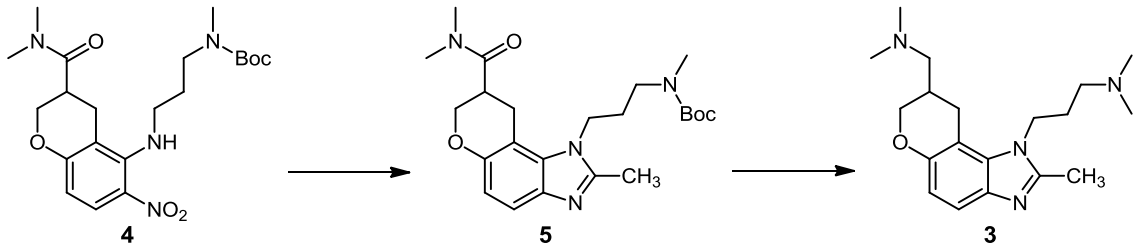
Supporting Information

Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches

Sergey M. Dibrov, Kejia Ding, Nicholas D. Brunn, Matthew A. Parker,B. Mikael Bergdahl, David L. Wyles, and Thomas Hermann

Compound Synthesis. The methyl analog **3**, which served as an inactive control compound, was obtained in two steps from the published intermediate **4** (1):



tert-Butyl-N-[3-[8-(dimethylcarbamoyl)-2-methyl-8,9-dihydro-7H-pyrano[2,3-g]benzimidazol-1-yl]-propyl]-N-methylcarbamate (**5**). To a flask containing 10% palladium on carbon (315 mg) was added a solution of **4** (350 mg, 0.80 mmol) in absolute ethanol (35 mL). The mixture was stirred for 90 minutes under 1 atm of H_2 at room temperature, at which point ^{1}H-NMR analysis showed complete reduction to the corresponding aniline derivative. The mixture was filtered through Celite. The Celite was washed with ethanol (3 x 10 mL) and all filtrates were combined. To this solution was added ethyl acetimidate hydrochloride (117 mg, 0.95 mmol) and the mixture was stirred at room temperature for 24 hours under argon. The solvent was evaporated, and the residue was partitioned between CH_2Cl_2 (50 mL) and 10% aq. K_2CO_3 (50 mL). The CH_2Cl_2 phase was separated, dried with Na_2SO_4, and evaporated. The residue was purified on silica using a gradient of 0–65% 2-propanol in ethyl acetate containing 1% triethylamine, giving 288 mg (83%) of **2** as a light tan oil that solidified into a beige crystalline solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 (d, Ar-*H*, *J* = 8.7 Hz, 1H), 6.77 (d, Ar-*H*, *J* = 8.7 Hz, 1H), 4.39 (m, OC*H*$_2$CH, 1H), 4.22 (t, NC*H*$_2$CH$_2$, *J* = 8.0 Hz, 2H), 4.00 (dd, OC*H*$_2$CH, *J* = 10.6, 10.6 Hz, 1H), 3.51 (m, C*H*CONMe$_2$, 1H), 3.70-3.25 (m, ArC*H*$_2$CH, 2H), 3.26 (m, *partly hidden*, NC*H*$_2$, 1H), 3.21 (m,

partly hidden, NCH_2, 1H), 3.19, 3.04, 2.87, 2.54 (4s, CH_3, 3H each), 2.06-1.86 (m, $CH_2CH_2CH_2$, 2H), 1.45 (s, *t*-Bu, 9H). ESI-MS $(M+H)^+$: 431.4.

