nature
structural biology

Structure of HCV IRES domain II determined by NMR

Peter J Lukavsky[1,2], Insil Kim[1], Geoff A Otto[1] & Joseph D Puglisi[1]

Complex RNA structures regulate many biological processes, but are often too large for structure determination by NMR methods. The 5′ untranslated region (5′ UTR) of the hepatitis C viral (HCV) RNA genome contains an internal ribosome entry site (IRES) that binds to 40S ribosomal subunits with high affinity and specificity to control translation. Domain II of the HCV IRES forms a 25-kDa folded subdomain that may alter ribosome conformation. We report here the structure of domain II as determined using an NMR approach that combines short- and long-range structural data. Domain II adopts a distorted L-shape structure, and its overall shape in the free form is markedly similar to its 40S subunit–bound form; this suggests how domain II may modulate 40S subunit conformation. The results show how NMR can be used for structural analysis of large biological RNAs.

RNA has a central role in biology as messenger RNA, as components of the ribosome and spliceosome and as small cellular RNAs. RNA can form complex tertiary structures that are often stabilized by protein interaction. X-ray crystallography has revealed the three-dimensional structures of several stably folded RNAs, including tRNA, ribozymes and the ribosome. RNAs are often conformationally heterogeneous, and this may be an essential feature of their functions. As such, RNAs have been frequently difficult to crystallize. NMR spectroscopy provides a powerful tool for structural characterization of dynamic RNAs; however, most RNA molecules are larger (>20 kDa) than the size limit accessible to earlier NMR methods. Here, we present an NMR approach that allowed the determination of the structure of a 25-kDa RNA domain, and provides qualitative information on a functional 100-kDa RNA from the hepatitis C virus. These results demonstrate the quantitative and qualitative power of NMR for structural characterization of large biologically important RNAs.

Initiation of translation of the HCV polyprotein is mediated by an IRES element located in the 5′ UTR of the viral RNA genome[1]. Canonical initiation in eukaryotes is a cap-dependent process, which requires a 5′-capped mRNA and the full complement of initiation factors[2]. The 5′ cap is recognized by eukaryotic initiation factor 4E followed by binding of other initiation factors and the 43S particle. This assembly then scans the 5′ UTR to locate the first AUG start codon to form the 48S preinitiation complex that establishes the proper mRNA reading frame[2]. Recruitment of the 43S particle to HCV RNA requires neither a 5′ cap nor scanning; only a subset of core initiation factors is required[1]. Both mRNA binding and proper P-site placement of the AUG start codon are accomplished by high-affinity (K_d = 1.9 nM) interaction between the IRES RNA and the 40S ribosomal subunit near the E site[3,4]. Surface ribosomal proteins mediate this interaction[5].

The functions of HCV IRES domains (I–IV) have been mapped[1] (**Fig. 1**). Domains II–IV are essential for IRES activity. Isolated domains III and IV alone bind the 40S subunit with near wild-type affinity[3] (K_d = 3.1 nM) and interact with the 40S subunit in the E site as revealed by cryo-electron microscopy (cryo-EM) of the HCV IRES RNA–rabbit 40S subunit complex[4]. Domain II does not contribute substantially to IRES-40S subunit affinity, but is required for IRES activity[3]. The cryo-EM structure shows that domain II contacts the 40S subunit in the head region of the E site. It adopts an elongated L-shaped structure, allowing the apical loop to reach deeply into the mRNA cleft near the coding RNA in the ribosomal P site[4]. Here, we present the solution structure of the HCV IRES domain II, determined by NMR spectroscopy, which reveals that this bound conformation of HCV IRES domain II is preformed in the absence of 40S subunits.

RESULTS

NMR structure determination

NMR structure determination of a large RNA, like the 25-kDa IRES domain II RNA, is challenging. Severe spectral overlap from ^{1}H resonances of the four different nucleotides, each containing at least ten protons, makes unambiguous assignments impossible. Excision of thermodynamically stable secondary structure elements, such as helices and loops, out of their larger structural context can overcome the overlap problem. In favorable cases, these fragments maintain the conformation adopted within the larger RNA. Based on initial NMR studies of a 77-nucleotide (77-nt) RNA (nt 45–117 of domain II), we identified such stable local RNA motifs within domain II and designed a 34-nt RNA (nt 69–98 of domain II, termed IIb) and a 55-nt RNA (nt 45–69 and 98–117 of domain II, termed IIa) covering the entire domain II in two parts (**Fig. 1b**). Chemical shifts, which are a sensitive measure of the local chemical environment of a nucleus, were compared to ensure that the local RNA motifs IIa and IIb adopt the same conformation within the context of the entire domain II (**Table 1**). Assigned chemical shifts are available online at http://puglisi.stanford.edu.

