



**HAL**  
open science

## Compaction of RNA Hairpins and Their Kissing Complexes in Native Electrospray Mass Spectrometry

Josephine Abi-Ghanem, Clémence Rabin, Massimiliano Porrini, Frédéric Rosu,  
Valérie Gabelica

### ► To cite this version:

Josephine Abi-Ghanem, Clémence Rabin, Massimiliano Porrini, Frédéric Rosu, Valérie Gabelica. Compaction of RNA Hairpins and Their Kissing Complexes in Native Electrospray Mass Spectrometry. *Journal of The American Society for Mass Spectrometry*, 2020, 31 (10), pp.2035-2043. <10.1021/jasms.0c00060>. <hal-03321844>

**HAL Id: hal-03321844**

**<https://hal.science/hal-03321844v1>**

Submitted on 18 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## **Compaction of RNA Hairpins and their Kissing Complexes in Native Electrospray Mass Spectrometry**

Josephine Abi-Ghanem<sup>1</sup>, Clémence Rabin<sup>1</sup>, Massimiliano Porrini<sup>1</sup>,  
Frederic Rosu<sup>2</sup>, Valérie Gabelica<sup>1\*</sup>

<sup>1</sup>Université de Bordeaux, Inserm & CNRS, Laboratoire Acides Nucléiques: Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320), IECB, 2 rue Robert Escarpit, 33607 Pessac, France.

<sup>2</sup> Université de Bordeaux, CNRS & Inserm, Institut Européen de Chimie et Biologie (IECB, UMS3033, US001), 2 rue Robert Escarpit, 33607 Pessac, France.

\* Corresponding author: [v.gabelica@iecb.u-bordeaux.fr](mailto:v.gabelica@iecb.u-bordeaux.fr)

### **Abstract:**

When electrosprayed from native solution conditions, RNA hairpins and kissing complexes acquire charge states at which they get significantly more compact in the gas phase than their initial structure in solution. Here we show the limits of using force field molecular dynamics to interpret the gas-phase structures of nucleic acid complexes in the gas phase, and we suggest that higher-level calculation levels should be used in the future.

## **Introduction:**

Ion mobility spectrometry (IMS) separates ions according to charge and shape, based on their electrophoretic mobility in a drift tube filled with a buffer gas such as helium.<sup>1-2</sup> The measured physical quantity is the collision cross section (<sup>exp</sup>CCS), and the interpretation of this structural information requires theoretical calculations (<sup>calc</sup>CCS) on ion structures *in vacuo*. Upon native electrospray ionization, intact molecules as well as noncovalent complexes can be preserved. However during the transfer in gas phase, structural rearrangements may occur.<sup>3-8</sup> Experimental and theoretical approaches initially suggested that Watson-Crick hydrogen bonds and stacking are preserved,<sup>9-10</sup> but we recently showed that DNA and RNA duplexes in the gas phase with low charge states undergo a huge compaction.<sup>11</sup> This raises the concern that other higher-order DNA or RNA structures containing double helix motifs may also end up more compact in the gas phase than they were in solution. DNA hairpins, pseudoknots and cruciforms were studied by ion mobility before,<sup>12</sup> but at high charge-to-mass ratios produced from 5 mM aqueous NH<sub>4</sub>OAc. Also the molecular dynamics trajectories at the time were 2 ns. Here we tested RNA structures sprayed from aqueous solutions at physiological ionic strength (150 mM NH<sub>4</sub>OAc), and performed much more extensive force field molecular dynamics calculations.

In particular, we studied RNA hairpins and RNA-RNA kissing complexes structures (Figure 1) in the gas phase. Hairpins (also called “stem-loops”) are formed by a duplex part (the stem) and a loop of unpaired nucleotides. Their flexibility depends on their composition and on the size of the stem and loop. The hairpin is a frequent motif in natural RNA folding and interactions.<sup>13</sup> For example, the tRNA structure involves three hairpins.<sup>14</sup> Also, hairpins with tetranucleotide loops are naturally present in a variety of RNAs such as rRNAs, group I introns or ribozymes.<sup>15-16</sup> RNA-RNA kissing complexes are formed when the nucleotides in the loops of two RNA hairpins are complementary. The smallest loop-loop interaction is formed by only two nucleotides in each loop,<sup>17</sup> but usually involve more nucleotides. Here we studied the RNAI:RNAII kissing complex involved in the replication of the ColE1 plasmid in *Escherichia coli* (actually their inverse sequences RNAIi and RNAIi, which form a more stable complex, often used for model biophysical studies),<sup>18-19</sup> and the binding of TAR RNA sequence from the mRNA of the HIV-1 virus with the R06 aptamer (itself a hairpin), designed to bind TAR and prevent the binding to the TAT protein.<sup>20</sup> An NMR structure of RNAIi:RNAIi (PDB: 2BL2)<sup>21</sup> and an X-ray crystal structure

of TAR:R06 (PDB: 1JLT)<sup>22</sup> have been published, and will be used as starting points to examining the differences between gas-phase and solution structures.



