Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex

Hélène Van Melckebeke†, Matthew Devany†, Carmelo Di Primo‡§, François Beaurain‡§, Jean-Jacques Toulmé‡§, David L. Bryce†, and Jérôme Boisbouvier‡†

1Institut de Biologie Structurale (IBS) Jean-Pierre Ebel, 41 Rue Jules Horowitz, Commissariat à l’Energie Atomique (CEA), Centre National de la Recherche Scientifique (CNRS), Université Joseph-Fourier, 41 Rue Jules Horowitz, F-38027 Grenoble, France; †INSERM U869, Institut Européen de Chimie et Biologie, 2 Rue Robert Escarpit, 33607 Pessac Cedex, France; ‡Université Victor Segalen, 146 Rue Léo Saignat, 33076 Bordeaux Cedex, France; and †Department of Chemistry, University of Ottawa, 10 Marie Curie Private, Ottawa, ON K1N 6N5, Canada

Edited by Adriana Bax, National Institutes of Health, Bethesda, MD, and approved April 14, 2008 (received for review January 3, 2008)

Transactivation-response element (TAR) is a stable stem–loop structure of HIV RNA, which plays a crucial role during the life cycle of the virus. The apical loop of TAR acts as a binding site for essential cellular cofactors required for the replication of HIV. High-affinity aptamers directed against the apical loop of TAR have been identified by the SELEX approach. The RNA aptamers with the highest affinity for TAR fold as hairpins and form kissing complexes with the targeted RNA through loop–loop interactions. The aptamers with the strongest binding properties all possess a GA base pair combination at the loop-closing position. Using liquid-crystal NMR methodology, we have obtained a structural model in solution of a TAR–aptamer kissing complex with an unprecedented accuracy. This high-resolution structure reveals that the GA base pair is unilaterally shifted toward the 5’ strand and is stabilized by a network of intersugar hydrogen bonds. This specific conformation of the GA base pair allows for the formation of two supplementary stable base-pair interactions. By systematic permutations of the loop-closing base pair, we establish that the identified atomic interactions, which form the basis for the high stability of the complex, are maintained in several other kissing complexes. This study rationalizes the stabilizing role of the loop-closing GA base pairs in kissing complexes and may help the development or improvement of drugs against RNA loops of viruses or pathogens as well as the conception of biochemical tools targeting RNA hairpins involved in important biological functions.

The transactivation of the genome transcription of HIV-1 requires the binding of the Tat protein to a structured segment of the viral RNA called transactivation-response element (TAR). TAR is a 57-nt-long RNA that folds into a stable stem–loop structure (1, 2). In the absence of Tat, the transcription of the virus is initiated, but the RNA polymerase II disengages prematurely from the template (3). Mutations in the apical loop of TAR that do not interfere with Tat binding do indeed modulate the transactivation of the genome transcription of HIV-1 re-
Temperature is observed between the TAR*GA-TAR (27.2°C) and TAR-TAR*UA complexes as a function of temperature. The experiments were performed for 1 μM each oligomer RNA in a 10-mM sodium phosphate buffer at pH 6.6, 50 mM NaCl, and 0.01 mM EDTA.

The Aptamer Loop-Closing GA Base Pair Is of the N1-N1 Carbonyl-Amino Type. The imino region of the 1D NMR spectra is composed of 16 signals, which have been assigned based on the observable sequential imino/imino NOE connectivities. The observation of the signals for the 16 different imino protons indicated they are all protected from solvent exchange by base-pairing interactions. To further identify nucleotides that are base-paired, experiments correlating 15N and 1H nuclei linked by indirect JNN couplings across the hydrogen bonds were recorded (28). The 5 bp of the TAR stem, the first 4 bp of the TAR*GA stem, and the 12 bases of the interacting loops exhibit chemical shifts, JNN couplings, and 1H-1H NOEs characteristic of canonical Watson–Crick base pairing, confirming the secondary structure proposed for the TAR/TAR*GA kissing complex (Fig. 2a). Analysis of JNN connectivity of the TAR*GA G5′ nucleotide revealed that the imino proton H1 of G5′ forms a hydrogen bond with N1 of A*12 [supporting information (SI) Fig. S1f], confirming that the loop-closing G and A nucleotides are base-paired together. An additional NOE correlation peak observed between H1 of G5′ and H2 of A*12 indicates the close proximity of these two atoms (Fig. S1b). The GA base-pairing geometry has thus been identified as the N1-N1 carbonyl-amino type (Fig. S1a), because all other base-pairing modes are incompatible with the NMR data.

