J Mol Model (2004) 10:408–417
DOI 10.1007/s00894-004-0216-7

ORIGINAL PAPER

Jérôme Golebiowski · Serge Antonczak · Juan Fernandez-Carmona · Roger Condom · Daniel Cabrol-Bass

Closing loop base pairs in RNA loop–loop complexes: structural behavior, interaction energy and solvation analysis through molecular dynamics simulations

Received: 3 May 2004 / Accepted: 10 September 2004 / Published online: 22 October 2004

Abstract Nanosecond molecular dynamics using the Ewald summation method have been performed to elucidate the structural and energetic role of the closing base pair in loop–loop RNA duplexes neutralized by Mg^{2+} counterions in aqueous phases. Mismatches GA, CU and Watson–Crick GC base pairs have been considered for closing the loop of an RNA in complementary interaction with HIV-1 TAR. The simulations reveal that the mismatch GA base, mediated by a water molecule, leads to a complex that presents the best compromise between flexibility and energetic contributions. The mismatch CU base pair, in spite of the presence of an inserted water molecule, is too short to achieve a tight interaction at the closing-loop junction and seems to force TAR to reorganize upon binding. An energetic analysis has allowed us to quantify the strength of the interactions of the closing and the loop–loop pairs throughout the simulations. Although the water-mediated GA closing base pair presents an interaction energy similar to that found on fully geometry-optimized structure, the water-mediated CU closing base pair energy interaction reaches less than half the optimal value.

Keywords RNA · Molecular dynamics · Mismatch · Energy · Simulation

J. Golebiowski (✉) · S. Antonczak · J. Fernandez-Carmona · D. Cabrol-Bass
Laboratoire Arômes, Synthèses, Interactions, Faculté des sciences, Université de Nice – Sophia Antipolis,
06108, Nice Cedex 2, France
e-mail: jerome.golebiowski@unice.fr
Tel.: +33-4-92076103
Fax: +33-4-92076125

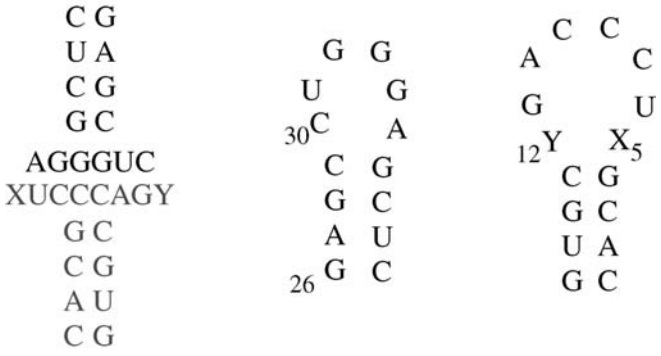
R. Condom
Laboratoire de Chimie Bio-organique, UMR UNSA-CNRS 6001, Université de Nice – Sophia Antipolis,
06108 Nice Cedex 2, France

Introduction

Some RNA elements are crucial for the gene expression of human immunodeficiency Virus type 1 (HIV-1) [1]. A segment of the HIV-1 RNA genome, consisting of residues 1–59, named the Transactivation Responsive (TAR) region, adopts a stem-loop secondary structure that presents a highly conserved hexanucleotide loop, providing a homing site for cellular proteins. Primary sequences of a three-nucleotide bulge and a six-nucleotide loop of the TAR RNA stem structure are critical for the activity of the transactivation protein Tat [2, 3]. Other proteins (Cyclin T1, CDK9 or P-TEFb) are also involved in control of the transcriptional regulation, depending on TAR RNA structure [4–6]. HIV-1 TAR RNA can therefore be considered as a potential target for the design of HIV-1 inhibitors. TAR RNA can potentially interact with other hairpin-loop nucleic acids to form kissing-loop complexes. Such complexes are generated during a highly selective process, through Watson–Crick interactions between complementary bases of each loop of the nucleic acids. In 1994, Chang and Tinoco characterized a complex between TAR and its rationally designed complementary RNA called TAR* [7]. The structure was further elucidated and confirmed the complementarity between the six nucleotides of each loop [8]. Stability of such macrocyclic complexes is dominated by the complementarity between bases of each loop [9]. The orientation of the loop [10] as well as the sequence of each stem next to the loop and the closing-loop base pair also have an effect on the stability of those entities [11].

Duncongé et al. [12] identified in vitro RNA hairpin aptamers specific for HIV-1 TAR RNA (see Scheme 1).

Aptamers can be defined as oligonucleotides usually identified by some selection process to bind to a given target specifically [13, 14]. The aptamers of Ducongé et al. [12] contain a stem and an eight-base sequence forming a loop in which the six central bases are complementary to the entire TAR loop. In their study, Ducongé et al. [12] showed, by thermal denaturation and



Scheme 1 Secondary structures of RNAs involved in this study. TAR model, XY aptamer (XY being the closing loop pair), and XY aptamer/TAR model duplex.

surface plasmon resonance, that the nature of the two closing-loop bases strongly influences the strength of the complex formed with TAR. Among all the other closing base pairs, the mismatch GA closing-loop base pair was shown to lead to the best stability of the TAR/aptamer complexes. The complexes' stabilities decrease as in the following order GA = AG > GG > GU > AA > GC (~TAR*) > UA >> CA~CU. The same ranking was observed previously for the hexanucleotide hairpin loops [15] internal mismatches, [16] or internal loops [17]. Dimerization of DIS RNA kissing-loop aptamer complexes containing nine nucleotides has also been investigated recently [18]. In this study, the authors observed that, although the closing GA base pair did not lead to the more stable dimer, it could allow formation of stable heterodimers. In a recent study, the role of loop-closing residues in these complexes have been investigated through molecular dynamics simulations [19]. The authors considered the GA, CA and UA loop-closing residues and concluded that inter-backbone hydrogen bonds, together with optimized stacking interactions at the stem–loop junction can explain the increased stability of the TAR–aptamer complex, where the aptamer loop is closed by a GA base pair.

