Journal of Biomolecular NMR, 6 (1995) 375–389
ESCOM



J-Bio NMR 311

# Assignment and modeling of the Rev Response Element RNA bound to a Rev peptide using $^{13}$C-heteronuclear NMR

John L. Battiste[a], Ruoying Tan[b], Alan D. Frankel[b] and James R. Williamson[a,*]

[a]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*
[b]*Department of Biochemistry and Biophysics and Gladstone Institute of Virology and Immunology, University of California at San Francisco, San Francisco, CA 94143, U.S.A.*

Received 6 July 1995
Accepted 30 August 1995

*Keywords:* RNA structure; Rev Response Element; RNA–protein interactions

## Summary

The Rev Response Element (RRE) RNA–Rev protein interaction is important for regulation of gene expression in the human immunodeficiency virus. A model system for this interaction, which includes stem IIB of the RRE RNA and an arginine-rich peptide from the RNA-binding domain of Rev, was studied using multidimensional heteronuclear NMR. Assignment of the RNA when bound to the peptide was obtained from NMR experiments utilizing uniformly and specifically $^{13}$C-labeled RNA. Isotopic filtering experiments on the specifically labeled RNA enabled unambiguous assignment of unusual nonsequential NOE patterns present in the internal loop of the RRE. A three-dimensional model of the RNA in the complex was obtained using restrained molecular dynamics calculations. The internal loop contains two purine–purine base pairs, which are stacked to form one continuous helix flanked by two A-form regions. The formation of a G-G base pair in the internal loop requires an unusual structure of the phosphate backbone. This structural feature is consistent with mutational data as being important for the binding of Rev to the RRE. The G-G base pair may play an important role in opening the normally narrow major groove of A-form RNA to permit binding of the Rev basic domain.

## Introduction

The RRE RNA–Rev protein interaction is essential for replication of HIV (human immunodeficiency virus). Binding of Rev to the RRE is thought to result in inhibition of splicing and/or facilitation of transport of mRNAs from the nucleus to the cytoplasm (Malim and Cullen, 1993; Ruhl et al., 1993; Fischer et al., 1994; Stutz and Rosbash, 1994). Biochemical studies have identified a high-affinity binding site for a single Rev protein in a small stem–loop structure in the RRE, termed stem–loop IIB (Cook et al., 1991; Kjems et al., 1991; Tiley et al., 1992). Further studies showed that a peptide spanning the RNA-binding domain of Rev was sufficient for specific binding to RRE IIB, provided that the peptide was in an α-helical conformation (Kjems et al., 1992; Tan et al., 1993). The simplified minimal complex between the RRE IIB RNA and the Rev peptide is a good model system for recognition of the RRE by the Rev protein and is suitable for structural studies using NMR spectroscopy.

The sequences of the RNA and peptide used in this study are given in Fig. 1, along with biochemical data identifying the nucleotides and amino acids that are important for binding. The peptide used here (suc-$Rev_{34-50}$-AAAAR-am) contains the basic region of Rev ($Rev_{34-50}$) with modifications at both the N- and C-termini. These modifications were introduced to increase the helical content of the peptide, resulting in increased binding specificity (Tan et al., 1993). Earlier NMR studies of RNAs derived from stem IIB of RRE showed that a specific complex could be formed upon binding to suc-$Rev_{34-50}$AAAAR-am (Battiste et al., 1994; Peterson et al., 1994). In these studies, the secondary structure of the bound RNA was characterized, identifying a G-A and a

*To whom correspondence should be addressed.

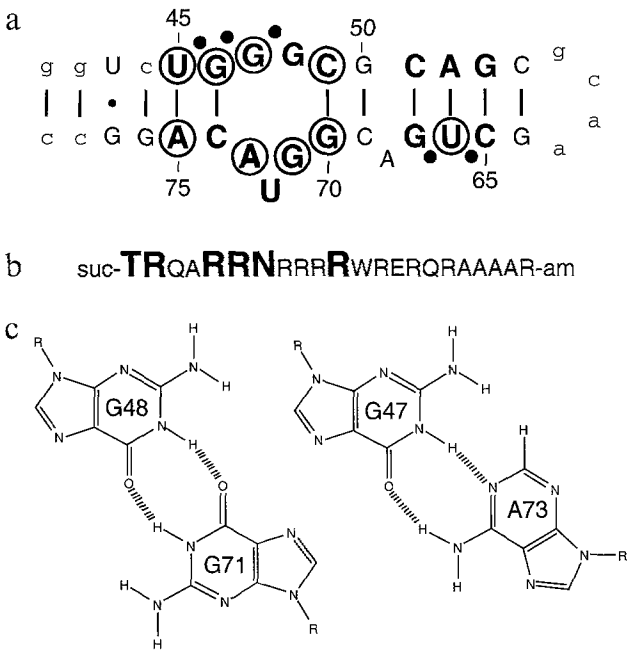


Fig. 1. (a) Sequence and secondary structure of the RRE RNA used for the NMR studies described here. The sequence is derived from stem IIB of RRE, and the numbering is discontinuous after the tetraloop to keep the numbering in the internal loop region the same as that of wild type. Small letters indicate non-wild-type nucleotides. Filled circles represent phosphates that interfere with peptide or protein binding upon chemical modification. Open circles represent bases that interfere with binding upon chemical modification (Kjems et al., 1992). Bold letters represent nucleotides that are conserved in RNAs generated by in vitro selection for Rev binding (Bartel et al., 1991). (b) Sequence of the suc-$Rev_{34-50}$AAAAR-am peptide used for our NMR studies. -suc and -am are succinyl and amide groups attached to the N- and C-terminus, respectively. Bold letters indicate positions where mutation to alanine significantly decreased the specificity of binding to RRE IIB RNA (Tan et al., 1993). (c) Hydrogen-bonding patterns of the two purine–purine base pairs in the internal loop, determined from previous NMR studies (Battiste et al., 1994; Peterson et al., 1994).

G-G base pair thought to be critical for the RRE–Rev interaction. Figure 1c shows the hydrogen-bonding interactions for these two base pairs, determined from NMR studies. Imino protons for many base pairs, including the purine–purine pairs, were only observed in the bound form of the RNA, indicating that the RNA structure may significantly change upon binding (Battiste et al., 1994). Therefore, structural studies to identify important features of the RRE RNA for protein binding cannot be performed on the free form of the RNA.

Obtaining high-resolution structures of large RNA molecules by either X-ray crystallography or NMR has proven difficult; however, recent technical advances have been made in both fields (Doudna et al., 1993; Nikonowicz and Pardi, 1993). For NMR in particular, advances in isotopic labeling methods have extended the size of RNA molecules that can be studied in detail (Batey et al., 1992; Nikonowicz et al., 1992; Nikonowicz and Pardi, 1993). Here we have used multidimensional heteronuclear NMR to determine the structure of the RNA in the minimal RRE RNA–Rev peptide complex. Due to spectral overlap from a large number of arginines, sequence-specific assignments of the peptide have not yet been obtained. However, assignment and structure determination of the bound RNA were achieved through uniform and specific isotopic labeling of the RNA with $^{13}C$. In particular, isotopic filtered NOESY spectra of a specifically $^{13}C$-labeled RNA sample were used to assign an unusual sequential NOE pathway observed in the purine-rich internal loop of the RRE. A three-dimensional model of the RNA was obtained using restrained molecular dynamics with NMR-derived NOE and torsion restraints. The resulting structure contains an unusual feature in the internal loop, involving formation of the G-G base pair.

## Materials and Methods

*Sample preparation*

Three RRE RNA samples were prepared for this study: unlabeled, uniformly $^{13}C$-labeled, and specifically $^{13}C$-labeled at guanosine residues. All RNA was prepared by in vitro transcription from oligonucleotide templates and purified as previously described (Battiste et al., 1994). The final concentrations of the samples in 500 µl of NMR buffer (10 mM sodium phosphate, pH 6.5, 50 mM sodium chloride and 0.1 mM EDTA) were 1.0, 1.5 and 1.2 mM for unlabeled, uniformly $^{13}C$-, and $^{13}C$-guanosine-labeled RRE RNA, respectively. The NTPs for uniformly $^{13}C$-labeled RNA were prepared as described previously (Batey et al., 1992). For the preparation of $^{13}C$-guanosine-labeled RRE RNA, the individual $^{13}C$-labeled NMPs were first separated by HPLC on a 7.8×100 mm HP-PEI anion exchange column (Interaction Chemicals, Inc., Mountain View, CA). The flow rate was 1 ml/min and the nucleotides were monitored by UV absorbance at 280 or 300 nm. Buffer A was 50 mM ammonium formate, pH 3.0, and buffer B was 500 mM ammonium formate, pH 2.5. Isocratic elution with 100% buffer A for 20 min was followed by a linear gradient to 100% buffer B over 20 min. The retention times were 8, 16, 33 and 42 min for CMP, AMP, UMP and GMP, respectively. Up to 60 mg of nucleotides could be injected, while maintaining baseline resolution. Pooled fractions were lyophilized to dryness. GMP was enzymatically converted to GTP following the standard procedure (Batey et al., 1992), except that nucleoside monophosphate kinase was omitted, the concentration of guanylate kinase was increased to 0.05 units/ml, and the catalytic amount of unlabeled ATP was increased to 0.3 mM. Standard T7 transcription was then performed with unlabeled UTP, CTP, ATP, and $^{13}C$-GTP.

The synthesis and purification of the Rev peptide and the formation of a 1:1 complex with each of the RRE RNAs were performed as described previously (Tan et al., 1993; Battiste et al., 1994).

*Proton NMR*

All NMR spectra were recorded at 25 °C on a Varian VXR 500 MHz instrument with a Varian inverse detection probe, and data were transferred to a Silicon Graphics Personal Iris for processing using FELIX software (Biosym, Inc., San Diego, CA). Standard NOESY, $^{31}$P-decoupled DQF-COSY, and TOCSY experiments on the unlabeled complex in 99.996% $D_2O$ were acquired as described before (Battiste et al., 1994). To monitor NOE buildups, four NOESY experiments were acquired with mixing times of 50, 100, 200 and 400 ms. A ROESY experiment was performed to identify potential conformational exchange peaks. The band width of the spin-lock mixing time (100 ms) was 3 kHz, and the flip angle used was 30° to help remove TOCSY artifacts (Kessler et al., 1987). The sweep width in both dimensions of the ROESY experiment was 5000 Hz, with 2048 complex points in the acquisition dimension and 256 $t_1$ increments. Per increment, 48 scans were collected with a relaxation delay of 2 s.

