A potential RNA drug target in the hepatitis C virus internal ribosomal entry site

ROSCOE KLINCK,¹ ERIC WESTHOF,² STEPHEN WALKER,¹ MOHAMMAD AFSHAR,¹
ADAM COLLIER,¹ and FAREED ABOUL-ELA¹

¹RiboTargets Ltd. Granta Park, Abington, Cambridge, CB1 6GB, United Kingdom

²Institut de Biologie Moléculaire et Cellulaire/Centre National de la Recherche Scientifique, Strasbourg, F-67084, France

ABSTRACT

Subdomain III_d from the hepatitis C virus (HCV) internal ribosome entry site (IRES) has been shown to be essential for cap-independent translation. We have conducted a structural study of a 27-nt fragment, identical in sequence to III_d, to explore the structural features of this subdomain. The proposed secondary structure of III_d is comprised of two 3 bp helical regions separated by an internal loop and closed at one end by a 6-nt terminal loop. NMR and molecular modeling were used interactively to formulate a validated model of the three-dimensional structure of III_d. We found that this fragment contains several noncanonical structural motifs and non-Watson–Crick base pairs, some of which are common to other RNAs. In particular, a motif characteristic of the rRNA α -sarcin/ricin loop was located in the internal loop. The terminal loop, 5'-UUGGGU, was found to fold to form a trinucleotide loop closed by a *trans*-wobble U•G base pair. The sixth nucleotide was bulged out to allow stacking of this U•G pair on the adjacent helical region. In vivo mutational analysis in the context of the full IRES confirmed the importance of each structural motif within III_d for IRES function. These findings may provide clues as to host cellular proteins that play a role in IRES-directed translation and, in particular, the mechanism through which host ribosomes are sequestered for viral function.

Keywords: 5' UTR; bioinformatics; drug design; G•U pair; HCV IRES; NMR; RNA motif; S-motif; sarcin-ricin loop (SRL)

INTRODUCTION

Infection by hepatitis C virus (HCV), the main causative agent of posttransfusion hepatitis, frequently leads to liver cirrhosis and hepatocellular carcinoma (Saito et al., 1990). An estimated 170 million people are infected by the virus, of whom 10–20% are expected to suffer these life-threatening complications (Cohen, 1999). The viral proteins are translated from a 9.5-kb single-stranded positive sense RNA, which is flanked by 5' and 3' untranslated regions (UTRs) (Houghton et al., 1994). The 5' UTR represents the most conserved region of the genome (Bukh et al., 1992). It forms the highly structured internal ribosome entry site (IRES) (nt 40–370), which directs translation in a cap-independent manner (Reynolds et al., 1996), a mechanism first observed in picornaviruses (Jackson et al., 1990).

The current secondary structure model of the HCV 5' UTR indicates the presence of four major structural domains (Fig. 1a), which, except for Domain I, all con-

tribute to IRES activity (Brown et al., 1992). In particular, mutational analysis indicates that several subdomains within Domain III, including III_d (Fig. 1b, left), are critical in HCV IRES mediated translation (Kieft et al., 1999). Thus III_d represents a potential target for structure-based drug design.

Until now structure-based drug design has been limited largely to protein targets, because of the relative lack of RNA structural information. However, the availability of high-resolution three-dimensional RNA structures has undergone a dramatic increase over recent years (Ramos et al., 1997; Batey et al., 1999). Analysis of these coordinates reveals that, although novel RNA motifs are frequently being reported, several structural motifs are repeated in different contexts throughout the structure database. Examples of these include tetraloops, U-turns, dinucleotide platforms, and the α -sarcin/ricin loop (SRL) motif.¹ The available repertoire of RNA motifs, albeit incomplete, can provide considerable insight into the three-dimensional structure of RNAs.

Reprint requests to: Fareed Aboul-ela, RiboTargets Ltd. Granta Park, Abington, Cambridge, CB1 6GB, United Kingdom; e-mail: fareed@ribotargets.com.

¹This motif has also been referred to as the "S-motif," "loop E motif," or "S-turn." Because our data is most closely comparable to that previously reported for the SRL, we use the term SRL motif in this paper.

FIGURE 1. **A:** HCV IRES sequence showing the predicted secondary structure (Honda et al., 1999). The IIIId stem-loop (253–279) and the element in Domain II that contains a putative SRL motif are boxed. **B:** IIIId fragment used for NMR (left), mutant pIIIId264-T-loop (center), and pIIIId275CUC (right). The naturally occurring sequence variant of IIIId is shown in the inset. This variant preserves the formation of the secondary structure, and the presence of consecutive G-C and G-U base pairs closing the stem at the base of the terminal loop. **C:** Secondary structure and base pairing geometry of the SRL motif. **D:** Plasmid construct used for translational assay.

