Structure Determination of Large Biological RNAs

By Peter J. Lukavsky and Joseph D. Puglisi

Abstract

Complex RNA structures regulate many biological processes but are often too large for structure determination by nuclear magnetic resonance (NMR) methods. We determined the solution structure of domain II of the hepatitis C viral internal ribosome entry site (HCV IRES), a 25-kDa RNA, using a novel NMR approach. Conventional short-range, distance, and torsion angle NMR restraints were combined with long-range, angular restraints derived from residual dipolar couplings (RDCs) to improve both the local and global precision of the structure. This powerful approach should be generally applicable to the NMR structure determination of large, modular RNAs.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy has become a powerful tool for high-resolution structure determination of RNA oligonucleotides up to 15 kDa (Lukavsky and Puglisi, 2001; Lynch et al., 2000; Varani and Timoco, 1991; Varani et al., 1996). As with proteins, RNA NMR structure determination requires uniform isotopic labeling with $^{13}$C and $^{15}$N to achieve complete resonance assignment using multidimensional double- and triple-resonance NMR experiments. Unambiguous resonance assignments form the basis for extraction of structural information from nuclear Overhauser effect spectroscopy (NOESY)—type experiments that yield local distance information between protons separated by less than 5 Å and from NMR experiments that measure torsion angles along the backbone. A maximum achievable number of local restraints is then used in a simulated annealing protocol followed by restrained molecular dynamics (MD) calculations to generate ensembles of structures that satisfy the restraints.

The precision of the structural ensemble depends on both the number and “quality” of the restraints. Long-range NOEs, between residues that are far apart in sequence or link regions of secondary structure, add more to the global structural precision compared to intraresidual NOEs (Allain and Varani, 1997). In general, global precision is more difficult to achieve for RNA molecules, since they often form extended structures, which yield...
only a limited number of long-range restraints compared to globular, compactly folded proteins. The number of restraints for RNAs is also often smaller than in proteins of similar molecular weight. Whereas proteins are composed of 20 different amino acids with an average molecular weight of 130 Da, RNA is composed of only four different nucleosides with an average molecular weight of 340 Da. The molecular weight difference per residue results in a lower density of protons/Dalton for RNA (1/3) compared to proteins, and therefore less structural restraints per residue compared to proteins. The difference of four RNA residues versus 20 protein residues is also reflected in less favorable chemical shift dispersion for all RNA nuclei. This makes unambiguous resonance assignments more difficult in RNA compared to proteins, where most NMR experiments gain resolution by the favorable backbone amide nitrogen chemical shift dispersion of about 30 parts per million (ppm). NMR experiments that make use of the 100% natural abundance of $^{31}$P in an RNA backbone suffer from severe $^{31}$P spectral overlap, small backbone two- and three-bond couplings (3–10 Hz) (Marino et al., 1999), and unfavorable chemical shift anisotropy parameters of the $^{31}$P nuclei, which result in short transverse relaxation times. The severe resonance overlap of ribose nuclei and small heteronuclear couplings makes it more difficult to extract backbone torsion angle restraints from NMR experiments for RNAs.

These NMR spectroscopic shortcomings of RNA molecules become even more severe for larger RNAs. Most biological RNAs are far larger than 15 kDa, which is the size at which the methods described above work effectively. RNA molecules often form more extended structures than proteins of similar molecular weight, resulting in slower overall tumbling times and therefore shorter transverse relaxation times. The concomitant increase in linewidth exacerbates the severity of spectral overlap for all RNA nuclei and eventually makes unambiguous assignments impossible. Many biologically interesting RNA molecules have therefore not been accessible to a high-resolution NMR structure determination. Instead, large biological RNAs have been reduced to collections of smaller, thermodynamically stable subdomains, such as helices and loops, taken out of their larger structural context. In favorable cases (Allain et al., 1996, 2000; Battiste et al., 1996; Fourmy et al., 1996), these subdomains maintained the conformation adopted in the larger RNA and helped to overcome the size and resonance overlap problem. High-resolution structures obtained from subdomains can be used to model the larger RNA molecule (Butcher et al., 1999; Cai and Tinoco, 1996). Without additional structural information from the larger RNA molecule, which would define the orientation of subdomains within their larger structural context, defining the structure of the large RNA is fraught with problems. Lack of interdomain NOE
restraints (Skrynnikov et al., 2000) in multidomain proteins causes a similar problem of domain orientation (Skrynnikov et al., 2000).

