Convergence of natural and artificial evolution on an RNA loop–loop interaction: The HIV-1 dimerization initiation site

J. STEPHEN LODMELL,¹ CHANTAL EHRESMANN, BERNARD EHRESMANN,
and ROLAND MARQUET

Institut de Biologie Moléculaire et Cellulaire, Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique, 67084 Strasbourg, France

ABSTRACT

Loop–loop interactions among nucleic acids constitute an important form of molecular recognition in a variety of biological systems. In HIV-1, genomic dimerization involves an intermolecular RNA loop–loop interaction at the dimerization initiation site (DIS), a hairpin located in the 5′ noncoding region that contains an autocomplementary sequence in the loop. Only two major DIS loop sequence variants are observed among natural viral isolates. To investigate sequence and structural constraints on genomic RNA dimerization as well as loop–loop interactions in general, we randomized several or all of the nucleotides in the DIS loop and selected *in vitro* for dimerization-competent sequences. Surprisingly, increasing interloop complementarity above a threshold of 6 bp did not enhance dimerization, although the combinations of nucleotides forming the theoretically most stable hexanucleotide duplexes were selected. Noncanonical interactions contributed significantly to the stability and/or specificity of the dimeric complexes as demonstrated by the overwhelming bias for noncanonical base pairs closing the loop and covariations between flanking and central loop nucleotides. Degeneration of the entire loop yielded a complex population of dimerization-competent sequences whose consensus sequence resembles that of wild-type HIV-1. We conclude from these findings that the DIS has evolved to satisfy simultaneous constraints for optimal dimerization affinity and the capacity for homodimerization. Furthermore, the most constrained features of the DIS identified by our experiments could be the basis for the rational design of DIS-targeted antiviral compounds.

Keywords: kissing loop complex; noncanonical interactions; RNA–RNA interaction; RNA dimer; SELEX

INTRODUCTION

Loop interactions among nucleic acids likely represent one of the most basic and ancestral forms of molecular recognition (Tomizawa, 1993). Loop–loop binding combines the fundamental base-pairing compatibilities of nucleic acids with the added specificity of a constrained three-dimensional structure, as the 5′ and 3′ ends of the loop are involved in secondary structure. The particular structure of the loops may increase specificity or binding affinity for another nucleic acid ligand, or pro-

vide a recognition site for a third-party ligand, such as a protein.

RNA loop–loop interactions have been demonstrated as essential in many biologically important systems. The RNA pseudoknot (Pleij et al., 1985), which represents the most extensively studied type of loop interaction, is a vital component in a vast array of biological processes (Dam et al., 1992; Tinoco, 1997). The RNA kissing loop complex is a special case of the pseudoknot, in which complementary sequences in apical loops of two hairpins base pair to form a stable or semi-stable structure. The kissing loop interaction was first observed in the crystal structure of tRNA^{Asp} (Moras et al., 1980, 1985) and its functional importance demonstrated in the control of replication of plasmids ColE1 and R1 (Tomizawa, 1984; Eguchi et al., 1991; Wagner & Simons, 1994). Kissing loop interactions have been

Reprint requests to: Roland Marquet, Institut de Biologie Moléculaire et Cellulaire, Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique, 15 rue René Descartes, 67084 Strasbourg, France; e-mail: r.marquet@ibmc.u-strasbg.fr.

¹Present address: Division of Biological Sciences, The University of Montana, Missoula, Montana 59812, USA.

described for a variety of other systems as well. Intermolecular kissing loop interactions are important for dimerization of retroviral RNAs, (reviewed in Paillart et al., 1996b) and multimerization of a localized mRNA (Ferrandon et al., 1997), and intramolecular interactions help to organize RNA structure within enterovirus and poliovirus genomes (Pilipenko et al., 1996; Mirmomeni et al., 1997) and within the *Neurospora* ribozyme (Rastogi et al., 1996). Although RNA loop-loop interactions often recruit a protein cofactor, the initial binding event does not require additional factors. Thus, this type of interaction represents a likely mechanism for intermolecular recognition in the proposed prebiotic RNA world.

A kissing loop interaction promotes dimerization in vitro of HIV-1 subgenomic RNA fragments (Laughrea & Jetté, 1994; Paillart et al., 1994; Skripkin et al., 1994; Muriaux et al., 1995). The essential stem-loop, located in the noncoding leader sequence some 270 nt from the 5' end of mature viral RNA, is called the dimerization initiation site (DIS). Mutations in the DIS of infectious HIV-1 molecular clones dramatically reduce the replication rate and the infectivity of the resulting viruses (Berkhout & van Wamel, 1996; Haddrick et al., 1996; Paillart et al., 1996a). A significant fraction of the genomic RNA inside the mutant virions is monomeric (Haddrick et al., 1996), and the encapsidated dimers appear abnormal when assessed by nondenaturing gel electrophoresis (Clever et al., 1996).

