J Mol Model (2004) 10:60–68
DOI 10.1007/s00894-003-0170-9

ORIGINAL PAPER

Jérôme Golebiowski · Serge Antonczak · Audrey Di-Giorgio · Roger Condom · Daniel Cabrol-Bass

Molecular dynamics simulation of hepatitis C virus IRES IIId domain: structural behavior, electrostatic and energetic analysis

Received: 28 May 2003 / Accepted: 31 October 2003 / Published online: 23 December 2003

Abstract The dynamic behavior of the HCV IRES IIId domain is analyzed by means of a 2.6-ns molecular dynamics simulation, starting from an NMR structure. The simulation is carried out in explicit water with Na^+ counterions, and particle-mesh Ewald summation is used for the electrostatic interactions. In this work, we analyze selected patterns of the helix that are crucial for IRES activity and that could be considered as targets for the intervention of inhibitors, such as the hexanucleotide terminal loop (more particularly its three consecutive guanines) and the loop-E motif. The simulation has allowed us to analyze the dynamics of the loop substructure and has revealed a behavior among the guanine bases that might explain the different role of the third guanine of the GGG triplet upon molecular recognition. The accessibility of the loop-E motif and the loop major and minor groove is also examined, as well as the effect of Na^+ or Mg^{2+} counterion within the simulation. The electrostatic analysis reveals several ion pockets, not discussed in the experimental structure. The positions of these ions are useful for locating specific electrostatic recognition sites for potential inhibitor binding.

Keywords Simulation · AMBER · HCV · IIId domain · RNA

J. Golebiowski (✉) · S. Antonczak · D. Cabrol-Bass
Laboratoire Arômes Synthèses Interactions,
Faculté des Sciences de Nice Sophia-Antipolis,
06108 cedex 2 Nice, France
e-mail: Jerome.Golebiowski@unice.fr
Tel.: +33 04 92 07 61 03
Fax: 33 04 92 07 61 25

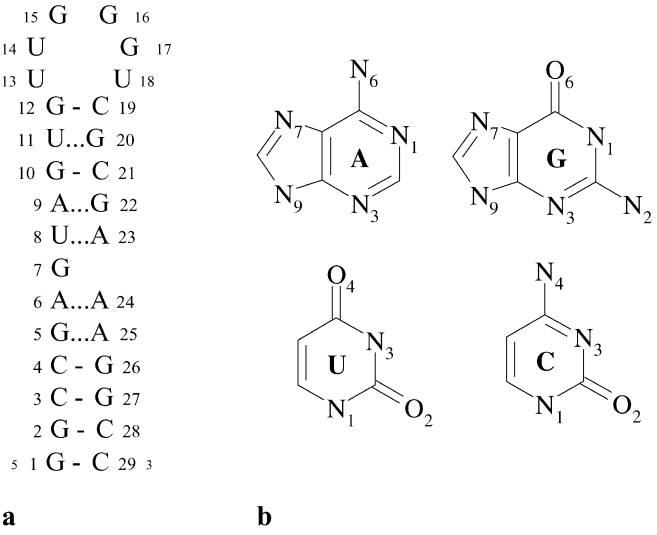
A. Di-Giorgio · R. Condom
Laboratoire de Chimie Bioorganique,
Faculté des Sciences de Nice Sophia-Antipolis,
06108 cedex 2 Nice, France

Introduction

Initiation of transcription of the hepatitis C virus (HCV) occurs through the mediation of an element called IRES (internal ribosome entry site), [1] located in the 5' untranslated region (5'-UTR) of the genomic RNA. This 5'-UTR adopts a folded structure that forms characteristic domains, composed of stem-loop structures. All the domains forming the IRES region (i.e. domains II to IV) appear to contribute strongly to IRES activity, [2] and particularly domains III and IV, which are critical for viral transcription. [3] Domain III constitutes several stem-loop motifs, denoted IIIa to IIIf, of which IIId has been extensively studied owing to its implication in IRES activity. [4] Indeed, IIId is directly involved in the binding of the 40S ribosomal sub-unit, essential for the initiation of transcription, and thus represents an interesting target for potential inhibitors. [5]

Secondary and tertiary structures of IIId have recently been described. [4, 6, 7, 8] It is composed of two helical regions, separated by a bulge and closed by an apical loop made up of six nucleotides (see Scheme 1), and which contains a highly conserved GGG triplet that was shown to play a key role in the recognition of ribosome 40S subunit. [4]

Two IRES IIId tridimensional structures have been elucidated by two different approaches. An experimental NMR structure was proposed by Lukavsky et al., [7] and Klinck et al. proposed another structure using a motif-based approach, combining motif prediction, NMR and mutation analysis. [8] The two structures are very similar concerning the helical part of the IIId domain, assigning a loop-E motif between G5 and A9, flanked by two A-form helices. The only difference in the apical loop concerns the description of the U13 U14 sequence. According to Klinck et al., U13 interacts closely with G17, and the turn of the loop is attributed to a U-turn motif for U14 G15 G16, whereas Lukavsky et al. consider that U14 is located in the major groove of the loop, stacking loosely on the 5' side residues.



a **b**

Scheme 1 **a** HCV IRES IIId RNA, including the first additional GC step. **b** Atom definition used in the following

Molecular dynamics simulations have proved to be a powerful tool for analyzing the details of the dynamical evolution of biomolecular structures, as well as the solvent and counterion interactions with the structure. The rational design of an inhibitor requires a good knowledge of the dynamic behavior of the target structure. Up to now, no dynamic analysis of the IRES IIId domain in explicit water phase has been reported. Ion-location analysis should be helpful in the understanding of the potential binding sites as well as the effect of ions in the tertiary fold on the structure. In the present work we investigate the behavior of the IRES IIId domain, using molecular dynamics simulations, starting from an experimental structure.