The application of residual dipolar couplings (RDC) to biomolecular NMR supplements local torsion and NOE restraints. RDCs yield orientational restraints that improve the global precision of NMR structures (Bax et al., 2001). Application of RDC-derived restraints to the structure determination of small protein and RNA molecules significantly improves both global and local precision of the structural ensembles. In addition, global structures of multidomain proteins can also be defined (Skrynnikov et al., 2000). Similarly, RDC-derived restraints should also benefit the structure determination of large RNAs, where NMR structures of subdomains can be solved to high resolution, but no structural information is available to define their relative orientation in the context of the larger RNA molecule. Here, we describe an NMR approach that uses RDC-derived angular restraints to improve local structures in RNA and to define the overall shape of the RNA molecule. We present the application of this method to the high-resolution structure determination of the hepatitis C viral (HCV) internal ribosome entry site (IRES) domain II RNA, a 25-kDa RNA (Lukavsky et al., 2003).

Design and Validation of RNA Oligonucleotides for Structural Studies

RNA oligonucleotides for structural studies are designed based on a correctly predicted RNA secondary structure. Although the initial secondary structure of the HCV IRES RNA had been determined carefully using a combination of comparative sequence analysis of related pestiviral 5′-untranslated regions (5′-UTR), thermodynamic modeling, and enzymatic footprinting (Brown et al., 1992), the secondary structure of its domain II had to be revised several times (Honda et al., 1999; Zhao and Wimmer, 2001) (Fig. 1). NMR spectroscopy is a powerful tool to study RNA secondary structure, since the imino proton “fingerprint” region of homonuclear two-dimensional (2D) NOESY spectra between 10 and 15 ppm provides information not only about the type of base pair formed, but also their sequential neighbors (Heus and Pardin, 1991). NMR spectroscopy on an oligonucleotide comprising the entire HCV IRES RNA domain II allowed us to confirm the correct secondary structure (Fig. 1B). Figure 1C shows the region of imino–imino NOE cross-peaks of a 2D S-NOESY spectrum (Smallcombe, 1993), which helped to establish the secondary structure of domain II. A strong imino–imino cross-peak arising from two base-paired uracil residues (U64 and U103) and an NOE to an adjacent G102–C65 Watson–Crick base pair as well as an A-form-like imino–ribose NOE
FIG. 1. Secondary structure of HCV IRES RNA and HCV IRES domain II RNAs. (A) Schematic representation of the secondary structure of HCV IRES RNA and its domain organization (Brown et al., 1992). (B) RNA NMR constructs of domain II and subdomains IIa and IIb used for NMR structure determination (Lukavsky et al., 2003). Numbering according
pattern (Heus and Pardi, 1991) (not shown) indicated the formation of a continuous helix below domain IIb rather than the predicted formation of a pyrimidine-rich internal loop or a three-way junction (Fig. 1C). In addition, the G107–U61 base pair showed two NOEs to adjacent G–C Watson–Crick base pairs, which was consistent only with the latest predicted secondary structure shown in Fig. 1B (Zhao and Wimmer, 2001).

Based on these initial NMR studies of domain II, smaller RNA oligonucleotides that correspond to subdomains could be correctly designed. RNA oligonucleotides IIb [34 nucleotides (nt)] and IIa (55 nt) comprising nt 69–98 and nt 45–69 and 98–117 of domain II, respectively, were designed for high-resolution NMR structure determination. Both domain IIb and IIa also contained two additional G–C base pairs analogous to domain II (Fig. 1B), and the apical end of domain IIa was capped by an additional C–G base pair and a UUCG-tetraloop to aid resonance assignments and to serve as a nucleation site for proper folding (Cheong et al., 1990). The addition of stabilizing G–C pairs and capping of helices by tetraloops are standard approaches to RNA oligonucleotide design and provide a stable context for secondary-structure formation.

