

# Development of Site-Specific Mg<sup>2+</sup>–RNA Force Field Parameters: A Dream or Reality? Guidelines from Combined Molecular Dynamics and Quantum Mechanics Simulations

Lorenzo Casalino,<sup>†</sup> Giulia Palermo,<sup>‡</sup> Nodira Abdurakhmonova,<sup>†,§</sup> Ursula Rothlisberger,<sup>‡</sup> and Alessandra Magistrato<sup>\*,||</sup>

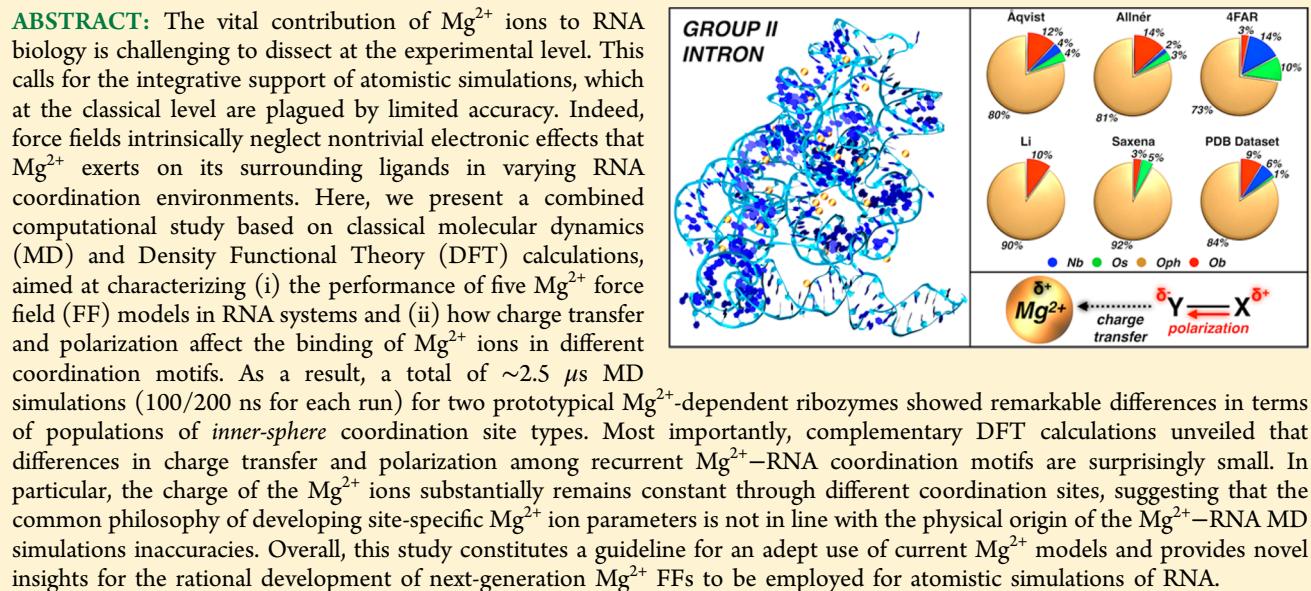
<sup>†</sup>International School for Advanced Studies (SISSA), Trieste, Italy

<sup>‡</sup>Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

<sup>§</sup>Università degli Studi di Trieste, Trieste, Italy

<sup>||</sup>CNR-IOM-Democritos National Simulation Center c/o SISSA, via Bonomea 265, Trieste, Italy

## S Supporting Information



## 1. INTRODUCTION

Mg<sup>2+</sup> is the most abundant alkali-earth metal in the biosphere with a high concentration in living cells.<sup>1</sup> Besides being operative in ATPases, ATP synthases and Mg<sup>2+</sup>-dependent enzymes processing nucleic acids,<sup>2–4</sup> this ion is ubiquitously present in RNA, playing a key function in both tuning RNA folding and catalysis. Due to their high charge density, Mg<sup>2+</sup> ions afford the unique capability of effectively screening the negative charge of the RNA phosphate backbone, contributing to shape its folding landscape and conferring stability to RNA tertiary structures.<sup>5</sup> Their crucial role is emphasized by the fact that RNA filaments can reach their native folded conformation only in the presence of Mg<sup>2+</sup> ions.<sup>6</sup> Moreover, Mg<sup>2+</sup> ions can enhance catalysis in ribozymes by stabilizing active sites and by properly orienting reactants and/or polarizing reactive groups.<sup>7–9</sup>

Mg<sup>2+</sup> ions accomplish this wide range of functions by binding to RNA in different manners. While diffuse Mg<sup>2+</sup> ions interact with RNA via long-range electrostatic interactions mediated by several layers of water molecules, specific Mg<sup>2+</sup>–RNA binding sites can form either when Mg<sup>2+</sup> ions come in direct contact with RNA atoms (hereafter referred to as *inner-sphere* coordination sites) or when Mg<sup>2+</sup>–RNA interactions are mediated by one shell of water molecules (hereafter referred to as *outer-sphere* coordination sites).<sup>10</sup> In both cases, when embedded in a RNA filament, Mg<sup>2+</sup> ions directly/indirectly interact with several different combinations of RNA atoms (i.e., *coordination patterns*, CPs), which can constitute specific and recurrent structural architectures (i.e., *binding motifs*). These are

Received: September 14, 2016

Published: December 8, 2016

specific structural arrangements provided by RNA for  $Mg^{2+}$  binding that are found in multiple RNA molecules.<sup>11,12</sup>

Recently, a comprehensive study considering all RNA crystal structures deposited in the protein data bank (PDB) classified the  $Mg^{2+}$ -RNA CPs and binding motifs, assaying also the statistical binding preferences of this metal toward different RNA ligands.<sup>11</sup> This analysis pointed out 41 and 95 distinct types for *inner-sphere* and *outer-sphere* CPs, respectively, among which 13 recurrent binding motifs were observed.<sup>11</sup>

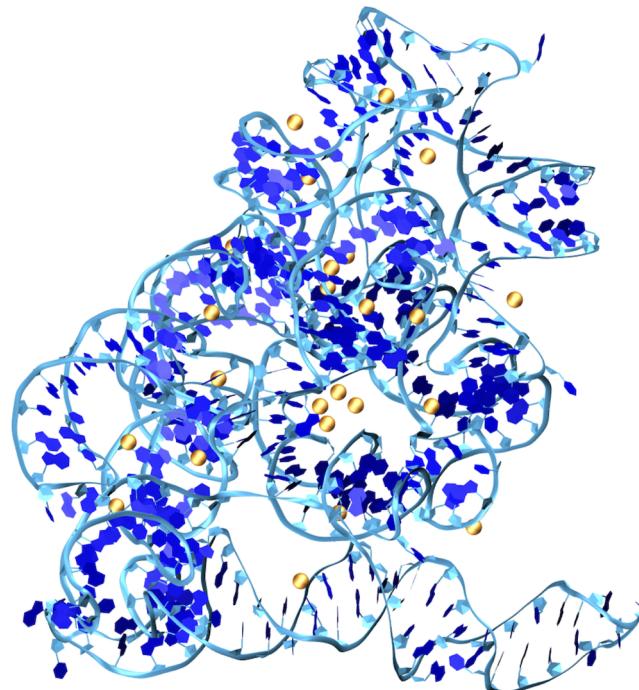
In spite of the critical role of  $Mg^{2+}$  ions for RNA functions, the experimental characterization of  $Mg^{2+}$ -RNA binding sites is currently limited for several reasons: (i)  $Mg^{2+}$  ions are difficult to detect via X-ray crystallography given their anomalous X-ray scattering and their isoelectronicity with water molecules and  $Na^+$  ions.<sup>11</sup> (ii) These ions are silent to most spectroscopic techniques, such complicating the dissection of their binding site composition on the basis of the ligands spectroscopic signature. (iii) Catalytic RNA is most often crystallized in the presence of metals inhibiting catalysis, thus biasing the identification of catalytic  $Mg^{2+}$  ions in nucleic acids.<sup>13</sup> (iv) A detailed understanding of the dynamical interplay between  $Mg^{2+}$  and RNA requires an atomistic resolution, which may be difficult to access in most wet-lab experiments.<sup>14</sup>

In this respect, molecular simulations can constitute a valuable support in the characterization of  $Mg^{2+}$ -RNA binding sites. Force field (FF)-based molecular dynamics (MD) simulations are largely employed for the study of biological macromolecules, including systems containing  $Mg^{2+}$  ions.<sup>3,4,7,8,15,16</sup> In current FFs, these ions are typically represented by fixed-point charges, bearing a doubly positive charge that electrostatically interacts with RNA macromolecules. However,  $Mg^{2+}$  can also exert a significant amount of nontrivial interactions such as polarization and charge transfer (CT) effects,<sup>5,17,18</sup> which are difficult to capture with simplified empirical FFs. Although some recently developed polarizable FFs may partially overcome these issues, their application has been so far mostly limited to model  $Mg^{2+}$ -water interactions.<sup>19–21</sup> Instead, an *ad hoc* bonded parametrization of  $Mg^{2+}$ , in which the bonds between  $Mg^{2+}$  and its ligands are explicitly defined, constitutes an impractical solution. Indeed, at variance with transition metal ions,  $Mg^{2+}$  ions are ubiquitous in RNA, bind in a wide range of CPs<sup>1</sup> and do not form coordination bonds with their d orbitals.<sup>12</sup> An alternative model is represented by the multisite ion approach (hereafter referred to as *cationic dummy atom* (CDA)), in which the  $Mg^{2+}$  ion is replaced by a central atom covalently bound to six dummy sites placed in the direction of coordinating atoms (i.e., at the vertexes of an octahedron) and parametrized to reproduce both the geometrical and energetic features of  $Mg^{2+}$ . The 2+ charge is differently (i.e., depending on the CDA type) distributed among the central atom and the dummies.<sup>22–24</sup> Finally, another recent solution has been proposed on the basis of a modified “12–6–4” van der Waals (vdW) potential for divalent metal ions.<sup>25,26</sup>

An accurate parameter-free description of  $Mg^{2+}$  binding to RNA can be achieved via mixed quantum mechanics/molecular mechanics (QM/MM) methods, which allow taking explicitly into account the electronic structure of a small QM part, while the rest of the system is described at a classical level.<sup>27</sup> These methods have been extensively employed to study ribozyme catalysis.<sup>7,8,28–31</sup> However, QM/MM studies are restricted in the size of the QM region and in the accessible time scales

(*pico*-seconds). This hampers the characterization of many  $Mg^{2+}$  binding sites (due to the large size limit), as well as of their influence in the structural and folding properties (due to the time-scale limit). Evidently, FF-based MD still represents the most useful tool to gain insights into the long-time scale conformational and compositional changes of the  $Mg^{2+}$  coordination sites, as well as on the associated RNA folding properties. As such, the development of reliable FFs for  $Mg^{2+}$  ions is urgently needed.

