Molecular dynamics reveals the stabilizing role of loop closing residues in kissing interactions: comparison between TAR–TAR* and TAR–aptamer

François Beaurain¹,²,*, Carmelo Di Primo¹,³, Jean Jacques Toulmé¹,³ and Michel Laguerre¹,²

¹Institut Européen de Chimie et Biologie, ²CNRS UMR 5144, 16 Avenue Pey Berland, F-33607 Pessac Cedex, France and ³INSERM U386, IFRR66 Pathologies Infectieuses et Cancers, Université Victor Segalen, 146 Rue Léo Saignat, F-33076 Bordeaux Cedex, France

Received October 31, 2002; Revised January 21, 2003; Accepted May 10, 2003

ABSTRACT

A RNA aptamer (R06) raised against the trans-activation responsive (TAR) element of HIV-1 was previously shown to generate a loop–loop complex whose stability is strongly dependent on the selected G and A residues closing the aptamer loop. The rationally designed TAR* RNA hairpin with a loop sequence fully complementary to the TAR element, closed by U,A residues, also engages in a loop–loop association with TAR, but with a lower stability compared with the TAR–R06 complex. UV absorption monitored thermal denaturation showed that TAR–TAR*(GA), in which the U,A kissing residues were exchanged for G,A, is as stable as the selected TAR–R06 complex. Consequently, we used the TAR–TAR* structure deduced from NMR studies to model the TAR–R06 complex with either GA, CA or UA loop closing residues. The results of the molecular dynamics trajectories correlate well with the thermal denaturation experiments and show that the increased stability of the GA variant results from an optimized stacking of the bases at the stem–loop junction and from stable interbackbone hydrogen bonds.

INTRODUCTION

Loop–loop interactions between RNA hairpins participate in the regulation of a variety of cellular processes, including gene expression (1). In Escherichia coli, replication of the CoIE1 plasmid is regulated by the interaction of two RNA transcripts, RNA I and RNA II, that fold as hairpins (2,3). The interaction starts with base pairing between the complementary loops of these RNAs and leads to the formation of a double-stranded RNA along the entire length of RNA I that disrupts RNA II hybridization with the template DNA required for replication (4). In 1995, Marino et al., published the structure of the RNA II–RNA II duplex derived from the RNA I–RNA II complex by inverting the loop orientation (5,6). The structure revealed three A-type helices that stack from the RNA IIi stem helix to the RNA III one through the loop–loop helix. All seven residues of the loops are involved in base pairing. This implies structural torsion at the two stem–loop junctions to cross the major groove of the loop–loop helix.

Several other structures of the kissing complex were solved by NMR spectroscopy or crystallography. The first crystallographic structure of a kissing complex was the anticondon–anticodon interaction seen in rRNA (7). Another example is found in the RNA of the 50S ribosomal subunit between residues 412–428 and 2438–2454 (8). Both loops are made of 7 nt, but only six are engaged in base pairing. The unpaired residue provides the link between the loop and the stem to cross the major groove of the loop–loop helix. In 2000, the solution structure of the H3 hairpin dimer from the moloney murine leukemia virus genome was published (9). These 5′-GACG-3′ tetraloops interact through two GC base pairs only. The complex is further stabilized by stacking interactions of the unpaired G and A residues. More recently, the dimerization initiating site (DIS) structure of the HIV-1 genome was solved by crystallography (10) and offers an example of the largest loop–loop interaction of a known structure. The DIS loop contains 9 nt, six of which define an auto-complementary sequence flanked by three conserved purine residues. This structure differs slightly from the NMR one (11) published in 1998 in which the kissing complex is kinked and the unpaired purine residues bulged in.

In 1994, Chang and Tinoco (12,13) published the solution structure of the TAR16–TAR16 kissing complex. The trans-activation responsive (TAR) element of HIV-1 is a 59 nt hairpin located at the 5′ untranslated end of all retroviral mRNAs. This RNA, which plays a crucial role during the life cycle of the virus, trans-activates the transcription of its genome (14). A truncated version of TAR, TAR16, was used in the NMR studies and GC base pairs were added in the stem to stabilize the stem–loop structure. TAR16 is a RNA hairpin, showing a 4 bp stem with a 5′-UCCCAG-3′ loop sequence fully complementary to the TAR one, rationally designed by the investigators for the sake of structural studies. It turned out that this loop connected to a different stem is found within the gag gene of HIV-1 (nucleotides 1096–1109). No information

*To whom correspondence should be addressed. Tel: +33 5 40 00 22 12; Fax: +33 5 40 00 22 15; Email: f.beaurain@iecb-polytechnique.u-bordeaux.fr