*Heteronuclear NMR*

Table 1 lists the acquisition parameters for the heteronuclear experiments performed on both labeled RRE RNAs complexed with unlabeled Rev peptide. In all experiments $^{13}$C decoupling was achieved with a WALTZ (Shaka et al., 1983) scheme during acquisition and a 180° $^{13}$C pulse in the middle of the $t_1$ proton evolution times. The $^{13}$C transmitter was set to ~80 ppm, between the C4' and C2'/C3' resonances, for all experiments focusing on coherence transfer in the ribose ring. For all NOESY experiments the transmitter was set to ~105 ppm, just downfield of the C5 resonances, to include the aromatic carbon resonances. In addition, for the NOESY experiments the INEPT delay was set using an average of the aromatic and ribose one-bond carbon–proton coupling constants (160 Hz). The relaxation delay was between 1.6 and 2.0 s for all experiments. The total acquisition times were 3.0–3.5 days for 3D experiments and 18–24 h for 2D experiments. Phase cycling for all experiments was performed as described in the references in Table 1. Quadrature detection was achieved using the TPPI-States method (Marion et al., 1989) for all experiments, except the double-half-filtered NOESY, which used the States method (States et al., 1982). In addition, the phase cycling for the double-half-filtered NOESY was performed in an interleaved fashion.

*NMR-derived restraints for molecular modeling*

*Hydrogen bonds* The hydrogen-bonding patterns of the base pairs, including the purine–purine base pairs in the internal loop, were previously determined from analy-

TABLE 1
DATA ACQUISITION PARAMETERS FOR HETERONUCLEAR NMR EXPERIMENTS

| Molecule | Experiment | $\omega_3$[a] | $\omega_2$[a] | $\omega_1$[a] | $J_{CH}$[b] (Hz) | No. of scans | Other | Ref. |
|---|---|---|---|---|---|---|---|---|
| Uniformly $^{13}$C-labeled RRE RNA | 3D NOESY-HMQC | 512 pts 4000 Hz $^1$H | 64 pts 5000 Hz $^{13}$C | 64 pts 4000 Hz $^1$H | 160 | 8 | 250 ms[c] | (Clore et al., 1990) |
| | 3D HCCH-DIPSI | 512 pts 2500 Hz $^1$H | 40 pts 5000 Hz $^{13}$C | 64 pts 2000 Hz $^1$H | 142 | 16 | 25 ms[d] | (Clore et al., 1990) |
| | 3D HCCH-COSY | 512 pts 2500 Hz $^1$H | 32 pts 5000 Hz $^{13}$C | 64 pts 2500 $^1$H | 142 | 16 | | (Clore et al., 1990) |
| | HSQC-CT | – | 1024 pts 5500 Hz $^1$H | 112 pts 5000 Hz $^{13}$C | 142 | 32 | 25 ms[e] | (Santoro and King, 1992) |
| $^{13}$C-guanosine RRE RNA | 3D NOESY-HMQC | 512 pts 5000 Hz $^1$H | 32 pts 6000 Hz $^{13}$C | 128 pts 5000 Hz $^1$H | 160 | 8 | 250 ms[c] | (Clore et al., 1990) |
| | HSQC-CT | – | 1024 pts 5000 Hz $^1$H | 128 pts 6000 Hz $^{13}$C | 142 | 64 | 25 ms[e] | (Santoro and King, 1992) |
| | Double-half-filtered NOESY | – | 1024 pts 5500 Hz $^1$H | 256 pts 5500 Hz $^1$H | 160 | 16 | 300 ms[c] | (Otting and Wüthrich, 1989,1990) |

[a] Listed for each dimension are the number of complex data points acquired, the spectral width, and the nucleus detected.
[b] Given is the value of the one-bond carbon–hydrogen coupling constant used for determination of the delay for transfer of magnetization between proton and carbon. For the NOESY experiments, the value used was an average of the aromatic and ribose coupling constants.
[c] The mixing time of the period for NOE transfer.
[d] The mixing time of the carbon spin-lock pulse for transfer of magnetization through the ribose ring.
[e] Value of the constant-time interval for evolution of the carbon chemical shift in $\omega_1$.

sis of NOESY spectra in $H_2O$ (Battiste et al., 1994). Hydrogen bonds were included as distance restraints between the proton and hydrogen-bond acceptor using upper and lower bounds of 1.70–2.20 Å, which is the range of hydrogen-bond distances observed in nucleotides (Saenger, 1984). Five additional hydrogen bonds, not directly derived from our experiments, were included for the GCAA tetraloop whose structure has been previously published (Heus and Pardi, 1991). All of our NOE and chemical shift data in the tetraloop region are consistent with the published tetraloop structure.

*NOE distance restraints* Distance restraints involving nonexchangeable RNA protons were derived from visual inspection of cross-peak intensities in NOESY experiments in 99.996% $D_2O$ with mixing times of 50, 100 and 200 ms. The following criteria were used for classification of NOEs into three distance bound ranges: (i) strong, 1.8–3.0 Å, medium to strong intensity NOEs at 50 ms; (ii) medium, 2.5–4.0 Å, all additional NOEs present at 100 ms; and (iii) weak, 3.5–5.0 Å, all additional NOEs present at 200 ms. NOEs involving exchangeable proton resonances were classified as strong, medium, or weak based on visual inspection of intensities from a NOESY in $H_2O$, mixing time 250 ms. NOEs resolved only in the NOESY-HMQC spectrum were classified into the same ranges by visual inspection of cross-peak intensities for a 250 ms mixing time.

*Torsion restraints* The sugar pucker was estimated from analysis of the H1'-H2' coupling constants in the $^{31}$P-decoupled DQF-COSY spectrum. Nucleotides with a H1'-H2' coupling constant of > 8 Hz in the COSY spectrum were classified as C2'-endo. Nucleotides with no COSY and TOCSY cross peaks between the H1'-H2' protons (J < 3 Hz) were classified as C3'-endo. Some nucleotides had weak H1'-H2' cross peaks in the TOCSY spectrum, but no COSY cross peaks. These are most likely nucleotides with some mixed population of C2'-/C3'-endo conformations. The ribose puckers for these nucleotides were left unrestrained during molecular dynamics. The sugar puckers for nucleotides G46, A68, G71 and U72 were restrained at a C2'-endo pucker. G48, C56, A57 and G67 were left unrestrained. All other sugar puckers were restrained to be C3'-endo.

The glycosidic torsion angle $\chi$ was determined from the intensity of the intranucleotide H8-H1' NOE. Nucleotides with a *syn* glycosidic conformation have a characteristic short H8-H1' distance of ~2.5 Å (vs. 3.8 Å for *anti*) that can be observed in NOESY spectra with short mixing times. Therefore, nucleotides were assigned as *anti* with $\chi$ loosely restrained between –125 to –190° if no intranucleotide H8-H1' NOEs were observed at a mixing time of 50 ms. Where possible, other intranucleotide base-to-ribose NOEs were checked to confirm the glycosidic conformation. The presence of a strong base-to-H3' NOE for nucleotides with C3'-endo sugar pucker confirms an *anti* conformation (1.8–3.5 Å for *anti* vs. 4.3–5.2 Å for *syn*). The presence of a strong base-to-H2' NOE for nucleotides with C2'-endo sugar pucker confirms an *anti* conformation (2.0–3.8 Å for *anti* vs. 3.4–4.6 Å for *syn*).

Additional torsion angles were added for A-form regions of the RNA, using standard values for the angles $\alpha$–$\zeta$ and $\chi$ (see Discussion). Criteria applied to determine if a nucleotide was consistent with an A-form geometry were: (i) the nucleotide is involved in a Watson–Crick base pair; (ii) the sugar pucker is C3'-endo; (iii) the NOE from the H8 proton to the H2' proton of the 5'-nucleotide is strong (< 2.5 Å); and (iv) for adenosines, the characteristic cross-strand H2 to H1' NOE is present. Based on these criteria, A-form torsions for $\alpha$–$\zeta$ were included for the nucleotides 41–42, 44–45, 49–54, 64–67, 69–70, 75–76, and 78–79. A 10° range around the following values for the torsion angles was used during restrained molecular dynamics: $\alpha$: –68°, $\beta$: 178°, $\gamma$: 54°, $\delta$: 82°, $\varepsilon$: –153°, $\zeta$: –71° and $\chi$: –158°. At junctions between A-form and non-A-form nucleotides the following torsion angles were removed. At a junction where the A-form nucleotide was on the 5' side, the torsion angles $\varepsilon$ and $\zeta$ were removed. At a junction where the A-form nucleotide was on the 3' side, the torsion angles $\alpha$, $\beta$, and $\gamma$ were removed for the A-form nucleotide.

Additional pseudo-torsions were added to keep base pairs from distorting too far from planarity. For A-form base pairs, the twist was set to a 10° range around the canonical A-form values. For all other base pairs, the torsion was set to a 40° range between –20° and 20° of planarity. This was necessary since, during the early stages of dynamics, only a quartic repulsive term was used for nonbonding interactions; therefore, no attractive force was present for stacking of the base pairs.

*Molecular modeling*

Molecular modeling was performed with NMRchitect within INSIGHT 2.30 (Biosym, Inc.), using DGII for distance geometry calculations, and DISCOVER with an AMBER force field for molecular dynamics calculations.

*Distance geometry* Using the distance and torsion restraints obtained from NMR, 20 different RNA structures were embedded. Triangle bound smoothing was performed before embedding, using an independent metrization protocol with constant majorization. No optimization using the DGII algorithm was performed.