The nine-membered loop of the DIS contains an autocomplementary hexanucleotide sequence flanked nearly always by two 5' and one 3' purines. Interestingly, among all natural isolates of HIV-1, only two autocomplementary sequences are commonly found: either GUGCAC or GCGCGC, represented in subtypes A and B, respectively (St Louis et al., 1998; Fig. 1). Such a high degree of phylogenetic conservation in an otherwise rapidly mutating genome suggests the sequence and/or structure of the DIS are constrained for function. Furthermore, it has been determined that not all autocomplementary sequences promote dimerization in vitro (Clever et al., 1996) or in vivo (Laughrea et al., 1999).

Here, we sought to determine which sequences and structures promote dimerization. We randomized some or all of the loop positions in a model DIS, and, using an iterative selection-amplification scheme (Joyce, 1989; Ellington & Szostak, 1990; Tuerk & Gold, 1990), we characterized dimerization-competent constructs. The artificial phylogeny generated by this method elucidated the importance of loop size, complementarity, noncanonical base pairs, and tertiary interactions in the stabilization of the DIS dimer, and indeed for loop-loop interactions in general. Furthermore, we found that a small number of sequence solutions promoted homodimerization, suggesting that there is limited potential for mutation at this site.

RESULTS AND DISCUSSION

Design of model RNAs

The model DIS RNAs were constructed as described in Materials and methods. These constructs were designed to mimic the DIS stem-loop of HIV-1 except that the asymmetric bulged loop found in the lower stem of natural HIV-1 isolates was removed to eliminate the possibility that sequences might be selected that recognize the internal rather than the apical loop on a partner molecule. The stem nucleotides comprising the 7 bp below the loop are identical in sequence to those of the HIV-1 MAL isolate. The constructs were 53 nt in length with 9-nt randomized loops and 22 bp of invariant HIV-1-like stem. The selection scheme used for the in vitro evolution is presented in Figure 1B and is described in detail in Materials and methods. For each of the selections, the protocol was run in parallel using three separate samples to control for potential biases in the pool composition resulting from contamination or other artifacts.

Partial randomization of the loop

In a first set of experiments, we targeted five of the nine DIS loop positions for randomization. This allowed us to test specific inter- and intraloop interactions while maintaining the overall context of the two naturally occurring DIS loop sequences. Previous experiments suggested the existence of a noncanonical A-A base pair between positions 1 and 9 of the loop, and a base triple between nucleotide G₂ with the intermolecular C₆:G₅ base pair (Paillart et al., 1997). Accordingly, positions 1, 2, 5, 6, and 9 were randomized so that the cycle 0 (preselection) RNA population had a loop consensus sequence of NNGUNNACN (subtype A context) or NNGCNGCN (subtype B context; Fig. 1). Furthermore, the entire selection for subtype A dimerization sequences was repeated three times (varying the magnesium concentration between 0.1 and 5 mM during dimerization) with similar results.

The proportion of dimerization-competent sequences increased steadily with each round of selection and amplification (Fig. 2). The apparent binding affinity reached a plateau after 5–7 cycles in each case (the number of cycles depended on the construct and the stringency of selection). The apparent dissociation constant (K_d) for the population of RNAs approached that of the wild-type HIV-1 DIS RNA of similar size, in the low nanomolar range. After 5–7 cycles of selection and amplification, DNA intermediates were cloned into plasmid vectors and sequenced. The sequences promoting dimerization in both contexts were highly constrained and fell into several easily identifiable categories (Fig. 3). Results are based on analysis of 146 individual clone sequences. K_d measurements of selected individual clones and dimer pairs is shown in Table 1.

FIGURE 1. A: The two major sequence variants of DIS stem-loops found in different HIV-1 viral isolates are represented by subtypes A and B of HIV-1 Group M (the Major phylogenetic group). The nucleotides flanking the autocomplementary sequence are highly conserved, most commonly as adenines and almost always as purines (Korber et al., 1997). Sequence differences between subtype A and B loops are in bold, and loop nucleotides are numbered for clarity. One of the partially randomized constructs used in this study is shown at the right. This construct was designed to test potential interactions between nt 1, 2, 5, 6, and 9 while keeping the overall structure reminiscent of the Subtype A DIS. Similar constructs were made that maintained the Subtype B sequence context. Additionally, constructs were made with each of the nine loop nucleotides randomized. **B:** The selection scheme used to isolate and amplify dimerization-competent sequences is shown.