Here, we report a comparative study relying on extensive (a total of  $\sim 2.5 \mu s$ , 100/200 ns for each run) classical MD simulations, based on the AMBER-ff12SB FF,<sup>32,33</sup> in combination with five different nonpolarizable  $Mg^{2+}$  FFs in order to benchmark their relative performances. These simulations are carried out on two prototypical ribozymes, namely, the group IIC intron (GII-I)<sup>13</sup> (Figure 1) and the



**Figure 1.** Group IIC intron (GII-I) ribozyme from *Oceanobacillus iheyensis* (PDB entry 4FAQ),<sup>13</sup> including 24  $Mg^{2+}$  binding sites. The ribozyme is depicted using blue ribbons, while  $Mg^{2+}$  ions are represented as orange spheres.

hepatitis delta virus (HDV),<sup>34</sup> at two different  $Mg^{2+}$  concentrations. Density Functional Theory (DFT) calculations on small models of representative  $Mg^{2+}$  coordination sites (both of *inner-* and *outer-sphere*) have been further performed, providing a systematic characterization of electronic effects occurring at *inner-* and *outer-sphere*  $Mg^{2+}$  coordination sites. These calculations unveil for the first time the general principles driving the binding of  $Mg^{2+}$  ions to varying RNA motifs and pin down possible sources of errors in the current FFs. This information can help for an adept use of currently available  $Mg^{2+}$  FFs and in the rational development of next-generation  $Mg^{2+}$  FF models.<sup>26,35</sup>

## 2. METHODS

**2.1. Model Systems.** MD simulations of GII-I ribozyme from *Oceanobacillus iheyensis* were done starting from the X-ray structure of the reactive adduct (PDB code 4FAQ, solved at

3.11 Å resolution).<sup>13</sup> We built two model systems of GII-I with different Mg<sup>2+</sup> concentrations, namely, [Mg<sup>2+</sup>] = 10 and 25 mM. The former is the experimental concentration at which the self-catalyzed splicing reaction occurs,<sup>7</sup> while the latter was selected to estimate the effect of Mg<sup>2+</sup> concentration on its coordination properties. Both the simulations were carried out at 150 mM for monovalent ions. The first system was solvated with a 18 Å layer of TIP3P water molecules,<sup>36</sup> reaching a total of ~330 000 atoms and containing 24 Mg<sup>2+</sup>, 20 K<sup>+</sup> (originally present in the X-ray structure) and 327 Na<sup>+</sup> ions. The second system was solvated with a 14 Å layer of TIP3P waters,<sup>36</sup> thus leading to ~270 000 atoms,<sup>3</sup> with 48 Mg<sup>2+</sup>, 20 K<sup>+</sup>, and 279 Na<sup>+</sup> ions. The X-ray structure of the reactive adduct was solved in the presence of the catalytically inactive Ca<sup>2+</sup> ions that in our simulations were substituted with the biologically active Mg<sup>2+</sup> ions. We decided to focus on the reactive adduct, such allowing the characterization of the structural and electronic features of a two-Mg<sup>2+</sup>-ion catalytic site, which is rare in RNA but extremely important for ribozymes catalysis. To assess the reliability of our model, we verified that the occupancy and the distribution of the divalent cation sites in the reactant state were retained in the X-ray structure of the first-step-splicing product (PDB entry 4FAR),<sup>13</sup> which instead was solved in the presence of Mg<sup>2+</sup> ions (Figure S1, Supporting Information (SI)).

MD simulations of the HDV ribozyme were based on the PDB entry 1VC7 (2.45 Å resolution),<sup>34</sup> in which the crystallized Sr<sup>2+</sup> ions were replaced by Mg<sup>2+</sup> ions. In conformity with GII-I, we built two models considering [Mg<sup>2+</sup>] = 10 and 25 mM. The first system (10 mM of Mg<sup>2+</sup>, including 6 Mg<sup>2+</sup> and 61 Na<sup>+</sup>) was solvated with an 18 Å layer of TIP3P waters, resulting in a total of ~83 000 atoms. The second system (25 mM of Mg<sup>2+</sup>, including 20 Mg<sup>2+</sup> and 33 Na<sup>+</sup>) was solvated with a 24 Å layer of TIP3P waters, corresponding to a total of ~110 000 atoms. The data harvested out of MD simulations on HDV were used as complement of the simulations of GII-I, and, as such, they are mostly reported in the SI.

**2.2. Classical MD Simulations.** MD was used to equilibrate the systems at physiological conditions. The AMBER-ff12SB<sup>32,33</sup> (ff99+bsc0+χOL3) was adopted for the RNA. In order to compare different Mg<sup>2+</sup> ions parametrizations, we considered nonbonded fixed point charge and CDA models. Among the former, we selected the parametrization due to Åqvist,<sup>37</sup> Allnér et al.,<sup>38</sup> and Li et al.<sup>39</sup> (see SI Section 1.1 and Table S1 for details on the parametrizations). Among the CDA models, we have considered the ones from Oelschlaeger et al.<sup>23</sup> and Saxena et al.<sup>22</sup> Monovalent ion parameters were taken from Joung et al.<sup>40</sup> MD simulations were carried out using an integration time step of 2 fs, keeping all bonds with hydrogen atoms fixed with SHAKE.<sup>41</sup> Temperature control (300 K) was performed by Langevin dynamics,<sup>42</sup> with a collision frequency  $\gamma = 1$ . Pressure control was accomplished by coupling the system to a Berendsen barostat,<sup>43</sup> at a reference pressure of 1 atm and with a relaxation time of 2 ps. All the simulations were carried out with the following protocol. First, the systems were subjected to energy minimization to relax the water molecules and the Na<sup>+</sup> counterions, keeping the RNA, Mg<sup>2+</sup>, and K<sup>+</sup> ions fixed with harmonic position restraints of 300 kcal/mol·Å<sup>2</sup>. Then, the systems were heated up from 0 to 100 K in the canonical ensemble (NVT), by running two NVT simulations of 5 ps each, imposing position restraints of 300 kcal/mol·Å<sup>2</sup> on the key elements of the catalytic site (Mg<sup>2+</sup>, K<sup>+</sup>, RNA ligands, and waters coordinating the two Mg<sup>2+</sup> ions) and of 100 kcal/mol·Å<sup>2</sup> on the remaining Mg<sup>2+</sup> and K<sup>+</sup> ions. The

temperature was further increased to 200 K in 100 ps of MD in the isothermal-isobaric ensemble (NPT), in which the restraint on the catalytic site was reduced to 25 kcal/mol·Å<sup>2</sup>. Subsequently, all the restraints were released, and the temperature of the system was ultimately raised up to 300 K in a single NPT simulation of 1 ns. After ~1.1 ns of equilibration, ~10 ns of NPT production was carried out allowing the density of the system to stabilize around 1.01 g/cm<sup>3</sup>. Finally, production runs were carried out in the NVT ensemble, collecting ~100/200 ns, depending on the system. A list of all the simulations performed as well as their length is provided in Table S2. Overall, a total of ~2.5 μs of classical MD simulations has been done by using the Amber12<sup>44</sup> code in its GPU CUDA accelerated PMEMD version. Although several parametrizations have been recently proposed to overcome some of the identified problems,<sup>33,45–47</sup> it is well known that RNA FFs experience problems for long MD simulations. Hence, in order to increase the statistics of the Mg<sup>2+</sup> binding sites, avoiding the incurrence of long-time RNA FF instabilities, we decided to perform several simulations of limited length (~100/200 ns) with different starting conditions (assigning different random velocities). We remark that the time scale of Mg<sup>2+</sup> association/dissociation to/from RNA is of the order of milliseconds,<sup>14</sup> which is currently not accessible to standard MD simulations. As such, we would not have been able to directly observe such events even by extending the lengths of the MD runs.

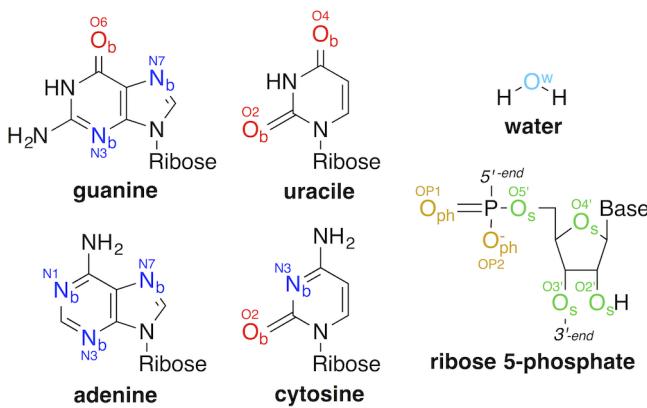
**2.3. Analysis of Structural Data.** Analysis of the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration ( $R_g$ ) and Radial Distribution Function ( $g(r)$ ) have been performed (see SI) with the *cpptraj* module of Amber12.<sup>44</sup> The coordination number (CN) analysis of all Mg<sup>2+</sup> ions was performed with Plumed 2.0<sup>48</sup> based on the switching function reported in eq 1:

$$s = \frac{1 - \left( \frac{r - d_0}{r_0} \right)^n}{1 - \left( \frac{r - d_0}{r_0} \right)^m} \quad (1)$$

where  $r_0$  is the cutoff distance of the first coordination shell, corresponding to the first peak of the calculated  $g(r)$  of Mg<sup>2+</sup> versus oxygen and nitrogen coordinating atoms, while  $d_0$  is assigned to zero;  $n$  and  $m$  are set to 50 and 100, respectively. In this analysis, we have divided the interacting RNA atoms in four groups (Figure 2): the RNA phosphate OP1 and OP2 oxygen atoms (labeled as O<sub>ph</sub>), the O2', O3', O4', O5' atoms of the ribose sugar (labeled as O<sub>s</sub>), the oxygen and the nitrogen atoms of the bases (labeled as O<sub>b</sub> and N<sub>b</sub>, respectively). Oxygen atoms of water molecules are labeled as O<sub>w</sub>. In order to make our analysis independent from the reference structure, we have also calculated the normalized interaction frequency of Mg<sup>2+</sup>-ligand contacts,  $F(X)$ , as defined in Zheng et al.<sup>11</sup> and reported in eq 2:

$$F(X) = \frac{p(\text{Mg}-X)}{p(X)} \quad (2)$$

Here,  $p(\text{Mg}-X)$  is the fraction of a given coordination type, namely, the number of the Mg<sup>2+</sup>-X interactions (where X is one atom type among O<sub>ph</sub>, N<sub>b</sub>, O<sub>b</sub>, O<sub>s</sub>, and O<sub>w</sub>) with respect to the total number of Mg<sup>2+</sup> interactions with all ligands;  $p(X)$  represents the fraction of atoms, i.e., the number of atoms of X type with respect to the total number of atoms in the data set



**Figure 2.** Donor atoms of RNA and water interacting with  $Mg^{2+}$  ions. Nitrogens ( $N_b$ ) and oxygens ( $O_b$ ) from nucleobases are shown in blue and red, respectively. Ribose oxygens ( $O_s$ ) are shown in green, while phosphate oxygens ( $O_{ph}$ ) in gold. Water oxygens ( $O_w$ ) are depicted in light blue.