Liquid-Crystal NMR Structure of the TAR-TAR*GA Complex. The hydrogen-bonding network of Watson–Crick and N1-N1 carbonyl-amino type base pairs determined experimentally for the TAR-TAR*GA complex served as a basis for the use of a set of 62 distance restraints. Three hundred ninety-two proton–proton distance constraints were extracted from NOESY experiments. To supplement the local information gained from NOE data, we measured 143 13C-1H residual dipolar couplings (RDCs) to provide long-range orientational information for both base planes and sugar rings. These data were completed by 171 dihedral angle restraints applied to the glycosidic angles and phosphodiester angles (SI Text). An average of 24 constraints per nucleotide (Table 1) were used to determine a structure ensemble of 17 conformers, which superimposed with an average pairwise rmsd of 0.26 ± 0.07 Å (Fig. 2b). Liquid-crystal NMR parameters are particularly sensitive to bond orientation and provide a unique opportunity to define precisely molecular conformation (29). The absence of significant violation of RDCs in the calculated structures indicates that the dipolar coupling data can be fit by a single structure. Analysis of the structure ensemble revealed that the puckering phases of the helices of the mean structure are depicted by thin lines.
different ribose rings can be determined with an average precision lower than $2^\circ$, whereas the buckle, propeller twist, opening, tilt, roll, and twist angles can be derived with an average precision on the order of $1^\circ$ (Tables S1–S3). To further investigate the accuracy of the calculated structures, we backcalculated a set of structures with a small fraction of RDCs omitted to derive the quality factor $Q_{\text{free}}$ in the calculated structures. We backcalculated a set of structures with 13% of all RDCs, and this resulted in lower than 2° deviations of the calculated structures from the experimental structures (Table 1).

**Table 1. Structural statistics**

| Distance constraints (intra/seq./long dist., j > i + 1) | 392 (149/186/57) |
| Hydrogen bond restraints | 62 |
| Dihedral restraints | 171 |
| $^1$C–H RDCs (class 1 / class 2 / class 3) | 143 (90/31/22) |


deviations from experimental restraints$^1$

| RDCs class 1 (Hz) | 1.41 ± 0.03 |
| dihedral angles ($^\circ$) | 0.33 ± 0.02 |
| NOEs (Å) | 0.0487 ± 0.0001 |
| bonds (Å) | 0.0038 ± 0.0001 |
| angles ($^\circ$) | 0.58 ± 0.005 |
| impropers ($^\circ$) | 1.45 ± 0.03 |

Cross validation of a subset of the class 1 RDCs ($Q_{\text{free}}$)

| TAR stem helix (2–5, 12–15) | 0.13 ± 0.08 |
| TAR*UA helix (2*-5*, 12*-15*) | 0.26 ± 0.11 |
| Intermol. helix (6–11, 6*-11*) | 0.12 ± 0.06 |
| Overall (2–15, 2*-15*) | 0.26 ± 0.07 |

$^1$The statistics are given as an average for the 17 lowest-energy structures.

