Structure and Dimerization of HIV-1 Kissing Loop Aptamers

J. Stephen Lodmell1,2, Chantal Ehresmann1, Bernard Ehresmann1 and Roland Marquet1*

1UPR 9002 du CNRS, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg cedex, France
2Division of Biological Sciences, The University of Montana, Missoula, MT 59812, USA

Dimerization of two homologous strands of genomic RNA is an essential feature of the retroviral replication cycle. In HIV-1, genomic RNA dimerization is facilitated by a conserved stem-loop structure located near the 5' end of the viral RNA called the dimerization initiation site (DIS). The DIS loop is comprised of nine nucleotides, six of which define an autocomplementary sequence flanked by three conserved purine residues. Base-pairing between the loop sequences of two copies of genomic RNA is necessary for efficient dimerization. We previously used in vitro evolution to investigate a possible structural basis for the marked sequence conservation of the DIS loop. In this study, chemical structure probing, measurements of the apparent dissociation constants, and computer structure analysis of dimerization-competent aptamers were used to analyze the dimers' structure and binding. The selected aptamers were variants of the naturally occurring A and B subtypes. The data suggest that a sheared base-pair closing the loop of the DIS is important for dimerization in both subtypes. On the other hand, the open or closed state of the last base-pair in the stem differed in the two subtypes. This base-pair appeared closed in the subtype A DIS dimer and open in subtype B. Finally, evidence for a cross-talk between nucleotides 2, 5, and 6 was found in some, but not all, loop contexts, indicating some structural plasticity depending on loop sequence. Discriminating between the general rules governing dimer formation and the particular characteristics of individual DIS aptamers helps to explain the affinity and specificity of loop-loop interactions and could provide the basis for development of drugs targeted against the dimerization step during retroviral replication.

Abbreviations used: HIV-1, human immunodeficiency virus type 1; DIS, dimerization initiation site; DEPC, diethylpyrocarbonate; DMS, dimethyl sulfate.

E-mail address of the corresponding author: r.marquet@ibmc.u-strasbg.fr

Introduction

Dimerization of retroviral RNA genomes prior to or concomitant with viral encapsidation and budding is a vital step in the retroviral replication cycle. An essential component of the dimerization process in HIV-1 is a stem-loop structure called the dimerization initiation site (DIS) located in the upstream leader sequence of the genomic RNA.1,2 The loop contains an autocomplementary hexanucleotide sequence comprised almost always of either GUGCAC (subtype A) or GCGCGC (subtype B) that, along with non-canonical interactions provided by conserved flanking purine nucleotides, is required for stable dimerization.3 Stable dimerization proceeds through an RNA loop-loop kissing interaction whereby the loop autocomplementary sequences form intermolecular base-pairs. Deletion or mutation of this sequence results in mutant viruses with markedly diminished infectivity and replication kinetics, and specific defects have been demonstrated in genomic RNA dimerization, encapsidation, and proviral DNA synthesis.4–9

Extensive chemical and enzymatic probing, site-directed mutagenesis, and molecular modeling have been used to construct a three-dimensional model of the subtype A kissing loop complex.3,10
Preliminary NMR data on subtype A model constructs provided several distance constraints consistent with this model, although a complete structure based solely upon the NMR data was not presented.11 In addition, an NMR-derived model for the subtype B DIS dimer has been reported.12 Interestingly, the NMR data of the subtype A and B dimers suggest significantly different conformations despite relatively modest sequence differences. In particular, the NMR model of the subtype B dimer suggests a melting of the last helical base-pair, and unpaired, but extensively stacked, adenines flanking the autocomplementary sequence. On the other hand, solution structure methods have suggested that the helical stems of subtype A dimers remain closed (i.e. base-paired) and that the purines flanking the autocomplementary sequence are involved in non-canonical base-pairing.3,10 Differences in the biochemical behavior of subtypes A and B dimers have been noted as well, and may be attributable in part to the presence or absence of a specific magnesium binding site among the two subtypes.10 Knowledge of the detailed three-dimensional structure could ultimately lead to rational drug design targeted against specific features of the DIS stem-loop. In addition, detailed analysis of retroviral RNA dimerization could lend insight into other RNA loop-loop interactions, which are found in many other biological systems, such as in the control of replication of plasmids ColE1 and R1,13–15 multimerization of cellularly localized mRNA,16 the organization of RNA structure within enterovirus and poliovirus genomes,17,18 and within the Neurospora ribozyme.19 Understanding the mechanisms of this type of intermolecular recognition and binding event requires both structural and functional studies of the kissing loop complexes.

The basic mechanism underlying the kissing loop interaction involves a primary recognition event between one or several loop nucleotides followed by an extension of the base-pairing interaction to include the entire autocomplementary loop sequence.20,21 Under some circumstances, the loop-loop base-pairing advances farther to include part of or the entire stem sequence surrounding the loop to form an extended duplex (for example, see reference 22). In vitro studies on HIV-1 DIS RNA have shown that an extended duplex can form, depending on the presence or absence of nucleocapsid protein, incubation temperature, ionic conditions, and the sequence of the stems.22–28 Whether or not the extended duplex occurs in vivo is unclear, as advantages associated with formation of an increasingly stable dimeric structure might be overcome by topological problems involved in twisting two very large RNAs around each other. In either event, the initial recognition between the two RNA molecules appears to be a loop-loop interaction, or kissing complex, and this is the subject of our current study.

We previously performed an in vitro selection/evolution study on model DIS stem-loops randomized at some or all of the loop positions in an effort to understand the sequence and structural constraints on this conserved motif.29 Results of that study revealed some fundamental sequence and structural requirements for loop-loop interactions in general and for dimerization of HIV-1 RNA in particular. Specifically, constraints were identified for loop size, autocomplementary sequence identity, autocomplementary sequence size, and non-canonical interactions involving the nucleotides flanking the autocomplementary sequence. Interestingly, there were very few autocomplementary sequences capable of promoting homodimerization. On the other hand, numerous families of aptamers were isolated that were incapable of homodimerization because their loops did not contain perfect autocomplementary sequences. However, these species were capable of efficiently dimerizing with appropriate complementary partner species isolated from the same pool.

