ARTICLE

**Fabien Kieken · Eric Arnoult · Florent Barbault
Françoise Paquet · Tam Huynh-Dinh · Jacques Paoletti
Daniel Genest · Gérard Lancelot**

HIV-1$_{Lai}$ genomic RNA: combined used of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin

Abstract The genome of all retroviruses consists of two identical copies of an RNA sequence associated in a non-covalent dimer. A region upstream from the splice donor (SL1) comprising a self-complementary sequence is responsible for the initiation of the dimerization. This region is able to dimerize in two conformations: a loop-loop complex or an extended duplex. Here, we solve by 2D NMR techniques the solution structure of a 23-nucleotide sequence corresponding to HIV-1 SL1$_{Lai}$ in which the mutation G12→A12 is included to prevent dimerization. It is shown that this monomer adopts a stem-loop conformation with a seven base pairs stem and a nine nucleotide loop containing the G10 C11 A12 C13 G14 C15 sequence. The stem is well structured in an A-form duplex, while the loop is more flexible even though elements of structure are evident. We show that the structure adopted by the stem can be appreciably different from its relaxed structure when the adenines A8, A9 and A16 in the loop are mechanically constrained. This point could be important for the efficiency of the dimerization. This experimental study is complemented with a 10 ns molecular dynamics simulation in the presence of counterions and explicit water molecules. This simulation brings about information on the flexibility of the loop, such as a hinge motion between the stem and the loop and a labile lattice of hydrogen bonds in the loop. The bases of the nucleotides G10 to C15 were found outside of the loop during a part of the trajectory, which is certainly necessary to initiate the dimerization process of the genuine SL1$_{Lai}$ sequence.

Introduction

Among all virus families, retroviruses are unique in having a diploid genome (Coffin 1997). Interaction between these two RNA copies occurs near their 5′-end (Bender and Davidson 1976; Bender et al. 1978; Chien et al. 1980; Murti et al. 1981). A region upstream from the splice donor has been identified as being responsible for the initiation of a homodimeric complex in the case of HIV-1$_{Lai}$. This 35-base sequence, called SL1, is highly conserved and can adopt a stem-loop structure with a self-complementary sequence, GCGCGC, in the nine-residue loop. The biological consequences of the genomic RNA dimerization seem to be important because dimerization is strongly associated with different steps of the replication cycle of the retroviruses, like recombination during reverse transcription (Panganiban and Fiore 1988; Hu and Temin 1990; Temin 1991), expression of gag (Bieth et al. 1990; Baudin et al. 1993) and encapsidation (Darlix et al. 1990, 1995). It has been shown that the SL1 sequence dimerizes spontaneously and that this complex can be converted by heating the loop-loop complex at around 55 °C, or in the presence of NCp7 protein at lower temperature, into a more stable dimer (Muriaux et al. 1995, 1996; Laughrea and Jette 1996). It has been proposed that the conformations of the unstable dimer and of the stable dimer involve a loop-loop complex and an extended dimer, respectively. In order to understand the mechanisms which govern the transitions monomer→loop-loop complex and loop-loop complex→extended dimer, it is necessary to determine the structure and dynamic properties of the three proposed conformations. It has been shown that, in vitro, HIV-1 dimer formation was initiated by the annealing of the autocomplementary sequence via a loop-loop interaction between the two RNA molecules, resulting in the formation of an unstable loop-loop

F. Kieken · E. Arnoult · F. Barbault · F. Paquet
J. Paoletti · D. Genest · G. Lancelot (✉)
Centre de Biophysique Moléculaire,
CNRS affiliated to Orléans University and to INSERM,
rue Charles Sadron, 45071 Orléans Cedex 02, France
E-mail: lancelot@cnrs-orleans.fr

T. Huynh-Dinh
Laboratoire de Chimie Organique, Institut Pasteur,
28 rue du Docteur Roux, 75724 Paris Cedex15, France

complex ($T_m \approx 35$ °C) (Muriaux et al. 1995). The structure of this so-called "kissing complex" has been solved by NMR (Mujeeb et al. 1998) and recently by X-ray crystallography (Ennifar et al. 2001) on a 23-nucleotide base sequence. These two structures are completely different. Two structural features of the solution structure of the kissing complex are important for understanding its low stability in solution. First, a fully unwound loop-loop helix and, second, a triple base stacking in which an adenine is stacked between two residues: a cytosine and an adenine. This induces mechanical constraints at the stem-loop junction. The stable dimer conformation involves the whole nucleotide sequence in an intermolecular interaction, leading to a zipper-like motif AA.A (Girard et al. 1999). This motif enables the bases to be positioned closer and allows the formation of an adjacent G-C base pair at the end of the stem. Thermal fusion together with NMR studies (Theilleux-Delalande et al. 2000) showed that, during the transition kissing complex→extended duplex, the loop-loop interaction is maintained while the intramolecular base pairing of the stem is destroyed before rearranging into an intermolecular base pairing within the extended duplex.

It can be pointed out that the putative stem-loop conformation of the monomer was derived from the sequence analysis but has never been observed experimentally. The goal of this study is to present the NMR structure of the 23-nucleotide monomer together with its dynamics properties. Because the wild sequence spontaneously homodimerizes, even at low concentration, we have introduced a single mutation in the SL1 loop, changing the second guanine of the GCGCGC sequence into an adenine. Under these conditions, it is shown that the molecule is mainly in a monomeric form and adopts a stem-loop structure. A 10 ns molecular dynamics simulation in the presence of water and of counterions allows us to investigate its dynamics properties on this time scale.

Materials and methods

Sample preparation

The synthesis of r(CUUGCUGAAGCACGCACGGCAAG) was carried out using phosphoramidite chemistry, as already described (Girard et al. 1999). The mutated adenine is quoted with an asterisk and the sequence will be referred as SL1* in the following. The purity (~100%) was checked with capillary electrophoresis, analytical HPLC and mass spectrometry. For NMR experiments, the sample was passed through a Chelex 100 column in order to remove paramagnetic impurities, concentrated at 2 mM by lyophilization and finally dissolved in $H_2O:D_2O$ (90:10) or in a 100% D_2O argon-degassed solvent.

Dimerization essays of oligoribonucleotides

The same protocol as for SL1 dimerization was used (Girard et al. 1999). Salt-free RNA samples (1–10 μM) were denatured at 90 °C for 3 min in 12 μL of milliQ water, then allowed to refold by snap freezing on ice for 10 min or by slow cooling until 4 °C and dissolved in 3 μL of 50 mM Tris-HCl, 100 mM NaCl buffer at

pH 7.5. The samples (1–10 μM) were then incubated for 15 min at the required temperature and analysed on 12 or 15% polyacrylamide gels with 1× Tris borate ethylenediethylamine buffer. Gels were stained with ethidium bromide and the relative amount of monomer and dimer determined by fluorescent scanning.