[1]Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305-5126, USA. [2]Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Correspondence should be addressed to J.D.P. (puglisi@stanford.edu).

Published online 26 October 2003; doi:10.1038/nsb1004

To address whether domain II also forms the same structure in the context of the entire HCV IRES, we prepared a segmentally (nt 40–104) ^{15}N-labeled IRES RNA by enzymatic ligation using T4 RNA ligase[6]. Chemical shifts of imino ^{1}H and ^{15}N resonances observed for domain II alone and in the context of the 100 kDa HCV IRES were virtually identical (**Fig. 1b**). This demonstrates that domain II forms an independently folded subdomain in the intact IRES[6].

To determine the structure of domain II by NMR, high-resolution restraints from the subdomains IIa and IIb were combined with residual dipolar coupling[7] data of the subdomains and the full domain II (**Supplementary Table 1**). The precise overlap of chemical shifts in the subdomains and overall domain II strongly supports this approach to structure determination (**Table 1**). Numerous NOE and scalar coupling restraints, obtained after complete resonance assignments of domains IIa and IIb, defined the local structures of these domains, but contained no information to define the global conformation of domain II: the atomic r.m.s. deviations for the 20 best IIb and 29 best IIa structures were 2.43 Å and 4.91 Å, respectively, whereas the entire domain II was poorly defined with an r.m.s. deviation of 7.48 Å for the best 20 structures (**Fig. 2** and see **Supplementary Tables 2–4** online). The precision of the subdomain structures was improved using residual dipolar couplings (RDCs)[7]; the magnitudes of the axial (D_a) and rhombic (R) components of the alignment tensor were determined using standard methods[7–9] (see **Supplementary Table 1** online). The local conformational restraints were then combined with 60 RDCs from the overall domain II to calculate the final structures. Details of the structure determination are outlined in Methods and **Supplementary Methods** online.

The final ensemble of 12 domain II structures has an r.m.s. deviation of 2.18 Å (**Table 1** and **Fig. 2**) and the r.m.s. deviations of the local motifs are 1.15 Å (IIb) and 1.62 Å (IIa), respectively (see **Supplementary Tables 2–4** online). The overall domain II RDCs also improve the local precision of the structures. A control calculation carried out on domain II using only the domain IIa and domain IIb RDCs yielded a very similar local precision of subdomains (1.43 Å for IIb and 2.63 Å for IIa, respectively) compared with the isolated subdomains (**Table 1**). In contrast, the global precision of the ensemble of domain II structures was much worse, with an r.m.s. deviation of 5.79 Å. This confirms that the 60 domain II RDCs are necessary to define the global structure of domain II (**Table 1**).