(i.e., the system in analysis). Hence,  $F(X)$  measures the frequency of a particular  $Mg^{2+}$ –ligand interaction normalized by the frequency of the considered atom type in the overall structure.

**2.4. Density Functional Theory (DFT) Calculations.** To perform a systematic analysis of the charge transfer and polarization effects taking place between  $Mg^{2+}$  and its first and second shell ligands, DFT calculations have been done on a set of small model systems accounting for most of the  $Mg^{2+}$  binding motifs reported by Zheng et al.<sup>11</sup> In particular, from our MD simulations, we identified and extracted 16 models including from 19 to 72 atoms, in which the number of RNA ligands in the *inner-sphere* and in the *outer-sphere* range from zero to four, with the remaining ligands being water molecules. We remark that for *inner-sphere* coordination sites a “ligand” refers to an RNA moiety or water molecule directly interacting with  $Mg^{2+}$ , and the “donor atom” is the atom actually involved in the contact, while for *outer-sphere* coordination sites a “ligand” is intended as a functional group of a nucleotide (i.e., phosphate or base) interacting with a hexa-hydrated  $Mg^{2+}$  ion.

All the models were subjected to geometry optimizations and vibrational frequencies analysis at room temperature using the Gaussian 09 code.<sup>49</sup> For the *inner-sphere* models, we initially imposed a positional constraint on terminal carbon atoms of RNA ligands. However, in some cases, the RNA– $Mg^{2+}$  coordination distances of the models extracted from MD simulations were too short/long being affected by the FFs inaccuracies (Table S3), potentially biasing the analysis of the electronic effects. Thus, we treated all the *inner-sphere* sites without any constraint. Instead, for the *outer-sphere* models, since the RNA ligands are not directly interacting with  $Mg^{2+}$  ions, we run a first optimization cycle imposing a positional constraint on terminal carbons followed by an additional unconstrained optimization (starting from the obtained minima). DFT calculations were performed with the M06<sup>50</sup> and B3LYP<sup>51,52</sup> exchange correlation functionals and using the 6-311++G\*\* basis set. The Polarizable Continuum Model (PCM)<sup>53</sup> was used with water as a solvent (dielectric constant,  $\epsilon = 78.355$ ). For the sake of completeness, for the *inner-sphere* systems, we have also investigated the effect of a dielectric constant of 4 (i.e., representing the RNA environment).<sup>54</sup> For the geometry optimization the convergence criterion of the RMS force was set to  $1 \times 10^{-5}$ .<sup>55</sup> Natural Bond Orbital (NBO)

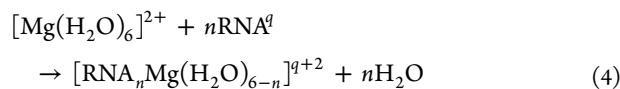
charges<sup>56</sup> have been calculated using the NBO 6.0 program.<sup>57</sup> For comparison, Bader charge analyses were also performed using the program developed by Tang et al.<sup>58</sup> To check the dependence of NBO charges on the basis set employed, we also run single point calculations with the 3-21G basis set on the geometries optimized with the 6-311++G\*\* basis set.

Charge transfer (CT) to  $Mg^{2+}$  ion from first/second shell ligands was quantified on the basis of the effective (NBO/Bader) charge of the  $Mg^{2+}$  ion with respect to its formal 2+ charge. For the *outer-sphere* sites, the CT occurring between the second and the first shell of coordination is determined as the difference between the formal charge of the second shell ligands and their NBO/Bader charge when interacting with the hexa-hydrated  $Mg^{2+}$ . In our schematic model, when  $Mg^{2+}$  ion coordinates its ligands, it exerts a polarization effect on them, triggering a general ligand charge rearrangement (LCR) that eventually includes also the charge transfer toward the metal. In this scenario, it is difficult to clearly dissect the polarization from CT effects. Thus, hereafter, we discuss the polarization by estimating the LCR ( $\Delta q$ ). This is calculated by taking the difference between the NBO/Bader charge of the isolated ligands and the charge they assume when bound to  $Mg^{2+}$ . This operation is done considering the coordinating and non-coordinating atoms separately. In this manner  $\Delta q$  provides a qualitative and simplified picture of the polarization occurring throughout the ligands.

For each model, we have also calculated its free energy of formation  $\Delta G_{form}$ . For the *inner-sphere* coordination sites,  $\Delta G_{form-is}$  is calculated as the RNA ligand/water exchange free energy accordingly to eq 3,

$$\Delta G_{form-is} = [G_{Mg\text{-motif-is}} + nG_{\text{wat}}] - [G_{RNA\text{-is}} + G_{Mg\text{-wat}}] \quad (3)$$

where  $G_{Mg\text{-motif-is}}$  is the total free energy of the specific  $Mg^{2+}$  *inner-sphere* coordination site;  $G_{Mg\text{-wat}}$  is the free energy of  $Mg^{2+}$  hydrated by six explicit water molecules;  $G_{RNA\text{-is}}$  is the free energy of the same RNA motif without the  $Mg^{2+}$  ions bound, while  $nG_{\text{wat}}$  is the free energy of the  $n$  water molecules that are released during the formation of the  $Mg^{2+}$  coordination site. Considering, for example, the chemical equality (eq 4) for the formation of  $[RNA_nMg(H_2O)_{6-n}]^{q+2}$  coordination site:

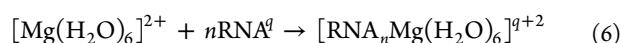


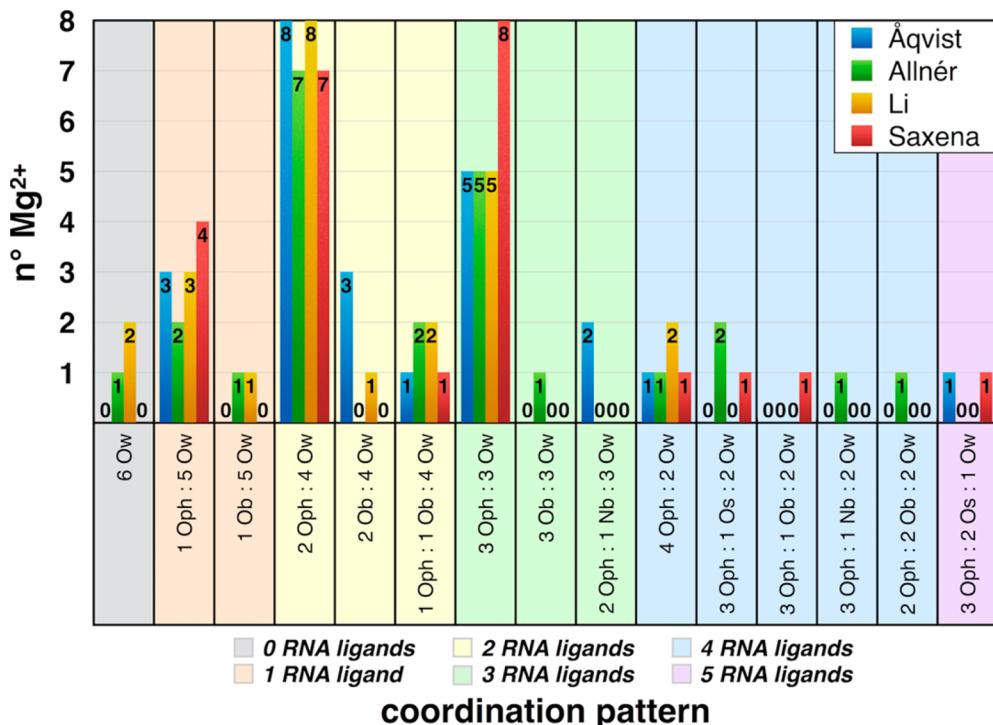
$G_{Mg\text{-motif-is}}$  is the free energy of the  $[RNA_nMg(H_2O)_{6-n}]^{q+2}$  *inner-sphere* coordination site;  $G_{Mg\text{-wat}}$  is the free energy of the  $[Mg(H_2O)_6]^{2+}$  complex,  $G_{RNA\text{-is}}$  is the free energy of  $nRNA^q$ , and  $nG_{\text{wat}}$  is the energy of  $nH_2O$ . In eq 4,  $q$  is the charge of RNA ligands.

For *outer-sphere* coordination sites,  $\Delta G_{form-os}$  is equal to (eq 5):

$$\Delta G_{form-os} = [G_{Mg\text{-motif-os}}] - [G_{RNA\text{-os}} + G_{Mg\text{-wat}}] \quad (5)$$

where  $G_{Mg\text{-motif-os}}$  is the total free energy of the specific  $Mg^{2+}$  *outer-sphere* coordination site;  $G_{Mg\text{-wat}}$  is the free energy of  $Mg^{2+}$  hydrated by six explicit water molecules;  $G_{RNA\text{-os}}$  is the free energy of the same RNA motif without the  $Mg^{2+}$  ions bound. This equation is consistent with the following chemical equality (eq 6):





**Figure 3.** Histograms showing the population of each  $Mg^{2+}$ –RNA CP observed during MD of GII-I performed with the Åqvist, Allnér, Li, and Saxena parametrizations at  $[Mg^{2+}] = 10$  mM. Bars of different colors are used to identify the FF model, as specified in the top right corner legend. The x-axis reports the CPs identified from MD simulations. The number of RNA ligands (from 0 to 5) is highlighted with different colors, as indicated in the bottom legend. The y-axis reports the number of  $Mg^{2+}$  ions (i.e., population) having a specific CP.

Corrections for basis set superposition errors (BSSE) and zero point energies have been applied, and we have also considered the entropic contribution (translational, rotational, and vibrational) to the free energy of formation.