NMR experiments and structure calculations

NMR experiments were carried out on an AMX-500 spectrometer operating at 11.74 Tesla and processed on a X32 computer. DQF-COSY, NOESY and TOCSY were recorded and processed, as already described (Girard et al. 1999). The H2 resonances of the adenines were located on the 1D spectra through their long T_1 relaxation time by using a recovery time of 3.8 s and a relaxation delay of 20 s. For experiments in H_2O, a sweep width of 13,000 Hz was used in the F1 and F2 dimensions. A watergate sequence (Piotto et al. 1992) was used in order to remove the water signal. ^{31}P resonances were measured from sodium phosphate at pH 7.0. In D_2O, five NOESY spectra corresponding to 0, 50, 100, 200 and 450 ms mixing times were collected at 27 °C with a recovery delay of 2 s without removing the sample or changing the gain. The NOE values were obtained by dividing the volume of the cross-peaks by the corresponding value of the diagonal obtained at 0 ms mixing time.

Molecular dynamics simulation

A 10 ns simulation in the NPT ensemble was performed with the AMBER package (Case et al. 1997) and PARM98 force field (Cheatham et al. 1999) using the PME summation (Darden et al. 1993) for computing electrostatic forces. The SHAKE algorithm (Ryckaert et al. 1977) was used on all covalent bonds containing a hydrogen atom and the time increment was 2 fs. The RNA molecule was surrounded by 4313 water molecules and 22 Na^+ ions. The starting positions of the counterions were such that they were 6 Å from the phosphorus atoms, and at more than 5 Å from any solute atom. Five hundred steps of energy minimization were first performed (100 steps using the steepest descent algorithm and the others the conjugate gradient method). Water was heated to 300 K by steps of 50 K, for 2 ps each, with a reassignment of velocities every 0.5 ps, followed by a 30 ps thermalization of water at 300 K. During this process the counterions and the solute were kept frozen. At 300 K an additional equilibration of 10 ps was done, during which the counterions were free to move, keeping all solute atoms frozen. Then a 10 ns production period at 300 K, with all atoms free to move, was performed during which the coordinates of counterions and solute atoms were stored every 0.5 ps and those of water every 3 ps.

Analysis of trajectories

The first 200 ps have been discarded for the analysis. In our analysis we have considered that a hydrogen bond (HB) is present when the donor-acceptor distance is smaller than 3.5 Å and the angle donor-H-acceptor is smaller than 45°, except when explicitly stated. Flexibility maps were obtained as follows. The root mean square (RMS) fluctuations of all the interatomic distances in SL1* were calculated and arranged in a two-dimensional array. They were then displayed as black, grey or white on a grid according to their value: smaller than $\sigma 1$, comprised between $\sigma 1$ and $\sigma 2$ or larger than $\sigma 2$ ($\sigma 1 < \sigma 2$), respectively, $\sigma 1$ and $\sigma 2$ being a priori chosen cut-offs. The root-mean-square deviation (RMSD) between two fitted conformations is given by:

$$RMSD = \left\{ (1/N) \sum \langle \Delta r^2 \rangle \right\}^{1/2} \qquad (1)$$

where N is the number of atoms and $\langle \Delta r^2 \rangle$ is the time-averaged square of the position difference of an atom in both conformations, while the sum is performed over all the atoms taken into account.

Results

NMR solution structure of SL1*

Assignment of ^1H resonances and structural information

The H5-H6 connectivities of the cytosines and of the uracils were located on TOCSY and the six H2 resonances of the adenines were found by using their long T_1 relaxation rates by inversion-recovery experiments on 1D spectra. Two of them exhibited strong cross-peaks with imino resonances characteristic of A-U Watson-Crick pairing on NOESY maps (Fig. 1). Moreover, both these H2 resonances are connected together (Fig. 2), indicating they are adjacent in the sequence. It is likely that they belong to A21 and A22 instead of A8 and A9, and under this hypothesis they are hydrogen bonded to U2 and U3. The sequential assignment of exchangeable protons justified this assumption, and allowed us to assign imino resonances of G23, G4, G19 and G18 from imino-imino NOESY cross-peaks (Fig. 1), and therefore the amino, H5 and H6 resonances of C1, C5, C17 and

C20. The assignment showed that Watson-Crick base pairing exists between fragments 1–5 and 19–23 as well as between 7 and 17. Moreover, the strong N1H(G6)-N3H(C18) connectivity demonstrated the formation of the wobble base-pair U6-G18. The imino resonances of G10 and G14 and the amino resonances of C11, C13 and C15 are not observed whatever the temperature and the pH in the range of 5.0–7.0, indicating that these residues are not involved in Watson-Crick interactions.

The sequential assignments of non-exchangeable protons started from H2 of A22 and used the NOESY connectivities between the H1′ of each nucleotide and aromatic protons of its 3′-end neighbour (Fig. 3) and between aromatics of adjacent residues. Uncertainties due to overlapping in the region 7.6–7.8/5.6–5.7 ppm were raised by running spectra at different temperatures. In Table 1 are reported the chemical shifts of the different protons which have been assigned.
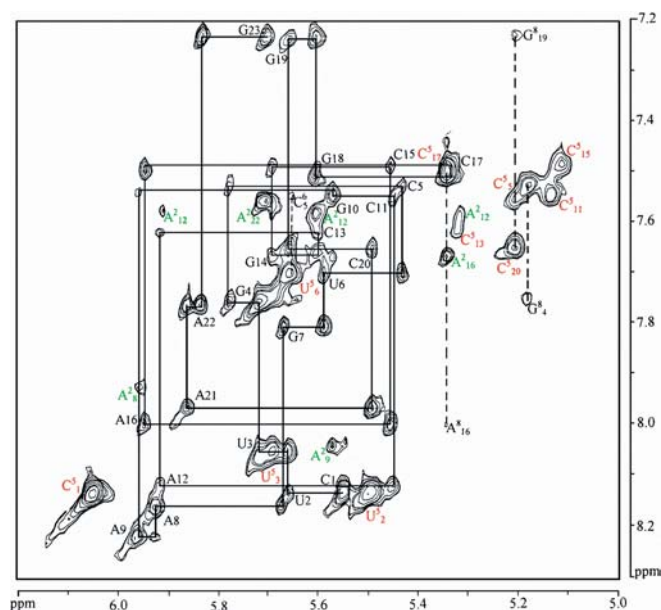
The above results strongly suggest that the SL1* sequence adopts a stem-loop structure in which the seven first bases are paired with the seven last, while residues 8 to 16 are not paired. The observation of continuous aromatic-H1′ inter-nucleotide connectivities gives evidence that the loop is not completely disordered. In addition, connectivities are observed between the aromatic protons of A8, A9 and A16 [medium for H2(A9)-H2(A16), small for H2(A8)-H2(A9) and H2(A8)-H8(A9)]. However, the cross-peaks H5(C11)-H8(G10), H5(C13)-H8(A12) and H5(C15)-H8(G14) were small or absent, indicating either a special relative position of the bases in the loop or an important