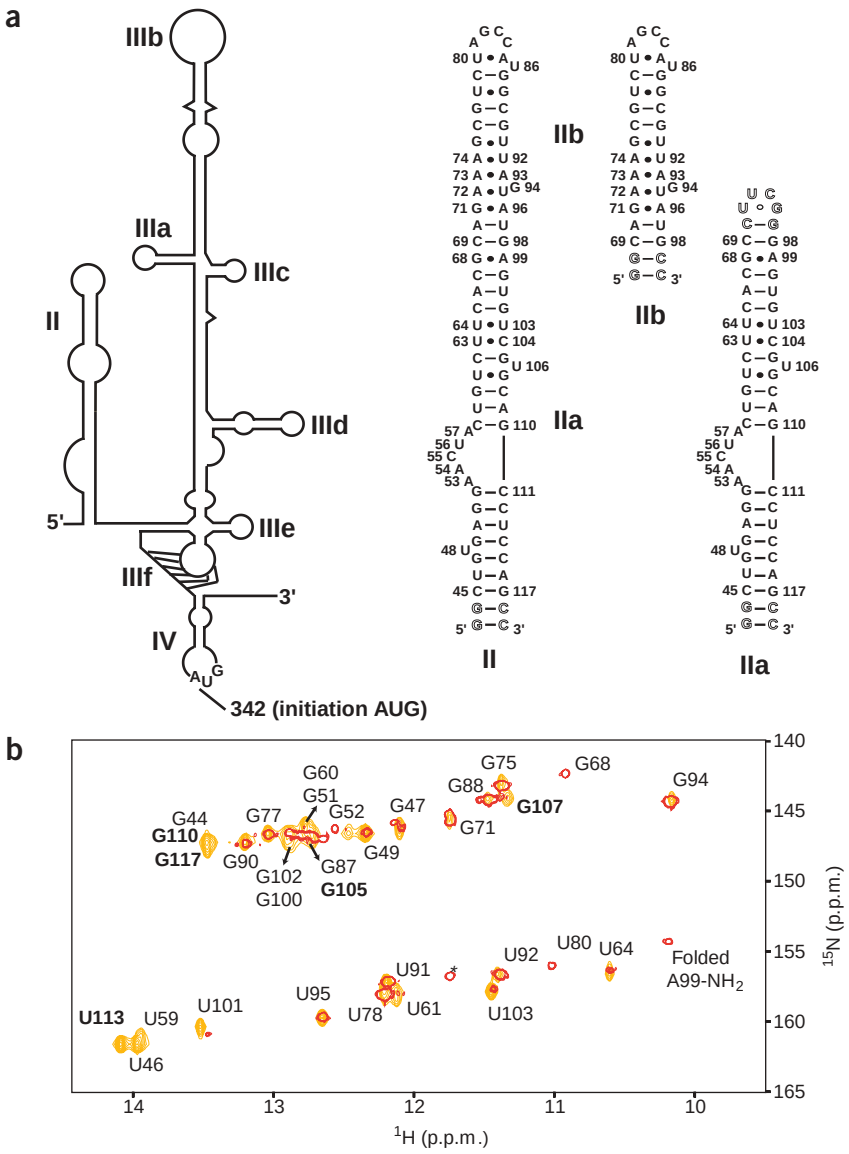

Figure 1 NMR structure determination of the HCV IRES domain II structure. (**a**) Secondary structure of HCV IRES RNA domain II. Left, secondary structure of HCV IRES RNA (nucleotides 1–367 of HCV genotype 1b). Domains are numbered according to ref. 10. Right, sequence and secondary structure of the HCV IRES domain II RNA oligonucleotides used for NMR structural studies; secondary structure of domain II is presented according to ref. 11. Outlined nucleotides were changed to improve transcription efficiency. The RNA oligonucleotides IIb (34 nt) and IIa (55 nt) comprised nt 69–98 and nt 45–69 and 98–117 of domain II, respectively. Both IIb and IIa contained two additional G-C base pairs analogous to II; the apical end of IIa was capped by an additional C-G base pair and a UUCG tetraloop. (**b**) Domain II alone adopts the same conformation in the context of the 100 kDa HCV IRES: comparison of ^{15}N-HMQC (heteronuclear multiple quantum coherence) spectra of imino N-H correlations of isolated ^{13}C-^{15}N-labeled domain II (25 kDa) and segmentally (nt 40–104) ^{15}N-labeled IRES (100 kDa). Domain II resonances are yellow, whereas resonances from domain II in the context of the IRES are red. Domain II residues that are not ^{15}N-labeled in the segmentally labeled sample, are shown in bold. Resonances of U46, U59 and U61 (segmentally labeled sample) and resonances of G68 and U80 (domain II sample) are observed at lower thresholds of the spectra (data not shown). An unassigned resonance is marked by an asterisk.

Overall structure and domain IIa

Domain II forms a bent structure (**Fig. 3a**). The observed secondary structure of domain IIa differs from the predicted secondary structure[10], but agrees with a recent proposal based on phylogenetic comparison of HCV isolates[11] (**Fig. 1b**). A stretch of noncanonical and Watson-Crick base pairs creates a continuous helix with the apical part of domain II. The single-stranded nucleotides A53, A54,

C55, U56 and A57 form an asymmetric internal loop, which introduces a bend (bend angle = 84.5 ± 9.6°) between the lower and upper helical regions of domain II (**Fig. 3b**). A53 stacks on top of the last G52-C111 base pair of the lower stem and base stacking continues throughout A54 and C55. A change in backbone direction occurs at residue U56 (non A-form values for torsion angles α = *gauche*$^+$, γ = *trans*, δ = *trans* and ζ = *gauche*$^+$), which allows stacking of A57 below the last C58-G110 base pair of the upper helix (**Fig. 3b**).

Mg^{2+} stabilizes the conformation of the hinge bulge (A53–A57). Addition of 5 mM Mg^{2+} causes upfield shifts of A54, C55 and A57 base ^{1}H resonances and reduced solvent exchange of the first two base pairs of the upper stem. The metal ion binding site(s) within the hinge bulge were identified through paramagnetic line broadening in the presence of substoichiometric amounts of Mn^{2+}, which can bind to similar RNA sites as Mg^{2+} (ref. 12). This paramagnetic probe caused specific broadening of resonances C2-H2 of A54 (**Fig. 3c**) and C8-H8 of G51 and G52, identifying one or potentially two metal ion binding sites in the major groove of the internal loop (**Fig. 3b**) with adenine N1 or N3 and guanine N7 positions as possible ligands for the metal ions. The location of the binding site is further supported by the fact that no broadening was observed for resonance C2-H2 of A53, as its Watson-Crick face points toward the minor groove (**Fig. 3b**).