Nucleic Acids Research, Vol. 31 No. 14 © Oxford University Press 2003; all rights reserved
TAR*16 adopts an overall three-dimensional conformation. The TAR 16–TAR*16 solution structure was used as the stabilizing role of the residues closing the aptamer loop. In the work presented herein, MD was used to rationalize the behavior of HIV-1 TAR complexed with an argininamide residue that mimicks Tat protein binding to the viral RNA was investigated at physiological magnesium concentration by systematic permutation (21). It was shown for R0624, the aptamer of highest affinity, that the nature of the residues closing the loop of the aptamer was crucial for kissing complex stability at low magnesium concentrations: the selected GA combination leads to the most stable complex while the CA one, for instance, drastically decreases the stability of the kissing complex formed between TAR and the aptamer. TAR has already been studied by molecular dynamics (MD), but only for its interaction with the Tat protein. The behavior of HIV-1 TAR complexed with an argininamide residue that mimicks Tat protein binding to the viral RNA was investigated (22). MD simulations of the BIV Tat–TAR complex were also run to understand the role of a key U-A-U base triple at the TAR–Tat interface (23). More recently two mutants of the HIV-1 DIS kissing loop structures have been identified by Beglov and Roux (28). All these published structures are made of two interacting RNA hairpins. In 2000, the first structural analysis of a DNA–RNA heterodimer kissing complex formed between the TAR RNA and a DNA aptamer (17) was published (18). The structure, still not entirely solved, shares with the RNA–RNA kissing complexes some common features, such as the quasi-continuous stacking from one stem to the other through the loop–loop helix. More recently Barbault et al., published the DNA DIS–DIS kissing complex structure (19), which differs from its RNA counterpart. Only 4 bp are involved in the loop–loop interaction; moreover, the loop–loop helix and the stem helices are not coaxial.

In an attempt to artificially control the gene expression of HIV-1, RNA aptamers specific for the TAR element of HIV-1 were identified by in vitro selection (SELEX) (20). The high affinity aptamers isolated at the end of the selection process were shown to fold as imperfect hairpins and form kissing complexes with the targeted RNA. The apical loop of all these aptamers corresponds to the 5'-GUCCAGA-3' consensus sequence, the six central bases of which are complementary to the entire TAR loop, whereas two G and A residues close the loop. The role of these conserved G and A residues was investigated at physiological magnesium concentration by systematic permutation (21). It was shown for R0624, the aptamer of highest affinity, that the nature of the residues closing the loop of the aptamer was crucial for kissing complex stability at low magnesium concentrations: the selected GA combination leads to the most stable complex while the CA one, for instance, drastically decreases the stability of the kissing complex formed between TAR and the aptamer. TAR has already been studied by molecular dynamics (MD), but only for its interaction with the Tat protein. The behavior of HIV-1 TAR complexed with an argininamide residue that mimicks Tat protein binding to the viral RNA was investigated (22). MD simulations of the BIV Tat–TAR complex were also run to understand the role of a key U-A-U base triple at the TAR–Tat interface (23). More recently two mutants of the HIV-1 DIS kissing loop structures have been investigated by MD (24). This is the first example of MD study of a kissing complex. The authors tried to gain a deep insight of the whole kissing complex formation and dimerization process.

In the work presented herein, MD was used to rationalize the stabilizing role of the residues closing the aptamer loop. The TAR16–TAR*16 solution structure was used as the starting structure to model the TAR–R0624 kissing complex. The model was first validated by thermal denaturation experiments. Then, MD simulations were run under conditions where the kissing complex was alternatively highly stable or unstable by using TAR*16 variants mutated in the loop closing residues. The results show that the increased stability of the selected TAR–aptamer complex over the TAR16–TAR*16 one is due to an optimized stacking and hydrogen bonding of the residues at the stem–loop junctions of the interacting hairpins.

MATERIALS AND METHODS

Oligonucleotides preparation

All RNA molecules were synthesized on an Expedite 8908 synthesizer and purified by electrophoresis on denaturing polyacrylamide gels. The pure samples were desalted on Sephadex G-10 spin columns. To avoid repeated thawing and freezing of the stock solutions, the samples were aliquoted at a volume and a concentration suitable for each experiment and stored at 20°C. Prior to the experiments, each RNA sample was heated at 95°C for 1 min and then put on ice for 10 min to avoid the formation of intermolecular complexes.

Thermal denaturation of RNA complexes

RNA complexes were prepared in 20 mM sodium cacodylate buffer, pH 7.3 at 20°C, containing 140 mM potassium chloride, 20 mM sodium chloride and 0.3 mM magnesium chloride (R’ buffer). Thermal denaturation was monitored in a Cary 1 spectrophotometer interfaced with a Peltier effect device that controls temperature within ±0.1°C. RNA samples were prepared at 1 μM final concentration in the mixture. They were mixed at room temperature and allowed to interact for 20 min before cooling down to 4°C. The experiment then started after 20 min at this temperature. Denaturation of the samples was achieved by increasing the temperature at 0.4°C/min from 4 to 90°C and was followed at 260 nm. A cuvette containing R’ buffer was used as the reference.

Computational procedure

Simulations were carried out using the CHARMM program (25) with periodic boundary conditions. The CHARMM27 parameters for RNA (26) were used along with the CHARMM-modified TIP3P water model (27) and the sodium parameters from Beglov and Roux (28).