Identity constraints of the central dinucleotide

The overall conservation of C or G at positions 5 and 6 was striking, occurring 96% and 89% of the time, respectively, in subtype A, and 61% and 83% of the time, respectively, in subtype B. In both contexts, clones with A or U at *both* positions 5 and 6 were recovered with a very low frequency. However, clones harboring one A

or U at either position 5 or 6 represented half (52%) of the sequenced subtype B clones, whereas they were virtually absent from subtype A clones (Fig. 3). In subtype B, the central dinucleotide is flanked by GC base pairs, whereas in subtype A it is flanked with AU base pairs (Fig. 1). These results suggest that a minimum number of GC base pairs among the four central positions is necessary for efficient dimerization, and GC

FIGURE 2. Evolution of the apparent dissociation constants of successive pool populations of subtype A randomized RNAs. The fraction of material migrating as a dimer is plotted against the log of the concentration of the input (monomeric) RNA in the reaction mixture for cycles 0 (preselection), 1, 3, 6, and 7. The curve for nonrandomized wild-type sequence RNA is shown for comparison. Comparable results were obtained for the subtype B and whole-loop randomized selections.

pairs at the central dinucleotide are of disproportionate importance. Indeed, a subtype A clone harboring a UA central dinucleotide formed poorly stable dimers in vitro (Table 1). The same dinucleotide introduced into the center of the subtype B DIS by site-directed mutagenesis allowed formation of stable dimers and did not abrogate viral replication (Laughrea et al., 1999). The identity of the central dinucleotide could be conserved because it interacts with other nucleotides in the formation of a stable dimer (see covariation data below), and/or because the central GC base pairs serve as essential nucleation sites during the first steps of dimerization (Eguchi & Tomizawa, 1991).

Selection of homodimers and heterodimers

In both the subtype A and B construct selections, we recovered the wild-type autocomplementary sequence

TABLE 1. Measured dissociation constants (K_d) of several selected dimerization families.

Loop complementary sequence(s)	K_d (nM)
GUGCAC homodimer	6.2 ± 5.1
GUUAAC homodimer	>200
GUGGAC/GUCCAC	6.8 ± 4.7
GCGCGC homodimer	2.4 ± 1.7
GCCGGC homodimer	$<1^a$
GCGGGC/GCCCGC	$<1^a$
GCAGGC/GCCUGC	1.3 ± 0.5

^a K_d values smaller than 1 nM were near the detection limit (see Materials and methods) and are therefore not reported with greater precision.

in significant proportions (Fig. 3). This result demonstrates that the sequences retained during natural evolution of HIV-1 form among the most stable dimers and lends credence to the idea that this motif has been retained precisely for this purpose. In addition, sequences capable of homodimerization but which are not found among natural isolates of HIV-1 were identified. The predominant sequence of this type in both contexts contained an inversion of the central dinucleotide. Why these autocomplementary sequences, which form dimers in vitro with wild-type affinity (Table 1), are not found naturally probably stems from the fact that putative single-mutant intermediates between the -GC- and -CG- forms would be incapable of homodimerization. Alternatively, the wild-type sequence has been retained not only for its ability to dimerize, but also to perform another unknown function. Interestingly, binding of nucleocapsid protein to short DNA and RNA oligonucleotides indicated that this protein preferentially binds to GXC (especially GUG) sequence motifs (Fisher et al., 1998; Vuilleumier et al., 1999). This motif is present in the wild-type DIS loops, but is lost upon inversion of the central dinucleotide.

Not all clones isolated from the dimerization-competent pools were capable of *homodimerization*, but their pairing partners were readily identified in the same aptamer populations by their complementary loop sequences and were verified experimentally. By far, the CC/GG type of heterodimer was predominant. AG/CU and UG/CA as well as some other lightly represented combinations were also observed, mostly in the subtype B context (Fig. 3). Although these heterologous combinations formed stable dimers (Table 1), they would obviously be of limited utility to a virus, because a stable interaction between two such motifs would require that a single cell be infected by two different, complementary strains.

A striking observation is the very dissimilar selection of homo- *versus* heterodimers in the subtype A and B construct selections (Fig. 3). In subtype A, 55% of the clones contain a 6-nt self-complementary sequence, whereas homodimers only account for 15% of the subtype B clones. Although this result was unexpected *a priori*, it is a direct consequence of the sequence constraints at positions 5 and 6 of the loop. In the subtype A construct selection, dinucleotides containing one or two A or U are strongly disfavored. Among the four dinucleotides containing only G and/or C that are selected, two allow homodimerization. In the subtype B construct selection, at most one A or U is tolerated at position 5 or 6. Among the 12 dinucleotides selected at these positions, only two can promote homodimerization.