In this study, we have measured the apparent dimerization equilibrium constant of individual dimers derived from subtypes A and B, and have applied chemical probing techniques to investigate commonalities and differences in the structures of dimers of different sequences. Using our library of aptamers possessing similar dimerization activity yet different sequences, we were able to determine the contributions of particular nucleotides to dimer structure and stability. Chemical structure probing is in agreement with the proposed sheared base-pair between positions 1 and 9 of the loop and showed protections consistent with a base triple interaction in some loop contexts. Surprisingly, we also found the status of the last base-pair in the stem was dependent upon the loop sequence.

Results
Isolation of dimerization-competent clonal RNAs

The library of dimerization-competent model DIS RNA aptamers used in this study was isolated by in vitro evolution.29 Briefly, positions 1, 2, 5, 6, and 9 of the DIS loop sequences derived from HIV-1 subtypes A or B were randomized (yielding pre-selection “randomer” pools with loop sequences NNGU NNA CN or NNG C NN GCN, respectively; the hexanucleotide sequences involved in intermolecular Watson-Crick base-pairing are underlined throughout). RNAs were selected and amplified according to their ability to dimerize. The resulting aptamers fell into several easily recognizable families (Figure 1). The subtype A aptamers formed predominately homodimers with respect to their autocomplementary sequences (either GUCGAC or GUGCAC) and otherwise formed heterodimers, mostly of the type GLUG-GAC-GUCCAC. The subtype B dimers were more diverse, forming a minority of homodimers of the types GCCGGC or GCGGC and several categories of heterodimers (Figure 1).29
The DIS model constructs were designed to examine the kissing complex rather than an extended duplex. The DIS stem in HIV-1 contains an internal loop that affects the conversion of the kissing complex to an extended duplex. In order to focus exclusively on the loop-loop interaction we replaced the internal loop with canonical base-pairs (Figure 1). This deletion strengthened the stem and also avoided the possibility that a loop from one molecule might interact with the internal loop of another. Rapid dissociation kinetics of five representative molecules, including the subtype A and B wild-types, upon dilution suggested that the interaction was solely loop-loop and did not involve the extended duplex. Dimeric complexes comprised of labeled RNAs were diluted 400-fold with buffer and aliquots were withdrawn at timed intervals and fractionated on a gel to determine the ratio of dimeric to monomeric RNA. The repartitioning between the dimeric and monomeric pools was rapid, most of it occurring in the first seconds after dilution. In contrast, a model extended duplex construct made of a sense and an antisense strand of RNA exhibited much slower dissociation after dilution, requiring minutes to hours to re-equilibrate (data not shown).

Within each of the families of dimers, there were variants with respect to positions 1, 2, and 9 although certain nucleotides at these positions were highly favored while others were strongly excluded. The ensemble of these variants provided a study system to ascertain the relative contributions of each position to overall dimer stability and/or structure. This was especially true for heterodimers that allowed a number of combinations to be tested, including some that formed fairly unstable complexes, and hence were unlikely to be formed during the selection process.

**K_d determinations**

During the selection procedure, binding partner(s) of individual aptamers within the total population were not explicitly identified. Thus, it
was not until individual clones were isolated and pure clonal RNA was produced that we could precisely measure the binding affinities of individual clones with various potential binding partners. We purified RNA from several individual clones from each group and combined these RNAs to measure their affinities for each other by maintaining a small, constant quantity of labeled RNA and adding increasing amounts of unlabeled homologous or heterologous RNA and measuring the distribution of dimeric versus monomeric RNA on a native polyacrylamide gel. \( K_d \) values were determined as described in Materials and Methods and exemplified in Figure 2. The \( K_d \) determinations were reproduced two to five times. Standard deviations were typically less than 50% of the mean values. Within a given aptamer subtype, \( K_d \) values were similar, probably because the populations of aptamers were selected solely for their ability to form dimers robust enough to survive gel electrophoresis.

### Effects of base substitutions at positions 5 and 6

Sequence analysis of selected aptamers indicated a high preference for G and C residues at positions 5 and 6 (Figure 1). Indeed, all of the stable homodimers contained a GC (wild-type-like) or CG ("inverted") central dinucleotide. One A-U base-pair was tolerated at these positions in heterodimers, but only in the subtype B aptamers, which have C and G at positions 4 and 7, respectively (Figure 1). In order to evaluate the importance of the nucleotides at positions 5 and 6, we measured the stability of homodimers and heterodimers of subtype B aptamers containing one or two G-U wobble base-pairs at these positions.

The aptamer with loop sequence AGgcUGgcA formed poorly stable homodimers (Table 1, line 1). On the contrary, the single isolated member of the homologous GU-family did not dimerize, even at very high concentrations of RNA (Table 1, line 4). Thus, the presence of a UG/GU tandem base-pair within the autocomplementary sequence weakly favored dimerization, while a GU/UG tandem abolished dimerization. This result is in line with the results of Turner's group, who have shown that the order and nearest neighbor context of tandem GU base-pairs significantly affect their thermodynamic contributions to helix stability. Both the AGgcUGgcA and the AAgcGUgcA aptamers formed fairly stable heterodimers containing one or two G-U wobble base-pairs at these positions.

The aptamer with loop sequence AGgcUGgcA formed poorly stable homodimers (Table 1, line 1). On the contrary, the single isolated member of the homologous GU-family did not dimerize, even at very high concentrations of RNA (Table 1, line 4). Thus, the presence of a UG/GU tandem base-pair within the autocomplementary sequence weakly favored dimerization, while a GU/UG tandem abolished dimerization. This result is in line with the results of Turner's group, who have shown that the order and nearest neighbor context of tandem GU base-pairs significantly affect their thermodynamic contributions to helix stability. Both the AGgcUGgcA and the AAgcGUgcA aptamers formed fairly stable heterodimers containing one or two G-U wobble base-pairs at these positions.

