules (thin cartoons). The pseudoknot domain is colored as in Fig. 1. (G) Atomic displacement factors (B-factors) of the refined model are displayed on the crystal structure; the spectrum goes from low B-factors (blue, minimum=90 $Å^2$) to high B-factors (red, maximum=220 $Å^2$). Simplified structural representations of (H) the pseudoknot crystal structure and (I) a computational model from the Weeks laboratory (Lavender et al., 2010). The Lavender *et al.* model was determined using pseudoatoms; therefore only the phosphates of each structure are shown as spheres, for the sake of comparison. Secondary structure elements are colored as in Fig. 1. (J) Surface representation of the pseudoknot-domain crystal structure, with secondary structure elements colored as in Fig. 1.

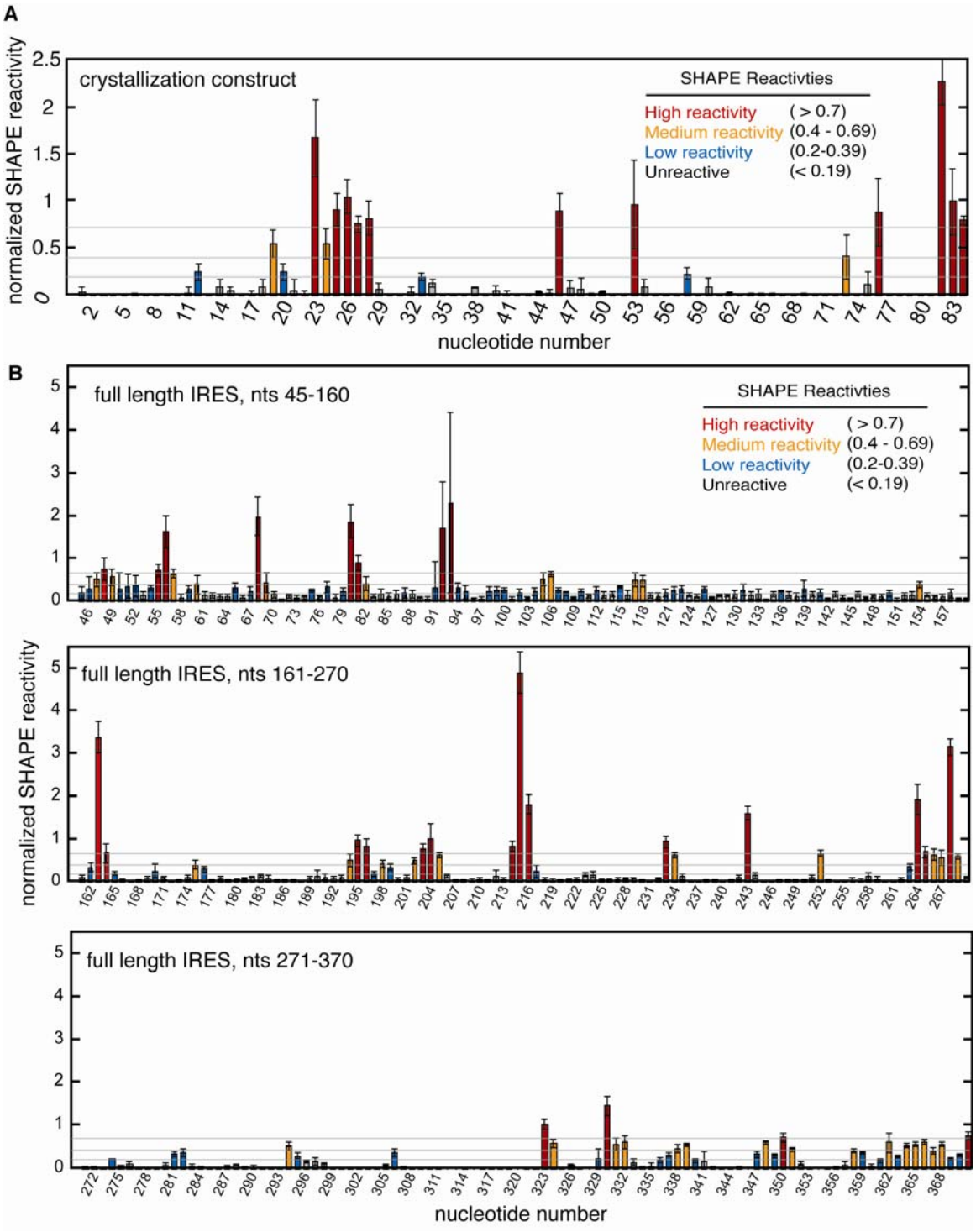

Figure S2, related to Figure 4. SHAPE reactivities of (A) the crystallization construct, (B) full-length IRES, and (C) full-length IRES in the presence of 40S ribosomal subunits. The average SHAPE reactivity of each nucleotide between three replicate experiments is shown. Bars are colored based on reactivity range (see legend) and light grey horizontal lines indicate the thresholds for these

groupings. Error bars represent the standard deviation between these three replicates. SHAPE reactivities mapped onto secondary structures. SHAPE modification reactivities of individual nucleotides from (D) the pseudoknot crystallization construct, (E) the full-length IRES and (F) the full-length IRES in the presence of 40S subunits superimposed on the secondary structure diagrams. Nucleotides for which the SHAPE reactivity was too inconsistent between replicates to estimate its true reactivity are shown in grey. The secondary structure of domain II is taken from Zhao and Wimmer, 2001, and is consistent with the NMR structure of domain II (Lukavsky et al., 2003)