### 3. RESULTS

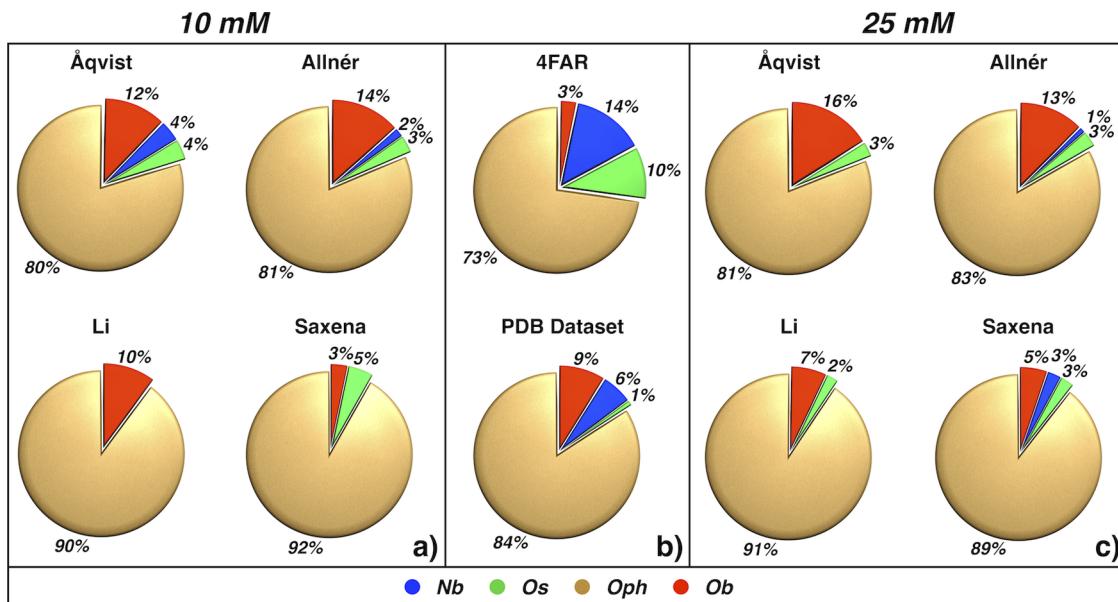
**3.1. Classical Force Field Models.** *3.1.1.  $Mg^{2+}$ -Coordination Sites Reproduced by FFs Models.* Classical MD simulations have been initially performed on GII-I considering five different  $Mg^{2+}$  FF parametrizations (i.e., Åqvist, Allnér, Li, Saxena, and Oelschlaeger). Besides its biological relevance as a model system of the eukaryotic spliceosome, we selected GII-I as a prototypical system to study  $Mg^{2+}$ –RNA interactions because of the following: (i) It is among the largest RNA macromolecules of known structure.<sup>6</sup> (ii) It presents a large number of  $Mg^{2+}$  binding sites (i.e., 24), including a catalytic bimetal site rarely resolved in ribozymes.<sup>6,8</sup> The transferability of the statistical analysis obtained for GII-I was tested by performing MD simulations also on HDV,<sup>34</sup> a well-characterized  $Mg^{2+}$ -dependent ribozyme. During the production runs, the structural stability of the two ribozymes has been evaluated by calculating RMSD, RMSF, and  $R_g$ . Except for a few cases, GII-I (Figures S2–S4) and HDV (Figures S14–S16) remain stable throughout all simulations, showing that different  $Mg^{2+}$  parametrizations and ionic strengths do not affect their overall structural stability.

Analysis of  $g_{Mg-X}(r)$  (Figure S5, Table S3) clearly shows that the distances between  $Mg^{2+}$ , and its ligands are underestimated in all cases with respect to the corresponding DFT values obtained in this study, with the only exception of the  $Mg^{2+}$ –N<sub>b</sub> distance, which is instead overestimated. For the Oelschlaeger parametrization, a different  $g(r)$  is observed, respect to all other models. This corresponds to unrealistically large distances

between  $Mg^{2+}$  and its ligands (see SI, Section 1.2 for additional details).

Examination of the coordination numbers (CN) at both  $[Mg^{2+}] = 10$  and 25 mM discloses that all the models reproduce the typical octahedral coordination sphere (i.e., CN = 6, data not shown), which is assumed by  $Mg^{2+}$  in water solution and constitutes its dominant coordination mode when bound to biological systems. An exception is again represented by the Oelschlaeger model, which often leads to an expanded coordination spheres (i.e., CN = 7/8). However, this model has been parametrized to reproduce the  $Mg^{2+}$  coordination sphere in protein enzymes (i.e., DNA polymerase  $\beta$ ).<sup>8,23</sup> For this reason, this DCA was not further considered here. For all the employed  $Mg^{2+}$  FFs, the CN stabilizes within the first ~70 ns of MD simulation, highlighting the convergence of our MD simulations with respect to  $Mg^{2+}$  coordination properties. However, a peculiar oscillatory behavior of the CN was overall observed for the  $Mg^{2+}$ –N<sub>b</sub> contacts.

The impact of the FF parametrization on the  $Mg^{2+}$  capability to account for multiple  $Mg^{2+}$ –RNA binding modes was investigated by performing a statistical analysis of the ligands composition of the  $Mg^{2+}$  inner-sphere binding sites, obtained from the MD simulations of GII-I at  $[Mg^{2+}] = 10$  mM (Figure 3). As a result, the Åqvist, Allnér, and Li FFs most frequently reproduce two CPs: the most populated CP is constituted by two O<sub>ph</sub> donor atoms and four water ligands (2O<sub>ph</sub>:4O<sub>w</sub>), while the second most populated one is composed by three O<sub>ph</sub> donor atoms and three water ligands (3O<sub>ph</sub>:3O<sub>w</sub>). These CPs are also well represented by the Saxena model, although their relative population is inverted. Remarkably, the Allnér model shows the widest range of possible CPs, although most of them with low population. Interestingly, in all simulations the most recurrent CPs are characterized by the presence of only O<sub>ph</sub>



**Figure 4.** Statistical distribution of RNA donor atoms in the  $\text{Mg}^{2+}$  *inner-sphere* (expressed as percentages) as obtained from MD simulations of GII-I performed with the Åqvist, Allnér, Li, and Saxena  $\text{Mg}^{2+}$  models at  $[\text{Mg}^{2+}] = 10 \text{ mM}$  (a) and  $25 \text{ mM}$  (c). Data for the 4FAR X-ray structure of GII-I and for the entire PDB data set relative to the *inner-sphere* sites are also reported (b).<sup>11,13</sup>

donor atoms as nonwater ligands, while the CPs with at least one  $\text{O}_b$  or one  $\text{O}_s$  donor atom are less frequent and the CPs with one  $\text{N}_b$  are rare.

A comparison with the ligand composition of  $\text{Mg}^{2+}$  sites in the X-ray structure of GII-I in the first-step-splicing product state (PDB entry 4FAR)<sup>13</sup> (Figure S6 vs Figure S8), shows that the crystallographic  $\text{Mg}^{2+}$  sites are only partially reproduced by the MD simulations. Sites 1 and 2, belonging to the catalytic binuclear site, are well accounted only by the Åqvist and Saxena FFs (see SI, Sections 1.3 and 2.3 for a detailed discussion on catalytic two- $\text{Mg}^{2+}$ -aided active site). Notably, the crystallographic sites in which  $\text{Mg}^{2+}$  is bound to  $\text{N}_b$  (i.e., sites 4, 16, 17, and 21) are not reproduced by any parametrization. Unfortunately, the limited resolution of the X-ray structure (i.e., 2.86 Å), which affects a complete characterization of the  $\text{Mg}^{2+}$  coordination spheres, hampers a well-defined structural match to assess the reliability of the  $\text{Mg}^{2+}$  FFs employed (Figure S6). However, an interesting comparison can be done between the MD CPs populations (Figure 3) and the histogram of the CPs identified in the whole PDB data set analysis<sup>11</sup> (Figure S7). This unveils that only five CPs (i.e.,  $1\text{O}_{\text{ph}}:5\text{O}_w$ ,  $2\text{O}_{\text{ph}}:4\text{O}_w$ ,  $1\text{O}_{\text{ph}}:1\text{O}_b:4\text{O}_w$ ,  $3\text{O}_{\text{ph}}:3\text{O}_w$ ,  $4\text{O}_{\text{ph}}:2\text{O}_w$ ) among those identified by Zheng et al.,<sup>11</sup> are reproduced by all the  $\text{Mg}^{2+}$  FFs here employed. Remarkably, the composition of the coordination sphere of the 24  $\text{Mg}^{2+}$  sites is conserved only for two sites (i.e., site 3 and 11) among all the  $\text{Mg}^{2+}$  parametrizations (Figure S8).

Finally, to establish if the  $\text{Mg}^{2+}$  models herein considered can reproduce the ligands distribution observed in RNA crystal structures, we performed an extensive statistical analysis comparing the percentages of the  $\text{Mg}^{2+}$  *inner-sphere* donor atoms, as obtained from our MD simulations (Figure 4a), with the data extracted by Zheng et al.<sup>11</sup> and with the crystal structure (4FAR)<sup>13</sup> (Figure 4b). From this analysis, we determined the following: (i) For all the  $\text{Mg}^{2+}$  FFs, the majority of the donor atoms is represented by  $\text{O}_{\text{ph}}$ , in agreement with Zheng et al.<sup>11</sup> (ii) The Åqvist, Allnér, and Li models overestimate the number of interactions with  $\text{O}_b$  atoms.

(iii) The Li and Saxena FFs do not account for the interactions with  $\text{O}_s$  (Li) and  $\text{N}_b$  (Li and Saxena). Interestingly, although in the crystal structure the number of  $\text{N}_b$  donor atoms (14%) is higher than in the PDB data set (6%), this is markedly reduced in all the simulated systems and completely disappears in the Li and Saxena models, clearly pinpointing an underestimation of the  $\text{Mg}^{2+}-\text{N}_b$  contacts by all the employed FFs. Overall, the Åqvist and Allnér parameters similarly account for the  $\text{Mg}^{2+}$ -RNA ligands contacts, showing a distribution that is in better compliance with the PDB data set.<sup>11</sup> To further check that these results were not biased by an arbitrary kinetic trapping of  $\text{Mg}^{2+}$  to the nearest close by minimum, we performed two additional MD simulations with different starting conditions for both Allnér and Åqvist models, confirming the trend discussed above (Figure S9).

By extending this analysis to the water ligands ( $\text{O}_w$ , Figure S10), which are not easily captured by X-ray crystallography, we found that water constitutes the majority of the *inner-sphere* ligands with the Åqvist and Allnér parametrizations exhibiting an almost identical statistical distribution. Since these models appeared to be the most reliable in reproducing the relative abundance of  $\text{Mg}^{2+}$ -RNA contacts present in the PDB data set, we considered only them in the following tests.

In a similar study,<sup>59</sup> the performance of the Åqvist, Allnér, Li FFs, and two additional models based on a modified 12–6–4 LJ potential developed by Li et al.<sup>25</sup> and by Panteva et al.<sup>26</sup> have been tested on the RNA stem-loop, which exhibits a  $\text{Mg}^{2+}$  ion-dependent conformational shift. All the FFs promoted the transition toward a  $\text{Mg}^{2+}$ -bound conformation. However, the formation of likely artificial chelated interactions in the 12–6–4 models interfered with folding of the RNA stem-loop. Interestingly, these 12–6–4 potentials showed the highest occupancy of the four key  $\text{Mg}^{2+}$ -RNA binding sites with a better description of  $\text{Mg}^{2+}-\text{N}_b$  interactions. Instead, consistent with our data, the point charge models exhibited a comparable occupancy centered on phosphate atoms throughout all the simulations, with a general underestimation of  $\text{Mg}^{2+}-\text{N}_b$  contacts.