Fig. 1a, b Portion of the 500 MHz NOESY spectrum for the 23-mer r(C1 U2 U3 G4 C5 U6 G7 A8 A9 G10 C11 A12 C13 G14 C15 A16 C17 G18 G19 C20 A21 A22 G23) in $H_2O:D_2O$ (90:10) solution at 9 °C. **a** Connectivities between imino protons for 150 ms mixing time and scheme of the stem-loop. **b** Connectivities between imino protons and aromatics or amino protons for 120 ms mixing time

a

b
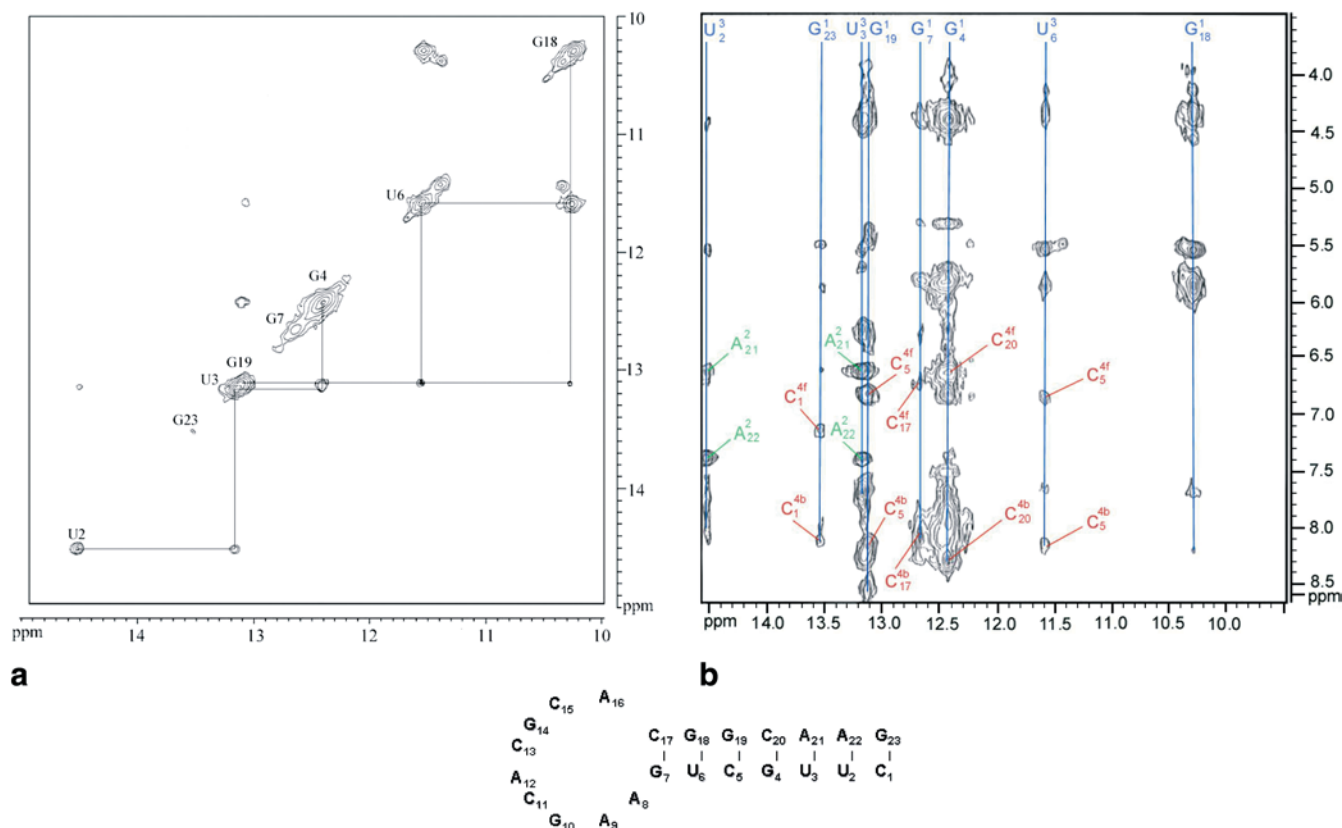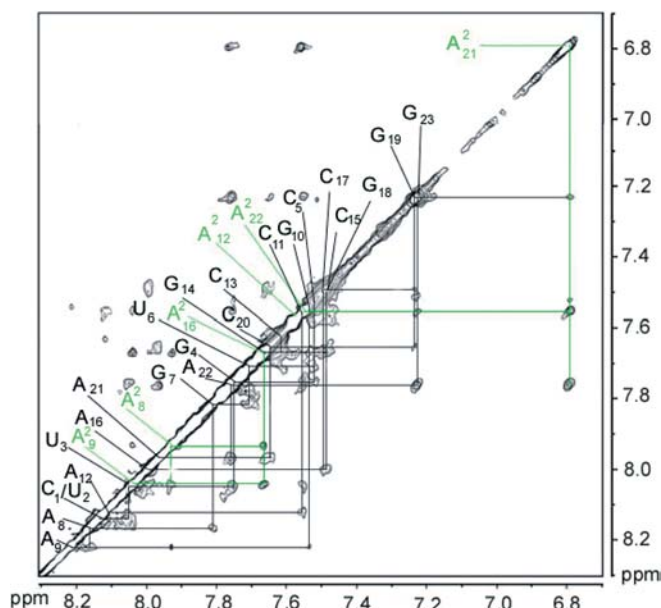
Fig. 2 Expansion of the aromatic to aromatic region of the NOESY spectrum of the 23-mer A12 mutated stem-loop in D_2O at 27 °C for 250 ms mixing time. The *grey lines* show the H2 resonances. The H2(A8)-H2(A16) and H2(A9)-H2(A16) connectivities are clearly seen

Fig. 3 Expansion of the aromatic to the H1′ region of the NOESY spectrum of the 23-mer A12 mutated stem-loop in D_2O at 27 °C for 450 ms mixing time. The H2 resonances are labeled in *grey*

Table 1 Assigned chemical shifts of the A12 mutated 23-mer SL1 stem-loop. Exchangeable protons were assigned at 9 °C at pH 5 in $H_2O:D_2O$ (90:10) and non-exchangeable protons at 27 °C in D_2O, pH 7.0

