Influence of Na\(^+\) and Mg\(^{2+}\) ions on RNA structures studied with molecular dynamics simulations

Nina M. Fischer\(^1\), Marcelo D. Polêto\(^1,2\), Jakob Steuer\(^1,3\) and David van der Spoel\(^1,\ast\)

\(^1\)Uppsala Centre for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden, \(^2\)Center of Biotechnology, Universidade Federal do Rio Grande do Sul, Bento Gonçalves 9500, BR-91500-970 Porto Alegre, Brazil and \(^3\)Department of Chemistry, University of Konstanz, Universitätsstraße 10, D-78457 Konstanz, Germany

ABSTRACT

The structure of ribonucleic acid (RNA) polymers is strongly dependent on the presence of, in particular Mg\(^{2+}\) cations to stabilize structural features. Only in high-resolution X-ray crystallography structures can ions be identified reliably. Here, we perform molecular dynamics simulations of 24 RNA structures with varying ion concentrations. Twelve of the structures were helical and the others complex folded. The aim of the study is to predict ion positions but also to evaluate the impact of different types of ions (Na\(^+\) or Mg\(^{2+}\)) and the ionic strength on structural stability and variations of RNA. As a general conclusion Mg\(^{2+}\) ions are found to conserve the experimental structure better than Na\(^+\) and, where experimental ion positions are available, they can be reproduced with reasonable accuracy. If a large surplus of ions is present the added electrostatic screening makes prediction of binding-sites less reproducible. Distinct differences in ion-binding between helical and complex folded structures are found. The strength of binding (\(\Delta G\‡\) for breaking RNA atom-ion interactions) is found to differ between roughly 10 and 26 kJ/mol for the different RNA atoms. Differences in stability between helical and complex folded structures and of the influence of metal ions on either are discussed.

INTRODUCTION

Positively charged ions play an essential role for the structural stability of RNA molecules. Especially, Mg\(^{2+}\) ions facilitate high structural complexity and folding arrangements that allow RNA molecules to perform various cellular functions (1). Apart from canonical functions assigned to RNA molecules such as being involved in protein synthesis, like messenger (m) or transfer (t) RNA, it is nowadays well established that RNAs act in many other biological processes. RNA molecules are for instance involved in gene regulation (e.g., small nuclear (sn), micro (mi) and small interfering (si) RNAs) and in enzymatic activity (e.g. ribozymes and ribonucleoprotein). In eukarya they also play a role in resistance to pathogenic and parasitic invaders (2,3). Some of these functions depend on the presence of metal ions. Mg\(^{2+}\) ions do not only stabilize specific RNA structures (4), but do also help to recognize binding partners and mediate catalytic processes (5–7). Hammerhead ribozymes are one well-known example that require metal ions to be present both for obtaining the correct three-dimensional fold and performing the ribozymes’ function (8–12).

The need for positively charged ions in close proximity to RNA molecules is not surprising given their negatively charged backbone. Each RNA nucleotide contains one phosphate group that carries one negative charge. Positive ions shield negative charges on the RNA backbone by reducing repulsive forces, thereby allowing intramolecular interactions and compact RNA-biopolymers. RNA molecules are stabilized internally by hydrogen bonds between nucleotides in the same plane and by base stacking. Positively charged ions can be divided into two main groups: (a) ions that bind to structurally well-defined sites in direct contact with or close to the RNA and (b) ions that form a cloud surrounding the RNA molecule (13). The classification of Mg\(^{2+}\) binding has been further refined (1,14,15): inner-sphere ions that form direct bonds with RNA atoms, outer-sphere ions that bind via a single hydration shell to the RNA, diffuse Mg\(^{2+}\) ions that bind via multiple hydration shells, and free ions where the RNA’s charge has no direct effect on the ions. Monovalent ions can also be part of the first group and can bind sequence-specific to electronegative pockets formed by RNA structures (16–18); they should not just be considered to be part of a diffuse ionic cloud as described by Manning (19). Folding studies have shown that tRNA thermodynamic stability increases when adding monovalent (in particular Na\(^+\) and K\(^+\)) and divalent (Mg\(^{2+}\)) ions (20–23). Mg\(^{2+}\) ions are however the
Wealsodeletedone

separatesingleRNAstrand (chain

ditionsforallstructuresinthedataset.Inthecaseof1D4R

or other small molecules to ensure identical starting con-

perimental structures, we removed water molecules, ions,

of the twelve structures were obtained by X-ray crystallog-

raphy include Mg$^{2+}$ and 10% Na$^+$ ion positions. Less than

percent of the structures solved by NMR contain Mg$^{2+}$
or Na$^+$ ions. An important reason for this is that Mg$^{2+}$ ions,

ions, and water molecules have the same number of
electrons. Thus, it is difficult to distinguish them from one
another in electron density maps alone. They can only be as-

signed unambiguously in very high-resolution X-ray crystal
structures (13,33–35). During the last years special NMR
protocols have been developed to be able to study metal ion
binding to RNA. However, a special sample preparation is
needed to be able to detect Mg$^{2+}$ and Na$^+$ ions (36).