Although Mg^{2+} stabilizes the hinge bulge, it is not a rigid structural motif. The ensemble of the 12 final structures contains two distinct conformations with the base moiety of residue U56 either stacked in (25%) or looped out (75%). Both conformations are equally supported by 23 inter-residue NOE restraints.

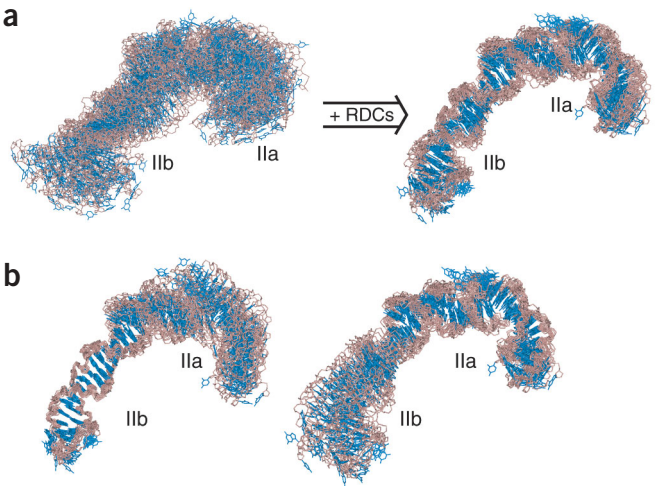

Figure 2 Structure determination of domain II using RDCs. (**a**) Heavy-atom superposition of the final 20 and 12 structures of domain II calculated with and without RDCs. Bases are blue and ribose-phosphate backbone is pink. (**b**) Local heavy-atom superposition of domain IIb and IIa of the final 12 structures of domain II calculated with RDCs.

Domain IIb

Domain IIb contains a hairpin loop and a loop E motif that are very similar to motifs previously found in HCV IRES domain IIId[13] (**Fig. 4a**). The loop E motif, formed by noncanonical base pairs and a local backbone reversal (S-turn), has a similar structure to other loop E motifs observed in RNAs[14]. Despite differences in loop sequence and size, both IIb and IIId hairpin loops adopt similar folds in which the 3′-side uracil residue (U269 in IIId and U86 in IIb, respectively) is looped out. In domain IIId, both the looped-out uracil and S-turn from loop E are located on the 3′ side of the domain, whereas they reside on opposite sides in domain IIb (**Fig. 4a**). This creates very different binding surfaces in both subdomains for the interaction with the 40S particle.

The helical stretch between the loop E motif and the asymmetric internal loop introduces a second, less pronounced bend (bend angle = 46.8 ± 13.2°) and contributes to the interaction surface with the 40S subunit through several noncanonical base pairs (**Fig. 4b**), such as U64•U103, U63•C104, U61•G107 flanked by a dynamic, looped out residue U106 and a previously unobserved G•A base pair between G68 and A99, involving hydrogen bonds between the O6 and N7 of G68 and N6 and N1 of A99 (**Fig. 4b**). The N1 position of A99 is protonated at pH 6.40, as shown by unusual proton and nitrogen chemical shifts,