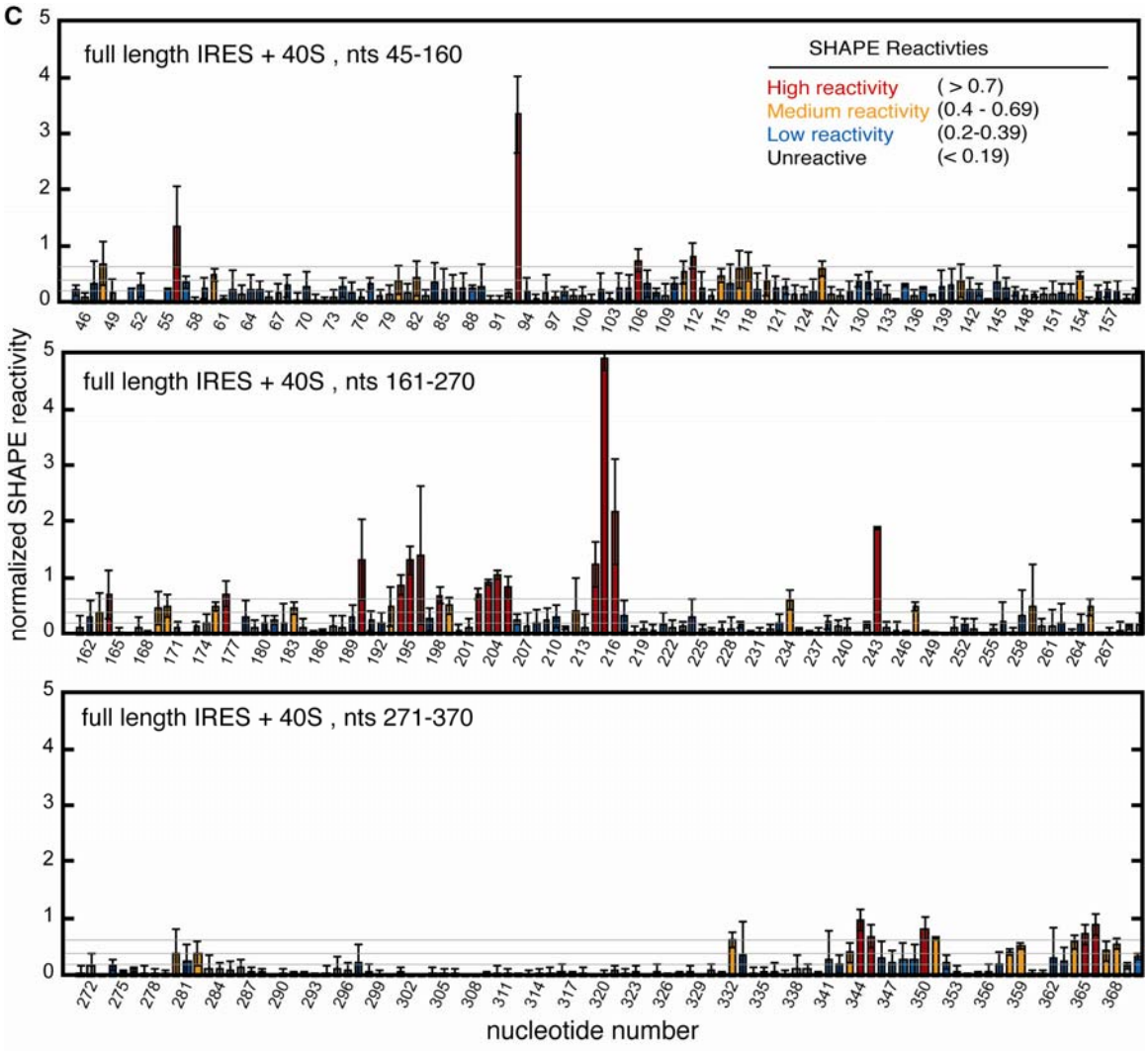

Figure S2 (continued). See figure legend on page 4.