Indeed, it is often possible to correctly predict the presence of known motifs by combining phylogeny and sequence-specific criteria relating to their three-dimensional structure (Moore, 1999; Westhof & Fritsch, 2000). It is thus reasonable to propose that motif prediction, used in conjunction with experimental evidence, can be used as a tool for structure prediction.

Here we report the use of a motif-based approach, which combines motif prediction, NMR, and mutational analysis, to reveal features of the three-dimensional structure of the HCV 5' UTR subdomain IIIId. We demonstrate that IIIId can be described as the juxtaposition of six independent motifs, including standard A-form helices, an SRL motif, and two non-Watson-Crick base pairs. Using available structural data, NMR validates the motif predictions and their arrangement within IIIId, and site directed mutagenesis confirms the importance of these motifs in the context of the whole IRES. The findings presented here have significant implications for the functional role and mechanism of HCV IRES-directed translation, and the simultaneous use of bioinformatics, NMR, and molecular biology for characterizing RNA structures has great potential for the field of RNA structure-based drug design.

RESULTS

Internal ribosome entry site and IIIId in hepatitis C virus

The primary sequence of IIIId shows only two changes between the six major HCV genotypes, both of which represent covariant substitutions (Fig. 1B). To investigate the functional role of IIIId, two mutants were constructed (see Materials and Methods; Fig. 1). The substitution mutants, pIIIId264-T-loop and pIIIId275CUC, replaced structural elements in the terminal and internal loop, respectively (Fig. 1B, center and right). In vivo analysis indicates that pIIIId264-T-loop abrogated IRES activity to 7% of wild-type levels and pIIIId275CUC resulted in a reduction to 25%. Dual substitution mutants within the IIIId terminal loop demonstrated reductions in activity in the range of 18–42% for the U264:U269 series (Table 1).

Single-point mutants within the IIIId internal loop at A260 and A276 caused reductions in activity of 12–42% and 45–57%, respectively. For this series, the mutant A260G produced the most dramatic abrogation of IRES activity for a single substitution mutant. In ad-

TABLE 1. Summary of translational activities of HCV subdomain III_d mutations, relative to wild type.

Mutation	Activity (%)
A260C	40
A260C; G insertion	1
A260U	47
A260G	12
A276U	57
A276G	45
A276G; U insertion	7
U264G, U269G	25
U264C, U269A	27
U264A, U269C	42
U264A, U269A	27
U264C, U269C	27
U264G, U269A	18
U264G, U269C	27
U264C, U269G	18
pIII _d 264-T-loop	7
pIII _d 275CUC	25

dition, dual mutants consisting of a single substitution with the insertion of a residue resulted in extreme abrogation of activity (Table 1).

In summary, multiple and point substitution mutations within both the III_d internal and III_d terminal loop indicate the crucial role of III_d in HCV IRES activity.

α-Sarcin/ricin loop motifs in the HCV IRES

The SRL motif, defined as an asymmetric internal loop, (Fig. 1C), was first identified as a recurrent structural element in 1985 (Branch et al., 1985). The motif is present in both the large subunit rRNA SRL and the loop E region of eukaryotic 5S rRNA, with isosteric sequence variations (Leontis & Westhof, 1998a). The

structure of the motif, characterized by a series of three non-Watson–Crick base pairs, has been studied extensively using NMR and X-ray crystallography (Wimberly et al., 1993, Szewczak & Moore, 1995, Correll et al., 1999). The essential structural characteristics of the motif, and the underlying mechanisms leading to its formation have been analyzed systematically (Leontis & Westhof, 1998a). As a result, though the motif is defined by conformation rather than by base sequence, the nucleotide requirements for the formation of the motif can be predicted. Based on this analysis, the motif was predicted to occur twice in the HCV IRES, in Domain II_b and III_d (Fig. 1A, boxed regions).

For III_d, this prediction was confirmed by NMR spectroscopy. Data were acquired on a 27-nt fragment of the IRES, identical in sequence to nt 253–279 (Fig. 1B, left). The distinct SRL geometry leads to a number of nonstandard chemical shift and NOE patterns (Szewczak & Moore, 1995). We searched the spectra of III_d for these distinctive patterns, and initiated the assignment process by postulating that they could be identified with the analogous positions within the proposed III_d SRL motif. Virtually without exception, every diagnostic signal expected for the SRL motif is identified in the III_d fragment (Table 2 and Fig. 2).