Methods

The starting structure was taken directly from the experimental NMR data given by Lukavsky et al. [7] Molecular dynamics simulations were carried out using the AMBER 6 [9] program in the isothermic–isobaric thermodynamic ensemble at 300 K. The helix charge was neutralized by addition of 28 Na^+ counterions, placed in the most electronegative region of space using the LEAP module. The system was then embedded in a periodic box containing 7687 TIP3P [10] water molecules. The water phase was extended to a distance of 15 Å from every solute atom. Simulation was carried out with the SANDER module using the SHAKE algorithm 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 adding the ions and the water molecules to the minimized structure, 1,000 steps of minimization keeping the complex and the ions fixed were performed using particle mesh Ewald (PME) summation for electrostatic interactions. PME parameters were chosen to obtain a grid spacing close to 1 Å and a 9-Å direct space cutoff. Equilibration was continued with 50-ps dynamics, keeping the solute fixed. Then, 1,000 steps of minimization and 10 ps of MD simulation using a restraint of 20 kcal $(\text{mol Å}^2)^{-1}$ on the solute atoms were performed, followed by four rounds of 1,000 steps minimization reducing the restraints by 5 kcal $(\text{mol Å}^2)^{-1}$ at each round, with 10-ps MD simulation. Further, the system was slowly heated from 100 to 300 K over a period of 20 ps. The equilibration was kept over 100 ps. Finally, a 2.6-ns production phase was started, saving the trajectory each picosecond. The same procedure was applied for a 1-ns simulation with 14 Mg^{2+} counterions.

Root mean square deviations (RMSd) calculated on heavy atoms over the trajectories were obtained with the Carnal module. Interaction energy analysis between selected NA bases was 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 the 2,600 structures taken from the trajectory and estimated as follows:

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

where E^n are the energies of the NA bases being in interaction. $E^{abc...n}$ is the energy of the supermolecule composed of several bases.

The energy was obtained using the Cornell et al. force field implemented in Gaussian 98, [11] using the parm96.dat potential (see for example [12]). Estimation of such interaction energies at this level of theory was shown to be accurate enough to lead to reasonable conclusions on the stability of the base pairing. [13, 14]

About simulations lengths

The question of the relation between a simulation length and its significance is regularly addressed. Although it seems clear that better sampling is obtained by means of long simulation, Auffinger and Westhof have pointed out that sampling conformational space over long time scales using an inappropriate description of the system or inappropriate protocols is useless. [15] 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. [16] More precisely, in nucleic acids systems, intermolecular interactions are strong enough to ensure a relatively good structural description, related to an experimental structure. The conformational sampling of double stranded NA can then be considered as ergodic within nanosecond time scale MD simulations. Moreover, solvation pattern and monovalent ion binding analysis can reasonably be realized by means of few nanoseconds MD simulations, but only on a semiquantitative basis, since the residence time of such binding takes place into a nanosecond regime (see for example [15, 17, 18]). In contrast, divalent ion desolvation processes occur within ~10 μs, and characteristics related to such processes 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 such an ion-binding analysis provides additional pertinent insights on the highly hydrophilic regions of the nucleic acid.

Results and discussions

The HCV IRES IIId domain can be decomposed into three sub-domains, each of them showing specific characteristics. First, the structure is composed of a right-handed A-form helix (nt 1–12; 19–29), which can be subdivided into two parts: a regular helical part that covers nt 1–4 and 26–29 plus stems 10–21 and 11–20, and a loop-E motif that stretches from G5 to A9 and from G22 to A25. Note also that the major groove cannot be

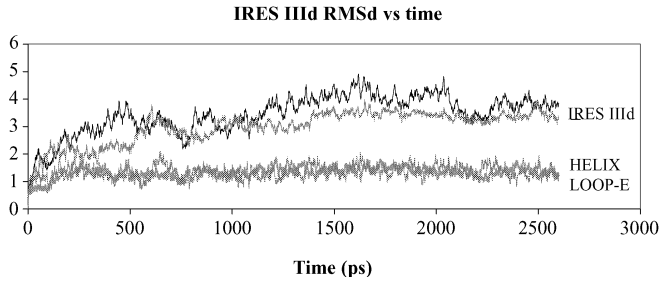


Fig. 1 Root mean square deviation RMSd of HCV IRES IIId RNA and selected substructures, compared to experimental values (see [7]) along 2.6-ns MD simulation. See text for details

followed all along the structure, since it is crossed by an AUG triple interaction. Finally, the third sub-domain is composed of a hexanucleotide loop (nt 13 to 18).

Root mean square deviation (RMSd)

During the whole production phase, the IIId domain remains faithful to its starting structure although an average RMSd calculated over all the heavy atoms of the experimental structure along the simulation reaches 3.4 Å. To gain insight into the dynamics of each motif of the macromolecule, RMSd has also been computed on the three parts of IIId cited above. Figure 1 reports RMSd plots for the whole structure, the regular helical part of the structure (nt 1–4; 26–29 and 10–12, 19–21), the loop-E motif (nt 5–9; 22–25) and the hexanucleotide loop (nt 13–18).