Recent studies describe monovalent- and divalent ion
binding and the influence of different ions on RNA struc-
tures. Molecular dynamics (MD) simulations of have been
reported suggesting that many Mg$^{2+}$ ions are strongly asso-
ciated with RNA, but not directly bound (15). A review by
Lipfert et al. (37) describes in detail the difference between
direct binding of ions and longer range (‘ion atmosphere’) asso-
ciation to nucleic acids and how this influences structure
and stability of RNA and DNA.

In this paper, we apply explicit solvent molecular dynam-
ics simulations on a dataset of twenty-four RNA structures.
Therefore, we can compare the implications of applying
Na$^+$ or Mg$^{2+}$ ions on helical and complex folded structures,
solved by either X-ray crystallography or NMR techniques.
By combining results from different structures and running
simulations in triplicate we can statistically distinguish true
effects of ions from stochastic fluctuations inherent to MD
simulations. Obviously the findings are still dependent on
force field quality and force fields for RNA have not been
scrutinized (38) to the same extent as those for proteins (39).

**MATERIALS AND METHODS**

**Dataset of RNA structures**

A dataset of 24 RNA structures was selected consisting
of twelve helical and twelve complex folded ones, (Table
1) from the Protein Data Bank (PDB, [http://www.rcsb.org/
pdb/](http://www.rcsb.org/pdb/)) (40) (Supplementary Figure S1). In both groups six
of the twelve structures were obtained by X-ray crystallog-
raphy and six by NMR experiments. When present in
experimental structures, we removed water molecules, ions,
or other small molecules to ensure identical starting con-
ditions for all structures in the dataset. In the case of 1D4R
(41) we also deleted one separate single RNA strand (chain
C) and in 1Q0 (42) the smaller RNA double helix (chain
A and B).

We define helical structures as those that form one dou-
ble helix composed of either one or two nucleotide strands.
In these structures, there are only very few unpaired nu-
cleotides present, i.e. single nucleotides at the strand’s ends
(1D4R (41), 4K31 (43) and 413D (44)). In the case of one
nucleotide strand forming a double helix there are unpaired
nucleotides in the loop region (1A4D (45), 2LPS (46), 2LV0
(47) and 2LJ2 (48)). In two structures, one single nucleotide
is sticking out of the helical main structure (2LPS (46) and
2QEK (49)). In contrast to helical RNAs, complex folded
RNA structures typically have a more globular shape, are
often multi-helical RNAs, and may be categorized as, e.g.,
ribozymes or pseudoknots.

**Ion selection and parameters**

MD simulations that intend to mimic cellular conditions
should use K$^+$ as monovalent ion, since it is the monova-

lent ion primarily found inside cells. However, we observed
formation of salt crystals of K$^+$ and Cl$^-$ ions in test simula-
tions (data not shown). The same issue was previously de-
scribed by other groups (65) when using K$^+$ instead of Na$^+
ions. For this reason, we chose Na$^+$ as representative for
monovalent ions in most of the simulations. Furthermore, we
used for Mg$^{2+}$ force field parameters refined by Allner et al. (66), that reproduce Mg$^{2+}$ hydration free energies and
exchange rates well.

Although it is known that the identity of the counter-ion
matters to the composition of the ion cloud around nucleic
acids (67) we have only used Cl$^-$ ions in this work. Further-
more, we note that other cation force field parameters have
been proposed for use in conjunction with nucleic acids (68),
however rather than comparing many different ion param-
eter sets we here focus on comparing different RNA struc-
tures.

**Molecular dynamics simulations**

RNA topologies were built using the parm99 (69) force field
with the GROMACS simulation package (version 4.6.7)
(70). First, an in vacuo minimization was carried out. Then,
each structure is placed in a rhombic dodecahedron box
filled with TIP3P water molecules to reach a ratio of 350
water molecules/nucleotide. During this step four different
systems are created (Supplementary Table S1), either with
just counterions (Cl$^-$) or at physiological salt (PS) concen-
tration:

- Na$^+$/Cl$^-$, just Na$^+$ counterions,
- Na$^+$/PS, Na$^+$ counterions plus 0.15 M/l NaCl,
- Mg$^{2+}$/Cl$^-$, just Mg$^{2+}$ counterions,
- Mg$^{2+}$/PS, Mg$^{2+}$ counterions plus 0.15 M/l NaCl.

All ions were placed randomly in the simulation box. We
monitored that Mg$^{2+}$ ions maintain a specific initial dis-
tance (>2 Å to any RNA atom) to the RNA and that direct
interactions did not occur during the minimization and first
four ns of the equilibration phase.
Table 1. Dataset of 24 RNA structures, twelve X-ray (six helical and six complex folded) and twelve NMR (six helical and six complex folded) structures, taken from the Protein Data Bank (PDB, http://www.rcsb.org/pdb/) (40). The fraction of single nucleotides (Nuc.) is calculated for each structure. Helical structures that have only base pairs have a fraction of 0.