*High-temperature dynamics* Dynamics at 1000 K was performed on the 20 starting structures obtained from the distance geometry calculation. The force constants for all restraints and energy terms were scaled to 1% of their normal value, except for the chirality terms, which were set to 50 times their normal value. The values of the force constants used for distance and torsion restraints were 10 kcal/mol Å$^2$ and 60 kcal/mol deg$^2$, respectively. The covalent, distance/dihedral, and nonbonded force constants

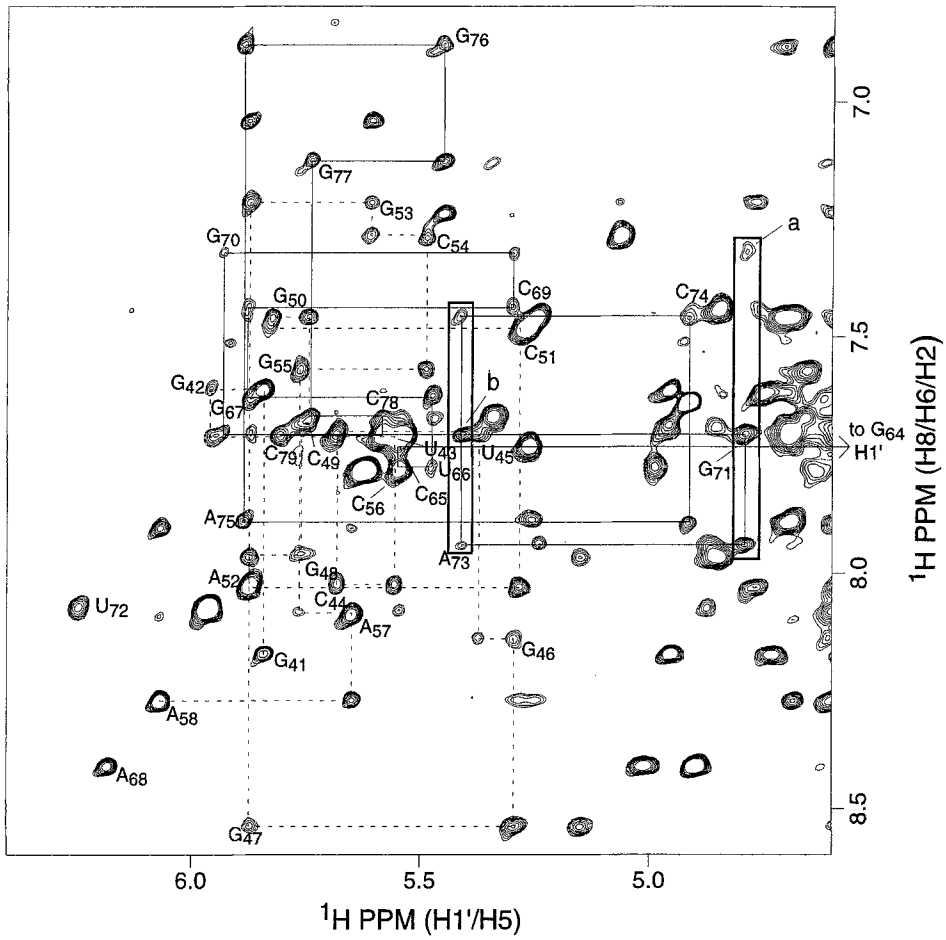


Fig. 2. Base to H1' region of a NOESY spectrum of the RRE RNA–Rev peptide complex in 99.996% $D_2O$ at 25 °C. The sequential assignment pathway of the RNA is highlighted. The dashed line traces the nucleotides G41 to A58, and the solid line traces the nucleotides G64 to C79. The numbering indicates intranucleotide H8/H6 to H1' NOEs. Two H1's that had NOEs to three H8/H6 protons are boxed. The nonstandard sequential H8 to H1' NOEs that were observed are indicated by peaks **a** and **b**. Peak **a** is G70 H8 to G71 H1', and peak **b** is G71 H8 to A73 H1'. The G64 H1' resonance is shifted downfield at 3.75 ppm, and is not included in this plot. This shift is a result of the tetraloop structure and has been described previously (Heus and Pardi, 1991).

were sequentially scaled to full value over 24 ps of simulation time. A repulsive quartic potential was used for nonbonding interactions, and Coulombic interactions were ignored. Minimization was performed before and after dynamics. High-temperature dynamics was then continued at 1000 K in 20 ps intervals with all force constants scaled to full value, except the nonbonded terms, which were set to 1% of their full value for 10 ps, scaled to full value over 5 ps, and kept at full value for the remaining 5 ps. Minimization was performed before and after dynamics. No Coulombic interactions and a repulsive quartic nonbonding potential were used, except for the last round of minimization where a Lennard-Jones nonbonding potential was used. This 20 ps dynamics cycle was repeated five times.

*Annealing* At this stage, the eight structures with the lowest restraint violation energies were selected for continued refinement. The structures were cooled to 10 K over a 10 ps time period (1 ps at 1000 K, 1 ps at 500 K, 1 ps at 200 K, 1 ps at 100 K, 1 ps at 50 K, 2 ps at 20 K, and 3 ps at 10 K). Minimization was performed before and after dynamics. The same nonbonding potentials as described for the last step of the high-temperature dynamics were used during the annealing. The annealing was repeated until the temperature of the molecule during dynamics remained equilibrated with the temperature bath. This was to ensure that the molecules had settled in a local minimum before beginning low-temperature dynamics.

*Low-temperature dynamics* At this stage of refinement, Coulombic interactions were used during dynamics with a distance-dependent dielectric constant, and an attractive Lennard-Jones potential was used for nonbonding interactions. Dynamics was performed at 10 K for 10 ps, followed by 10 ps at 300 K, and then another 10 ps at 10 K. Minimizations were performed before, between and after dynamics. An average structure was generated by superimposing all eight structures, then averaging the coordinates. The resulting average structure was then refined by 1000 steps of conjugate gradient

minimization, annealed with restrained molecular dynamics by cooling from 300 K to 10 K over 5 ps, and subjected to another 1000 steps of conjugate gradient minimization. In order to examine the effect of the A-form torsion restraints, a final 5 ps of dynamics at 300 K was performed on the eight final structures with the indirect A-form and planar torsions turned off. An average structure for the RNA with the indirect torsions removed was generated as described above.

## Results

*Sequential NOE pattern of RRE*

Sequence-specific assignment of the RRE RNA complexed with the Rev peptide was not obtainable using only conventional homonuclear 2D NMR experiments; therefore the use of isotopic labeling and heteronuclear NMR experiments was required. This was due to resonance overlap as well as the presence of unusual sequential NOE patterns. Sequential assignment of nucleic acids relies on a particular pattern of NOEs between base and H1' protons that is observed in Watson–Crick duplexes. When other types of structures are present, this pattern may be altered, making it difficult to unambiguously determine the assignments. The sequential NOE pathway finally determined for RRE is shown in Fig. 2. The details of the heteronuclear experiments performed to obtain the assignments of RRE are described below. Sequential i,i–1 NOEs from H8/H6 protons to H1' protons of the 5' neighboring nucleotide were observed along almost the entire length of both strands of the hairpin, including the internal loop region. Two breaks in the sequential connectivity for U72 and A68 were observed; however, non-sequential (i,i–2) NOEs were observed from A73 to G71 and from C69 to G67. This pattern is consistent with unstacking of U72 and A68 as single nucleotide bulges

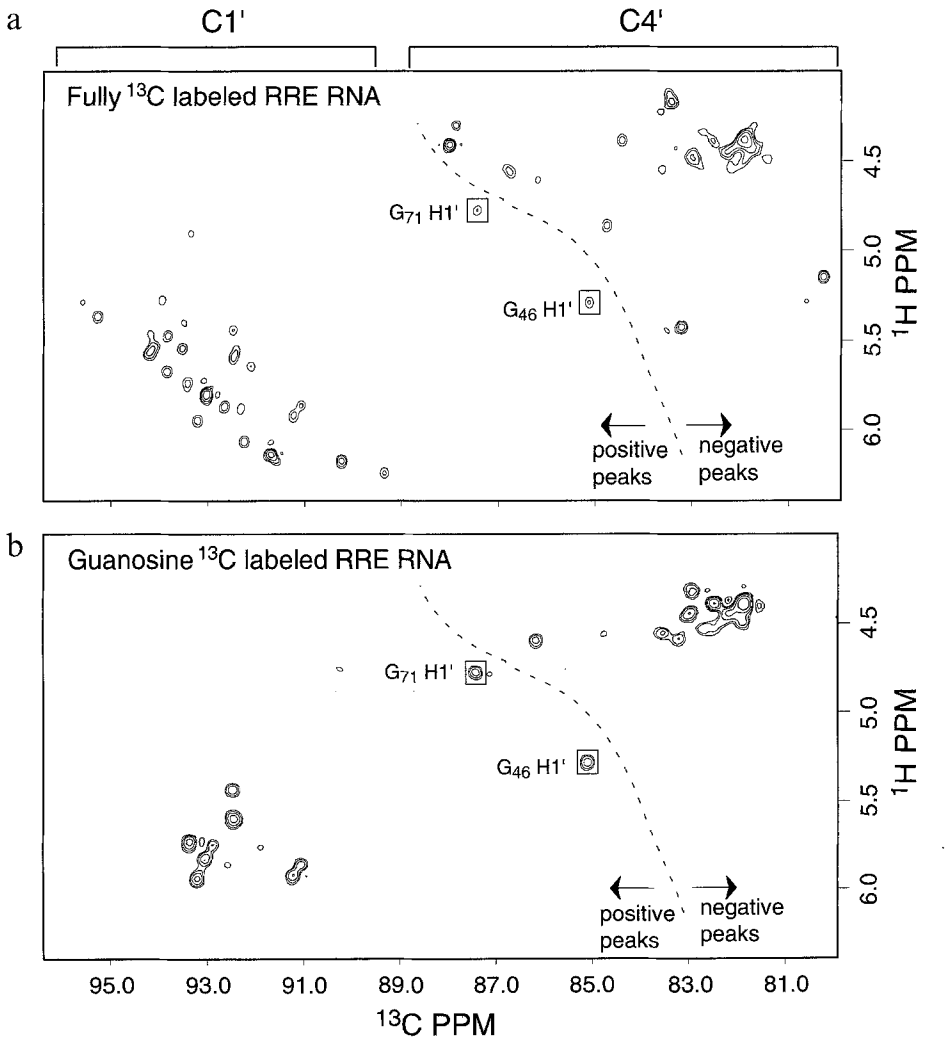


Fig. 3. (a) C4' and C1' regions of a heteronuclear single-quantum coherence experiment performed in the constant-time mode (HSQC-CT) on the uniformly $^{13}$C-labeled RRE RNA–Rev peptide complex. Positive and negative peaks are indicated by their position relative to the dashed line. Two C1's are boxed that have unusual carbon chemical shifts. (b) Same region of an HSQC-CT on the $^{13}$C-guanosine-labeled RRE RNA–Rev peptide complex. Note that a few cross peaks listed in the assignment table are absent from these constant-time experiments, due to short carbon T2s. These resonances were observed either in plots of the HSQC-CT closer to the noise level or in 3D NOESY-HMQC/HCCH-DIPSI experiments.