**Figure 1:** Sequences and secondary/tertiary structures of the TAR:R06 and RNAIi:RNAIli kissing complexes.

## **Experimental:**

### **Oligonucleotides**

Lyophilized oligonucleotides were purchased from Integrated DNA Technology (Leuven, Belgium) with standard desalting. They were dissolved in RNase-free water (Ambion, Fisher Scientific, Illkirch, France), then a stock solution at 100  $\mu$ M of each hairpin was prepared in 150 mM NH<sub>4</sub>OAc (itself prepared from BioUltra  $\approx$  5 M stock solution, Fluka, Sigma-Aldrich, Saint-Quentin Fallavier, France). This stock was desalted using Amicon Ultra-0.5 3K ultrafiltration devices (Merck Millipore, Saint-Quentin en Yvelines, France). The individual hairpins were then heated at 90°C for 90 s and then quickly cooled on ice to obtain the hairpin structure. The injected solutions were 10  $\mu$ M of each hairpin in 150 mM NH<sub>4</sub>OAc. 2  $\mu$ M of dT<sub>6</sub> was added as a reference for quantitation. The kissing complexes were prepared by mixing equimolar amounts of hairpins.

### **Native Electrospray Ionization Mass Spectrometry (ESI-MS)**

Experiments were performed on an Agilent 6560 DTIMS-Q-TOF instrument (Agilent Technologies, Santa Clara, CA), with the dual-ESI source operated in negative ion mode. The syringe pump flow rate was 180  $\mu$ L/h. The drift tube was filled with helium (temperature = 296  $\pm$  1 K). The pressure in the drift tube was 3.89  $\pm$  0.01 Torr, and the pressure in the trapping funnel is 3.67  $\pm$  0.01 Torr (this ensures that only helium is present in the drift tube). In our default conditions,

the source temperature and fragmentor voltage were set at 200°C and 350 V, and the Trap Entrance Grid Delta fixed at 12 V (Trap entrance Grid low 105 V, Trap Entrance Grid High 117 V). These parameters are not the softest possible on our instrument,<sup>23</sup> but such degree of in-source activation was necessary to ensure a good declustering (i.e. to get rid of most non-specific ammonium adducts) and transmission of ions, and did not cause any apparent dissociation of the kissing complexes. To produce more activating conditions, the fragmentor voltage was set to 600 V. Alternatively, to produce softer conditions that preserve some ammonium adducts, the Trap Entrance Grid Delta was set to 4 V.

### **MD in solution and representative structures**

The MDs in solution were performed using as starting point the PDB coordinates 2JLT<sup>22</sup> for TAR/R06 and 1BJ2 for RNAi/RNAii.<sup>21</sup> The structures were minimized and equilibrated using the module pmemd.MPI of Amber 15<sup>24</sup> suite of packages using the parmbsc0 force field with  $\chi$ OL3 correction.<sup>25-27</sup> The complexes were neutralized with Na<sup>+</sup> ions and hydrated with TIP3P water molecules in a truncated octahedron.<sup>28-29</sup> Periodic boundary conditions at constant temperature (300 K) and pressure (1 bar) using the Berendsen algorithm were applied.<sup>30</sup> Covalent bonds involving hydrogens were restrained using SHAKE,<sup>31</sup> with a 2-fs integration time-step. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME)<sup>24</sup> algorithm with a radial cutoff of 9 Å ; the same cut-off was used for the van der Waals interactions. The non-bonded pair-list was updated heuristically and the center-of-mass motion removed every 10 ps. Water molecules and counter-ions were energy-minimized (steepest descent followed by conjugate gradient) and equilibrated at 100 K for 100 ps at constant volume and temperature while RNA molecules were positionally restrained. The whole system was then heated from 100 K to 300 K over 10 ps by 5-K increments with harmonic positional restraints on solute atoms (5.0 kcal/mol/Å<sup>2</sup> force constant). The simulation was continued at constant pressure and temperature. The positional restraints were gradually removed over 250 ps.

MD production runs of 300 ns to 1  $\mu$ s were started from the end point of the equilibration. For the 16-mer hairpin TAR and the 18-mer hairpin R06, a multiple molecular dynamic (MMD) approach was used. From the end point of the equilibration, 25 trajectories of 10 ns each (total 250 ns) were launched with new velocities for each trajectory. The cpptraj<sup>32</sup> module of AmberTools15 was used

to calculate the radius of gyration and the representative structure. The statistical analysis was carried out using the R package.<sup>33</sup> Image rendering was performed with tachyon in VMD.<sup>34</sup>

### **Localized charges**

After the representative structure in solution was chosen, we need to decide which phosphate groups will be neutralized by protons. We applied an in-house python script to generate 1000 structures with random protonation schemes. Using Amber module sander we calculated a single point energy for each of these structures (no geometry optimization in the gas phase at this stage). The 200 structures showing the lowest energy were chosen and their protonation scheme was displayed as histograms, to deduce the most probable locations (further information below in the results). We then placed the protons on the representative structure according to those locations.