Figure 2A shows the pattern of NOEs connected with the U259 imino resonance (vertical slice on the right), confirming the formation of a *trans*-Hoogsteen type U•A base pair. The NOE between the U259 imino proton resonance (12.5 ppm) and the A274 H8 resonance (7.4 ppm) is indicative of the reverse Hoogsteen base pairing predicted by the SRL motif. The NOEs involving the two amino resonances of A274 (centered around 6.5 ppm) are also predicted by the base-pairing scheme. Moreover, the chemical shift at which the amino resonances appear is itself an indication of the formation of a nonstandard structure.

TABLE 2. Comparison of chemical shifts observed for analogous proton and carbon resonances for the sarcin-ricin loop RNA and the SRL motif within HCV III_d.^a

	H8/H6	H5/H2	H1'	C1'	H1/H3	NH ₂
G273(G19)	7.8(7.51)		5.66(5.17)	87.6(87.5)	10.81(11.5)	
A260(A12)	7.49(7.54)	6.85(6.94)	5.63(5.73)	90.9(90.8)		
U259(U11)	7.96(7.96)	5.85(5.89)	5.47(5.63)	no(no)	12.51(12.79)	
A274(A20)	7.43(7.55)	7.78(7.91)	5.93(5.98)	92.2(91.4)		6.61/6.53 (6.82/6.45)
G258(G10)	8.1(no)		5.9(5.96)	81.8(81.3)	10.11(10.18)	
A257(A9)	8.21(8.54)	7.79(7.7)	5.62(5.66)	85.7(85.6)		(8.72/6.72)
A275(A21)	7.73(7.82)	7.14(7.84)	5.61(5.57)	89.7(no)		

^aResonances for nonexchangeable proton and C1' carbon resonances have been measured at 303 K, and exchangeable proton chemical shifts have been measured at 278 and 283 K for the III_d and SRL elements, respectively. Assignments from Szewczak and Moore (1995) are reported in parentheses. Additional III_d assignments for aromatic, H1', C1', exchangeable proton and some additional sugar resonances are listed in Table 3. "No" indicates that the resonance in question could not be assigned. Italicized assignments are featured in Figure 3. Note that buffer conditions in the SRL study differed somewhat from those used here.

FIGURE 2. NMR data demonstrating presence of an SRL motif in HCV IIId. **A:** The region of a NOESY (150 ms mixing time) spectrum acquired in H_2O , illustrating the NOEs between imino proton and aromatic/amino proton resonances. A one-dimensional slice through the spectrum corresponding to the U259 imino resonance is presented in the inset at right, and the corresponding horizontal slice corresponding to the A274 H8 resonance is shown underneath the two-dimensional spectrum. **B:** Region of a NOESY spectrum (120 ms mixing time) acquired in D_2O showing NOEs between aromatic proton resonances. The characteristic NOE between the H2 proton resonances of A260 and A274 is highlighted. **C:** Natural abundance ^1H - ^{13}C HMQC spectrum of IIId RNA. In this spectrum, the carbon spectral window has been chosen such that carbon signals occurring below 85 ppm, which normally include nonanomeric sugar resonances, appear folded in as negative peaks (red), and the anomeric sugar and aromatic carbons appear as positive signals (blue). One folded peak appears in the anomeric proton region (5.9 ppm). Anomalous carbon and proton chemical shifts are highlighted (see text).

An intense cross-strand A260H2-A274H2 NOE, indicative of a sheared A260 $\cdot$ G273 pair stacking on the *trans*-Hoogsteen U259 $\cdot$ A274 pair is shown in Figure 2B. This signal confirms that the overall geometry of the motif is conserved. Four unusual sugar carbon and proton shifts are highlighted in the region of the ^1H - ^{13}C correlation spectrum shown in Figure 2C. The three signals associated with the putative SRL motif are in very good agreement with the reported chemical shifts (Szewczak & Moore, 1995).

Confirming this postulated set of assignments was then trivial (see Materials and Methods).

The aromatic proton or sugar proton and carbon resonances that differ significantly between the HCV IIId and SRL fragments appear at either end of the region defined by the SRL motif, as expected based on differences in flanking base pairs (Table 2). The largest chem-

ical shift difference is observed at A275 H2, which appears well over 0.5 ppm upfield from the analogous resonance within the SRL.