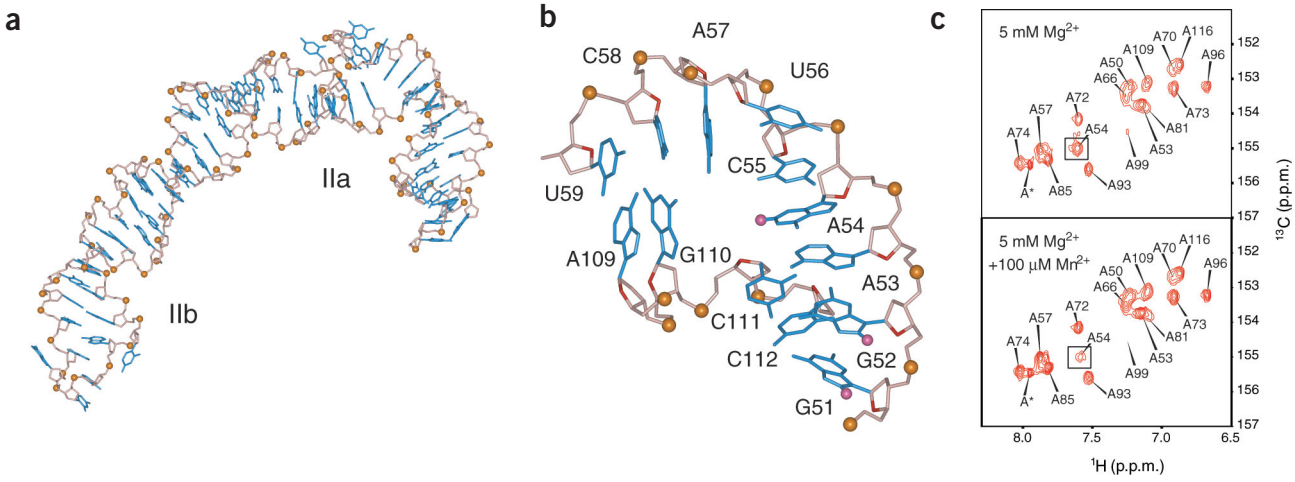

Figure 3 Solution structure of HCV IRES RNA domain II and magnesium binding sites. (**a**) Energy-minimized average structure of domain II. Color scheme as in **Figure 2**; phosphorus atoms are yellow. (**b**) Single representative view of the multinucleotide bulge from the major groove. Numbering according to **Figure 1a**. Color scheme as in **Figure 3a**; ribose O4′ atoms are red. Protons that specifically broadened in the presence of a paramagnetic probe are purple. (**c**) ^{13}C-HSQC (heteronuclear single quantum coherence) spectra of adenine C2-H2 correlations in the absence (top) and presence (bottom) of Mn^{2+} ions. Numbering according to **Figure 1a**; the *n* + 1 residue is labeled A*. The C2-H2 resonance of A54, which broadens, is boxed. Additionally, the initially broad C2-H2 resonance of A99 (top) broadens beyond detection (bottom).

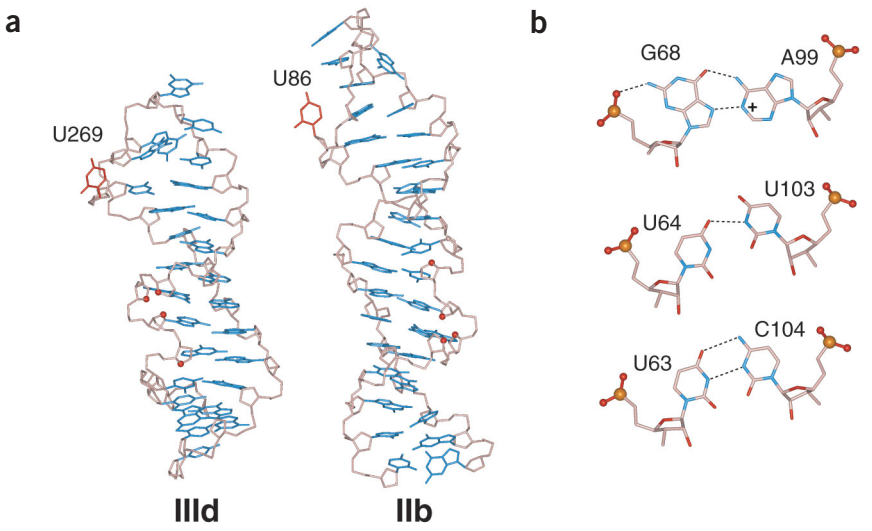

Figure 4 Comparison of subdomains IIb and IIId of the HCV IRES and base pairing schemes. (**a**) Single representative view of domains IIId and IIb. Color scheme as in **Figure 2**; ribose O4′ atoms within the loop E S-turn are red. The dynamic, looped-out uracil base of both hairpin loops is also red. (**b**) Base pairing schemes found in the helical stretch between the loop E motif and the asymmetric internal loop of domain II. Numbering according to **Figure 1a**. Base nitrogens are blue, ribose-phosphate backbone is pink, ribose O2′ and O4′ are red; phosphorus atoms and phosphate oxygens are yellow and red, respectively. Hydrogen bonds are shown as dashed lines.

15.58 p.p.m. for ^{1}H and 156.9 p.p.m. for ^{15}N, respectively. The stretch of noncanonical base pairs is the only major site of sequence variations among HCV isolates[15], yet these variations maintain the general features of the structure described above.

DISCUSSION

HCV IRES domain II is an independently folded domain of the IRES. The chemical shifts of resonances in segmentally labeled domain II in the intact, 100 kDa IRES are the same as those of the isolated domain II. The high-resolution NMR structure of domain II markedly resembles the 40S subunit–bound form as determined by cryo-EM[4] . In contrast, preliminary NMR studies of domain III in the free form indicate a rather dynamic structure as compared with domain II (data not shown). The flexibility of domain III might facilitate conformational changes required for its high-affinity interaction with the 40S subunit. Domain II interacts with 40S subunits only in the context of the intact IRES and docks its preformed solution conformation into the 40S subunit at the edge of head and platform[4] near the tRNA exit site. The global similarity of the domain II and tRNA structures is constrained by the shape of the ribosomal tRNA binding sites.