The starting coordinates were extracted from the NMR structure of the TAR16–TAR*16 kissing complex (13) (PDB code 1KIS). The RNA complex was surrounded by 30 Na+ counterions in order to neutralize the system and by 4503 TIP3P water molecules filling a 75 × 45 × 45 Å3 orthorhombic box. The counterions and the water were placed randomly according to a pre-equilibrated box (http://www.psc.edu/general/software/packages-charmm/tutorial/dna.html). Molecules distant by less than 1.6 Å from the RNA were removed. To treat the long-range electrostatic interactions, the particle mesh Ewald method (PME) was used (29). The charge grid spacing was chosen to be close to 1 Å. A 10 Å atom-based switch truncation method was applied to the Lennard Jones interactions. Simulations were carried out at constant volume and temperature using a 2 fs time step. The non-bonded list was updated every 10 steps and the SHAKE algorithm was used to constrain the X-H bonds (30).

The equilibration protocol started with the RNA structure fixed by 500 steps of steepest descent and 500 steps of adopted basis Newton–Raphson minimization followed by a 60 ps dynamic at 300 K to allow the water molecules to relax. Then, another set of minimizations was applied without constraints
on the solute and five 60 ps dynamics were run with positional constraints on the RNA ranging from 1 to 10 \(\text{Kcal/mol/Å}^2\). Finally, no positional constraints were applied to the system and the MD was resumed for a total simulation time of 3 ns. The calculation lasted ~3 weeks of CPU time on a quadiprocessor SGI R12000 Origin200 server.

The InsightII program (Accelrys Inc.) was used to generate the two TAR\(^*\)\(_{16}\) mutants, TAR\(^*\)\(_{16}(\text{GA})\) and TAR\(^*\)\(_{16}(\text{CA})\). Residue U\(^*5\) was replaced by a guanine or a cytosine, respectively (Fig. 1b). A short minimization was applied on these RNA variant hairpins with all nucleotides fixed except nucleotides closing the aptamer loop. Finally, MD of the complexes formed between TAR and these mutants were run under the conditions defined for TAR\(_{16}-\text{TAR}^*\)\(_{16}\).

All structure analyses were performed within the Correl module of CHARMM27 (25) and the X3DNA program (31). Static structures and trajectory dynamics were visualized using InsightII (Accelrys Inc.).

RESULTS

Thermal denaturation analysis of RNA complexes

In a previous work (20), R06\(_{24}\), the best aptamer arising from selection against the HIV-1 TAR RNA was proved to fold as a hairpin and to form a kissing complex with the target (Fig. 1a).

Surprisingly, the six central bases of the loop complementary to the TAR apical loop sequence were systematically flanked by two selected G and A residues. It was then proved that these G and A residues closing the loop are a key structural determinant that drives the stability of the TAR–aptamer complex (21). It was also demonstrated that these residues explained the increased stability of the complex formed with the aptamer over that formed with the TAR\(^*\)\(_{16}\) ligand at physiological magnesium concentration.

To ascertain whether TAR\(_{16}-\text{TAR}^*\)\(_{16}\) can be used as the starting structure to model the TAR–R06\(_{24}\) RNA aptamer kissing complex and to analyze the crucial role of the GA residues, an aptamer-like ligand of TAR was derived from TAR\(^*\)\(_{16}\) by replacing the TAR\(^*\)\(_{16}\) uracil 5 residue by a guanine (Fig. 1). Three other combinations of the residues closing the loop, GC, CU and CA, that were shown to decrease the stability of the TAR–aptamer complex (21) were also introduced in TAR\(^*\)\(_{16}\). These TAR\(^*\)\(_{16}\) variants were then evaluated for binding to TAR\(_{16}\), by thermal denaturation experiments. The melting temperatures (\(T_m\)) of the bimolecular complexes formed between the TAR\(^*\)\(_{16}\) variants and TAR\(_{16}\) are presented in Table 1.

When uracil 5 of TAR\(^*\)\(_{16}\) is replaced by a guanine residue, generating a GA instead of the parental UA combination, the \(T_m\) of the TAR\(_{16}-\text{TAR}^*\)\(_{16}\) complex increases from 18.4 to 32.5\(^\circ\)C. This increased stability almost matches that observed between miniTAR–R06\(_{24}\)(UA) and miniTAR–R06\(_{24}\)(GA) aptamer complexes (Table 1). Interestingly, the UA→AU inversion of the TAR\(^*\)\(_{16}\) loop closing residues has no significant effect on the stability of the TAR\(_{16}-\text{TAR}^*\)\(_{16}\) complex. The other inversions of closing residues minimally affect the stability of the TAR–aptamer complex (Table 1). A canonical GC closing pair increases the intrinsic stability of the TAR\(^*\)\(_{16}\) stem (data not shown) but does not increase the stability of TAR\(_{16}-\text{TAR}^*\)\(_{16}\), in perfect agreement with the results obtained with miniTAR and R06\(_{24}\). CA and CU introduced as closing residues destabilize the TAR\(_{16}-\text{TAR}^*\)\(_{16}\) complex to such a degree that the \(T_m\) could not be determined.