Discussion

The TAR*GA Loop-Closing Base Pair Conformation Is Stabilized by a Network of Intersugar Hydrogen Bonds and Induces the Formation of Two Additional Base Pairs. One major consequence of the introduction of a purine–purine base pair is an increase in $\Delta G$ of $\sim 2$ kcal/mol compared to the TAR-TAR*GA complex compared to the TAR-TAR*UA structure (14). Very interestingly, superimposition of the TAR*GA ensemble on the first 4 bp of the TAR*UA stem reveals a very good overlap between the two loop-closing A*12 bases, whereas the U*5 of TAR*UA superimposes on the six-atom cycle of TAR*GA G*5 (Fig. 4). The resulting unilateral shift of G*5 is stabilized by a network of new hydrogen bonds. The H2 proton of G*5 is placed in the center of a cluster of three hydrogen bond acceptors located on the TAR phosphodiester backbone between C5 and C6 (namely O5, O1P, and O2P). Analysis of the different possible rotameric states around the C2–O2 bond indicates that H2 of G*5 can form a stable hydrogen bond with O2P or O1P acceptors of C6 (Fig. 5). Because this proton was not detectable by NMR (32, 33), the exact nature of the TAR accepting group cannot be determined directly by NMR. To add experimental evidence that the H2 proton of G*5 participates in an intermolecular hydrogen bond, 2'-OMe modified versions of TAR*GA were synthesized to perform thermal denaturation experiments. The replacement of the H2 of the aptamer G*5 by a methyl group induces a drop of the melting temperature by 10°C (Table S4). In a control experiment, no effect was detected on the stability of the kissing complex when a similar modification to the sugar of A*12 was introduced. These results strongly support the notion that this intermolecular hydrogen bond involving H2 of G*5 plays a substantial role in the stabilization of the kissing complex, as suggested earlier by molecular dynamics simulations (17). Similar analysis of the liquid-crystal NMR structure ensemble indicates that the H2 of C5 may form a supplementary weak intermolecular hydrogen bond with U6-O2P. The G*5 stretched conformation may be also stabilized by the C2–endo conformation of G*4, which places its H2 proton in position to form a strong hydrogen bond with the O2P acceptor of G*5. The sugar ring of G*4 is itself stabilized by a hydrogen bond between its O2P and O2 acceptor of C6.

![Fig. 3. Detailed view of the base pair stacking at the TAR*GA loop-closing position. The superposition of the 17 structures of the NMR ensemble has been performed on the lowest energy structure for all base and sugar heavy atoms of nucleotides C6, G*4, G*5, G*11, A*12, and C*13 with a rmsd equal to 0.13 Å.](image-url)
the kissing complex were detected when the H₂ of G₄ and U₃ were replaced by methyl groups (Table S4). These results indicate that these two intramolecular hydrogen bonds are likely not critical for the stabilization of the kissing complex but may potentially stabilize the aptamer both in the TAR/TAR*GA complex and in its free form.

The stabilization of the G₅ stretched conformation also allows the concomitant stabilization of two supplementary base pairs: the N1-N1 carbonyl-amino loop-closing G₅-A¹₂ pair itself, and the adjacent intermolecular base pair connecting the 5’ end nucleotide of the aptamer loop (U¹⁶) with the 3’ end nucleotide of the TAR loop (A¹¹). These two imino protons are not NMR-detectable in a complex formed between TAR and TAR*UA (13), indicating these hydrogen bonds are not protected from chemical exchange with water by a stable base-pairing interaction. Together, the presence of these four supplementary base–base hydrogen bonds, along with the formation of the intersugar hydrogen bond between TAR*GA and TAR described above (Fig. 5), are key factors responsible for the 50-fold decrease in the dissociation constant of the aptamer–TAR kissing complex induced by the introduction of GA as a loop-closing base pair (16). The atoms involved in these intermolecular hydrogen bonds or stabilizing interactions should be preserved in any chemically modified aptamers to maintain the high affinity of the complex.