The structural features of domain II, such as the loop E motif and noncanonical pairs, may provide a stable point for RNA-RNA or RNA-protein interactions. In contrast to the protein-rich surface of the ribosome, the environment around domain II probably includes RNA. Dynamic looped-out uracil residues may allow for minor conformational adjustments upon 40S subunit binding. Interaction of domain II closes the mRNA cleft on the small ribosomal subunit and thus probably aids proper start-codon selection, suggesting that domain II functions like an initiation factor, or assists host initiation factors in their function. The structure of HCV IRES domain II indicates a modular organization of the HCV IRES with subdomains that have distinct functional roles during IRES-mediated initiation.

Biological RNAs are often much larger than the current size limits for NMR. By combining high-resolution structural information, long-range conformational restraints and segmental labeling strategies, we have shown that structures of 25-kDa RNAs can be determined and 100-kDa RNAs are accessible to NMR measurements. Our results underscore the power of NMR to define secondary and tertiary structures of RNA, and map RNA-ligand interactions in large biological RNAs such as messenger RNAs.

METHODS

Sample preparation. RNA oligonucleotides were prepared by transcription from linearized plasmid DNA using phage T7 RNA polymerase, followed by purification using PAGE[16], elution from the gel and equilibration with centrifugal devices against final buffer[17] (10 mM sodium phosphate, pH 6.4, 100 mM sodium chloride). Unlabeled and ^{13}C-^{15}N-labeled RNA oligonucleotides were prepared. Labeled nucleoside triphosphates were prepared in our laboratory using published methods[18]. NMR samples were prepared in a Shigemi NMR tube (250 µl containing 4% or 100% D_2O and 0.25 mM d_{12}-EDTA) at RNA concentrations of 0.8–2.5 mM. $MgCl_2$ was added from a 1 M stock solution up to a concentration of 5 mM. Weakly aligned NMR samples were prepared by addition of 8 mg ml^{-1} (II) or 10 mg ml^{-1} (IIa and IIb) of filamentous phage Pf1 (ref. 19).

NMR spectroscopy. NMR data were acquired at 15, 25 or 30 °C on Varian Inova 500 MHz and 800 MHz NMR spectrometers equipped with triple-resonance *x*-, *y*- and *z*-axis gradient probes. ^{1}H, ^{13}C, ^{15}N and ^{31}P assignments were obtained using standard homonuclear and heteronuclear methods optimized for RNA structure determination (RnaPack, http://www.varianinc.com) following detailed published procedures[17]. NMR data were processed with VNMR 6.1C (http://www.varianinc.com) and spectra were analyzed with SPARKY[20]. Base pairing schemes were established using the HNN-COSY experiment[21]. ^{1}H, ^{13}C, ^{15}N and ^{31}P chemical shifts are available online at http://puglisi.stanford.edu. NOEs from exchangeable protons were characterized as either strong (1.8–3.5 Å), medium (1.8–4.5 Å), weak (1.8–6 Å) or very weak (3.5–6.5 Å) based on their NOE peak intensities at mixing times of 50 and 120 ms, whereas NOEs from nonexchangeable protons were characterized as either strong (1.4–3.0 Å), medium (1.8–4 Å), weak (1.8–5 Å) or very weak (3.5–6.5 Å) based on their NOE peak intensities at mixing times of 50, 150 and 250 ms. Dihedral torsion restraints were obtained from DQF-COSY, 3D HMQC-TOCSY, HP-COSY and 3D HCP experiments as described[17]. Heteronuclear one-bond $^{1}J_{NH}$ and $^{1}J_{CH}$ couplings were measured in the absence and presence of filamentous phage Pf1 using phase cycle–modified TROSY[22,23] pulse sequences, which allowed detection of either the upfield or downfield component of the downfield ^{15}N or ^{13}C doublet, respectively. A constant-time ^{13}C evolution was applied for the measurement of $^{1}J_{CH}$ couplings. The RDC values were calculated as $^{1}D_{XH} = {}^{1}J_{XH}$ (aligned) $- {}^{1}J_{XH}$ (isotropic), and are available online at http://puglisi.stanford.edu. On the basis of the length of the time domain data and the standard deviation of at least two individual measurements, the accuracy of the measured couplings was estimated at ±2.5 Hz ($^{1}D_{CH}$) and ±1.25 Hz ($^{1}D_{NH}$), respectively. The $^{1}D_{NH}$ values and errors were normalized to ^{13}C.

