Solution RNA Structures of the HIV-1 Dimerization Initiation Site in the Kissing-Loop and Extended-Duplex Dimers

Seiki Baba[1], Ken-ichi Takahashi[1,2], Satoko Noguchi[1], Hiroshi Takaku[1], Yoshio Koyanagi[3], Naoki Yamamoto[4] and Gota Kawai[1,*]

[1]*Department of Life and Environmental Sciences, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016;* [2]*Department of Bioscience, Faculty of Bioscience, Nagahama Institute of Bio-Science and Technology, 1266 Tamura-cho, Nagahama, Shiga 526-0829;* [3]*Institute for Virus Research, Kyoto University, Kyoto 606-8507; and* [4]*AIDS Research Center, The National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640*

Dimer formation of HIV-1 genomic RNA through its dimerization initiation site (DIS) is crucial to maintaining infectivity. Two types of dimers, the initially generated kissing-loop dimer and the subsequent product of the extended-duplex dimer, are formed in the stem-bulge-stem region with a loop including a self-complementary sequence. NMR chemical shift analysis of a 39mer RNA corresponding to DIS, DIS39, in the kissing-loop and extended-duplex dimers revealed that the three dimensional structures of the stem-bulge-stem region are extremely similar between the two types of dimers. Therefore, we designed two shorter RNA molecules, loop25 and bulge34, corresponding to the loop-stem region and the stem-bulge-stem region of DIS39, respectively. Based upon the chemical shift analysis, the conformation of the loop region of loop25 is identical to that of DIS39 for each of the two types of dimers. The conformation of bulge34 was also found to be the same as that of the corresponding region of DIS39. Thus, we determined the solution structures of loop25 in each of the two types of dimers as well as that of bulge34. Finally, the solution structures of DIS39 in the kissing-loop and extended-duplex dimers were determined by combining the parts of the structures. The solution structures of the two types of dimers were similar to each other in general with a difference found only in the A16 residue. The elucidation of the structures of DIS39 is important to understanding the molecular mechanism of the conformational dynamics of viral RNA molecules.

Key words: DIS, HIV-1, NMR, RNA, structure.

Abbreviations: DIS, dimerization initiation site; HIV-1, human immunodeficiency virus type 1.

Two molecules of viral genomic RNA are packaged in a dimeric state in the virion of human immunodeficiency virus type 1 (HIV-1), and this dimer formation is crucial to maintaining their infectivity (*1–4*). Accumulating evidence from both *in vivo* and *in vitro* experiments has shown that the specific sequence, the dimerization initiation site (DIS) located close to the 5′ terminus of the genomic RNA, is required for spontaneous dimerization of HIV-1 RNA. DIS can form a stem-loop structure with a self-complementary sequence in the loop and a bulge in the stem (*5, 6*). The dimerization of DIS forms the kissing-loop dimer as the first step; then, their intramolecular stems are converted into intermolecular stems, generating the extended-duplex dimer (*7, 8*). This two step dimerization process is called the kissing-loop mechanism. The kissing-loop dimer is converted into the extended-duplex dimer by incubation at 55°C (*9, 10*) or by incubation at physiological temperature with the HIV-1 nucleocapsid protein, NCp7, which includes two basic regions and two zinc-fingers (*11*). A number of experiments have been performed to gain an understanding of the role of the zinc-fingers as well as the basic regions (*12–16*). Our previous

results show that, for the two step dimerization from the kissing-loop dimer to the extended-duplex dimer, the two basic regions surrounding the N-terminal zinc finger of NCp7 have RNA chaperone activity by themselves, and the zinc fingers increase the efficiency of the activity (*17, 18*).

A number of three dimensional structural analyses using NMR and X-ray methods have been performed to determine the conformation of each region of DIS, the loop region in the kissing-loop (*19, 20*) or extended-duplex dimers (*21–24*), as well as the bulge-out region (*25–27*). However, our previous studies suggested that the 39mer RNA sequence, DIS39, which covers the entire bulge and loop regions, is necessary and sufficient for the two step dimerization (*28, 29*). Thus, it is still relevant to determine the structures of the kissing-loop and extended-duplex dimers for DIS39 with the same sequence and conditions.

In the present study, we designed two shorter RNA molecules, loop25 and bulge34; loop25 includes the loop-stem region of DIS39, and bulge34 includes the stem-bulge-stem region (Fig. 1), respectively we then determined the solution structures of loop25 in each of the kissing-loop and extended-duplex dimers as well as bulge34. By combining the structure parts, the solution structures of DIS39 in the kissing-loop and extended-duplex dimers were able to be determined.

---

*To whom correspondence should be addressed. Tel/Fax: +81-47-478-0425, E-mail: gkawai@sea.it-chiba.ac.jp
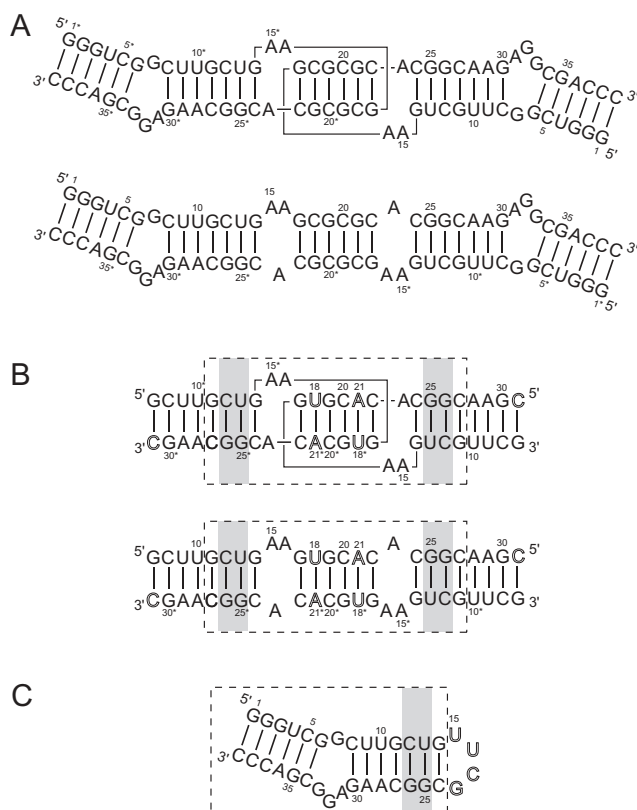
Fig. 1. **Secondary structure of a 39mer RNA corresponding to the dimerization initiation site (DIS39) and its fragments used in this study.** (A) The kissing-loop and extended-duplex dimers of DIS39. (B) The kissing-loop and extended-duplex dimers of loop25, which is composed of the loop and stem of DIS39. Modified residues are indicated by open characters. The sequence of the self-complementary loop was modified to increase the dispersion of NMR signals, and a base pair was added to the stem. The broken box indicates the part to be used for structure calculation. (C) Bulge34 consists of the stem-bulge-stem region of DIS39 and the connecting UUCG loop. The broken box indicates the part to be used for structure calculation. Gray shading indicates the two base pairs, C12–G26 and U13–G25, that are superimposed to combine the structures of the kissing-loop or extended-duplex dimer region and the stem-bulge-stem region. Asterisks indicate residues in the other strand.