**3.1.2. Effect of Mg<sup>2+</sup> Ions Concentration.** Specific Mg<sup>2+</sup> concentrations are important to modulate the activity of self-splicing ribozymes and to affect the stability of RNA molecules<sup>13,60</sup> since many body effects arise in the close presence of Mg<sup>2+</sup> ions.<sup>61</sup> However, given the difficulty to reproduce the experimental metal ions concentrations in MD simulations, the majority of the computational studies on biomolecules are usually done at high ionic strengths.<sup>3,62</sup> Simulations performed at [Mg<sup>2+</sup>] = 25 mM show the appearance of novel Mg<sup>2+</sup> CPs and the disappearance of some sites lowly populated at [Mg<sup>2+</sup>] = 10 mM (Figure S11). In particular, novel low populated sites (i.e., not identified by Zheng et al.<sup>11</sup>) come out by using Allnér (five sites) and Åqvist (two sites) models (Figure S11). Additionally, the relative abundance of the most populated patterns is altered. The Allnér model favors the O<sub>ph</sub>:5O<sub>w</sub> CP, which is the most abundant also in X-ray studies,<sup>11</sup> whereas the Åqvist model prefers the 2O<sub>ph</sub>:4O<sub>w</sub> pattern (Figure S11). Notably, the Mg<sup>2+</sup> concentration can also affect the composition of the original 24 coordination sites with the Allnér model being the less sensitive to the change in concentration (Figure S12).

The RNA donor atoms distribution in the *inner-sphere* sites (Figure 4b) also shows remarkable differences at the higher Mg<sup>2+</sup> ions concentration: (i) The N<sub>b</sub> interacting atoms are no longer accounted by the Åqvist model and are substituted by an increased percentage of O<sub>b</sub> donor atoms. (ii) The Saxena and Li models recover N<sub>b</sub> and O<sub>s</sub> contacts, respectively, but their relative abundances are rather far from what can be expected from the PDB data set. (iii) The statistical distribution of RNA ligands for the Allnér model does not change significantly, confirming its minor sensitivity to the ionic strength. (iv) By further including the water ligands in this analysis (Figures S12 and S13), some new *outer-sphere* sites appear and the number of Mg<sup>2+</sup>–O<sub>w</sub> contacts increases.

**3.1.3. MD Simulations on HDV.** To complement our statistical analysis, we performed MD simulations on the HDV ribozyme with the Åqvist and the Allnér FFs. At [Mg<sup>2+</sup>] = 10 mM, the number of sites was too low to be statistically relevant, and the simulation has been discussed only as a complement of the result of GII-I (Figure S20). Instead, at [Mg<sup>2+</sup>] = 25 mM, the same relative abundance of the most populated CPs, which emerged for GII-I, is encountered: (i) For the Allnér model the most recurrent pattern is 1O<sub>ph</sub>:5O<sub>w</sub> followed by 2O<sub>ph</sub>:4O<sub>w</sub>. (ii) For the Åqvist model, it is 2O<sub>ph</sub>:4O<sub>w</sub> followed by 1O<sub>ph</sub>:5O<sub>w</sub> (Figure S17). Consistent with what observed for GII-I, the composition of the binding sites is different with the two parametrizations employed, enlightening that our results are not system specific. Additionally, the simulations on the HDV ribozyme assess again the sensitivity of the *inner-sphere* coordination sites to varying Mg<sup>2+</sup> concentrations and the general difficulty of the currently available force fields in accounting for Mg<sup>2+</sup>–N<sub>b</sub> contacts (Figures S18 and S19). Obviously, the same holds true combining the statistics of both RNA systems investigated here (Figure S20).

In order to make our analysis independent from the structures investigated, we have also calculated the normalized interaction frequency F(X) between Mg<sup>2+</sup> and RNA donor atoms (Table 1), as defined by Zheng et al.<sup>11</sup> and detailed in the Methods section. Table 1, which displays the results for both GII-I and HDV ribozymes, shows that contacts with O<sub>ph</sub> and O<sub>s</sub> are reasonably reproduced by both Åqvist and Allnér parametrizations, while binding to O<sub>b</sub> atoms is overestimated (i.e., the Mg<sup>2+</sup>–O6@G contacts are overestimated in both

**Table 1. Normalized Interaction Frequency F(X)<sup>11</sup>**  
Calculated for *inner-sphere* Mg<sup>2+</sup>–RNA Contacts for the Whole PDB Data Set<sup>13</sup> and Obtained from Our Simulations Performed with the Åqvist and Allnér Parametrizations at [Mg<sup>2+</sup>] = 10 and 25 mM on GII-I and HDV<sup>a</sup>

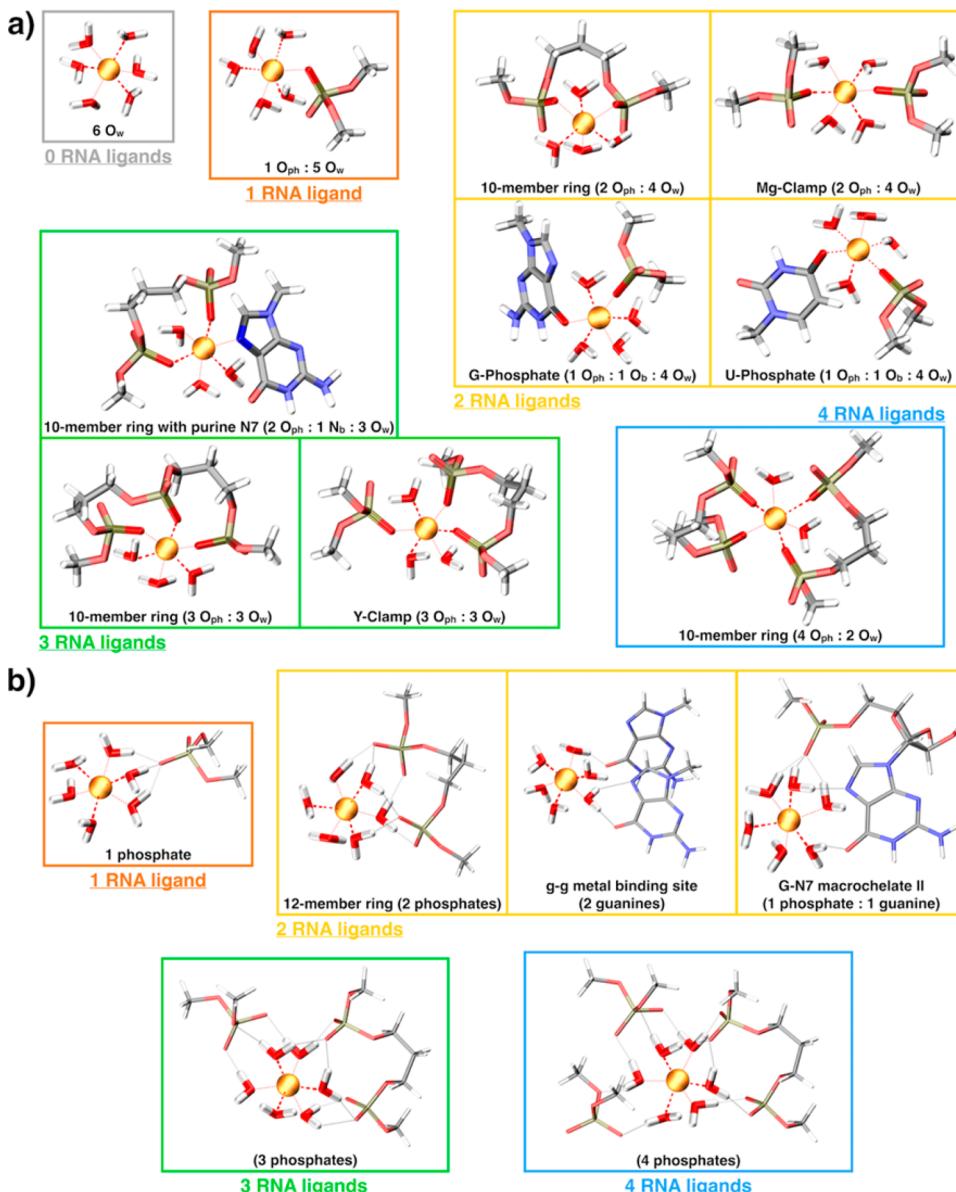
		F(X)					
		RNA donor atoms	Åqvist 10 mM	Allnér 10 mM	Åqvist 25 mM	Allnér 25 mM	PDB data set
O <sub>s</sub>	O <sub>ph</sub>	OP1	3.99	4.14	3.94	3.90	4.19
		OP2	4.60	4.75	5.02	5.22	4.99
		O2'	0.31	0.31	0.31	0.26	0.07
		O3'	0.15	—	—	—	0.04
		O4'	—	—	—	—	0.004
		O5'	—	—	—	—	0.04
A	N <sub>b</sub>	N1	—	—	—	—	0.04
		N3	—	—	—	—	0.003
		N7	0.58	0.58	—	—	0.74
G	O <sub>b</sub>	O6	2.95	3.44	2.97	2.55	1.46
	N <sub>b</sub>	N3	—	—	—	—	0.002
		N7	0.49	—	0.25	0.28	1.35
C	O <sub>b</sub>	O2	0.70	0.70	0.71	0.81	0.14
	N <sub>b</sub>	N3	—	—	—	—	0.01
U	O <sub>b</sub>	O2	0.76	0.76	0.38	0.44	0.08
		O4	1.53	0.76	1.92	1.76	2.33

<sup>a</sup>Reported data are comprehensive of the two studied systems and grouped according to the Mg<sup>2+</sup> model and the concentration conditions. RNA donor atoms, including the phosphate oxygens (O<sub>ph</sub>), the O3', O5' O2', O5' atoms of the sugar (O<sub>s</sub>), and the nitrogen (N<sub>b</sub>) and oxygen (O<sub>b</sub>) atoms of the bases are specified in the first column.

models and the Mg<sup>2+</sup>–O2@C contacts in Allnér, irrespective of the Mg<sup>2+</sup> concentration). Coherent with this observation, we remark that the Allnér model accounts for four CPs, characterized by the presence of at least one coordination to O<sub>b</sub> atoms, that are not observed in the PDB data set<sup>11</sup> (Figure S11). In addition, in the simulations with the Åqvist FF, one unprecedented site involving 4O<sub>b</sub> is noticed (Figure S11). These results may be related to lowly populated CPs that have not yet been identified in the PDB data set or, more likely, to inaccuracies in the Mg<sup>2+</sup> or RNA vdW parameters, as already pointed out in other studies.<sup>47</sup>

**3.2. Ab Initio Models.** A systematic analysis of the electronic effects taking place between Mg<sup>2+</sup> and its first and second shell ligands was done by performing DFT calculations on a set of cluster models representative of the recurrent Mg<sup>2+</sup>–RNA binding architectures (Figure 5). In the following, we have quantified the CT effect and the LCR as an estimate of the polarization effect, as detailed in the Methods section.