**A GA Base Pair Closing the Loop: A General Rule to Stabilize Kissing Complexes?** Using thermal denaturation, it was demonstrated that the combination of two purines (and especially the GA base pair) located at the stem–loop junction of the aptamer is the key factor enabling the stability of the TAR–aptamer kissing complex. A ranking of the stabilizing effects of the different base pair at the junction between the aptamer stem and loop has been obtained from these data: GA > GG > GU > AA > GC > UA > CA, CU (Table 2; refs. 16 and 17). To test whether these conclusions are restricted to this aptamer or whether they are more general, we investigated the effect of the change of the loop-closing base pair of TAR hairpin. This study reveals that the GA base pair located at the stem–loop junction of the TAR sequence also confers the highest stability for this kissing complex (Table 2). An increase of the melting temperature by 14°C is obtained on average, when replacing a Watson–Crick base pair by a GA base pair in the TAR loop-closing position.

*In vitro* selection of aptamers against the anticodon hairpin of yeast tRNA*⁰₆* also converged to the formation of kissing complexes (34). Interestingly, the authors showed that a GA base pair is preferred at the stem–loop junction of the tRNA*⁰₆* aptamer. To complement these results, we have performed a denaturation analysis of the tRNA*⁰₆*-G₄-aptamer complex. The results show that the replacement of the GA base pair closing the aptamer loop by a GC base pair results in a drop of Tₘ by +10°C (Table S5 and Fig. S3). Similar studies have been performed on the RNAI–RNAII kissing complex involved in the control of the colE1 plasmid replication. This kissing complex has been shown to possess no linking residues between the 7-nt loop–loop helix and the intramolecular helices (23, 35). Again, a mutation of the GC closing base pair to a GA base pair leads to a stabilization of the complex (16).

Despite the fact that all these previous examples indicate that the addition of a GA base pair to close the aptamer loop seems to be a general rule to stabilize kissing complexes, some slightly different results have been observed for complexes that possess some unpaired nucleotides. In the DIS-DIS HIV-1 kissing complex, switching from a purine–purine to a purine–pyrimidine base pair at the loop-closing position still induces destabilization of the DIS kissing complex (36). However, an AA combination is preferred at this position compared with the GA pair in this kissing complex containing a 6-nt loop–loop helix and three unpaired purines.
Curvature of Kissing Complexes Lacking Unpaired Residues. It must be noted that standard NMR data, based on the detection of NOEs and indirect scalar couplings, mainly define local geometries and do not provide direct evidence for global bending of the overall helical axis of the complex (14). The introduction of restraints based on RDCs (37–39) greatly improves this situation, allowing a precise determination of the bending of nucleic acids (40–46). The solution structure of a kissing complex determined here by liquid-crystal NMR technology is a particularly high resolution, allowing us to assess precisely the global bending of the kissing complex in solution. For these reasons, the global bending of the TAR*GA-TAR structures should preferentially be compared with x-ray structures of RNA–RNA complexes.

As for the TAR-TAR*GA complex, the structures of these kissing complexes share a common quasicontinuous helix fold formed by two intramolecular and one intermolecular helices but differ in several respects. First, the number of paired loop residues can vary from two up to seven (22, 23, 35). In addition, some of the kissing complexes possess nonpaired bases (22, 24, 25, 47–49). For example, the crystal structures of the HIV-1 MAL and LAI kissing complexes (49), the anticodon–anticodon interaction in tRNA^Ala (47), and a loop–loop interaction in 23S RNA (48) show coaxial alignment of the two opposite stems due to flanking residues on the 5′ side. The stabilization observed in the UV-melting data for this type of 5′ configuration is likely to be caused by coaxial stacking (36). The unpaired bases provide a linker between the loop and the stem to cross the major groove of the loop–loop helix and give enough flexibility at the junction to maximize the stacking between each helical segment.

In contrast, the absence of a linker implies the existence of severe structural constraints at the stem–loop junctions for the HIV TAR aptamer complex. Using liquid-crystal NMR technology, we have shown that a unilateral stretching of the loop-closing GA base pair located at the stem–loop junction alleviates the tensions created by the lack of unpaired linking residues in the TAR-TAR*GA complex, consequently providing a higher stability. This results in a 45° bending of the TAR-TAR*GA kissing complex toward the major groove (Fig. 2b). Similar bending has been documented for a kissing complex formed between CoE1 plasmid specific RNA I and RNA II transcripts, comprising seven intermolecular base pairs and no unpaired nucleotides (23, 35).