### **MD in gas phase**

The whole system was heated from 100 K to 300 K during 100 ps, and 50 independent attempts were compared (further discussion in the results section). Then gas phase MD simulations were run with the sander module of Amber15, in canonical ensemble (at T = 298 K) using a Langevin thermostat, implementing a direct Coulomb summation for evaluating the electrostatic interactions, using an integration time-step of 1 fs and no radial cut-off for the non-bonded forces. Because the screening effect of the solvent is missing *in vacuo*, we set no radial cut-off for both Coulomb and van der Waals interactions and the PME algorithm cannot be applied. In order to calculate the CCS from the gas phase MD simulations, snapshots were collected every 50 ps. Theoretical CCS values were calculated with EHSSrot algorithm<sup>35</sup> with the parameterization of Siu et al.,<sup>36</sup> which accuracy was validated previously on rigid nucleic acid structures (G-quadruplex, G-duplex) of similar size.<sup>37-38</sup>

### **DFT calculations**

DFT calculations were run on TAR<sup>4-</sup> were performed using Gaussian09<sup>39</sup> (M06-2X functional,<sup>40</sup> Grimme D3 dispersion correction<sup>41</sup> and 6-31G(d,p) basis-set). A geometry optimization was first carried out on the a truncated hairpin TAR14<sup>3-</sup> lacking the terminal base pair, because the optimization could not start converging when starting from the full 16-mer hairpin. Then, the terminal base pair was added, one more phosphate group was deprotonated, and the entire 16-mer TAR<sup>4-</sup> could be geometry optimized. The electrostatic potentials were calculated from the total electronic density.

## **Results and Discussion**

### **RNA hairpins and kissing complexes are significantly more compact in the gas phase than in solution**

The experimental collision cross sections in helium are determined from the arrival time distributions. Repeated measurements ( $n = 3$  minimum) did all fall within less than 1%. The standard relative uncertainty on the collision cross section of the peak center, estimated by propagation of uncertainty on temperature, pressure, tube length and slope of the  $t_a=f(1/\Delta V)$  linear regression is estimated at  $\sim 1.2\%$ . However, the peaks are broader than the instrumental function, suggesting that the gas-phase ensemble contains multiple conformations which span a range of CCS values. The full width at half maximum of the CCS distributions was calculated as reported previously.<sup>42</sup> The values are reported in Table 1 and representative experimental distributions are shown in Figure 2.

To calculate the CCS values corresponding to the structures in solution, we used the coordinates from the published structures of the kissing complexes. The hairpins were extracted from their respective kissing complexes. The calculated collision cross sections in helium of these starting structures are given in Table 1. The CCS values are all  $\sim 20$  to 30% smaller in the gas phase experiments than in the starting solution structures. The entire experimental CCS distribution is below the expected one in solution.

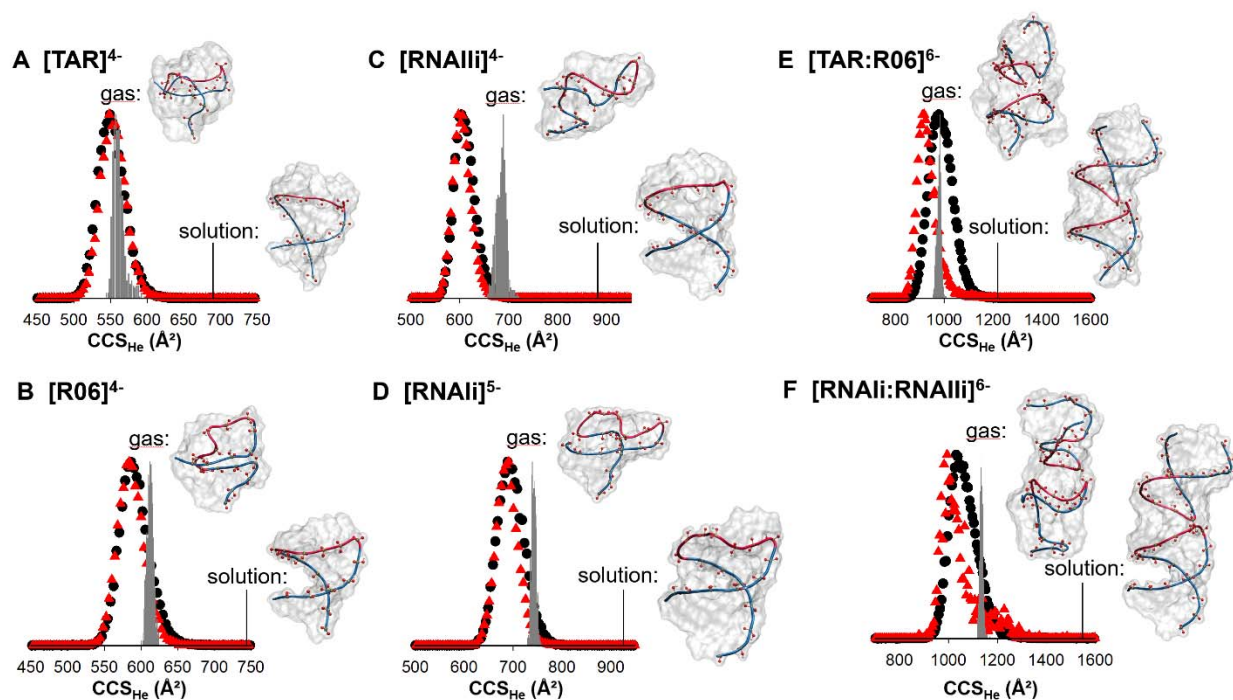
The compaction can occur at various stages of electrospray: i.e. when solvent molecules are still around, or when counter-ions are still around, or after full desolvation and declustering if our collisional activation conditions are not soft enough. We thus tested different pre-mobility activation conditions. CCS distributions in our default conditions (fragmentor voltage at 350 V) are shown in black and those in harsh conditions (fragmentor voltage at 600 V) are shown in red. For the hairpins, the differences are minor, with just a slight decrease in the peak width at the larger CCS side. For the kissing complexes, the compaction upon activation is more substantial. It means that softer conditions are essential to preserve the solution structure, and that maybe our initial conditions are already too energetic. We also tested softer conditions than the default ones, which led to the preservation of numerous ammonium ion adducts (supporting information Figure S1). The arrival time distribution for the adducted species is however not much different from that of

the naked hairpin or complex. Thus, even in soft conditions leading to mass spectra of a lesser quality due to numerous non-specific ammonium adducts, the gas-phase structures are already more compact than the solution ones. This suggests that the rearrangements occurred at even earlier stages, i.e. either during the electrospray process or upon desolvation.