MATERIALS AND METHODS

RNA Synthesis, Purification and Preparation—Non-labeled loop25 was synthesized chemically by the phosphoramidite method with an automatic DNA/RNA synthesizer, Expedite model 8909 (PerSeptive Biosystems Inc., MA, USA). The protection groups were removed with ammonia and tetra-*n*-butylammonium fluoride. Non-labeled DIS39 and bulge34 were synthesized enzymatically by the *in vitro* transcription reaction method with AmpliScribe T7 transcription kits (Epicentre Technologies Co., WI, USA). Purification for each RNA sample was performed by PAGE using 30 cm × 40 cm glass plates (Nihon Eido Co. Ltd., Tokyo, Japan) under denaturing conditions, and extensive desalting by ultrafiltration (Centricon YM3, Amicon Inc., MA, USA) was carried out. For stable isotopic labeling by the *in vitro* transcription with ^{13}C- and ^{15}N-labeled NTPs (Nippon Sanso, Tokyo, Japan), we used

DIS39 rather than shorter loop25 and bulge34 because the efficiency of *in vitro* transcription is better for larger RNA.

For the preparation of the kissing-loop dimer, DIS39 or loop25 in water was incubated at 368 K for 5 min and chilled on ice for 5 min. Then, the solvent was adjusted to 1× PN-buffer [10 mM sodium phosphate (pH 7.0) and 50 mM NaCl] by adding concentrated buffer. For the preparation of the extended-duplex dimer, DIS39 or loop25 in 1× PN-buffer was incubated at 368 K for 5 min and slowly cooled to room temperature. Bulge34 was annealed by heating at 363 K for 5 min and snap-cooling on ice. To confirm the formation of the hairpin structure, the samples were subjected to a native PAGE experiment. For NMR measurements, RNA samples were dissolved in 10 mM sodium phosphate buffer (pH 7.0) containing 50 mM NaCl. The final concentration of chemically synthesized loop25 was 1.8 mM. The concentrations of DIS39 and bulge34 (transcripts) were 1.0 and 0.5 mM, respectively. The concentration of the kissing-loop and extended-duplex dimers of [G-^{13}C/^{15}N] and [A-^{13}C/^{15}N] DIS39 were 0.4, 0.3, 0.2 and 0.1 mM, respectively.

NMR Measurements—NMR spectra were recorded on Bruker DRX-500 and DRX-600 spectrometers. Spectra were recorded at probe temperatures of 283 to 303 K and NMR data at 298 K were used for structure calculation. The imino proton signal of the G and U residues in H$_2$O were distinguished from each other by the HSQC selected and HSQC filtered 1D spectra measured with ^{13}C and ^{15}N-labeled DIS39. Exchangeable proton NOEs were determined by 2D NOESY in H$_2$O with a mixing time of 150 ms using the jump-and-return scheme and gradient pulses for water suppression. For resonance assignments, well-established procedures were used (*30*). The H2 protons of adenosine were assigned based on a 2D HSQC experiment with natural abundance ^{13}C. NOE distance restraints from non-exchangeable protons were obtained from 2D NOESY experiments (mixing times of 50, 100, 200, and 400 ms) in D$_2$O. The intensities of the NOEs between exchangeable protons were interpreted as distances of 2.1–5.0 Å. For loop25, distances were estimated by analyzing the initial slope of NOE intensities for mixing times of 25, 50, 100, 200 ms. Judgment of intermolecular NOE is described in the result section. Two restraints (>5 Å) were added to the distance restraints based on the absence of NOE cross peaks in the case of the kissing-loop dimer. For bulge34, the intensities of NOEs due to nonexchangeable protons were interpreted as distances with a margin of −1.5 to +1.5 Å for the 100 ms 2D NOESY and −1.0 to +2.0 Å for the 200 ms 2D NOESY. Two restraints (>5 Å) were added to the distance restraints based on the absence of NOE cross peaks. The formation of hydrogen binding of G:C, A:U or G:U base pairs is interpreted as distance constraints as 1.8–2.1 Å for hydrogen and acceptor atoms and 2.8–3.2 Å for donor and acceptor atoms; G11:C27 to G14:C24, G11*:C27* to G14*:C24* and G17:C22* to C22:G17* for loop25 in the kissing-loop dimer, G11:C27* to G14:C24*, G11*:C27 to G14*:C24 and G17:C22* to C22:G17* for loop25 in the extended-duplex dimer, and G1:C39 to C5:G35 and U9:A29 to G14:C24 for bulge34. Dihedral restraints were obtained as described below. The absence of crosspeaks between H1′–H2′ in the 2D TOCSY and DQF-COSY experiments was interpreted as the residue being in the C3′-*endo*

conformation. On the other hand, the presence of strong crosspeaks between H1′–H2′ in the 2D TOCSY and DQF-COSY experiments was interpreted as the residue being in the C2′-*endo* conformation. The correction of sugar puckering is interpreted as dihedral restraints for ν_2 as 40.00 ± 20.00° (C3′-*endo*) or −35.00 ± 20.00° (C2′-*endo*). Based on the sequential connectivity of the Watson-Crick and G-U base pairs, the RNA-A conformation was assumed for the stem region and dihedral restraints were introduced for backbone torsion angles (α, β, γ, δ, ε and ζ) as the ideal conformation with a margin of ±10.00°. For loop25 in the kissing-loop dimer, information about the C3′-*endo* conformation (G11–G14, G17–C27), the C2′-*endo* conformation (A16) and RNA-A conformation in the stem region (G11–U13, U18–A21, G25–C27) was used as the dihedral restraints. For loop25 of the extended-duplex dimer, information about the C3′-*endo* conformation (G11–G14, G17-C27) and RNA-A conformation in the stem region (G11–U13, U18–A21, G25–C27) was used as the dihedral restraints. For bulge34, the information about the C3′-*endo* conformation (G1–G14, C24–A31, C34–C39) and RNA-A conformation in the stem region (G1–C5, G11–G14, C24–C27, G35–C39) was used as the dihedral restraints.

Structure Calculation—A set of 100 structures was calculated using the simulated annealing protocol described below with the InsightII/Discover package, and the amber force field was used. The force constants were 100 kcal mol⁻¹ Å⁻² for distance restraints and 100 kcal mol⁻¹ rad⁻² for dihedral restraints. The starting coordinates were randomized, and the randomized structures were heated to 2,000 K in 5 ps, and the temperature was kept to 2,000 K for another 5 ps. After that, all restraints were increased to full values in 20 ps, then, decreased to 1/10 of full values in 5 ps at 2,000 K. Van der Waals radii were increased from 0.1 to 0.825 in 20 ps at 2,000 K. All restraints were increased to full value again in 10 ps at 2,000 K. Scalings for non-bond interactions were increased to full value in the next 20 ps at 2,000 K, and the temperature was kept to 2,000 K for another 5 ps. Then, the temperature was gradually scaled to 300 K in 10 ps. After that, the structure was heated from 300 to 1,000 K in 5 ps, and the van der Waals radii were increased from 0.825 to 1 at 1,000 K, and then decreased from 1 to 0.825 at 1,000 K. An additional 5 ps of dynamics was performed at 1,000 K, and the temperature was gradually scaled to 300 K for 10 ps. A final minimization step was performed, which included a Lennard-Jones potential and electrostatic terms with a dielectric constant of 7. The ten final structures with the lowest total energies were chosen.