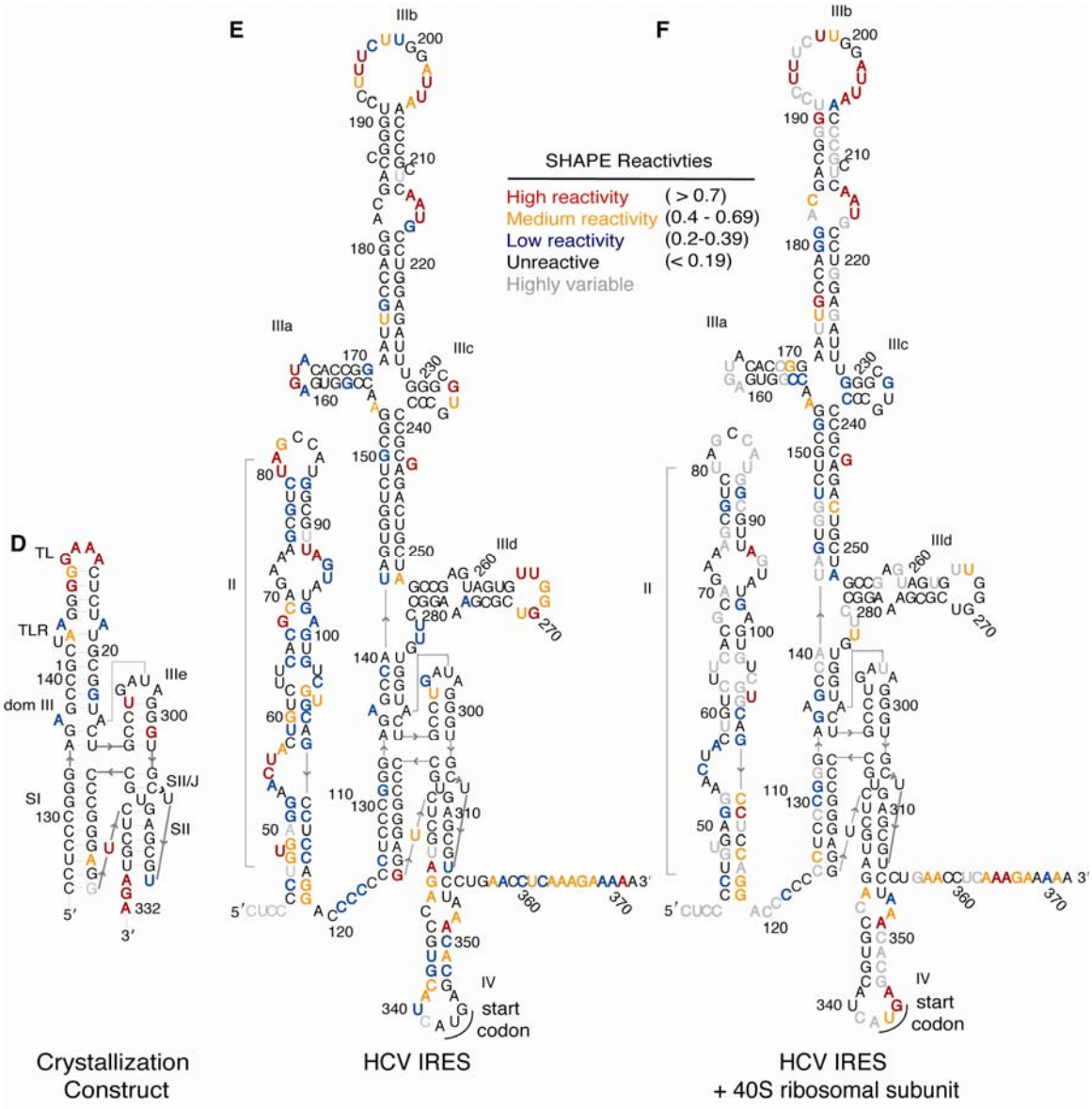

Figure S2 (continued). See figure legend on page 4.