The IIId terminal loop

Six-nucleotide terminal loops in RNA are often closed by a non-Watson-Crick or G $\cdot$ U base pair, leaving a U-turn or one of several known tetraloop motifs to facilitate the backbone turn. The task of assigning the terminal loop, nt 263–270, was greatly simplified by the previous assignments of the resonances from the SRL region, and analysis of the NMR data revealed novel structural features in this region. A paucity of internucleotide NOEs involving U269 indicated that this nucleotide was bulged out of the loop structure. NOEs between G268 and C270, shown in Figure 3A, suggest

TABLE 3. Chemical shift assignments of III_d at 303 K (imino resonances at 278 K).

	H2/H5	H6/H8	H1'	H2' (H3')	H1/H3	C1'
G253	NA ^a	8.14	5.73	4.79	—	—
C254	5.28	7.74	5.62	4.51	NA	—
C255	5.16	7.23	5.36	4.43	NA	—
G256	NA	7.48	5.64	4.20	10.53	88.42
A257	7.79	8.22	5.62	3.78 (4.56)	NA	85.71
G258	NA	8.10	5.90	4.90 (4.94)	10.11	81.79
U259	5.85	7.96	5.47	—	12.51	93.22
A260	6.85	7.49	5.63	—	NA	90.89
G261	NA	7.44	3.84	—	13.49	—
U262	5.45	7.49	5.42	4.22	11.50	—
G263	NA	7.84	5.75	4.47	12.93	—
U264	5.36	7.57	5.63	4.30	11.66	—
U265	5.64	7.71	5.62	4.34	10.70	—
G266	NA	7.80	5.54	4.64	—	—
G267	NA	7.76	5.67	4.51	—	—
G268	NA	7.95	5.89	4.88	10.98	—
U269	5.89	7.81	5.58	4.37	—	—
C270	5.75	7.80	5.50	4.55	NA	—
G271	NA	7.54	5.71	4.61	10.81	—
C272	4.88	7.14	5.36	4.52	NA	—
G273	NA	7.80	5.66	—	11.73	87.58
A274	7.78	7.43	5.93	5.21/4.18	NA	92.20
A275	7.14	7.73	5.61	4.62	NA	89.68
A276	8.31	7.49	6.15	4.82	NA	90.51
G277	NA	7.65	4.13	—	12.99	—
G278	NA	7.16	5.69	—	13.21	—
C279	5.20	7.46	5.74	3.98	NA	—

^aNA: not applicable.

not only that the two bases are stacked, but also that G268 undergoes a localized backbone inversion. The primary evidence for this conformation was the characteristic box pattern of normal and “backwards” aromatic to H1' NOEs (Fig. 3A) and a G268H1'–C270H1' NOE (not shown) at 400 ms mixing time. Though some of these signals are weak, all of them also appear at 120 ms mixing time (data not shown), and thus are unlikely to result solely from spin diffusion. The backbone inversion at G268 is accommodated by, and dependent upon, the bulged out nt U269. The arrangement of G268, U269, and C270, gives rise to an S-shaped backbone geometry, which has previously been observed in the SRL motif, and the NMR structure of the rev response element (Battiste et al., 1996).

Nucleotide G268 is also involved in a base pair with U264, indicated by imino proton NOEs shown in Figure 3B. As the moderate intranucleotide G268H8–H1' NOE (Fig. 3A) precludes a *syn* glycosidic angle for this nucleotide, the base pair must be a locally parallel *trans*-wobble U•G. This configuration should be contrasted with that observed in the UUCG tetraloop (Allain & Varani, 1995), involving a 2' hydroxyl and a bifurcated hydrogen bond between the O2 of uracil and imino and amino protons on guanosine. The latter base pairing leads to a very different pattern of chemical shifts. Both the U264 and G268 imino resonances show NOEs to

one another and to G263H1, consistent with stacking of the U264•G268 pair on the G263–C270 pair.

With U264 and G268 base paired, U265, G266, and G267 remain to accomplish the loop backbone turn. The U-turn motif, found in tRNA and in a number of RNA crystal and NMR structures (Moore, 1999) is an obvious candidate for this turn. Though predicted anomalous phosphorus shifts were not observed (Moore, 1999), all of the observed proton–proton NOEs, and the chemical shift of the putative U265 imino proton resonance were consistent with this motif (Fountain et al., 1996; Jang & Patel, 1998). Thus the U-turn is the most plausible model for the folding of this element.

A validated model for the three-dimensional structure of subdomain III_d

We have used NMR data in conjunction with a motif-based approach to model the three-dimensional structure of III_d (Fig. 4). Six motifs were used to construct the model: (1) an A-form double helix, (2) a sheared G•A pair, (3) the SRL motif, (4) a localized backbone inversion, (5) a *trans*-wobble U•G base pair, and (6) a U-turn. Each of these motifs has a characteristic NMR signature, which provided an experimental basis for its use.