<table>
<thead>
<tr>
<th>Method/PDB id</th>
<th>Classification/system</th>
<th>Fraction unpaired Nuc.</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>U</th>
<th>Mg²⁺</th>
<th>K⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X-ray</strong></td>
<td>Helical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1D4R (41)</td>
<td>Single recognition particle RNA</td>
<td>0.04</td>
<td>54</td>
<td>8</td>
<td>18</td>
<td>22</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>1QCO (42)</td>
<td>Plasmid copy control related RNA</td>
<td>0.00</td>
<td>38</td>
<td>6</td>
<td>13</td>
<td>13</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2QEK (49)</td>
<td>HIV-1 dimerization initiation site</td>
<td>0.04</td>
<td>46</td>
<td>14</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>4K31L (43)</td>
<td>rRNA A-site</td>
<td>0.05</td>
<td>44</td>
<td>8</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>413D (44)</td>
<td>A-form RNA double helix</td>
<td>0.04</td>
<td>26</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>420D (50)</td>
<td>RNA with (anti-G)(syn) mispairs</td>
<td>0.00</td>
<td>32</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td><strong>X-ray</strong></td>
<td>Complex folded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4BSR (51)</td>
<td>SAM-I riboswitch</td>
<td>0.30</td>
<td>94</td>
<td>28</td>
<td>22</td>
<td>32</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4FMc (52)</td>
<td>Guanine riboswitch aptamer</td>
<td>0.31</td>
<td>67</td>
<td>16</td>
<td>17</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4FRG (53)</td>
<td>Cobalamin riboswitch aptamer</td>
<td>0.34</td>
<td>84</td>
<td>27</td>
<td>19</td>
<td>22</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>4FJ2 (54)</td>
<td>Class II preQ1 riboswitch</td>
<td>0.37</td>
<td>76</td>
<td>19</td>
<td>19</td>
<td>22</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>4KQY (55)</td>
<td>S-box (SAM-I) riboswitch</td>
<td>0.34</td>
<td>119</td>
<td>40</td>
<td>27</td>
<td>33</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>4PSJ (56)</td>
<td>tRNA-like structure</td>
<td>0.36</td>
<td>83</td>
<td>17</td>
<td>28</td>
<td>22</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td><strong>NMR</strong></td>
<td>Helical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A4D (45)</td>
<td>Loop D/Loop E arm of 5S rRNA</td>
<td>0.07</td>
<td>41</td>
<td>6</td>
<td>10</td>
<td>17</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>2D18 (57)</td>
<td>HIV-1 dimerization initiation site</td>
<td>0.00</td>
<td>34</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2KYD (58)</td>
<td>A-form RNA double helix</td>
<td>0.00</td>
<td>32</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2L2J (48)</td>
<td>R/G stem loop RNA</td>
<td>0.10</td>
<td>42</td>
<td>7</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2LPS (46)</td>
<td>a5(gamma) group II intron</td>
<td>0.12</td>
<td>34</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>2LV0 (47)</td>
<td>Stem–loop from 23S rRNA</td>
<td>0.17</td>
<td>24</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td><strong>NMR</strong></td>
<td>Complex folded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1YMO (59)</td>
<td>Telomerase RNA pseudoknot complex</td>
<td>0.36</td>
<td>47</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>2ADT (60)</td>
<td>Tetraloop–receptor complex</td>
<td>0.31</td>
<td>86</td>
<td>22</td>
<td>16</td>
<td>26</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>2LKR (61)</td>
<td>U2/U6 snRNA</td>
<td>0.37</td>
<td>111</td>
<td>29</td>
<td>22</td>
<td>25</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>2MHI (62)</td>
<td>CR4/5 domain of telomerase RNA</td>
<td>0.25</td>
<td>53</td>
<td>7</td>
<td>15</td>
<td>17</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>2MTK (63)</td>
<td>Ribozyme’s III-IV-V junction</td>
<td>0.36</td>
<td>47</td>
<td>7</td>
<td>13</td>
<td>16</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>2MSK (64)</td>
<td>Telomerase RNA pseudoknot</td>
<td>0.30</td>
<td>48</td>
<td>12</td>
<td>9</td>
<td>6</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Average</td>
<td></td>
<td>56</td>
<td>13</td>
<td>14</td>
<td>16</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

All systems were minimized once more to eliminate any possible clashes and bad contacts. Subsequently, seven equilibration steps are carried out to provide a careful equilibration protocol. First, an NVT ensemble was conducted for 2 ns using position restraints with a force constant of 1000 kJ/(mol×nm²) to all heavy atoms. During this step the system was heated up to 300 K. Then, six NPT ensembles are conducted at 1 bar and 300 K for 26 ns in total. The number of restrained RNA atoms and the restraining force constant were gradually reduced while ions were given time to occupy preferred binding sites.

Finally, production runs were carried out for 50 ns at 300 K and 1 bar, with no restraints. An integration step of 2 fs was applied and all bonds were constrained using the LINCS (71,72) algorithm. A cutoff of 10 Å was used for Lennard–Jones and short-range Coulomb interactions and the particle mesh Ewald (PME) method (73) for long-range electrostatic interactions. Velocity rescaling (74) was used for temperature coupling with a time constant of 0.1 ps in order to ensure correct temperature fluctuations. For simulations at constant pressure we used the Parrinello-Rahman pressure coupling algorithm (75) with a time constant of 2 ps.

Each of the four systems was simulated three times in order to ensure statistical significance of our analyses. This resulted in 288 simulations and an overall simulation time of 14.4 Hs for all production phases.