Chemical shifts were used to confirm that tertiary structures within domain II are accurately represented by the isolated, smaller subdomains IIa and IIb. Chemical shifts are a sensitive measure of the local chemical environment of a nucleus especially in RNA (Cromsigt et al., 2001; Furtig et al., 2004), and the chemical shifts of aromatic base protons are mainly affected by the shielding effects generated by the ring currents of the 5′-neighboring base (Cromsigt et al., 2001). For domain II, aromatic 1H, 13C, and 15N chemical shifts (II versus IIa or IIb) were compared. Similar chemical shifts in the different RNA oligonucleotides were reflected in low root-mean-square deviations (RMSD) for chemical shifts of aromatic C–H groups [0.099 (13C) and 0.074 (1H), respectively] and N–H groups [0.055 (15N) and 0.028 (1H), respectively]. Larger RMSDs were observed only for the C69–G98 Watson–Crick base pair, since the 5′-neighboring nucleotides differed in all three constructs, but close inspection of NOE patterns and cross-peak intensities in S-NOESY spectra indicated that the same structure was formed (data not shown). These data demonstrate that subdomains IIa and IIb adopt the same conformation within the context of the entire domain II (Lukavsky et al., 2003).
Use of Segmental Isotope Labeling for NMR Studies of Large RNAs

Segmental isotope labeling of RNA reduces the complexity of NMR spectra and therefore allows more detailed NMR studies of subdomains of very large RNAs (Kim et al., 2002; Xu et al., 1996). Segmental labeling can also be used to confirm the validity of NMR data acquired on model oligonucleotides. Isotopic labeling of domain II in the context of the entire 100-kDa HCV IRES RNA was used to address the question of whether domain II forms an individually folded domain within the HCV IRES.

The most cost-efficient method of choice for the preparation of a segmentally labeled RNA oligonucleotide is T4 RNA ligase-catalyzed joining of a 3′ “donor” RNA having a 5′-terminal monophosphate and a 5′ “acceptor” RNA terminating in a 3′-hydroxyl. The product of such a ligation has a standard 3′,5′-phosphodiester linkage (Romaniuk and Uhlenbeck, 1983). The efficiency of ligation depends on the structure of the 5′ and 3′ ends, and it is efficient only for RNAs that have a single-stranded region as the site of ligation. To avoid intramolecular ligation and the formation of by-products, both sides of the 5′-fragment should be dephosphorylated and both sides of the 3′-fragment should be phosphorylated. This can be easily achieved using hammerhead ribozymes to engineer the RNA 5′ and 3′ ends of the donor and acceptor in a way such that only the desired product can be formed (Kim et al., 2002).

Based on the secondary structure of domain II (Fig. 1B), which was confirmed by NMR methods, only two single-stranded regions emerged as potential candidates for optimum ligation conditions, namely the single-stranded region in domain IIa (A53–A57) and the apical hairpin loop (U80–U86). Since both sites would allow only 20% or 50%, respectively, of domain II to be 15N labeled in the context of the HCV IRES, we chose nucleotides C104 and G105 as the point of ligation. While this region is base paired in the context of domain II, individually transcribed parts comprising nt 40–104 and nt 105–354, respectively, should be single stranded at their ligation site. Both the 5′ and 3′-RNA fragments were transcribed in vitro using transcription with 15N-labeled nucleotide triphosphates for the 5′-fragment and unlabeled nucleotide triphosphates for the 3′-fragment. The purified RNA fragments were then ligated using T4 RNA ligase at a maximum efficiency of 50%, yielding a 0.1 mM NMR sample of segmentally 15N-labeled HCV IRES RNA (Kim et al., 2002). Chemical shifts of imino 1H and 15N resonances observed for domain II alone and in the context of the 100-kDa HCV IRES were almost identical, confirming that domain II forms an independently folded subdomain in the intact IRES (Fig. 2).
Measurement of RDCs

In proteins, RDCs are usually obtained for amide backbone one-bond $^1$H–$^{15}$N couplings and side-chain one-bond $^1$H–$^{13}$C couplings (Tjandra et al., 1997). This set can be supplemented by additional one-bond $^{15}$N–$^{13}$C, $^{13}$C–$^{13}$C, two-bond $^{13}$C–$^1$H, and three-bond $^{13}$C–$^{15}$N RDCs to provide a high density of RDC-derived restraints (Skrynnikov et al., 2000). In RNA, on the other hand, one-bond $^1$H–$^{15}$N couplings from imino protons usually yield only a limited number of restraints and backbone three-bond $^1$H–$^{31}$P and four-bond $^{13}$C–$^{31}$P couplings are small and cannot be accurately measured. Orientational restraints for RNA structure determination are therefore commonly derived from RDCs measured from one-bond $^1$H–$^{15}$N interactions supplemented by a few derived from $^1$H–$^{15}$N interactions (Hansen et al., 1998). Bacteriophage Pf1 is the ideal liquid-crystalline medium for RDC measurements of RNA, because negatively charged RNA oligonucleotides are aligned by steric interactions with the negatively charged phage particles rather than by direct binding to the phage, which would increase the overall tumbling time, and, correspondingly, the
linewidths of RNA resonances (Hansen et al., 1998). One-bond coupling constants in RNA are usually measured in isotropic and aligned media (upon addition of up to 25 mg/ml of Pf1 phage (Hansen et al., 1998)), and the RDC value can then be extracted from the difference between the value (in Hertz) of the coupling constant obtained in aligned and isotropic media.