### Table 1. Melting temperature (\(T_m\)) of RNA kissing complexes

<table>
<thead>
<tr>
<th>Loop closing residues of TAR(^*)(<em>{16}) or R06(</em>{24}) ligands</th>
<th>(T_m) (°C) TAR(<em>{16}-\text{TAR}^*)(</em>{16}) complex</th>
<th>miniTAR–R06(_{24}) complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA</td>
<td>18.4 ± 0.4</td>
<td>29.9 ± 0.6</td>
</tr>
<tr>
<td>AU</td>
<td>17.6 ± 1.5</td>
<td>31.4 ± 0.6</td>
</tr>
<tr>
<td>GC</td>
<td>17.3 ± 0.6</td>
<td>Not detected</td>
</tr>
<tr>
<td>CG</td>
<td>11.7 ± 1.1</td>
<td>Not detected</td>
</tr>
<tr>
<td>CU</td>
<td>Not detected</td>
<td>16.8 ± 1.1</td>
</tr>
<tr>
<td>UC</td>
<td>Not detected</td>
<td>21.0 ± 0.0</td>
</tr>
<tr>
<td>CA</td>
<td>Not detected</td>
<td>11.4 ± 0.5</td>
</tr>
<tr>
<td>AC</td>
<td>32.5 ± 0.8</td>
<td>47.3 ± 0.3</td>
</tr>
<tr>
<td>GA</td>
<td>26.0 ± 0.7</td>
<td>42.8 ± 0.4</td>
</tr>
</tbody>
</table>

All experiments were performed at a concentration of 1 \(\mu\)M each RNA. The experiments with TAR\(_{16}-\text{TAR}^*\)\(_{16}\) were performed at 0.3 mM Mg\(^{2+}\). As an indication, melting temperatures of miniTAR–R06\(_{24}\) kissing complexes at 1 mM Mg\(^{2+}\) extracted from the work of Ducongé et al. (21) are given in the right column. Melting temperatures are the average and standard deviation of two or three experiments.
again in agreement with the destabilizing effect previously reported for the miniTAR–aptamer complex (21). These results indicate that rather than the presence of a purine or a pyrimidine residue at position 5 of TAR*, the combination of the loop closing residues is the key factor for the stability of the kissing complex.

Finally, to demonstrate that the loop complementarity between TAR16 and TAR*16 plays a key role, as shown for the TAR–aptamer complex (21), TAR*16 was evaluated for binding to TARmal, a variant that differs by a U7®C7 mutation in the TAR loop. No complex formation could be observed with any of the TAR*16 variants (data not shown). If the complementarity between the loops were restored by a compensatory mutation on the TAR*16 side (A10®G10), a bimolecular complex was detected in the presence of TARmal (data not shown).

The thermal denaturation experiments clearly demonstrate that the TAR16–TAR*16 and TAR–R06 complexes behave similarly. In particular, the residues closing the loop play an equivalent role in the complex, driving the stability of the interaction. Consequently, TAR16–TAR*16 can be used as a starting structure to model the TAR–aptamer kissing complex and to study the structural role of the G and A closing residues.

The TAR16–TAR*16 molecular dynamics study

The simulated structure is more regular than the solution structure. The average root mean square positional deviation (RMSD) of all non-hydrogen atoms of the RNA structure after simulation against the starting structure is found to be ~2.7 Å. The first major difference observed between the starting structure and that after simulation is the angle between the two stem helix main axes which rapidly decreases during the MD run (Fig. 2). The initial structure derived from the NMR spectroscopic study bends towards the major groove of the loop–loop helix with an angle of ~30°. This curvature can be observed in every kissing complex structure derived from NMR spectroscopy experiments (6,9,11,12). During the simulation run the curvature disappeared and was replaced during the first nanosecond of simulation by an almost straight coaxial structure, as observed in the crystal structure of the 23S RNA kissing complex (8) and in the DIS–DIS one (10). The distance between the barycenter of each phosphate atom cluster (a 7 Å equilateral triangle formed between phosphates 5–7 on each strand) is then reduced, decreasing from 12 Å in the initial structure to 11 Å in the simulated one.

As observed in some oligonucleotide structures determined by NMR spectroscopy, the starting structure of TAR16–TAR*16 base pairs adopts a conformation far from canonical ones. Nucleotides have dihedral angles far from the value expected in the classical A-form and base pairs present abnormal values of intrinsic parameters (shear, buckle, stretch, propeller twist, stagger and opening). For instance, only 30 out of the possible 43 hydrogen bonds between the 16 Watson–Crick base pairs (11 GC and 5 AU) were present in the published structure (13). However, after a few picoseconds of MD the dihedral angles were close to values expected for the RNA A-form and all 43 expected hydrogen bonds were present. Hence, compared to the starting structure, the simulated one offers a more regular and canonical one.