Conclusion

The high-resolution structure of HIV-1 TAR complexed with a high-affinity RNA aptamer has allowed us to elucidate the molecular origins of the gain in binding affinity associated with the aptamer GA loop-closing base pair, compared with rationally designed hairpin ligands containing only Watson–Crick base pairs. The larger interglycosidic distance imposed by this purine–purine base pair induces a unilateral shift of the G base, allowing for the formation of two supplementary stable base-pair interactions. This specific conformation is furthermore stabilized by a network of intersugar hydrogen bonds. Systematic permutation of the closing base pair and comparisons with other kissing complexes indicate that the stabilization role of the GA base pair seems to be a general rule for kissing complexes lacking unpaired nucleotides between the interacting loop and the intramolecular stems. Thus, the precise atomic interactions crucial for the stability of RNA–RNA kissing complexes, which are revealed by this liquid-crystal NMR study, may serve as a basis to produce modified aptamers with improved nuclease resistance and that maintain the highest affinity for the HIV TAR sequence. This knowledge will also be useful for the design of high-affinity RNA aptamers against RNA loops of viruses or pathogens as well as in the conception of biochemical tools targeting RNA hairpins involved in important biological functions.

Table 2. Effects of the loop closing base pair on the stability of the kissing complex

<table>
<thead>
<tr>
<th>TAR/aptamer</th>
<th>UA</th>
<th>GA</th>
<th>CU</th>
<th>GC</th>
<th>GU</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>UA</td>
<td>13.0 ± 0.8</td>
<td>33.2 ± 1.6</td>
<td>&lt;10</td>
<td>18.7 ± 1.9</td>
<td>23.4 ± 1.0</td>
<td>&lt;10</td>
</tr>
<tr>
<td>GA</td>
<td>24.5 ± 0.5</td>
<td>41.8 ± 2.1</td>
<td>17.3 ± 0.2</td>
<td>31.5 ± 0.6</td>
<td>33.8 ± 2.6</td>
<td>23.2 ± 0.8</td>
</tr>
<tr>
<td>CU</td>
<td>&lt;10</td>
<td>11.0</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
<td>&lt;10</td>
</tr>
<tr>
<td>CG</td>
<td>&lt;10</td>
<td>33.1 ± 1.2</td>
<td>&lt;10</td>
<td>14.9 ± 0.3</td>
<td>20.0 ± 0.8</td>
<td>&lt;10</td>
</tr>
</tbody>
</table>

Melting temperatures (Tm) (°C) of complexes formed by different mutants of TAR (vertically) and SELEX identified aptamer (horizontally). The mutations concern the stem–loop closing base pair for both RNAs and are indicated in the different columns. A short version of TAR without the stem-stabilizing GC base pair (G3-C14, 14 nt), and a nontruncated aptamer (R0624, refs. 16 and 17) has been used in this study. The given values are the mean and standard deviation of two or three measurements performed for 1 μM of each oligomer RNA in a 20 mM sodium cacodylate buffer (pH 7.3 at 20°C) containing 20 mM sodium chloride and 140 mM potassium chloride, and in the presence of 0.3 mM MgCl2.

Materials and Methods

Sample Preparation. Unlabeled RNA molecules used for this study were synthesized on a solid phase. 15N-13C-labeled (98%) 16-nt TAR and TAR*GA were prepared in vitro by using T7 RNA polymerase (Sillantes). RNA samples were dialyzed against 50 mM NaCl at pH 6.6 in 0.01 mM EDTA, 10 mM sodium phosphate buffer containing 0.4 g liter−1 sodium azide. Only one of the partners was labeled with 15N and 13C in TAR/TAR*GA samples used for NMR studies. The final sample concentrations of the labeled species were between 0.4 and 0.8 mM, and the relative concentration of the unlabeled species in the mixed samples was 2:1. All of the NMR data were acquired at 25°C in D2O buffer, except experiments involving exchangeable protons, which were recorded at 10 and 5°C in H2O buffer. Further details on assignment, dihedral, and distance restraint determination are given in SI Text and Table S6.