The same compaction happens to RNA double helices. Supporting Information Figure S2 shows the CCS distributions of control duplexes having the same composition as the kissing complexes. The CCS distributions overlap. Hence the CCS distributions are also useless to distinguish double helix from kissing complex structures. Together with our previous study on DNA and RNA double helices,<sup>11</sup> our results show that, at the charge states naturally produced from aqueous NH<sub>4</sub>OAc mimicking physiological ionic strength, double stranded regions in nucleic acids undergo rearrangements that cannot be captured in source conditions leading to decent mass spectra.

**Table 1:** Theoretical and calculated collision cross sections in helium.

	<sup>EHSS</sup> CCS <sub>He</sub> (Å <sup>2</sup> ) (solution starting structure)	<sup>DT</sup> CCS <sub>He</sub> (Å <sup>2</sup> ) of peak center (soft pre-IM conditions)	FWHM of experimental CCSD (Å <sup>2</sup> ) and relative FWHM (%)	% CCS change in gas phase (soft conditions) compared to in solution
[TAR] <sup>4-</sup>	690	552	44 (8%)	-20 %
[R06] <sup>4-</sup>	744	590	44 (7.5%)	-20 %
[TAR/R06] <sup>6-</sup>	1219	986	154 (16%)	-19 %
[RNAi] <sup>5-</sup>	926	695	64 (9%)	-25 %
[RNAii] <sup>4-</sup>	882	607.5	46 (7.5%)	-31 %
[RNAi/RNAii] <sup>7-</sup>	1548	1055	120 (11%)	-32 %



**Figure 2:** Experimental gas-phase collision cross section distributions for the hairpins (A–D) and kissing complexes (E, F). The effect of pre-ion mobility activation is seen by comparing black distributions (fragmentor at 350 V) and red curves (fragmentor at 600 V).

### Can molecular dynamics (MD) with molecular mechanics force field account for the gas-phase compaction?

The parmbsc0 force field with  $\chi$ OL3 correction is appropriate to model RNA structures in solution. Can it account for the rearrangements that occur in the gas phase? The general approach is to (1) generate solution structures (either by taking published coordinates or by running short solution dynamics starting from them), then (2) strip the structures of their solvent and reduce the number of charges to match the ones observed by ESI-MS, then (3) prepare the system for gas-phase dynamics and finally (4) run them long enough or at high temperature enough for the experimentally rearrangement to occur. At each step, choices have to be made. We will show here that only making choices that maximize compaction at each step that the force field MD simulations can render a compaction that starts to match, although not always perfectly, with the experimentally observed one.

In the first step, we explored whether hairpins and kissing complexes might adopt a wider variety of conformations in solution, by running solution force field MD simulations starting from the

published coordinates. To sample a variety of conformations we ran multiple molecular dynamic (MMD) starting from the experimental structure and running 25 trajectories for 10 ns each. The kissing complexes are stable systems in solution and MD simulation were done for 300 ns. A first clustering was based on the radius of gyration, as its calculation is easier to implement on-the-fly, and the correlation of  $r_g$  with CCS, although not perfect,<sup>43</sup> suffices to capture starting points which are the most compacted conformation (low  $r_g$ ), and compared the results with less compact starting points (high  $r_g$ ). Although the visual differences are not huge, as we will show below this choice influences the final results.

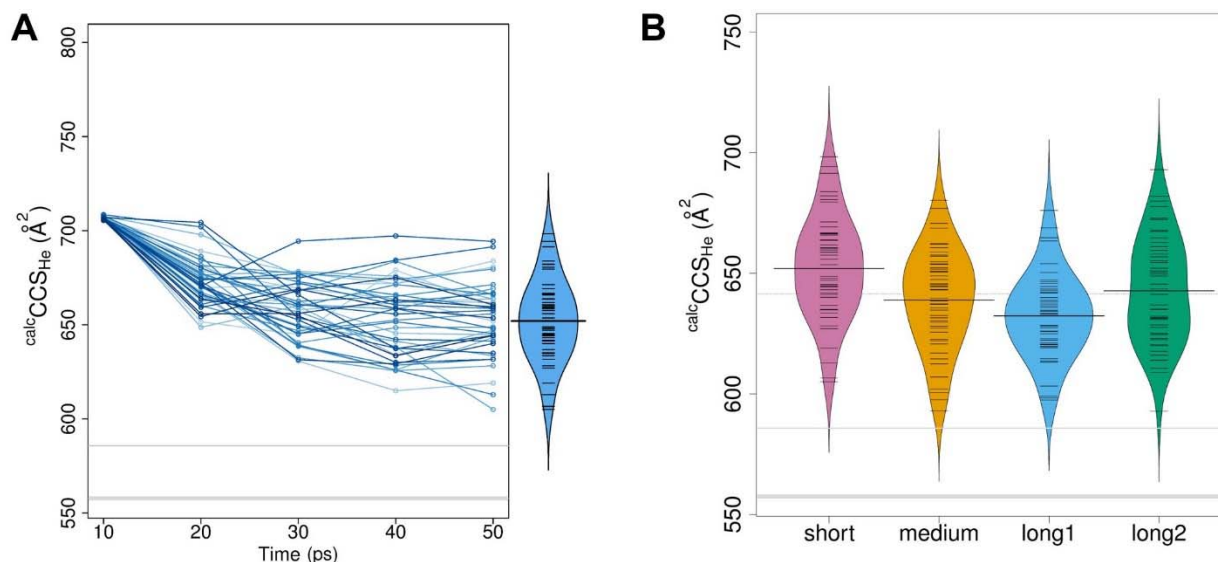
Next, we tested different protonation schemes. Once the representative structure is selected, the choice is whether to localize the charge (LC) or use a distributed charge model (DC).<sup>44</sup> With the DC model (taking TAR as an example), the net charge of all the 15 phosphate groups is reduced and the total charge of the hairpin will be of -4 and distributed equally along the phosphate group of the RNA. With the LC model, assuming that only the backbone phosphate groups can be protonated, one needs to add 11 protons out of the 15 phosphate groups with 1365 different ways of positioning them, for a given conformation. Here we randomly protonate the hairpin 1000 times. A single point energy calculation with a classical force field was used to classify the resulted structures (Supplementary Figure S3). Among the 200 structures with the lowest energy, we observed the sites that are the most often deprotonated, to define the most probable (MP) protonation. The least probable (LP) localized charge distribution was obtained by filtering the 200 structures with the highest energy.