RESULTS AND DISCUSSION

Analysis of the NMR Spectra of DIS39, Loop25 and Bulge34—Our previous NMR study revealed that the two types of dimers of DIS39 prepared as described in "MATERI-ALS AND METHODS" correspond to the kissing-loop and extended-duplex dimers (*31*). NMR spectra of DIS39 in each of the kissing-loop and extended-duplex dimers were measured in D₂O, and the signals due to H1′, H6/H8 were assigned by the sequential assignment method (Fig. 2). Figure 3A shows the difference in the chemical shift of H1′, H6/H8 between the two types of dimers. It was found that the difference is concentrated in the loop region.

Interestingly, structures of the stem-bulge-stem region of the kissing-loop and extended-duplex dimers were extremely similar, even though the stems are formed by intra and inter molecules. This was also shown by analysis of the TOCSY spectrum; differences are located in the loop regions. Most residues were adapted to the C3′-*endo* conformation except for G32, G33 in the bulge-out region of both forms, A16 in the kissing-loop dimer and A15, A16 in the extended-duplex dimer, which might be a mixture of the C2′-*endo* and C3′-*endo* conformations.

To reveal further authentic structure, two RNA molecules were designed; loop25 includes the loop region and bulge34 includes the stem-bulge-stem region (Fig. 1, B and C). Loop25 was constructed to determine the authentic structure of the loop region. In order to increase the dispersion of the NMR signals, the sequence of the loop was modified from GCGCGC to GUGCAC. One base pair was added by replacing A31 by C31 in the stem to increase the stability of the kissing-loop dimer. It is noted that the loop sequences of GCGCGC and GUGCAC correspond to those of HIV-1 subtypes B and F (*32*), respectively, and both sequences have dimerization activity (*6*, *9*, *10*). The chemical shifts of loop25 were compared with those of DIS39 in each of the kissing-loop and extended-duplex dimers (Fig. 3, B and C). For both conformations, the chemical shifts for most of the stem region and A15, A16 and A23 were strikingly similar between the loop25 and DIS39. Due to the base alterations, the chemical shifts of the self complement loop were slightly different for both dimers. The chemical shift of H8 was shifted more than 0.2 ppm due to the addition of the terminal base pair. It is noteworthy that the chemical shift difference in loop25 between the kissing-loop and extended-duplex dimers (Fig. 3D) was almost identical to that of DIS39 (Fig. 3A). These results indicate that the structures of loop25 in the kissing-loop and extended-duplex dimers are essentially identical to those of DIS39. Upon analysis of the TOCSY spectrum, it was found that most of the residues were adapted to the C3′-*endo* conformation except A15 and A16 for the extended duplex dimer and A16 for the kissing loop dimer, and these results also agree with the results for DIS39.

Bulge34 was constructed to determine the authentic structure of the stem-bulge-stem region. Bulge34 consists of the stem-bulge-stem region of DIS39 and the connecting UUCG loop. The NMR signals of bulge34 were assigned by the sequential assignment technique. The chemical shift of H1′, H6/H8 of bulge34 were compared to those of DIS39 in the kissing-loop dimer (Fig. 3E). The chemical shifts for the stem-bulge-stem regions of bulge34 and DIS39 were identical, although the chemical shifts of the residues adjacent to the loop were slightly different by reflecting the difference in the closing loop sequences. Upon analysis of the TOCSY spectrum, it was found that most residues were adapted to the C3′-*endo* conformation except for G32, G33 in the bulge-out region and C in the UUCG loop, and that the conformation in the stem-bulge-stem region also agreed with that of DIS39. These results indicate that the structure of the stem-bulge-stem region of bulge34 is identical to that of DIS39.