Several NMR methods have been developed for the accurate measurement of one-bond couplings. Initially, one-bond $^1$H–$^{15}$N and $^1$H–$^{13}$C couplings were measured by simply recording 2D heteronuclear single-quantum correlation (HSQC) spectra without decoupling during the $t_1$ evolution period. Since this method doubles the number of resonances in already crowded RNA 2D HSQC spectra, only a limited number of coupling constants could be measured. More recently, one-bond couplings are measured using either IPAP-HSQC (in-phase and antiphase) experiments, which yield either the upfield or downfield component of the $F_1$ $^{15}$N–$^1$H doublet (Ottiger et al., 1998) or sets of $J_{CH}$-modulated 2D HSQC spectra (Tjandra and Bax, 1997). Both methods alleviate the problem of spectral overlap, and the latter has been applied to the study of the Sarcin–Ricin loop (Warren and Moore, 2001) and the hammerhead ribozyme RNA (Bondensgaard et al., 2002). For larger RNAs, like the HCV IRES domain II, 2D IPAP HSQC and 2D $t_1$-coupled HSQC cannot be applied, since the upfield component of the $F_1$ $^{15}$N–$^1$H or $^{13}$C–$^1$H doublet is too broad to allow accurate measurement of the coupling constant. Figure 3 shows an adenine C2–H2 cross-peak from a spectrum acquired with a 2D $t_1$- and $t_2$-coupled $^1$H–$^{13}$C HSQC experiment on the 77-nt RNA comprising domain II (see Fig. 1B). Only the downfield components in the $^{13}$C dimension yield sufficiently sharp resonances, while the upfield components of the multiplet exhibit severe broadening (Fig. 3). A similar situation is encountered for the $^1$H–$^{15}$N multiplet of imino resonances (data not shown). In addition, $^1$H–$^{13}$C correlation experiments require a constant time (CT) frequency-editing period to eliminate one-bond $^{13}$C–$^{13}$C couplings during $t_1$ evolution (17 ms for base and 25 ms for sugar carbons). In large RNAs, where short transverse relaxation times are encountered, these long CT periods can lead to significant losses in signal intensity and thereby make measurement lengthy or even impossible.

These problems can be overcome by using TROSY-based methods, which significantly reduce signal loss and allow selection of the sharper, downfield $^{13}$C or $^{15}$N components of the $^1$H–$^{13}$C and $^1$H–$^{15}$N multiplets. For the one-bond $^1$H–$^{15}$N coupling measurements of HCV IRES domains II, IIa, and IIb, the spin-state selective, gradient- and sensitivity-enhanced 2D $^1$H–$^{15}$N TROSY by Weigelt (1998) or the original 2D WATERGATE $^1$H–$^{15}$N TROSY (Pervushin et al., 1997) were therefore used. One-bond
$^1$H–$^{13}$C couplings were measured using a $^{13}$C version of the Weigelt-TROSY (Weigelt, 1998), which included a CT evolution period (see Fig. 3). The latter $^1$H–$^{13}$C TROSY version was preferred over a previously published version (Brutscher et al., 1998) due to its excellent H$_2$O suppression. All measurements were performed in the same buffer (10 mM sodium phosphate buffer at pH 6.40, 100 mM sodium chloride, and 5 mM magnesium chloride with 4% D$_2$O) and repeated at least twice in the absence or presence of 8 (II) or 10 (IIa and IIb) mg/ml Pf1 phage at 800 MHz field strength. To ensure complete alignment at such high salt versus low Pf1 phage concentrations (Zweckstetter and Bax, 2001), the samples were prealigned on top of the 800-MHz magnet overnight. The spinner with the sample tube was simply inserted into a piece of cardboard with a hole in the middle to prevent the sample from falling into the magnet and was placed on top of the sample inlet tube. The uncertainty in the measurement was estimated based on the quotient of linewidth and signal-to-noise ratio.
(Bax et al., 2001) to be 2.5 Hz for $^{13}\text{C}^{1}\text{H}$ and 1.25 Hz for $^{15}\text{N}^{1}\text{H}$ couplings.