The aptamer with loop sequence AGgcUGgcA formed poorly stable homodimers (Table 1, line 1). On the contrary, the single isolated member of the homologous GU-family did not dimerize, even at very high concentrations of RNA (Table 1, line 4). Thus, the presence of a UG/GU tandem base-pair within the autocomplementary sequence weakly favored dimerization, while a GU/UG tandem abolished dimerization. This result is in line with the results of Turner's group, who have shown that the order and nearest neighbor context of tandem GU base-pairs significantly affect their thermodynamic contributions to helix stability. Both the AGgcUGgcA and the AAgcGUgcA aptamers formed fairly stable heterodimers containing one or two G-U wobble base-pairs at these positions.
This observation is in keeping with the proposal that the second nucleotide of the loop may be involved in non-canonical interactions with the central 5:6 and 6:5 base-pairs (refers to the partner molecule in a homo- or heterodimer pair) during the dimerization process, and that those interactions differ between the A and B subtypes.

Surprisingly, among the subtype A aptamers, clones with the wild-type loop sequence (AGGUGCACA) formed the least stable dimer (Table 2, pairs 5-8). This indicates that the conserved C2 is not particularly stabilizing within the DIS dimeric complex and suggests that this particular nucleotide has been evolutionarily conserved to fulfill another function.

### Complementary/autocomplementary sequence span length

Wild-type DIS sequences typically have a hexanucleotide autocomplementary sequence within the loop, and in the wild-type dimers there are six canonical intermolecular base-pairs between the autocomplementary sequences. However, we isolated aptamers that contained sequences that could potentially form more than six intermolecular base-pairs (Figure 3(a)). One such example is the aptamer with the loop sequence AUguGCaC, which has the potential to form homodimers containing eight intermolecular base-pairs. However, these dimers are less stable than those formed with aptamers with an A or a C at position 2 (Table 2, pairs 1-4) suggesting that extending the complementarity above six nucleotides does not favor dimerization.

To test this hypothesis, we measured the stability of heterodimers formed by aptamers of the "GG/CC" groups of subtype A and B (Figure 1). Not surprisingly, the $K_d$ of subtype B dimers were significantly lower than those of subtype A dimers, reflecting the stabilization provided by the C4-G7 and G7-C4 base-pairs (Table 3). Within the subtype A heterodimers, the $K_d$s of those having no potential to form Watson-Crick base-pairs between positions 2 and 9 ranged from 1.7 to 50 nM (Table 3, pairs 1, 2, 6, and 7). The $K_d$s of the heterodimers that could form one (pairs 3, 4, and 8) or two (pair 5) base-pair(s) fell in the same range, indicating that potential base-pairing between positions 2 and 9 was not the major factor affecting the stability of dimers. Rather, the identity of the nucleotide at position 1, and to a lesser extent at position 9, seemed to play a major role in this respect: those dimers with a G1 were significantly less stable than those with A1 (Table 3, compare pairs 1-5 and 6-8).

However, pairs 1-5 and 6-8, which are identical with respect to position 1, allowed a more detailed analysis of the influence of base-pairing between positions 2 and 9. Heterodimer no. 3, which differs from no. 1 only at N2, had a twofold higher $K_d$, even though it could form a U9:A2 base-pair. Similarly, heterodimer no. 4, which allows formation of a U2:A9 base-pair, had a lower $K_d$. This conclusion also held true when comparing pairs 6 and 8 that only differed at N2 and could potentially form six or seven intermolecular Watson-Crick base-pairs (Table 3).

The subtype B heterodimers were more difficult to compare because the clones often did not allow variation of N2 (and N2') independently of N1 and N9 (and N1' and N9') (not shown). Moreover, most $K_d$ values were within a rather narrow range. However, pairs 9 to 12, which include heterodimers with extended Watson-Crick and/or wobble pairing potential, showed no correlation between...
dimer stability and complementary sequence span length (Table 3). Similar data were obtained with aptamers of the GC and CG groups (Figure 1), although in this case heterodimerization competed with homodimerization (data not shown).

Taken together, these data suggest that the optimal size of the auto-complementary sequence is six nucleotides. The adjacent flanking nucleotides are not efficiently utilized for canonical intermolecular base-pairing.

**Table 3.** $K_d$ measurements of heterodimer pairs with different base-pairing potentials between positions 2 and 9

<table>
<thead>
<tr>
<th>Pair no.</th>
<th>Loop sequences of dimer pairs</th>
<th>Potential N2:N9 and N9:N9 base-pairs</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AGuGGuU/ACguCCacA</td>
<td>A2:A9', U9:U2'</td>
<td>1.7 ± 0.8</td>
</tr>
<tr>
<td>2</td>
<td>AGuGGuU/ACguCCacC</td>
<td>A2:C9', U9:U2'</td>
<td>13 ± 4.5</td>
</tr>
<tr>
<td>3</td>
<td>AGuGGuU/ACguCCacA</td>
<td>A2:A9', U9:A2'</td>
<td>3.6 ± 1.6</td>
</tr>
<tr>
<td>4</td>
<td>AGuGGuU/ACguCCacA</td>
<td>U2:A9', U9:C2'</td>
<td>2.8 ± 1.2</td>
</tr>
<tr>
<td>5</td>
<td>AGuGGuU/ACguCCacA</td>
<td>U2:A9', U9:C2'</td>
<td>7.2 ± 3.5</td>
</tr>
<tr>
<td>6</td>
<td>GCguCCacU/ACguCCacA</td>
<td>C2:A9', U9:U2'</td>
<td>13 ± 5</td>
</tr>
<tr>
<td>7</td>
<td>GCguCCacU/ACguCCacA</td>
<td>C2:C9', U9:C2'</td>
<td>50 ± 13</td>
</tr>
<tr>
<td>8</td>
<td>GCguCCacU/ACguCCacA</td>
<td>C2:A9', U9:A2'</td>
<td>30 ± 1</td>
</tr>
<tr>
<td>9</td>
<td>AcGCacU/AAGuCGacA</td>
<td>A2:A9', A9:A2'</td>
<td>0.8 ± 0.3</td>
</tr>
<tr>
<td>10</td>
<td>UGcGGuC/AGuCCacC</td>
<td>U2:C9', C9:U2'</td>
<td>1.9 ± 1</td>
</tr>
<tr>
<td>11</td>
<td>AGcGGuC/AAGuCCacC</td>
<td>A2:C9', A9:U2'</td>
<td>0.2 ± 0.1</td>
</tr>
<tr>
<td>12</td>
<td>AGcGGuC/AGuCCacC</td>
<td>U2:C9', C9:G2'</td>
<td>1.5 ± 0.7</td>
</tr>
</tbody>
</table>