The average RMSd for the three sub-domains of IRES IIId are 1.3 Å, 1.3 Å and 2.9 Å for the regular helical part, the loop-E motif and the hexanucleotide loop, respectively. Note that a small RMSd variation is reported for both the helical part and the loop-E motif, although the latter bears some non-canonical interactions. This indicates the presence of rather strong intramolecular interactions in this region, contributing to a low structural variation of this part of the structure. In contrast, the terminal loop structure is clearly more flexible; nucleotides from G15 to U18 undergo large variations of their positions along the simulation with RMSd values rising up to 3.7 Å. Also a rise of the RMSd is observed between 500 and 700 ps. This point will be discussed further.

Dynamical structural analysis

Figure 2 presents the superposition of IRES IIId structures along the 2.6-ns simulation.

Along the simulation, only non-standard interactions have been reported at the apical loop level. More precisely, the U18 base stays outside the loop during the whole simulation and does not interact with any other nucleotide. The three consecutive guanine bases (G15, G16, G17 in Scheme 1) show different behaviors. Two of them (G15, G16) are stacked together and located outside the loop, above the minor groove. According to Eq. (1), the trajectory-averaged stacking energy between these two bases has been estimated as −3.7 kcal mol^{-1} (standard deviation: 1.0 kcal mol^{-1}). Due to their specific position, the Watson–Crick (WC) sites as well as the N_7 atoms of these two bases do not interact with other elements of the structure along the simulation. This point confirms the key role of the loop in the intermolecular recognition site involving IRES IIId.

Fig. 2 Superposition of 14 structures representative of the evolution of IRES IIId RNA along 2.6-ns MD simulation

Analysis of the third guanine base (G17) puts forward the existence of two different positions clearly separated in time. During the first 700 ps, G17 is located outside the loop, above the minor groove, giving rise to a close interaction between its WC face and the phosphodiester linker of G15, as is reported in the experimental NMR structure of Lukavsky et al. [7] The resulting average energy of this interaction has been estimated to be −25 kcal mol^{-1} over the period considered. Moreover, the N_7 and O_6 atoms of the G17 base are in close interaction with water molecules, making a bridge between G17 and the two other guanines (G15 and G16). Those bridging molecules often exchange with bulk and show a short residence time (less than 100 ps). This short residence time reveals a non-specific interaction within the guanine triplet site, compared to other sites. A structural reorganization of the loop is observed after ~700 ps of simulation. This structural change leads to the formation of an H-bond between H_3 of U13 and O_6 of G17. The WC face of G17 now points into the major groove. This interaction within the major groove is highly favored by the formation of an ion pocket where an Na^+ ion interacts with N_7 and O_6 atoms of G17, as shown in Fig. 3. This structural feature is conserved for the rest of the simulation. Considering the mobility of this guanine

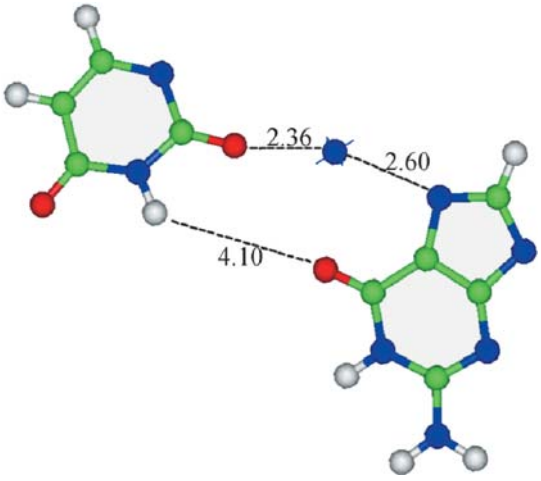


Fig. 3 Snapshot of a typical Na^+ mediated U13–G17 interaction along the simulation

base, a few comments can be put forward. The residence time of a conformational switching of a weaker water-mediated CU base pair has been estimated to be at least 1 ns. [19] Then, the conformational change around G17 qualitatively suggests a mobility of this part of the loop, confirmed by the different G17 positions within the loop in the NMR experimental structures. [7] Also, the starting structure has been taken from amongst the 25 experimental structures, for which the G17 location was different. The simulation confirms such a mobility, as shown in Fig. 2 and further analyses. The evolution of G17 goes towards an interaction with U13, and the structure is then similar to the structure given by Klinck, or other NMR structures given by Lukavsky, different from our starting structure. Thus, the simulation data are fully consistent with the experimental facts concerning G17 mobility. These findings justify a dynamic approach concerning the structural analysis of IRES IIId, that cannot be fully described using a single structure. These results also suggest that this conformational change of G17 could be reversible in a time that cannot be quantified, since it depends on the presence of a Na^+ ion in the surroundings of the G17–U13 base pair.

Finally, concerning the two last bases of the loop, one can note that U14 stays in the major groove of the loop and interacts with U13 through stacking interactions.

H-bond analysis

To analyze the interactions between bases in such a structure in more depth, H-bond occupancy was computed (Table 1). An H-bond has been considered when the distance between two heavy atoms was less than 4 Å and the angle (X–H...O or N with X= N or O) was larger than 60°. The atom assignments are given in Scheme 1.