The next step is the heating of the molecular system up to 300 K to prepare it for the MD in the gas phase. Note that this section describes the heating process that prepares the system for the actual (300 ns to 1  $\mu$ s) MD runs at 300 K. We repeated the heating process and observed different outcomes, leading to a broad distribution of CCS values (Figure 3). The gas-phase CCS always decreases compared to the solution, already during the heating step. Thus, depending on the heating trajectory, starting points for the gas-phase MD could be closer to or further from the experimental CCS value. We also tested different heating protocols (Table 2). The resulting distributions for TAR<sup>4+</sup> (starting from the most compact solution conformation, localized charges, most probable proton distribution) are shown in Figure 3B. The compaction observed from 0 to 50 ps (Figure 3A) continues slightly to longer times. However a slower heating did not lead to

particularly more compact conformations (compare long2 and long1). For the remainder of the paper, we settled on the “medium” heating conditions defined in Table 2.

**Table 2:** Parameters of the four heating processes tested herein. The change of temperature is applied over steps/iterations Istep1 through Istep2.

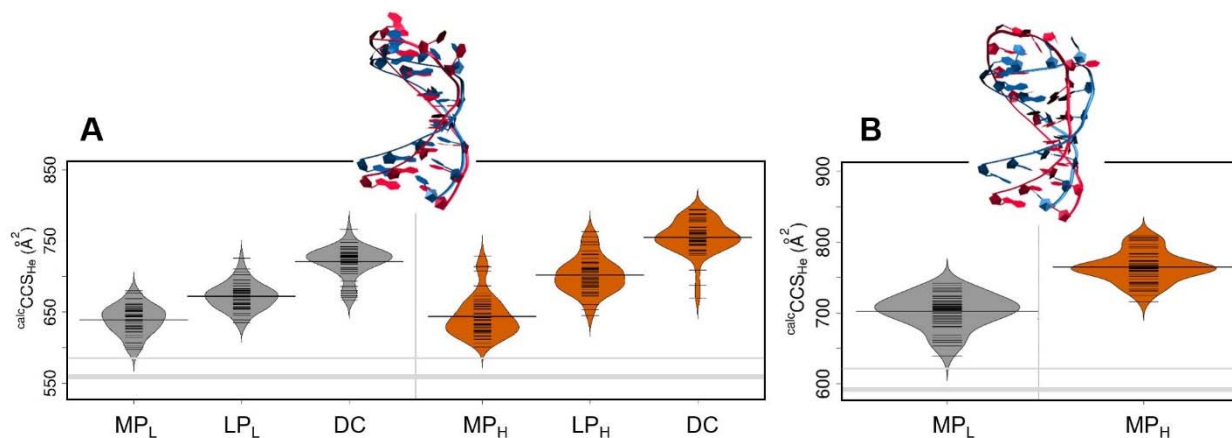
	From 0 to 300 K		Total length (ps)
	Istep1 (ps)	Istep2 (ps)	
Short	0	20	50
Medium	0	20	100
Long1	0	20	200
Long2	0	100	200



**Figure 3:** Effect of the of the heating process on the CCS distribution. A) Evolution of the CCS along the 50 ps heating process (50 replicates) and resulting CCS distribution represented as a beanplot, where each small black line is one heating trajectory, and the median is shown. B) CCS distribution following different heating processes (defined in Table 2). The bold grey line indicates the center of the experimental distribution, and a thin grey line its half-maximum on the high-CCS side.