8-(Dimethylaminomethyl)-1-(3-dimethylaminopropyl)-2-methyl-8,9-dihydro-7H-pyrano[3,2-e]benzimidazole (**3**). Compound **5** (200 mg, 0.47 mmol) was dissolved in THF (10 mL) under argon. Lithium aluminum hydride (660 mg, 17 mmol) was added and the mixture was stirred at 60 °C for 1.5 hours. The mixture was cooled to room temperature and diluted with anhydrous ether (15 mL), then quenched by careful dropwise sequential addition of 2-propanol (0.5 mL), 15% aq. NaOH (0.5 mL), and water (1.7 mL). The solids were removed by filtration and washed with anhydrous ether (3 x 15 mL). The filtrate and washings were combined, dried with Na_2SO_4, and evaporated to give the crude product (140 mg, 91%) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.38 (d, Ar-*H*, *J* = 8.6 Hz, 1H), 7.00 (d, Ar-*H*, *J* = 8.6 Hz, 1H), 4.29 (m, *partly hidden*, OCH_2CH, 1H), 4.27 (t, NCH_2CH_2, *J* = 7.7 Hz, 2H), 3.86 (dd, OCH_2CH, *J* = 10.6, 7.0 Hz, 1H), 3.25 (dd, ArCH_2CH, *J* = 15.8, 4.7 Hz, 1H), 2.86 (dd, ArCH_2CH, *J* = 15.8, 7.0 Hz, 1H), 2.56 (s, CCH_3, 3H), 2.32-2.28 (2m, *partly hidden*, 2 x NCH_2, 2H each), 2.28 (m, OCH_2CH, 1H), 2.25, 2.24 (2s, N(CH_3)$_2$, 6H each), 1.96-1.84 (m, $CH_2CH_2CH_2$, 2H). The *bis*-hydrochloride salt was obtained as follows: A portion of the crude product (35 mg, 0.11 mmol) was dissolved in anhydrous ethanol (0.75 mL). Anhydrous HCl in ethanol (0.067 mL of 3.5M HCl; 0.23 mmol) was added. The salt was then precipitated by addition of anhydrous Et_2O and recrystallized from ethanol/Et_2O/EtOAc (Et_2O:EtOAc = 1:1). The resulting tan crystalline powder (13.1 mg) was quite hygroscopic, quickly becoming sticky in air but reverting to a free-flowing powder under argon. ESI-MS $(M+H)^+$: 331.3.