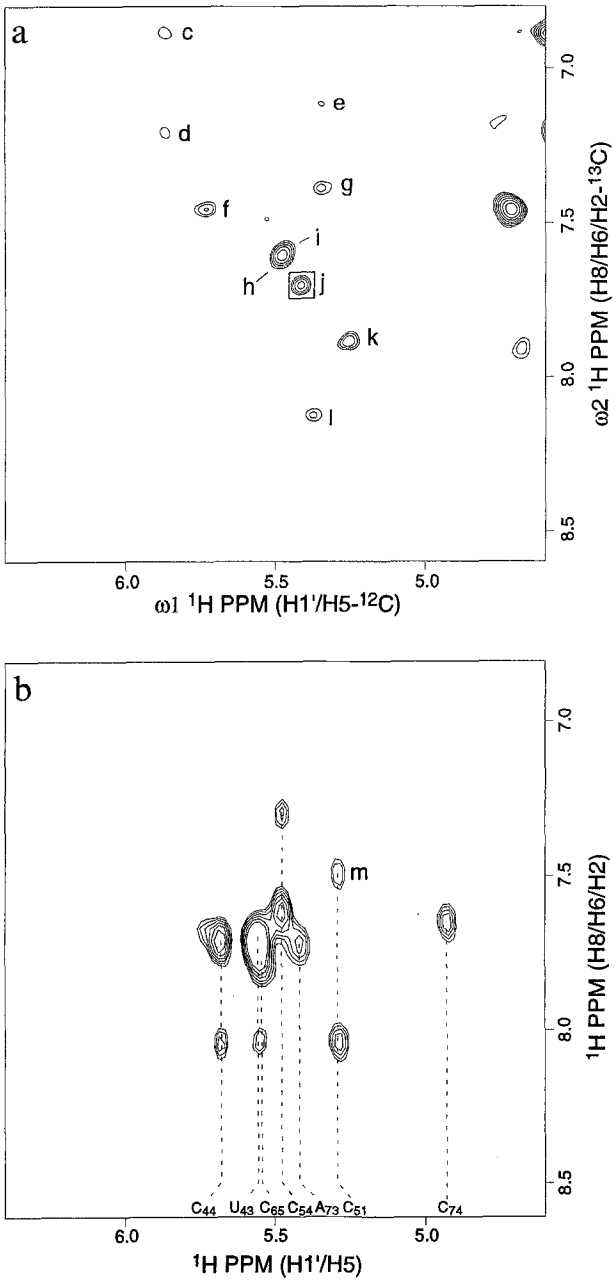


Fig. 4. (a) $\omega_2$-selected, $\omega_1$-filtered subspectrum of a double-half-filtered NOESY on the $^{13}$C-guanosine-labeled RRE RNA–Rev peptide complex. This subspectrum selects protons attached to $^{13}$C on the $\omega_2$ axis and filters out protons attached to $^{13}$C on the $\omega_1$ axis. All H8 to H5/H1' NOEs are lettered as follows: **c**=G76 H8-A75 H1', **d**=G53 H8-A52 H1', **e**=G77 H8-C78 H5, **f**=G50 H8-C49 H1', **g**=G70 H8-C69 H1', **h**=G67 H8-U66 H1', **i**=G55 H8-C54 H1', **j**=G71 H8-A73 H1', **k**=G64 H8-C65 H5, **l**=G46 H8-U45 H1'. (b) One 2D plane of a 3D NOESY-HMQC experiment in the C1' region at a $^{13}$C chemical shift of 93.7 ppm. Assignments of the H1's present at this slice are given. Peak **m** is a cross peak not resolvable in a 2D NOESY experiment, due to overlap (see text). Note that all NOEs to base protons for H1 of A73 and C74 are not observed, since this is not the plane of optimal intensity for these protons.

along a continuous helix. In addition, there were two 'reverse' connectivities from H8 protons of G70 and G71 to H1' protons of G71 and A73, respectively (peaks **a** and **b** in Fig. 2). The experiments performed to elucidate this unusual NOE pathway and the structural features that give rise to these NOEs are discussed below.

*Identification of H1' resonances*

To unambiguously identify all H1' protons for assignment of the sequential H8/H6 to H1' pathway, the constant-time variant of the Heteronuclear Single Quantum Coherence (HSQC-CT) experiment was performed (Santoro and King, 1992). When the constant-time interval is set to an odd integer multiple of $1/J_{CC}$ (the one-bond carbon–carbon coupling constant), the sign of the cross peak for the C1' and C5' resonances is opposite to that of the C2', C3' and C4' resonances (Santoro and King, 1992). This is important because in more complex RNA structures, $^{1}$H and $^{13}$C chemical shifts alone may not be sufficient to identify H1' proton resonances, which can be shifted upfield into the H2'/H3'/H4'/H5'/H5" region. Figure 3a shows the C1' and C4' region of an HSQC-CT spectrum on a uniformly $^{13}$C-labeled RRE RNA complexed with Rev peptide. Highlighted are two C1'/H1' cross peaks that are shifted upfield in both the proton and carbon dimensions. These two resonances were unambiguously assigned as C1'/H1' correlations by the sign of the peak in the HSQC-CT spectrum. Similar experiments on the aromatic region of the RNA can distinguish the pyrimidine from the purine resonances due to the difference in carbon–carbon couplings between the C8/H8 and C6/H6 spin systems (data not shown).

*Specific $^{13}$C-labeling to resolve NOE ambiguities*

To further aid in the RRE RNA assignment, heteronuclear experiments were performed on a sample where only the guanosine residues were $^{13}$C-labeled. If an RNA is labeled specifically by nucleotide, the ribose protons for the labeled nucleotide can be unambiguously identified by the fact that they are attached to a $^{13}$C-carbon (ribose spin systems are identical, regardless of the nucleotide attached). Figure 3b shows the HSQC-CT spectrum for the $^{13}$C-guanosine RRE RNA–Rev peptide complex. In this spectrum, all of the guanosine H1' protons can be identified, which helps reduce the ambiguity in assigning the through-space H8/H6 to H1' NOE sequential pathway. Due to the presence of the unusual H8-H1' NOEs in the internal loop, this information was important for confirming the assignments of RRE RNA.

An additional advantage of labeling RNA with only one type of $^{13}$C-nucleotide is that isotopic filtering experiments can be performed (Otting and Wüthrich, 1989, 1990). In a double-half-filtered NOESY spectrum, the conventional NOESY spectrum is divided into four different subspectra, filtering for or against protons bonded to $^{13}$C in each dimension (Otting and Wüthrich, 1989,1990). One subspectrum from a double-half-filtered NOESY of the RRE-Rev peptide complex is shown in Fig. 4a. In this

spectrum, one proton involved in the NOE is bonded to $^{12}C$ ($\omega_1$ axis), while the other is bonded to $^{13}C$ ($\omega_2$ axis). The region shown is the same as in Fig. 2, and all cross peaks are between guanosine H8 protons bonded to $^{13}C$ and H1'/H5 protons bonded to $^{12}C$, which must be internucleotide NOEs. Distinguishing between inter- and intranucleotide NOEs was important for assignment of the internal loop region of the RNA, where an ambiguous set of sequential NOEs was observed. For instance, A73 H1' exhibited three NOEs to H8/H6 protons, one intranucleotide and two internucleotide (see Fig. 2). The 'reverse' i,i+2 NOE between G71 H8 and A73 H1' (peak **j**) can be identified from this double-half-filtered subspectrum, since it is the only NOE of the three that involves a guanosine H8. The other two cross peaks, a sequential i,i−1 internucleotide NOE from C74 H6 to A73 H1' and an intranucleotide NOE from A73 H8 to A73 H1', are absent from this double-half-filter subspectrum, since all the protons involved in the NOEs are bonded to $^{12}C$. Instead, these NOEs were observed in a different double-half-filter subspectrum that only contains NOEs between protons that are both bonded to $^{12}C$ (data not shown).

*Three-dimensional experiments to alleviate overlap*

With both uniformly and specifically $^{13}C$-labeled RRE RNA, 3D heteronuclear NMR experiments were performed to alleviate spectral overlap by resolving the $^1H$ 2D NOESY, COSY, and TOCSY spectra in a third dimension according to $^{13}C$ chemical shift. A representative plane through the C1' region from a 3D NOESY-HMQC spectrum (Clore et al., 1990) is shown in Fig. 4b. Although this is the most crowded C1' plane, there are seven fairly resolved H1' resonances. One example of an overlap problem that was alleviated by this experiment is illustrated by peak **m**, which is an NOE from C51 H6 to C51 H1' that was obscured by a strong H5-H6 cross peak in the 2D NOESY spectrum (Fig. 2). The H1' assignments obtained from the filtered NOESY and NOESY-