We now compare the distributions of CCS values at the end of the heating process, for the different protonation schemes and with different starting solution structures. Figure 4A shows that, for

TAR<sup>4-</sup>, the MP charge distribution combined with the low starting radius of gyration gives the CCS distribution at the lowest values, i.e. closest to the experimental ones. As expected, the structure with the least probable protonation scheme does not compact as much. The one with distributed charges does not match well with experimental values either. This was initially surprising, because distributed charges were reported to lead to better base-base interaction energies and better DNA double helix preservation in the gas-phase models,<sup>45</sup> but we note that the structures modeled in that paper were all extended ones. Our results suggest that the localized charges scheme is more relevant for gas-phase RNA at the compact charge states produced in native MS conditions. For TAR<sup>4-</sup>, the distributions do not significantly depend on the initial structure chosen, but they do for R06<sup>4-</sup> (Figure 4B): depending on the starting structure, the conformations converge towards a different ensemble, and starting from a more compact structure converges towards a more compact gas-phase ensemble.



**Figure 4:** A) Distribution of CCS at the end of the heating process for different protonation schemes of TAR<sup>4-</sup> for a low (10.7 Å, grey) or high (11.7 Å, orange) radius of gyration starting point: MP = most probable localized charges, LP = least probable localized charges, DC = distributed charges. B) Distribution of CCS at the end of the heating process for R06<sup>4-</sup>, for the most probable localized charge distribution applied to a low (11.2 Å, grey) or high (12.9 Å, orange) gyration radius starting point. The structures with low radius of gyration are shown in blue and with high radius of gyration in pink. The bold grey line indicates the center of the experimental distribution, and a thin grey line its half-maximum on the high-CCS side.

Note that after the heating steps, all calculated CCS distributions are still much higher than the experimental ones. During the next step, i.e. long gas-phase MD trajectories, the conformations will have a chance to get even more compact by passing some rearrangement barriers (Figure S4). However, the final ensemble of conformations will depend on the combination of choices made for steps (1)—(3).

The CCS distributions shown in Figure 1 were obtained by running gas-phase force field MD on the most compact gas-phase conformer produced by heating of the most probable localized charge model, itself built from the most compact solution conformer. But even with all these human-imposed biases (because we knew the gas-phase CCS values before doing the modeling), when the hairpins become larger, the mismatch between calculated and experimental CCS increases. An interesting observation nevertheless is that for a compaction to occur in the models, the loop must close on the stem, and this outcome was influenced by the choice of the proton location.

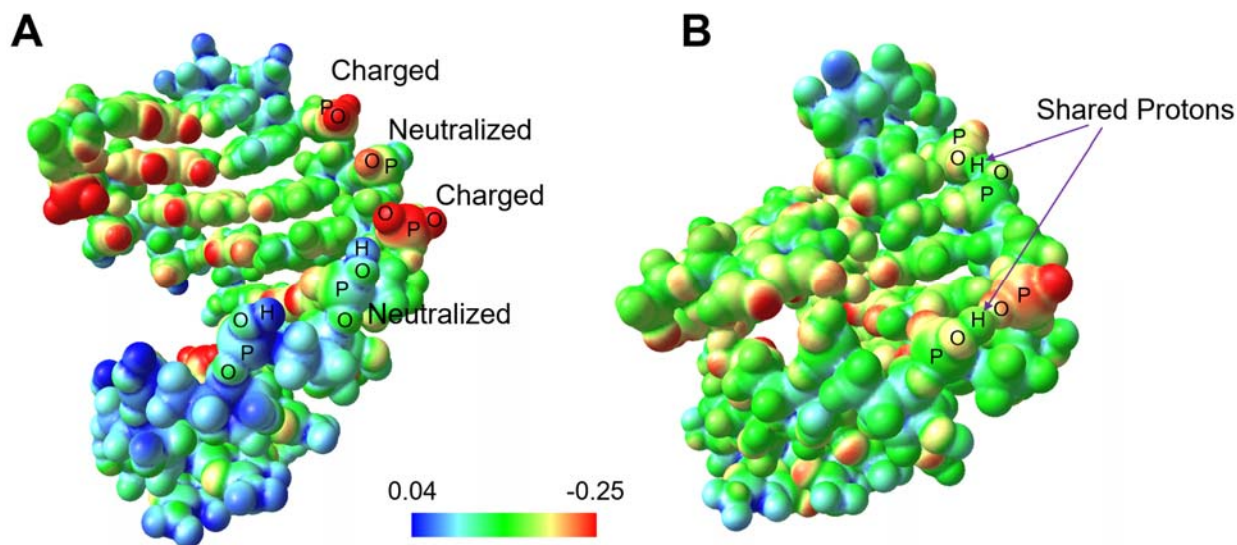
For the kissing complexes, the workflow described above for hairpins gave even larger discrepancies. The experimentally observed degree of compaction could not be reproduced by the models. In particular, the stems and loops could not come close to one another. We thus applied further restraints during the heating step, in a similar way as described previously for DNA double helices:<sup>11</sup> protonated and deprotonated phosphate groups from the stem and loop that faced one another were forced to come close to one another during the heating (Figure S5). With that procedure, the models had collision cross sections in the range of the experimentally observed ones. It does not mean that the structures shown in Figure 1 are the gas phase structures, but it means that the gas-phase structures are as compact as those shown.

## **Discussion: the limitations of force fields to describe electrosprayed nucleic acids**

The general conclusion of the force field molecular dynamics study is that this approach is not viable to interpret the experimental collision cross sections in terms of structure. We also learned that the proton location matters, because the compaction is due to the formation of new (non-native) hydrogen bonds between phosphate groups. However, the global compactness is generally underestimated by force fields in the gas phase, and this is problematic to interpret ion mobility spectrometry results.