Thus, the structures of DIS39 for two types of dimers can be determined by determining the structures of loop25 and bulge34, and combining them.
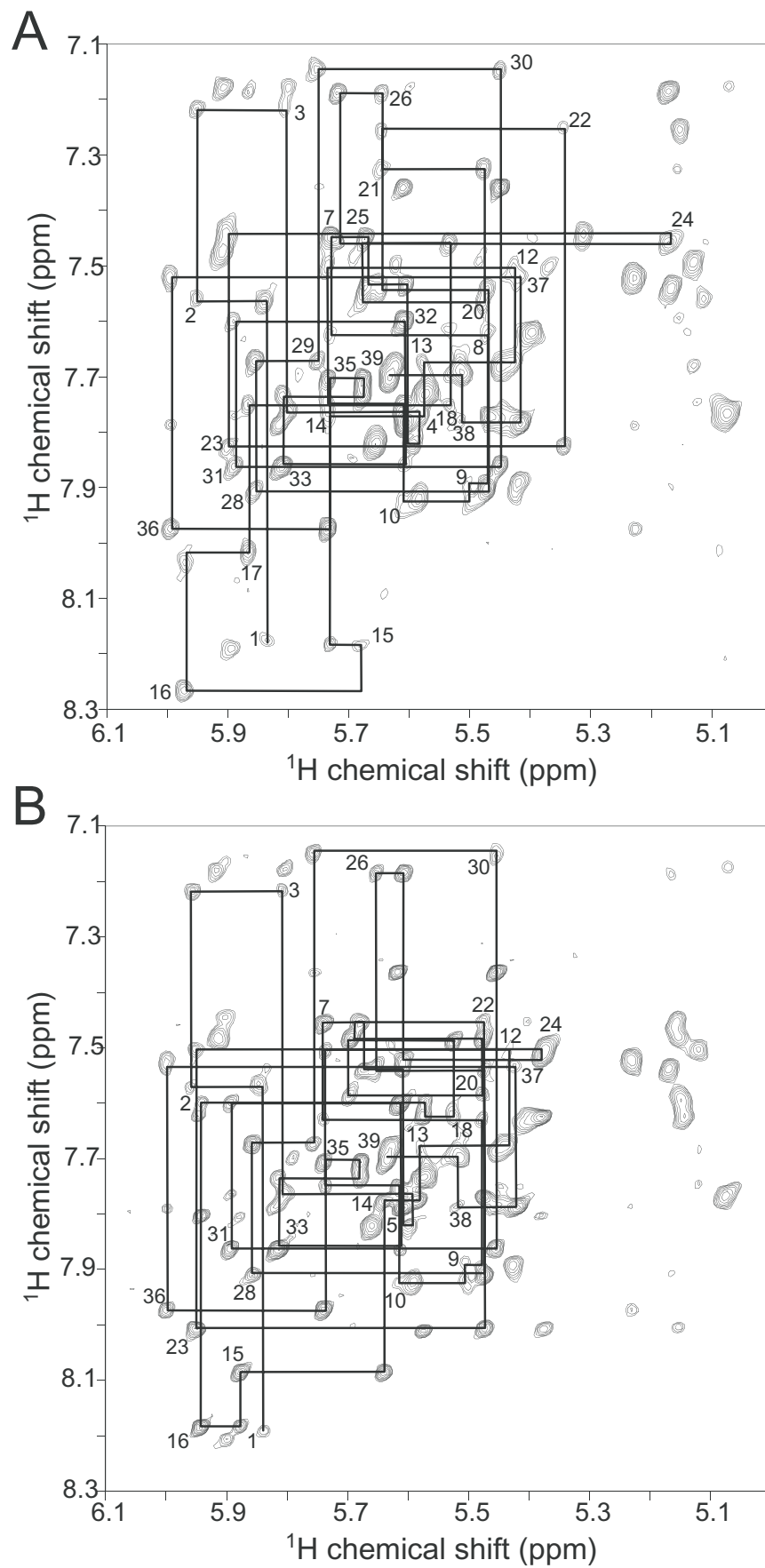
Fig. 2. **2D NOESY spectra of the (A) kissing-loop and (B) extended-duplex dimers of DIS39 measured in D₂O at 25°C with a mixing time of 200 ms.** Cross-peaks between aromatic H6/H8 protons and ribose H1′ protons are shown, and the sequential NOE connectivity is indicated.
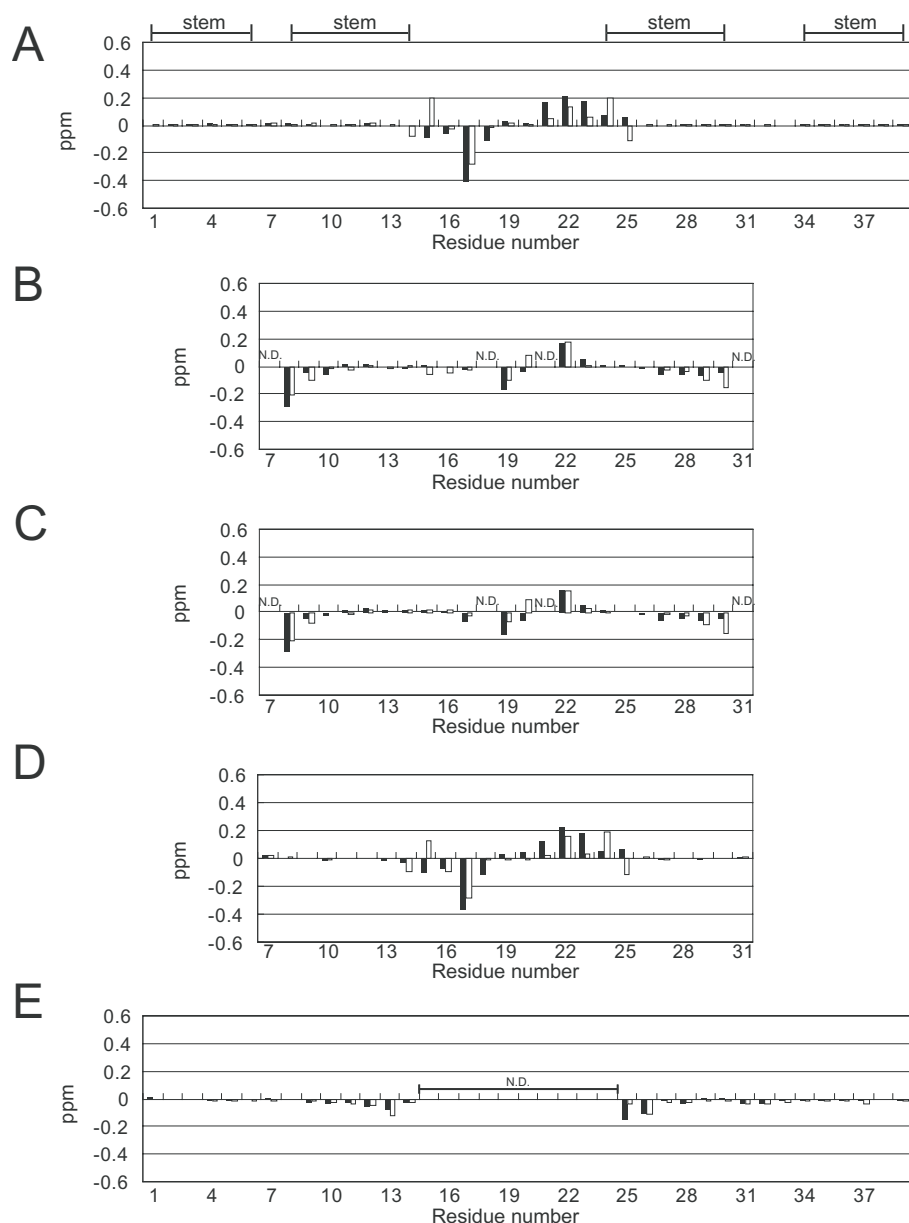
Fig. 3. **Chemical shift differences for H6/H8 and H1′.** Filled and open bars indicate H6/H8 and H1′, respectively. (A) Chemical shift differences between the kissing-loop and extended-duplex dimers of DIS39. Lines above the graph indicate the stem regions. (B) Chemical shift differences between DIS39 and loop25 in the kissing-loop dimer (data for replaced residues 7, 18, 21 and 31 are not shown). (C) Chemical shift differences between DIS39 and loop25 in the extended-duplex dimer (data for replaced residues 7, 18, 21 and 31 are not shown). (D) Chemical shift differences between the kissing-loop and extended-duplex dimers of loop25. (E) Chemical shift differences between DIS39 and bulge34 in the kissing-loop dimer (data for residues 15–24 are not shown).

Structure Determination—The loop region of loop25 in the kissing-loop dimer: To determine the structure of the loop region of DIS39 in both the kissing-loop and extended-duplex dimers, the NMR signals of loop25 were further analyzed and structural information was collected. The structure of the loop region consisting of the nine nucleotide loop and the stem with four base pairs was determined as shown by the broken box in Fig. 1B. A total of 286 distance restraints, 76 hydrogen bonding distance restraints, 140 dihedral restraints (Table 1), and 136 chiral restraints were used for the structural calculation. Three NOEs in the loop region, H2(A21)–H1′(U18), H2(A21)–H1′(G19) and H2(A21)–H8(G19), were judged to be intermolecular by analysis of the imino proton spectra. Four NOEs in the stem-loop linking region were considered to be intermolecular or intramolecular based on the results of the isotope filter NMR measurement (data not shown), and it was concluded that two NOEs, H2(A23)–H1′ (G17),

H2(A16)–H1′ (G16), are intermolecular and three NOE, H8(A16)–H1′ (A16), H8(A16)–H2′ (A16), are intramolecular. One NOE in the stem-loop linking region was considered to be intermolecular or intramolecular in the structure calculation, and it was concluded that this NOE, H2(A23)–H2(A15), is intramolecular. Each restraint is used twice for two molecules. The structures were calculated by the restrained molecular dynamic calculation with the simulated annealing method. The structure was defined with a heavy atom r.m.s.d. of 2.14 Å for the ten converged structures (Fig. 4A, left panel), and the minimized average structure is shown in Fig. 4A (right panel). Although the overall convergence was not very good, the self-complementary region was well defined with 0.76 Å, and the stem-loop linking region was defined with 1.86 Å. The structural statistics are summarized in Table 1.