**3.2.1. Electronic Signature of Mg<sup>2+</sup>–RNA Inner-Sphere Coordination Sites.** The DFT-NBO analysis of the *inner-sphere* models shown in Figure 5a reveals that no coordination bonds are formed between Mg<sup>2+</sup> and first shell ligands, confirming that Mg<sup>2+</sup> interactions with the surrounding ligands is based exclusively on electrostatics and CT and polarization effects. In particular, when Mg<sup>2+</sup> ion coordinates its six ligands, it induces a charge movement from the more distant atoms toward those directly coordinated to Mg<sup>2+</sup>, resulting in their polarization. Simultaneously, part of the charge accumulated on the coordinating atoms is transferred to Mg<sup>2+</sup> ion (CT) (Figure 6d). As a consequence of this LCR, the more distant ligand atoms become more positive (i.e., display a positive Δq, Figure



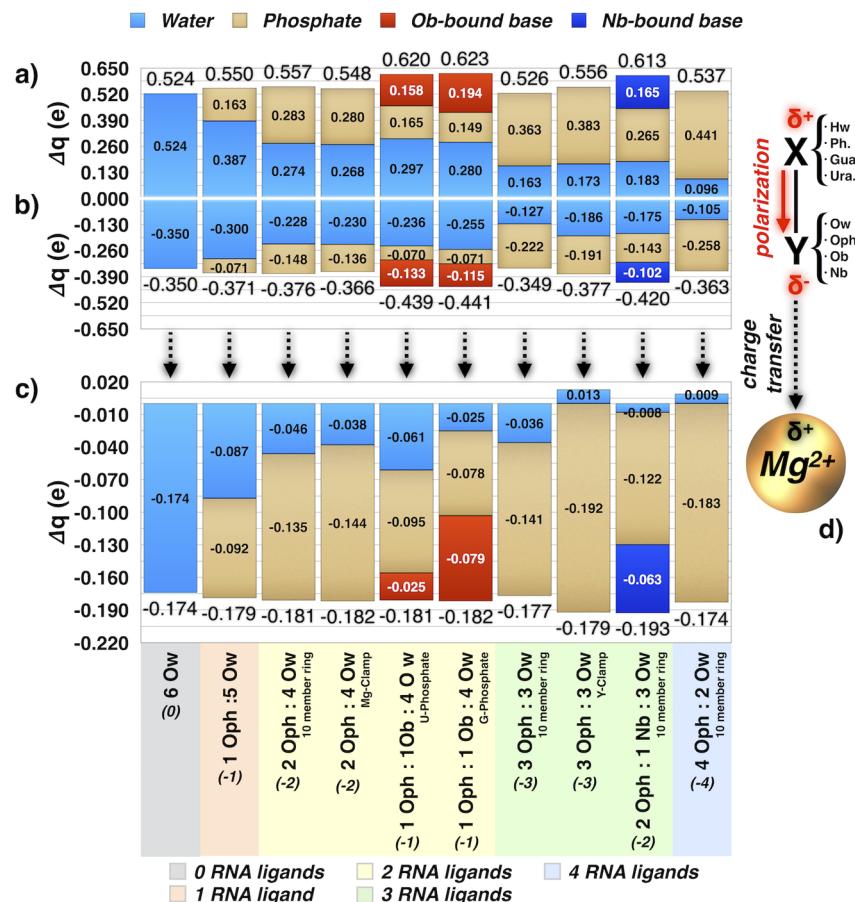
**Figure 5.** Model systems of  $\text{Mg}^{2+}$ -RNA binding architectures. Sixteen models have been selected, including both *inner-sphere* (a) and *outer-sphere* (b) coordination sites, in which the number of RNA ligands increases from 0 to 4. Boxes of different colors are used to identify the model systems characterized by 0 (gray), 1 (orange), 2 (yellow), 3 (green), and 4 (blue) RNA ligands. RNA ligands and water molecules are shown as sticks, while  $\text{Mg}^{2+}$  ions are shown as orange spheres. For each model, the motif name (if present) and the specific CP (in parentheses) are reported.

6a), while the directly coordinating donor atoms become more negative (i.e., show a negative  $\Delta q$ , Figure 6b). Here, we have monitored in detail the amount of CT and of LCR for each selected coordination site.

The LCR of the noncoordinating atoms (Figure 6a) is remarkable and similar in all models, ranging from +0.52 to +0.62 e. Obviously, the average variation of the partial charge of each atom is significantly smaller. By looking at the contribution of each ligand type, it arises that the charge rearrangement of the water hydrogen atoms ( $\text{H}_w$ ) conspicuously decreases when a nonwater ligand (phosphate, base, or both) is introduced in the coordination sphere (a decrease of 45% is registered for each  $\text{H}_w$  in the  $4\text{O}_{\text{ph}}:2\text{O}_w$  10-member ring<sup>11</sup> motif with respect to the  $6\text{O}_w$  model). This enlightens the dominant role played by RNA ligands, with the water molecules largely adapting their charge distribution to each specific binding site. The adaptive behavior of water ligands

stands out again as their strength as charge donors rapidly decreases to zero when more than two nonwater ligands are present in the coordination site. Interestingly, in the Y-Clamp ( $3\text{O}_{\text{ph}}:3\text{O}_w$ ) and 10-member ring ( $4\text{O}_{\text{ph}}:2\text{O}_w$ ) motifs,<sup>11</sup> the oxygens of the water molecules even accept from  $\text{Mg}^{2+}$  part of the charge donated by phosphate ligands, displaying a more negative  $\Delta q$ . Indeed, the coordinating atoms of the phosphates and the nucleobases (i.e.,  $\text{O}_{\text{ph}}$ ,  $\text{O}_b$ , and  $\text{N}_b$ ) of these models transfer a significant amount of charge to the  $\text{Mg}^{2+}$  ion (Figure 6b). As expected, the largest LCR is observed for the motifs containing aromatic nucleobases (i.e., guanine or uracil in the G-Phosphate, U-Phosphate and  $2\text{O}_{\text{ph}}:1\text{N}_b:3\text{O}_w$  motifs) due to the high delocalization of the electrons over the aromatic ring.

The analysis of the CT (Figures 6c and 7a) shows a relevant decrease in the  $\text{Mg}^{2+}$  NBO charge when a hexa-coordinated  $\text{Mg}^{2+}$  complex is formed ( $6\text{O}_w$  model). The  $\text{Mg}^{2+}$  ion is indeed capable of withdrawing  $-0.17$  e from its six ligands, reaching a



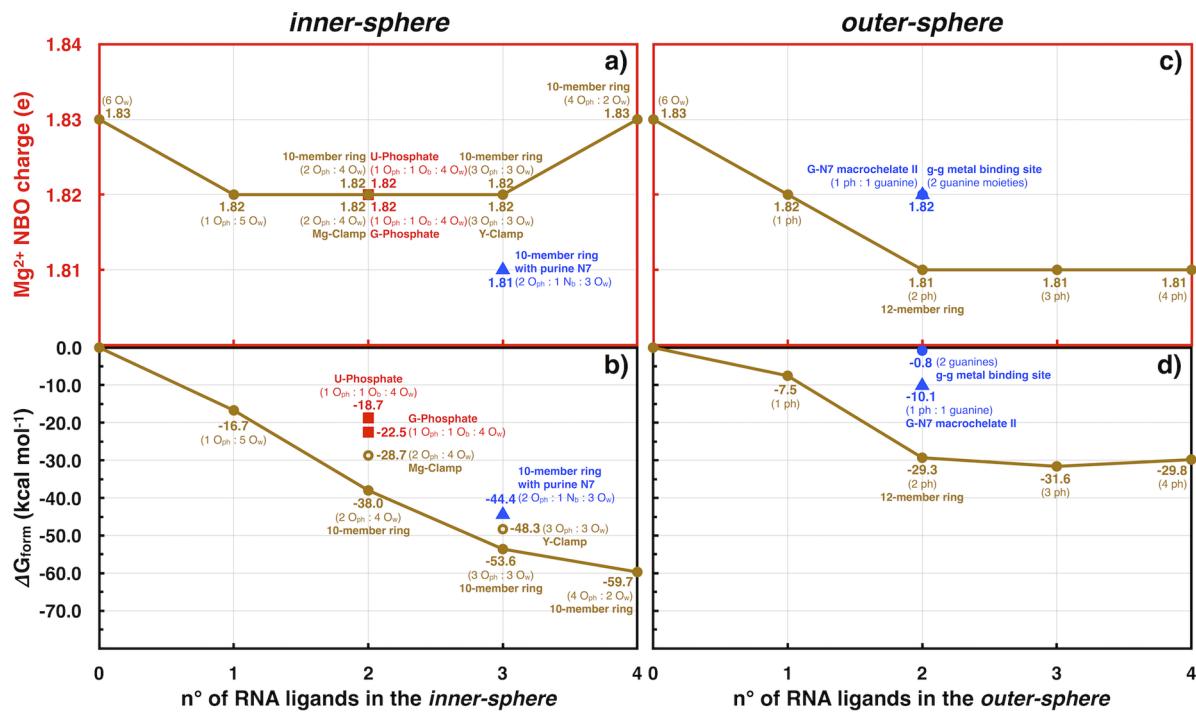
**Figure 6.** Charge rearrangements ( $\Delta q$ , e) of the (a) non- $Mg^{2+}$ -coordinated and (b)  $Mg^{2+}$ -coordinated ligands atoms in the *inner-sphere* coordination sites; (c) amount of charge transferred ( $\Delta q$ , e) from the ligands toward  $Mg^{2+}$  ion calculated from the NBO charge distribution and the M06 functional with 6-311++G\*\* basis set. Each contribution is dissected by atom type with light blue, gold, red, and dark blue referring to water ( $O_w$ ) and phosphate ( $O_{ph}$ ) and to nucleobases coordinated via  $O_b$  or  $N_b$  atoms, respectively. (d) Schematic picture of the polarization and charge transfer effects exerted by a  $Mg^{2+}$  ion. The formal charge of each CP is reported in parentheses. Despite the charge distribution accuracy is significantly lower, we report here fractional charges with three decimal digits in order to have a complete balance among the charges received by  $Mg^{2+}$  ions and the ones distributed over the ligands.

net charge of +1.83 (Figure 7a). Intriguingly, there are no differences in the amount of CT between the  $6O_w$  and  $4O_{ph}:2O_w$  models, and only very subtle variations exist in the other coordination sites. This clearly pinpoints that the amount of CT from the first shell ligands toward the  $Mg^{2+}$  ion does not depend on the number of nonwater ligands introduced but occurs in a limited and saturated manner (i.e., ~ 0.18 e transferred), consistently with a similar NBO study.<sup>12</sup> We remark that in a different study<sup>31</sup> on a two- $Mg^{2+}$  ion system the withdrawn charge, calculated with the Mulliken scheme, was ~0.43 e. Figures 6c and 7a strikingly show that the effective charge of  $Mg^{2+}$  is essentially constant, while that of the ligands is strongly adapting to the coordination environment.