Structure calculation. The protocol for structural calculations included simulated annealing of starting structures with random angles within the X-PLOR

Table 1 Structural statistics and atomic r.m.s. deviations for final domain II, IIa and IIb structures calculated with RDCs

	<SA-II>	<SA-II*>r	<SA-IIa>	<SA-IIb>
Final forcing energies (kcal mol^{-1})[a]				
Distance restraints (1744 [II], 1146 [IIa], 831 [IIb])	16.9 ± 1.1	17.8	14.0 ± 1.3	8.6 ± 0.6
Dihedral restraints (523 [II], 419 [IIa], 253 [IIb])	8.1 ± 1.3	6.9	5.7 ± 0.5	4.0 ± 0.9
Residual dipolar coupling restraints[b] (261 [II], 136 [IIa], 105 [IIb])	31.5 ± 3.2	31.1	18.4 ± 2.7	6.8 ± 1.1
R.m.s. deviations from experimental restraints				
Distance restraints (Å)	0.01 ± 0.001	0.01	0.02 ± 0.001	0.01 ± 0.001
Dihedral restraints (°)	0.67 ± 0.05	0.62	0.07 ± 0.03	0.7 ± 0.09
Residual dipolar coupling restraints (Hz)[b]				
$1D_{NH}$ and $1D_{CH}$ of II (60 [II])	2.07 ± 0.11	1.91	–	–
$1D_{NH}$ and $1D_{CH}$ of IIa (112 [II], 136 [IIa])	2.43 ± 0.06	2.34	2.5 ± 0.05	–
$1D_{NH}$ and $1D_{CH}$ of IIb (89 [II], 105 [IIb])	2.26 ± 0.12	2.30	–	2.23 ± 0.06
R.m.s. deviations from idealized covalent geometry				
Bonds (Å)	0.003 ± 0.00004	0.003	0.003 ± 0.00003	0.003 ± 0.00005
Angles (°)	0.82 ± 0.004	0.82	0.8 ± 0.006	0.8 ± 0.003
Impropers (°)	0.36 ± 0.03	0.35	0.3 ± 0.02	0.3 ± 0.01
Heavy-atoms r.m.s. deviations (Å)				
All RNA	2.18	1.68	2.34	1.35
IIa (G43–C69, G98–C119)	1.62	1.78	2.22	
IIb (A70–U97)	1.15	1.03	–	1.35
IIa internal loop (G52–C58, G110–C111)	1.01	0.77	1.25	
IIb hairpin loop (C79–G87)	1.50	1.27	–	1.47
IIb loop E motif (G71–A74, U92–A96)	0.41	0.33	–	0.40

The notation of the NMR structures is as follows: <SA-II>, <SA-IIa>, and <SA-IIb> are the final 12, 22 or 20 simulated annealing domain II, IIa or IIb structures, respectively; SA-II* is the mean domain II structure obtained by averaging the coordinates of the individual SA-II structures best-fitted to each other; <SA-II*>r is the regularized mean structure obtained by restrained minimization of the mean structure SA-II*. The number or terms for the various restraints is given in parentheses; for final force values see Methods. Heavy-atom r.m.s. deviations are defined as the average difference between <SA> or <SA*>r versus SA*.

[a]None of the final structures showed distance violations >0.25 Å and dihedral violations >6.5°. Only structurally useful intraresidue distance restraints, involving protons separated by more than three bonds, are included. [b]The r.m.s. deviations of chemical shifts (p.p.m.) for IIa/IIb versus II are as follows: 0.099 (^{13}C) and 0.074 (^{1}H) for aromatic C-H groups and 0.055 (^{15}N) and 0.028 (^{1}H) for aromatic N-H groups, respectively. All $^{1}D_{NH}$ residual dipolar couplings are normalized to ^{13}C. The axial (D_a) and rhombic (R) components of the individual final alignment tensors are $D_a = -12.93$ and $R = 0.45$ for IIa, $D_a = -15.28$ and $R = 0.17$ for IIb and $D_a = -26.40$ and $R = 0.29$ for II, respectively.

3.1 package[24] and restrained molecular dynamics (rMD) refinement within the CNS package[25]. starting structures (200 domain II or 100 domain IIa and IIb) were generated and subjected to a simulated annealing protocol excluding RDC restraints (for more details, see **Supplementary Methods** online). Structures that met the acceptance criteria (no NOE violation >0.5 Å and no torsion angle violation >10°) were then used to estimate D_a and R of the alignment tensor by determining the best fit of observed RDCs to ensemble members using the singular value decomposition method[26] implemented in PALES[27]. First, iterative refinements of domains IIa and IIb with RDC restraints were carried out until the rhombicity of the lowest-energy structures met acceptance criteria ($\Delta R < \pm 0.015$). These final alignment tensors for domains IIa and IIb (**Table 1**) were then used for the refinement of domain II together with a third alignment tensor for the domain II RDC restraints, which was subsequently improved using the same iterative procedure. The final 12 structures were chosen that had the lowest total and restraint violation energies. Structures were analyzed using the programs CURVES[28] and MolMol[29] and structure figures were generated using InsightII (http://www.accelrys.com).