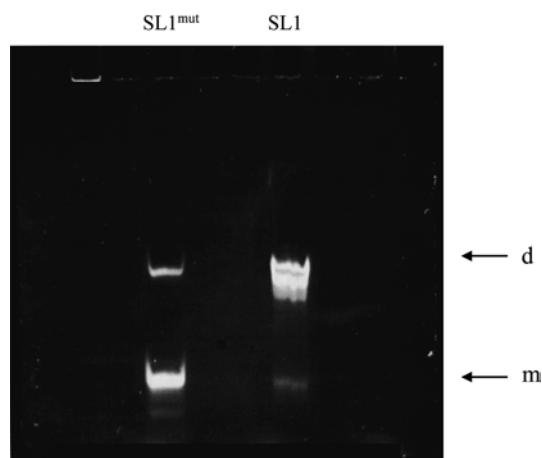|	NH	NH$_2$	H6/H8	H5	H2	H1′	H2′
C1		7.18/8.14	8.14	6.04		5.55	4.47
U2	14.53		8.14	5.50		5.65	4.53
U3	13.17		8.06	5.69		5.71	4.58
G4	12.44	7.84/7.94	7.76			5.77	4.47
C5		6.86/8.19	7.53	5.18		5.43	4.55
U6	11.59		7.70	5.65		5.58	4.22
G7	12.67	–[a]	7.81			5.66	4.30
A8		–[a]	8.17		7.93	5.92	4.64
A9		–[a]	8.23		8.05	5.96	4.58
G10		–[a]	7.54			5.57	4.64
C11		–[a]	7.55	5.14		5.45	4.46
A12		–[a]	8.13		7.59	5.91	4.41
C13		–[a]	7.62	5.32		5.60	4.27
G14		–[a]	7.67			5.69	4.58
C15		–[a]	7.49	5.11		5.45	4.52
A16		–[a]	8.00		7.67	5.94	4.53
C17		6.74/8.08	7.49	5.34		5.31	4.44
G18	10.30	–[a]	7.51			5.60	4.68
G19	13.11	–[a]	7.24			5.66	4.45
C20		6.85/8.32	7.66	5.21		5.49	4.51
A21		7.72/6.27	7.97		6.80	5.86	4.58
A22			7.76		7.56	5.83	4.41
G23	13.54	–[a]	7.23			5.70	3.93

[a] Not assigned

^{31}P NMR

Most of the ^{31}P resonances were concentrated within a narrow chemical shift range from –1.4 to –2.0 ppm, which are standard values for regular helices (Legault and Pardi 1994). Only one resonance was downfield shifted at –0.5 ppm. Therefore it seems that there are not too large differences between the backbone conformation of all the different nucleotides except one. Unfortunately, hetero TOCSY experiments did not allow the assignment of this particular resonance.

SL1* is monomeric

Two 10 µM of SL1 and SL1* respectively were heat denatured at 90 °C for 3 min and snap frozen on ice, then incubated at 4 °C for 15 min and run on gel electrophoresis at 4 °C (Fig. 4). The relative amount of dimer was more than 90% for the wild sequence and less than 15% for the mutated one. Additional gels performed at 37 °C indicated that the small amount of SL1* dimer adopts an extended conformation.

At higher concentration (2 mM), the lack of complementary base pairing for the central bases, as evidenced by NMR, is also an indication that SL1* is mainly in a monomeric form. This is confirmed by fitting the build-up of the H5-H6 NOE connectivities of C5 and C20 by using the RELAZ program (Lancelot et al. 1989). A global correlation time of 3 ns is obtained, which is consistent with molecular species of about 7000 daltons (SL1* molecular weight), while it was

dynamics in its central part. COSY revealed no connectivity between H1′ and H2′, implying that all the sugar rings adopted a North conformation, except for A8, A9 and G23. Furthermore, the H1′(n–1)-H8/H6(n) NOESY cross-peaks of the stem were small while the H2′(n)-H6/H8(n–1) ones are large. All these data show that the stem is in an A-form duplex.

Fig. 4 Comparison of dimer formation of wild-type 23-mer SL1 (*right*) and A12 mutated 23-mer SL1 (*left*); *d* shows the dimer and *m* the monomer. Heat denatured RNA samples were incubated for 15 min at 4 °C and run at 4 °C (see Materials and methods)

shown that the dimeric form of the wild sequence gave a correlation time of 6 ns under the same conditions (Girard et al. 1999).

Experimental NMR constraints

Part of the interproton distances were first estimated from volume integration of the NOESY cross peaks in D_2O at 50, 90 and 120 ms mixing times. The volumes were converted into distances by assuming a 6th-power dependence and taking the H5-H6 distances of cytosine (2.46 Å) as reference. A tolerance of 15% was assigned to the distances involving protons belonging to the stem (1–7 and 17–23 nucleotides). The possible local difference in correlation time and the possible chemical exchange – even if it was not detected in the spectra – could alter the distances computed from the NOEs for protons belonging to nucleotides owning high flexibility. For this reason, the distance restraints involving the nucleotides 8 to 16 were taken as: $0.5d < r < 1.3d$, where d is the

distance computed from NOE. When $0.5d$ was less than 1.8 Å, this last value was used. For distances involving exchangeable protons for which NOESY was performed in H_2O, only an upper limit distance was used: 3.5, 5 and 6 Å for strong, medium and weak cross-peaks, respectively. Finally, the distances corresponding to cross-peaks seen only for NOESY experiments at 300 ms mixing time are assigned to be in the range 4–7 Å. A 3′-*endo* conformation was assigned to all ribose rings, except A8, A9 and G23, by restraining the endocyclic torsion angles. The other five backbone dihedral angles were maintained at their A-form value within a window of $\pm 15°$ for the nucleotides of the stem and of $\pm 45°$ for the nucleotides of the loop, except A8, A9 and A16.

A total of 249 distances and 155 angles was used as constraints in the determination of the SL1* structure (see Table 2 for details). This corresponds to about 17.5 constraints per nucleotide.