Table 1 clearly shows that the two strands of the regular helix part of the structure remain in close interaction throughout the simulation. The G7–U8–A23 triplex bonds are highly conserved and confirm the potential role of the triplex as a homing site for molecular recognition. The average interaction energy of the triplex has been evaluated as -20.5 kcal mol^{-1} with a standard deviation of 0.2 kcal mol^{-1}. This three-body interaction can be split into three contributions. The energy is −11.9 and −9.1 kcal mol^{-1} for the AU and GU interactions, respectively. For the third contribution, since the uracyl is the central base, no direct interaction is observed between A23 and G7 and the resulting interaction energy is close to zero (+0.5 kcal mol^{-1}). The N_7 atom of G7, which points towards the bulk, is not involved in any H-bond and contributes towards defining a Sarcin–Ricin Loop (SRL) type recognition site together with the phosphate linkers of G22 and C21. [20] H-bond occupancy of ~50% between U13 and both C19 and G17 reveals the flexibility

Table 1 Selected IRES IIId base pairs H-bond occupancy (HB) along 2.6-ns MD simulation. Only values greater than 50% are presented

Base pair	Bonded atoms	HB (%)
G1–C29	O_6...H_{42}–N_4	99.8
	N_1–H_1...N_3	98.0
	N_2–H_{21}...O_2	97.7
G2–C28	O_6...H_{42}–N_4	100.0
	N_1–H_1...N_3	100.0
	N_2–H_{21}...O_2	99.9
C3–G27	N_4–H_{42}...O_6	100.0
	N_3...H_1–N_1	100.0
	O_2...H_{21}–N_2	100.0
C4–G26	N_4–H_{42}...O_6	99.5
	N_3...H_1–N_1	99.9
	O_2...H_{21}–N_2	100.0
G5–A25	N_3...N_6–H_{62}	99.9
	N_2–H_{21}...N_7	100.0
	O_2'...H_{62}–N_6	100.0
A6–A24	N_6–H_{62}...N_7	100.0
	N_7...H_{62}–N_6	100.0
	O_1P...H_{61}–N_6	96.7
U8–A23	N_3–H_3...N_1	100.0
	O_2...H_{62}–N_6	85.0
G7–U8	N_2–H_{22}...O_4	100.0
G7–A23	N_2–H_{21}...O_2P	98.5
	N_2–H_{21}...O_5'	95.0
	N_1–H_1...O_2P	98.1
A9–G22	N_6–H_{62}...N_3	99.6
	N_6–H_{62}...O_2'	99.7
	N_7...H_{22}–N_2	100.0
G10–C21	O_6...H_{41}–N_4	100.0
	N_1–H_1...N_3	89.4
	N_2–H_{21}...O_2	100.0
U11–G20	N_3–H_3...O_6	89.6
	O_2...H_1–N_1	99.6
G12–C19	O_6...H_{41}–N_4	83.6
	N_1–H_1...N_3	97.7
	N_2–H_{21}...O_2	98.0
U13–C19	N_4...H_{41}–N_4	82.7
	N_3–H_3...N_3	50.0
	O_4...N_4–H_{41}	51.7
U13–G17	N_3–H_3...O_6	55.1

in this part of the loop. Note that this percentage takes into account interaction distance up to 4 Å and explains how the U13–C19 H-bond occupancy can be accounted for, even after the 700-ps reorganization discussed above. H-bond analysis also confirms that both G15 and G16 remain fully accessible since they do not interact through their Watson–Crick sites, whose H-bond occupancy never reach the value of 10% (data not shown) and are completely accessible for molecular recognition. These results are in accordance with the chemical probing given by Lukavsky et al., indicating that the three guanine Watson–Crick faces are almost completely protected after 40S ribosome sub-unit binding. [7] Nonetheless, on the autoradiograph of kethoxal probing proposed by Lukavsky et al., the G17 Watson–Crick face appeared to behave slightly differently from G15 and G16, suggesting a different behavior upon binding of the ribosome 40S subunit. Indeed, Kieft et al. suggested that the G17C mutant behaved differently from G15C or G16C upon enzyme cleavage. [6] Our results might rationalize those findings.

Sugar puckering

Sugar puckering angle analysis allows analysis of the structural feature of the RNA double helix along the simulation. It is representative of the helical skeleton distortion as well as the "structural rigidity" of a nucleoside considered. Figure 4 shows the dynamic evolution of the sugar puckering pseudorotation angles along the 2.6-ns simulation. The angles are represented on a trigonometric circle, where the radius represents the time evolution along the simulation.

Analysis of the helix part, the loop-E motif and then the apical loop substructure of the RNA are reported. The helical part of IRES IIId (up to step G5–A25, and steps 10 and 11) adopts a quite rigid structure, with all but the starting G1–C29 angles presenting classical *A*-RNA C3'-endo values (between 0 and 36°). In general, a loop-E motif can present several sequences, leading to different rigidity. [21] The one encountered in IRES IIId is flanked by two Watson–Crick GC base pairs and does not show particular mobility along the simulation. In detail, A6 (the beginning of the loop-E) motif, is the first nucleoside showing a reverse angle pucker with respect to classical RNA. G7, involved in the AUG triplex has a C2'-endo pucker with instantaneous values largely deviating from this average angle, revealing a larger flexibility of the loop-E RNA substructure than a classical A-form RNA. Recall that even if the RMSd is identical to the helical region, it has been evaluated on fewer stems than the latter. One can also notice that U8–A23 pseudorotation angles are a little lower than the C3'-endo reference value, since they oscillate around 0°. The apical loop substructure presents several pucker angle variations along the simulation. At the beginning of the loop, G12 and U13 sugars oscillate largely around their mean value, with angles reaching −20° to +40° for G12 and up to 80° for U13. The S-turn made up of U14, G15 and G16 leads to C2'-endo pucker for those three sugars, with relatively small variations around this value. The G17 sugar backbone presents three different structural features along the simulation, going from C4'-endo (252 to 216°) to C1'-exo (108 to 144°) at the end of the simulation. These puckering differences are attributed to the base orientation, which binds U13 after ~700 ps of simulation and whose WC face is located in the major groove, as discussed previously. This large evolution of the sugar backbone suggests a high flexibility of the loop substructure and confirms that the S-turn provides a unique surface for specific interaction with other RNA elements or with proteins. One can also note the huge structural