The model was then carefully analyzed to ensure concordance with the NMR data. All theoretically probable and improbable NOEs expected from the terminal loop region of the model were compared to the NMR data. In all cases the NMR data was consistent with the model. Moreover, inspection of the model explained the few apparent discrepancies between the chemical shifts observed within the SRL motif within III_d and the analogous resonances in the SRL itself. Specifically, the upfield shift of the A275 H2 proton in III_d mentioned above is explained by its position immediately above the six-membered ring of A276. As in the earlier SRL NMR study (Szewczak & Moore, 1995), we could find no direct proof of the A257•A275 base pairing pictured in Figure 1. However, NOEs such as those observed between A275 H2 and A276 H2, and between A275 H8 and A276 H8, confirm the overall geometry of this region (data not shown). In addition, imino proton resonances could not be identified for G266, G267, and U269, which in our model are completely exposed to the solvent.

DISCUSSION

HCV IRES SRL motifs and implications for IRES function

III_d is the first reported occurrence of the SRL motif in a mammalian viral RNA of clinical importance. Our mutation data and the results of Doudna and coworkers (Kieft et al., 1999) show the functional relevance of the

FIGURE 3. NMR spectra of terminal loop. **A:** NOESY (400 ms mixing time) region showing the G268H1'–C270H6, G268H8–C270H5, and G268H8–C270H6 NOEs, consistent with G268/C270 base stacking. The geometry of the nucleotides G268, U269, and C270 is illustrated in the schematic on the right. The distances between G268H8 and C270H1' and G268H1' and C270H6 are 3.24 Å and 2.41 Å, respectively, in the three-dimensional model of IIIId (Fig. 4). **B:** Imino–imino proton region of NOESY (150 ms mixing time) in H₂O, showing the NOEs between imino proton resonances of U262 and G271, and between U264 and G268 imino resonances. In each case, the NOE is evidence of G•U base pairing. Also indicated is an upfield imino resonance tentatively assigned to U265. These data indicate the formation of a *trans*-wobble base pairing geometry for U264•G268, as shown on the right.

IIIId SRL motif and U•G *trans*-wobble pair within the HCV 5' UTR, as mutations that disrupt either element inhibit IRES activity (Table 1). Interactions with cellular factors are critical for IRES activity and analysis of the IIIId structure may provide clues to likely protein ligands. A number of SRL motifs have already been shown to act as substrates for ribosomal proteins (Leontis & Westhof, 1998a). Moreover the prokaryotic L25 and eukaryotic L5 proteins bind related loop E elements of their respective 5S rRNAs. It has also been suggested that the IIIId terminal loop is involved in tertiary interactions with subdomain IIIe (Kieft et al., 1999).

With regard to the second SRL motif predicted to be located in the internal loop of helix IIb (Fig. 1A), Rijnbrand et al. (1996) have shown that the mutant A96G significantly abrogates HCV IRES activity. This mutation would disrupt the predicted SRL motif, indicating that this motif is also critical for IRES function.

The structure of IIIId and implications for drug design

In isolation, the folding of the IIIId RNA fragment is unaffected by temperature or the addition of magnesium, and signals indicative of alternative conformers are absent under all of these conditions (data not shown). Thus, the fragment forms a robust secondary structure. Moreover NMR (Stoldt et al., 1999) and X-ray (Lu & Steitz, 2000) studies have shown that the SRL/loop E non-Watson–Crick base pairs are maintained in the presence of the L25:RNA complex, and in the 50S ribosomal subunit itself (Ban et al., 1999; Cate et al., 1999). Where the U-turn motif has been observed in tRNAs, and in the GTPase activating region hexaloop (Conn et al., 1999), the motif is maintained in the presence of tertiary interactions with other RNA loops. Thus it seems highly probable that

FIGURE 4. Model of subdomain IIIId structure. Schematic diagram of the modular strategy for building the IIIId model (left). The component building blocks for the starting structure have been two A-form helical regions (green), the SRL motif (red), the trans-wobble U-G base pair (yellow), and the U-turn (orange). Stereoview (DRAWNA 2.1) of IIIId model (right).

the IIIId structure studied here remains intact within the full length IRES.