Nonetheless, the detailed effect of the closing base pair, particularly mismatch ones, on helices behaviors is not yet fully elucidated, in spite of recent work, inspecting the structural aspects of such complexes bearing loop or loop-loop motifs [19–23]. Work dealing with single GA mismatches inserted in a regular helix is quite rare, since tandem GA-GA mismatches are more usually investigated. Our work should give more insights on single GA mismatches through a molecular dynamics approach. The influence of solvent molecules may be crucial in structuring mismatch base pairs, especially when the latter are located in a rigid region, as might be the case at the loop–loop junction. Water molecules can be inserted between the two mismatch bases. Such configurations are called water-mediated base pairs and have been reported in several studies [24–26]. Although the structural factor probably partly drives the stability of nucleic acid complexes, its quantification together with the specific solvation energetics has not been described accurately. The rational design of more effective RNA-type inhibitors requires knowledge of the effect of such closing base pairs at a molecular level. In the present work, we investigate the behavior of TAR/aptamer complexes in water phase when the closing base pair's is GA, GC and CU by means of molecular dynamics simulations and energetic analysis. The influence of closing base pairs is evaluated on the basis of their structural, energetic and solvation specificities. Our work should provide a complementary study to the work of Beaurain et al. [19], more particularly considering our solvation and energetic analysis, and could give additional descriptions of RNA/aptamer systems.

Methods

Starting structures were constructed using experimental data from the TAR/TAR* complex [8]. The complexes were neutralized by addition of 14 Mg^{2+} counterions, placed in the most negative region of space using the LEAP module. The use of Mg^{2+} ions has been considered since Chang and Tinoco [7] noticed that their presence has a significant role in the stabilization of their TAR/TAR* loop–loop complex. The requirement of Mg^{2+} for stable complex formation was also reported for the ColE1 complex [10]. Furthermore, note that the experimental data are given at physiological concentration of Mg^{2+} [12]. Each system was embedded in a periodic box containing 7957, 7776 and 7691 TIP3P [27] water molecules for the GA, GC and CU complexes, respectively. The water phase was extended to a distance of 15 Å from any solute atom. Molecular dynamics simulations were carried out using the AMBER6 [28] program in the isotherm–isobar thermodynamic ensemble at 300 K, with the SANDER module using SHAKE on bonds involving hydrogen atoms. A time step of 2 fs was applied. An 8 Å cutoff was applied to non-bonded van der Waals interactions and the non bonded pair list was updated every 15 steps. After having added the ions and the water molecules to the minimized complexes, 1000 steps of minimization keeping the complexes and the ions fixed were performed using particle mesh Ewald (PME) summation. PME parameters are chosen to obtain a grid spacing close to 1 and 9 Å direct space cutoff. The equilibration runs continued by an equilibration of 50 ps of PME dynamics, keeping the solute fixed. Then, 1000 steps of minimization and 10 ps of MD simulation using a restraint of 20 kcal $(mol\cdot Å^2)^{-1}$ on the solute atoms were performed, followed by four rounds of 1000 steps minimization reducing the restraints by 5 kcal $(mol\cdot Å^2)^{-1}$ at each round, with 10 ps MD simulation. Further, the system was slowly heated from 100 to 300 K over a period of 15 ps. The equilibration was continued over 100 ps after these 15 ps. The 2 ns production phase was started for each complex, saving the trajectory each picosecond. Root mean square deviations (RMSd) calculated on heavy atoms over the trajectories were obtained with the Carnal module. Further, interaction energy analyses between selected base pairs were performed by extracting the coordinates of the considered bases and replacing the sugar backbone phosphate by a hydrogen atom whose charge was adjusted to keep neutrality. The interaction energies were obtained considering 2000 separate single point calculations on structures taken from each trajectory and estimated as follows:

$$E^{\text{interaction}} = E^{abc\ldots n} - \sum_n E^n \quad (1)$$

where E^n are the computed energies of the separated nucleic acid bases involved in the interaction. $E^{abc\ \ldots\ n}$ is the single-point energy of the structure made up of n nucleic acid bases.

The energy was obtained using the Cornell et al. force field implemented in Gaussian98 [29], using the parm96.dat potential, as in a preceding studies [30]. Estimation of such interaction energies

at this level of theory has been shown to reproduce fairly well post-HF ab initio calculations and can thus considered to be accurate [31, 32].

About simulations lengths

The question about the relation between a simulation length and its significance is regularly addressed. Considering MD simulations of biological systems, reliable simulations can be generated on the nanosecond range time scale, especially when systems are close to equilibrium or located in a conformational attractor [33]. More precisely, in nucleic acid systems, intermolecular interactions are strong enough to ensure a relatively good structural description, related to an experimental structure. The conformational sampling of double-stranded nucleic acids starting from an experimental structure can then be considered to be satisfactory within a nanosecond time-scale MD simulation, even if modern force-fields are now accurate enough to avoid structural denaturation, to ensure a good sampling of the structure, starting from a valid initial structure [34]. Moreover, solvation pattern and monovalent ion-binding analysis can reasonably be realized by means of MD simulations of a few nanoseconds, but only on a semiquantitative basis, since the residence time of such binding takes place in a nanosecond regime (see for example [35–37]). On the contrary, divalent ion desolvation processes occur within ~10 μs, and characteristics related to such process can hardly be reproduced by simulations. For the moment, the binding-site exchange of divalent cations cannot be observed systematically by means of nanosecond regime simulation, since they show much stronger electrostatic interactions, very difficult to overcome at room temperature. Note that the scope of this study focuses on the RNA behavior and that ion-binding analysis nonetheless provides additional pertinent insights on their potential effect on complex stability, although the force field approach does not reproduce induction effects such as the polarization energy or charge transfer. Such effects could only be taken into account through a polarizable force field or a quantum mechanical treatment, far beyond the scope of this study. Note that the amount of non-additivity contribution in the first ligand shell of Mg^{2+} can be very different between molecular mechanics and quantum mechanics or polarizable models [38]. Nonetheless there is a substantial degree of compensation of errors when moving from a quantum chemical treatment to a nonpolarizable molecular mechanics force field.

Results and discussion

Choice of the starting structure

The TAR and the loop–loop part structures of the three complexes were taken from the experimental structure given by Chang et al. [8] on the so-called TAR/TAR* kissing-loop dimer. The remaining stem of the three aptamers was built on the basis of an A-form RNA. In the case of the GC closing base pair, a classical Watson–Crick interaction was considered.