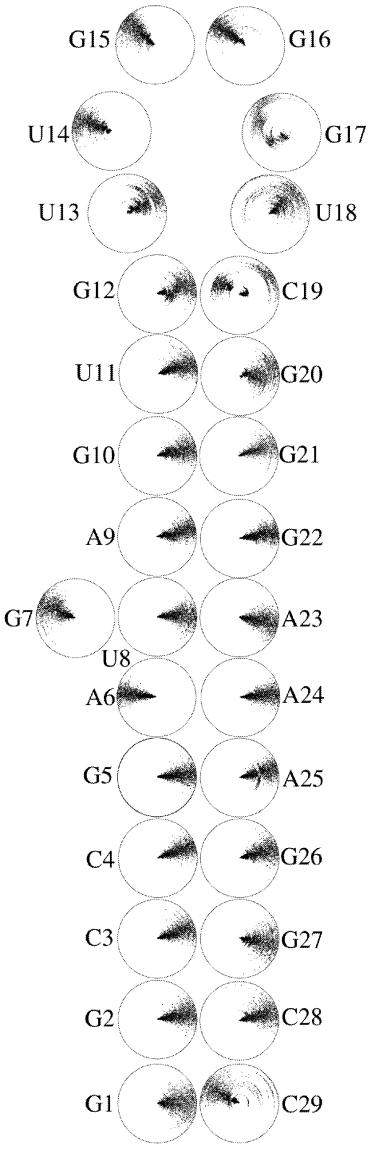


Fig. 4 Polar plot of the nucleosides sugar pseudorotation puckering angle along the 2.6-ns simulation. The angle is measured in a trigonometric circle, the horizontal right axis representing a zero angle. The radial component corresponds to time. *t*=0 corresponds to the center and 2.6 ns to the boundary of the circle

Table 2 2.6-ns averaged groove widths in IRES IIId RNA, in Å, standard deviation in parentheses. Minor groove width is measured by means of O_2'–O_2' distance and major groove width is measured through interphosphate distance

	Minor groove width	Major groove width
12–19	11.1 (0.6)	19.7 (0.9)
11–20	10.8 (0.7)	19.7 (0.8)
10–21	10.9 (0.4)	19.0 (0.5)
9–22	10.4 (0.2)	15.8 (0.5)
8–23	13.3 (0.3)	12.9 (0.3)
6–24	13.9 (0.3)	11.9 (0.2)
5–25	10.3 (0.3)	15.9 (0.4)
4–26	11.1 (0.4)	19.1 (0.4)
3–27	11.3 (0.4)	19.5 (0.6)
2–28	11.1 (0.4)	18.8 (0.5)

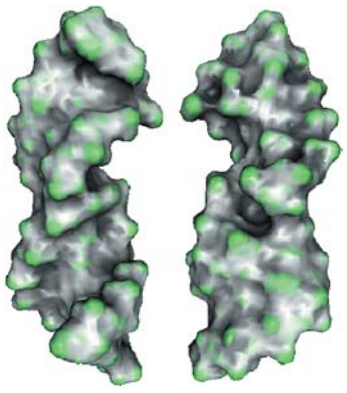


Fig. 5 IRES IIId minor (*right*) and major (*left*) groove, colored by curvature

evolution of the U18 bulged out nucleoside, which leads to a large perturbation of C19.

This analysis reveals large fluctuations around their mean value for a majority of the nucleosides. One can also notice that fluctuations of a considered stem strongly influence its neighbors, suggesting that any perturbation provided by an intermolecular recognition would result in a structural reorganization of the whole structure.

Accessibility of the minor and major grooves

The large number of non-canonical interactions within IRES IIId lead to a structure in which the grooves are largely distorted. For each stem of the helical part of IRES IIId, Table 2 reports the interphosphate as well as the O_2'–O_2' distances, which are representative of the major and minor groove widths. By inspecting these values, one can first note that since the standard deviations are rather small (~5%), no drastic change has been found for each groove width along the simulation. The largest deviation is reported for 11–20 and 12–19 stems, whose interactions are influenced by the structural reorganization within the loop.

Concerning the major groove width, two different trends depending on the stem involved can be reported. While the stems located at each extremity of the helical part show interphosphate distances slightly larger than the values found in classical RNA (18 Å), the four stems located in the loop-E structure, namely 5–25, 6–24, 8–23 and 9–22, exhibit narrower width, revealing an important narrowing of the major groove, as depicted in Fig. 5. This narrowing is strengthened by a close interaction between the phosphate group of A6 and the NH_2 group of A24 conserved along the simulation, together with the presence of two Na^+ ions, near nucleotides A24 and A25.