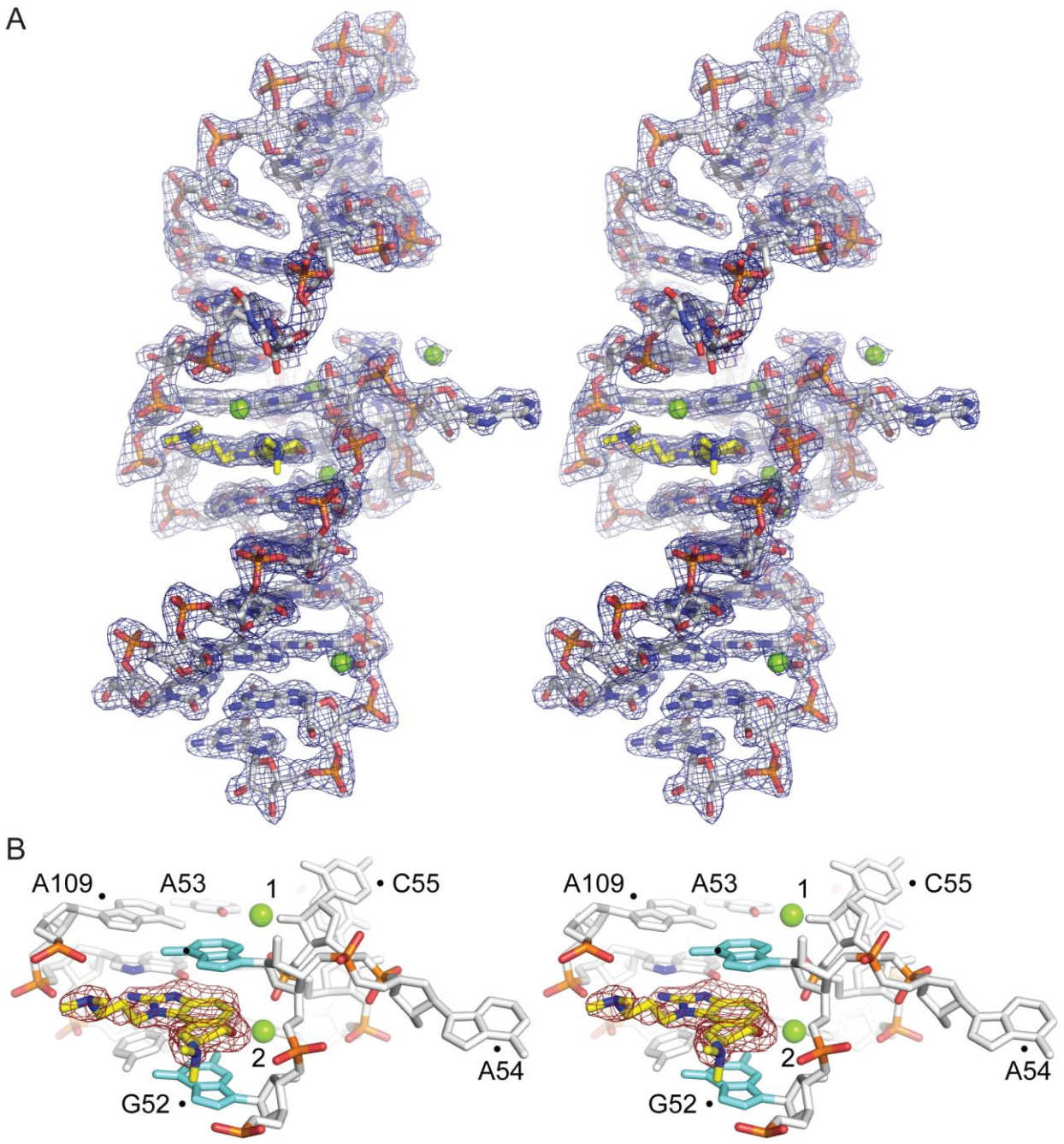


Fig. S1. Stereo view of the subdomain IIa RNA in complex with the benzimidazole ligand. (*A*) Overall structure overlaid with a 2Fo-Fc electron density map contoured at 1σ. The ligand is shown in yellow, magnesium ions in green. (*B*) Detail view of the ligand binding site. Overlaid on the ligand is a Fo-Fc electron density map contoured at 3σ. The density map was calculated from experimental phases with the ligand omitted from the RNA complex model.