**Figure 3.** Investigation of optimal loop interactions. (a) Whereas the size of intermolecular kissing helix is conserved at six base-pairs in the natural DIS isolates, some of the clones isolated in the *in vitro* selection were potentially capable of forming an additional intermolecular base-pair if nucleotides at positions 2 and 9' were complementary. (b) Likewise, when complementarity existed between positions 1 and 9 of the same molecule, this could extend the stem and reduce the loop size by two nucleotides. Both of these possibilities were tested by measuring dimerization affinities and reactivities toward chemical structure probes.
DIS monomer loop size and the 1:9 base-pair

The highly conserved HIV-1 DIS loop almost always contains a hexanucleotide autocomplementary sequence flanked by two purines on the 5’ side and one purine on the 3’ side. The flanking purines are typically depicted as unpaired because they are not capable of forming canonical base-pairs across the loop. However, for both subtypes, there was a strong overrepresentation of As at position 1 and 9 of the selected aptamers. An overwhelming proportion (62 %) of selected clones presented an A1:A9 combination, as compared to A1:C9 (10 %), A1:U9 (9 %), and A1:G9 (5 %) (data not shown and reference29). This composition bias is in agreement with structural probing and mutagenesis studies that suggested that nucleotides 1 and 9 may form a non-canonical base-pair.3,10

A small proportion of the RNAs isolated by in vitro evolution contained nucleotides at positions 1 and 9 that were complementary. If these nucleotides formed a base-pair, they would in essence reduce the loop size to seven nucleotides, while extending the stem length by one base-pair (Figure 5b). To examine whether the capacity to form Watson-Crick base-pairs across positions 1 and 9 would affect dimerization, we compared the Kd’s of dimers comprised of monomer bases that could potentially form this 1:9 base-pair with those that could not. The species measured were from the subtype A CC/GG group and the subtype B AG/CU group that provided a number of clones with different 1:9 combinations.

The subtype A heterodimers with A1:U9 or A1’U9’ combinations displayed two- to fourfold decreased stability, as compared to the homodimer with A1:A9 and A1’A9’ (Table 4, pairs 1-4). These results indicated that potential intramolecular Watson-Crick base-pairing between nucleotides 1 and 9 does not increase the stability of the dimers. The same conclusion was drawn from the subtype B heterodimers, even though the range of the Kd values was narrower (Table 4, pairs 9-11). This conclusion fits with the observation that no sequenced aptamer had the potential to form a C1:G9 or G1:C9 base-pair (data not shown). Our data also showed that A1:C9 and G1:U9 combinations decreased the dimer stability, and that the effect of these combinations was additive (Table 4, pairs 5-8).

Thus, although the potential to form an intramolecular Watson-Crick base-pair between nucleotides 1 and 9 is not favored, the identity of these nucleotides can significantly influence dimer stability. This observation suggests that these nucleotides adopt a non-canonical conformation, even when their identities would suggest the potential to form Watson-Crick base-pairs.

Structural constraints on the G0:C10 base-pair closing the DIS loop

Due to occasional errors introduced during the reverse transcription and PCR steps of the in vitro evolution of the aptamers, we recovered several clones that contained deletions or substitutions at non-degenerate positions.29 Clones with deletions in the loop were not analyzed in detail, because the rules governing dimerization of these RNAs may considerably differ from those of the original DIS.

Otherwise, we noticed an unexpectedly high frequency of substitution of C10, the nucleotide immediately 3’ to the loop, in the cloned subtype B aptamers (40 %, data not shown). This variability contrasted with the almost absolute conservation (94 %) of this nucleotide in the subtype A aptamers. These data suggested that while there was a strong pressure to maintain the G0:C10 base-pair in subtype A aptamers, this pressure was apparently relaxed in subtype B.

The status of the last G:C base-pair of the stem that closes the DIS loop is of considerable interest, since in the NMR model of the subtype B DIS, these two nucleotides were seen as unpaired,12 while in the subtype A DIS, these two nucleotides have been modeled as base-paired. Thus, we tested whether subtype B aptamers containing substitutions at position 10 were able to form stable dimers (Table 5). First, we observed that

<table>
<thead>
<tr>
<th>Pair no.</th>
<th>Loop sequences of dimer pairs</th>
<th>Potential N1:N9 and N1’:N9’ base-pairs</th>
<th>Kd (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype A</td>
<td>AAguGCaCA homodimer</td>
<td>A1:A9, A1’:A9’</td>
<td>0.9 ± 0.4</td>
</tr>
<tr>
<td>2</td>
<td>AAguGCaCA/AAguGCacU</td>
<td>A1:A9, A1’:U9’</td>
<td>3.6 ± 1.5</td>
</tr>
<tr>
<td>3</td>
<td>AAguGCaCA/AAguGCacU</td>
<td>A1:A9, A1’:U9’</td>
<td>1.7 ± 0.8</td>
</tr>
<tr>
<td>4</td>
<td>AAguGCaCA/AAguGCacU</td>
<td>A1:A9, A1’:U9’</td>
<td>2.8 ± 1.2</td>
</tr>
<tr>
<td>5</td>
<td>AAguGCaCA/AAguGCacU</td>
<td>A1:C9, A1’:U9’</td>
<td>1.3 ± 0.5</td>
</tr>
<tr>
<td>6</td>
<td>AAguGCaCA/AAguGCacU</td>
<td>A1:C9, A1’:U9’</td>
<td>15 ± 5.5</td>
</tr>
<tr>
<td>7</td>
<td>AAguGCaCA/GCguGCacU</td>
<td>A1:A9, G1:U9’</td>
<td>13 ± 5</td>
</tr>
<tr>
<td>8</td>
<td>AAguGCaCA/GCguGCacU</td>
<td>A1:C9, G1:U9’</td>
<td>50 ± 13</td>
</tr>
<tr>
<td>Subtype B</td>
<td>AAgcAgGcA/AAgcCUgA</td>
<td>A1:A9, A1’:A9’</td>
<td>0.9 ± 0.4</td>
</tr>
<tr>
<td>10</td>
<td>AAgcAgGcA/UAgcCUgA</td>
<td>A1:A9, U1’:A9’</td>
<td>1.5 ± 0.7</td>
</tr>
<tr>
<td>11</td>
<td>AAgcAgGcA/UAgcCUgA</td>
<td>A1:U9, U1’:A9’</td>
<td>2.1 ± 0.9</td>
</tr>
</tbody>
</table>
substitution of C10 with U decreased the stability of the wild-type subtype B homodimer by tenfold (Table 5, lines 1 and 2, and Figure 2(a)). The magnitude of this effect is the same as we observed for substitutions of A2 (Table 2). Furthermore, we observed that the C10 to U substitution significantly increased the stability of some heterodimers (Table 5, compare lines 3 and 4). Indeed, not only C and U, but also G and A allowed formation of stable heterodimers (Table 5, lines 5 and 6). However, not all nucleotide combinations were tolerated, as shown by the poor stability of the heterodimer containing A both at position 10 and 10′ (Table 5, line 7).