TABLE 2
NONEXCHANGEABLE PROTON AND CARBON CHEMICAL SHIFTS (ppm) FOR THE RRE RNA BOUND TO Rev PEPTIDE

| Residue | H8,H6/C8,C6 | H5,H2/C5,C2 | H1'/C1' | H2'/C2' | H3'/C3' | H4'/C4' |
|---|---|---|---|---|---|---|
| G41 | 8.17/139.3 | – | 5.84/90.5 | 4.96/74.8 | 4.73/(74.5) | 4.55/83.5 |
| G42 | 7.62/136.9 | – | 5.96/93.1 | 4.70/75.5 | 4.49/72.3 | 4.54/82.4 |
| U43 | 7.71/140.1 | 5.56/104.6 | 5.56/93.4 | 4.27/75.7 | 4.56 | 4.45 |
| C44 | 8.03/142.1 | 5.85/97.3 | 5.68/93.7 | 4.33/75.2 | 4.36/72.3 | 4.48/81.2 |
| U45 | 7.71/140.9 | 4.53/102.8 | 5.38/95.1 | 4.22/75.0 | 4.46/71.6 | 4.34/82.6 |
| G46 | 8.14/138.9 | – | 5.29/85.1 | 4.38/78.5 | 4.49/77.1 | 4.60/86.1 |
| G47 | 8.54/141.4 | – | 5.87/92.5 | 4.15/75.7 | 5.15/73.9 | 4.44/82.9 |
| G48 | 7.97/138.4 | – | 5.77/87.7 | 4.86/77.6 | 4.15 | 4.53/84.7 |
| C49 | 7.69/142.1 | 4.95/97.9 | 5.74/94.7 | 4.72/75.1 | 4.34 | (4.59) |
| G50 | 7.45/135.2 | – | 5.83/92.9 | 4.52/75.1 | 4.55 | 4.49/81.7 |
| C51 | 7.48/139.9 | 5.28/97.5 | 5.28/93.8 | 4.43/74.9 | 4.52/71.4 | (4.38)/(81.7) |
| A52 | 8.03/139.4 | 7.04/152.4 | 5.87/92.5 | 4.57/75.7 | 4.76/72.0 | (4.25) |
| G53 | 7.22/135.6 | – | 5.61/92.4 | 4.28/75.4 | (4.40)/(71.7) | (4.43) |
| C54 | 7.27/139.6 | 5.04/97.6 | 5.48/93.7 | 4.49/75.7 | 4.36/75.6 | (4.07) |
| G55 | 7.57/136.5 | – | 5.77/93.3 | 4.53/76.1 | 4.64/73.2 | 4.39/82.3 |
| C56 | 7.78/143.4 | 5.63/97.8 | 5.54/92.4 | 4.43/75.7 | 4.24/74.8 | 4.11/83.2 |
| A57 | 8.09/140.7 | 7.90/155.4 | 5.65/92.1 | 4.35/76.9 | 4.60/76.0 | 4.23/83.6 |
| A58 | 8.28/141.5 | 8.10/154.8 | 6.07/92.1 | 4.69/77.0 | 5.28 | 4.50/82.6 |
| G64 | 7.88/137.5 | – | 3.75/92.8 | 4.33/74.8 | 4.18/(74.2) | (4.33) |
| C65 | 7.73/135.4 | 5.26/96.8 | 5.55/94.0 | 4.45/75.0 | | |
| U66 | 7.77/141.4 | 4.99/103.2 | 5.48/94.0 | 4.41/75.1 | (4.35) | (4.39) |
| G67 | 7.63/136.8 | – | 5.87/91.0 | 4.21/78.2 | 4.75/74.6 | (4.40) |
| A68 | 8.41/142.7 | 8.30/155.7 | 6.18/90.1 | 4.90/75.8 | 5.0275.3 | 4.41/84.3 |
| C69 | 7.44/143.3 | 4.85/96.1 | 5.29/95.6 | 4.04/75.7 | 4.17 | 4.43/83.2 |
| G70 | 7.32/134.6 | – | 5.93/91.1 | 4.47/76.4 | (4.45) | 4.36/81.7 |
| G71 | 7.70/137.7 | – | 4.78/87.4 | 3.93/79.1 | | (4.16) |
| U72 | 8.07/144.4 | 5.97/105.5 | 6.24/89.3 | 4.46/75.4 | 4.29/78.8 | 4.86/84.5 |
| A73 | 7.94/141.1 | 7.63/152.4 | 5.41/93.4 | 4.69/75.1 | 4.31 | 4.54/82.0 |
| C74 | 7.45/139.6 | 5.24/97.1 | 4.91/93.2 | 4.18/74.9 | 4.47 | 4.31/81.5 |
| A75 | 7.89/139.3 | 7.24/152.6 | 5.88/92.2 | 4.59/76.0 | 4.70/72.0 | (4.40)/(81.4) |
| G76 | 6.89/135.8 | – | 5.45/92.4 | 4.32/75.7 | 4.25/74.2 | 4.38/82.0 |
| G77 | 7.13/136.6 | – | 5.74/93.1 | 4.66/75.1 | 4.33/72.6 | 4.44/82.0 |
| C78 | 7.67/140.8 | 5.34/97.1 | 5.57/94.1 | 4.24/75.7 | (4.46) | (4.16) |
| C79 | 7.71/141.9 | 5.53/97.9 | 5.81/92.9 | 4.05/77.6 | 4.18 | 4.43/83.2 |

Numbers in parentheses indicate a tentative assignment. The errors in chemical shift are ±0.01 and ±0.10 ppm for proton and carbon, respectively. Proton shifts are referenced to internal TSP. Carbon shifts are referenced to an external TSP sample.

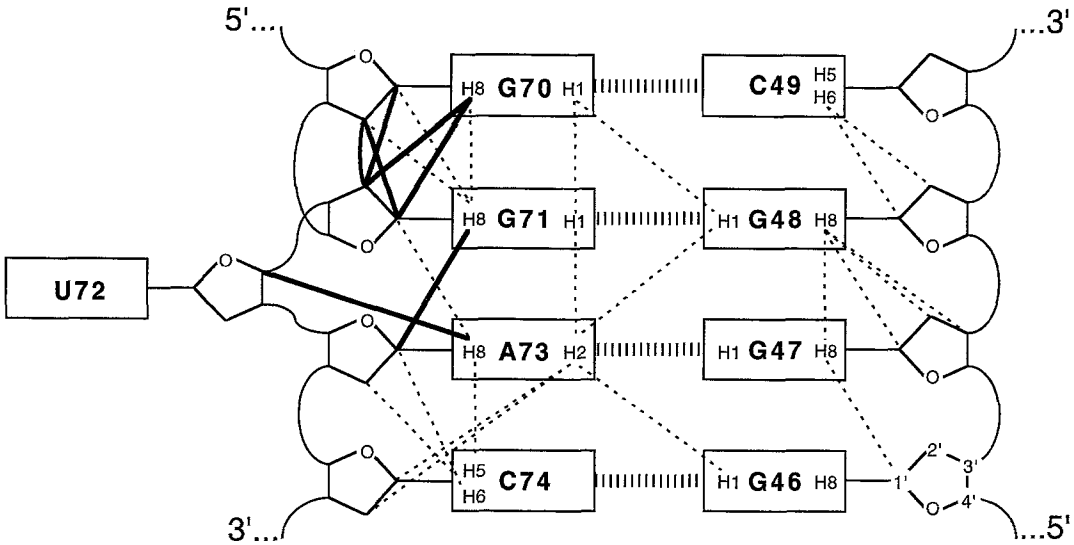


Fig. 5. Schematic diagram of the NOEs observed in the internal loop region. Thick dashes between nucleotides indicate base-pair formation. Thin dashed lines indicate observed NOEs. The unusual sequential and ribose–ribose NOEs observed are highlighted by thick solid lines. NOEs to ribose protons are drawn to the respective positions on the ribose ring. As an example, the numbering for the ribose of G46 is given.

HMQC experiments were extended to other ribose protons using 3D HCCH-COSY and HCCH-DIPSI experiments (Clore et al., 1990; Nikonowicz and Pardi, 1993) to obtain through-bond correlations. The nonexchangeable proton and carbon assignments for the RRE RNA bound to the Rev peptide are given in Table 2. All base, H1', H2' and most of the H3' and H4' protons were assigned. Coherence transfer from the H5'/H5" to the H1' protons in the HCCH-DIPSI experiment was poor; therefore, very few H5'/H5" protons were assigned. Our assignments for the base, H1', and H2' protons are nearly identical to those for a similar RNA sequence bound to the same peptide, as previously reported (Peterson et al., 1994). The only significant differences are in the assignments for the H1' of G71 and A73; based on the data presented in Figs. 2–4, we believe these new assignments are correct.

*Identification of a minor RNA conformer*

Several additional peaks were observed in NOESY spectra of the complex, due to slow exchange between different conformations. Exchange peaks were distinguished from NOEs by the opposite sign of cross peaks in a ROESY spectrum (data not shown). The conformational exchange was very local, including resonances A68 H2, A52 H8, A52 H1', C69 H1', and C69 H5 in the upper stem. The RRE–Rev peptide complex exists in a major (90–95%) and minor (5–10%) conformation. The chemical shifts of the minor conformation do not coincide with free RNA; therefore, the exchange must be with another RNA conformation bound to the peptide. In the major conformation, A68 is clearly unstacked, and we believe that the minor conformer may involve stacking of A68 into the upper stem, due to the nature of the chemical shift changes of the exchange peaks. Mutation of A68 to U68 results in a complex whose NOESY spectra are nearly identical to the wild type, except for the absence of the exchange peaks (H. Mao and J. Williamson, unpublished results). This is consistent with the hypothesis that the exchange peaks result from stacking, since single-nucleotide pyrimidine bulges have a much lower propensity to stack than single-nucleotide purine bulges.