The  $O_{ph}$  donors are responsible for the largest amount of CT, prevailing over the  $O_w$ ,  $O_b$ , and  $N_b$  (with the size of the base playing a noteworthy role in the last two cases, i.e., the largest CT occurs for guanine with respect to uracil). The trend observed for the NBO charges is independent from the exchange correlation functional employed, i.e., M06 vs B3LYP (Figures S24 and S25). Although a small dependence on the basis set (i.e., 6-311++G\*\* vs 3-21G) is observed for the amount of CT and LCR (Figure S26), the general trend is maintained. For the sake of completeness, the charge distribution analysis was done also with the Bader partitioning

scheme.<sup>58</sup> This shows a peculiar behavior for the hexa-hydrated  $Mg^{2+}$  ion motif ( $6O_w$ ), in which, at variance with all the other models, only a small CT takes place (Figure S27). However, all NBO results discussed here and obtained with an extended basis set are consistent with those observed from the Bader analysis.

For each model, we have also calculated the free energy of formation considering the water/nonwater ligand exchange reaction ( $\Delta G_{\text{form-is}}$  Figure 7b) (see Methods section). This gives a measure of how each type of binding site contributes to the stability of a folded RNA macromolecule.  $\Delta G_{\text{form-is}}$  shows an almost linear trend until three nonwater ligands interact with magnesium, while slightly deviating from the linearity when a fourth  $O_{ph}$  ligand is introduced. This is perfectly in line with the small decrease in both polarization and charge transfer occurring in this motif. If we consider the stabilization energy per RNA ligand in the coordination sphere (calculated as  $\Delta G_{\text{form-is}}$  divided by the number of RNA ligands), we clearly see that the contribution of the first phosphate is the largest, while all the subsequent ligands similarly contribute to the stability of the coordination site (Figure S28). Interestingly, remarkable differences among the models displaying the same number of RNA ligands (two or three) are observed. These are mostly due (i) to the electrostatic contribution associated with the formal



**Figure 7.** Mg<sup>2+</sup> charge (e) and free energy of formation ( $\Delta G_{\text{form}}$ , kcal/mol) of *inner-sphere* Mg<sup>2+</sup> coordination sites, (a) and (b), respectively, and *outer-sphere* Mg<sup>2+</sup> coordination sites, (c) and (d), respectively, plotted as a function of the number of RNA ligands and calculated at the DFT/M06/6-311++G\*\* level for the models shown in Figure 5. The Natural Bond Orbital (NBO) charge is used to estimate the charge. Gold circles, red squares, and blue triangles refer to model systems characterized by the presence of O<sub>ph</sub>-only and at least one O<sub>b</sub> or one N<sub>b</sub> as nonwater ligands, respectively. Model systems characterized by O<sub>ph</sub> ligands only but corresponding to a different geometrical isomer are indicated with golden empty circles. For each model system, the CP is reported in parentheses.

charge of the ligands (for example, between 3O<sub>ph</sub>:3O<sub>w</sub> and 2O<sub>ph</sub>:1N<sub>b</sub>:3O<sub>w</sub> exhibiting a -3 and -2 charge, respectively), (ii) to the entropic contribution in case of geometrical isomers like, for instance, a 10-member ring (2 consecutive phosphates) and Mg-Clamp (two separated phosphates) with the same 2O<sub>ph</sub>:4O<sub>w</sub> CP. Indeed, the negative entropic contribution upon Mg<sup>2+</sup> binding is clearly smaller for the former model, resulting in a more negative  $\Delta G$ . By changing the solvent dielectric constant to 4, thus simulating the RNA environment, we note a similar trend of the  $\Delta G_{\text{form-is}}$  even if in a larger extent, and an almost equal CT (~0.1% variation) toward Mg<sup>2+</sup> ions (Figure S29).

**3.2.2. Electronic Signature of Mg<sup>2+</sup>-RNA Outer-Sphere Coordination Sites.** A DFT-NBO analysis has been performed also for the *outer-sphere* models (Figure 5b). The polarization and CT effects occurring in each system are extended also to the second coordination shell (Figure 8e). Indeed, upon interacting with the hexa-hydrated Mg<sup>2+</sup> ion, the second shell ligands get polarized and transfer part of their charge to the first shell water molecules, resulting in a peculiar LCR of H<sub>w</sub> and O<sub>w</sub>. By virtue of this contribution from the *outer-sphere*, the  $\Delta q$  of the water hydrogens (H<sub>w</sub>) decreases (becomes less positive) with respect to the 6O<sub>w</sub> motif, while the O<sub>w</sub> atoms accumulate a large amount of negative charge. For the sake of clarity, while the CT taking place between the second and the first shell of coordination is evaluated for the *inner-sphere* sites, the polarization effect exerted on the *outer-sphere* RNA moieties is not detailed.

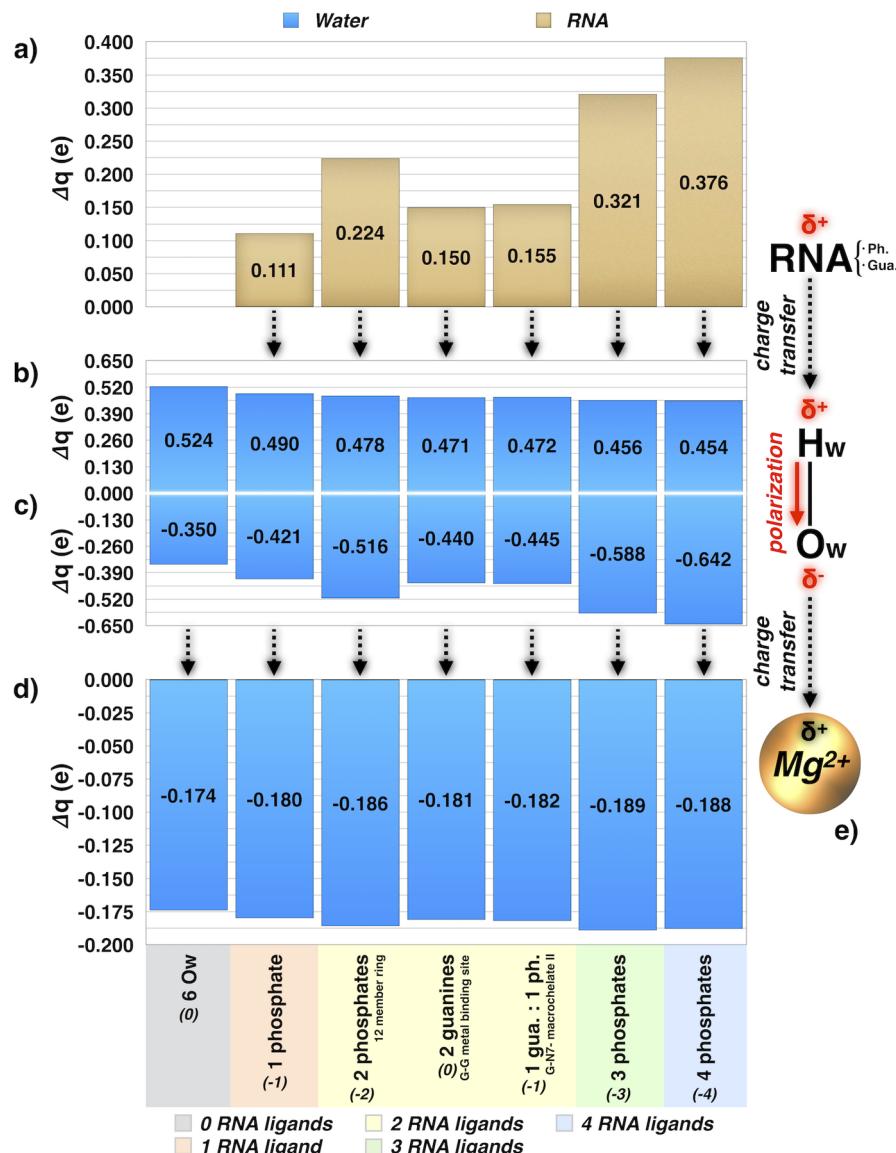
By looking at the CT taking place from the *outer-* to the *inner-sphere* ligands (Figure 8a), an almost linear contribution of each phosphate is observed, with the nucleobases participating to a lesser extent. We remark that the amount

of CT might depend on the number of atoms in the *outer-sphere* ligands directly interacting with the *inner-sphere* water molecules. The effect of introducing a second shell of ligands is clearly perceptible in the LCR of the first shell (Figure 8b and c). In fact, in the model with four phosphates in the outer shell, the  $\Delta q$  for each H<sub>w</sub> decreases by 13% (less positive), while the  $\Delta q$  for each O<sub>w</sub> becomes 2-fold more negative than in the free 6O<sub>w</sub> motif. Interestingly, the amount of charge transferred to the Mg<sup>2+</sup> ion (Figures 8d and 7c) is similar to that observed for the *inner-sphere* models. Also in this case, a constant NBO charge ranging from -0.18 to -0.19 e is withdrawn by the metal, independently on the number of the RNA moieties in the *outer-sphere*. This saturated amount of CT makes the LCR of the water molecules extremely relevant.

As expected, the  $\Delta G_{\text{form-os}}$  of the *outer-sphere* sites is smaller than the corresponding *inner-sphere* models (Figure 7d). After a second RNA ligand is introduced in the second shell of coordination, the free energy of formation reaches a plateau, suggesting that more than two second shell RNA ligands do not contribute to an extra stabilization, conversely to what observed for the *inner-sphere* sites. However, our results highlight the importance of *outer-sphere* coordination sites in the stabilization of RNA structures, even if in a lesser extent with respect to the *inner-sphere* ones. Finally, the  $\Delta G_{\text{form-os}}$  contribution per RNA ligand (Figure S28) reaches its maximum when two phosphates are present, then notably decreasing in the models with three and four RNA ligands.

## 4. CONCLUSIONS

The extensive benchmarks performed from MD simulations of GII-I and HDV ribozymes have shown remarkable differences



**Figure 8.** (a) Charge transfer ( $\Delta q$ , e) from the second shell ligands (phosphates or guanines) in the *outer-sphere* coordination sites; charge rearrangement ( $\Delta q$ , e) of  $H_w$  (b) and of  $O_w$  (c) of first shell water molecules; (d) amount of charge ( $\Delta q$ , e) transferred from the waters toward  $Mg^{2+}$  ion calculated from the NBO charge distribution and the M06 functional with 6-311++G\*\* basis set; (e) schematic picture of the polarization and charge transfer effects exerted by a  $Mg^{2+}$  ion. The formal charge of each CP is reported in parentheses.

in the performances of distinct  $Mg^{2+}$  FF models. This indicates that a conscious use of the classical FF parameters for  $Mg^{2+}$  ions is essential when running MD simulations of RNA filaments. This choice rigorously depends on the structural features of the specific  $Mg^{2+}$  site in RNA and on the experimental working conditions relative to  $Mg^{2+}$  concentration.