Identity constraints of the flanking nucleotides

Inspection of the flanking nucleotides among the different groups of dimers suggests roles for these posi-

FIGURE 3. Summary of results of in vitro selection of dimerization-competent sequences starting from HIV-1 subtypes A or B loop sequence contexts. Clones are grouped according to the identity of the central dinucleotide (circled). The percentage of the total clones of type A or type B falling into each dinucleotide group is indicated above the group name. The percentage identity of each nucleotide at positions 1, 2, and 9 of each group is indicated.

tions in the process of dimerization. First, there was not an overall tendency toward complementarity between positions 2 and 9 among homodimers or among heterodimerizing partners, which would increase the size of the autocomplementary sequence and the pairing potential between the two loops (Fig. 3 and Table 2B). Thus, the 6-nt autocomplementary span conserved among all natural isolates is also clearly the best suited to dimerization in these artificially evolved sequence contexts. Interestingly, structural studies on ColE1-based kissing hairpins (bearing inverted loop sequences relative to the natural ColE1 system) form seven loop-loop base pairs (Eguchi & Tomizawa, 1991; Gregorian & Crothers, 1995; Marino et al., 1995), and contain no flanking unpaired loop nucleotides. Thus, the optimal number of loop base pairs is influenced by the structural context surrounding the complementary sequences.

Second, the bias of position 1 toward A is remarkable (Fig. 3 and Table 2). Overall, approximately 80% of all clones had A at this position (one would expect 25% As were this position not constrained). The high incidence of A also at position 9 argues strongly for a noncanoni-

cal base pairing interaction between these two positions or at least that canonical base pairs between positions 1 and 9 are disfavored overall for formation of dimers (Table 2A). This result indicates that a 9-nt loop is better suited for this interaction than a 7-nt loop. The identity of the 1:9 pair was also influenced by the identity of the central dinucleotide, suggesting that the non-canonical closing pair of the loop was specifically tailored to the sequence and structure of the rest of the loop (Fig. 3).

Covariations of the central and flanking nucleotides

Position 2 tended toward purines in general (67% overall, 87% in subtype B), but was differently biased depending on the nucleotides present at positions 5 and 6. Computer modeling of the DIS dimer suggested an intermolecular base triple interaction between position 2 and the intermolecular 6:5' base pair in the subtype A wild-type context, and an intramolecular contact between positions 2 and 5 in subtype B (Paillart et al., 1997). Surprisingly, despite the conservation of posi-

TABLE 2. Covariation analysis of selected loop positions using sequence data from all selections.^a

A					
9	A	C	G	U	
1					
A	104	19	11	15	
C	8	2	—	3	
G	4	—	—	5	
U	1	2	—	2	
B					
9	A	C	G	U	
2					
A	68	12	8	9	
C	23	1	—	2	
G	25	—	2	3	
U	3	10	1	10	
C					
6	A	C	G	U	
2					
A	3	38	41	15	
C	—	15	5	6	
G	2	3	22	1	
U	1	9	13	1	

^aNumbers represent the number of clones having the particular combination of nucleotides shown in the row and column headings. **A:** analysis of position 1 with respect to 9; **B:** analysis of position 2 with respect to 9; and **C:** analysis of position 2 with respect to position 6.

tion 2 as G or A among natural subtype A isolates, wild-type-like variants were recovered with each of the four bases represented at position 2, and the substitution of the wild-type G₂ with A, C, or U did not adversely affect binding affinity. When the central dinucleotide was inverted, there was a strong bias toward A₂. When the central dinucleotide was GG or CC, U₂ or C₂ was prevalent, respectively (Fig. 3). Covariations were likewise observed in the subtype B context. For example, only in the CG and UG groups was G₂ highly represented; otherwise A₂ was predominant (Fig. 3).

Table 2C shows that certain combinations of nucleotides were favored in covariation analysis of position 2 and the central dinucleotide (positions 5 and 6). For example, position 6 was biased toward C and away from G when position 2 was C. This could indicate a direct interaction of position 2 with positions 5 or 6, or with positions 5' or 6' of the partner loop. Similarly, a G at position 2 correlated with a strong bias toward G at position 6. According to the available models of the subtype A DIS (Paillart et al., 1997; Dardel et al., 1998; Jossinet et al., 1999), close proximity of position 2 and the central dinucleotide in the major groove would afford the opportunity for a variety of inter- and intra-

molecular interactions, depending on the specific sequence context.

Stabilization of the dimers by canonical and noncanonical interactions

As already observed, A and U are almost totally excluded from positions 5 and 6 of the subtype A clones, and only one A or U is tolerated at these positions in the subtype B clones. This observation strongly suggests that the stability of the canonical base pairs formed by the central 6 nt of the DIS loop (positions 3–8) is a major factor governing dimerization. To evaluate its importance, we calculated the stability of the duplexes formed by the six central nucleotides of the DIS loops of subtypes A and B for all complementary dinucleotides at positions 5 and 6 (Fig. 4). Calculations were performed using the “individual nearest neighbor–hydrogen bonding” (INN-HB) model and thermodynamic parameters determined by Turner and coworkers (Xia et al., 1998). In the case of the subtype A construct selection, all hexanucleotides forming dimers with a calculated ΔG lower than -11.1 kcal/mol were selected. Conversely, the sequences forming dimers with a calculated ΔG higher than -10.2 kcal/mol were not (or marginally) selected. The calculated energy difference between selected and unselected sequences is also very small for the subtype B construct (0.7 kcal/mol; see Fig. 4). Indeed, only the most stable of the heterodimers containing one A-U base pair at positions 5 and 6 were selected. Thus, these calculations

FIGURE 4. Calculated stability of the six central base pairs of the selected and unselected dimers. The stability of the six central base-pairs of subtype A and subtype B dimers was calculated for all combinations of nucleotides at positions 5 and 6. The selected and unselected sequences are indicated for both subtypes.

confirm that the stability provided by base pairing of six central nucleotides of the DIS loop is of crucial importance for the overall stability of the dimers.