The context dependence of the geometry of RNA motifs is presently not well appreciated. If there is any, it may be most pronounced at the interfaces between the motif and the helical regions. For example, based on analogy to the prokaryotic loop E motif (Correll et al., 1998), one might have predicted the formation of a face to face G256-A276 base pair. However, a detailed analysis of other related motifs (Correll et al., 1999) shows that the Y-Y pairs seen below the *trans* A-A pairs are geometrically closer to a sheared G-A pair than to the Watson-Crick/Watson-Crick G-A pairs (Leontis & Westhof, 1998b). This ambiguity is fully removed by the NMR data that confirm the phylogenetic and structural analysis. The data in Table 2 and Figure 2 indicate a remarkable conservation of the chemical environment and three-dimensional structure of the core of the SRL motif within two distinct RNA fragments. Moreover, this core of the SRL motif, consisting of the contiguous sheared and *trans*-Hoogsteen base pairs, folds in a conserved three-dimensional structure in a number of sequence contexts (Wimberly et al., 1993; Dallas & Moore, 1997; Correll et al., 1998; Leontis & Westhof, 1998a,b). The internal loop region must be considered the region of the IIIId model that presents the most definitive prediction of the actual three-dimensional structure of the molecule, whereas the largest degree of uncertainty applies to the U-turn region of the terminal loop.

The task of applying structure-based methodology to RNA targets is made more daunting by the necessity of reducing time scales for drug development, and by the large size of the functional units formed by folded RNAs. The model presented herein, the first detailed characterization of the three-dimensional structure of an HCV IRES subdomain, was constructed in a fraction of the time that would have been required for a *de novo* NMR or crystal-structure determination. Yet the end product is a very useful substrate for a drug design project. As more RNA three-dimensional structures solved by NMR appear in the literature, parallel databases of conserved RNA structure motifs and their NMR signatures are now being built. The combined use of NMR, structure/sequence analysis and molecular modeling promises to rapidly advance the field of RNA structure-function analysis and structure-based drug design.

MATERIALS AND METHODS

Plasmids

All substitution mutants were created by oligonucleotide site-directed mutagenesis using the template plasmid pTZ18:5442-16-1. This contains the HCV 1a 5' UTR (nt 18-356) in the *Bam*HI site of pTZ18U (Collier et al., 1998). 5' UTR mutants were subcloned into the *Bam*HI site of the dual reporter pRT9. The plasmid pRT9 contains the HIV-1 LTR (-340 to +78) that transcribes the bicistronic mRNA encoding renilla luciferase, the HCV 5' UTR and firefly luciferase upstream of renilla luciferase (Fig. 1D).

In vivo analysis of translation

Twenty-four hours prior to transfection, 35 cm² dishes were seeded with C63 cells (a gift from Dr. Jonathan Karn, MRC Laboratory of Molecular Biology, Cambridge). Transfections were performed with 1 µg of pRT9 or 5' UTR mutant derivative using calcium phosphate. After 40 h posttransfection, cells were harvested, lysed, and assayed for firefly and renilla luciferase activity using the Promega Dual-Luciferase assay according to the manufacturer's instructions (Promega).

NMR sample preparation

The 27-nt III_d analog was prepared by T7 RNA polymerase transcription using synthetic DNA templates (Milligan & Uhlenbeck, 1988). The transcripts were purified and dialyzed, as described (Varani et al., 1996). Final concentration of the NMR sample was 1.2 mM, with 8 mM Na-phosphate, pH 6.6, in 200 µL. Addition of sodium or magnesium chloride had no effect on the spectra.

NMR spectroscopy

All NMR spectra were recorded on Bruker DRX500 and DMX600 spectrometers. For analysis of the exchangeable protons, NOESY experiments were run at 5 and 25 °C, with 150 ms mixing time. A jump-return-WATERGATE sequence was used for water suppression (Liu et al., 1998). NOESY (60, 120, and 400 ms mixing times) TOCSY, and COSY-DQF experiments were run at 25 °C and 30 °C in D₂O. A natural abundance proton-carbon correlation experiment, ¹H-¹³C-HMQC, was performed at 30 °C. Proton chemical shifts were referenced to the residual water peak (4.77 ppm at 25 °C).

Assignment of proton and carbon resonances

The imino resonances were assigned using a NOESY spectrum at 5 °C. Sequential imino-imino connectivities were observed from G278 to G256. The 5' terminal G253 imino proton was not observed. The chemical shift of G256H1 was consistent with a sheared-type base pairing with A276 (Cheng et al., 1992). Sequential NOEs were also observed from U259 through the double helix up to U264, which was base paired to G268. This segment included the sheared-type A260-G273 and U262-G271 wobble pairs. The two remaining resolved imino resonances were then tentatively assigned to G258 and U265.