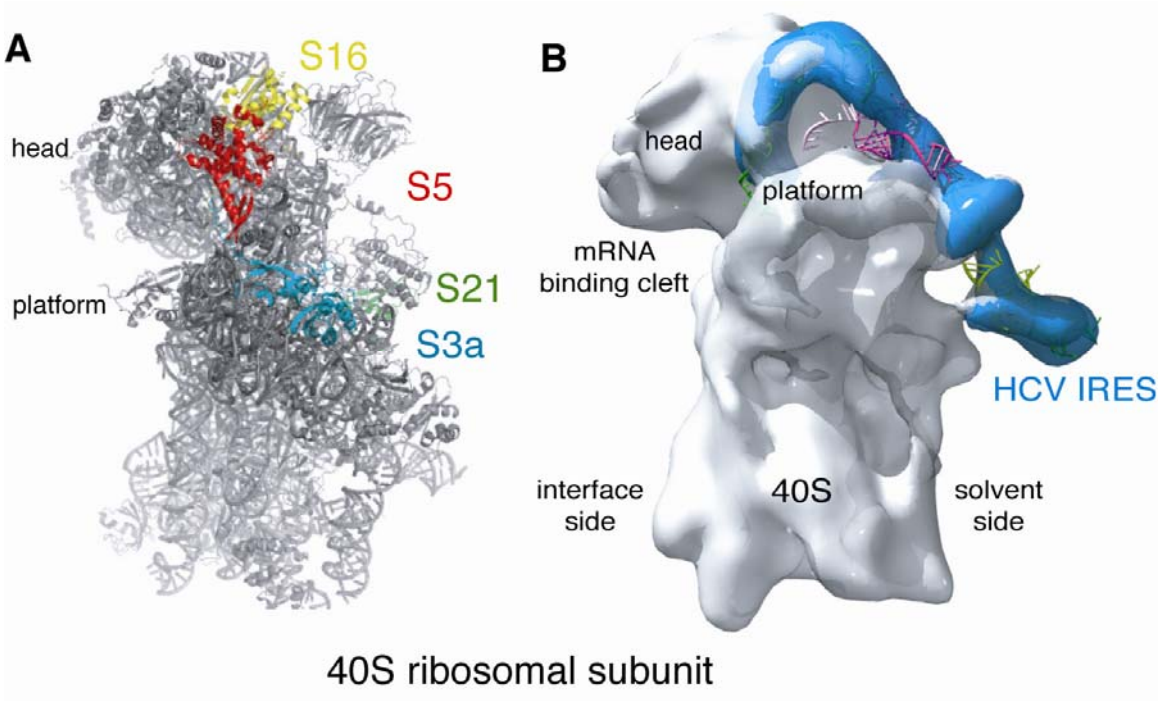

Figure S3, related to Figure 5. Comparison 40S ribosomal subunit crystal structure to docked IRES model. (A) Crystal structure of the *T. thermophila* 40S subunit (2XZM) (Rabl et al., 2011), in the same orientation of (B) the docked HCV IRES-40S subunit cryo-EM-guided model with pseudoknot domain in purple (from Fig. 5). The positions of proteins S16, S5, and S3a are highlighted; these proteins were found to interact with the IIIe tetraloop by site-specific crosslinking (Laletina et al., 2006). While p40, a ribosome-associated protein which also crosslinks to IIIe was not present in the crystal structure, this protein binds to S21 (Sato et al., 1999), the location of which is also highlighted. Proteins S5 and S16 crosslinked the strongest to the IIIe tetraloop, and p40 and S3a crosslinked to a lesser extent. S3a sits on the back of the platform, very close to the proposed binding site of the IIIe tetraloop, whereas S5 and S16 are in the head, somewhat distant from where we hypothesize the IIIe loop would bind, although each of these contact the mRNA binding cleft to some extent.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Crystallization Construct Design and Cloning

Crystallization constructs were cloned between the EcoRI and BamHI restriction sites in pUC19, using standard methods. The sequence for the RNA to be used in crystallography was inserted behind a T7 promoter and between a 5′ hammerhead ribozyme and 3′ hepatitis delta ribozyme. Approximately 16 initial constructs were designed, varying the presence of dom IV, dom IIId, a tetraloop and tetraloop receptor or a U1A binding site. After screening of these constructs in the Nucleix crystallization screen (Qiagen), the best construct was further optimized with respect to length and sequence of 5′- and 3′-overhangs and the space region. The sequence of pKB267, the optimal construct, which was inserted into pUC19, including the T7 promoter and 5′ and 3′ ribozymes was: 5′-GAATTCTAATACGACTCACTATAGGGGAGACTCCCGGGAGGCTGATGAGT CCGTGAGGACGAAACGGTACCCGGTACCGTCCCTCCCGGGAGAGCCGCTA AGGGGGAAACTCTATGCGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCC GGGAGGTCTCGTAGAGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCG GCTGGGCAACACCATTGCACTCCGGTGGCGAATGGGACGGATCC-3′.