A new undescribed interbackbone hydrogen bond. A secondary structure of the TAR16–TAR*16 duplex is presented in Figure 3. Compared to other kissing complexes (DIS–DIS, 23S and H3–H3, see Introduction), for which at least one unpaired residue provides the link between the stem and the loop–loop helix, all residues from either the TAR or TAR* loops are paired. This implies severe structural constraints at the stem–loop junctions to cross the major groove of the loop–loop helix. In TAR*16 these constraints are initially sustained by the α*6, β*6 and ξ*5 dihedral angles (see Table 2). These

---

**Figure 2.** Stereo views of TAR16–TAR*16. Starting from top left, structures at times 0, 360, 800 and 3000 ps. TAR is represented in blue and TAR* in red. After 800 ps, the structure is almost linear and remains linear until the end of the run.
angles change during the simulations and new values of $\alpha$ and $\beta$ angles allow the reorientation of atom O$_{1p}$ from the phosphate moiety of residue U$^*$5 towards the sugar group of residue C6. This change, coupled with the closer distance seen between the phosphate clusters, results in the formation of a hydrogen bond between atoms U$^*$5(O$_4$) and A$^*$12(N$_7$) that was not described in the published structure (12). This hydrogen bond, which appeared at the end of the equilibration step, is not described in the published structure (12). This hydrogen bond. The U$^*$6/A11 pair also opens at time 2.5 ns and adopts a conformation with a bridged water molecule between the two residues. Base pair opening is not supposed to occur during MD on a nanosecond timescale (32–34). However, when the solution structure of the TAR$_{16}$–TAR$^*$$_{16}$ kissing complex was solved by NMR spectroscopy (13), distance constraints between U$^*$6/A11 and U$^*$5/A$^*$12 base pairs residues were explicitly added by the authors, strongly indicating that these pairs are opened or not stable and consequently could open during the simulation. As mentioned above, residues 5 and 6 of each strand support all the constraints to connect the different helices. These residues are involved in AU and GC base pairs on each side of the stem–loop junction. The G$^{*11}$/C6 and G12/C5 pairs remained in a Watson–Crick conformation as indicated by the RMSD (Fig. 4), while the A$^*$12/U$^*$5 and U$^*$6/A$^*$11 pairs deviate from canonical distances. Structural freedom to absorb the torsion is then provided by UA pairs, thermodynamically less stable than GC ones.

The TAR$_{16}$–U$^*$5/A$^*$12 closing pair opening at time 2500 ps is correlated with a high interglycosidic distance between the two C$_1^*$ atoms of both sugars of the base pair. Indeed, during the last 500 ps, A$^*$12(C$_1^*$)/U$^*$5(C$_1^*$) shows a surprisingly large distance of ~14 Å. All other C$_1^*$–C$_1^*$ distances remained between 10.5 and 10.9 Å, as expected for canonical base pairs. This opening is also correlated with the reorganization of the dihedral angles of the stem–loop junction on the TAR side. Figure 5 shows the superimposition of 20 structures of the last 500 ps of dynamics for the A$^*$12/U$^*$5 base pair with its first hydration shell. The same water molecule bridges atoms U$^*$5(O$_4$) and A$^*$12(N$_7$) over the 500 ps. A hydrogen bond

**Figure 4.** Average RMSD of each base pair versus an ideal AU or GC with error bars. On the x-axis, the names of the different base pairs are ordered according to their relative position in space, starting from the TAR helix on the left to the loop–loop helix and finally the TAR$^*$ helix on the right. RMSD for the mutated GA and CA TAR$^*$12–TAR$^*$5 were calculated against the mean conformation over the last nanosecond of simulation.
between U*5(N3) and A*12(N6) is also observed. This conformation agrees well with the one-dimensional imino spectrum (12) that showed no U*5 imino proton peak. This absence can be interpreted as an exposure to the solvent of the imino proton, as suggested in our structure (Fig. 5).

This largest A*12(C1¢)/U*5(C1¢) distance also allows a continuous stacking of residues U*5 and C6 which was interrupted in the previously published structure (12), where only residues C6 and G*11 can stack due to the very high twist between A*12/U*5 and G*11/C6. The distance A*12(C1¢)–U*5(C1¢) being larger, A*12/U*5 adopts a different position in space allowing a continuous stacking over the junction between the TAR*16 stem and the loop–loop helix (Fig. 6). This results in a 2.7 kcal/mol gain for the stacking energy between the initial and the final structure (Table 3). On the other hand, the opening of the A*12/U*5 pair breaks the stacking between residues A*12 and its neighbors, C*13 and G*11. On the TAR strand, in contrast, the distance C5(C1¢)–G12(C1¢) remained unchanged, the base pair kept its canonical conformation and no significant larger stacking between residues C5 and U*6 (Fig. 6 and Table 3) was observed.

It should be noted that, as for crystal structures, which can be distorted by crystal artifacts, structures from NMR spectroscopy may not provide unambiguous data. As all dynamic trajectories depend on the starting structure, there is always a doubt about the conformation obtained during the run. So, in order to limit this sampling artifact, three other MD runs (1.2, 1.5 and 3 ns) were also performed with slightly different protocols. The three other MD simulations were performed starting from the same box and the same structure but for the first 3 ns one the integration time step used was 1.5 fs instead of 2 fs. For the remaining two, one was performed according to the procedure described in Materials and Methods but with a different value for the NTRFRQ parameter within the CHARMM package. NTRFRQ is the step frequency for stopping the rotation and translation of the molecule. Values of 2000 and 2000 were tried instead of 2000 in all other dynamics, without any noticeable effects. All of them confirm the weakness of the A*12/U*5 and A11/U*6 base pairs. Moreover, they all exhibit a continuous stacking on the TAR*16 side and a particularly large A*12/U*5 C1¢–C1¢ distance. We particularly focused on this large interglycosidic distance and on the stabilizing effect of purine–purine combinations for kissing complexes. In all known GA base