Assignments of the aromatic and anomeric resonances were confirmed using NOESY, TOCSY, COSY, and natural abundance ¹H-¹³C-HMQC experiments. A sequential H8/H6-H1' NOE pathway was assigned for the length of the molecule with the exception of G258 and U269, where significant deviations from A-type geometry occurred. In the former case, G258 showed only an intranucleotide H8 to H1' NOE, whereas a "sequential" A257H1'-U259H6 NOE could be observed. In addition, a nonstandard A257H8-U259H1' NOE was observed, suggesting a localized strand reversal for A257 that would require G258 to be excluded from the body of the helix. This unusual geometry is characteristic of the SRL motif. A similar localized strand reversal occurs at G268, which excludes U269 from the loop structure. The five H2 resonances,

all occurring in the internal loop region of the molecule, were confirmed on the basis of their characteristic carbon C2 chemical shifts. A260, A274, A275, and A276 all showed unambiguous sequential NOEs to the H1' in the 3' direction. In addition, at 400 ms NOESY mixing time, an intranucleotide H2-H1' NOE was observed for these residues. H2-H2 NOEs were identified between A260 and A274 and between A275 and A276. A cross-strand A260H2-A274H1' was also observed. The remaining resonance, A257H2, was assigned by elimination, as spectral overlap precluded the identification of any inter- or intranucleotide NOEs.

Three-dimensional model of subdomain III_d structure

Motif (i) was built using idealized coordinates (InsightII Biopolymer module, MSI). Motifs (ii) and (iii) were taken from the crystal structure of the SRL (PDB 1QA6; Correll et al., 1998). Motif (iv) was extracted from the NMR structure of the internal loop of the rev response element (PDB 1ETG; Battiste et al., 1996). The U-turn motif (vi) was extracted from the GTPase RNA crystal structure (PDB 1QA6; Conn et al., 1999). Motif (v) was generated by hand using idealized base planarity and hydrogen bonding distances and angles. A sheared G-A base pair (motif (ii), nt 256 and 276) was added manually, maintaining acceptable C255O3'-G256P and A276O3'-G277P distances. The A257-A260 and G273-A275 SRL motif [motif (iii)] was then positioned. Idealized A-form coordinates were then used to build the G261-G263 double helix, with some manual adjustment to incorporate the U262-G271 base pair. Motif (iv), representing nt G268 and U269 was then positioned in such a way as to optimize the G268-C270 stacking and allow for suitable orientation of G268 for the positioning of U264. Nucleotide U264, using A-form coordinates, was then positioned to stack on G263 and form a *trans*-wobble pair with G268. The U-turn motif was then positioned to complete the loop sequence between U264 and G268. All the components of the model were ligated using the InsightII biopolymer module and the resulting structure was energy minimized to remove unfavorable bond lengths and angles using Charmm 25.2 with the PARM22 force field. An additional minimization held all residues rigid except 264, 268, and 269, and included NOE-based distance restraints involving protons on residues 264, 268, and 270 as referred to in Figure 3. A final minimization released the NOE restraint and resulted in negligible adjustments to the positions of residues 264 and 268.

NOTE ADDED IN PROOFS

The coordinates have been deposited in the Protein Data Bank (1fqz).

ACKNOWLEDGMENTS

We would like to thank Prof. R. Elliott for the kind gift of plasmid pTZ18:5442-16-1 and Dr. Jonathan Karn for helpful discussions.

Received May 10, 2000; returned for revision June 6, 2000; revised manuscript received June 30, 2000