Structure refinement

Two hundred random structures were generated with XPLOR (v. 3.851) (Brünger 1992) and submitted to a simulated annealing restrained molecular dynamics (MD) protocol in vacuum. The force field consisted of bond lengths, bond angles, dihedral angles, improper angles, single quadratic van der Waals repulsion and experimental constraints. One thousand steps of Powell minimization (Powell 1977) to remove possible atomic overlap were first performed, followed by 10 ps of equilibration by MD simulation at 1500 K with all atomic masses equally set to 10 a.u. During an additional 2 ps simulation at the same temperature, the physical terms of the force field were gradually increased to their maximum values, while the constrained terms were kept constant. Then during 20 ps the temperature was progressively decreased to 300 K. The protocol was ended with 1000 steps of Powell minimization. At this point, 37 structures exhibiting both a low potential energy and a low violation of the interproton distances

Table 2 NMR restraints and statistics for the A12 mutated 23-mer SL1 stem-loop

A. Distance and dihedral constraints	
Intra-residue distance restraints	50
Inter-residue distance restraints	163
Hydrogen bonding distance restraints	36
Total distance restraints	249
Torsion angle restraints for sugar puckers	46
Backbone torsion angle restraints	109
Total angle restraints	155
Total restraints	404
B. Precision of lowest-energy structures	
Number of structures	19
Maximum NMR distance restraint violations (up to 0.3 Å) in structures	0
RMSD[a]	
Stem (bases C1-G7 and C17-G23)	0.15 ± 0.08 Å
Loop (bases A8-A16)	1.02 ± 0.39 Å
Whole molecule	1.39 ± 0.70 Å

[a]RMSD: root-mean-square deviation ± standard deviation in position of heavy atoms compared to the global average structure

were selected and submitted to 10,000 steps of Powell energy minimization with full electrostatic and van der Waals terms in the force field. Superimposition of these 37 structures gives a family of hairpins with a great variability of the loop conformation (Fig. 5).

The 37 selected structures were then averaged and energy minimized. The resulting structure was used later as the initial conformation for an unrestrained MD simulation in the presence of water and counterions.

Calculated chemical shifts

The set of converged structures was submitted to the NUCHEMICS software (Wijmenga et al. 1997) in order to calculate the proton chemical shifts of H8, H6, H5, H2 and H1′. H2 and H5 have chemical shifts very sensitive to the conformation, because they are close to a five- or six-membered ring, and small displacements can induce a large variation of their chemical shifts (Giessner-Prettre et al. 1976). For all the protons of the stem and for all the 37 structures, a good agreement is obtained between calculated and experimental chemical shifts (less than 0.3 ppm difference). On the contrary, none of the structures gives simultaneously a good fit for all the H5 and H2 resonances of the loop, although the fit could be reached for each individual resonance depending on the structure. This is again a consequence of the flexibility of the loop. Nevertheless, as experimentally found, it appears that for many structures the calculated H5 chemical shifts of C11, C13 and C15 are upfield shifted relatively to the H5 resonance of a free cytosine (5.1–5.3 ppm for SL1* compared to 6.13 ppm for the free cytosine). This is in agreement with a stacking of these cytosines.

Nineteen of these structures gave a good agreement between calculated and experimental H2 chemical shifts of A8, A9 and A16 (less than 0.3 ppm difference). Their

PDB accession number is 1JTJ. For these structures, the plane of A16 is roughly parallel to the G7-C17 base pair plane and A8, A9 and G10 are stacked with their plane approximately perpendicular to the plane of A16 (Fig. 6).

Molecular dynamics

RMS difference

Figure 7 shows the time evolution of the RMSD between the simulated and the initial structures. For the whole molecule or for the loop, an equilibrium structure is not reached, even after 10 ns of simulation, due to conformational transitions in the loop between metastable structures. On the other hand, a stationary behaviour is rapidly reached for the stem with an average value of 2.8 Å. Therefore the stem converges to a stable conformation. The program CURVES (Lavery and Sklenar 1989) shows a curvature of the stem during the simulation, while it is straight in the initial structure, explaining the rather high value of the RMSD. Displaying the helix axis of the stem at different times of the simulation shows that the bending corresponds mainly to a kink at each extremity. Therefore skipping the first and last base pair of the stem decreases the RMSD of the stem to 1.7 Å.

Hydrogen bonds

A complex fluctuating network of transient hydrogen bonds (HBs) is observed in the loop during the time course of the simulation, involving both base and backbone atoms. Only N4H(C15)→O2P(G14) seems to be sufficiently stable to remain present all along the trajectory. This HB appears also in the NMR structure.
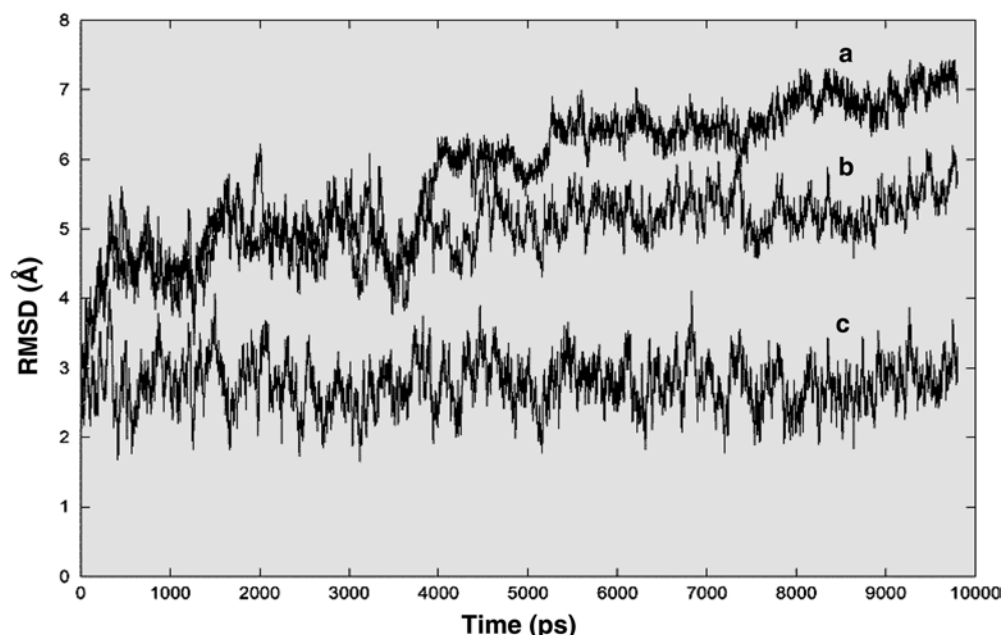

Fig. 5 Superposition of 19 structures of SL1*
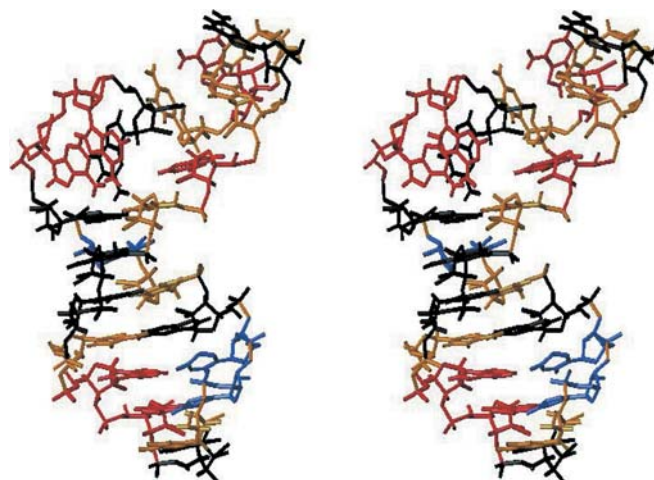
Fig. 6 Stereoscopic view of the mutated SL1 stem-loop. The stacking of A8, A9 and G10 can easily be observed

Fig. 7 RMSD time series of the loop (*a*), the whole molecule (*b*) and the stem (*c*). The initial NMR structure is taken as reference

The O2′ atoms of each residue act often as donors in two types of intramolecular HB. First, O2′H(i)→O5′($i+1$) with various occupancy percentages according to the residue are observed. Second, an O2′H(i)→O4′($i+1$) HB type, less represented than the previous one, is observed during a large fraction of time for some nucleotides ($i=11$ and 12). For each nucleotide, the existence of these bonds between nucleotide pairs fluctuates during the simulation.