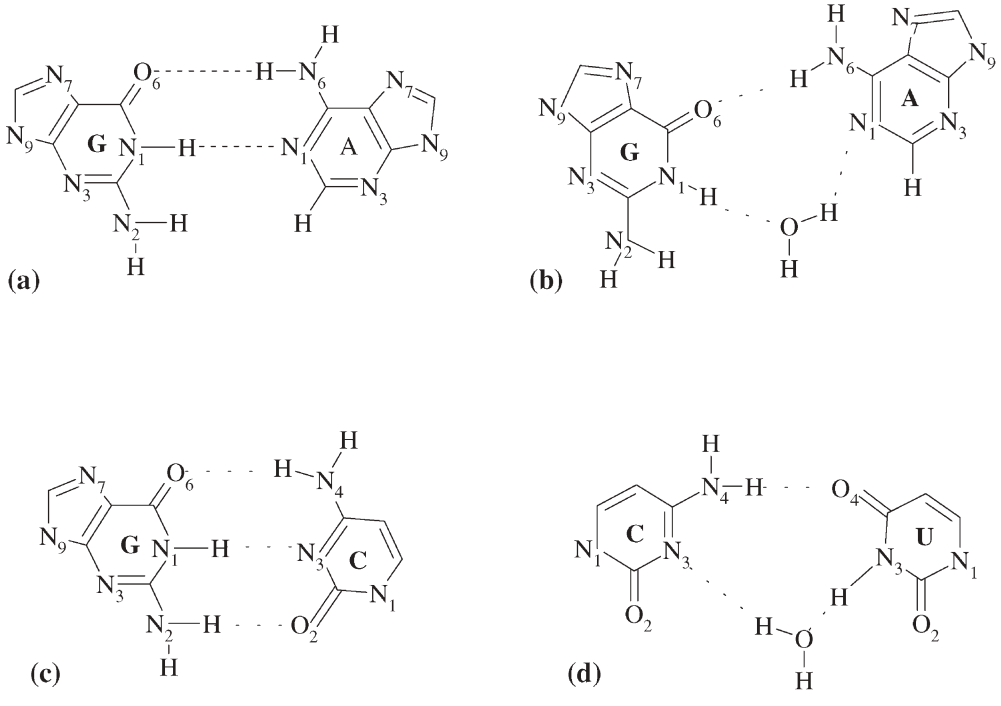
The structure of mismatch GA closing base pairs is not obvious. Structures of several single GA mismatches in RNA have been determined by X-ray crystallography and NMR. Although in symmetric tandem GA mismatches the GA base pair structure depends on the interaction with the base pair 5′ of G [39], the structure of a single GA base pair does not follow this rule. One can observe sheared GA pairs in structures containing tandem mismatch base pairs or non-classical interactions such as

5′ G G A 3′
3′ C A U 5′,

in which AU is in a reversed-Hoogsteen structure [40, 41],

5′ G G C G N
3′ C C G A A,

where N in the hairpin is C or A, [42] or



Scheme 2 Base pairs found in the simulation: **a** GA imino hydrogen-bounded, **b** water mediated GA **c** Watson-Crick GC **d** water mediated CU.

5′ C G A 3′
3′ G A A 5′

[43]. Imino hydrogen-bonded GA can be found in structures such as GAAA tetraloops [44], in helices [45, 46] or near other mismatches, for example in the

G G G
G A C

sequence [47] (Scheme 2).

Ducongé et al. [12] suggested that the GA mismatch base pair could not be in a sheared structure since the GA aptamer was much more efficient in TAR binding than its isosteric AA one. AA would be exchangeable with GA if a sheared pair occurred [48]. The fact that the AA pair is the less stable of the PuPu base pairs, suggests that PuPu pairs interact through their Watson–Crick sites. Note also that the FMN cofactor RNA-aptamer presents an imino hydrogen-bonded GA base pair at the junction between the stem and the internal loop [49]. We have conducted a simulation where the GA base pair was in the sheared conformation. Although a 1-ns simulation lead to an apparently stable complex, we observed a regular distortion of the structure that ended in a structure in which the GA closing base pair was no longer in direct interaction, leading in fact to large structural deviations. On the basis of this knowledge, we have considered a GA imino hydrogen-bonded structure, where the oxygen atom of the guanine base interacts with the H atom of the NH_2 terminal group of the adenine.

Concerning the CU complex, it has appeared in most experimental structures that CU base pairs present a water molecule inserted between the two N_3 atoms of uracil and cytosine. A high-resolution example of such water-mediated base pairs was given at 1.8 Å resolution (NDB id AR0005) [50]. The bases are connected by direct hydrogen bonds and additional water-mediated H-bond interactions, as depicted in Scheme 2d. This type of mismatch interaction has already been studied by ab initio calculations and molecular dynamics simulations in the case of a "regular" A-form helix [24, 26]. The starting structure of the CU complex takes into account this type of water molecule. As for the GA complex, a simulation conducted with a direct CU H-bond resulted in a denaturation of the complex and confirmed a wrong choice of starting structure. Also, although GA and water-mediated CU are known to covary at the top of the anticodon stem of tRNAs, [51] their role at the aptamer loop closure has been shown to be clearly specific [12] and suggests a radically different behavior in this type of structure.

Structural analysis

Root mean square deviations (RMSd) along the trajectories, taking the starting structures as a reference, were used to analyze the stability of the simulations. They are shown in Fig. 1 for the three complexes, whose average structures are also shown.

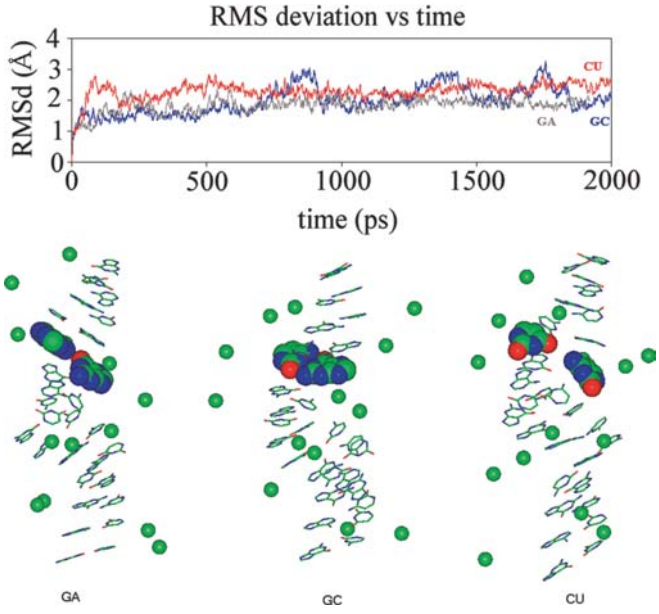


Fig. 1 Root mean square deviation of the three complexes, calculated on the heavy atoms along the 2 ns simulation (grey, GA; red, CU; blue, GC) and average structures showing Mg^{2+} ions.