The RDC values were calculated as the difference of the measured coupling (Hertz) in aligned and isotropic media (for $^{1}\text{H}^{15}\text{N}$ RDCs, the negative gyromagnetic ratio has to be taken into account). RDCs, whose standard deviation from the average of at least two individual RDC values exceeded the uncertainty of the measurement (see above), were not used for subsequent structure calculations. Large errors in some RDCs originated from partial overlap or water suppression artifacts near the residual water line. In addition, RDCs from three dynamic uracil residues (U48, U56, and U106) were also eliminated. A final set of 136 (IIa), 105 (IIb), or 60 (II) RDCs was used for the structure refinement of subdomains (IIa and IIb) and domain II.

Use of RDCs in Structure Calculation

To determine the structure of domain II by NMR, high-resolution restraints from the subdomains IIa and IIb were combined with RDC data of the subdomains and the full domain II. Assignment procedures for domains IIa and IIb followed previously published protocols using RNA-Pack NMR pulse sequences on VARIAN NMR spectrometers (Lukavsky and Puglisi, 2001). Complete resonance assignments of domains IIa and IIb yielded 1146 NOE and 419 scalar coupling restraints for domain IIa, 831 NOE and 253 scalar coupling restraints for domain IIb, and a total of 1744 NOE and 523 scalar coupling restraints for domain II, respectively. Random starting structures were generated and subjected to a simulated annealing protocol excluding RDC restraints, followed by restrained molecular dynamics simulations (Lukavsky et al., 2003). The atomic RMSDs for the 20 best IIb and 29 best IIa structures were 2.43 and 4.91 Å, respectively; however, the entire domain II was poorly defined with an RMSD of 7.48 Å for the best 20 structures, since no RDC-derived restraints were used to define the global conformation of domain II.

Calculating structures without RDC-derived restraints was necessary for two reasons. First, NOE restraints tend to favor convergence through cooperative contributions to the folding energy during simulated annealing, thereby creating a funnel-like energy landscape, whereas angular restraints tend to compete with one another and, thus, are not usable during the initial in-silico folding of the structures (Bax et al., 2001). Second, initial structures calculated with RDCs can be used to determine the magnitude of the axial ($D_a$) and rhombic (R) components of the alignment tensor to extract orientational restraints for the subsequent NMR structure refinement from measured RDCs. In addition to one model-free approach,
several model-based prediction methods have been proposed to determine an initial alignment tensor for the refinement of NMR structures (see below), and initial structures calculated without RDCs can be used for this purpose.

Prediction of the Alignment Tensor

Several methods for prediction of the magnitude of the $D_a$ and $R$ components of the alignment tensor have been tested on RNA molecules (McCallum and Pardi, 2003; Warren and Moore, 2001). These include laborious grid search approaches (Clore et al., 1998a), prediction of the alignment tensor from the three-dimensional shape of the biomolecule (Zweckstetter and Bax, 2000), structure-independent prediction from the histogram of the RDCs (Clore et al., 1998b), and singular value decomposition (SVD), which determines $D_a$ and $R$ values consistent with a given RDC data set and structure using a fitting procedure (Losonczi et al., 1999).

Analysis of the histogram of the RDC data was used as a model-free approach for the determination of an initial alignment tensor for NMR refinement of the subdomains IIa and IIb as well as domain II (Table I). All histograms displayed the characteristic powder pattern with RDC values ranging from $-44.5$ to $+37.0$ for II, $-23.8$ to $+22.6$ for IIa, and $-23.8$ to $+19$ for IIb, respectively. In domain IIb, RDCs from aromatic carbon–proton pairs, which are almost perpendicular to the helix axis, clustered in the right-hand side indicative of a roughly helical shape of domain IIb. For domains II and IIa, on the other hand, negative RDCs for aromatic carbon–proton pairs in the lower stem indicated a bend within domain IIa. Analysis of the structural ensembles of domains IIa, IIb, and II calculated without RDCs using the SVD method yielded very similar $D_a$ and $R$ values (Table I). Independent fits performed on helical regions of the subdomains IIa and IIb as well as subdomains IIa and IIb in the context of domain II also confirmed that these parts align as a single species (Losonczi and Prestegard, 1998).