For the five terminal base pairs of the stem, the usual Watson-Crick HBs are present during more than 90% of the time. The non-canonical base pair U6-G18 is strongly maintained by two HBs (N1H→O2 and N3H→O6) which are present at more than 97% of the time. The G7-C17 pair located at the stem-loop junction exhibits the three Watson-Crick HBs, but one of them (N4H→O6) is weaker than the others according to the following criteria: (1) it is only present during 80% of the time while both others are present at more than 94%, (2) the maximum time between two successive disruptions is shorter than 70 ps, while it is more than 200 ps for the others, and (3) if the definition of a HB is more restricted (D-H···A = 20° instead of 45°), its percentage of existence decreases to 50% while for both the others it remains higher than 85%. The O2′H of C17 is involved in a HB with either O4′ or O5′ of G18 during all the simulation, with an occupancy ratio larger than 90%. This is also observed in the NMR structure. On the other hand, the O2′H of the other stem residues are mainly involved in HBs with water molecules.

The hydration pattern of each residue is in full agreement with already published data for RNA sequences (Auffinger and Westhof 1997). The bases of the loop are more involved in HBs with water than those of the stem as expected. A particular behaviour is found at the wobble U6-G18 base pair: (1) the average number of HBs with solvent is twice for N2(G18) than for the other guanines of the stem and (2) O2(U6) and O4(U6) are involved in 0.32 and 1.80 HBs, respectively, on average, while for the other uracils of the stem these proportions become 1.08 and 0.86, respectively. On the average, 30% of the water molecules hydrogen bonded to the solute bridge two atoms belonging to SL1*.

Flexibility

An example of a flexibility map for SL1* is shown in Fig. 8 for $\sigma 1 = 0.4$ Å and $\sigma 2 = 0.8$ Å (see Materials and methods section). Each strand of the stem is globally rigid, and the set of atoms belonging to both strands as a whole is also a significant rigid group. On the other hand, the loop appears to be much more flexible and the map shows also important flexibility between the stem and the loop. In Table 3 are reported the mean fluctuation of interatomic distances for the different regions of the map. The average fluctuations of interatomic distances are more than three times larger for the loop than for the stem. Residues belonging to the stem are submitted to highly correlated motions, whereas the motions of any nucleotide of the loop are much less coupled to those of other nucleotides. In addition, the RMS fluctuations of the glycosidic and backbone dihedral angles are larger in the loop than in the stem by a factor ranging between 1.5 and 2.5. On the other hand, the fluctuations of the pucker angles are not significantly different in the stem and in the loop.

Many conformational transitions between metastable states are observed during the simulation, affecting the global shape of the loop backbone and the orientation of
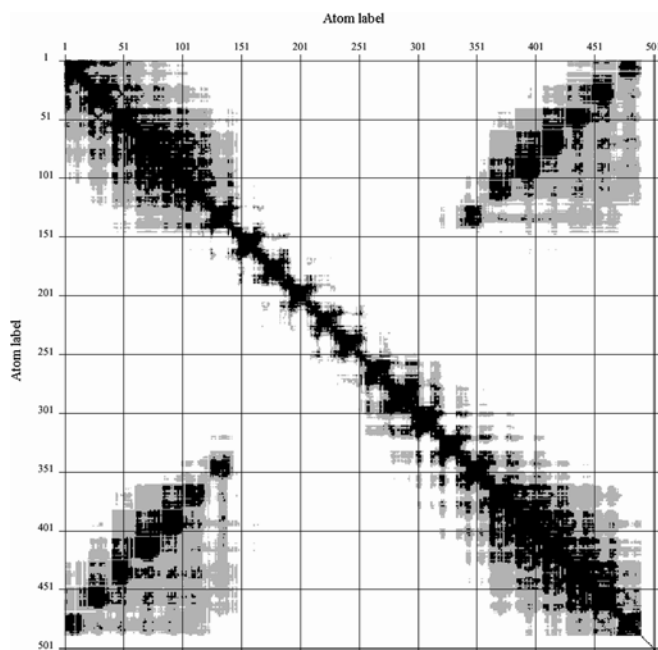
Fig. 8 Flexibility map of the RNA stem-loop for $\sigma1 = 0.4$ Å and $\sigma2 = 0.8$ Å (see Materials and methods section for details). First strand of the stem corresponds to atoms 1–143, loop to atoms 144–337 and second strand of the stem to atoms 338–490. *Black*, *grey* and *white* mean rigid, semi-rigid and flexible, respectively. Hydrogen atoms are not taken into account

Table 3 Root-mean-square (RMS) fluctuations of interatomic distances in the different subunits of the stem-loop. For subunit names with a slash (/), distances are between an atom belonging to the first part of the name and an atom belonging to the second part

Subunit	RMS (Å)
Strand 1	0.62
Strand 2	0.63
Whole stem	0.70
Loop	2.26
Strand 1/loop	2.48
Strand 2/loop	2.16
Strand 1/strand 2	0.78
Whole molecule	1.72

the bases, modulated by rearrangements of an intramolecular HB network.

Stem/loop orientation

In order to monitor the general orientation of the stem relative to the loop, the geometric centers of the three groups of atoms defined by (1) the six terminal base pairs (residues 1–6 and 18–23), (2) the nucleotides of the loop (residues 8–16) and (3) the junction base pair (residues G7 and C17), respectively, have been considered. More precisely, we have calculated the time series of the angle stem-junction-loop (not shown). Large slow fluctuations of this angle between 95° and 145° are observed, flanking the values 116° and 132° calculated

from the NMR structures of the kissing complex (Mujeeb et al. 1998) and of the monomer (this work), respectively.