To reveal the underlying reasons, we tested the gas-phase structural remodeling of a biologically relevant hairpin at a much higher level of theory: density functional theory (DFT) with the M06-2X functional (well adapted to nucleic acids),<sup>46</sup> and a sufficient basis set (6-31G(d,p)). The largest hairpin for which it could be done is TAR<sup>4+</sup>. Figure 5A shows the starting structure, with the electrostatic potential calculated from the total electron density. The charges in the initial structure are localized (deprotonated phosphate groups n° 1, 6, 12 and 14 are clearly in red). Figure 5B shows the optimized structure: it has compacted significantly, the base pairs are nevertheless well preserved, and new hydrogen bonding have now formed between neutralized P—O—H groups and charged P—O<sup>-</sup> groups. As a result, the charges are more distributed (no more clear red phosphate groups).

In summary, in the final structures, the charges are more delocalized owing to shared protons, but starting from an appropriate localized charge distribution was probably necessary for the compaction to happen. The collision cross section of the final (DFT-optimized) structure is 592 Å<sup>2</sup>, still on the high side of the experimental one, showing that further compaction occurs in practice. The final structure indeed has many void spaces, and proper conformational exploration would be required, but this is not attainable at this level of theory on our computers.



**Figure 5:** Starting structure (A) and optimized structure (B) of the truncated hairpin TAR<sup>4</sup>. The color coding represents the electrostatic potential calculated from the total electronic density (blue: positive charges; red: negative charges).

### **Conclusion:**

In conclusion, RNA structures such as hairpins and kissing complexes rearrange when passing from the solution to the gas phase, at least at the low charge states obtained from native MS. It does not necessarily mean that all aspects of the solution structure are lost, but the backbone rearrangements are responsible for a significant compaction. Regarding nucleic acids modeling in the gas phase for the interpretation of ion mobility spectrometry measurements, we suggest that, instead of force field molecular dynamics simulations, higher-level calculations would be required, ideally Born-Oppenheimer molecular dynamics, as shown recently on short rigid nucleic acid duplexes.<sup>38</sup> We also anticipate that different starting geometries and different starting proton locations will lead to different final structures. The variety of ways in which nucleic acid structures can rearrange upon or after electrospray probably account for the width of the collision cross section distribution. The current limiting factor for using higher-level calculations to interpret ion mobility spectrometry data is computer power, but it is clear that this will become increasingly feasible in the future. Meanwhile, we will explore in future works how structure optimizations and

molecular dynamics simulations with intermediate levels of accuracy, such as semi-empirical calculations, could perform to interpret collision cross section distributions.

**Acknowledgement:** This work was funded by the European Research Council (grant ERC-2013-CoG-616551-DNAFOLDIMS to VG).

### **Associated content:**

Additional experimental and computational results as described in text (PDF).

### **References:**

1. Gabelica, V.; Marklund, E., Fundamentals of Ion Mobility Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51-59.
2. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Ion mobility-mass spectrometry. *J. Mass Spectrom.* **2008**, *43* (1), 1-22.
3. Rolland, A. D.; Prell, J. S., Computational insights into compaction of gas-phase protein and protein complex ions in native ion mobility-mass spectrometry. *Trends Anal. Chem.* **2019**, *116*, 282-291.
4. Jhingree, J. R.; Bellina, B.; Pacholarz, K. J.; Barran, P. E., Charge Mediated Compaction and Rearrangement of Gas-Phase Proteins: A Case Study Considering Two Proteins at Opposing Ends of the Structure-Disorder Continuum. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (7), 1450-1461.
5. Hoaglund, C. S.; Counterman, A. E.; Clemmer, D. E., Anhydrous protein ions. *Chem. Rev.* **1999**, *99*, 3037-3079.
6. Clemmer, D. E.; Russell, D. H.; Williams, E. R., Characterizing the Conformationome: Toward a Structural Understanding of the Proteome. *Acc. Chem. Res.* **2017**, *50* (3), 556-560.
7. Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E., Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew. Chem. Int. Ed.* **2014**, *53* (30), 7765-9.
8. Allison, T.; Barran, P.; Cianferani, S.; Degiacomi, M.; Gabelica, V.; Grandori, R.; Marklund, E.; Menneteau, T.; Migas, L.; Politis, A.; Sharon, M.; Sobott, F.; Thalassinou, K.; Benesch, J. L., Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *ChemRxiv* **2020**, *preprint* <https://doi.org/10.26434/chemrxiv.11439558.v1>.
9. Gabelica, V.; De Pauw, E., Collision-Induced Dissociation of 16-mer DNA Duplexes with Various Sequences: Evidence for Conservation of the Double Helix Conformation in the Gas Phase. *Int. J. Mass Spectrom.* **2002**, *219* 151-159.
10. Gidden, J.; Baker, E. S.; Ferzoco, A.; Bowers, M. T., Structural motifs of DNA complexes in the gas phase. *International Journal of Mass Spectrometry* **2005**, *240* (3), 183-193.