The three RMSd vary below 3 Å, revealing stable trajectories. The CU complex, however, shows a slightly larger RMSd variation around the mean value, indicating a higher deformation along the simulation. To analyze the behavior of the loop–loop helix sub-structure within the three complexes more closely, we have computed the RMSd of the loop–loop helix along the simulations. The values obtained are 1.3, 1.4 and 1.7 Å for the GA, GC and CU complexes, respectively. Such values emphasize that this helical part within the complexes remains faithful to its starting structure.

At first sight, the three structures are similar to the TAR/TAR* kissing-hairpin dimer [8] but dynamic analysis of their behavior along the simulation led to interesting structural conclusions that can hardly be reached through an experimental approach. Note that the complexes can be decomposed into three parts: two classical A-form helices, linked by a central right-handed helix composed of nucleotides of the two loops. A detailed analysis of the central helix reveals a structure different from the A-form. This difference originates from gaps between residues 12 and 13 in the aptamers as well as between A35 and G36 of TAR. These gaps are connected by only one phosphodiester bond, leading to a bend of the structure. The resulting complexes can nonetheless be considered as a quasi-continuous bent helix made-up of these coaxially stacked structures. In the following, discrepancies between the three complexes are revealed through sugar puckering, axis curvature and inter-base distance analysis.

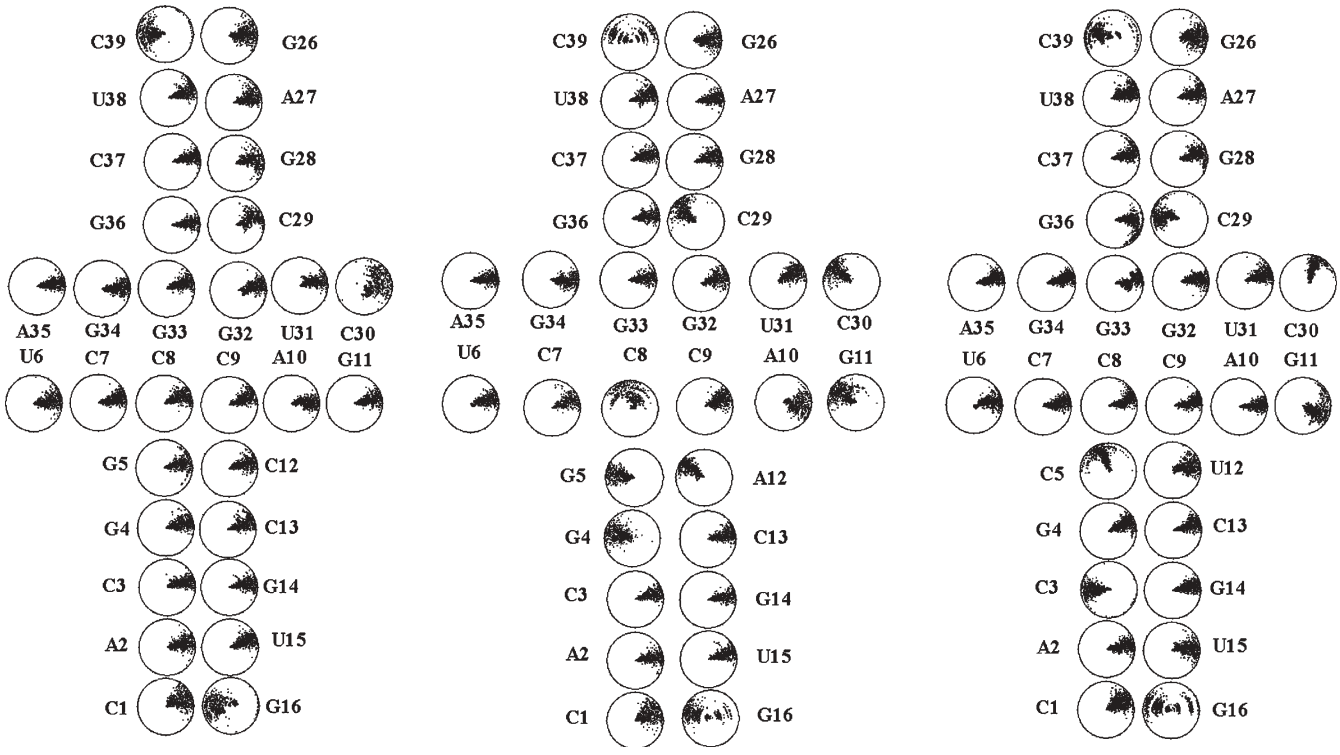


Fig. 2 Polar plot of the nucleosides sugar pseudorotation puckering angle along the 2 ns simulation, for GC, GA and CU complexes respectively. The angle is measured in a trigonometric circle, the horizontal right axis representing a zero angle. The radial component corresponds to time. t=0 corresponds to the center and 2 ns to the boundary of the circle.

Table 1 Overall axis-bend angle between TAR, loop–loop and aptamer helices. Comparison with the TAR/TAR* experimental structure

	GC complex	GA complex	CU complex	TAR/TAR* (exp.)
TAR/Loop°	34.8	64.2	65.6	28.9
Loop/Aptamer°	76.9	40.0	23.1	52.1

Sugar puckering

Examination of the backbone structure through the sugar-puckering angle is shown in Fig. 2. A majority of the sugar rings adopts a classical A-form RNA C_3'-endo pucker in all three complexes.

Concerning the GC complex, the analysis reveals an A-form-like RNA of the whole complex (pucker angle of ~20°, C3′-endo), with few structural variations at the loop–loop junction. The GA complex shows a C2′-endo (~160°) pucker angle at its mismatch base pair, which induces a distortion of two nucleosides at the loop–loop junction (G11-C30 and A12), together with a nucleoside in TAR (C29). These sugar structures are the consequence of an opening of the major groove of the loop–loop helix, discussed in the following. The CU complex shows a large variation of the C5 nucleoside, indicating a perturbation of the first stem of the loop-loop helix. A second stem, i.e. G11-C30, is also distorted with respect to an A-form RNA, and this perturbation is also encountered in the TAR helix, through its C29-G36 and G28 nucleosides, especially in the third part of the simulation (1500 ps to the end). Then the CU complex seems to necessitate a larger reorganization of TAR upon binding. Such an allosteric effect could be the origin of the weaker affinity between TAR and the CU aptamer, since it is much larger than in the other complexes considered.

Axis curvature

Analysis of the bend angles and the average complexes' axis curvature between the three coaxially stacked helices was performed with the Curves5.3 software [52]. The average values were estimated from the 2 ns simulations and are compared to those calculated on the experimental TAR–TAR* structure in Table 1. The table provides two angles calculated between the three axis. A graphical representation of the three axis is shown in Fig. 3.