RNA purification for crystallization

RNA was generated by T7 transcription according to standard methods. Transcription reactions (10 mL) contained HindIII-digested template plasmid (1 mg; HindIII used to facilitate separation of the HDV ribozyme from the desired product), RNasin Plus RNase inhibitor (~10 μL; Promega), NTPs (×4, 5 mM of

each), T7 polymerase (0.1 mg/mL), pyrophosphatase (5 μg; Roche), DTT (10 mM) in 30 mM Tris HCl, pH 8.1, 25 mM $MgCl_2$, 2 mM spermidine, and 0.01% Triton X-100 (Ke and Doudna, 2004). Transcription reactions were incubated at 37°C for 2-3 hours, after which cycles of ribozyme cleavage were performed. For these cycles, transcription reactions were brought to 60 mM $MgCl_2$ (with an additional 35 mM) and a DNA oligonucleotide (5′-TACGAGACCTCCCGGGGCA CTCG-3′; 2 mM final concentration) was added, intended to disrupt the structure of the pseudoknot such that the 5′ and 3′ ends would be free for interaction with the ribozymes. Reactions were heated at 75°C for 5 min, room temperature for 12 min and 42°C for 8 min. This cycle was repeated three times to allow for complete cleavage of the ribozymes. Ribozyme-cleaved RNA was purified using denaturing gel electrophoresis (0.5× TBE, 10% acrylamide, 8M urea) and recovered by passive elution from frozen and thawed crushed gel slices. Acrylamide was filtered away and RNAs were washed by concentration and dilution in 50 mM KCl (15 mL) for three cycles. Concentrated RNA was stored at -80°C prior to annealing.

For annealing, RNA was diluted to 30 ng/μL in Annealing Buffer (20 mM Hepes 7.5, 50 mM KCl, 2.5 mM $MgCl_2$) and heated to 65°C for 5 min, then placed immediately at room temperature. After annealing, RNAs were concentrated and brought to 2-4 mg/mL in Annealing Buffer (but with only 15-30 mM KCl) for crystallization.