**Chemical structure probing**

Whereas the interactions between the autocomplementary sequences of the two monomers comprising a dimer were predictable, the potential roles of the conserved flanking nucleotides were less clear. We used chemical structure probing to gain information regarding specific molecular contacts between nucleotides in homom and heterodimers. We paid particular attention to the possible hydrogen bonding arrangements of these flanking nucleotides, and to the structure of the base-pair closing the DIS loop. Existing models of the DIS dimer provided a framework to test the proposed base-pairing schemes. In particular, we sought to test the geometry of the proposed 1:9 non-canonical base-pair as well as a potential base triple interaction between nucleotides 2, 5′, and 6 (where 5′ represents the nucleotide at position 5 of the partner monomer). In addition, we aimed to learn about the apparent differences in structure between the subtypes A and B dimer as seen in the NMR structures, especially concerning the base-pair closing the DIS loop.

Dimethyl sulfate (DMS) was used in conjunction with hydrazine to examine the reactivity of the N3 position of cytosines. When involved in a canonical base-pair, this position is unreactive toward DMS. Likewise, the reactivity of the N7 positions of adenines and guanines was probed using diethylypyrocarbonate (DEPC) and the nickel compound Ni-CR, respectively. We focused our chemical probing efforts on aptamers of the wild-type-like groups, and the CC/GG heterodimer groups of subtypes A and B (see Figure 1). Representative probing experiments are shown here.

A consistent result of the chemical probing experiments was the decrease of reactivity of nucleotides in the autocomplementary sequence when the labeled RNA was involved in a dimeric complex relative to the monomeric form. Under high RNA dilution conditions promoting the monomeric species, the loop cytosines were reactive to DMS, whereas they were much less reactive under dimeric conditions (Figure 4). This was especially evident in the GG:CC families of clones, where several loop cytosines could be monitored simultaneously when the -CC- species was labeled (Figure 4(b)). These results show, not unexpectedly, that loop self-complementary nucleotides were unpaired in the monomer but were less reactive in the dimer, most likely because they formed Watson-Crick base-pairs.

**Reactivity of the proposed 1:9 base-pair**

DEPC probing was well suited to study the accessibility of the proposed 1:9 base-pair, since in most aptamers of A and B subtypes, these two nucleotides were adenines (see above and reference29). A conformational change involving nucleotides A1 and A9 accompanied the conversion from monomer to dimer of aptamers of subtype A (Figure 5) and B (data not shown). The same pattern of reactivity was observed for homodimers and heterodimers, irrespective of the identity of the nucleotides at positions 2, 5, and 6 (Figure 5, and data not shown). Namely, the reactivity toward DEPC of the N7 position of A9 strongly decreased upon dimerization, while the reactivity of the N7 position of A1 remained unchanged (Figure 5). The decreased reactivity of the N7 position of A9 might be due either to increased base stacking or to direct hydrogen bonding in the kissing loop complex.

We next used a combination of DEPC and DMS/hydrazine probing to gain insight into the reactivity of the 1:9 base-pair of the aptamers harboring a A1:C9 combination, which was the second most frequent pair (see above). We tested subtype A and B aptamers, belonging to the GC and GG/CC families (Figure 1). In all cases, the reactivity of

<table>
<thead>
<tr>
<th>Dimer no.</th>
<th>Sequences of the dimer loops and stem-closing base-pairs</th>
<th>Potential N0:N10 and N0′:N10′ base-pairs</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gAAgCCgCcAc homodimer</td>
<td>g0:u10, g0′:c10′</td>
<td>0.3 ± 0.14</td>
</tr>
<tr>
<td>2</td>
<td>gAAgCCgCaAu homodimer</td>
<td>g0u10, g0′:u10′</td>
<td>2.9 ± 1.3</td>
</tr>
<tr>
<td>3</td>
<td>gAAgCCgCcAc/gAAgCCgCcAc</td>
<td>g0:c10, g0′:c10′</td>
<td>0.8 ± 0.3</td>
</tr>
<tr>
<td>4</td>
<td>gAAgCCgCcAc/gAAgCCgCcAu</td>
<td>g0:c10, g0′:a10′</td>
<td>0.15 ± 0.08</td>
</tr>
<tr>
<td>5</td>
<td>gAAgCCgCcCg/gAAgCcCcCg</td>
<td>g0:c10, g0′:g10′</td>
<td>0.20 ± 0.10</td>
</tr>
<tr>
<td>6</td>
<td>gAAgCCgCcAc/gAAgCCgCcAc</td>
<td>g0:a10, g0′:c10′</td>
<td>0.33 ± 0.15</td>
</tr>
<tr>
<td>7</td>
<td>gAAgCCgCcAc/gAAgCCgCcAc</td>
<td>g0:a10, g0′:a10′</td>
<td>360 ± 180</td>
</tr>
</tbody>
</table>

The nucleotides corresponding to the original loop-closing base-pair are indicated in bold.
the N7 position of A1 towards DEPC was found to be unaffected by the monomer-dimer interconversion (data not shown). At the same time, the reactivity of the N3 position of C9 towards DMS dramatically increased upon dimerization (Figure 4(b)). This increased reactivity at the Watson-Crick position (N3) of C9, the enhanced reactivity of the Watson-Crick position (N1) of A9 previously reported for wild-type subtype A and B dimers, are strongly suggestive of a sheared geometry between these positions.