First, note that the values for the bend angles of the GC complex show the same trend as found for the TAR–TAR* structure, but more pronounced. This is especially

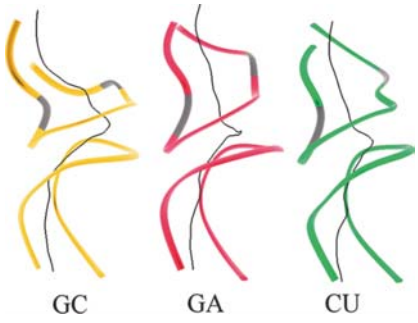


Fig. 3 Representation of the average helical axis and ribbon for the three considered complexes. The *gray colored ribbon* indicates the location of the aptamer closing base pair.

true for the kink localized on the loop/aptamer side, showing that in this case, the addition of the GC stem essentially perturbs the nucleosides within the aptamer RNA without having an important structural role on the TAR structure.

A contrario, values computed for the GA and CU complexes show the opposite trend to TAR/TAR*. Both structures exhibit axes less bent at the loop–aptamer junction than that of the TAR–loop junction. These unexpected trends reflect the importance that an additional stem can have on the global axis behavior, even if the TAR structure remains conserved.

The stem-by-stem axis curvatures were evaluated for each step of the global helix within the three complexes. The evolution from TAR to the aptamer is shown in Fig. 4 and represents the localized axis curvature between each stem of the helix.

We can see from Fig. 4 that the evolution of the stem-by-stem axis curvature is quite similar along TAR and the loop–loop parts of the three structures considered. The curvature decreases in the TAR helix and then regularly increases at the beginning of the loop–loop sub-structure. Then, a difference clearly occurs. The effect of the GC closing base pair is less pronounced. It produces only a slight curvature of the axis (respectively 8 and 5.3° for U6/G5 and G5/G4), in agreement with the kink angles discussed above. Although the axis curvature for the two preceding helices is comparable for the GC and GA complexes, the presence of the GA closing base pair increases the local curvature at the loop closure. The curvature angle reaches 25.2 and 15.8° for U6/G5 and G5/G4, respectively, in the GA complex. The CU closing base pair produces a radically different behavior. In this complex, the curvature increases from 16.6 to 26° for the U6/C5 and C5/G4 stems, suggesting a large distortion of the aptamer bearing the CU closing base pair with respect to the other systems.

Closing loop inter-base distance

Distortion of the stem–loop junction backbone can be evaluated from a structural point of view by means of a C_1'–C_1' distance analysis (Fig. 5).

Let us first consider the GA complex. The average C_1'-C_1' distance is 14.6 Å, with instantaneous values

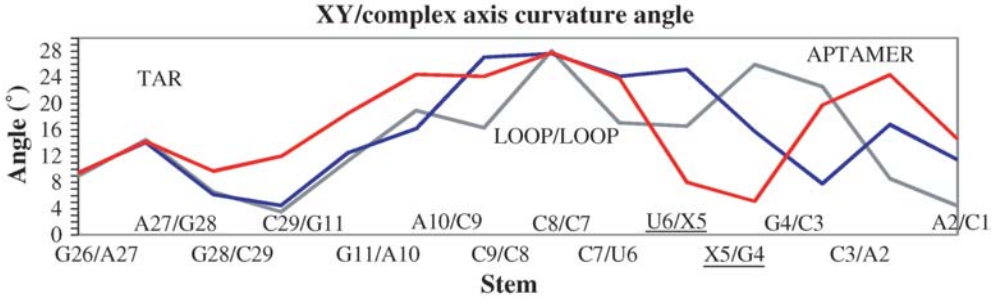


Fig. 4 Evolution of the three complexes axis curvature for one step relative to the next. GC: blue, GA: grey, CU: red

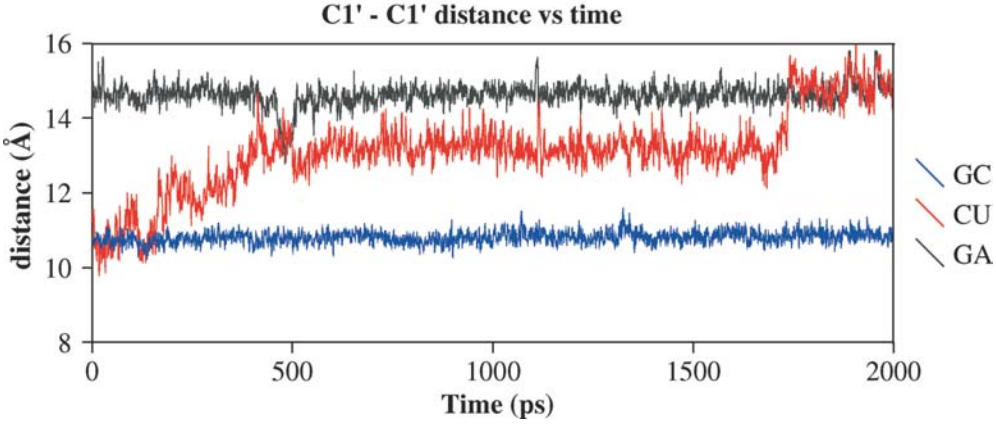


Fig. 5 Evolution of the C_1'-C_1' inter-closing loop base pair distance in the three complexes.

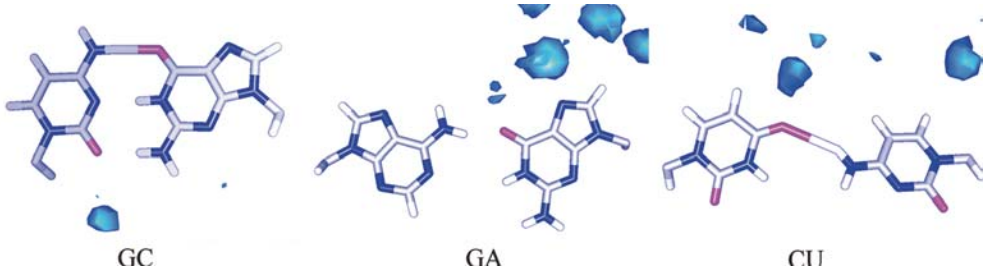


Fig. 6 Average interaction structures of the X5-Y12 mismatches with preferential solvation zones.

presenting few variations (standard deviation: SD=0.29 Å). These small variations around the mean value are the result of an exchange of water molecules between the bases. During a specific period of the simulation (400–500 ps), the inter-base distance (~ 13 Å) is shorter than the mean value. This period corresponds to a two H-bond-interaction pattern within the guanine and the adenine, but occurs during a very short lifetime (<100 ps). The presence of an inserted water molecule between the N_1 and N_1–H_1 atoms of the adenine and guanine, respectively, was found regularly until the end of the simulation. This water-mediated interaction pattern has already been studied theoretically [25]. The authors indicated that water-mediated GA base pairs can be considered as structurally autonomous building blocks in RNA, considering that the inserted water molecule has both a donor and acceptor nature. A high-resolution experimental inter-base distance of a water-mediated GA base pair is found to be 14.8 Å in loop E of 5S rRNA (PDB code: 354d) [47], in agreement with our average value. Note that this inter-base distance is one of the largest observed for an RNA base pair [51].