**Force field evaluations**

Since most of our simulations were done using a somewhat old force field, an updated set of force field parameters for nucleic acids was tested, namely parmbsc0 (82) in conjunction with parmOL (83,84) for a subset of four RNA struc-
RESULTS

Structural changes

Figure 1 illustrates how the surrounding environment, espe-

500 ps from the equilibration phase and every 250 ps from the production phase trajectory. These structures were superimposed to the experimental RNA structure while only considering atoms with <4 Å root mean square fluctuation (RMSF) values. RMSF values are obtained to the RNA structure closest to the average structure of the second half of the production run. The input parameters for MobyWat that differ to the default parameters are the following: the maximum and minimum distance limits were set to 6.0 and 1.0 Å and the clustering tolerance to 1.5 Å. The results are based on the MER clustering algorithm that yield the best results comparing experimental- and predicted ion binding sites. The top 50 predicted ion binding sites were used and RMSD values with respect to the experimental ones calculated for each of them. The predicted ion binding site with the smallest RMSD to an experimental ion binding site was considered as a potential binding site.

Ion binding sites in RNA structures

The occupancies of Na\(^+\) and Mg\(^{2+}\) ions in close proximity of RNA structure were computed using the program MobyWat (80,81). Of the seven RNA structures with experimentally determined ion positions, structures were taken every 10 ps of the 50 ns production run. For each of the four ionic conditions.

Radial distribution functions

The radial distribution functions (RDFs) were determined using GROMACS and trajectories with structures taken every 10 ps of the 50 ns production run. For each of the four systems the RDFs are calculated between each of the RNA base (A-N1, A-N3, A-N6, A-N7, A-N9, G-O6, G-N1, G-N2, G-N3, G-N7, G-N9, C-O2, C-N1, C-N3, C-N4, C-O2, U-O4, U-N1 and U-N3), the two phosphate oxygen (O1P, O2P), or sugar oxygen (O2′, O3′, O4′ and O5′) atoms and positively charged ions (Na\(^+\) or/and Mg\(^{2+}\)) present in the system. In all structures, O1P is the atom that points towards the solvent and O2P (particularly in helical structures) towards the minor groove. The average RDFs are calculated for each RNA atom over the 12 helical and 12 complex folded structures.

Free energy of activation

The same RNA atoms and positively charged ions as described in the RDF analysis were used to determine the free energy of activation \(\Delta G\) for contact breaking. To specify the contact distance that is required as input parameter between an ion and a certain RNA atom we used the minimum between the first and second maxima from the corresponding RDF values for each structural replica. Similarly, the minimum between the second and third maxima determines the contact distance for second shell contacts. When no peak could be detected within a certain cutoff distance (3.5 or 6.0 Å) the contact distance for this RNA atom was calculated as the average over all minima of all other RNA atoms in this structure.

Ion binding sites in RNA structures

The occupancies of Na\(^+\) and Mg\(^{2+}\) ions in close proximity of RNA structure were computed using the program MobyWat (80,81). Of the seven RNA structures with experimentally determined ion positions, structures were taken every 10 ps of the 50 ns production run. For each of the four systems the RDFs are calculated between each of the RNA base (A-N1, A-N3, A-N6, A-N7, A-N9, G-O6, G-N1, G-N2, G-N3, G-N7, G-N9, C-O2, C-N1, C-N3, C-N4, C-O2, U-O4, U-N1 and U-N3), the two phosphate oxygen (O1P, O2P), or sugar oxygen (O2′, O3′, O4′ and O5′) atoms and positively charged ions (Na\(^+\) or/and Mg\(^{2+}\)) present in the system. In all structures, O1P is the atom that points towards the solvent and O2P (particularly in helical structures) towards the minor groove. The average RDFs are calculated for each RNA atom over the 12 helical and 12 complex folded structures.

Free energy of activation

The same RNA atoms and positively charged ions as described in the RDF analysis were used to determine the free energy of activation \(\Delta G\) for contact breaking. To specify the contact distance that is required as input parameter between an ion and a certain RNA atom we used the minimum between the first and second maxima from the corresponding RDF values for each structural replica. Similarly, the minimum between the second and third maxima determines the contact distance for second shell contacts. When no peak could be detected within a certain cutoff distance (3.5 or 6.0 Å) the contact distance for this RNA atom was calculated as the average over all minima of all other RNA atoms in this structure.

RESULTS

Structural changes

Figure 1 illustrates how the surrounding environment, espe-

500 ps from the equilibration phase and every 250 ps from the production phase trajectory. These structures were superimposed to the experimental RNA structure while only considering atoms with <4 Å root mean square fluctuation (RMSF) values. RMSF values are obtained to the RNA structure closest to the average structure of the second half of the production run. The input parameters for MobyWat that differ to the default parameters are the following: the maximum and minimum distance limits were set to 6.0 and 1.0 Å and the clustering tolerance to 1.5 Å. The results are based on the MER clustering algorithm that yield the best results comparing experimental- and predicted ion binding sites. The top 50 predicted ion binding sites were used and RMSD values with respect to the experimental ones calculated for each of them. The predicted ion binding site with the smallest RMSD to an experimental ion binding site was considered as a potential binding site.