Two sets of structure calculations with RDC restraints were performed on each subdomain using the two individually determined alignment tensors for domains IIa and IIb. Implementation of RDC restraints in the CNS package (Brunger et al., 1998) requires definition of a tetraatomic pseudomolecule OXYZ, which represents the alignment tensor (Bax et al., 2001). The pseudomolecule was added to the simulated annealing structures calculated without RDCs at a fixed position in space away from the RNA and the refinement procedure including RDC restraints was performed as follows: (1) 500 steps of restrained energy minimization; (2) restrained molecular dynamics (rMD) at 2000 K, while increasing the torsion angle.
force constant; (3) rMD at 2000 K while increasing the RDC force constant keeping the torsion angle force constant low and (4) rMD at 2000 K while keeping the RDC force constant at 0.25 kcal/Hz and increasing the torsion angle force constant; (5) rMD while cooling to 300 K; and finally (6) 5000 cycles of energy minimization, which included a Lennard–Jones potential, but no electrostatic terms. The ensemble of structures that had the lowest average total energy and final RDC restraint violation energy was then used to determine a new alignment tensor by SVD fitting of observed RDCs to each ensemble member. The average values for $D_a$ and $R$ obtained from the fitting were then used in another round of structure calculations to improve the alignment tensor iteratively until the average value for the rhombic component was unchanged and met acceptance criteria ($\Delta R < \pm 0.015$) (Fig. 4), yielding final alignment tensors for Ia and Ib (Table I). Figure 4 illustrates the iterative improvement of the alignment tensor for domain Ia. The initial $D_a$ and $R$ values determined by the SVD method yielded the ensemble with the lowest final RDC restraint violation and total energy and even an unrefined alignment

<table>
<thead>
<tr>
<th>Method</th>
<th>$D_a$(Hz)</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular value decomposition (SVD) method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain Ia</td>
<td>−9.44</td>
<td>0.40</td>
</tr>
<tr>
<td>Upper helix of Ia (58–69, 98–110)</td>
<td>−11.26</td>
<td>0.56</td>
</tr>
<tr>
<td>Lower helix of Ia (43–52, 111–119)</td>
<td>−8.92</td>
<td>0.62</td>
</tr>
<tr>
<td>Average</td>
<td>−9.88 ± 1.23</td>
<td>0.53 ± 0.11</td>
</tr>
<tr>
<td>Domain Ib</td>
<td>−11.97</td>
<td>0.15</td>
</tr>
<tr>
<td>No hairpin loop of Ib (67–79, 87–100)</td>
<td>−12.30</td>
<td>0.19</td>
</tr>
<tr>
<td>Average</td>
<td>−12.13 ± 0.23</td>
<td>0.17 ± 0.03</td>
</tr>
<tr>
<td>Domain II</td>
<td>−19.86</td>
<td>0.44</td>
</tr>
<tr>
<td>Subdomain Ia of II (43–69, 98–119)</td>
<td>−23.07</td>
<td>0.42</td>
</tr>
<tr>
<td>Subdomain Ib of II (70–97)</td>
<td>−21.06</td>
<td>0.54</td>
</tr>
<tr>
<td>Average</td>
<td>−21.52 ± 1.38</td>
<td>0.43 ± 0.09</td>
</tr>
<tr>
<td>Histogram of the observed RDCs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain Ia</td>
<td>−11.90</td>
<td>0.60</td>
</tr>
<tr>
<td>Domain Ib</td>
<td>−11.90</td>
<td>0.40</td>
</tr>
<tr>
<td>Domain II</td>
<td>−22.25</td>
<td>0.44</td>
</tr>
<tr>
<td>Final tensor after iterative refinement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain Ia</td>
<td>−12.93</td>
<td>0.45</td>
</tr>
<tr>
<td>Domain Ib</td>
<td>−15.28</td>
<td>0.17</td>
</tr>
<tr>
<td>Domain II</td>
<td>−26.40</td>
<td>0.29</td>
</tr>
</tbody>
</table>

$^a$See text for details.
tensor already significantly improved the definition of the overall structure of domain IIa (RMSD of 4.91 Å without RDCs versus 2.98 Å after the first round of calculations). The final stage was reached after five rounds of calculations and the final ensemble of domain IIa structures had an RMSD of 2.34 Å. The structure of domain IIb was refined in the same manner improving the RMSD from 2.43 to 1.35 Å.