*Summary of NMR data*

In the stem region of RRE, there were NOEs typical of A-form RNA helices, except for the internal loop region. The pattern of NOEs for the internal loop region of RRE is shown schematically in Fig. 5. A normal H8/H6 to H1' sequential pattern of NOEs was observed from residues G46 to G48; however, a much more extensive network of NOEs was observed from G70 to C74. The two unusual 'reverse' sequential H8 to H1' NOEs are highlighted by bold lines in Fig. 5. In addition, there are several unusual ribose to ribose NOEs between G71 and G70, which are also highlighted. While there are no internucleotide NOEs involving the H6 and H1' protons of the bulged U72, there is one NOE from A73 H8 to U72 H4', positioning the ribose of U72 in the major groove. No intranucleotide H8/H6 to H1' NOEs were observed in a 50 ms NOESY, indicating that all glycosidic torsions are in the *anti* conformation. In another study, the G71 H8 to G71 H1' NOE was identified as strong (< 3.0 Å), suggesting that G71 was in the *syn* conformation (Peterson et al., 1994). According to our assignments and the filtered NOESY spectra, this strong cross peak represents an *inter*nucleotide NOE between G71 H8 and A73 H1'. The intranucleotide H8-H1' cross peak we have identified for G71 is not present at a mixing time of 50 ms. In addition, this nucleotide has a strong H8-H2' NOE at 50

TABLE 3
SUMMARY OF RESTRAINTS USED FOR MOLECULAR MODELING

| Type of restraint | Number of restraints[a] | Restraints per nucleotide |
|---|---|---|
| **Distance restraints** | | |
| Hydrogen bond | 44 | 1.3 |
| Exchangeable proton NOEs | 35 | 1.0 |
| Intranucleotide nonexchangeable NOEs | 184 (41) | 5.4 |
| Internucleotide nonexchangeable NOEs | 113 (38) | 3.3 |
| Total distance restraints | 376 (79) | 11.0 |
| **Torsion restraints** | | |
| Glycosidic bond | 34 | 1.0 |
| Sugar pucker[b] | 60 | 1.8 |
| A-form[c] | 70 | 2.1 |
| Planar base pairs | 41 | 1.2 |
| Hydrogen bonded aminos[d] | 24 | 0.7 |
| Total torsion restraints | 229 | 5.9 |
| Total all restraints | 605 | 17.8 |

[a] Numbers in parentheses are the number of distance restraints that could only be resolved in a 3D NOESY-HMQC.
[b] The torsions $\nu_2$ and $\nu_4$ were set to restrain the sugar puckers.
[c] Criteria for designation of a torsion as A-form are described in the text.
[d] These torsions are to prevent the amino protons involved in hydrogen bonds from flipping to the other amino proton position.

ms, consistent with a nucleotide in an *anti*/C2'-endo conformation (see Materials and Methods). Intranucleotide base to ribose NOE intensities consistent with an *anti*/C2'-endo conformation were also observed for the other nucleotide in the G-G base pair (G48). Since the symmetric G71-G48 base pair in RRE would be expected to have one of the nucleotides in a *syn* conformation (Bartel et al., 1991; Iwai et al., 1992; Saenger, 1984), some unusual structure must be present to allow this base pair to form with both nucleotides in an *anti* conformation.

*Structure determination*

Once assignments for the RRE RNA were obtained, NOE and coupling constant data were converted to distance and torsion restraints for distance geometry and molecular dynamics calculations. A summary of all the restraints used in modeling is given in Table 3. From 20 initial DG structures, eight low-energy structures were obtained after the molecular dynamics calculations. The average restraint violation energy of the final structures was 29.8 kcal/mol and the total energy including restraints was −136.8 kcal/mol. The single largest distance and torsion violations in the eight structures were 0.36 Å and 4.4°, respectively. A complete listing of the rmsd values for the restraint and idealized geometry violations is given in Table 4. To assess the agreement of the structures obtained from restrained molecular dynamics, a stereoview of the superposition of the eight structures to the average structure is shown in Fig. 6. The average rmsd of the nucleotides in the helical region not including the A68/U72 bulges and the tetraloop was 1.94 Å, which is consistent with a moderately well-defined structure. All pairwise rmsd values of the eight structures superimposed are given in Table 5.

To analyze the effects of the A-form torsions on the resulting RNA structures, an additional round of dynamics was performed with the A-form torsions removed. The resulting average structure was then compared to

TABLE 4
STRUCTURAL STATISTICS OF LOW-ENERGY RRE RNA STRUCTURES

| Parameter | <RRE>[a] | RRE-AVG[b] |
|---|---|---|
| **Rmsd of experimental restraints** | | |
| All distance restraints (376) (Å) | 0.0399 ± 0.0033 | 0.0394 |
| Intranucleotide distance restraints (184) (Å) | 0.0232 ± 0.0030 | 0.0255 |
| Internucleotide distance restraints (192) (Å) | 0.0509 ± 0.0049 | 0.0491 |
| All torsion restraints (229) (°) | 0.7534 ± 0.0337 | 0.6861 |
| All IL[c] distance restraints (106) (Å) | 0.0603 ± 0.0091 | 0.0556 |
| Intranucleotide IL distance restraints (50) (Å) | 0.0329 ± 0.0071 | 0.0392 |
| Internucleotide IL distance restraints (56) (Å) | 0.0770 ± 0.0121 | 0.0669 |
| Maximum distance restraint violation (Å) | 0.25 ± 0.05 | 0.25 |
| Maximum torsion restraint violation (°) | 3.7 ± 0.5 | 2.8 |
| Number of distance violations >0.1 Å | 17 ± 1 | 17 |
| Number of torsion violations >1.0° | 27 ± 3 | 27 |
| **Rmsd from idealized geometry** | | |
| Bonds (Å) | 0.0049 ± 0.0001 | 0.0048 |
| Angles (°) | 0.2666 ± 0.0038 | 0.2673 |
| Impropers (°) | 0.3641 ± 0.0160 | 0.3777 |

[a] <RRE> indicates average statistics for the eight structures ± SD.
[b] RRE-AVG indicates the statistics for the average structure.
[c] IL stands for internal loop, i.e., nucleotides G46-C49 and G70-C74.

TABLE 5
PAIRWISE RMS DIFFERENCES BETWEEN THE EIGHT LOW-ENERGY STRUCTURES AND THE AVERAGE STRUCTURE OF RRE RNA

| | RRE-1 | RRE-2 | RRE-3 | RRE-4 | RRE-5 | RRE-6 | RRE-7 | RRE-8 | RRE-AVG |
|---|---|---|---|---|---|---|---|---|---|
| RRE-1 | – | 1.68 | 2.46 | 2.56 | 2.18 | 2.93 | 3.01 | 1.63 | 1.54 |
| RRE-2 | 1.20 | – | 2.00 | 2.36 | 1.64 | 1.96 | 2.92 | 1.96 | 1.33 |
| RRE-3 | 2.04 | 2.00 | – | 1.68 | 2.92 | 3.31 | 1.73 | 2.35 | 1.87 |
| RRE-4 | 2.15 | 2.04 | 1.34 | – | 3.26 | 3.41 | 2.28 | 2.58 | 2.06 |
| RRE-5 | 1.47 | 1.44 | 1.95 | 1.95 | – | 2.18 | 3.65 | 2.12 | 1.85 |
| RRE-6 | 1.16 | 1.10 | 2.09 | 2.17 | 1.67 | – | 4.41 | 3.31 | 2.55 |
| RRE-7 | 2.40 | 2.34 | 1.33 | 1.74 | 2.21 | 2.49 | – | 2.69 | 2.63 |
| RRE-8 | 1.68 | 1.66 | 2.04 | 2.17 | 1.21 | 1.81 | 2.28 | – | 1.70 |
| RRE-AVG | 1.23 | 1.22 | 1.29 | 1.56 | 1.40 | 1.29 | 1.63 | 1.58 | – |

Numbers above the diagonal are rmsd values for the stem region of the RRE, which comprises all nucleotides except G55-A58, A69 and U72. Numbers below the diagonal are rmsd values for the internal loop nucleotides G46-C49, G70-G71 and A73-C74. The average stem rmsd to RRE-AVG = 1.94 Å; the average internal loop rmsd to RRE-AVG = 1.40 Å; the average stem pairwise rmsd = 2.54 Å (not including the AVG structure); and the average internal loop pairwise rmsd = 1.83 Å (not including the AVG structure).

that generated including the A-form torsions. The two structures were very similar, with the rmsd between the two being 1.69 Å for the stem of the RNA. The average total and restraint violation energies in the eight structures with the A-form torsions removed were −151.6 and 26.0 kcal/mol, respectively. All models subsequently shown are those obtained from dynamics with the A-form torsion restraints included.

## Discussion

One significant limitation in NMR structural determination of large oligonucleotides at this time is the difficulty in obtaining torsion restraints to define the conformation of the phosphate backbone. In the standard 2D experiments for determination of coupling constants, there is a large amount of cancellation of the antiphase $^{1}H$-$^{1}H$ and $^{1}H$-$^{31}P$ COSY correlations, due to the increased line width of larger molecules such as studied here (~15 kDa). In particular, the line widths of the $^{31}P$-resonances in the bound RRE were too broad for obtaining high-quality spectra. While the 3D HCCH-COSY and HCCH-TOCSY experiments can aid in the assignment of ribose protons, there is no coupling constant information in these spectra to define any of the torsions $\alpha$–$\zeta$ in nucleic acids. Some new methods have recently been developed

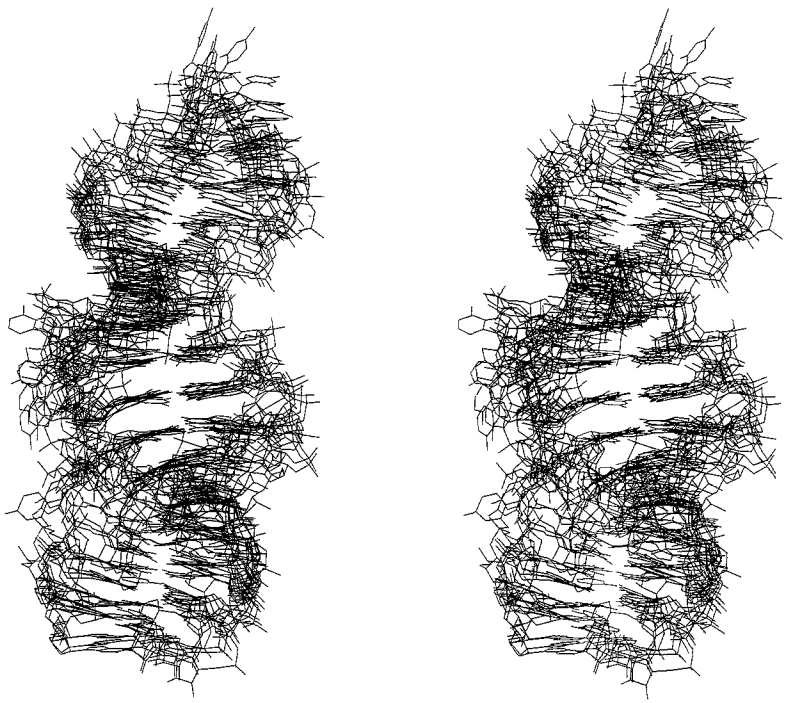

Fig. 6. Stereoview of the superpositions of the eight structures obtained from the molecular dynamics calculations (heavy atoms only). The structures were all superimposed to the average structure, which is not shown. The superposition was done on the helical stem of RRE, which comprises all nucleotides except for G55-A58, A68 and U72.