When considering the GC complex, we note a closing base pair distance (10.8 Å) very similar to that in a classical stem (10.5 Å) due to the strong H-bonds. Weak fluctuations around the mean value indicate that the GC base pair is in a tight interaction (SD=0.17 Å). The distance is far smaller than in the GA complex and avoids large flexibility (or deformability) at the loop–loop junction.

Finally, similar analysis of the CU complex indicates that the inter-base distance varies along the simulation, as already noticed on the isosteric AU apatmer MD simulation [19]. In a first 300 ps period, this distance oscillates around a value of 11–12 Å, roughly comparable to the experimental value of the same mismatch when inserted into a regular helix (11.8 Å) [53]. Then, during 200 ps, the distance increases slowly to finally fluctuate around 12.9 Å (SD=0.83 Å) until 1700 ps. This large value is the result of water exchange between uracil and cytosine together with a loose UC interaction. Schneider et al.[26] evaluated the C1′–C1′ distance when water-mediated UC base pairs were located in a double helix RNA duplex [26]. Their values are on average ~1 Å shorter than ours. This discrepancy emphasizes that the water-mediated CU closing base pair is structurally not suited for closing the loop in this type of complex. Indeed, after 1700 ps, the groove is opened widely, with a mean C_1'–C_1' distance of 14.9 Å until the end of the simulation, suggesting the beginning of the complex distortion, through denaturation of the CU closing base pair.

On the basis of the structural results, it has appeared that the UC closing base pair does not interact strongly enough to keep a close contact. The bases are not close enough to generate a tight interaction at the loop–loop junction, even with the presence of an inserted water molecule. Conversely, both GC and GA present a tight interaction between the bases (three direct H-bond for GC and at least one direct H-bond for GA plus the inserted water molecule). However, the larger GA inter-base distance leads to a larger flexibility while keeping tight interactions between all residues.

Solvation analysis

Figure 6 shows the average interaction pattern between the mismatch base pairs, together with their preferential solvation zone. The mismatch GA base is also a rich solvation and ion-binding site. Despite chemical acylation experiments suggesting a poor specific recognition of the deep narrow major groove of A-RNA [54], experimental facts suggest that ions can interact specifically with the major groove, more precisely through interaction with N_7 atoms of guanine bases [53]. Actually, we do observe such interactions in the GA structure. The major groove O_6 and N_7 atoms of the mismatch guanine base are largely solvated and present an interaction with a counterion. These interactions probably tend to stabilize the structure of this aptamer with respect to the GC and the CU aptamers.

A bridging water molecule can be observed between TAR and the GA aptamer. It is located in the sharp turn between the aptamer helix and the loop–loop part of the complex, in close interaction between the phosphodiester oxygen O_{2P} atom of the C29 residue, the N_4–H_{42} atom of C30 and the O_2' atom of G5. This intermolecular bridging solvent molecule is in close interaction with the complex 100 ps after the beginning of the production phase and until the end of the simulation. This residence time (more than 1.1 ns) is long enough to be considered as relatively important for the stabilization of the edifice in the GA complex. Note that the average residence time for a water molecule in interaction with a helical structure of a nucleic acid between *n* and *n+1* stems (O_n ... *W*...O_{n+1}) has

been estimated by MD simulation to be less than 400 ps [27, 30].

In GC and CU complexes, such a remarkable interaction bridging the TAR and the aptamer parts of the complex throughout the simulation is not observed. Even if a water molecule shows an interaction between the same two residues as mentioned above (G5/C30), its residence time is no longer than 200 ps and the water exchanges regularly with bulk, in agreement with the residence time computed on classical Watson–Crick helices [35, 36].

Indeed, at the closing base pair location of the GC complex, no rich solvation site is characterized. The solvation of the CU base pair is much more important than in the GC one, probably due to regular exchange of water molecule around the structure, revealing a loose interaction of the base pair. Although the first residence time is about 140 ps (60–200 ps), the longest of the simulation, the molecule considered regularly binds the CU base pair, but leaves the binding site within less than 60 ps. This rapid exchange is the result of a loose interaction at the closing mismatch CU stem level. Cytosine and uracil are too small to achieve a tight interaction in this configuration. This requires a large closing base pair even with an inserted water molecule. This behavior is different from that found when CU mismatches are found in a regular A-form helix, with residence times for water-mediated CU mismatches of more than 400 ps [26]. These residence times are consistent with their shorter inter-base C_1'–C_1' distance, confirming a tighter C_wU interaction in a regular helix than in our loop-loop complex.

The presence of Mg^{2+} ions around the structure appears to have an effect on the dynamic equilibrium of the structures investigated. Nonetheless, as noted previously, such ions were proved to be required for stable complex formation [7, 10]. We note that all the Mg^{2+} counterions are located in the major groove of the helix, in connection with one or two oxygen atoms of the phosphodiester linkers since the major groove is flanked by two negatively charged phosphate clusters. The strong electrostatic interactions lead to clearly localized Mg^{2+} ions along the simulations. The resulting Mg–O distances are 1.97 Å whatever the number of oxygen atoms in contact with the cation. The ions location is roughly the same within the three structures considered. An ion is located near the loop–loop helix, modifying the dynamic interaction between one or two stems of the helix. The effect of the ions appears to be particularly important at the junction between the TAR stem and the loop–loop helix and between the loop–loop helix and the aptamers stems. The torsion induced by the loop–loop formation leads to the presence of several phosphodiester linkers in the same region, and can be considered as a preferential binding site for divalent cations [8].