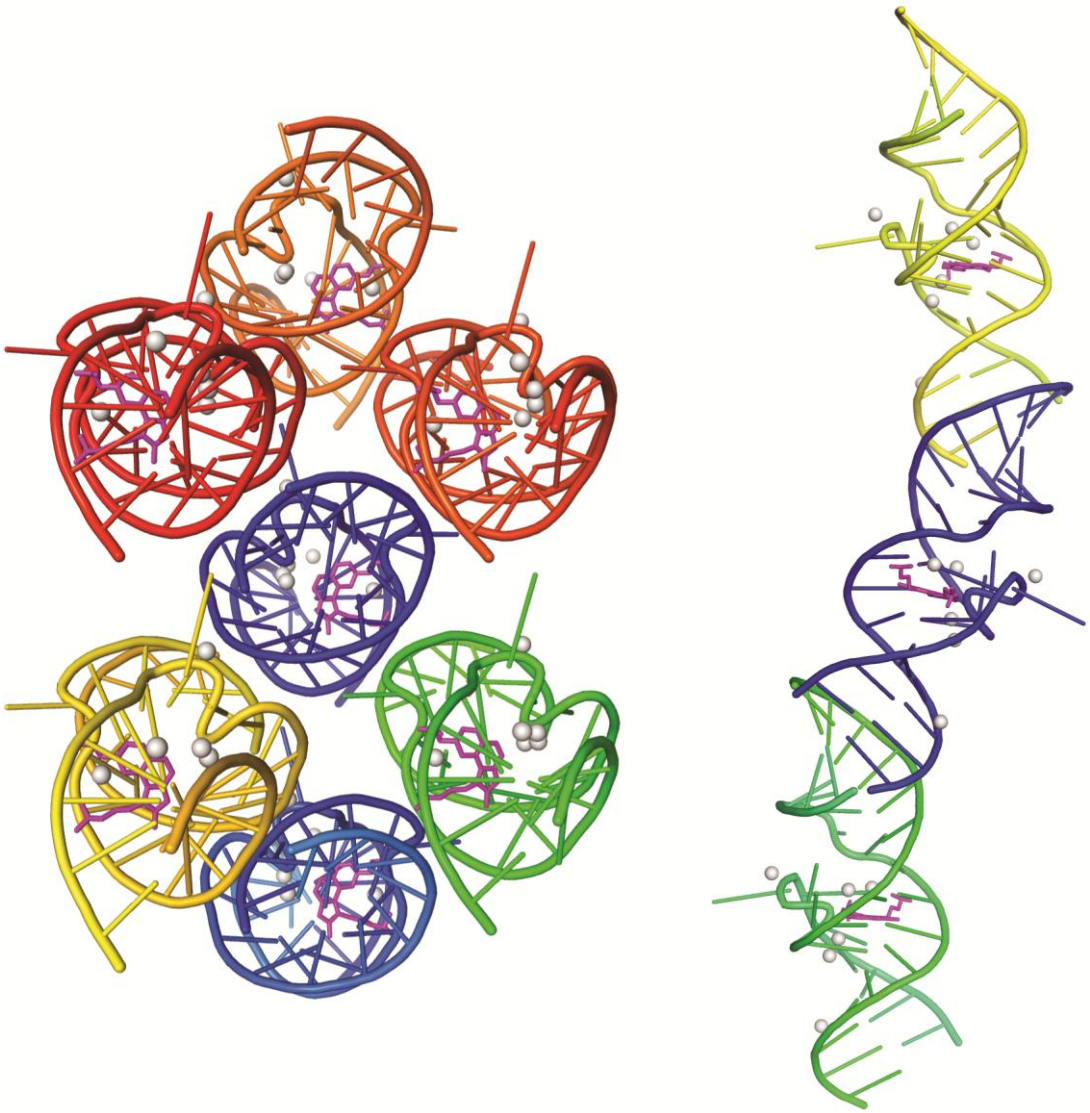


Fig. S2. Packing of the subdomain IIa ligand complex in the crystal lattice. Side-by-side packing of RNA helices does not involve contacts with the ligand binding pocket. Helices are stacked end-to-end to from a pseudo-continuous arrangement that is stabilized by intermolecular base pair formation between 3' overhang nucleotides in adjacent RNAs. The benzimidazole ligand **2** is shown in magenta.