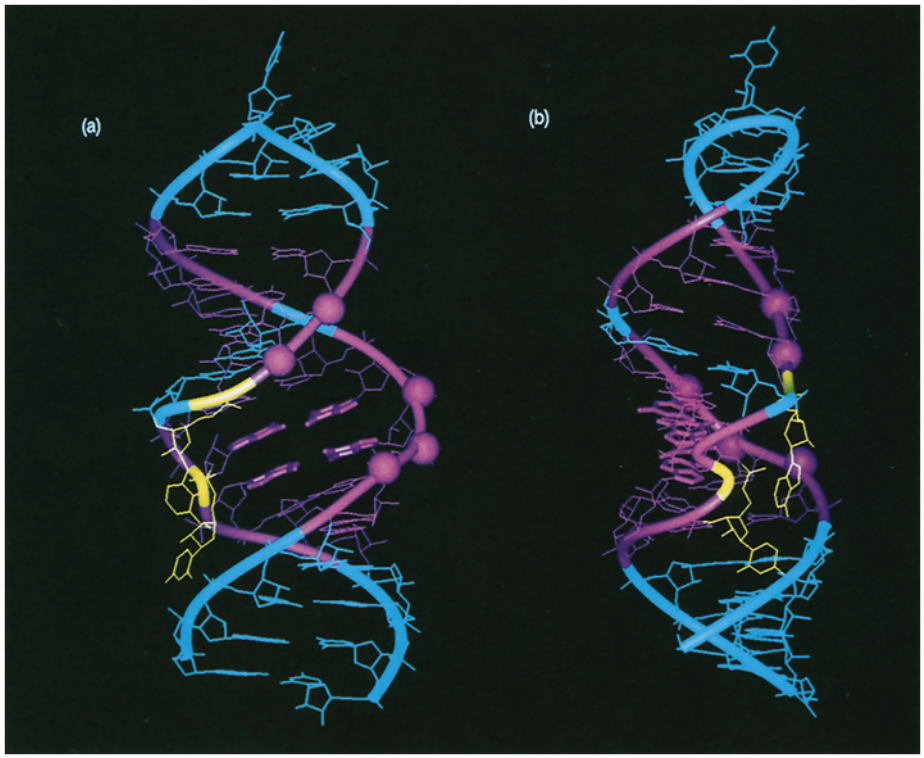


Fig. 7. Two views of the average structure of RRE RNA: (a) the major groove of the internal loop region; and (b) the structure rotated 90°, showing the phosphate backbone near the bulged U72 in the internal loop. In both views, nucleotides determined to be important for binding from biochemical studies (Bartel et al., 1991; Kjems et al., 1992) are shown in magenta. The important phosphates (Kjems et al., 1992) and the purine–purine base pairs are shown as large spheres and thick sticks, respectively. The two bulged nucleotides are colored yellow.

utilizing E.COSY-type correlations, which are less sensitive to line width problems (Hines et al., 1994; Schwalbe et al., 1994). However, these methods still suffer from serious overlap problems for larger RNAs, particularly in A-form regions.

For RRE RNA, the only coupling constants directly determined were for the H1'-H2' protons, which crudely define the sugar pucker. In the absence of backbone torsion restraints, it is difficult for molecular dynamics calculations to properly define the structure of a large nucleic acid based on the relatively small number of internucleotide distance restraints obtained in this study. Therefore, to help with the refinement process, additional torsion restraints were added for nucleotides strongly suggested by NOE data to be in an A-form helical conformation (see Materials and Methods). These included all of the Watson–Crick base pairs in RRE, except for G46-C74 immediately below the internal loop. Care was taken to remove A-form torsions for nucleotides next to non-A-form regions, such as bulges and other non-WC base pairs. Including canonical A-form torsions will bias the resulting structures such that small distortions from A-form geometry that might be present will be lost. It is unlikely, however, that there are sufficient internucleotide NOEs in this study to uniquely define such small deviations. Additionally, upon removal of A-form torsions at the end of the dynamics calculations, the RNA structures did not change significantly. Perhaps most importantly, the internal loop region changed very little upon removal of the A-form restraints, with the average rmsd before and after dynamics being 1.34 Å. The bias imposed by performing the previous molecular dynamics calculations including the A-form torsion restraints is not necessarily eliminated by simply removing the A-form restraints at the end, yet this shows that there are no distance restraints that were conflicting with the torsion restraints. The unusual region of the RRE that is the primary interest of this study is the internal loop. No indirect torsion restraints were included in the internal loop region, and the resulting structure of the internal loop is defined solely by the NMR data.

The general features of the RRE RNA structure are consistent with the model, proposed on the basis of biochemical experiments, that the peptide is recognizing the RRE in the major groove around the purine-rich internal loop. Two views of the average structure, rotated 90° with respect to one another, are shown in Fig. 7. Highlighted in magenta are the nucleotides that are important for binding of the peptide as determined from biochemical studies (Bartel et al., 1991; Kjems et al., 1992). The phosphates that interfere with peptide binding upon ENU modification (Kjems et al., 1992) are shown as large

spheres. These phosphates flank the major groove around the internal loop region, consistent with the fact that the peptide is recognizing the major groove of the RRE. Preliminary analysis of RNA–peptide NOEs was also consistent with peptide binding in the major groove (data not shown).

The most striking feature of the RNA is a large kink in the backbone around the bulged U72 nucleotide. This kink in the phosphate backbone of RRE RNA is due to the fact that the strand polarity reverses direction for nucleotide G71, enabling the G71-G48 base pair to form with both glycosidic torsions *anti*. A detailed view of the nucleotides in the internal loop involved with the kink of the phosphate backbone is shown in Fig. 8. Between A73 and G71, which are stacked, the backbone goes into the major groove for the bulged U72, then loops around to the ribose of G71. The ribose of G71 is oriented in the opposite direction from the rest of the strand, which can be easily discerned by looking at the position of the O4' oxygen. In this conformation, the positions of the H8 and H1' protons (green and magenta spheres, respectively) are reversed with respect to the major and minor groove for this one nucleotide. This results in the unusual sequential connectivity patterns observed in the NMR spectra. The sequential NOEs for G71 H1' are highlighted with white arrows to show how the H8 protons in either direction along the strand are close enough in space to observe an NOE. The reversal of the backbone orientation for the G71 ribose also explains the unusual ribose to ribose NOEs observed. In the RRE RNA structure, the H2' proton of G71 is positioned toward the ribose of G70 as a result of the reversal, and exhibits NOEs to the H1' and H2' protons of G70. Interestingly, a very similar structure with a bulged nucleotide reversing the strand polarity for one nucleotide has been observed in the structures of the sarcin/ricin and E loops of ribosomal RNA determined by NMR (Szewczak et al., 1993; Wimberly et al., 1993). In both cases, an *anti–anti* purine–purine base pair was formed, where a *syn–anti* base pair would have been expected. It appears that this type of structure may be a common fold for RNA.

The geometric constraints on formation of the G-G base pair are shown schematically in Fig. 9. The two faces of a nucleotide base are stereochemically distinct, as represented in the figure by black and white faces. In a standard right-handed antiparallel helix, a given face is oriented in opposite directions on each strand (Fig. 9a). In order to form the G-G base pair observed in RRE, the same face of both guanosine residues must be oriented in the same direction (Figs. 9b,c). There are two possible ways to accommodate this geometry into an antiparallel helix. One is to change the orientation of one glycosidic bond to the *syn* conformation (Fig. 9b). Structural predictions for the RRE had assumed that one of the guanosine positions in the RRE must be *syn*. Alternatively, the orientation of the phosphate backbone can be changed such that the strands are effectively parallel for the G-G base pair, while the glycosidic conformation remains *anti*, as was illustrated in the RRE structure in Fig. 8. Some interesting results regarding these possible orientations

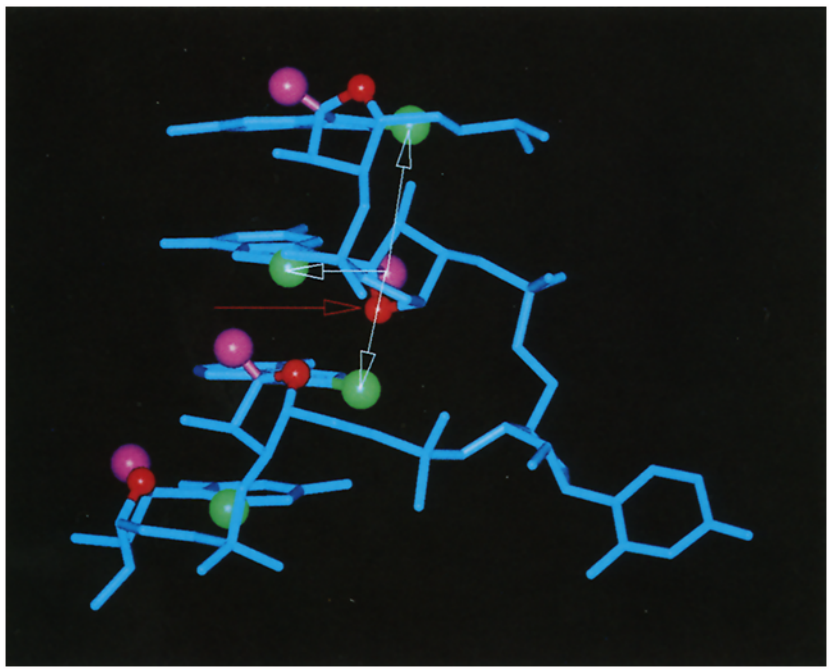

Fig. 8. View along the phosphate backbone of the nucleotides G70-C74. H6/H8 protons, H1' protons and O4' oxygens are highlighted as green, magenta and red spheres, respectively. The red arrow highlights the orientation of the O4' in G71 compared to the rest of the riboses in the helix. The white arrows highlight the three NOEs to base protons observed for G71 H1'.