Analysis of the minor groove accessibility is of primary importance for the potential recognition of the structure. The minor groove is rich in H-bond acceptor/donor atoms and numerous works have been devoted to the design of selective ligands for a specific reading of this groove, especially in DNA. Hence, the major groove narrowing mentioned above results in an opening of the minor groove, with a distance of ~14 Å between both A6 and A24 sugar O_2' atoms. This opening is at the origin of a large accessibility of the Watson–Crick faces of A23, A24 and A25, just before the AUG triplex motif (see Fig. 5).The triplex formed by G7–U8 and A23 bases causes the minor groove to be opened widely and suggests a possible recognition site for a potential inhibitor. Then, the four first canonical nucleotides lead to a classical A-form RNA deep minor groove, but note that the depth of the minor groove is very irregular across the remaining structure.

Concerning the apical loop of the structure, one can observe that its minor groove only contains the stacked G15 and G16 bases, which are free from interaction. The other groove is wide open and U13, C19 and G17 bases are also largely accessible. This explains the different role of G17 related to G15 and G16 in the IRES IIId recognition process pointed out by Kieft et al. [6] The authors clearly indicated that their enzyme-cleavage pattern of the G17C mutant is different from either G16C and G15C. The latter ones have completely different behavior and are not accessible through the same groove of the loop.

Electrostatic analysis

A better knowledge of the electrostatics of the surface atoms of the IRES IIId structure would be helpful in the rationalization and ranking of the potential interaction sites. To this purpose, we have performed an analysis of both the electrostatic potential mapped on the vdW surface of the 2.6-ns averaged IRES IIId structure (after minimization) and of the behavior of counterions. Note first that the major groove exhibits a large negative potential, around -25 $kT\ e^{-1}$, while the potential in the minor groove does not exceed $-6\ kT\ e^{-1}$. The most negative potential can be found in the loop-E motif, at the S-turn. Indeed, one can notice the very low electrostatic potential created by the vicinity of the phosphate linkers of the nucleotides involved in the S-turn of the helix (nt 5–8, *red* colored in Fig. 6).

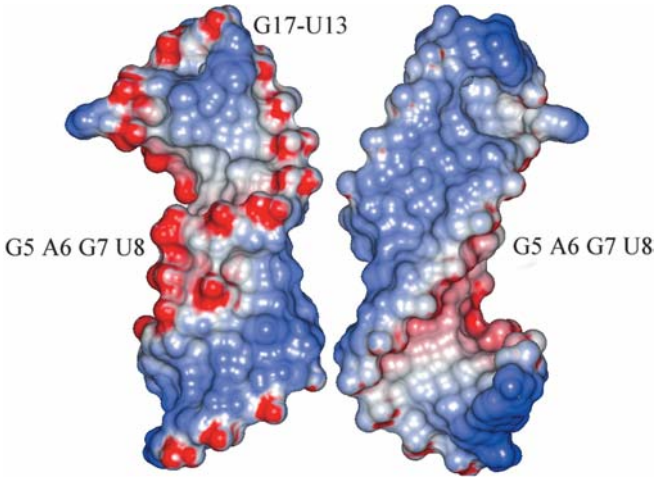


Fig. 6 Minor (*left*) and major groove (*right*) view of the non-linear Poisson–Boltzmann calculated negative electrostatic potential, mapped on the van der Waals surface of the average minimized structure. From −25 $kT\,e^{-1}$ (*red*) to −4 $kT\,e^{-1}$ (*blue*). The location of the phosphate linkers of nucleotides 5 to 8 as well as the 13–17 bases (*blue*, on the *left*) is indicated

The apical loop substructure bears a quite low electrostatic potential, around −5 $kT\,e^{-1}$, for the majority of the accessible sites between U13 and G17. Nonetheless, the N_7 and O_6 atoms of G15 and G16 display a stronger negative potential, suggesting that they can be targeted at a relatively long distance, as well as the phosphate linkers present in the S-turn motif of the loop. Note that, in spite of this stronger electrostatic field, the residence time of water molecules around the guanine bases does not exceed the already short time of 60 ps, revealing that no desolvation penalty will occur upon intermolecular recognition.

Let us now focus on specific cation–base interaction sites, which can reveal interesting characteristics connected to molecular recognition. A large majority of the counterions, whose location was determined at the beginning of the simulation, diffused throughout the simulation box during the trajectory. Nonetheless, ions were found to be located in particular positions with quite long residence times and a few interesting comments can be made regarding their positions. The phosphates pointing into the major groove lead to a more negative electrostatic potential than in the minor groove. We naturally observe that the Na^+ ions placed in the more hydrophobic minor groove at the beginning of the simulation tend to be more mobile than those found in the major groove. Despite chemical acylation experiments suggesting a poor specific recognition of the deep narrow major groove of A-RNA, [22] experimental facts suggest that ions can specifically interact with the major groove, more precisely through interaction with N_7 atoms of guanine bases. [23] Actually, we do observe such interactions in the helical part of the structure, particularly when purine bases are present.

Indeed, in the helical part, three preferential Na^+ binding sites can be considered. G22 O_2P, N_7 and O_6 atoms and G7 O_6 and N_7 atoms are close to each other within the major groove and contribute to form an ion pocket where Na^+ counterions are strongly bound. We have observed two distinct ions in interaction in this site. An Na^+ occurs in this position after 1 ns of simulation and remains for ~1 ns. Then an Na^+–Na^+ ion exchange over a period of ~50 ps is observed and the second ion remains in close interaction until the end of the simulation. A second binding site is found just under the AUG triplex, near the S-turn motif. The large torsion of the helix brings the A6 O_2P atom close to the sugar backbone O_2' atom and the phosphodiester linker oxygen atoms of U8 and G7. An Na^+ ion is found in this strongly negative pocket from t=1.1 ns to the end of the trajectory. The third ion pocket is found in the regular *A*-form helical part. It has been shown that purine–purine steps form ion pockets in *A*-RNA, due to the closeness of N_7, O_2' and O_2P atoms in the major groove. [17, 24] This structural pattern is found both in the 6–5 (ApG) and 26–27 (GpG) steps. Unexpectedly, although GpG steps are considered as stronger binders than ApG ones [17], no specific ion interaction has been found involving G26, G27 and any ion. This ion in the A6 G5 step is observed from t=1 ns to the end of the trajectory. The fact that three ion sites are found around the loop-E motif confirms that the loop-E motif is at the origin of special recognition features that could not be expected considering classical *A*-form helixes.