We observed no differences between subtype A and subtype B aptamers regarding the reactivity pattern of the 1:9 base-pair. Although one cannot formally exclude that the same reactivity pattern is generated by distinct conformations, our results strongly suggest that nucleotides 1 and 9 adopt the same conformation in both subtypes. Furthermore, the probing results are consistent with a sheared A1:A9 and A1:C9 conformation (see also Figure 7).

**The G0:C10 base-pair closing the stem**

Published structural data and sequence analysis of our aptamer clones (see above) suggested that the G0:C10 base-pair adopts different conformations in subtypes A and B dimers. Thus, we used DMS/hydrazine probing to compare the status of C10, and thus of the 0:10 base-pair, in subtype A and B aptamers.

The closing base-pair of the subtype B clones of the CC/GG group we tested was more reactive than the equivalent pair in the subtype A aptamers (compare C10 reactivity in Figure 4(b) versus Figure 4(a), and data not shown). The G0:C10 base-pair of the subtype A dimers appeared closed, however, inasmuch as the C was not reactive to DMS treatment (Figure 4(a), and data not shown). In addition, our data indicate that in some subtype B aptamers, the G0:C10 base-pair did not open upon dimerization, but was already disrupted in the monomer: C10 was clearly reactive in the monomeric form of the AUgcCCgcC (Figure 4(b)) and CUgcCCgcC (not shown) aptamers, but poorly reactive in the AGgcCCgcA monomer (not shown).

It should be noted that although our chemical structure probing results with the subtype B aptamers were in general agreement with the NMR model of the subtype B dimer, the stem sequences were different. The NMR study used constructs harboring a U:G base-pair at the penultimate position of the stem, which is the usual base-pair for subtype B stems, while in this study the equivalent pair was a U:A canonical pair, also found in some natural isolates. The stem used in the NMR study also significantly deviated from the wild-type sequence below the penultimate U:G base-pair, but this apparently does not affect the status of the G0:C10 base-pair.

**Cross-talk between nucleotides 2, 5, and 6**

Based on probing data of wild-type subtype A and B RNA and a limited set of mutants, we previously proposed a triple interaction between nucleotides 2, 5 and 6, or 2, 6 and 5. The exact nature of the base triple depended on the identity of these nucleotides: we proposed inter- and intra-molecular non-canonical interactions for the wild-type subtype A and B, respectively. This conclusion was supported by a co-variation analysis of our selected aptamers. Here, we used Ni-CR to probe the N7-position of guanines of a series of aptamers, mainly from the subtype A families.

Unexpectedly, we observed a strong reactivity of G5 or G6 towards Ni-CR in some of the dimers formed by the selected aptamers. In fact, it has been shown that the N7 position of G is not normally reactive in a regular RNA helix, and that modification of a paired G requires widening of the major groove. Furthermore, the reactivity of
position N7 of guanosine residues is very sensitive to base stacking.\textsuperscript{40} Thus, protection of G5 or G6 in some but not all aptamers, depending on the identity of position 2, does not unambiguously implicate direct hydrogen bonding between nucleotide 2 and the protected position at the central dinucleotide of the autocomplementary sequence.

Nevertheless, Ni-CR provided evidence for cross-talk between nucleotides at positions 2, 5, and 6 within the subtype A family of aptamers harboring a guCGac autocomplementary sequence.

The aptamer with loop sequence AAduGGAcA was radioactively labeled and probed as a monomer or paired with several other partners from the same family. Position N7 of G6 of this aptamer was protected only in the heterodimer with the loop sequence aptamer ACGuGGAcA, and not with other aptamers harboring either adenine or uracil at position 2 (Figure 6(a)). This result indicated that C2 modulated the accessibility of G6 in the kissing loop complex, \textit{via} either a direct interaction, a close proximity resulting in steric hindrance, or a decrease in the width of the major groove.
When testing the accessibility of the ACGuCGacA aptamer in complex with a series of partners varying at their N2 and N9 positions, the N7 position of G6 was protected upon heterodimerization with all partners (Figure 6(b)). Thus, it seems that in the AAguCGacA/ACguCGacA dimer, G6 of the first aptamer was indirectly protected, either as a consequence of an intramolecular interaction between the second and sixth nucleotides in the loop of the second aptamer, or by the narrowing of the major groove.

In certain other instances, we noted a lack of evidence for a specific interaction between position 2 and the 5-6 base-pair. For example, in the subtype A wild-type group, the reactivity of G5 was diminished among homodimers relative to monomers as previously observed for the wild-type sequence using DMS. However, the diminution was approximately equal regardless of the identity of position 2 (not shown). This situation contrasts with the results described above for the subtype A CG family. It suggests that N2 was inserted within the major groove of the intermolecular helix formed by the self-complementary sequences, irrespective of its identity, or that the width of the major groove significantly differed between the wild-type and CG families.

**Figure 6.** NiCR probing to measure the reactivity of position N-7 of guanosine residues. (a) A labeled subtype A clone of loop sequence AAguCGacA was incubated in the presence of Ni-CR without added unlabeled RNA (monomer lane) or with added unlabeled partner RNAs of general sequence NNguCGacN as indicated. Protection of the G6 position is specific to the heterodimerization partner bearing a C at position 2, indicating an intermolecular base-triple interaction could occur. (b) As above, except that labeled RNA ACguCGacA was incubated alone or with the indicated partners. The reactivity of G6 is uniformly diminished in all the dimers, suggesting either a decreased accessibility of G6 to NiCR or that an intramolecular hydrogen bond protects position G6.
Analysis of the 1:9 base-pair geometry by ISOPAIR

The computer algorithm ISOPAIR developed by Gautheret and Gutell was used to search for plausible geometries for non-canonical base-pairing interactions between positions 1 and 9. This program examines the phylogenetic variation or covariation between specific nucleotide positions and attempts to find isosteric solutions based on the observed allowable (as well as disallowed) nucleotide combinations. Among several of the aptamer families, this algorithm selected (as either the best fit or one of the best fits) a base-pairing geometry that was consistent with the chemical probing results.