Additional model building and refinement

Anomalous difference maps contoured at 5 σ showed clear density for nine Ni(II) positions along the major groove of the RNA (Ni 1-9; **Supplementary Fig. 1**). Refinement of the model revealed these Ni(II) ions were ~2 Å from the N7 of guanine bases, indicating inner-sphere coordination of the nickel ion. An additional eight Ni(II) sites were identified during refinement, mostly coordinated to the Lewis basic N7 of guanine, with one coordinated to the N7 of an adenine (Ni 15) and one directly to a phosphate (Ni 16), with perhaps additional water-mediated coordinations.

Modeling with EM structures

The pseudoknot structure, along with previously determined crystal and NMR structures of dom II (1P5P) (Lukavsky et al., 2003), IIId (1F84) (Lukavsky et al., 2000), JIIIabc (1KH6) (Kieft et al., 2002), and IIIb (1KP7) (Collier et al., 2002), was manually modeled into the cryo-EM difference density from an HCV IRES-eIF3 reconstruction (Siridechadilok et al., 2005) and overlayed with a cryo EM reconstruction of the 40S-HCV IRES (Spahn et al., 2001) using Chimera (Goddard et al., 2007).

SHAPE RNA Preparation

5′ and 3′ handles (Wilkinson et al., 2006) were added to the pKB84 full-length IRES and pKB267 crystallization construct by PCR using the following primers: pKB267 forward: 5′-TAATACGACTCACTATAGGCCTTCGGGCCAACC

TCCCGGGAGAGCCGCTAAGG-3′, pKB267 reverse: 5′-GAACCGGACCGAAG CCCGATTTGGATCCGCCGAAGCGGATCGGTCTACGAGACCTCCCGGGGCA C-3′, pKB84 forward: 5′-TAATACGACTCACTATAGCT CCCCTGTGAGGAAC-3′, pKB84 reverse: 5′-GAACCGGACCGAAGCCCGATTTGGCCGGATCCGCCGA AGCGGATCGGGGCCTTTCTTTATGTTTTTGGCGTCTTCCATGAGGATCC-3′. Amplification of the full-length IRES from the IRES-FF luc construct contained the first 32 nts of the FF luc coding sequence in order for the RT traces to be clean and well-resolved in the dom IV region. The Taq PCR reaction (1 mL) was ethanol precipitated and one-third of this product was used in T7 in vitro transcription reactions (1 mL) without further purification. Transcribed RNA was purified using denaturing gel electrophoresis (0.5x TBE, 10% acrylamide, 8M urea) and recovered by passive elution from crushed gel slices.

SHAPE Reactions and Analysis

The crystallization construct was annealed (5 pmol in 9 µL) as above for crystallography. Full length IRES (5 pmol in 6.5 - 9 µL) was annealed in Annealing Buffer (20 mM Hepes 7.5, 50 mM KCl, 2.5 mM $MgCl_2$) by heating at 65°C for 5 min, then allowing it to cool slowly by placing the heat block at room temperature. Native gel analysis confirmed that RNA ran as a single band, even with 5′- and/or 3′-handles present. Folded RNA samples were placed at room temperature for 10 min before reaction with 1M7. For the full-length IRES, 40S ribosomal subunits (7.5 pmol) or 40S storage buffer was added to folded RNAs and 9 µL reactions were brought to final concentrations of 50 mM MES 6.5, 90

mM KCl, 2.3 mM MgCl2 and 1.0 mM DTT, heated at 37°C for 10 min for 40S activation and incubated at room temperature for 10 min prior to reaction with 1M7.

To each 9 µL reaction, 1M7 dissolved in DMSO (1 µL) or DMSO only (1 µL, used as a background control) was added. Final concentrations of 1M7 were 6.5 mM for the crystallization construct and 3.5 mM for the full-length IRES, with or without 40S subunits. These reactions were allowed to incubate at room temperature for 5 minutes, and reactions containing 40S ribosomal subunits (or mock reactions) were extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1, Sigma) and once with chloroform. All RNAs were concentrated after modification by ethanol precipitation. Modified RNA was recovered by ethanol precipitation [90 µL sterile water or 100 µL from aqueous layer, 5 µL NaCl (5 M), 1 µL glycogen (20 mg/mL), 400 µL ethanol; 30 min at -80°C] and resuspended in 10 µL of DEPC treated water.