REFERENCES

- Allain F, Varani G. 1995. Structure of the P1 helix from group I self-splicing introns. *J Mol Biol* 250:333–353.
- Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA. 1999. Placement of protein and RNA structures into a 5 angstrom-resolution map of the 50S ribosomal subunit. *Nature* 400:841–847.
- Batey RT, Rambo RP, Doudna JA. 1999. Tertiary motifs in RNA structure and folding. *Angewandte Chemie Int Ed* 38:2326–2343.
- Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. 1996. α -Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* 273:1547–1551.
- Branch AD, Bebebfeld BJ, Robertson HD. 1985. Ultraviolet light-induced crosslinking reveals a unique region of local tertiary structure in potato spindle tuber viroid and HeLa 5S RNA. *Proc Natl Acad Sci USA* 82:6590–6594.
- Brown EA, Zhang H, Ping LH, Lemon SM. 1992. Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* 20:5041–5045.
- Bukh J, Purcell RH, Miller RH. 1992. Sequence analysis of the 5' noncoding region of the hepatitis C virus. *Proc Natl Acad Sci USA* 89:4942–4946.
- Cate JH, Yusupov MM, Yusupova GZ, Earnest TN, Noller HF. 1999. X-ray crystal structures of 70S ribosome functional complexes. *Science* 285:2095–2104.
- Cheng J-W, Chou S-H, Reid BR. 1992. Base pairing geometry in GA mismatches depends entirely on the neighboring sequence. *J Mol Biol* 228:1037–1041.
- Cohen J. 1999. The Scientific challenge of hepatitis C. *Science* 285:26–30.
- Collier AJ, Tang S, Elliot RM. 1998. Translation efficiencies of the 5' untranslated region from representatives of the six major genotypes of hepatitis C virus using a novel bicistronic reporter assay system. *J Gen Virol* 79:2359–2366.
- Conn GL, Draper DE, Lattman EE, Gittis AG. 1999. Crystal structure of a conserved ribosomal protein-RNA complex. *Science* 284:1171–1174.
- Correll CC, Munishkin A, Chan Y-L, Ren Z, Wool IG, Steitz TA. 1998. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci USA* 95:13436–13441.
- Correll CC, Wool IG, Munishkin A. 1999. The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: The structure at 1.11 angstroms resolution. *J Mol Biol* 292:275–287.
- Dallas A, Moore PB. 1997. The loopE-loopD region of *Escherichia coli* 5S rRNA: The solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure* 5:1639–1653.
- Fountain MA, Serra MJ, Krugh TR, Turner DH. 1996. Structural features of a six-nucleotide RNA hairpin loop found in ribosomal RNA. *Biochemistry* 35:6539–6548.
- Honda M, Beard MR, Ping L-H, Lemon SM. 1999. 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.
- Houghton M, Selby M, Weiner A, Choo Q-L. 1994. Hepatitis C virus: Structure, protein products and processing of the polyprotein precursor. In: Liekola J, Lundsgaard-Hansen P, eds. *Current studies in hematology and blood transfusion*. Basel: Karger. pp 1–11.
- Jackson RJ, Howell MT, Kaminski A. 1990. The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem Sci* 15:477–483.
- Jang L, Patel DJ. 1998. Solution structure of the tobramycin-RNA aptamer complex. *Nat Struct Biol* 5:769–774.
- Kieft JS, Zhou K, Jubin R, Murray MG, Lau JYN, Doudna JA. 1999. The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *J Mol Biol* 292:513–529.
- Leontis NB, Westhof E. 1998a. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol* 283:571–583.
- Leontis NB, Westhof E. 1998b. Conserved geometrical base pairing patterns in RNA. *Q Rev Biophys* 31:399–455.
- Liu M, Mao X-a, Ye C, Huang H, Nicholson JK, Lindon JC. 1998. Improved watergate pulse sequences for solvent suppression in NMR spectroscopy. *J Mag Res* 132:125–129.
- Lu M, Steitz TA. 2000. Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-angstroms resolution. *Proc Natl Acad Sci USA* 97:2023–2028.
- Milligan JF, Uhlenbeck OC. 1988. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* 180:51–62.
- Moore PB. 1999. Structural motifs in RNA. *Annu Rev Biochem* 68:287–300.
- Ramos A, Gubser CC, Varani G. 1997. Recent solution structures of RNA and its complexes with drugs, peptides and proteins. *Curr Opin Struct Biol* 7:317–323.
- Reynolds JE, Kaminski A, Kettinen A, Carroll AR, Clarke BE, Rowlands DJ, Jackson RJ. 1996. Internal initiation of translation of hepatitis C virus RNA: The ribosome entry site is the authentic initiation codon. *RNA* 2:867–878.
- Rijnbrand RCA, Abbink TEM, Hassnoot PCJ, Spann WJM, Bredenbeek PJ. 1996. The influence of AUG codons in the hepatitis C virus 5' nontranslated region on translation and mapping of the translation initiation window. *Virology* 226:47–56.
- Saito I, Miyamura T, Ohbayashi A, Harada H, Katayama T, Kikuchi S, Watanabe T, Koi S, Onji M, Otha Y, Choo Q-L, Houghton M, Kuo G. 1990. Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci USA* 87:6547–6549.
- Stoldt M, Woehnert J, Ohlenschlaeger O, Goerlach M, Brown LR. 1999. The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition. *EMBO J* 18:6508–6521.
- Szewczak AA, Moore PB. 1995. The sarcin/ricin loop, a modular RNA. *J Mol Biol* 247:81–98.
- Varani G, Aboul-ela F, Allain FH-T. 1996. NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc* 29:51–127.
- Westhof E, Fritsch V. 2000. RNA folding: Beyond Watson-Crick pairs. *Structure* 8:R55–R65.
- Wimberly B, Varani G, Tinoco IJ. 1993. The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* 32:1078–1087.