The loop region of loop25 in the extended-duplex dimer: The loop region of loop25 in the extended-duplex dimer was

Table 1. **NMR restraints and structural statistics.**

	Number of restraints		
	loop25 in the kissing-loop dimer (17 mer × 2)	loop25 in the extended-duplex dimer (17 mer × 2)	bulge34 (30 mer)
Distance restraints	286	384	345
imino-imino	12	12	10
intra residue	154	182	163
intra molecule	106	174	170
inter molecule	12	16	–
>5 Å	2	0	2
Hydrogen bonding distance restraints	76	76	58
Dihedral restraints	140	138	126
3′-*endo*	30	30	28
2′-*endo*	2	0	0
RNA-A stems	108	108	98
r.m.s.d. from the idealized geometry (Å)			
Bonds (Å)	0.00897 ± 0.00004	0.00803 ± 0.00020	0.00775 ± 0.00015
Angle (°)	2.43 ± 0.23	2.33 ± 0.05	2.24 ± 0.07
Impropers (°)	1.57 ± 0.10	1.82 ± 0.64	1.53 ± 0.21
Heavy-atoms r.m.s.d. (Å)[a]			
All	2.14	1.45	1.98
Stem-loop linking region[b]	1.86	1.31	
Bulge region[c]			1.90

[a]Averaged r.m.s.d. between an average structure and the 10 converged structures were calculated. The converged structures did not contain experimental distance violations >0.2 Å or dihedral violations >5°, [b]The stem-loop linking region consists of residues 14 to 17, 22 to 24, 14* to 17* and 22* to 24*, [c]The bulge region consists of residues 6 to 10 and 28 to 34. Asterisks indicate residues in the other molecule.

determined (broken box in Fig. 1B). A total of 384 distance restraints, 76 hydrogen bonding distance restraints, 138 dihedral restraints (Table 1) and 136 chiral restraints were used for the structure calculation. For the stem-loop linking region, H2 of A23 was connected by intermolecular NOEs to H1′ and H2 of A15, H2 of A16 and H1′ of G17. The structures were calculated by the restrained molecular dynamic calculation with the simulated annealing method described above. The structure was well defined with a heavy atom r.m.s.d. of 1.45 for the ten converged structures (Fig. 4B, left panel), and the minimized average structure is shown in Fig. 4B (right panel). The stem-loop linking region was defined with 1.31 Å. The structural statistics are summarized in Table 1.

The stem-bulge-stem region of bulge34: A structural determination was performed for bulge34 except for the UUCG loop (broken box in Fig. 1C). A total of 345 distance restraints, 58 hydrogen bonding distance restraints, 126 dihedral restraints (Table 1) and 120 chiral restraints were used for the structure calculation. Two NOE restraints (>5 Å), H2(A31)–H1′(U9) and H1′(A31)–H1′(U9), were added to the distance restraints based on the absence of NOE cross peaks. The structures were calculated by the restrained molecular dynamic calculation with a simulated annealing protocol. The structure was defined with a heavy atom r.m.s.d. of 1.98 for the ten converged structures (Fig. 4C, left panel), and the minimized average structure is shown in Fig. 4C (right panel). Although the overall convergence is not very good, the stem regions are well defined with 0.83 or 0.78 Å, respectively. The bulge region was defined with 1.90 Å. The structural statistics are summarized in Table 1.

The two types of dimers of DIS39: Solution structures of DIS39 were then constructed by combining the structure parts. The structures of the kissing-loop or extended-duplex dimer region and stem-bulge-stem region were combined by superimposing two base pairs, C12:G26 and U13:G25 (Fig. 1, gray area). The left panels of Fig. 5 show the ten structures prepared by using the minimized average structure of the stem-bulge-stem region (Fig. 4C, right) and each of the ten lowest energy structures of the loop region (Fig. 4, A or B, left) superimposed by the loop region. The right panels of Fig. 5 show the structures prepared using the minimized average structure of the stem-bulge-stem region (Fig. 4C, right) and the loop region (Fig. 4, A or B, right). The relative angles between the stem-bulge-stem regions differ between the kissing-loop and extended-duplex dimers as shown in the right panels of Fig. 5. However, the fluctuations of the relative angles are rather large and the ranges overlap between the two dimers as shown in the left panels of Fig. 5. In fact, the values of the residual dipolar coupling for the stem-bulge-stem region are similar between the kissing-loop and extended-duplex dimers (to be published). A preliminary normal mode analysis suggested the existence of hinge motion, and, in order to reveal the dynamic properties of the dimers, a molecular dynamics analysis, as well as the thermodynamics analysis (*33*), is required. The most obvious local difference was observed for A16; for the kissing-loop dimer, A16 was close to the same residue in the other molecule (Fig. 6A, left) and did not stack above A15 of the same molecule nor G17 of the other molecule (Fig. 6A, right), whereas for the extended-duplex dimer, A16 was apart from the same residue of the other molecule (Fig. 6B, left) and stacked between A15 and G17 (Fig. 6B, right).

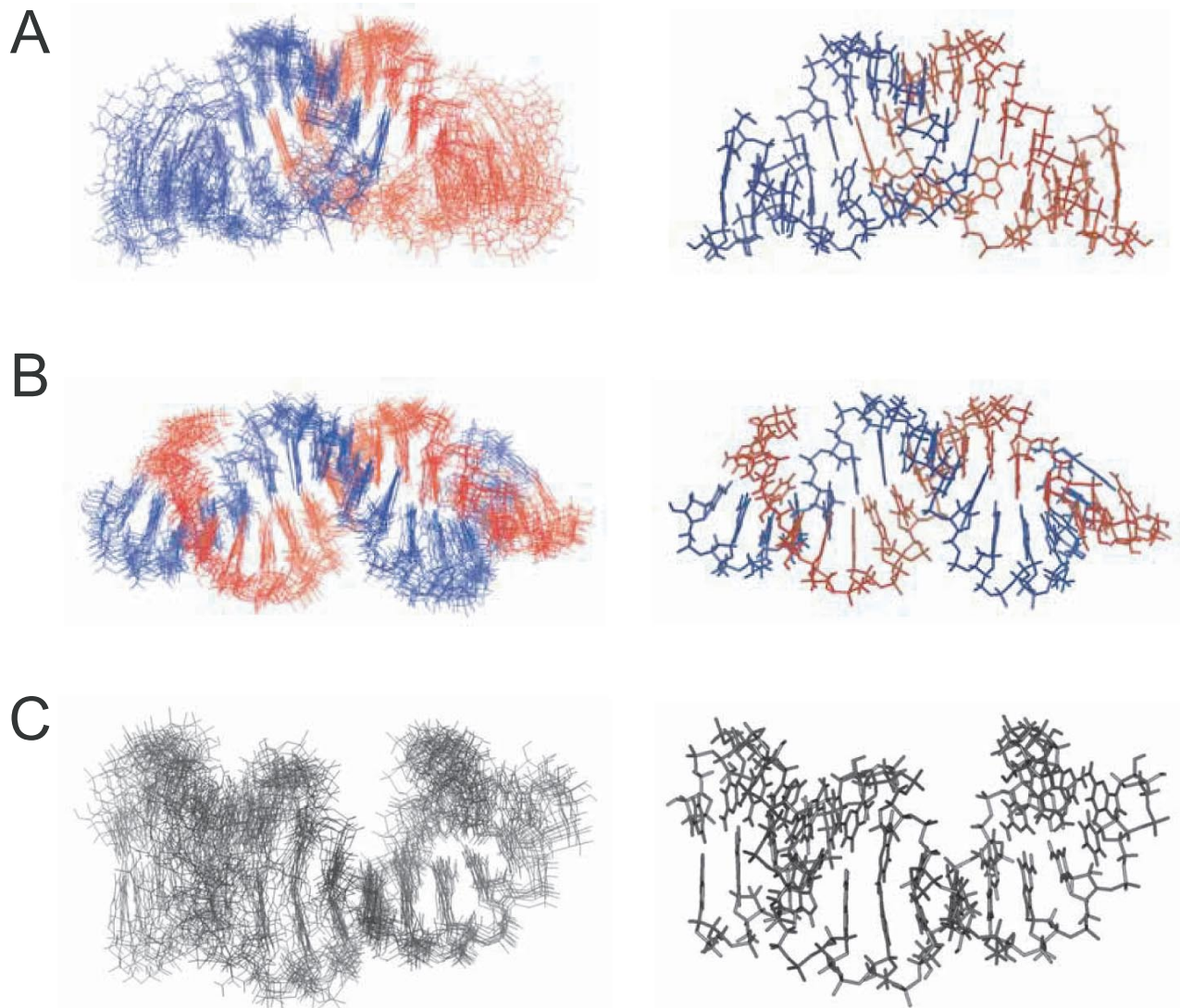*Structural Comparison with Related Structures—* Ennifar *et al.* (*20*) determined the crystal structure of