Calculated chemical shifts

Chemical shifts of aromatic protons were calculated with NUCHEMICS (Wijmenga et al. 1997) for each conformation produced by the simulation and averaged over the trajectory. For all the protons of the stem we found a good agreement with experiment, the differences being less than 0.3 ppm. For 11 of the 16 aromatic protons of the loop there is also a good agreement but for five of them, namely H8(A8), H5(C11), H5(C13), H5(C15) and H6(C15), differences higher than 0.6 ppm are observed. This has to be related to the large variability of the base orientation in the loop, which is not well sampled by a 10 ns simulation. For example, C11 is stacked with A12 for about the first half of the simulation time, but remains unstacked during the second half. On the contrary, C13 is always stacked with G14.

Discussion and conclusion

Until now, the SL1* stem-loop is, among the known NMR solution structures of RNA stem-loop sequences, certainly one of those containing the longest loop. This stem-loop comprises a seven base pairs double helix in the A-form, including the wobble U6-G18 for the stem, while the nine central residues are unpaired. This long unpaired loop exhibits obvious elements of structure, as shown by the continuous NOE pathway between ribose and the aromatic protons of successive residues. Moreover, NOESY cross-peaks are observed between the two ends of the loop, connecting the aromatic protons of A8, A9 and A16. Nevertheless, the lack or the weakness of the H5(n)-H8(n–1) NOE connectivities between the cytosines and the guanines of the loop implies that this interaction is subject to important dynamics. As a consequence, modelling using distance and angular NMR restraints leads to 19 almost identical structures within the stem, but with large differences within the loop. The riboses of the loop are all in the North conformation, except A8, A9 and G23.

The RMSD of the 19 refined structures was 0.23 ± 0.10 Å for the stem and 0.89 ± 0.57 Å for the loop (Table 2). This large difference in the RMSD values provides evidence for the high flexibility of the loop. In order to verify that this difference in dynamics is not a consequence of the difference in restraint distance tolerances between the stem and the loop, another structure refinement was made by using the same tolerance for the stem and for the loop (15%). In this case, although a slightly higher RMSD value (1.01 ± 0.26 Å) was obtained for the loop, and a higher value for the stem (0.56 ± 0.17 Å), a significant difference between both values is still observed. This indicates that the large

dynamics of the loop is not a consequence of the tolerance in the distance restraints, even if its amplitude is a function of this one. This is also indicative that too strong restraints on the proton distances in the loop decreased the structural precision of the stem. This point has been explored in another restrained MD, where only the restrained H-H distances involving the nucleotides G10 to C15 were used as upper limits because NOEs were observed between the adenines A9 and A16 belonging to the beginning and to the end of the loop, the other restraint distance tolerances being equal to 15%. In this case, a higher RMSD value (0.34 ± 0.16 Å) was obtained for the stem. We conclude that the constraint tolerances on A8, A9 and A16 act upon the precision structure of the stem. These data reflect the fact that the relaxed structure of the stem is not compatible with any structure which can be generated by mechanical constraints of adenines A8, A9 and A16. This point could be important for the efficiency of the dimerization. As a matter of fact, it has already been reported that the mechanical constraints imposed by the kissing complex formation disrupt the G7-C17 base pair (Mujeeb et al. 1998). The dynamics of the loops in our solution structures has been evaluated by fitting the stems of the different refined structures –and not their loops – before the RMSD calculation of the loops. A value of 2.62 ± 1.37 Å was found, indicating large variations in the position of the loop in relation to the stem. On the other hand, the RMSD value of 0.89 ± 0.57 Å obtained by fitting the loops indicated that no large variation in the relative position of the nucleotides occurred in the loops.

Owing to the differences in loop and stem flexibility, we undertook a 10 ns MD simulation of SL1* in the presence of explicit water and counterions. The main features of the NMR results are confirmed by the MD simulation: a well-structured A-form stem and a flexible loop. The Watson-Crick base pairing G7-C17 at the stem-loop junction, which was detected by NMR, was found to be stable in the simulation, although one of the HBs was weakened. This has to be compared to data observed in the kissing complex. In the case of the genuine SL1$_{Lai}$ loop (subtype B: GCGCGC self-complementary hexanucleotide), G7 and C17 are reported unpaired in a NMR-derived model (Mujeeb et al. 1998), while in the case of the isolat Mal (subtype A: GUG-CAC self-complementary hexanucleotide) they are paired, as shown by the reactivity of C17 toward DMS/hydrazine treatment (Lodmell et al. 2001).

The 19 selected structures have their computed chemical shifts in agreement with the experimental ones, except for the H5 of cytosines in the loop whose differences in chemical shift exhibit great variability. Thus it is difficult to report an accurate orientation of the bases in the GCACGC sequence; if it exists, the bases are certainly directed outward from the loop, as indicated by the lack of H5(n)-H8(n–1) connectivities. It is likely that some stacking occurs in this sequence, as shown by the upfield shifts of the H5 of the cytosines. MD simulation

brings about different explanations for the flexibility of the loop, which is responsible for the limited number of NOESY cross-peaks. A first reason comes from the large fluctuations of the orientation of the loop relative to the main direction of the stem, although the loop is always directed toward the same side of the stem. A second reason is the internal deformation of the loop, related to labile intramolecular hydrogen bonds due to the absence of Watson-Crick pairing, leading to jumps between metastable conformations. During the time course of the simulation it appeared that the bases A8 and A9 are pointing sometimes toward the outside of the loop (accessible to solvent) or sometimes inside the loop. In this last case, non-canonical HBs between A8 and A16 could be observed, as has already been reported (Jossinet et al. 1999). Some of these large deformations are slow processes ranging in the nanosecond time scale, so that a 10 ns simulation is probably not sufficient to properly sample the whole dynamics of the loop. Nevertheless, a previous analysis of this simulation in terms of NMR relaxation parameters (Fausti et al. 2001) showed that the NOESY H1′-H6/H8 cross-peaks calculated from the trajectory are in fair agreement with the experimental ones, suggesting that the 10 ns trajectory contains information on the major part of the internal dynamics of SL1*.

It is demonstrated that SL1* is mainly monomeric at low concentration as well as at high concentration in the conditions of the NMR experiments. Under the same experimental conditions it was shown previously that a genuine SL1 sequence is dimeric (Girard et al. 1999). Therefore, the mutation from a G12 into an adenine disrupts the stability of the pairing between the six nucleotides self-complementary sequence, which is now GCACGC. This is consistent with the results of Clever et al (1996), who reported that longer synthetic HIV-1 RNA sequences containing the same mutation in the palindrome do not dimerize.