Table 3. Intermolecular stacking energy at the junctions

<table>
<thead>
<tr>
<th></th>
<th>CA final</th>
<th>UA final</th>
<th>GA final</th>
<th>UA initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAR6</td>
<td>Electrost</td>
<td>5.3 ± 2.9</td>
<td>2.9 ± 0.6</td>
<td>-5.6 ± 3.4</td>
</tr>
<tr>
<td></td>
<td>Van der Waals</td>
<td>-5.1 ± 1.0</td>
<td>-5.0 ± 1.1</td>
<td>-6.4 ± 1.1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>0.2 ± 2.8</td>
<td>-3.0 ± 1.3</td>
<td>-12.0 ± 3.4</td>
</tr>
<tr>
<td>TAR*6</td>
<td>Electrost</td>
<td>-0.5 ± 0.9</td>
<td>-0.5 ± 0.7</td>
<td>-0.6 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>Van der Waals</td>
<td>-5.0 ± 1.0</td>
<td>-5.5 ± 1.1</td>
<td>-4.3 ± 1.2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>-5.5 ± 1.2</td>
<td>-6.0 ± 1.3</td>
<td>-4.8 ± 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average Van der Waals and electrostatic energies (kcal/mol) of dynamics for the three mutants over the third nanosecond (final) or during equilibration (initial). These energies were calculated between the aromatic atoms of residues TAR6 (respectively, TAR*6) and of TAR*5/TAR*12 (respectively, TAR5/TAR12). For the three variants, stacking energy at the junction between U*6 and C5/G12 did not change significantly. In contrast, for the other junctions, stacking energies are representative of the stability observed by thermal denaturation experiments.
pair conformations (35) only the cis Watson–Crick/Watson–Crick one can maintain a C1–C1′ interglycosidic distance >12 Å. This was taken into account to generate the starting structure of the TAR*16(GA) mutant. U*5 was mutated into a guanine and the generated GA base pair was positioned in the cis Watson–Crick/Watson–Crick conformation. After the minimization step that reduces the severe steric hindrance, the interglycosidic distance increased from 10.8 to 12.3 Å (Fig. 7). We also decided to simulate the CA variant, because it was known to be one of the worst aptamers for TAR and only one residue modification is needed. So, simply by changing the U5* residue to G or C it was possible to mimic three different kinds of kissing complex with stabilities ranging from ‘very stable’ to ‘unstable’. To generate the CA variant, the mutated base pair was set in a N6-N3 N6-carbonyl moiety of residue 5 towards the sugar group of residue 6 of the 23S rRNA (8), meaning that they might be a crucial stabilizing factor for loop–loop interactions. In the RNA Ii–RNA IIi kissing complex structure (6), the inter-backbone hydrogen bonds are sustained by H2–O1P hydrogen bond, namely TAR5H2–TAR6O1P, was also observed, but with a lower probability. TAR5H2–TAR6O1P had a probability of presence of ~70%, whereas that of TAR5H2–TAR6O1P was 40%, and it was possible to see these two hydrogen bonds at the same time in only 10% of the registered frames. It is worth noting that these two H2–O1P inter-backbone hydrogen bonds, not described in the TAR16–TAR*16 structure (12), are present in the kissing complex of the 23S rRNA (8), meaning that they might be a crucial stabilizing factor for loop–loop interactions. In the RNA Ii–RNA IIi kissing complex structure (6), the inter-backbone hydrogen bonds are sustained by H2 and O2P atoms. The TAR5H2–TAR6O1P hydrogen bond was also present in the TAR16–TAR*16(CA) complex, but with a lower probability (50% presence instead of 95% for TAR16–TAR*16). The stacking at the junction between residues C6 and C*5/A*12 was the worst observed over the three mutants (Table 3).

Compared with other cis Watson–Crick base pairs the intrinsic parameters of this unusual GA pair were characterized by large values of the three classical parameters: opening (average ~24.4° for the last nanosecond of simulation), buckle (15.3°) and propeller twist (~20.3°). The average interglycosidic distance of 13.0 Å was associated with a high stretch of 1.6 Å. Compared with other cis Watson–Crick GA base pairs (37–39), the main characteristic of the GA closing pair lies in the high propeller twist and buckle angles.

Compared with the TAR16–TAR*16 dynamics, one additional inter-backbone hydrogen bond forms during the (GA) simulation. The symmetric counterpart of the TAR5H2–TAR6O1P, hydrogen bond, namely TAR5H2–TAR6O1P, was also observed, but with a lower probability. TAR5H2–TAR6O1P had a probability of presence of ~70%, whereas that of TAR5H2–TAR6O1P was 40%, and it was possible to see these two hydrogen bonds at the same time in only 10% of the registered frames. It is worth noting that these two H2–O1P inter-backbone hydrogen bonds, not described in the TAR16–TAR*16 structure (12), are present in the kissing complex of the 23S rRNA (8), meaning that they might be a crucial stabilizing factor for loop–loop interactions. In the RNA Ii–RNA IIi kissing complex structure (6), the inter-backbone hydrogen bonds are sustained by H2 and O2P atoms. The TAR5H2–TAR6O1P hydrogen bond was also present in the TAR16–TAR*16(CA) complex, but with a lower probability (50% presence instead of 95% for TAR16–TAR*16), and its symmetric TAR5H2–TAR6O1P hydrogen bond was seen only twice for periods shorter than 100 ps.