At the apical loop substructure, due to their large mobility resulting in a diffuse electrostatic potential along the simulation, G15 and G16 bases do not show strong interactions with Na^+ ions. Through its O_6 atom, G17 participates, together with O_2 atoms of both U13 and C19, in the formation of an ion pocket. The ion was considered to be in the binding site when the distance was less than 4 Å from two of the three atoms. This ion has then been observed from t=600 ps to the end of the simulation (see Fig. 2). This is the only Na^+ specific site of the loop substructure that does not involve phosphodiester linker atoms.

The strong interactions of guanine O_6 atoms could explain the lower activity of the isosteric G15–17A mutant observed by Lukavsky et al., since adenines lack an H-bond acceptor atom in position 6. Moreover, the different dynamic behavior of the three guanines suggests that they do not play the same role in the specific IRES IIId loop recognition process. Contrary to G15 and G16 bases, which appear to be clearly dedicated to an external recognition process (since they do not interact through intramolecular interactions), G17, due to its O_6 oxygen atom plays a key role in the structural feature of the loop, through ion-mediated intramolecular interactions with U13 and C19.

IRES has been proven to have a specific folding depending on magnesium concentration and was shown to play a role in enzyme-cleavage patterns. [6] Thus for the purpose of comparison, a 1-ns molecular dynamics simulation was performed taking into account Mg^{2+}

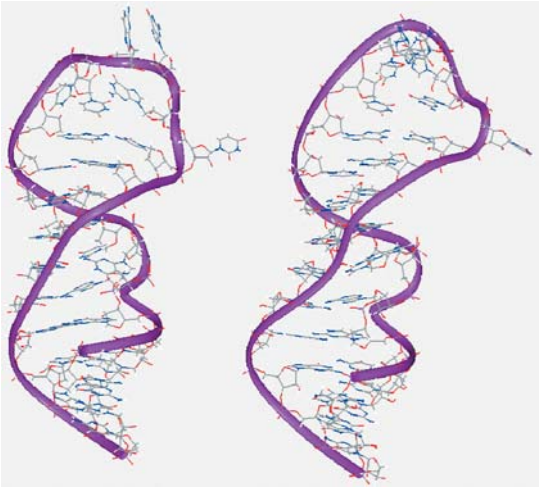


Fig. 7 Comparison between averaged structure of IIId IRES neutralized by Na^+ (2.6 ns averaged, *right*) or Mg^{2+} (1 ns averaged, *left*) cations

counterions, instead of Na^+. The starting structure is a little more conserved than with Na^+ ions, due to a larger electrostatic interaction with the doubly charged Mg^{2+} ions. The RMSd value never reached more than 3 Å. The position of the Mg^{2+} ions was in fact very similar to the Na^+ ones, but resulted in a slightly more compact structure, as shown in Fig. 7, especially in the vicinity of the S-turn motifs. Indeed, it has been shown that Mg^{2+} ions interact strongly with loop-E structures and suggested that their presence could be crucial for the conservation of this motif. [25]

The loop structure was more constrained and we have not observed any exchange of interactions involving G17, which remained in interaction with the phosphate linker of G15 throughout the simulation. Also, although G15 and G16 remained stacked with each other throughout the simulation, they have shown less mobility and a larger accessibility above the minor groove. The results clearly suggest a specific role played by the cations around the RNA structure and emphasize the difficulty in analyzing their influence on the structure of the helix rationally. Also, the ion pocket at G17 found in the Na^+ simulation, is partly reproduced by the Mg^{2+} simulation. A hexahydrated Mg^{2+} ion has been found interacting with G17 O_6 and N_7 atoms as well as with U14 O_1P and G15 O_6 atoms. This interaction is weaker than with the one found for the Na^+ ions (from t=0 to 600 ps). Then the ion leaves the interaction site to a distance of more than 8 Å away in the bulk. A second ion pocket, similar to that found with Na^+ ions, located in the helical part of the structure, more precisely in the S-turn of the loop-E motif (near linkers of nt 6, 7, 8), is perfectly determined. U8 is also involved in a second ion binding, together with the phosphate linker of G20. This interaction also contributes to a tighter folding of the IIId structure. Yet, the main results concerning the accessibility of special structural motifs (G15/G16 stacked guanines, G17 free N_7 atom, bulged out U18, SRL motif and A23–A24–A25 free WC face) remain qualitatively the same and confirm that IRES IIId domain bears several strongly negative electrostatic sites, which should be crucial for external recognition process.