Sequence analysis of various stem-loop sequences able to dimerize indicated a high preference for G and C residues at a central position (12 and 13) in the loop. Substitution of the G12 and C13 nucleotides by G12 and U13 increased the dissociation constant by a factor greater than 10^5 (Lodmell et al. 2001). Taking into account that G is potentially able to form a wobble pair with U, the difference in energy involved between a GC and a GU base pair is not sufficient to explain such a decrease in stability. Another source of loss in energy must be involved. The sequence of the stem is important for kissing complex formation. The dimer formed by loop-loop interaction is less stable than the extended duplex (Theilleux-Delalande et al. 2000), but high mutation in the stem stabilizes the kissing complex formation (Mujeeb et al. 1998). Nevertheless, this mutation changes the subtle equilibrium between the loop-loop formation and the stem formation by breaking the last G7-C7 base pair in the stem (Mujeeb et al. 1998), while this base pair has been detected in the Lai isolat (Theilleux-Delalande et al. 2000). Loop-loop complex

formation requires high structural compatibility between the stem conformation and the loop conformation. For example, conservation of the flanking purine nucleotides 8, 9 and 16 is required for stable dimerization (Paillard et al. 1997). Interstrand interactions between these residues are involved in both the kissing complex (Mujeeb et al. 1998) and the extended duplex (Girard et al. 1999). Consequently, it is easy to deduce that the substitution of G12 by A12, which prevents the formation of two central base pairs over the six possible ones in the wild type, deeply modifies the conformation of the putative loop-loop complex and requires special interaction of the flanking adenines to allow kissing complex formation. In this case, the lack of dimer formation is likely reinforced by the impossibility for the flanking A8, A9 and A16 to stabilize such a structure. These considerations are in agreement with our dynamics studies, since among the metastable conformations observed inside the loop, a number of them are not favourable to the formation of a kissing complex owing to the orientation of several bases in the loop, which are not in a position to establish Watson-Crick base pairs. On the contrary, in other conformations, most of the bases of the G10 to C15 sequence are oriented outside the loop and this is necessary for the dimerization process. This dynamic property is certainly a requisite to initiate the formation of the kissing complex in the wild type.

References

Auffinger P, Westhof E (1997) RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA^{Asp} anticodon hairpin. J Mol Biol 269:326–341

Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C (1993) Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA from defined structural domains. J Mol Biol 229:382–397

Bender W, Davidson N (1976) Mapping of polyA sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. Cell 7:595–607

Bender W, Chien YH, Chattopadhyay S, Vogt PK, Gardner MB, Davidson N (1978) High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. J Virol 25:888–896

Bieth E, Gabus C, Darlix JL (1990) A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis in vitro. Nucleic Acids Res 18:119–127

Brünger AT (1992) X-PLOR: a system for X-ray crystallography and NMR, version 3.1. Yale University Press, New Haven

Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, Weiner PK, Kollman PA (1997) AMBER. University of California, San Francisco

Cheatham TE III, Cieplak P, Kollman PA (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J Biol Struct Dyn 16:845–862

Chien YH, Junghans RP, Davidson N (1980) Molecular biology of RNA tumor viruses. In: Stephenson JR (ed). Academic Press, New York, pp 395–446

Clever JL, Wong ML, Parslow TG (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. J Virol 70:5902–5908

Coffin J (1997) Retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J (eds). Cold Spring Harbor, New York, pp 261–368

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092

Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barre-Sinoussi F (1990) *Cis* elements and *trans*-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. J Mol Biol 216:689–699

Darlix JL, Lapadat-Tapolsky M, Rocquigny H, Roques B (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. J Mol Biol 254:523–537

Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. Nat Struct Biol 8:1064–1068

Fausti S, La Penna G, Paoletti J, Genest D, Lancelot G, Perico A (2001) Modeling the dynamics of a mutated stem-loop in the SL1 domain of HIV-1$_{Lai}$ genomic RNA by ^{1}H-NOESY spectra. J Biomol NMR 20:333–349

Giessner-Prettre C, Pullman B, Borer PN, Kan LS, T'so PO (1976) Ring-current effects in the NMR of nucleic acids: a graphical approach. Biopolymers 15:2277–2286

Girard F, Barbault F, Gouyette C, Huynh-Dinh T, Paoletti J, Lancelot G (1999) Dimer initiation sequence of HIV-1$_{Lai}$ genomic RNA: NMR solution structure of the extended duplex. J Biomol Struct Dyn 16:1145–1157

Hu WS, Temin HM (1990) Retroviral recombination and reverse transcription. Science 250:1227–1233

Jossinet F, Paillart JC, Westhof E, Hermann T, Skripkin E, Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. RNA 5:1222–1234

Lancelot G, Guesnet JL, Vovelle F (1989) Solution structure of the parallel-stranded duplex oligonucleotide alpha-d(TCTAAAC)-beta-d(AGATTTG) via complete relaxation matrix analysis of the NOE effects and molecular mechanics calculations. Biochemistry 28:7871–7878

Laughrea M, Jette L (1996) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. Biochemistry 35:1589–1598

Lavery R, Sklenar H (1989) Defining the structure of irregular nucleic acids: conventions and principles. J Biomol Struct Dyn 6:655–667

Legault P, Pardi A (1994) ^{31}P chemical shift as a probe of structural motifs in RNA. J Magn Reson 103:82–86

Lodmell JS, Ehresmann C, Ehresmann B, Marquet L (2001) Structure and dimerization of HIV-1 kissing loop aptamers. J Mol Biol 311:475–490

Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. Nat Struct Biol 5:432–436

Muriaux D, Girard PM, Bonnet-Mathoniere B, Paoletti J (1995) Dimerization of HIV-1$_{Lai}$ RNA at low ionic strength. An autocomplementary sequence in the 5′ leader region is evidenced by an antisense oligonucleotide. J Biol Chem 270:8209–8216

Muriaux D, Fossé P, Paoletti J (1996) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. Biochemistry 35:5075–5082

Murti KG, Bondurant M, Tereba A (1981) Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. J Virol 37:411–419

Paillard JC, Westhof E, Ehresmann C, Ehresmann B, Marquet R (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. J Mol Biol 270:36–49

Panganiban AT, Fiore D (1988) Ordered interstrand and intra-strand DNA transfer during reverse transcription. Science 241:1064–1069

Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR 2:661–665

Powell MJD (1977) Restart procedures for the conjugate gradient method. Math Program 12:241–254

Ryckaert J P, Ciccotti G, Berendsen H J C (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Temin HM (1991) Sex and recombination in retroviruses. Trends Genetics 7:71–74

Theilleux-Delalande V, Girard F, Huynh-Dinh T, Lancelot G, Paoletti J (2000) The HIV-1$_{Lai}$ RNA dimerization: thermodynamic parameters associated with the transition from the kissing complex to the extended dimer. Eur J Biochem 267: 2711–2719

Wijmenga SS, Kruithof M, Hilbers CW (1997) Analysis of ^{1}H chemical shifts in DNA. Assessment of the reliability of ^{1}H chemical calculations for use in structure refinement. J Biomol NMR 10:337–350