Liquid-Crystal NMR Spectroscopy. For liquid-crystal NMR studies, the 15N-13C-labeled RNA complexes were aligned by adding to an isotropic sample a P1 filamentous phage solution (ref. 39; ASLA Biotech) to give a final concentration of 13 mg/ml. The 2H NMR splitting observed for 2H2O at 25°C in these samples was 14.0 ± 0.1 Hz for the complex containing 15N-13C-labeled TAR RNA and 12.4 ± 0.1 Hz for the sample containing 15N-13C-labeled TAR*GA RNA. Spin-state selective experiments (50–52) were used to extract one-bond 13C-1H couplings in isotropic and P1-aligned 15N-13C-labeled RNA complex samples. Extracted RDCs were scaled linearly with respect to the observed 2D NMR splitting, to take into account the small difference in the magnitude of the alignment of different samples. The linearity of measured dipolar couplings with the 2D splitting was verified experimentally by varying the amount of phages in the same sample of TAR/TAR*GA complex (data not shown). The RDCs were separated into three classes to account for overlap of the peaks in the spectra, with error bars estimated at 1 Hz (well resolved), 5 Hz (partial overlap), and 10 Hz (overlap).

Structure Calculation. A first estimate of the magnitude of the axial and rhombic components of the alignment tensor was obtained by using histogram methods (53–55). The values have been further refined by a grid search over the D2 and R values of the tensor to determine the minimal value of the dipolar energy for the set of structures calculated using Xplor-NIH (56) with a parameter file specifically optimized for nucleic acids (ref. 57; details in SI Text). Furthermore, the grid search procedure was repeated using only either RDCs for TAR or for TAR*GA molecules, independently. Both calculations converged on similar values of D2 and R, suggesting that the different parts of the complex are not subject to internal flexibility. This is corroborated by the absence of unpaired or highly mobile nucleotides at the junction between the three different helical segments. A single alignment tensor was then used to calculate a final set of 800 structures from...
which the 17 structures with lowest energies were selected to represent the structure ensemble, and these were deposited in the Protein Data Bank (PDB ID code 2nn1).

Cross-validation of the structures was carried out by selecting a subset of high-precision one-bond $^{13}$C-$^1$H RDCs (class 1), excluding them from the list of restraints enforced during the structure calculation, and then evaluating how well the excluded RDCs are predicted on the basis of the resulting structures. Specifically, 28 high-precision $^{13}$C-$^1$H RDCs were divided into different groups, and each of the four groups was used for cross-validation in succession. A similar procedure was performed independently for TAR.

Thermal denaturation of experiments was monitored on a UVikon XL (Bio-Tek Instruments) spectrophotometer interfaced with a Peltier effect device that allows to interact for 20 min before they were cooled to 4°C.

The experiment started after 20 additional minutes at this temperature. Denaturation of the samples was achieved by increasing the temperature at a rate of 0.4°C per min from 4 to 90°C and was followed to 260 nm.

ACKNOWLEDGMENTS. We thank Drs. Brutscher, Simorre, and Soulier for helpful discussions. This work was supported by the Commissariat à l’Energie Atomique, the Centre National de la Recherche Scientifique (CNRS), University of Grenoble, University of Ottawa Faculty of Science, Canadian Institutes of Health Research (CIHR), and Human Frontier Science Program (HFSP), H.V.M. and M.D. acknowledge the receipt of fellowships from the Agence Nationale de Recherche sur le Sida et les Hépatites Virales et CNRS. J.B. is a recipient of a Career Development Award from HFSP (Grant CD0029/2004-C); J.B and D.L.B. acknowledge funding from the Canada-Canada Research Group, the CHR-CNRS program, and the Jacques Cartier Center.