Fig. 4. **Solution structures of each part of DIS39.** Left panels show the superimposition of the 10 lowest energy structures and the right panels show the minimized average structures. (A) The loop region of loop25, as shown by the broken box in Fig. 1b, in the kissing-loop dimer. Each strand is colored in red or blue. (B) The loop region of loop25 in the extended-duplex dimer. (C) The stem-bulge-stem region of bulge34.

the kissing-loop dimer. The present structure is similar to the crystal structures in general, except for A15 and A16. In the present structure, A15 stacks on G14 and A16 interacts with the same residue in the other molecule (Fig. 6A, right). On the other hand, in the crystal structure, A15 and G16 are flipped out (20). It is noted that the numbering system of DIS39 is used for other structures for convenience, and position 16 is occupied by A or G depending on the strain. A15 and A16 (or G16) might be flexible and can be flipped out even in solution. Mujeeb *et al*. (19) determined the solution structure of the kissing-loop dimer. In this structure, A16 interacts with A15 and C24 in the other molecule, and, as a result, the distance between the two stems is relatively short. Thus, this restricted interaction makes the global structure different from the present structure and the crystal structure. However, the location of A15 is similar in the two solution structures. The difference in the conformation of A16 between the two solution

structures may reflect the difference in the sequence of the stem adjacent to the loop and/or in the sample condition, including the salt concentration. The NOE connectivity determined in the present study agrees in general with those of Dardel *et al*. who analyzed the structure of the stem-loop region in the kissing-loop dimer by NMR (34).

Girard *et al*. (21) and Mujeeb *et al*. (22) determined the solution structures of extended-duplex dimers. In these two structures, as well as in the present structure, A15, A16 and A23 form a zipper like structure (Fig. 6B, right). On the other hand, in the case of the crystal structure of the extended-duplex dimer, G16 forms a G:A base pair and A15 is flipped out, and it was assumed that this in-out bulge transconformation is magnesium-dependent (23).

Structures of the stem-bulge-stem region were shown by Lawrence *et al*. (26) and Yuan *et al*. (27). In the solution structure determined by Lawrence *et al*. (26), continuous stackings were formed for each strand, G6-C8 and
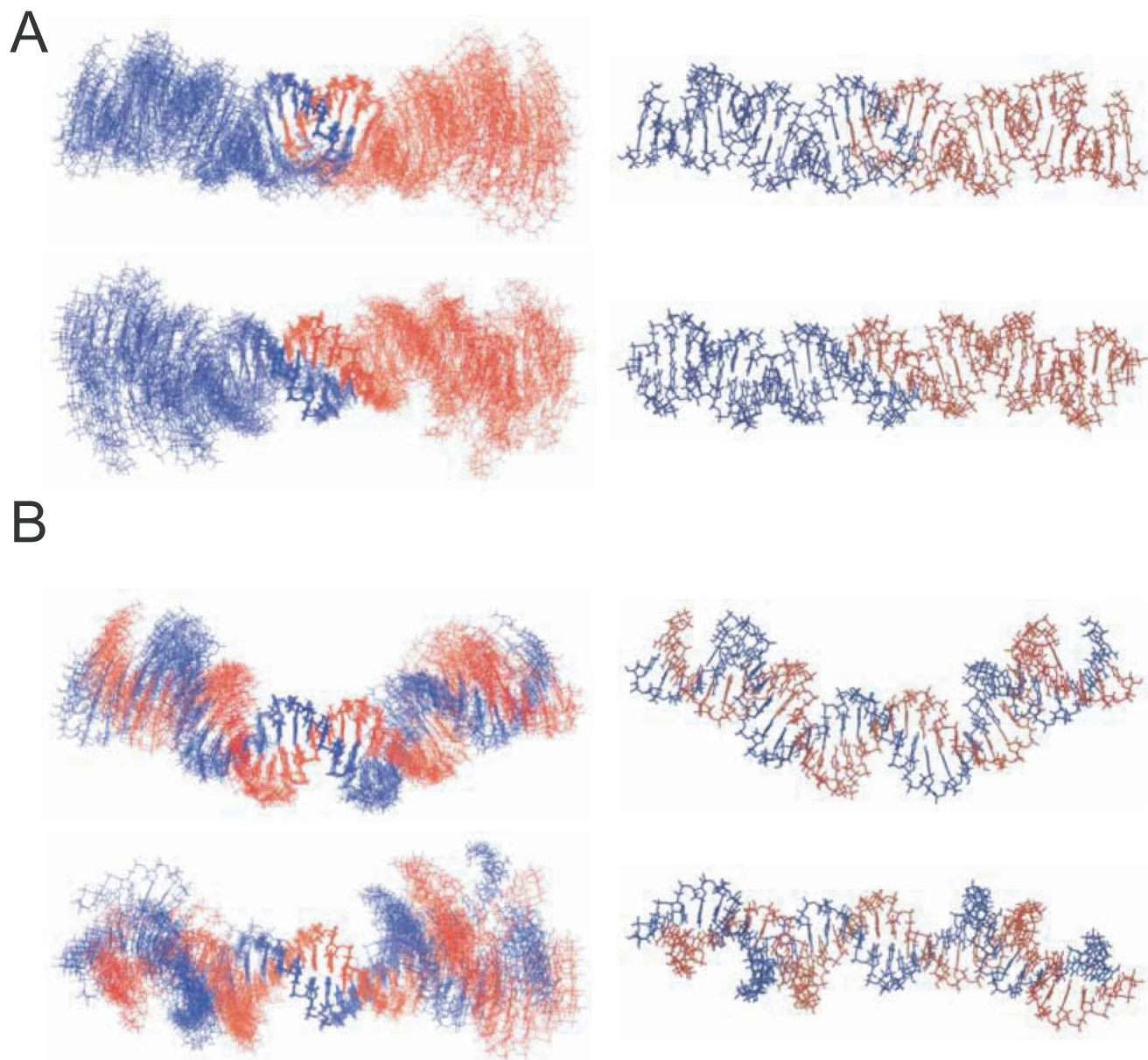
A