The specific role of the ions is difficult to analyze here, for the reasons mentioned above. More precisely, the Mg^{2+} simulation reveals a more rigid structure, faithful to the starting experimental one. We believe that this rigidity is caused by the very large electrostatic interactions of the ions with the structure, preventing the RNA from any structural reorganization. The structural sensitivity of the RNA to the surrounding counterions could only be rationalized on the basis of longer simulations, which are beyond the scope of this study. The spontaneous ion binding of IRES IIId could only be analyzed after a simulation bearing a mixture of Na^+ and Mg^{2+} ions. This study only focuses on the potential binding sites and has shown that they behave similarly whatever the ions and that the structure remains close to the experimental one, validating our simulation procedure. The only difference is for the G17 base, which appears to be more flexible in the Na^+ simulation. The structural reorganization of G17 during the Na^+ simulation is supported by the fact that the 25 experimental NMR structures (PDB id: 1F84) show very different G17 locations and suggests a structural flexibility at this location, confirmed by the simulation.

Conclusions

Water phase dynamic behavior of HCV IRES IIId RNA domain, neutralized by Na^+ or Mg^{2+} counterions has been studied by 2.6 ns of molecular dynamics simulation. The root mean square deviations of the three sub-domains of IRES IIId reveal a large structural flexibility of the hexanucleotide loop domain, while the rest of the structure (loop-E motif and regular helix part) remains quite rigid. Concerning the loop, two of the three guanines are stacked together and located outside the loop above the minor groove. The third guanine base undergoes relatively large structural variation, with a Watson–Crick face located in the major groove. This guanine base, together with uracyl U13 is at the origin of an ion pocket, which contributes to keeping the structural integrity of the loop structure. H-bond analysis reveals that the two other guanines do not interact though their Watson–Crick faces that are completely accessible for molecular recognition. The different behavior of G17 with respect to G15 and G16 suggests a radically different role in the recognition process. Major and minor groove accessibility has been analyzed and revealed that the minor groove, although accessible, presents a very different depth along the helical region. Moreover, the major groove of the structure is separated by the crossing of the SRL motif in the middle of the helical region. Electrostatic analysis has shown that several sites produce a large negative electrostatic potential where counterions are found. Localization of such electronegative chemical groups and the analysis of the dynamic behavior of the

structure could be useful for the rational design of selective and efficient inhibitors, that could target those potential binding sites. Such inhibitors should for example possess electrostatic complementarity with at least G15 and G16 bases, as well as with the N_7 atom of G17, and/or complementarity with the Loop-E motif phosphodiester linkers. Such inhibitors have actually been synthesized and activity with IRES should be reported soon.

Acknowledgements The CINES computing center is acknowledged for providing us with computing time. JG thanks Adrian Wiley for his help in the preparation of the manuscript and Juan Fernandez Carmona for the energy calculations. The referees are acknowledged for interesting suggestions.

References

1. Tsukiyama-Kohara K, Iizuka N, Kohara M, Nomoto A (1992) J Virol 66:1476–1483
2. Fukushi S, Katayama K, Kurihara C, Ishiyama N, Hoshino FB, Ando T, Oya A (1994) Biochem Biophys Res Commun 199:425–432
3. Honda M, Brown EA, Lemon SM (1996) RNA 2:955–968
4. Jubin R, Vantuno NE, Kieft JS, Murray MG, Doudna JA, Lau JYN, Baroudy BM (2000) J Virol 74:10430–10437
5. Kolupaeva VG, Pestova TV, Hellen CU (2000) J Virol 74:6442–6450
6. Kieft JS, Zhou K, Jubin R, Murray MG, Lau JYN, Doudna JA (1999) J Mol Biol 292:513–529
7. Lukavsky PJ, Otto GA, Lancaster M, Sarnow P, Puglisi JD (2000) Nature Struct Bio 7:1105–1110
8. Klinck R, Westhof E, Walker S, Afshar M, Collier A, Aboul-Ela F (2000) RNA 6:1423–1431
9. Case DA, Pearlman DA, Caldwell JW, Cheatham TEI, 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 (2000) AMBER 6. University of California, San Francisco, Calif.
10. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
11. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniles 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, Stefanov BB, Liu G, 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, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian, Pittsburg, Pa.
12. Kollman PA, Dixon R, Cornell W, Fox T, Chipot C, Pohorille A (1997) In: Wilkinson PWA, van Gunsteren W (ed) Computer simulation of biomolecular systems: theoretical and experimental applications, vol 3. Escom, The Netherlands, p 83
13. Hobza P, Kabelàc M, Šponer J, Mejzlik P, Vondrasek J (1997) J Comput Chem 97:1136–1150
14. Hobza P, Šponer J (1999) Chem Rev 99:3247–3276
15. Auffinger P, Westhof E (2000) J Mol Biol 300:1115–1133
16. Auffinger P, Westhof E (2001) Biopolymers 56:266–274
17. Auffinger P, Westhof E (2001) J Mol Biol 305:1057–1072
18. Auffinger P, Westhof E (1997) J Mol Biol 269:326–341
19. Schneider C, Brandl M, Sühnel J (2001) J Mol Biol 305:659–667
20. Yang X, Gerczei T, Glover L, Correll CC (2001) Nat Struct Biol 8:968–973
21. Leontis NB, Westhof E (1998) J Mol Biol 283:571–583
22. Weeks KM, Crothers DM (1993) Science 261:1574–1577
23. Gao YG, Robinson H, van Boom JH, Wang AJH (1995) Biophys J 69:559–568
24. Young MA, Jayaram B, Beveridge DL (1997) J Am Chem Soc 119:59–69
25. Serra MJ, Baird JD, Dale T, Fey BL, Retatagos K, Westhof E (2002) RNA 8:307–323