11. Porrini, M.; Rosu, F.; Rabin, C.; Darre, L.; Gomez, H.; Orozco, M.; Gabelica, V., Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent. Sci.* **2017**, *3* (5), 454-461.
12. Baker, E. S.; Dupuis, N. F.; Bowers, M. T., DNA Hairpin, Pseudoknot, and Cruciform Stability in a Solvent-Free Environment. *J. Phys. Chem. B* **2009**, *113* (6), 1722-1727.
13. Nowakowski, J.; Tinoco, I., Jr., RNA Structure and Stability. *Sem. Virol.* **1997**, *8*, 153-165.
14. Ladner, J. E.; Jack, A.; Robertus, J. D.; Brown, R. S.; Rhodes, D.; Clark, B. F. C.; Klug, A., Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc. Natl. Acad. Sci. USA* **1975**, *72* (11), 4414-4418.
15. Woese, C. R.; Winker, S.; Gutell, R. R., Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops" *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8467-8471.
16. Wedekind, J. E.; McKay, D. B., CRYSTALLOGRAPHIC STRUCTURES OF THE HAMMERHEAD RIBOZYME: Relationship to Ribozyme Folding and Catalysis. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 475-502.
17. Stephenson, W.; Asare-Okai, P. N.; Chen, A. A.; Keller, S.; Santiago, R.; Tenenbaum, S. A.; Garcia, A. E.; Fabris, D.; Li, P. T., The essential role of stacking adenines in a two-base-pair RNA kissing complex. *J Am Chem Soc* **2013**, *135* (15), 5602-11.
18. Gregorian, R. S., Jr.; Crothers, D. M., Determinants of RNA Hairpin Loop–Loop Complex Stability. *J. Mol. Biol.* **1995**, *248*, 968-984.
19. Marino, J. P.; Gregorian, R. S., Jr.; Csankovszki, G.; Crothers, D. M., Bent Helix Formation Between RNA Hairpins with Complementary Loops. *Science* **1995**, *268* (5216), 1448-1454.
20. Ducongé, F.; Toulmé, J.-J., In vitro selection identifies key determinants for loop–loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1. *RNA* **1999**, *5*, 1605-1614.
21. Lee, A. J.; Crothers, D. M., The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* **1998**, *6*, 993-1005.
22. Lebars, I.; Legrand, P.; Aime, A.; Pinaud, N.; Fribourg, S.; Di Primo, C., Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res* **2008**, *36* (22), 7146-56.
23. Gabelica, V.; Livet, S.; Rosu, F., Optimizing Native Ion Mobility Q-TOF in Helium and Nitrogen for Very Fragile Noncovalent Structures. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (11), 2189-2198.
24. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman, AMBER 2015. *University of California, San Francisco.* **2015**.
25. Abelson, A. K.; Delgado-Vega, A. M.; Kozyrev, S. V.; Sanchez, E.; Velazquez-Cruz, R.; Eriksson, N.; Wojcik, J.; Linga Reddy, M. V.; Lima, G.; D'Alfonso, S.; Migliaresi, S.; Baca, V.; Orozco, L.; Witte, T.; Ortego-Centeno, N.; group, A.; Abderrahim, H.; Pons-Estel, B. A.; Gutierrez, C.; Suarez, A.; Gonzalez-Escribano, M. F.; Martin, J.; Alarcon-Riquelme, M. E., STAT4 associates with systemic lupus erythematosus through two independent effects that

- correlate with gene expression and act additively with IRF5 to increase risk. *Ann Rheum Dis* **2009**, *68* (11), 1746-53.
26. Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M., Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *Journal of Chemical Theory and Computation* **2010**, *6* (12), 3836-3849.
  27. Zgarbova, M.; Otyepka, M.; Šponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* **2011**, *7* (9), 2886-2902.
  28. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
  29. Mahoney, M.; Jorgensen, W., A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **2000**, *112* (20), 8910-8922.
  30. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984**, *81* (8), 3684-3690.
  31. van Gunsteren, W. F.; Berendsen, H. J. C., Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics* **1977**, *34* (5), 1311-1327.
  32. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9* (7), 3084-3095.
  33. R core team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2015.
  34. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14* (1), 33-8, 27-8.
  35. Shvartsburg, A. A.; Mashkevich, S. V.; Baker, E. S.; Smith, R. D., Optimization of algorithms for ion mobility calculations. *J. Phys. Chem. A* **2007**, *111* (10), 2002-2010.
  36. Siu, C. K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W., Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses. *J. Phys. Chem. B* **2010**, *114* (2), 1204-1212.
  37. D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V., Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J Mass Spectrom* **2015**, *50* (5), 711-26.
  38. Swasey, S. M.; Rosu, F.; Copp, S. M.; Gabelica, V.; Gwinn, E. G., Parallel Guanine Duplex and Cytosine Duplex DNA with Uninterrupted Spines of Ag(I)-Mediated Base Pairs. *J. Phys. Chem. Lett.* **2018**, *9* (22), 6605-6610.
  39. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.;

- Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
40. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120* (1-3), 215-241.
41. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
42. Marchand, A.; Livet, S.; Rosu, F.; Gabelica, V., Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **2017**, *89*, 12674-12681.
43. Kulesza, A.; Marklund, E. G.; MacAleese, L.; Chiro, F.; Dugourd, P., Bringing Molecular Dynamics and Ion-Mobility Spectrometry Closer Together: Shape Correlations, Structure-Based Predictors, and Dissociation. *J. Phys. Chem. B* **2018**, *122* (35), 8317-8329.
44. Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M., The structure and dynamics of DNA in the gas phase. *J Am Chem Soc* **2003**, *125* (26), 8007-14.
45. Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M., The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* **2003**, *125*, 8007-8014.
46. Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D., Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theory Comput.* **2008**, *4*, 1996-2000.