Interaction energy

In order to gain a better understanding of the closing base pair evolution as well as to analyze more deeply their influence on the stability of the corresponding complex, several quantitative evaluations of base interactions have been performed using Eq. 1. The interaction-energy values between bases involved in the hexanucleotide loop–loop structure, and between closing base pairs are shown in Table 2. The large standard deviation associated with some base pairs indicates conformational heterogeneity. Beaurain et al. [19] have computed stacking-interaction energies and concluded that the GA mismatch led to the strongest stacking interaction amongst their three complexes studied. Here, we examine the strength of the loop–loop helix, to give additional insight on the complexes' behavior.

Considering the GA complex, the mean interaction energy computed for the six pairs of complementary bases contributes to the stabilization of the molecular structure. Note however, that a high standard deviation with a low interaction energy is reported for the G32-C9 and G34-C7 pairs. These fluctuations arise from the fact that an Mg^{2+} ion is located in the surroundings of these two stems. The small evolution of the distance between the ion and the base pair induces a large perturbation of the interacting scheme between the two bases, due to the charge of the ion.

Although the interaction between G and A bases is non-canonical, the two bases remain in interaction through at least one H-bond during all the simulation, with eventually the presence of an inserted water molecule. We have calculated the interaction energy involving, or not, a selected water molecule, whose behavior was considered as representative. As reported previously, solvent molecules can play crucial roles in the affinity between ligands [55] or for example on the conformation of nucleic acid mismatches or bulges. [56–58] The interaction energy with the inserted water molecule oscillates around −20.1 kcal mol^{-1}, in agreement with the value computed at the ab initio MP2 level (−25.8·kcal mol^{-1}) [25]. The system also samples configurations where no water molecule is inserted but where the base pair still shows a large C_1'-C_1' distance. This type of configuration is sampled during a total of ~300 ps of the 2 ns simulation, indicating that solvation of the closing base pair is not a static effect. The resulting interaction energy is about −7 kcal·mol^{-1}, half the ab initio optimal value (−14.2 kcal·mol^{-1}) [25].

Only two short periods within the simulation have been observed that show an interaction scheme similar to that found at the ab initio level, with an interaction energy reaching approximately −13 kcal·mol^{-1}. These two periods correspond to a double H-bond interaction pattern between G and A, i.e., no water molecule is inserted between them. Then the GA average interaction can be decomposed into several interacting schemes: single H-bonded interaction, water-mediated interaction and short periods of doubly H-bonded interactions.

The GC base pair is canonical and shows an optimal interaction along the simulation. The value obtained from our simulation (−27.1 kcal mol^{-1}) is very close to the ab initio energy calculated on a fully relaxed base pair (see Table 2), with small variations around this value. Concerning the loop–loop helix interaction energy, conclusions made for the GA complex are also valid, for we have also observed the presence of a Mg^{2+} ion near the fourth stem (G33-C8), resulting in a slide of the corresponding base pair.

Analysis of the CU base pair was performed following the procedure used for the GA complex. The CU interaction samples some very weak stabilization energies corresponding to the opening of the CU base pair, already observed on the C_1'-C_1' distance. The large distance between the bases leads to an average interaction energy of −5.6 kcal·mol^{-1}.

Although the CU base pair can be considered in a water-mediated interaction scheme, the water molecule present in the starting structure remains inserted for only ~45 ps, before its replacement by another solvent molecule. Indeed, after 100 ps, the original water molecule is far enough to be considered to be totally dissociated from the CU stem. The three-body interaction energy (C, U and the considered water) fluctuates around −15 kcal·mol^{-1} and represents about 66% of the value obtained from a relaxed water-mediated base pair at the ab initio level (−22.8 kcal·mol^{-1}) [24, 25]. The following stem of the loop–loop helix (C30-G11) shows a weaker interaction energy than in the other complexes, in agreement with the structural evolution of the backbone at the loop-loop junction. As for the two preceding complexes, one can note in Table 2 a weak average interaction energy in the last stem of the loop caused by the presence of an Mg^{2+} ion.

Conclusions

Three molecular dynamics simulations of 2 ns have been performed on RNA hairpin-loop complexes involving HIV TAR RNA and aptamers whose closing-loop base pairs were a Watson–Crick-type GA, a Watson–Crick GC and a water-mediated CU. The objectives of this study were to analyze the influence of the closing base pair on complexes bearing an hexanucleotide loop–loop interaction. Experimental data at physiological Mg^{2+} concentration had shown that the GA complex has a large affinity for HIV-1 TAR RNA, GC a medium affinity and CU the weakest one.

The three simulations show RMSD below 3 Å, revealing stable simulations, where the complexes are composed of a bent quasi-continuous helix composed of three coaxially stacked helices. Simulations considering other closing base pair structures (direct CU and reverse Hoogsteen GA) lead to a large structural deviation of the complex through denaturation of the closing base pair. From a structural point of view, several features can be pointed out, considering the three complexes with TAR:

1. The GA aptamer derivative appears to be more flexible than the other complexes. Indeed, it leads to the largest C_1'-C_1' distance of the three simulations, allowing more flexibility than in the GC complex. The CU complex shows an elongation of the C_1'-C_1' distance, but reveals a denaturation of the water mediated CU starting structure. The sugar-puckering analysis shows a large allosteric effect of the CU complex, suggesting the necessity for TAR to reorganize upon binding. Such reorganization seems less effective for the other complexes.
2. Although both GA and CU complexes show a widely opened major groove, allowing larger flexibility and larger solvation of the base pair, contrarily to the GC pair, only GA shows a structurally stable complex at the loop–loop closure. Such a feature indicates a better compromise between affinity for TAR and the intrinsic structural stability of the GA aptamer.

Energetic analysis of the simulations has allowed us to quantify the interaction energy of closing base pairs together with the interactions within the hexanucleotide loop–loop sub-structure. The Watson–Crick GC base pair remains in close interaction with an optimal interaction energy throughout the simulation. The GA base pair samples three types of interacting schemes, with a single H-bond, a water-mediated interaction and few two H-bond interactions. The CU base pair does not show a large energetic affinity along the simulation and this weak interaction results in a destabilization of the vicinal base pair in the loop–loop structure.

Considering the mismatch base pairs (GA and CU), we have shown that inserted water molecules exchange regularly with bulk but the residence time was longer within the GA mismatch, revealing a tighter interaction. Also, we have detected the presence of a bridging water molecule, found only in the GA complex, presenting a long residence time (more than 1.1 ns). This type of water molecule showed a much shorter residence time in the two other systems, and could be the origin of a relative increased stability of the complex composed of the GA mismatch, even if such a water molecule could go away from its binding site without any substantial harm. Such considerations should be more deeply investigated in further work.