RESULTS

Structural changes

Figure 1 illustrates how the surrounding environment, espe-

500 ps from the equilibration phase and every 250 ps from the production phase trajectory. These structures were superimposed to the experimental RNA structure while only considering atoms with <4 Å root mean square fluctuation (RMSF) values. RMSF values are obtained to the RNA structure closest to the average structure of the second half of the production run. The input parameters for MobyWat that differ to the default parameters are the following: the maximum and minimum distance limits were set to 6.0 and 1.0 Å and the clustering tolerance to 1.5 Å. The results are based on the MER clustering algorithm that yield the best results comparing experimental- and predicted ion binding sites. The top 50 predicted ion binding sites were used and RMSD values with respect to the experimental ones calculated for each of them. The predicted ion binding site with the smallest RMSD to an experimental ion binding site was considered as a potential binding site.

RESULTS

Structural changes

Figure 1 illustrates how the surrounding environment, espe-

500 ps from the equilibration phase and every 250 ps from the production phase trajectory. These structures were superimposed to the experimental RNA structure while only considering atoms with <4 Å root mean square fluctuation (RMSF) values. RMSF values are obtained to the RNA structure closest to the average structure of the second half of the production run. The input parameters for MobyWat that differ to the default parameters are the following: the maximum and minimum distance limits were set to 6.0 and 1.0 Å and the clustering tolerance to 1.5 Å. The results are based on the MER clustering algorithm that yield the best results comparing experimental- and predicted ion binding sites. The top 50 predicted ion binding sites were used and RMSD values with respect to the experimental ones calculated for each of them. The predicted ion binding site with the smallest RMSD to an experimental ion binding site was considered as a potential binding site.

RESULTS

Structural changes

Figure 1 illustrates how the surrounding environment, espe-
cially the presence or absence of Mg\textsuperscript{2+} ions, influences RNA structural changes during MD simulations. In agreement with other studies, root mean square deviation (RMSD) values are lower for RNA structures simulated with Mg\textsuperscript{2+} ions than without.

In addition to RMSD values, we calculated eRMSD values (76). This metric discriminates effectively between structurally and kinetically different RNA conformations. It directly describes variations in base-base interactions and therefore captures whether or not important structural characteristics, like base pairs, are preserved during the simulation. Bottaro et al. (76) showed that multiple different secondary RNA structures can be found within 4 Å RMSD of each other. Such two RNA structures with low RMSD values to a reference structure do not necessarily have the same secondary structures. Indeed, the base-base interactions could be completely lost in one structure and not in the other. This kind of structural differences is described by eRMSD values that takes structural information about base-pairing into account. An eRMSD of <0.8 indicates all base-base contacts are close to the native experimental structure and an eRMSD of >1 suggests non-native base-base contacts occur in the structure (76). There are more structures with an average (over three replicas) eRMSD >1 for systems simulated without Mg\textsuperscript{2+} than systems simulated with Mg\textsuperscript{2+} ions (Supplementary Table S2). The six complex folded NMR structures have almost always eRMSD values >1 regardless of the surrounding ionic environment, except for one structure (PDB id: 2ADT) and another structure (PDB id: 2MTK), when simulated with Mg\textsuperscript{2+} and NaCl. None of the structures obtained by X-ray crystallography have average eRMSD values >1 when Mg\textsuperscript{2+} ions were present during the simulation. In general, structures simulated with Mg\textsuperscript{2+} ions have lower RMSD and eRMSD values compared to structures simulated without Mg\textsuperscript{2+} ions.

We performed a statistical (kernel density) comparison test that compares the distributions of two-dimensional data points. It returns a p-value that is higher for better fits between the two distributions. Our null hypothesis is that the distributions are independent of whether Mg\textsuperscript{2+} ions are present in the simulations. When thus comparing simulations with a 0.15 M NaCl salt concentration with and without Mg\textsuperscript{2+} ions, the P-values for X-ray structures are both less than 0.001 (Figure 1 for individual P-values). This indicates that Mg\textsuperscript{2+} is significantly responsible for maintaining native base-base contacts during the simulations, at least for X-ray RNA structures. The P-values for NMR structures are higher and therefore statistically not significant.

**Ion binding**

In order to analyze where ions are located during the simulations, radial distribution functions (RDFs) were derived for positively charged ions present during the simulation and all RNA atoms. Direct contacts are identified between Na\textsuperscript{+} and 16 RNA atoms (O1P, O2P, sugar oxygen atoms, A-N1, A-N3, A-N7, G-O6, G-N7, C-O2, C-N3, U-O2 and U-O4). Mg\textsuperscript{2+} ions form direct contacts in the simulations only twice with one of the RNA atoms. In one case Mg\textsuperscript{2+} binds directly to an O1P atom and in another structure to a cytosine oxygen (C-O2). All other Mg\textsuperscript{2+} interactions with RNA occur indirectly via water molecules. Clearly recognizable first and second shell contacts between positively charged ions and RNA atoms can be determined for seven RNA atoms in RDFs (Figure 2).