However, comparison of the subtype A and subtype B constructs reveals that the stability provided by the central six nucleotides is lower in the selected subtype A dimers than in the selected subtype B dimers (Fig. 4). Because both selections were conducted under identical conditions, the actual stability of both types of selected dimers must be very close. Thus, our data suggest that the contribution of the noncanonical interactions to the overall stability of the dimers is greater for subtype A than for subtype B by about 3 kcal/mol (Fig. 4). This conclusion is in keeping with the finding that dimerization of wild-type subtype A RNA requires magnesium binding, unlike dimerization of wild-type subtype B RNA (Jossinet et al., 1999).

Randomization of the entire loop

In the third construct, all nine loop nucleotides were randomized, which enabled us to examine the overall sequence requirements for dimerization without imposing sequence context bias. When the entire loop was randomized, the fraction of dimerization-competent sequences increased rapidly over several cycles of selection, reaching an affinity plateau after five rounds of selection (data not shown). Analysis of individual clones after several rounds of selection showed that the population of RNAs was still very complex. In fact, of all of the full-length clones retained for analysis, none had a perfect autocomplementary sequence of 6 nt, and no sequence was observed more than once, indicating that heterodimerization was prevalent in the pool. It should be noted here that there was no selection against heterodimerization in our experiments, and the first set of experiments showed that the K_d s for heterodimer pairs were roughly equivalent to those of the most stable homodimers (Table 1; J.S.L., C.E., B.E., & R.M., in prep). Therefore, in light of the fact that the pool 0 population was very diverse, with all nine positions randomized, it was to be expected that we found only heterodimers.

The large variety of heterodimerization-competent sequences in the whole-loop selection made it difficult to separate the clones into natural families. However, when the consensus identities of the loop positions were analyzed for all clones, the results (Fig. 5) confirmed several of the identity rules gleaned from the partial-loop randomization experiments above and from the natural phylogenetic variation. First, A was overrepresented at positions 1 (79%) and 9 (55%) and potential canonical base pairs between 1 and 9 were very underrepresented at 7%. Second, purines were dominant (72%) at position 2. Gs and Cs heavily predominated in positions 3–8, with C + G averaging 83%. Third, the dearth of base pairing potential between positions 1 and 9 and

FIGURE 5. Results of completely randomized loop selection. The in vitro selection was carried out as described in Materials and methods, and the results shown are based on the analysis of 29 full-length clones. The consensus HIV-1 DIS loop sequence from natural isolates is shown within the loop. Arrows indicate the identities of loop positions among clones in the artificially generated phylogeny.

2 and 9 argues that the optimal loop size is not smaller than 9 nt, and most likely indicates that a region of complementarity between loops need not be larger than 6 nt (although without specifically identifying the pairing partners for each of these sequences this point is not formally proven by the artificial phylogeny). Finally, the consensus sequence strongly resembles the naturally occurring subtype B DIS loop.

Concluding remarks

The chemotherapeutic treatment of AIDS has dramatically improved with the advent of tritherapy and the increasing number of antiretroviral drugs approved by the U.S. Food and Drug Administration (Carpenter et al., 2000). However, as a result of long-term adverse effects of therapy and the rapid emergence of multiresistant HIV-1 strains, new antiretroviral drugs are urgently needed (Carpenter et al., 2000). Given its functional role, the DIS constitutes a promising new antiretroviral target (Berkhout & van Wamel, 1996; Haddrick et al., 1996; Paillart et al., 1996a).

The constraints elucidated in the in vitro selection reveal that evolution of the natural DIS resulted from the dual constraints for high affinity binding and the ability to self-recognize. It is important to note that only a small subset of sequences satisfies these needs. In addition, our selections identified the adenines at positions 1 and 9, a purine at position 2, and the guanine and cytosine at positions 5 and 6 as the main constraints for optimal dimerization. Understanding of these constraints, especially the noncanonical interactions that may be unique to the DIS, may prove invaluable for rational drug design targeted against the DIS motif in HIV genomic RNA.