By performing DFT calculations on a representative set of model systems of  $Mg^{2+}$ –RNA binding architectures, we have strikingly disclosed for the first time that  $Mg^{2+}$  exhibits similar electronic properties in varying  $Mg^{2+}$ –RNA binding sites, while remarkable differences are observed for its surrounding ligands (Figures 6, 7a and c, and 8). As such, these results stunningly point out that the development of  $Mg^{2+}$  site-specific FFs is not in line with the physical origin of the inaccuracies in the  $Mg^{2+}$ –RNA MD simulations.<sup>26</sup>

We report below a series of practical FFs user guidelines for an adept use of current  $Mg^{2+}$  models and some key insights

from electronic structure calculations, constituting a rationale for the development of next-generation  $Mg^{2+}$  FFs.

**4.1. Practical FFs User Guidelines.** As a general feature, our extensive benchmark among the available  $Mg^{2+}$  ion models has revealed that all the FFs tend to underestimate the  $Mg^{2+}$ – $N_b$  contacts. Both Åqvist and Allnér parametrizations better account for the experimental distribution of RNA ligands,<sup>11</sup> although overestimating the  $Mg^{2+}$ – $O_b$  contacts.<sup>47</sup> Additionally, while the Allnér model reproduces more diverse and hydrated  $Mg^{2+}$  binding sites, the Åqvist model better accounts for highly RNA-coordinated sites (Figures 3 and 4). As such, our findings can summarily suggest the following:

- (i) The Allnér parameters are less sensitive to variations in  $Mg^{2+}$  ion concentration.
- (ii) The Allnér parameters tend to reproduce more hydrated sites and are thus indicated for *outer-sphere* binding sites or for small RNA molecules in which few highly chelated phosphate sites are present.

- (iii) The Åqvist parameters reproduce highly phosphate coordinated sites and may be the best choice in large RNA macromolecules.
- (iv) The Saxena parameters reproduce the active site properties of a two-Mg<sup>2+</sup>-ion motif irrespective of the Mg<sup>2+</sup> concentration employed and are thus recommended for these highly buried coordination sites in which other divalent ions are close by.
- (v) Careful attention should be given to the experimental working conditions relative to Mg<sup>2+</sup> concentration, on the basis of the remarkable differences shown by the Åqvist and Allnér FFs in reproducing the Mg<sup>2+</sup>-RNA coordination sphere.

**4.2. Mg<sup>2+</sup> Force Fields Developer Guidelines.** Surprisingly, this study reveals that the electronic properties of Mg<sup>2+</sup> remain constant in all these varying RNA environments. Given these findings, a set of developer guidelines is suggested below:

- (i) The net amount of charge transferred to Mg<sup>2+</sup> ion is roughly constant after a first RNA ligand is coordinated. This suggests that also the vdW parameters of Mg<sup>2+</sup> should be maintained in different RNA environments.
- (ii) Charge rearrangements on first shell RNA ligands are quite similar for the different binding motifs investigated, with the largest contribution coming from the nucleobases.
- (iii) Special care must be devoted to water molecules, which demonstrated the unique capability to act as a buffer, adapting the amount of charge transferred to Mg<sup>2+</sup> ion and the extent of polarization to the specific coordination site and for which the largest differences in LCR among the different binding sites are observed.
- (iv) The common view/practice of adapting Mg<sup>2+</sup> ions to the specific binding site seems to be in contrast with the physical origin of inaccuracies in the Mg<sup>2+</sup>-RNA MD simulations. In this scenario, the development of site specific FF parameters for the different binding sites appears to be an immediate and practical but unphysical manner to indirectly account for the different electronic effects induced by Mg<sup>2+</sup> on the surrounding ligands.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jctc.6b00905](https://doi.org/10.1021/acs.jctc.6b00905).

Additional figures and tables. ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [alessandra.magistrato@sissa.it](mailto:alessandra.magistrato@sissa.it).

### ORCID

Lorenzo Casalino: [0000-0003-3581-1148](https://orcid.org/0000-0003-3581-1148)

Giulia Palermo: [0000-0003-1404-8737](https://orcid.org/0000-0003-1404-8737)

Nodira Abdurakhmonova: [0000-0003-3702-3241](https://orcid.org/0000-0003-3702-3241)

### Present Address

Giulia Palermo: University of California at San Diego, 9500 Gilman Dr., M/C 0365, La Jolla, California 92093-0365, United States.

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank ISCRA (Grant HP10BPN8BA) for computational resources. U.R. acknowledges the Swiss National Computing Center for computing resources. L.C. thanks European Social Fund 2007/2013, Project DOCTOR EUROPAEUS.

## ■ ABBREVIATIONS

FF, Force Field; MD, Molecular Dynamics; DFT, Density Functional Theory; PDB, Protein Data Bank; CP, Coordination Pattern; CDA, Cationic Dummy Atom; vdW, van der Waals; GII-I, Group IIC Intron; HDV, Hepatitis Delta Virus; CN, Coordination Number; HFE, Hydration Free Energy; PCM, Polarizable Continuum Model; BSSE, Basis Set Superposition Error; NBO, Natural Bond Orbital; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Fluctuation;  $R_g$ , Radius of Gyration;  $g(r)$ , Radial Distribution Function; CT, Charge Transfer

## ■ REFERENCES

- (1) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262–272.
- (2) Ryan, M. F. *Ann. Clin. Biochem.* **1991**, *28* (1), 19–26.
- (3) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. *Acc. Chem. Res.* **2015**, *48*, 220–228.
- (4) Sgrignani, J.; Magistrato, A. *ACS Catal.* **2015**, *5*, 3864–3875.
- (5) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* **2010**, *43*, 40–47.
- (6) Marcia, M.; Pyle, A. M. *RNA* **2014**, *20*, 516–527.
- (7) Casalino, L.; Palermo, G.; Rothlisberger, U.; Magistrato, A. *J. Am. Chem. Soc.* **2016**, *138*, 10374–10377.
- (8) Sgrignani, J.; Magistrato, A. *J. Phys. Chem. B* **2012**, *116*, 2259–2268.
- (9) Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. *Nucleic Acids Res.* **1997**, *25*, 3421–3427.
- (10) Casalino, L.; Magistrato, A. *Inorg. Chim. Acta* **2016**, *452*, 73–81.
- (11) Zheng, H. P.; Shabalin, I. G.; Handing, K. B.; Bujnicki, J. M.; Minor, W. *Nucleic Acids Res.* **2015**, *43*, 3789–3801.
- (12) Petrov, A. S.; Bowman, J. C.; Harvey, S. C.; Williams, L. D. *RNA* **2011**, *17*, 291–297.
- (13) Marcia, M.; Pyle, A. M. *Cell* **2012**, *151*, 497–507.
- (14) Kowerko, D.; Konig, S. L.; Skilandat, M.; Kruschel, D.; Hadzic, M. C.; Cardo, L.; Sigel, R. K. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3403–3408.
- (15) Perez, A.; Luque, F. J.; Orozco, M. *Acc. Chem. Res.* **2012**, *45*, 196–205.
- (16) Vidossich, P.; Magistrato, A. *Biomolecules* **2014**, *4*, 616–645.
- (17) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669–8681.
- (18) Sponer, J. E.; Sobalik, Z.; Leszczynski, J.; Wichterlova, B. *J. Phys. Chem. B* **2001**, *105*, 8285–8290.
- (19) Yu, H. B.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 774–786.
- (20) Kumar, M.; Simonson, T.; Ohanessian, G.; Clavaguera, C. *ChemPhysChem* **2015**, *16*, 658–665.
- (21) Panteva, M. T.; Giambasu, G. M.; York, D. M. *J. Comput. Chem.* **2015**, *36*, 970–982.
- (22) Saxena, A.; Sept, D. *J. Chem. Theory Comput.* **2013**, *9*, 3538–3542.

- (23) Oelschlaeger, P.; Klahn, M.; Beard, W. A.; Wilson, S. H.; Warshel, A. *J. Mol. Biol.* **2007**, *366*, 687–701.
- (24) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C. L. *J. Phys. Chem. B* **2014**, *118*, 4351–4362.
- (25) Li, P. F.; Merz, K. M. *J. Chem. Theory Comput.* **2014**, *10*, 289–297.
- (26) Panteva, M. T.; Giambasu, G. M.; York, D. M. *J. Phys. Chem. B* **2015**, *119*, 15460–15470.
- (27) Carloni, P.; Rothlisberger, U.; Parrinello, M. *Acc. Chem. Res.* **2002**, *35*, 455–464.
- (28) Mlynksy, V.; Walter, N. G.; Sponer, J.; Otyepka, M.; Banas, P. *Phys. Chem. Chem. Phys.* **2015**, *17*, 670–679.
- (29) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. *J. Chem. Theory Comput.* **2013**, *9*, 857–862.
- (30) Lee, T. S.; Silva-Lopez, C.; Martick, M.; Scott, W. G.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 325–327.
- (31) Boero, M.; Tateno, M.; Terakura, K.; Oshiyama, A. *J. Chem. Theory Comput.* **2005**, *1*, 925–934.
- (32) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughon, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (33) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., III; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.
- (34) Kapral, G. J.; Jain, S.; Noeske, J.; Doudna, J. A.; Richardson, D. C.; Richardson, J. S. *Nucleic Acids Res.* **2014**, *42*, 12833–12846.
- (35) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 628–639.
- (36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (37) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.
- (38) Allner, O.; Nilsson, L.; Villa, A. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502.
- (39) Li, P. F.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.
- (40) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (41) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (42) Turq, P.; Lantelme, F.; Friedman, H. L. *J. Chem. Phys.* **1977**, *66*, 3039–3044.
- (43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (44) Case, D. A.; Darden, T. A.; Iii, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I. W.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C. A.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; Cheatham, T. E.; Goetz, A. W.; Kolossvai, I. AMBER12; University of California: San Francisco, 2012.
- (45) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., III *RNA* **2015**, *21*, 1578–1590.
- (46) Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E., III; Galindo-Murillo, R.; Jurecka, P. *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.
- (47) Chen, A. A.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820–16825.
- (48) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (49) 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, J. A., Jr.; 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.; Yazev, 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, 2009.
- (50) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
- (51) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (52) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (53) Improta, R.; Barone, V.; Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2006**, *125*, 054103.
- (54) Misra, V. K.; Hecht, J. L.; Yang, A. S.; Honig, B. *Biophys. J.* **1998**, *75*, 2262–2273.
- (55) Li, X.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 835–839.
- (56) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899–926.
- (57) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Landis, C. R.; Weinhold, F. NBO 6.0; Theoretical Chemistry Institute, University of Wisconsin, Madison, WI, 2013.
- (58) Tang, W.; Sanville, E.; Henkelman, G. *J. Phys.: Condens. Matter* **2009**, *21*, 084204.
- (59) Bergonzo, C.; Hall, K. B.; Cheatham, T. E., III *J. Chem. Theory Comput.* **2016**, *12*, 3382–3389.
- (60) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. *Cell* **2005**, *121*, 1005–1016.
- (61) Zhu, Y.; Chen, S. *J. J. Chem. Phys.* **2014**, *141*, 055101.
- (62) Ho, M. H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2010**, *132*, 13702–13712.