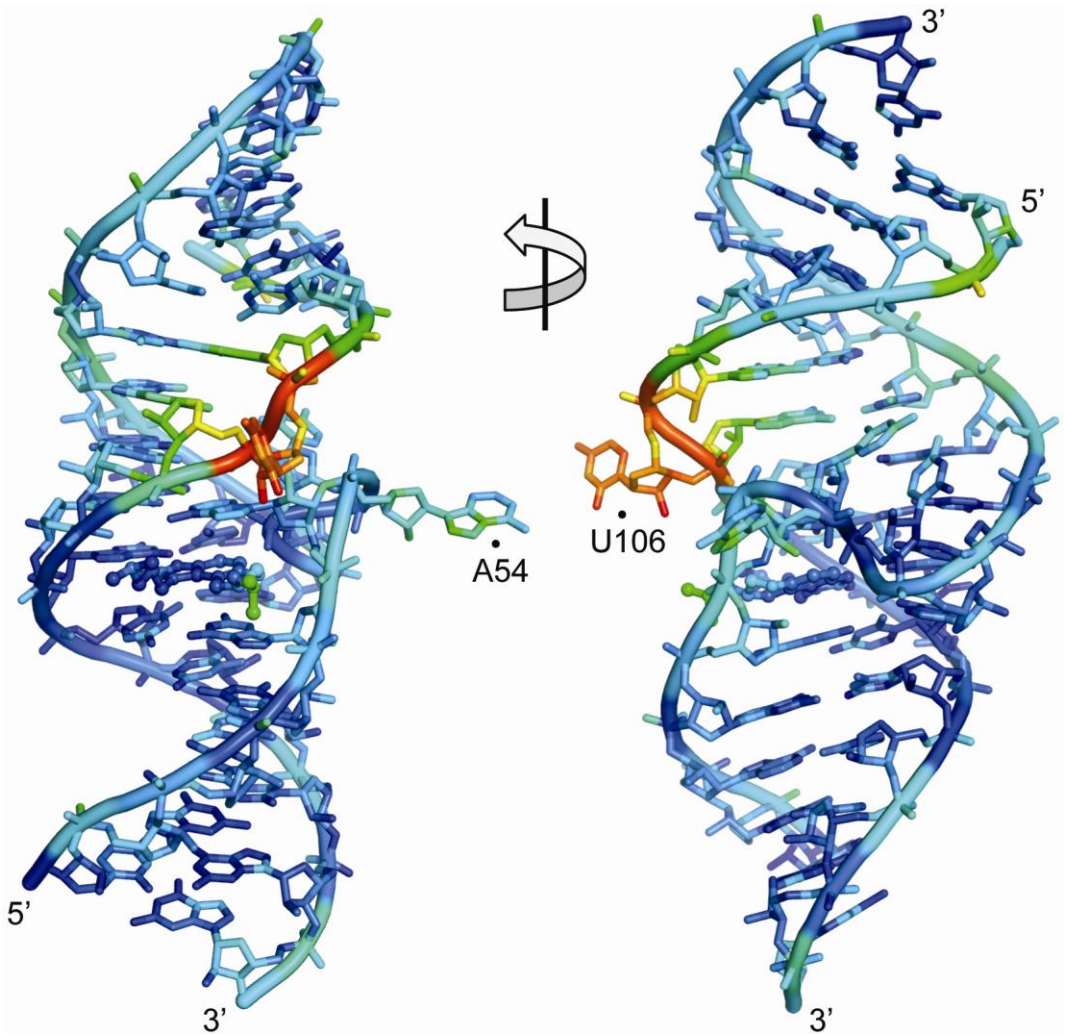


Fig. S3. Two views, rotated by 90, of the subdomain IIa ligand complex with B factors mapped as color coding on the structure (blue = low, red = high B factors). The benzimidazole ligand **2** is shown in ball-and-stick representation.

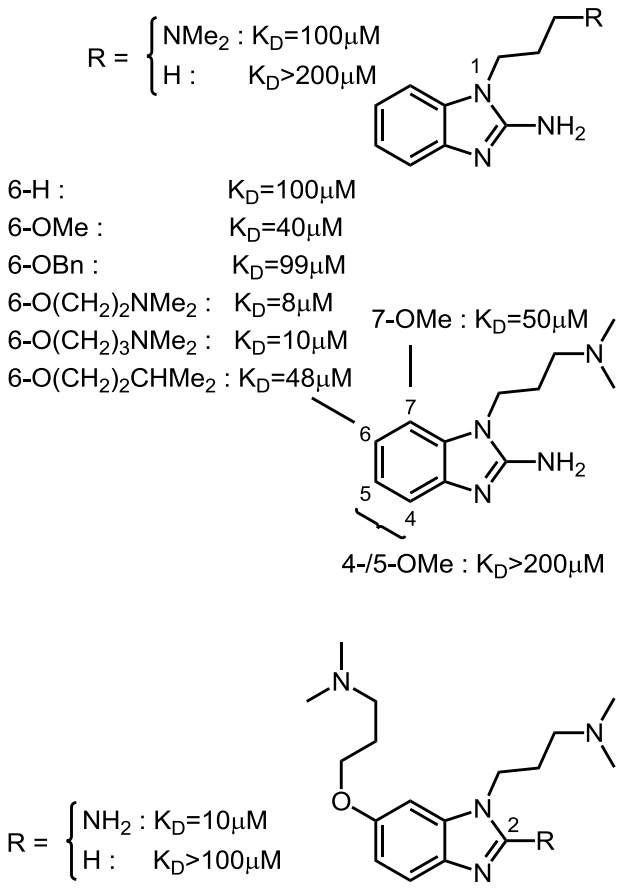


Fig. S4. Structure activity relationships of benzimidazole ligands of the subdomain IIa target. Affinities (K_D) were determined by a mass spectrometry assay for the IIa RNA (2).