B

Fig. 5. **Solution structures of the (A) kissing-loop and (B) extended-duplex dimers of DIS39.** Left panels show the structures constructed by combining the structures of the loop (the 10 lowest energy structures of the kissing-loop or extended-duplex dimers) and the stem-bulge-stem (minimized average structure) regions. Right panels show the structures constructed by combining the minimized average structures of the loop and stem-bulge-stem regions. The two regions were combined by superimposing two base pairs, C12–G26 and U13–G25 (Fig. 1, gray area). Each strand is colored in red or blue and views from two different directions are shown.

G30-C34. Yuan *et al*. (*27*) showed that G7 and A31 form a base pair, and that G33 is not always stacked on G32 or C34, and, in general, the present structure is identical to the latter structure. Greatorex *et al*. (*25*) showed that the bulge region is too flexible to determine the conformation. These conformational differences may be caused by differences in the stability of the terminal stem. Lawrence *et al*. (*26*) adopted a stable 7 base-pair stem, and their structure forms an ordered conformation in the bulge region. In contrast, Greatorex *et al*. (*25*) adopted an unstable 4 base-pair stem and the bulge region is flexible. Yuan *et al*. (*27*) adopted a 4 base-pair stem and a flanking adenosine

residue at the 3′ terminal that must stabilize the stem. In the present study, a 6 base-pair stem was used.

Mechanism of the Two Stem Dimerization—Between the kissing-loop and extended-duplex dimers, A16 shows the most drastic change in interaction with other residues, suggesting that A16 is the key residue in the two step dimerization reaction. The difference in the A16 conformation among structures with different sequences and determined under different conditions as described above, also suggests the importance of this residue. Mujeeb *et al*. (*19*, *22*) also pointed out the flexibility around the junction of the loop and the stem of DIS in the kissing-loop and
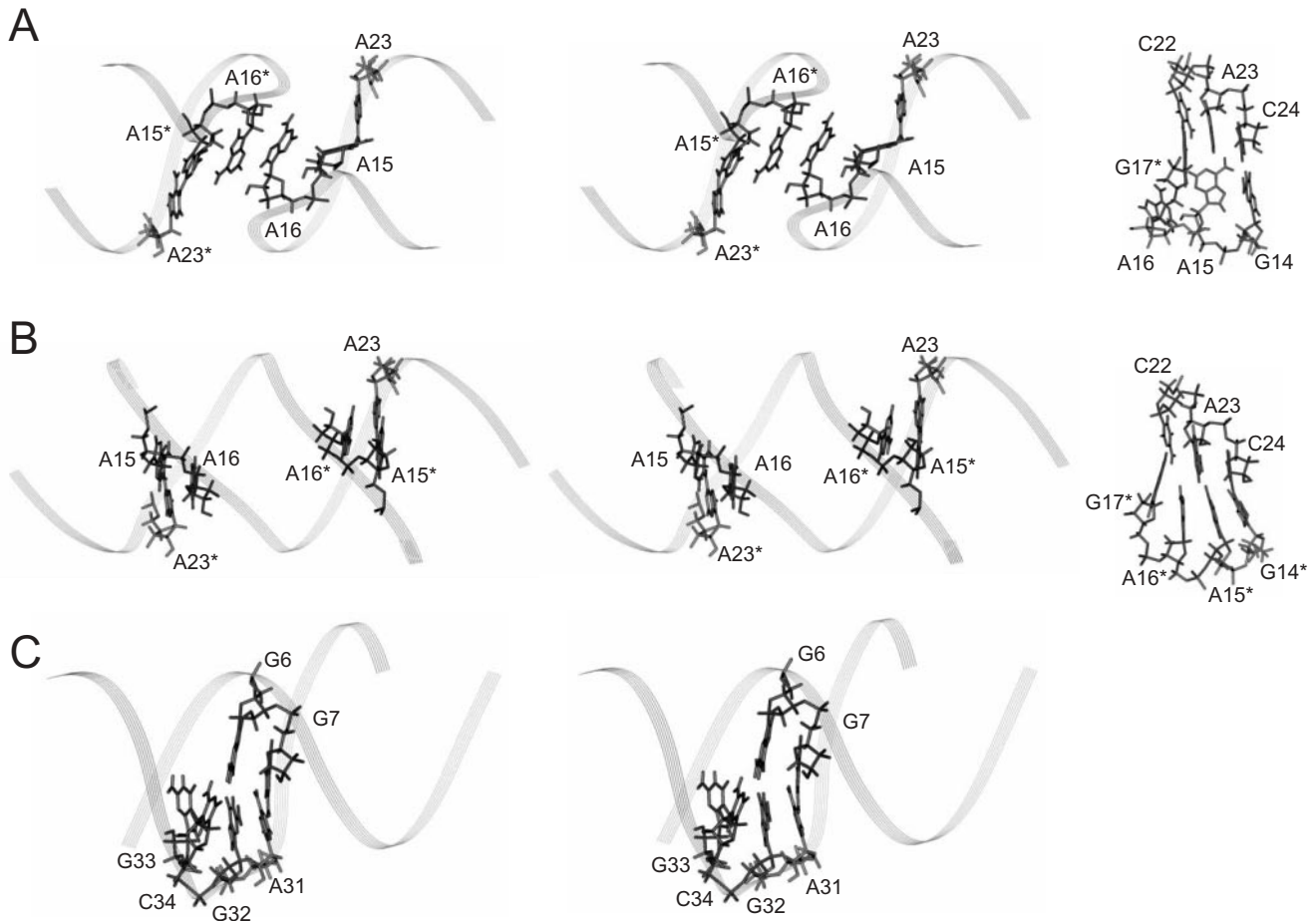
Fig. 6. **Structures of the linking regions.** (A) Regions linking the stem and loop in the kissing-loop dimer. The left panels show the positions of A15, A16, and A23 in the entire structure in a stereo view, and the right panels show residues linking the stem and loop. Asterisks indicate residues in the other strand. (B) Regions linking the stem and loop in the extended-duplex dimer. (C) The bulge region linking the two stems.

extended-duplex dimers. Imino proton signals due to U9:A29 and U10:A28 are much broader than other signals in the stem region, and no imino proton signal due to C8:G30 was observed. Thus, the stem between the loop and bulge is destabilized by the bulge region. Our previous experiments also showed that the bulge region is required for the two-step dimerization to adjust the thermal stability of DIS, and Greatorex *et al.* (*25*) also indicated that the flexibility of the bulge region is critical based on the fact that mutations in the bulge region strongly affect the melting temperature, as well as the fact that none of the wild-type sequences in the bulge region that increase the melting temperature is ever found in wild-type viruses. Thus, the conformational conversion from the kissing-loop dimer to the extended-duplex dimer might require two factors, the movement of A16 and the modest stability of the stem caused by the presence of the bulge region.