Hydrogen bonding for the (GA) simulation was correlated with the different combination of dihedral angles obtained at the stem–loop junctions. In fact, during the run two different combinations of dihedral angles appeared at the junctions and switched for periods of 300 ps up to >1 ns. The corresponding angles had very close values on both strands: strand 1, α60°, β6 130° and ζ5 60°; strand 2, α60°–40°, β6–140° and ζ5 160° (note that for the TAR16–TAR*16 dynamics a combination similar to the second one was observed on the TAR*16 strand). However, the same conformation cannot be present simultaneously on both strands. When one junction adopted one conformation the other had to adopt the other. Combination 2 allows the reorientation of atom O1P from the phosphate moiety of residue 5 towards the sugar group of residue 6 of the other strand and facilitates hydrogen bond formation. In contrast, conformation 1 was really less favorable for hydrogen bond formation. In this case the donor–hydrogen acceptor angle was close to 100°, reducing the hydrogen bonding.

**Figure 7.** Conformation of the starting TAR*16(GA) and TAR*16(CA) loop closing residues. Interglycosidic distance is 12.3 Å for A*12/G*5 and 10.8 Å for A*12/C*5. Dashed lines represent hydrogen bonds.
probability. As a matter of fact, during the entire run, at the junctions, the highly probable hydrogen bond always coexists with the less probable one, corresponding to dihedral combination 2 and 1, respectively.

Figure 8 shows the dihedral conformation switch of both strands that occurred at time 3.7 ns. Before this moment, the TAR*16 junction had its dihedral angles in conformation 2 and the TAR5*H2¢±TAR6O1P hydrogen bond was stable. Within a few picoseconds, the conformation switched. Conformation 2 is observed on the TAR16 side and the TAR5H2¢±TAR*6O1P hydrogen bond becomes the most probable one. During the whole simulation, conformation 2 was present for a global time of 3 ns on the TAR*16 side but for only 1 ns on TAR16. A similar phenomenon was observed during the TAR 16±TAR*16(CA) simulation. However, the two conformations of dihedrals were seen alternating with higher frequency, without correlation between the two strands.

DISCUSSION

For the formation of stable kissing complexes, loop complementarity together with base stacking, length of the linkers to cross the major groove of the loop–loop helix and specific binding of cations such as magnesium are key structural determinants. Unpaired bases, for instance, used as linkers to cross the major groove of a loop–loop helix were described in the TAR–DNA aptamer kissing complex (17) and in the DIS of HIV-1 (40). RNA I–RNA II and TAR–TAR*, in contrast, are kissing complexes without any unpaired nucleotides to connect the stem and the loop of each RNA strand. R06, an RNA aptamer identified by in vitro selection against TAR, folds as a hairpin and also engages in loop–loop interactions with the viral RNA motif. Compared to TAR*, the aptamer displays a loop that is not fully complementary to the target one. Two unpaired residues, G and A, as demonstrated by footprinting experiments (20), close the aptamer loop. These bases that could not have been predicted on a rational basis provide an increased stability to the resulting kissing complex, at physiological magnesium concentration, compared to the rationally designed TAR–TAR* one. The aim of this study was to relate these observations to structural features. For this purpose, MD was used to compare TAR–TAR* and TAR–R06 kissing complexes. We decided to run MD on three TAR16–TAR*16 variants in order to study the structural role of the closing base pair on the duplex stability. In a first step, we ran MD from the TAR16–TAR*16 structure solved by NMR (12) and analyzed its trajectory focusing on the stability of the A/U base pair located at the junctions. Then, by substituting the 5* residue by a G or a C we were able to reproduce and understand the gain or loss of stability observed in the thermal denaturation experiments.

The three-dimensional structure of TAR16–TAR*16 obtained by MD differs from the one solved by NMR spectroscopy (12). Firstly, the kissing complex is not bent towards the major groove of the loop–loop helix. This curvature was also reported for all other kissing complex structures determined by NMR experiments (6,11,41). The simulated structure of TAR16–TAR*16 adopts an overall conformation which actually agrees well with the coaxial shape obtained for kissing complexes analyzed by X-ray crystallography (8,10). Secondly, the A/U base pairs at the stem–loop junctions, A*12/U*5 and U*6/A*11, are opened. MD predicts distances...
between the hydrogens of these bases >5 Å and exposure to the solvent. This is compatible with the absence of NOE peaks between the residues of these pairs (12) and with the one-dimensional imino spectrum (12,13). Moreover, A*12/U*5 tends to have a larger C1–C1’ distance which allows the U*5 and C6 bases to stack. Finally, a rearrangement of dihedrals at the junctions allows the formation of an inter-backbone hydrogen bond as observed in other kissing complexes [RNA II–RNA III (6) and 23S rRNA (8)].