The RDF peaks are higher for Na\textsuperscript{+} ions in system Na\textsuperscript{+}CI compared to Na\textsuperscript{+} ions in system Na\textsuperscript{+}PS and Mg\textsuperscript{2+}PS and for Mg\textsuperscript{2+} ions in system Mg\textsuperscript{2+}CI compared to Mg\textsuperscript{2+} ions in system Mg\textsuperscript{2+}PS. In both systems (Na\textsuperscript{+}CI and Mg\textsuperscript{2+}CI) fewer positively charged ions are therefore present in the bulk water surrounding the RNA molecule. We use the same arbitrary definition of bulk water/ions (distance >20 Å to any RNA atom) as described by Hayes et al. (15). In system Mg\textsuperscript{2+}PS when both Na\textsuperscript{+} and Mg\textsuperscript{2+} ions are present more Mg\textsuperscript{2+} ions are found in the bulk solvent than in system Mg\textsuperscript{2+}CI. Nevertheless, fewer Na\textsuperscript{+} ions are found close to the RNA in system Mg\textsuperscript{2+}PS compared to system Na\textsuperscript{+}PS and also system Na\textsuperscript{+}CI. When comparing helical versus complex folded RDFs, fewer Na\textsuperscript{+} ions are in direct contact with helical phosphate oxygen atoms compared to complex folded ones. In both structure groups there is a preference...
for O2P over O1P for Na⁺ as well as Mg²⁺ ions. For the nitrogen atom in adenine (A-N7) we observe only differences of the RDFs of Na⁺ ions in the second and forth system between helical and complex folded structures. For A-N7 there seems to be a preference for Na⁺ first shell binding compared to Mg²⁺ second shell binding interactions. When comparing Na⁺ direct binding between helical and complex folded structures for both favoured guanine atoms (G-O6 and G-N7), there are more occurrences in complex folded structures with a slight preference for G N7. This is in contrast to second shell Mg²⁺ interactions, where G-O6 is the preferred atom. The peak for helical G-O6 atoms and Mg²⁺ ions in system Mg²⁺PS is the highest determined for all RDFs. For U-O2 atoms very few ion contacts were found, and most of them are found in complex folded structures. Since helical structures mostly form Watson-Crick base pairs U-O2 atoms lie in the minor groove and it is likely therefore they not easily accessible to ions (Figure 2). The other uracil oxygen atom (U-O4) is slightly preferred by Na⁺ ions in first shell interactions in complex folded structures and by Mg²⁺ ions in second shell binding for helical structures.

### Ion binding energetics

An analysis method that was developed to study kinetics of hydrogen bond breaking and forming (77) and thermodynamics of hydrogen bond breaking in different environments (78) was used here for studying ion-binding energetics. This method yields the Gibbs energy of activation $\Delta G^\ddagger$ for contact breaking and was previously applied on RNA-ion contacts in a study of viral RNA (79). The highest energy for breaking first shell contacts was found between one phosphate oxygen atom (O1P) and a Mg²⁺ ion (Figure 3). This is the result of one of the two direct interactions between a Mg²⁺ ion and an RNA atom that occurred in the simulations, as also observed in the RDF analysis. The energy of first shell contacts between Na⁺ ions and RNA phosphate oxygen atoms (O1P and O2P) is not as high compared to other RNA atoms (A-N7, G-O6, G-N7, C-O2, C-N3, U-O2 and U-O4). These results differ from the RDF results insofar that the RDF peaks for C-O2, C-N3, and U-O2 atoms are very low especially compared to the peaks of phosphate oxygen atoms. The main difference between helical and complex folded first shell contacts is for atoms that are only available for interactions in complex folded structures (A-N3, A-N9, G-N9, U-N1). A-N3 lies in the minor groove in helical structures and the other atoms are the base atoms that are closest to the sugar ring. Therefore, they are not easily accessible to ions in helical RNA structures.

### Ion binding positions

To investigate whether Mg²⁺ ions find experimentally identified binding sites, when initially placed randomly in the solvent (with a distance >2 Å to any RNA atom), we determined the occupancy of Na⁺ and Mg²⁺ ions during the simulation using the software MobyWat (80,81). Figure 4 shows the top 10 predicted binding sites for Mg²⁺ and Na⁺ ions for one of the three replicas of each system during the equilibration phase superimposed on the X-ray structure of 2QEK (49). We chose 2QEK as example structure, because both monovalent (K⁺) and divalent (Mg²⁺) ions are present in this structure. When only Na⁺ ions are present in the simulation both K⁺ and Mg²⁺ binding sites are occupied (Figure 4A and B). This can also be observed for other structures (Table 2). In some cases it seems as if the binding site can be occupied by both Na⁺ and Mg²⁺ ions. Mg²⁺ binding sites are more difficult to predict with MD simulations since the hydration layer around Mg²⁺ ions is almost never dismantled. When only Mg²⁺ are present in the simulation, ions are closer to RNA atoms compared to when both Mg²⁺ and Na⁺ are present. The closest distance between experimentally pre-
The RMSD between experimental and predicted binding sites are given in Å. The 10 top ranked ion binding sites predicted with MobyWat for one replica each for four systems is shown: (A) Na⁺Cl (green), (B) Na⁺PS (blue), (C) Mg²⁺Cl (orange), (D) in Mg²⁺PS there are Mg²⁺ (red) and Na⁺ (purple).