Coordinates. Structures have been deposited with the Protein Data Bank (accession codes 1P5M (ensemble of domain IIa), 1P5N (ensemble of domain IIb), 1P5O (ensemble of domain II) and 1P5P (average minimized domain II structure)).

Note: Supplementary information is available on the Nature Structural Biology website.

ACKNOWLEDGMENTS

We thank E. Lau for labeled nucleotide preparation and P. Sarnow, Y. Shibata-Lukavsky, R.L. Gonzalez, D. Daniels, C. Liu and E.V. Puglisi for helpful discussions. This work was supported by grants from the US National Institutes of Health, the Hutchison Foundation and Eli Lilly. Stanford Magnetic Resonance Laboratory is supported by the Stanford School of Medicine.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 28 April; accepted 5 September 2003
Published online at http://www.nature.com/naturestructuralbiology/

1. Pestova, T.V., Shatsky, I.N., Fletcher, S.P., Jackson, R.J. & Hellen, C.U. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **12**, 67–83 (1998).
2. Sachs, A.B., Sarnow, P. & Hentze, M.W. Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* **89**, 831–838 (1997).
3. Kieft, J.S., Zhou, K., Jubin, R. & Doudna, J.A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**, 194–206 (2001).
4. Spahn, C.M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* **291**, 1959–1962 (2001).
5. Otto, G.A., Lukavsky, P.J., Lancaster, A.M., Sarnow, P. & Puglisi, J.D. Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction. *RNA* **8**, 913–923 (2002).
6. Kim, I., Lukavsky, P.J. & Puglisi, J.D. NMR study of 100kDa HCV IRES RNA using segmental isotope labeling. *J. Am. Chem. Soc.* **124**, 9338–9339 (2002).
7. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
8. Clore, G.M., Gronenborn, A.M. & Tjandra, N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.* **131**, 159–162 (1998).
9. McCallum, S.A. & Pardi, A. Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J. Mol. Biol.* **326**, 1037–1050 (2003).
10. Honda, M., Beard, M.R., Ping, L.H. & Lemon, S.M. A phylogenetically conserved stem-loop structure at the 5′ border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.* **73**, 1165–1174 (1999).
11. Zhao, W.D. & Wimmer, E. Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. *J. Virol.* **75**, 3719–3730 (2001).
12. Gonzalez, R.L., Jr. & Tinoco, I. Jr. Identification and characterization of metal ion binding sites in RNA. *Methods Enzymol.* **338**, 421–443 (2001).
13. Lukavsky, P.J., Otto, G.A., Lancaster, A.M., Sarnow, P. & Puglisi, J.D. Structures of two essential RNA domains for hepatitis C virus internal ribosome entry site function. *Nat. Struct. Biol.* **7**, 1105–1110 (2000).
14. Correll, C.C. *et al.* Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. USA* **95**, 13436–13441 (1998).
15. Smith, D.B. *et al.* Variation of the hepatitis C virus 5′ non-coding region: implications

for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *J. Gen. Virol.* **76**, 1749–1761 (1995).
16. Puglisi, J.D. & Wyatt, J.R. Biochemical and NMR studies of RNA conformation with an emphasis on RNA pseudoknots. *Methods Enzymol.* **261**, 323–350 (1995).
17. Lukavsky, P.J. & Puglisi, J.D. RNAPack: an integrated NMR approach to RNA structure determination. *Methods* **25**, 316–332 (2001).
18. Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D. & Williamson, J.R. Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.* **20**, 4515–4523 (1992).
19. Hansen, M.R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**, 1065–1074 (1998).
20. Goddard, T.D. & Kneller, D.G. *Sparky 3*. (University of California, San Francisco, 2000).
21. Dingley, A.J. & Grzesiek, S. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J. Am. Chem. Soc.* **120**, 8293–8297 (1998).
22. Weigelt, J. Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *J. Am. Chem. Soc.* **120**, 10778–10779 (1998).
23. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366–12371 (1997).
24. Brunger, A.T. *X-PLOR, Version 3.1. A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, Connecticut, USA, 1992).
25. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
26. Losonczi, J.A., Andrec, M., Fischer, M.W. & Prestegard, J.H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138**, 334–342 (1999).
27. Zweckstetter, M. & Bax, A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122**, 3791–3792 (2000).
28. Lavery, R. & Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **6**, 655–667 (1989).
29. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55, 29–32 (1996).