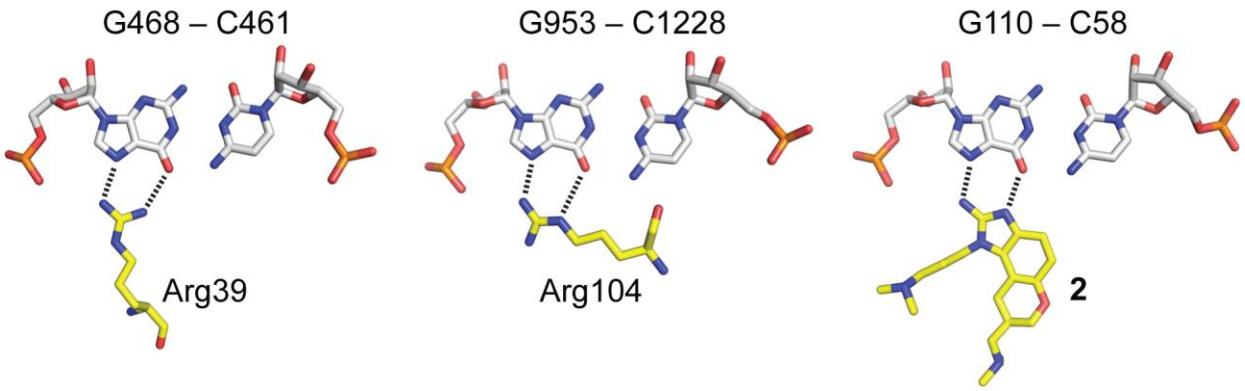


Fig. S5. Examples of Hoogsteen edge recognition of G-C pairs by arginine in the bacterial ribosome compared to the benzimidazole G110-C58 interaction in the subdomain IIa complex. Arg39 in ribosomal protein L34 binds at G468-C461 in the 23S rRNA; Arg104 in ribosomal protein S13 interacts with G953-C1228 in the 16S rRNA (3).

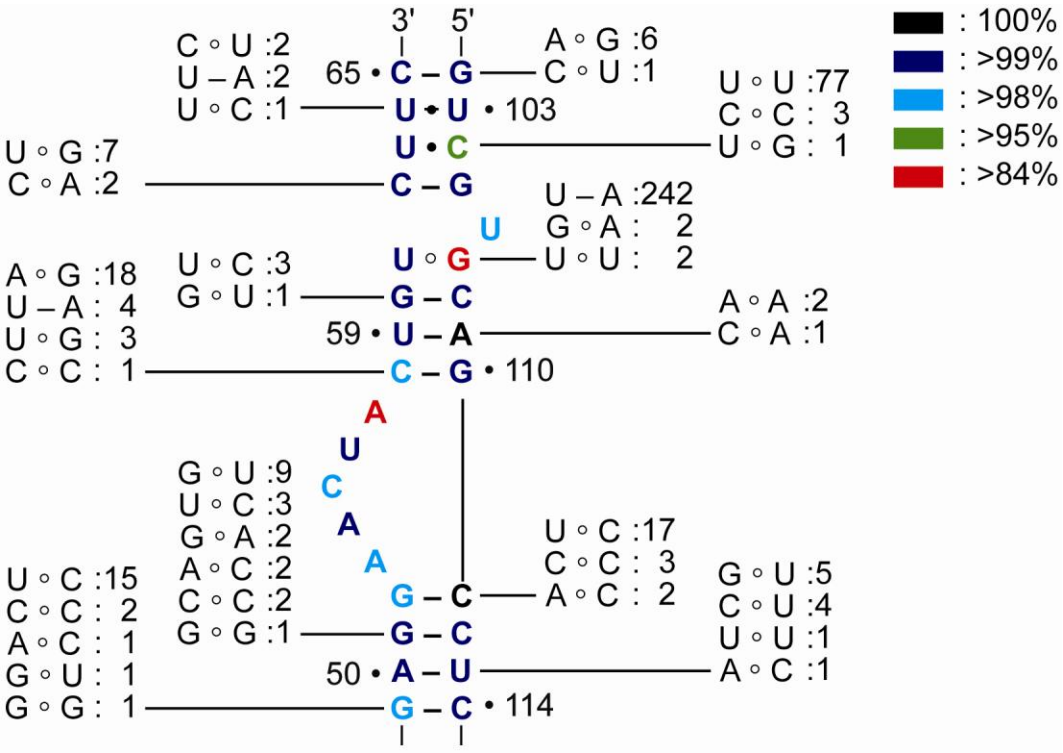


Fig. S6. Conservation of residues and base pairs in subdomain IIa. The conservation of residues, taken from Table S4, is shown in color coding. The occurrence of co-varying base substitutions is indicated for each pair. Conservation is uniform across all genotypes, except for the two lowest conserved residues, A57 and G107, in which the most prevalent polymorphisms occur predominantly in genotype 3 (U57) or all genotypes except 3 (A107). A total of 1,601 sequences from HCV clinical isolates were analyzed (4).