The results obtained here may be generally applicable to other RNA loop–loop interactions. Significantly, a 9-nt loop appears better suited for this type of interaction than a 7-nt loop. In addition, an 8-nt self-complementary sequence is disfavored as compared to a 6-nt sequence. At the same time, the strong selective pressure on the nucleotides flanking the self-complementary sequence indicates that the dimers are stabilized by noncanonical interactions. Selection of noncanonical interactions rather than an additional Watson–Crick base pair suggests that the noncanonical interactions stabilize RNA loop–loop interactions more than canonical pairings in this context. The importance of noncanonical base pairings in the maximal stabilization of loop–loop interactions was also apparent in the selection of aptamers binding the 6-nt TAR loop (Duconge & Toulme, 1999) or the tRNA anticodon loop (Scarabino et al., 1999).

Given the multitude of dimerization-competent clones described in this study, we can anticipate finding facsimile DIS loops in other biological systems. We expect that these systems will be very different from the well-studied loop–loop interactions involved in bacterial gene regulation (Eguchi et al., 1991; Wagner & Simons, 1994). In these systems, the kissing loops are strictly complementary and interact solely via Watson–Crick base pairs. Furthermore, the presence of a U-turn within the conserved YUNR of the loop is crucial for efficient interaction (Franch et al., 1999). The absence of this motif in the DIS-derived loops indicates that these two classes of loop–loop complexes are stabilized by different interactions. These differences might originate in the fact that in gene regulation of bacteria, the kinetics of loop–loop interaction is more important than the stability of the complexes. Thus, the natural antisense RNAs do not form the most stable kissing loop complexes with their target (Eguchi et al., 1991; Wagner & Simons, 1994; Franch et al., 1999), whereas the retroviral sequences have evolved to form more stable self-recognizing complexes.

MATERIALS AND METHODS

Production of pool 0 (preselection) RNAs

Three different pool 0 (preselection) DIS RNA transcript populations were generated. All had the same stem sequence, but they differed in their loop sequences. For the first two constructs, loop nt 1, 2, 5, 6, and 9 were randomized in both the subtype A and B contexts, giving pool 0 consensus loop sequences of 5'-N₁N₂G₃U₄N₅N₆A₇C₈N₉-3' or 5'-N₁N₂G₃C₄N₅N₆G₇C₈N₉-3', respectively (see Fig. 1). In the third construct, all nine loop nucleotides were randomized.

Pool 0 RNA transcripts were constructed by T7 transcription of PCR products degenerate at several or all loop positions. PCR oligonucleotides were from NAPS GmbH (Göttingen, Germany). PCR reactions were performed using constant 5' and 3' amplification primers (5'-GGAATTCT

AATACGACTCACTATAGGGATGTAGGACTCGCTTGCTG-3' and 5'-TCCCCCGGGACACAGGACTCGCTTGCT-3', respectively) containing a T7 promoter and *Eco*RI or *Sma*I restriction sites for accurate sizing and later cloning steps, and variable loop templates (5'-CTCGCTTGCTGTGTGCACCTCAGCAAGCGAGTCCTACATCCC-3' or 5'-CTCGCTTGCTGTGCGCGCCTCAGCAAGCGAGTCCTACATCCC-3') where several or all of the underlined nucleotides were degenerate. Loop template (0.1 pmol) was amplified with 100 pmol each of the amplification primers in a volume of 100 μ L using *Taq* polymerase. After purification, the PCR products were trimmed to uniform size using *Sma*I, repurified by phenol extraction and ethanol precipitation, and used as templates for transcription using T7 RNA polymerase to yield 53-nt-long model DIS constructs degenerate at several or all loop positions. The RNA transcripts were purified by denaturing gel electrophoresis. In addition, some of the PCR amplification products were cloned and sequenced at this step to verify the degeneracy of the indicated base positions. For example, 39 unselected (pool 0) clones of the subtype A construct were sequenced, and the distribution of nucleotides over randomized positions was found to be 37% A, 31% C, 16% G, and 16% U. Although these values deviate somewhat from the 25% theoretical value, they demonstrate that there was sufficient diversity in the starting pool to allow all combinations of nucleotides during the selection.