We note that a large majority the closing base pairs showing experimental strong affinities for TAR bear purine-type bases, suggesting that at least a large major groove is required for flexibility in such rigid complexes, but also that the direct affinity between bases is crucial. The GA closing base pair seems to allow such a compromise, allowing structural flexibility and energetic stability through direct or water-mediated bonding. The behavior of such building blocks should be taken into account in the design of loop containing RNA motifs.

Acknowledgements The authors acknowledge the CINES computing center for providing us with computing time. JG thanks Dr.

Martin Swain for his help in the preparation of the manuscript and the referees for clever suggestions.

References

1. Cullen BR (1994) Infect Agents Dis 3:68–76
2. Jones KA (1997) Genes Dev 11:2593–2599
3. Karn J (1999) J Mol Biol 293:235–254
4. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA (1998) Cell 92:451–462
5. Ivanov D, Kwak Y, Nee E, Guo J, García-Martínez L, Gaynor R (1999) J Mol Biol 288:41–56
6. Garriga J, Peng J, Parreno M, Price DH, Henderson EE, Grana X (1998) Oncogene 17:3093–3102
7. Chang KY, Tinoco I Jr (1994) Proc Natl Acad Sci USA 91:8705–8709
8. Chang KY, Tinoco I Jr (1997) J Mol Biol 269:52–66
9. Eguchi Y, Tomizawa JI (1990) Cell 60:199–209
10. Eguchi Y, Tomizawa JI (1991) J Mol Biol 220:831–842
11. Gregorian RSJ, Crothers DM (1995) J Mol Biol 248:968–984
12. Ducongé F, di Primo C, Toulmé JJ (2000) J Biol Chem 275:21287–21293
13. Ellington AD, Szostak JW (1990) Nature 346:818–822
14. Tuerk C, Gold L (1990) Science 249:505–510
15. Serra MJ, Lyttle MH, Axenson TJ, Schadt CA, Turner DH (1993) Nucleic Acids Res 21:3845–3849
16. Santa Lucia J, Kierzek R Jr, Turner DH (1991) Biochemistry 30:8242–8251
17. Schroeder SJ, Turner DH (2001) Biochemistry 40:11509–11517
18. Lodmell JS, Ehresmann C, Ehresmann B, Marquet R (2001) J Mol Biol 311:475–490
19. Beaurain F, di Primo C, Toulmé JJ, Laguerre M (2003) Nucleic Acid Res 31:4275–4284
20. Réblova K, Spackova N, Sponer JE, Koca J, Sponer J (2003) Nucleic Acid Res 31:6942–6952
21. Beaurain F, Laguerre M (2003) Oligonucleotides 13:501–514
22. Réblova K, Spackova N, Stefl R, Csaszar K, Koca J, Leontis NB, Sponer J (2003) Biophys J 84:3564–3582
23. Pattabiran N, Martinez H, Shapiro B (2002) J Biomol Struct Dyn 20:311–486
24. Brandl M, Meyer M, Sühnel J (1999) J Am Chem Soc 121:2605–2606
25. Brandl M, Meyer M, Sühnel J (2000) J Phys Chem A 104:11177–11187
26. Schneider C, Brandl M, Sühnel J (2001) J Mol Biol 305:659–667
27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
28. Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling C, Darden T, Merz KMJ, Stanton RV, Chen A, Vincent JJ, Crowley M, Tsui V, Radmer R, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner P, Kollman PA (1999) AMBER 6.0. University of California, San Francisco
29. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratman RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu C, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Gonzales C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian Inc, Pittsburgh
30. Golebiowski J, Antonczak S, Di-Giorgio A, Condom R, Cabrol-Bass D (2004) J Mol Mod 10:60–68
31. Hobza P, Kabelác M, Sponer J, Mejzlik P, Vondrasek J (1997) J Comp Chem 97:1136–1150
32. Hobza P, Sponer J (1999) Chem Rev 99:3247–3276
33. Auffinger P, Westhof E (2001) Biopolymers 56:266–274
34. Cheatham TE III, Young MA (2001) Biopolymers 56:232–256
35. Auffinger P, Westhof E (2000) J Mol Biol 300:1113–1131
36. Auffinger P, Westhof E (2001) J Mol Biol 305:1057–1072
37. Auffinger P, Westhof E (1997) J Mol Biol 269:326–341
38. Gresh N, Sponer JE, Spaková N, Leszczynski J, Sponer J (2003) J Phys Chem B 107:8669–8684
39. Wu M, Turner D (1996) Biochemistry 35:9677–9684
40. Wimberly B, Varani G, Tinoco I Jr (1993) Biochemistry 32:1078–1087
41. Szewczak AA, Moore PB, Chan YL, Wool IG (1993) Proc Natl Acad Sci U S A 90:9581–9585
42. Heus HA, Pardi A (1991) Science 253:191–194
43. Biou V, Yaremchuck A, Tukalo M, Cusack S (1994) Science 263:1404–1410
44. Rüdisser S, Tinoco I Jr (2000) J Mol Biol 295:1211–1223
45. Ye X, Gorin A, Ellington AD, Pate DJ (1996) Nat Struct Biol 3:1026–1033
46. Ennifar E, Yusupov M, Walter P, Marquet R, Ehresmann B, Ehresmann C, Dumas P (1999) Structure 7:1439–1449
47. Correll CC, Freeborn B, Moore PB, Steitz TA (1997) Cell 91:705–712
48. Westhof E, Fritsch V (2000) Structure 8:55–65
49. Fan P, Suri AK, Fiala R, Live D, Patel DJ (1996) J Mol Biol 258:480–500
50. Tanaka Y, Fujii S, Hiroaki H, Sakata T, Tanaka T, Uesugi S, Tomita KI, Kyogoku Y (1999) Nucleic Acids Res 27:949–955
51. Leontis NB, Westhof E (1998) Q Rev Biophys 31:399–455
52. Lavery R, Sklenar H (1988) J Biomol Struct Dyn 6:63–91
53. Gao YG, Robinson H, van Boom JH, Wang AJH (1995) Biophys J 69:559–568
54. Weeks KM, Crothers DM (1993) Science 261:1574–1577
55. Okimoto N, Tsukui T, Kitayama K, Hata M, Hoshino T, Tsuada M (2000) J Am Chem Soc 122:5613–5622
56. Zacharias M, Heinz S (1999) J Mol Biol 289:261275
57. Zacharias M (2000) Curr Opin Struct Biol 10:311–317
58. Zacharias M (2001) Biophys J 80:2350–2363