The general procedure of primer extension and data analysis for modified RNAs was outlined previously(Mortimer and Weeks, 2009). Briefly, a fluorescently labeled DNA primer ((5′ VIC- or NED-labeled GAA CCG GAC CGA AGC CCG; 3 µL, 0.3 µM) was annealed to the RNA (10 µL, from the previous step) by heating at 65°C (6 min) and 35°C (5 min). Reverse transcription buffer and Superscript III were added and the reactions incubated at 45°C for 1 min, 52°C for 25 min, and 65°C for 5 min. Reactions were quenched by adding 4 µL EDTA (50 mM, pH 8.0) and cDNA was recovered by ethanol precipitation, washed twice with 70% ethanol (800 µL), dried by vacuum for 10 min, and

resuspended in 10 µL de-ionized formamide. Dideoxy sequencing markers were generated using unmodified RNA and primers labeled with unique fluorophores (6-FAM or PET, 0.3 or 0.6 µM respectively), and by adding 1 µL of 2′,3′-dideoxythymidine or 2′,3′-dideoxyadenosine (10 mM) after addition of reverse transcription buffer. cDNA extension products, along with dideoxy sequencing markers were separated by capillary electrophoresis with an ABI 3730xl Genetic Analyzer, using GeneMapper 4.0 for fragment analysis. Raw traces were analyzed using ShapeFinder (Vasa et al., 2008). Data sets were normalized by excluding the 2% most reactive nucleotides and dividing by the average intensity of the next 8% most reactive nucleotides. On this scale, 1.0 is thus the average intensity of highly reactive positions. Data presented are the average of normalized data from at least three replicates.

Purification of 40S ribosomal subunits

40S ribosomal subunits were purified from HeLa cytoplasmic lysate (a gift from R. Tjian, University of California at Berkeley) as previously described (Fraser et al., 2007).

SUPPLEMENTAL REFERENCES

Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-Ela, F., Varani, G., and Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. Nat. Struct. Biol. *9*, 375-380.

Fraser, C.S., Berry, K.E., Hershey, J.W., and Doudna, J.A. (2007). eIF3j is located in the decoding center of the human 40S ribosomal subunit. Mol Cell *26*, 811-819.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J. Struct. Biol. *157*, 281-287.

Ke, A., and Doudna, J.A. (2004). Crystallization of RNA and RNA-protein complexes. Methods *34*, 408-414.

Kieft, J.S., Zhou, K., Grech, A., Jubin, R., and Doudna, J.A. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. Nat. Struct. Biol. *9*, 370-374.

Laletina, E.S., Graifer, D.M., Malygin, A.A., Shatskii, I.N., and Karpova, G.G. (2006). [Molecular environment of the subdomain IIIe loop of the RNA IRES element of hepatitis C virus on the human 40S ribosomal subunit]. Bioorg Khim *32*, 311-319.

Lavender, C.A., Ding, F., Dokholyan, N.V., and Weeks, K.M. (2010). Robust and generic RNA modeling using inferred constraints: a structure for the hepatitis C virus IRES pseudoknot domain. Biochemistry *49*, 4931-4933.

Lukavsky, P.J., Kim, I., Otto, G.A., and Puglisi, J.D. (2003). Structure of HCV IRES domain II determined by NMR. Nat. Struct. Biol. *10*, 1033-1038.

Lukavsky, P.J., Otto, G.A., Lancaster, A.M., Sarnow, P., and Puglisi, J.D. (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. Nat. Struct. Biol. *7*, 1105-1110.

Mortimer, S.A., and Weeks, K.M. (2009). Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. Nat Protoc *4*, 1413-1421.

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. Science *331*, 730-736.

Sato, M., Saeki, Y., Tanaka, K., and Kaneda, Y. (1999). Ribosome-associated protein LBP/p40 binds to S21 protein of 40S ribosome: analysis using a yeast two-hybrid system. Biochem Biophys Res Commun *256*, 385-390.

Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. Science *310*, 1513-1515.

Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A., and Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. Science *291*, 1959-1962.

Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. (2008). ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA *14*, 1979-1990.

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.

Zhao, W.D., and Wimmer, E. (2001). Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. Journal of virology *75*, 3719-3730.