Selection for dimerization-competent sequences

For the in vitro selection, an aliquot of each of the pool 0 degenerate RNAs was subjected to 3' end labeling with ³²P pCp using T4 RNA ligase (Pharmacia) according to the manufacturer's protocol. The labeled RNA was purified by denaturing gel electrophoresis and was eluted into 0.5 M NH₄OAc, then phenol and phenol-chloroform extracted and ethanol precipitated. A mixture of 3,000 cpm of ³²P 3'-end-labeled RNA (corresponding to less than 1 nM labeled RNA in the final dimerization reactions) and increasing quantities of unlabeled RNA (0–1 μ g) were denatured at 90 °C in 8 μ L H₂O for 2 min, then placed on ice for 2 min. Then fivefold concentrated buffer was added to adjust salt conditions to 50 mM sodium cacodylate, pH 7.5, 300 mM KCl, and 5 mM MgCl₂ in a final volume of 10 μ L. The mixtures were incubated at 37 °C for 15 min, then placed on ice. Two microliters of cold 40% glycerol loading buffer were added just prior to loading onto a 5.5% acrylamide/45 mM tris-borate/5 mM MgCl₂ gel. Samples were electrophoresed for 2 h at 15 V/cm with 45 mM tris-borate/5 mM MgCl₂ running buffer. Bands were visualized by phosphorimager analysis and bands corresponding to the RNA dimer in lanes containing approximately 25% dimer/75% monomer (w/w) were excised from the gel. Extracted RNA was subjected to one-tube RT/PCR using *Tth* polymerase (Promega) according to the manufacturer's protocol (reverse transcription temperature was 56 °C, then 25 cycles of PCR with 46 °C hybridization temperature). The resulting PCR product was trimmed with *Sma*I and used for a subsequent round of T7 transcription. After five to seven identical cycles of transcription and RT/PCR, the DNA intermediates were cloned into pUC18 and individual clones were sequenced. Specifically, 63, 90, and 29 full-length clones were retained for comparative analysis of the subtype A and sub-

type B and whole-loop context selections, respectively. Clones having spurious deletions or insertions that made loop size or alignment ambiguous were not retained for comparative analysis.

Determination of K_d values

A constant amount of radioactively labeled RNA (1,000–3,000 cpm, approximately 0.2 nM in the final reaction volume of 10 μ L) was mixed with increasing amounts (typically 0–1,000 nM) of unlabeled homologous or heterologous RNA in water. The mixtures were heated to 95 °C for 2 min, then quick-chilled on ice. Buffer conditions were adjusted to 50 mM sodium cacodylate, pH 7.0, 5 mM $MgCl_2$, and 300 mM KCl by adding a 5 $\times$ concentrated buffer mix and samples were incubated at 37 °C for 15 min. Samples were then placed on ice, glycerol loading buffer was added, and the samples were loaded onto a 5.5% acrylamide gel containing 0.1 mM $MgCl_2$ and 45 mM Tris-Borate buffer (0.5 $\times$ TB). Running buffer was also 0.5 $\times$ TB supplemented with 0.1 mM $MgCl_2$. Samples were electrophoresed at 4 °C for 2 h under constant voltage (15 V/cm). The proportion of material migrating as monomeric or dimeric RNA was quantified using a Fuji BAS 2000 phosphorimager equipped with MacBAS software. K_d s were then calculated using the best fit of the experimental data in which the fraction dimer is plotted against the log of the unlabeled RNA concentration in the reaction mixtures. Because the concentration of labeled RNA in the reactions was approximately 0.2 nM, we could not accurately measure K_d values less than about 1 nM, because the first points on the curve fit would already be significantly dimeric. In such instances, K_d s are reported as <1 nM.

ACKNOWLEDGMENTS

We thank Delphine Mignot for expert technical assistance and Eric Westhof and Jean-Christophe Paillart for critical reading of the manuscript. This work was supported by grants from the Agence Nationale pour la Recherche sur le SIDA (ANRS).