TAR*16(GA), the aptamer-like TAR ligand of highest affinity in vitro, displays the most stable behavior during the MD. No base pair opening was observed after an equilibration period. In this run, the stacking and the inter-backbone hydrogen bonds are optimized at the stem–loop junctions. Moreover, preliminary NMR experiments on TAR/TAR*16(GA) showed that the imino proton of G*5 was not exposed to the solvent and that the GA base pair is not in a sheared conformation as generally observed with purine/purine base pairs (J. Boisbouvier et al., unpublished results). These experimental results support the choice of the starting TAR*16(GA) structure. No experimental data were available to define the initial conformation of the TAR*16(CA) closing base pair. This obviously makes the results obtained for TAR/TAR*16(CA) more hypothetical. Nevertheless, with this variant we demonstrated that a closing base pair that cannot find a stable conformation with a 12 Å interglycosidic distance allows neither the formation of any stable hydrogen bond between the backbones of each strand nor a continuous stacking between the stems and the loops. Finally, TAR*16(UA) is an intermediate case. A UA closing base pair allows the formation of a hydrogen bond between the backbones of the structure and a continuous stacking between U*5 and C6. However, in this case, its loop closing base pair hardly finds a stable conformation with a large C1–C1’ distance. In addition the U*6/A11 base pair is unstable.

Preliminary results concerning the unresolved solution structure of the RNA I–RNA II kissing complex (5) show an absence of imino proton resonance in the RNA I(U7)/RNA II(A13) base pair located at one extremity of the loop–loop helix. This reflects base pair opening at the junction. In contrast, RNA II–RNA III, which is 350 times more stable in vitro than the wild-type complex (3), shows full base pairing. Differences in structure and dynamics for TAR16–TAR*16 complexes are also located at the junctions. Taking into account energetic parameters such as hydrogen bonding and stacking (Table 3), differences in stability observed in vitro between these complexes can be reasonably related to these structural features.

In TAR16–TAR*16(GA) there are no linking residues at the junction. This implies high structural constraints on the backbone between residues 5 and 6 to cross the major groove. In order to relax these constraints and/or to optimize the stacking, we suggest that the distance between residues of the loop closing base pair has to be as large as possible. Indeed, Ducongé and Toulmé (20) showed that in all closing base pair variants of R0624, purine–purine ones and more particularly the GA combination have the best stabilizing effect on the kissing complex. Purine–pyrimidine combinations are less favorable for stable loop–loop interaction and pyrimidine–pyrimidine ones drastically decrease the kissing complex stability. A purine–purine combination in a cis Watson–Crick/Watson–Crick conformation is able to close the base pair with a 12.3 Å interglycosidic distance. Such a distance is incompatible with a purine–pyrimidine combination and even more so with a pyrimidine–pyrimidine one. In these cases the link between the two pairs needs to be mediated by water molecules. To build a regular RNA or DNA helix, cis Watson–Crick/Watson–Crick pyrimidine–purine base pairs were preferred to maintain a constant interglycosidic distance.

The large interglycosidic distance between the GA cis Watson–Crick/Watson–Crick purine–purine base pair, which is known to disturb RNA helix (39), is here a crucial factor for dimer stability when located at the stem–loop junction.

In a previous work (21), in an attempt to determine whether the GA residues identified by in vitro selection against TAR (20) might be a rule to close hairpin loops and to lead to stable kissing complexes, the Watson–Crick GC pair of RNA II, next to the loop, was replaced by G and A residues. An increased stability of the RNA I–RNA II complex, monitored by thermal denaturation experiments, was observed. An in vitro selection against yeast tRNA\(^\text{Phe}\) also identified G and A residues closing the loop of the aptamer (42). However, a selection performed against the DIS of HIV-1 demonstrated that the AA sheared pair, crucial for dimerization of the viral genome, is preferred over other combinations, including GA (43,44). Preliminary thermal denaturation experiments performed on other kissing complexes with shorter or longer loops did not give any clue to predict when a GA combination might be preferred for stable loop–loop interactions (unpublished results). In the DIS of HIV-1, three purine bases are always present at the junction of all known mutants. Substitutions of these residues by pyrimidines are always detrimental to the stability of the loop–loop interaction (40). Even if GA does not always seem to be the best combination for stable dimers, purine bases seem to be preferred at the stem–loop junctions of RNA hairpins involved in loop–loop interactions.

In conclusion, the TAR16–TAR*16 simulated structure is consistent with the NMR data. The A*11/U6 and A*12/U*5 base pair opening reflects the very high tension and instability of the stem–loop junctions. Moreover, even if structures of the TAR*16 variants are not available, the MD results, starting from the TAR16–TAR*16 structure to generate a TAR–aptamer-like complex, correlate with the experimental \(T_m\) values obtained with TAR and the selected RNA aptamer. Inter-backbone hydrogen bonds, together with optimized stacking at the stem–loop junctions, seem to be structural factors that explain the increased stability of the TAR–aptamer complex when the loop of the aptamer is closed by G and A residues.

ACKNOWLEDGEMENTS

We are grateful to Justine Michel for the synthesis of the RNA and Fabien Darfeuille for helpful technical assistance and exciting discussion. We would like also to thank Eric Westhof for critical reading of the manuscript.

REFERENCES