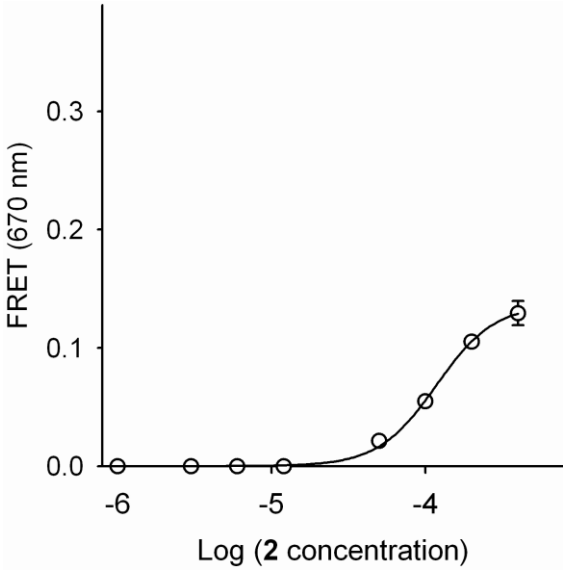


Fig. S7. FRET signal for titrations of Cy3/Cy5-labelled IIa RNA with benzimidazole **2** in the absence of Mg^{2+}. Fitting of a dose-response curve resulted in an EC_{50} value for the FRET signal increase of 117±8μM. Error bars represent ±1 SD calculated from triplicate experiments. Note that the final state (FRET signal = 0.13) does not correspond to the completely folded state of the RNA in the presence of Mg^{2+} (FRET signal > 0.3; see Figure S8 and ref. (5)).

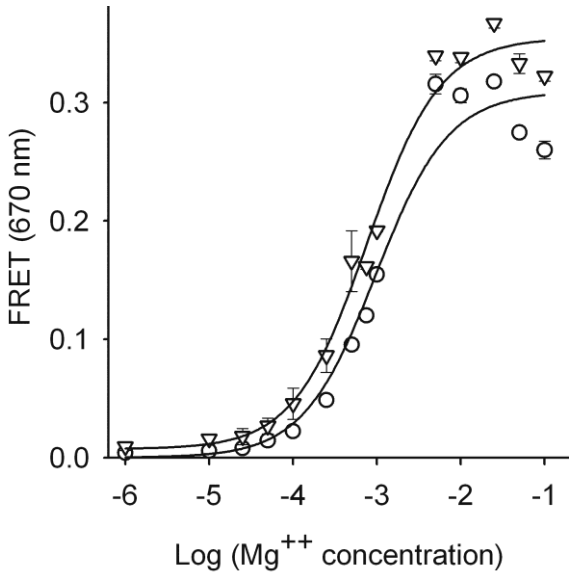


Fig. S8. FRET signal in a titration of Cy3/Cy5-labeled subdomain IIa RNA double mutants with Mg^{2+}, C58G/G110C (○) and C58U/G110A (□), demonstrating proficiency of the mutant RNA to fold into the L-shaped IIa architecture. Fitting of dose-response curves resulted in an EC_{50} value for Mg^{2+} binding of 970±140μM (C58G/G110C) and 760±90μM (C58U/G110A). Error bars represent ±1 SD calculated from triplicate experiments. In comparison, folding of the wt RNA occurred by Mg^{2+} binding with an EC_{50} value of 375±140μM (5). Error bars represent ±1 SD calculated from triplicate experiments.

Table S1. Oligonucleotides used for experiments.

Oligonucleotide	Sequence
Crystallization:	
IIa left strand	5’ CGA GGA ACU ACU GUC UUC CC
IIa right strand	5’ GGU CGU GCA GCC UCG G
2AP Fluorescence Experiments:	(6)
IIa left strand	5’ CGG AGG A[2AP]C UAC UGU CUU CAC GCC
IIa right strand	5’ GCG UGU CGU GCA GCC UCC GG
FRET Experiments:	(5)
IIa left strand	5’ [Cy3]-UCG GAG GAA CUA CUG UCU UCA CGC C
IIa right strand	5’ [Cy5]-UGC GUG UCG UGC AGC CUC CGG

Table S2. Crystallographic data collection and refinement statistics for the HCV IRES subdomain IIa benzimidazole complex.