Received March 30, 2000; returned for revision

May 16, 2000; revised manuscript received June 20, 2000

REFERENCES

- Berkhout B, van Wamel JL. 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol* 70:6723–6732.
- Carpenter CCJ, Cooper DA, Fischl MA, Gatell JM, Gazzard BG, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JSG, Richman DD, Saag MS, Schechter M, Schooley RT, Thompson MA, Vella S, Yeni PG, Volberding PA. 2000. Antiretroviral therapy in Adults. Updated recommendations of the international AIDS Society–USA panel. *JAMA* 283:381–391.
- Clever JL, Wong ML, Parslow TG. 1996. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* 70:5902–5908.
- Dam E, Pleij K, Draper D. 1992. Structural and functional aspects of RNA pseudoknots. *Biochemistry* 31:11665–11676.
- Dardel F, Marquet R, Ehresmann C, Ehresmann B, Blanquet S. 1998. Solution structure of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res* 26:3567–3571.
- Duconge F, Toulme JJ. 1999. In vitro selection identifies key determinants for loop–loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1. *RNA* 5:1605–1614.
- Eguchi Y, Itoh T, Tomizawa J. 1991. Antisense RNA. *Annu Rev Biochem* 60:631–652.
- Eguchi Y, Tomizawa J. 1991. Complexes formed by complementary RNA stem-loops: Their formation, structures and interaction with ColE1 Rom protein. *J Mol Biol* 220:831–842.
- Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822.
- Ferrandon D, Koch I, Westhof E, Nusslein-Volhard C. 1997. RNA–RNA interaction is required for the formation of specific bicoid mRNA 3' UTR–STAUFEN ribonucleoprotein particles. *EMBO J* 16:1751–1758.
- Fisher RJ, Rein A, Fivash M, Urbaneja MA, Casas-Finet JR, Medaglia M, Henderson LE. 1998. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J Virol* 72:1902–1909.
- Franch T, Petersen M, Wagner EG, Jacobsen JP, Gerdes K. 1999. Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J Mol Biol* 294:1115–1125.
- Gregorian RJ, Crothers DM. 1995. Determinants of RNA hairpin loop–loop complex stability. *J Mol Biol* 248:968–984.
- Haddrick M, Lear AL, Cann AJ, Heaphy S. 1996. Evidence that a kissing loop structure facilitates genomic RNA dimerization in HIV-1. *J Mol Biol* 259:58–68.
- Jossinet F, Paillart J-C, Westhof E, Hermann T, Skripkin E, Lodmell JS, Ehresmann C, Ehresmann B, Marquet R. 1999. Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA* 5:1222–1234.
- Joyce GF. 1989. Amplification, mutation and selection of catalytic RNA. *Gene* 82:83–87.
- Korber B, Hahn B, Foley B, Mellors J, Leitner T, Myers G, McCutchan F, Kuiken C., eds. 1997. *Human retroviruses and AIDS 1997: A compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos, New Mexico: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
- Laughrea M, Jetté 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.
- Laughrea M, Shen N, Jetté L, Wainberg MA. 1999. Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: Role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* 38:226–234.
- Marino JP, Gregorian RJ, Csankovszki G, Crothers DM. 1995. Bent helix formation between RNA hairpins with complementary loops. *Science* 268:1448–1454.
- Mirmomeni MH, Hughes PJ, Stanway G. 1997. An RNA tertiary structure in the 3' untranslated region of enteroviruses is necessary for efficient replication. *J Virol* 71:2363–2370.
- Moras D, Comarmond M, Fischer J, Weiss R, Thierry J, Ebel J, Giege R. 1980. Crystal structure of yeast tRNA Asp. *Nature* 288:669–674.
- Moras D, Dock A, Dumas P, Westhof E, Romby P, Ebel J, Giege R. 1985. The structure of yeast tRNA(Asp). A model for tRNA interacting with messenger RNA. *J Biomol Struct Dyn* 3:379–393.
- Muriaux D, Girard PM, Bonnet-Mathonière B, Paoletti J. 1995. Dimerization of HIV-1 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.
- Paillart JC, Berthou L, Ottmann M, Darlix JL, Marquet R, Ehresmann B, Ehresmann C. 1996a. 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.
- Paillart JC, Marquet R, Skripkin E, Ehresmann B, Ehresmann C. 1994. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J Biol Chem* 269:27486–27493.
- Paillart JC, Marquet R, Skripkin E, Ehresmann C, Ehresmann B. 1996b. Dimerization of retroviral genomic RNAs: Structural and functional implications. *Biochimie* 78:639–653.
- Paillart JC, Westhof E, Ehresmann C, Ehresmann B, Marquet R.

1997. Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol* 270:36–49.
- Pilipenko EV, Poperechny KV, Maslova SV, Melchers WJ, Slot HJ, Agol VI. 1996. *Cis*-element, oriR, involved in the initiation of (–) strand poliovirus RNA: A quasi-globular multi-domain RNA structure maintained by tertiary (“kissing”) interactions. *EMBO J* 15: 5428–5436.
- Pleij CW, Rietveld K, Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* 13:1717–1731.
- Rastogi T, Beattie TL, Olive JE, Collins RA. 1996. A long-range pseudoknot is required for activity of the *Neurospora* VS ribozyme. *EMBO J* 15:2820–2825.
- Scarabino D, Crisari A, Lorenzini S, Williams K, Tocchini-Valentini GP. 1999. tRNA prefers to kiss. *EMBO J* 18:4571–4578.
- Skripkin E, Paillart JC, Marquet R, Ehresmann B, 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.
- St Louis D, Gotte D, Sanders-Buell E, Ritchey D, Salminen M, Carr J, McCutchan F. 1998. Infectious molecular clones with the non-homologous dimer initiation sequences found in different subtypes of human immunodeficiency virus type 1 can recombine and initiate a spreading infection in vitro. *J Mol Biol* 72:3991–3998.
- Tinoco IJ. 1997. From RNA hairpins to kisses to pseudoknots. *Nucleic Acids Symp Ser* 1997:49–51.
- Tomizawa J. 1984. Control of ColE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* 38:861–870.
- Tomizawa J. 1993. Evolution of functional structures of RNA. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp 419–445.
- Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510.
- Vuilleumier C, Bombarda E, Morellet N, Gerard D, Roques BP, Mely Y. 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: A fluorescence study. *Biochemistry* 38:16816–16825.
- Wagner EG, Simons RW. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Annu Rev Microbiol* 48:713–742.
- Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* 37:14719–14735.