Refinement of subdomains IIa and IIb in the context of the entire domain II was then performed using final alignment tensors and RDCs from the domain IIa and IIb oligonucleotides (see Fig. 1B). The protocol for structural calculations of domain II was performed in the same two stages as described for domains IIa and IIb and used a total of 1744 distance restraints, including 21 NN hydrogen bond distance restraints, 261 RDC restraints (from domains IIa, IIb, and II), and 523 dihedral restraints. For each alignment tensor (final alignment tensors of IIa and IIb, and the initial alignment tensor of II), a separate pseudomolecule
OXYZ was added to the accepted simulated annealing structures of domain II at a fixed position in space away from the RNA. The refinement procedure including all RDC restraints was then performed very similarly; the only exception was that RDCs were introduced stepwise, first for IIb and II, then for IIa and II, and finally for all RDCs to improve convergence. After the first round of calculations, the individual ensembles were analyzed to improve iteratively the alignment tensor for domain II RDCs without changing the alignment tensors for domains IIa and IIb until the average value for the rhombic component was unchanged and met the same acceptance criteria ($\Delta R < \pm 0.015$) as for the individual subdomains IIa and IIb. The final ensemble of domain II structures (Fig. 5B) after five rounds of calculations showed both locally improved subdomains (1.15 Å for IIb and 1.62 Å for IIa) and a well-defined global shape of domain II (2.18 Å) compared to the ensembles obtained without RDCs (Fig. 5A) (RMSDs of 2.43 Å for IIb, 4.38 Å for IIa, and 7.48 Å for II).

Fig. 5. Final ensembles of domain II structures calculated with different sets of RDCs and corresponding local superpositions of subdomains IIa and IIb. (A) Final ensemble of structures calculated without RDCs. (B) Final ensemble of structures calculated with RDCs from domains IIa, IIb, and II. (C) Final ensemble of structures calculated only with RDCs from domains IIa and IIb. (D) Final ensemble of structures calculated only with RDCs from domain II.
improvement could be achieved only with the combination of RDCs from the subdomains with the global RDCs from domain II. Omitting the 60 RDCs from domain II yielded well-defined local domains (1.43 Å for IIb and 2.63 Å for IIa), while the overall shape of domain II improved only slightly (5.79 Å) as shown in Fig. 5C. Similarly, using only 60 RDCs from domain II also gave only a slight improvement of the overall definition (5.74 Å), since subdomains were less well-defined without RDCs (2.31 Å for IIb and 3.02 Å for IIa) as shown in Fig. 5D.

A limited number of RDC restraints was sufficient to define the global fold of domain II, consistent with prior data (Mollova et al., 2000). For domain II, only 60 additional RDCs, which is less than one RDC restraint per residue, were sufficient to define the global shape of domain II. Mollova et al. (2000) demonstrated that only 27 RDCs were sufficient to determine the global structure of tRNA, but only if the local structures of the helical stems is known. In the case of a de novo RNA NMR structure determination, where local and global structures are unknown, such well-defined subdomain structures can be obtained using an RDC-based refinement procedure. Only smaller subdomains of the larger RNA allow extraction of a sufficiently large number of RDCs to obtain well-defined structures, since NMR spectra are less crowded compared to larger RNAs. A small number of global RDCs is then sufficient to define orientation of these well-defined substructures and thereby the overall conformation of a large RNA.

Conclusions

We have outlined a powerful and general approach to structure determination of modular large RNAs using NMR spectroscopy. Dissection of a large RNA into subdomains allows extraction of the maximum number of conventional NOE- and scalar coupling-derived restraints, as well as a much larger number of RDCs. This leads to determination of high-quality, well-defined subdomain structures, which provide the basis for a global refinement of the larger RNA using RDCs. In the future, this approach should allow NMR structure determination of even larger RNAs, where several smaller RNA subdomains could be oriented within the larger RNA using segmental labeling techniques to determine a sufficient number of global RDC restraints. These approaches will further increase the power of NMR to probe the biological functions of RNA.

References