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	2.2
Low-resolution limit (Å)	16.8
Redundancy[a]	6.7 (4.5)
Completeness (%)[a]	99.15 (75.6)
$I/\sigma(I)$[a]	32.5 (6.0)
Total reflections	31264
Unique reflections	4664
Refinement	
Space group	$P2_12_12_1$
Cell dimensions (Å)	
a	32.05
b	33.58
c	81.05
α	90
β	90
γ	90
R_{work}/R_{free}	0.18 / 0.24
No. atoms	
RNA atoms	757
Anions	3 SO_4^{2-}
Ligand	24
Solvent atoms	81
Metal ions	6 Mg^{2+}
Mean *B* factors (Å^2)	
RNA	27.1
Anions	52.7
Ligand	25.0
Solvent	29.4
Metal	34.1
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.878
Dihedral angles (°)	17.72

[a]Numbers in parentheses are for the highest-resolution shell.

Table S3. Hydrogen bonds and metal coordination around the ligand binding site in the subdomain IIa benzimidazole complex.

Residue	1	Residue	2	Distance (Å)
A53	N3	C55	O2’	2.87
A53	N1	A109	N6	2.91
C55	OP2	U56	O2’	2.55
U56	O1P	C58	N4	2.93
A57	N1	C111	O2’	2.52
A57	N6	C111	O2	2.92
U56	O1P	Mg1		2.55
U59	O4	Mg1		2.78
U56	OP2	Mg2		2.70
A57	N7	Mg2		2.81
A53	N7	Mg3		2.86
C108	OP2	Mg3		2.81
G110	N7	ligand **2**	N2	2.70
G110	O6	ligand **2**	N3	2.69
A109	OP2	ligand **2**	N16	2.57

Table S4. Conservation of residues in the subdomain IIa.[*]

Residue	A	C	G	U	Residue	A	C	G	U
65	0.4[1]	99.6	0.0	0.0	**102**	0.0	0.0	99.9	0.1
64	0.0	0.1	0.0	99.9	**103**	0.1	0.1	0.0	99.8
63	0.0	0.2	0.0	99.8	**104**	0.0	95.1	0.1	4.8[3]
62	0.0	99.6	0.0	0.4	**105**	0.1	0.0	99.9	0.0
					106	0.9[4]	0.4	0.1	98.6
61	0.0	0.0	0.1	99.9	**107**	15.3[5]	0.0	84.6	0.1
60	0.0	0.0	99.8	0.2	**108**	0.0	99.9	0.0	0.1
59	0.1	0.1	0.0	99.8	**109**	100.0	0.0	0.0	0.0
58	1.1	98.4	0.0	0.4	**110**	0.2	0.1	99.7	0.0
57	84.6	0.3	0.1	15.1[2]					
56	0.1	0.2	0.3	99.4					
55	0.1	98.3	0.2	1.4					
54	99.8	0.0	0.1	0.2					
53	98.7	0.2	0.9	0.2					
52	0.1	0.2	98.6	1.1	**111**	0.0	100.0	0.0	0.0
51	0.1	0.1	99.6	0.2	**112**	0.1	99.3	0.1	0.6
50	99.4	0.2	0.3	0.1	**113**	0.0	0.1	0.0	99.9
49	0.1	0.1	98.9	0.9	**114**	0.0	99.9	0.1	0.1

[*]Total 1,601 sequences from the HCV database of clinical isolates (4); 0.1% corresponds to occurrence in maximally 2 sequences.

[1]genotype 1b only

[2]predominantly genotype 3

[3]genotypes 1, 2, 4, 6 only

[4]genotype 5 only

[5]all genotypes except 3

Supporting Information References

1. Parker MA, Satkiewicz E, Hermann T, & Bergdahl BM (2011) An efficient new route to dihydropyranobenzimidazole inhibitors of HCV replication. *Molecules* 16(1):281-290.

2. Seth PP*, et al.* (2005) SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J Med Chem* 48(23):7099-7102.

3. Selmer M*, et al.* (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313(5795):1935-1942.

4. Kuiken C, Hraber P, Thurmond J, & Yusim K (2008) The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res* 36(Database issue):D512-516.

5. Parsons J*, et al.* (2009) Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat Chem Biol* 5(11):823-825.

6. Dibrov SM, Johnston-Cox H, Weng YH, & Hermann T (2007) Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew Chem Int Ed Engl* 46(1-2):226-229.