Table 2. Average RMSD values between experimental and predicted binding sites during the production phase. The position of ions are predicted with MobyWat (80,81). The resulting top 50 ion positions are considered for each replica

<table>
<thead>
<tr>
<th>System</th>
<th>Na⁺Cl</th>
<th>Na⁺PS</th>
<th>Mg²⁺Cl</th>
<th>Mg²⁺PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion</td>
<td>Na⁺</td>
<td>Na⁺</td>
<td>Mg²⁺</td>
<td>Mg²⁺</td>
</tr>
<tr>
<td>1D4R</td>
<td>0.8</td>
<td>1.6</td>
<td>1.9</td>
<td>2.1</td>
</tr>
<tr>
<td>MG-90</td>
<td>0.5</td>
<td>0.7</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>MG-91</td>
<td>0.8</td>
<td>1.1</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-48⁴</td>
<td>0.6</td>
<td>0.8</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-49</td>
<td>0.8</td>
<td>1.1</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-50</td>
<td>0.5</td>
<td>0.7</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>MG-51</td>
<td>0.6</td>
<td>0.8</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-52</td>
<td>0.6</td>
<td>0.8</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>MG-53</td>
<td>0.8</td>
<td>1.1</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-49</td>
<td>0.5</td>
<td>0.7</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>MG-50</td>
<td>0.6</td>
<td>0.8</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-51</td>
<td>0.6</td>
<td>0.8</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>MG-52</td>
<td>0.6</td>
<td>0.8</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>MG-53</td>
<td>0.8</td>
<td>1.1</td>
<td>1.4</td>
<td>1.7</td>
</tr>
</tbody>
</table>

**DISCUSSION**

The structural analyses indicate (at least in X-ray structures and most NMR structures) that Mg²⁺ ions have a stronger stabilizing effect for helical structures than for complex folded structures (Figure 1). Although eRMSD values of complex folded structures, when simulated with and without Mg²⁺ ions are comparable, they are in general high (above 1), indicating that these structures do not maintain their native fold (76). This might be due to the fact that the quality of complex folded NMR structures is not as good as that of X-ray structures (Supplementary Table S1), for instance because structures that are inherently more flexi-

dicted Mg²⁺ binding sites and those observed in our simulation is 1-2 Å. Overall, Mg²⁺ binding sites are predicted better than K⁺ binding sites. This indicates a preference of Mg²⁺ ions to experimentally predicted Mg²⁺ ion binding sites. When both, Na⁺ and Mg²⁺ ions are present in the simulations the distances to experimentally predicted binding sites are higher compared to other systems. This is surprising since it does not correlate with lower RMSD or eRMSD values for those structures. It indicates that although specific ion positions are not found during MD simulations the overall structure maintains a native-like fold, potentially due to there being a 'sufficient' amount of screening of electrostatic interactions. In general we observe Mg²⁺ ions present along the minor groove of the RNA and in some specific binding sites.

The predicted binding sites are in good agreement with experimentally identified ion locations in close proximity to the RNA (Table 2). There are some cases for which the experimental binding site was not detected, however, in particular for ions directly bound to RNA. This is expected, due to the high barrier for desolvation of Mg²⁺ ions (66). Most of these sites are located at the surface of the RNA and only one RNA atom can be identified as potential contact site in experimentally predicted structures. For this reason RMSD values to the experimental binding sites are marked with an asterisk in Table 2 for these ions.
ble and difficult to solve by X-ray crystallography instead are solved by NMR techniques. Especially, the RNA backbone seems not to be as well defined for NMR structures based on the validation results of X-ray and NMR structures (Supplementary Table S1).

It has been reported (38) that helical RNA structures undergo irreversible structural changes in longer MD simulations (over 50 ns) when using parm99 and parmbsc0 (82). They change into a ladder–like structure, similar to what we observe in the majority of helical RNA structures with high RMSD values. The reason for this is that the glycoside torsion angle $\chi$ is shifted from the anti to the high-anti region. A specific force field parameter set, called parmOL (83,84), has been developed to eliminate this artifact. Since we did not use these parameters for most of the simulation, the correct backbone angle of some helical RNA structures was not maintained in this work. We did, however, use the combination of parmOL (83,84) and parmbsc0 (82) parameters specifically developed for RNA, for a subset of our structures. These structures undergo less structural changes and have lower RMSD and \( \epsilon \)RMSD values compared to structures that were simulated with the same ion conditions (Supplementary Figure S2). However, our observation that Mg\(^{2+}\) results in more stable simulations still holds. The comparison between Na\(^{+}\) and K\(^{+}\) as a counterion (Supplementary Figures S2 versus S3 and Supplementary Figures S8–S11) suggest potassium stabilizes the structures somewhat more than does sodium.

Both Na\(^{+}\) and Mg\(^{2+}\) ions bind sequence specific and also to specific binding sites (Figure 4). In both helical and complex folded structures certain RNA atoms are preferred. In complex folded structures atoms are available for binding that are not sterically accessible for ions in helical RNA structures. For example, one of the oxygen atom in uracil (U-O2) is hidden in the minor groove of a helical RNA with classical Watson-Crick base pair interactions. In complex folded structures we find this atom to be more accessible to ions (Figure 2), consistent with findings reported by Kirmizialtin et al. (18). We think it is appropriate to distinguish between adenine and guanine N7 atoms unlike what was done in previous studies (15,18). Doing so reveals that more ions are close to the guanine N7 atom than can be explained based just on accessibility and indeed the distributions are quantitatively different for both atoms (Figure 2).