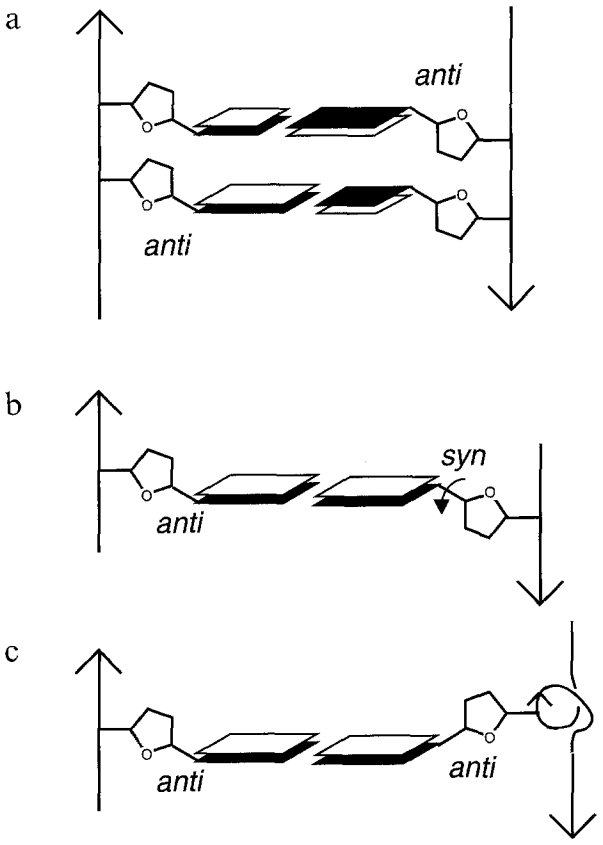


Fig. 9. A schematic of two different ways the G-G base pair observed in RRE can be incorporated into an antiparallel helix. Each base has two distinct faces, which are represented by black and white shadowing. (a) Orientation of the bases in an antiparallel Watson–Crick duplex. (b) G-G base pair with one of the guanosine glycosidic torsions *syn*. (c) G-G base pair with the duplex locally parallel stranded and both glycosidic torsions *anti*.

have been obtained from two molecular modeling studies of the RRE RNA, where three-dimensional models of the RNA were built using the available secondary structure data as restraints (Le et al., 1994; Leclerc et al., 1994). In both cases the glycosidic torsion of G71 was initially restrained as *syn*; however, during molecular dynamics without restraints the torsion flipped to the *anti* position, to give structures grossly similar to the structure presented here. Potentially in the context of an adjacent single-nucleotide bulge, the *anti–anti* G-G base pair may be more sterically favorable than the *syn–anti* form.

The orientation of the G-G pair described above places the N7 of G71 in the minor groove and G47 in the major groove. This is seemingly inconsistent with the chemical modification data that has been accumulated on the RRE–Rev interaction. DEPC modification of the N7 position at G71, but not G47, interfered with Rev binding (Kjems et al., 1992), suggesting that G71 N7 is interacting with the peptide in the major groove. However, subsequent studies with 7-deazaguanosine substitutions showed that neither N7 in the G-G base pair was critical for binding of the Rev protein (Pritchard et al., 1994). The bulky DEPC group attached to the N7 position was proposed to interfere with folding of the RNA, rather than directly inhibiting a contact with the protein (Pritchard et al., 1994). Our results suggest that the modification of G71 with DEPC was in fact not directly interfering with binding, since the N7 is positioned in the minor groove and preliminary NMR data suggest that the peptide is lying in the major groove (data not shown).

Interestingly, this leaves no biochemical evidence of any direct hydrogen-bonding interaction between Rev and the G-G base pair, even though this base pair is absolutely required. In vitro selection studies found that the only mutations allowed for the G-G base pair were covariations to A-A and C-A, which can form base pairs isosteric to the G-G pair (Bartel et al., 1991; Giver et al., 1993). In order to form the G-G or any isosteric base pair as shown in this study, a single-nucleotide bulge 3' to position 71 would be required for the reversal of the phosphate backbone. In vitro selection studies found that only the presence, but not the identity of the nucleotide at this position was important (Bartel et al., 1991). It was also observed that substitution of this nucleotide with a C3 linker that has no base moiety bound the Rev protein with wild-type affinity (Pritchard et al., 1994). This evidence, in conjunction with the structure presented here, suggests that the *structure* resulting from the formation of the G-G base pair rather than its sequence is important for binding of Rev to the RRE. One potential structural role for the G-G base pair is to widen the normally narrow major groove of A-form RNA, which has been recognized as a prerequisite for protein major groove recognition (Weeks and Crothers, 1993). From qualitative analysis of the RNA structure, it appears that the kink of the backbone in the internal loop may be reducing the helical twist in this region, resulting in opening up of the major groove.

## Conclusions

Assignments for the $^{1}H$ and $^{13}C$ resonances of the bound form of the RRE RNA were determined from heteronuclear experiments of $^{13}C$-labeled RNA. A three-dimensional model of the RNA was obtained from restrained molecular dynamics, including some model bias for the A-form regions of the molecule. An interesting structural feature in the internal loop was identified, involving a reversal of the phosphate backbone for G71. This structural feature provides an explanation for the observation that isosteric base pairs in combination with a single-nucleotide bulge, which preserve the structure of the internal loop, are the only mutations allowed for the G-G base pair. Even though we do not yet have detailed knowledge of the sequence-specific hydrogen-bonding con-

tacts of the peptide, it appears that the formation of the purine–purine base pairs may play largely a structural role in the specificity of Rev binding.

## Acknowledgements

This work was supported by Grant GM-39589 to J.R.W. and A.D.F. from the National Institutes of Health, and by a grant to J.R.W. from the Searle Scholars Program of the Chicago Community Trust. J.L.B. was supported in part by the NIH Interdepartmental Biotechnology Program (GM-08334). R.T. is a Scholar of the American Foundation for AIDS Research (70405-15-RF). We thank Jody Puglisi for critical comments on the manuscript.

## References

Bartel, D.P., Zapp, M.L., Green, M.R. and Szostak, J.W. (1991) *Cell*, **67**, 529–536.

Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J.D. and Williamson, J.R. (1992) *Nucleic Acids Res.*, **20**, 4515–4523.

Battiste, J.L., Tan, R., Frankel, A.D. and Williamson, J.R. (1994) *Biochemistry*, **33**, 2741–2747.

Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 8172–8184.

Cook, K.S., Fisk, G.J., Hauber, J., Usman, N., Daly, T.J. and Rusche, J.R. (1991) *Nucleic Acids Res.*, **19**, 1577–1582.

Doudna, J., Grosshans, C., Gooding, A. and Kundrot, C.E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7829–7833.

Fischer, U., Meyer, S., Teufel, M., Heckel, C., Luhrmann, R. and Rautmann, G. (1994) *EMBO J.*, **13**, 4105–4112.

Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A.D. (1993) *Nucleic Acids Res.*, **21**, 5509–5516.

Heus, H.A. and Pardi, A. (1991) *Science*, **253**, 191–194.

Hines, J.V., Landry, S.M., Varani, G. and Tinoco Jr., I. (1994) *J. Am. Chem. Soc.*, **116**, 5823–5831.

Iwai, S., Pritchard, C., Mann, D.A., Karn, J. and Gait, M.J. (1992) *Nucleic Acids Res.*, **20**, 6465–6472.

Kessler, H., Griesinger, C., Kerssebaum, R., Wagner, K. and Ernst, R.R. (1987) *J. Am. Chem. Soc.*, **109**, 607–609.

Kjems, J., Brown, M., Chang, D.D. and Sharp, P.A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 683–687.

Kjems, J., Calnan, B.J., Frankel, A.D. and Sharp, P.A. (1992) *EMBO J.*, **11**, 1119–1129.

Le, S.-Y., Pattabiraman, N., Jacob, V. and Maizel, J. (1994) *Nucleic Acids Res.*, **22**, 3966–3976.

Leclerc, F., Cedergren, R. and Ellington, A.D. (1994) *Nature Struct. Biol.*, **1**, 293–300.

Malim, M.H. and Cullen, B.R. (1993) *Mol. Cell. Biol.*, **13**, 6180–6189.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M. and Pardi, A. (1992) *Nucleic Acids Res.*, **20**, 4507–4513.

Nikonowicz, E.P. and Pardi, A. (1993) *J. Mol. Biol.*, **232**, 1141–1156.

Otting, G. and Wüthrich, K. (1989) *J. Magn. Reson.*, **85**, 586–594.

Otting, G. and Wüthrich, K. (1990) *Quart. Rev. Biophys.*, **23**, 39–96.

Peterson, R.D., Bartel, D.P., Szostak, J.W., Horvath, S.J. and Feigon, J. (1994) *Biochemistry*, **33**, 5357–5366.

Pritchard, C.E., Grasby, J.A., Hamy, F., Zacharek, A.M., Singh, M., Karn, J. and Gait, M.J. (1994) *Nucleic Acids Res.*, **22**, 2592–2600.

Ruhl, M., Himmelspach, M., Bahr, G.M., Hammerschmid, F., Jaksche, H., Wolf, B., Aschauer, H., Farrington, G.K., Probst, H., Bevec, D. and Hauber, J. (1993) *J. Cell Biol.*, **123**, 1309–1320.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York, NY.

Santoro, J. and King, G.C. (1992) *J. Magn. Reson.*, **97**, 202–207.

Schwalbe, H., Marino, J.P., King, G.C., Wechselberger, R., Bermel, W. and Griesinger, C. (1994) *J. Biomol. NMR*, **4**, 631–644.

Shaka, A.J., Keeler, J. and Freeman, R. (1983) *J. Magn. Reson.*, **53**, 313–340.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.

Stutz, F. and Rosbash, M. (1994) *EMBO J.*, **13**, 4096–4104.

Szewczak, A.A., Moore, P.B., Chan, Y.-L. and Wool, I.G. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9581–9583.

Tan, R., Chen, L., Buettner, J.A., Hudson, D. and Frankel, A.D. (1993) *Cell*, **73**, 1031–1040.

Tiley, L.S., Malim, M.H., Tewary, H.K., Stockley, P.G. and Cullen, B.R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 758–762.

Weeks, K.M. and Crothers, D.M. (1993) *Science*, **261**, 1574–1577.

Wimberly, B., Varani, G. and Tinoco Jr., I. (1993) *Biochemistry*, **32**, 1078–1087.