For example, the in vitro selection yielded a wild-type-like subtype A group with the following combinations of nucleotides at positions 1 and 9: A:A, A:C, C:A, G:A (reference and above). When these combinations were input into ISOPAIR, the program generated several possible geometries offering the least perturbation in the sugar phosphate backbone (Figure 7). The first choice set of isosteric structures is consistent with the chemical structure probing results, and is consistent with a recently proposed geometry for this base-pair. This sheared geometry for the 1:9 base-pair was proposed as the best or one of the best fits for the following groups: subtype A: wt group, CC group, subtype B: wt group, CG group, CU group, UG group. In addition, ISOPAIR suggested the same geometry for several of the other groups when minority species were removed from consideration (subtype A: CG group; subtype B: GG group, CC group; not shown). It is notable that this geometry for the 1:9 pair was returned by the program for both wild-type contexts.

Discussion

The conserved DIS sequence of HIV-1 has been shown to be a functionally important motif inasmuch as mutations or deletions in this region cause defects in infectivity and replication kinetics. Because this region of the viral RNA is conserved and functionally important, it is of interest to study its structure in some detail. However, high resolution structural studies of the DIS have been hampered somewhat by technical difficulties. NMR analysis of DIS model constructs based upon the subtype A and subtype B sequences suggest two significantly different structures. Furthermore, formally proving that the observed constraints arise from a kissing complex and not an extended duplex RNA appears problematic. Attempts to crystallize DIS model constructs have been successful, but the crystals represented the extended duplex form of dimer, and not the kissing loop conformation. It is possible that the very high concentrations of RNA required for crystallization and NMR analysis favor formation of the extended duplex over the kissing complex. Whereas the extended duplex form of dimer may indeed be physiologically significant, it would most likely arise through an initial kissing complex. Here solution structure probing techniques are especially valuable because concentrations and reaction conditions are more flexible. We used these conditions together with our large library of dimerization-competent clones to explore unanswered questions regarding the conformation of the monomeric and dimeric kissing loops of the DIS.

Many of the clones used in this study had no capacity for homodimerization, but dimerized efficiently with complementary partners isolated from the same pool. In fact, the $K_a$s for these clones were in the same range as those capable of homodimerization. Clone pairs of this type were useful for biochemical investigations, because one can completely control the monomeric/dimeric state of these RNAs by adding or withholding the complementary species from the reaction mixture. Using pairs such as these, we were able to see subtle differences in the dimer stability, and to follow the
changes in nucleotide reactivities toward chemical probes during the monomer-dimer transition. These were particularly important for testing the geometry of the proposed 1:9 base-pair.

Several lines of evidence suggest that nucleotides 1 and 9 adopt a sheared configuration. First, the in vitro selection overwhelmingly favored A:A juxtapositions. Second, analysis of the selected 1:9 pairs in the various families was conducted using ISOPAIR, an algorithm that finds “best-fit” isosteric geometries for given combinations of paired nucleotides. In many clone families, including the wild-type families, ISOPAIR suggested an A-A sheared geometry (or its isosteric counterparts) as the best or one of the best choices. Third, $K_d$ measurements in several instances demonstrated that loss of the A-A juxtaposition at positions 1 and 9 was detrimental to dimer stability. These data suggest that a sheared 1:9 base-pair favors dimerization, but does not allow us to distinguish between a transient role during the dimerization process and a structural role in the resulting kissing loop complex. Chemical structure probing was used to assess the relative reactivity of atomic positions on the 1 and 9 nucleotides. Although they do not formally prove its existence, the probing results of both A1:A9 and A1:C9 pairs were consistent with the sheared geometry shown in Figure 7. Remarkably, the potential sheared base-pair closing the loop is not a peculiarity of the natural DIS, but seems to exist in all selected nine nucleotide loop variants we tested, and thus it appears as a main feature of the kissing loop complex. Interestingly, selections of other kissing complex-forming loops also revealed the importance of non-canonical base-pairs closing the loop, although each might adopt a different geometry.

The status of the open or closed nature of the last base-pair of the stem below the DIS loop is also of interest because of the apparently disparate observations regarding this base-pair between subtypes A and B kissing complexes. The mutations we observed in the cloned aptamers at position 10, which should be conserved as a C, also suggested important differences between subtype A and subtype B aptamers. Indeed, mutations at this position seemed to be tolerated in subtype B, but not in subtype A aptamers. By assessing the reactivity of C10 (involved in the base-pair that closes the DIS loop) toward DMS/hydrazine treatment, it appears that this base-pair is closed in the subtype A monomer and kissing loop dimer. On the other hand, the same base-pair appears to be open in both the subtype B monomer and dimer. However, $K_d$ measurements revealed a complex picture, even among subtype B aptamers. In some aptamers, mutation of C10 was found to have little or even stabilizing effects, while in others, including the subtype B wild-type sequence, the same mutation decreased the dimer stability by an order of magnitude. Therefore, the sequence of the loop appears to affect the structure of the stem below. Given the role that the region below the apical loop appears to have in dimerization, this may be of significance in vivo as well.

The geometry of nucleotides 2, 5, and 6 was also of interest because molecular modeling of the subtype A DIS dimer suggested an intermolecular base triple interaction between these nucleotides. In this study, using a variety of sequence variants, we observed different probing patterns in the wild-type and CG families of subtype A. Our results showed that, in the CG family, a C at position 2 was required to observe protection of the central autocomplementary nucleotides towards modifications by Ni-CR. On the other hand these nucleotides were partially protected in the wild-type family, irrespective of the identity of N2. In the latter case, our data did not allow us to discriminate between close proximity of N2, narrowing of the major groove, or direct interaction with N2. However, the $K_d$ measurements are consistent with the existence of specific cross-talk between N2 and the central autocomplementary nucleotides in the wild-type families of both subtypes. The most stable wild-type-like dimers of subtype A had an A or a C at position 2, suggesting an important role for the amino group at position 6 of adenine or 4 of cytidine. In subtype B, the most stable dimers were observed for the aptamers having a purine at position 2, suggesting that base stacking could be an important determinant in dimer stability. Thus, the present data, as well as previous results, show that N2 might adopt a number of different conformations. Stabilizing interactions appear to differ not only between subtypes A and B, but also between the aptamer families of each subtype, and even within a given family. Given its versatility, this interaction may not represent a good target for synthetic DIS ligands designed to inhibit dimerization.