At low ion concentrations a larger fraction of the Na\(^{+}\) and Mg\(^{2+}\) ions are in direct contacts with the RNA in our simulations than at higher concentrations (2). When, however, both Na\(^{+}\) and Mg\(^{2+}\) ions are present, more Mg\(^{2+}\) ions are closer to the RNA (distance less than 10 Å) than Na\(^{+}\). This is in agreement with the ‘ion atmosphere’ as described by Lipfert et al. (37). It seems therefore that the overall salt concentration should be factored in when considering the properties of the ‘ion atmosphere’.

Zheng et al. (13) investigated Mg\(^{2+}\) ion binding sites experimentally, in particular the difference between first and second shell binding frequencies. Since we only observe two direct contacts for Mg\(^{2+}\) ions in our simulations we cannot compare our simulations with the first shell contact frequencies derived in that work (13). The main reason for this is that it is very difficult to replace the hydration shell around Mg\(^{2+}\) by direct contacts during explicit MD simulations (66). Although refined Mg\(^{2+}\) ion parameters (64) were used, the activation energy remains slightly higher and the ion–water exchange rate faster than experimental values (66,85).

When we compare our Gibbs activation energies for second shell dissociation/binding of Mg\(^{2+}\) ions (Figure 3) to the experimental frequencies reported in (13) we see a preference for the same RNA atoms. The calculations fit the results by Zheng et al. (13) remarkably well. The RNA atoms with the highest experimental frequencies are (starting from the highest): G-O6, G-N7, O2P, U-O4, A-N7, O1P and A-N6 (13). For helical structures the RNA atoms with highest $\Delta G^\dagger$ are for system Mg\(^{2+}\)Cl (starting from the highest): G-O6, G-N7, C-N4, U-O4, A-N6, A-N7, C-N4, O2P, U-O2 and O1P. For helical structures the RNA atoms with highest free energies of activation $\Delta G^\ddagger$ are for system Mg\(^{2+}\)Cl (starting from the highest): G-O6, G-N7, C-N4, U-O4, A-N6, A-N7, C-N4, O2P, C-O2 and O1P. Although the activation energy is higher for O2P than O1P it seems to be underestimated in all simulations compared to the energies calculated for other RNA atoms. The main difference between helical and complex folded structures is that in helical ones the activation energy is higher for C-N4. When we compare the activation energy of the Mg\(^{2+}\) ion directly in contact with O1P (30.0 kJ/mol) to experimentally predicted activation energies $\Delta G^\ddagger$ between Mg\(^{2+}\) ions and DNA (53.1–55.7 kJ/mol) (85) it is quantitatively underestimated.

After the equilibration phase we could reproduce all experimentally predicted ion binding sites with good accuracy (Table 2, Figure 4). Especially when Na\(^{+}\) or Mg\(^{2+}\) ions are present (system Na\(^{+}\)PS and Mg\(^{2+}\)PS) without any additional salt concentration the binding sites are reproduced well using the MobyWat (80,81) analysis (Supplementary Table S3). The reason for this is likely that in simulations at low ionic strength the ions are found in close proximity to the RNA. We find, however, that the occupancy of the experimental ion binding sites calculated from the simulation is not reproduced with the same accuracy as the positions. A similar study focused on ion-binding to helical DNA was able to reproduce experimental ion-counts quantitatively (86), possibly because of improved cation force field parameters (67).

An interesting study by Lemkul et al. (87) applied grand-canonical Monte Carlo-MD (GCMC-MD) in order to predict ion-binding for four different RNA molecules. Although this approach most likely is more suited to find first shell binding locations than the MD approach we used, the use of pure MD allows to deduce time-dependent properties such as $\Delta G^\dagger$ for contact breaking (Figure3). A combination of the two techniques, prediction using GCMC-MD followed by regular MD would therefore yield a more complete picture of binding thermodynamics and kinetics. Nevertheless it seems the quality of binding site predictions is similar in both methods. Ion binding site prediction is inherently difficult for these systems with long exchange times.
It is likely as well that ion binding sites are missed by any structural analysis since ion-binding and conformational flexibility are interdependent. In fact, it is remarkable that Mg$^{2+}$ ions are predicted so close to experimental binding sites in normal simulation, while they maintain their hydration shell.

In comparison to previous studies our dataset contains a large number (24) of structures yielding rigorous results. Binding site positions and kinetics can be studied, and the relative influence of different ions studied. Based on our results (e.g. Figure 1) there is no justification for using Na$^+$ ions rather than Mg$^{2+}$ ions in RNA simulations, unless, as in this work, the purpose of the study is to investigate the difference in RNA properties due to the ‘ion atmospheres’ (37). Further improvement of force fields for RNA, water and ions remain needed to describe the complex energy landscape formed by these flexible biomolecules.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR online.

**FUNDING**

Swedish research council [2013-5947 to D.S.]; eSSENCE e-Science collaboration [to N.M.F]; Coordination for the Improvement of Higher Education Personnel (CAPES) [to M.D.P.]; Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing [SNIC2016-34-44]. Funding for open access charge: Vetenskapsrådet; [2013-5947]; faculty funding.

**Conflict of interest statement.** None declared.

**REFERENCES**