In the present study, a set of structures corresponding to the initial and final structures of the two-step dimerization of DIS are provided; these structures will promote studies to elucidate the molecular mechanism of the conformational change in the two-step dimerization, including an analysis of the interaction between DIS and NCp7, in addition to the molecular dynamics approach.

Coordinates: The structure has been deposited in the Protein Data Bank (accession code 2D17: the stem-bulge-stem region of bulge34, 2D18: the extended-duplex dimer of loop25, 2D19: the kissing-loop dimer of loop25, 2D1A: the extended-duplex dimer of DIS39 and 2D1B: the kissing-loop dimer of DIS39).

REFERENCES

1. Hoglund, S., Ohagen, A., Goncalves, J., Panganiban, A.T., and Gabuzda, D. (1997) Ultrastructure of HIV-1 genomic RNA. *Virology* **233**, 271–279
2. Laughrea, M., Jette, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M.A. (1997) Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.* **71**, 3397–3406
3. Clever, J.L. and Parslow, T.G. (1997) Mutant human immunodeficiency virus type 1 genomes with defects

in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414

4. Paillart, J.C., Berthoux, L., Ottmann, M., Darlix, J.L., Marquet, R., Ehresmann, B., and Ehresmann, C. (1996) A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J. Virol.* **70**, 8348–8354

5. Laughrea, M. and Jette, L. (1994) A 19-nucleotide sequence upstream of the 5′ major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry* **33**, 13464–13474

6. Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl. Acad. Sci. USA* **91**, 4945–4949

7. Fu, W. and Rein, A. (1993) Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J. Virol.* **67**, 5443–5449

8. Fu, W., Gorelick, R.J., and Rein, A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.* **68**, 5013–5018

9. Laughrea, M. and 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

10. Muriaux, D., Fosse, P., and 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

11. Muriaux, D., Girard, P.M., Bonnet-Mathoniere, B., and Paoletti, J. (1995) Dimerization of HIV-1Lai 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

12. Laughrea, M., Shen, N., Jette, L., Darlix, J., Kleiman, L., and Wainberg, M.A. (2001) Role of distal zinc finger of nucleocapsid protein in genomic RNA dimerization of human immunodeficiency virus type 1; No role for the palindrome crowning the R-U5 hairpin. *Virology* **281**, 109–116

13. de Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 ψ–RNA recognition element. *Science* **279**, 384–388

14. Amarasinghe, G.K., de Guzman, R.N., Turner, B.G., Chancellor, K.J., Wu, Z.R., and Summers, M.F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to Stem-Loop SL2 of the ψ–RNA packaging signal. Implications for Genome recognition. *J. Mol. Biol.* **301**, 491–511

15. Berkowitz, R., Fisher, J., and Goff, S.P. (1996) RNA packaging. *Curr. Top. Microbiol. Immunol.* **214**, 177–218

16. Darlix, J.L., Lopez-Lastra, M., Mély, Y., and Roques, B. (2003) Nucleocapsid protein chaperoning of nucleic acids at the heart of HIV structure, assembly and cDNA synthesis. In *HIV Sequence Compendium 2002* (Kuiken, C., Foley, B., Freed, E., Hahn, B., Marx, P., McCutchan, F., Mellors, J.W., Wolinsky, S., and Korber, B., eds.) pp. 69–88, Los Alamos National Laboratory, Los Alamos, NM

17. Takahashi, K., Baba, S., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2001) Two basic regions of NCp7 are sufficient for conformational conversion of HIV-1 dimerization initiation site from kissing-loop dimer to extended-duplex dimer. *J. Biol. Chem.* **276**, 31274–31278

18. Baba, S., Takahashi, K., Koyanagi, Y., Yamamoto, N., Takaku, H., Gorelick, R.J., and Kawai, G. (2003) Role of the Zinc Fingers of HIV-1 Nucleocapsid Protein in Maturation of Genomic RNA. *J. Biochem.* **134**, 637–639

19. Mujeeb, A., Clever, J.L., Billeci, T.M., James, T.L., and Parslow, T.G. (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.* **5**, 432–436

20. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001) Crystal Structures of Coaxially-Stacked Kissing Complexes of the HIV-1 RNA Dimerization Initiation Site. *Nat. Struct. Biol.* **8**, 1064–1068

21. Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J., and Lancelot, G. (1999) Dimer Initiation Sequence of HIV-1Lai Genomic RNA: NMR Solution Structure of the Extended Duplex. *J. Biomol. Struct. Dyn.* **16**, 1145–1157

22. Mujeeb, A., Parslow, T.G., Zarrinpar, A., Das, C., and James, T.L. (1999) NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett.* **458**, 387–392

23. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure Fold Des.* **7**, 1439–1449

24. Ennifar, E., Walter, P., and Dumas, P. (2003) A Crystallographic Study of the Binding of 13 Metal Ions to Two Related RNA Duplexes. *Nucleic Acids Res.* **31**, 2671–2682

25. Greatorex, J., Gallego, J., Varani, G., and Lever, A. (2002) Structure and Stability of Wild-Type and Mutant RNA Internal Loops from the SL-1 Domain of the HIV-1 Packaging Signal. *J.Mol.Biol.* **322**, 543–557

26. Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z., and Summers, M. F. (2003) Structure of the Intact Stem and Bulge of HIV-1 Psi-RNA Stem-Loop Sl1. *J. Mol. Biol.* **326**, 529–542

27. Yuan, Y., Kerwood, D.J., Paoletti, A.C., Shubsda, M.F., and Borer, P.N. (2003) Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1×3 internal loop. *Biochemistry* **42**, 5259–5269

28. Shen, N., Jette, L., Liang, C., Wainberg, M.A., and Laughrea, M. (2000) Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.* **74**, 5729–5735

29. Takahashi, K.I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **6**, 96–102

30. Varani, G., Aboul-era, F., and Allain, F.H.-T. (1996) NMR investigation of RNA structure. *Prog. NMR Spect.* **29**, 51–127

31. Takahashi, K., Baba, S., Hayashi, S., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) NMR analysis on intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J. Biochem.* **127**, 681–639

32. St.Louis, D.C., Gotte, D., Sanders-Buell, E., Ritchey, D.W., Salminen, M.O., Carr, J.K., and McCutchan, F.E. (1998) Infectious molecular clones with the nonhomologous dimer initiation sequences found in different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection in vitro. *J. Virol.* **72**, 3991–3998

33. Weixlbaumer, A., Werner, A., Flamm, C., Westhof, E., and Schroeder, R. (2004) Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes. *Nucleic Acids Res.* **32**, 5126–5133

34. Dardel, R., Marguet, R., Ehresmann, C., Ehresmann, B., and Blanquet, S. (1998) Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res.* **26**, 3567–3571