This study provides new insights into the constraints that allow high affinity dimerization of HIV-1 DIS kissing loops. It points towards conserved features among the selected aptamers, such as the conformation of nucleotides 1 and 9, as well as features restricted to one of the two DIS subtypes, such as the opening of the last base-pair (0:10) of the stem, or to individual aptamers, such as the cross-talk between nucleotides 2, 5, and 6. This detailed knowledge of the features governing dimerization of the DIS loop, is a step further towards the development of antiviral strategies that exploit these characteristic interactions.

Materials and Methods

Generation of pure clonal RNAs

The RNAs used in this study were selected from randomized pools for their ability to dimerize as described previously. Individual pUC18 plasmid clones containing unique aptamer sequences were propagated in DH5α cells. Individual sequenced plasmids were prepared and used as substrates for T7 RNA polymerase transcription.
after digestion with Smal. RNAs (53 nt) were gel purified on 8% (w/v) denaturing polyacrylamide gels prior to use in dimerization and structure probing experiments.

Where necessary, RNAs were radioactively labeled with [32P]labeled RNA aptamers. Briefly, 10 pmol of RNA were mixed with 20 μCi [32P]labeled RNA aptamers and two units of T4 RNA ligase (Pharmacia) in a buffer containing 0.1 mM ATP, 50 mM Tris-HCl (pH 7.5), 1 mM MgCl2, 2 mM DTT, and 100 μg/ml BSA in a final volume of 10 μl. The mixture was incubated overnight at 4°C, and labeled product was purified on a denaturing 8% polyacrylamide gel followed by passive elution into 0.5 M ammonium acetate, phenol:chloroform extraction, and ethanol precipitation.

**Determination of Kds**

A constant amount of radioactively labeled RNA (1000-3000 cpm, approximately 0.2 -0.5 nM in the final reaction volume of 10 μl) was mixed with increasing amounts (typically 0-1000 nM) of unlabeled homologous or heterologous RNA in water. The mixtures were heated to 95°C for two minutes, then quick-chilled on ice. Buffer conditions were adjusted to 50 mM sodium cacodylate (pH 7.0), 5 mM MgCl2, 300 mM KCl by adding a 5x concentrated buffer mix and samples were incubated at 37°C for 15 minutes. Samples were then placed on ice, glycerol loading buffer was added, and were loaded onto a 5.5% acrylamide gel containing 0.1 mM MgCl2 and 45 mM Tris-borate buffer (0.5x TB). Running buffer was 0.5x TB supplemented with 0.1 mM MgCl2. Samples were electrophoresed at 4°C for two hours under constant voltage (15 V/cm). The proportion of material migrating as monomeric or dimeric RNA was quantified using a Fuji BAS 2000 phosphorimager of material migrating as a dimer, described.26 Briefly, Kds were determined using a least squares curve fit of the experimental data using equation (1) for homodimerization and equation (2) for heterodimerization, where f0 is the weight/weight fraction of material migrating on the gel as a dimer, β is the maximum fraction of RNA able to dimerize, M0 is the initial concentration of unlabeled oligonucleotide, C0 is the concentration of labeled oligonucleotide, and Kd is the equilibrium dissociation constant:

\[
K_d = \frac{(4βM_0 + K_d) - \sqrt{8M_0βK_d + K_d^2}}{4M_0} \quad (1)
\]

\[
K_d = \frac{(βC^* + M_0 + K_d - \sqrt{(βC^* + M_0 + K_d)^2 - 4βC^*M_0})}{2C^*} \quad (2)
\]

**Chemical structure probing of dimeric complexes**

The chemical structure probes DMS, DEPC,44 and Ni-CR35 were used to investigate the accessibility of particular atomic positions on the nucleotide bases of the aptamers. The sites of chemical modification were visualized directly rather than by primer extension. Briefly, 32P-radiolabeled RNA aptamers were incubated under conditions promoting dimerization, then the chemical probe was added such that, on average, less than one nucleotide was modified per molecule. Following the modification, the RNAs were incubated under conditions promoting cleavage at the modified positions. The treated RNA was then electrophoresed on a denaturing 12% polyacrylamide gel, using a partial alkaline hydrolysis ladder and/or a ribonuclease T1 hydrolysis ladder for comparison to determine the site(s) of cleavage as previously described.36,45

**Inferring isosteric nucleotide pairings using ISOPAIR**

ISOPAIR is an algorithm developed by Gautheret and Gutell42 that allows a user to input combinations of nucleotides that are observed among phylogenetic variants. The program generates probable isosteric geometries for the nucleotide combinations that would minimally perturb the sugar-phosphate backbone. The Linux version of the program was used according to documentation supplied by the authors (igs-server.cnrs-mrs.fr in directory/pub/ISOPAIR).

**Acknowledgments**

Delphine Mignot is acknowledged for skillful technical assistance. This work was supported by a grant from the Agence Nationale de Recherches sur le SIDA (ANRS). J.S.L. was a fellow of the ANRS.

**References**


HIV-1 Kissing Loop Aptamers


31. Liang, C., Rong, L., Quan, Y., Laughrea, M., Kleiman, L. & Wainberg, M. A. (1999). Mutations within four distinct gag proteins are required to restore replication of human immunodeficiency virus type 1 after deletion mutagenesis within the dimerization initiation site. *J. Virol.*, 73, 7014-7020.


38. Zhang, P., Burrows, J. & Rokita, S. E. (1998). Nickel- and cobalt-dependent reagents identify structural features of RNA that are not detected by...
dimethyl sulfate or RNase T1. Biochemistry, 37, 2207-2214.

Edited by D. E. Draper

(Received 15 March 2001; received in revised form 22 June 2001; accepted 25 June 2001)