Magnesium-binding architectures in RNA crystal structures: validation, binding preferences, classification and motif detection

Heping Zheng1,2,3,4,†, Ivan G. Shabalin1,2,3,4,5,†, Katarzyna B. Handing1,3,4,6, Janusz M. Bujnicki6,7 and Wladek Minor1,2,3,4,5,*

1Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908-0736, USA, 2Center for Structural Genomics of Infectious Diseases (CSGID) Consortium, USA, 3Midwest Center for Structural Genomics (MCSG) Consortium, USA, 4New York Structural Genomics Research Consortium (NYSGRC), USA, 5Enzyme Function Initiative (EFI), USA, 6Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw 02-109, Poland and 7Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland

Received February 20, 2014; Revised February 28, 2015; Accepted March 04, 2015

ABSTRACT

The ubiquitous presence of magnesium ions in RNA has long been recognized as a key factor governing RNA folding, and is crucial for many diverse functions of RNA molecules. In this work, Mg2+-binding architectures in RNA were systematically studied using a database of RNA crystal structures from the Protein Data Bank (PDB). Due to the abundance of poorly modeled or incorrectly identified Mg2+ ions, the set of all sites was comprehensively validated and filtered to identify a benchmark dataset of 15334 ‘reliable’ RNA-bound Mg2+ sites. The normalized frequencies by which specific RNA atoms coordinate Mg2+ were derived for both the inner and outer coordination spheres. A hierarchical classification system of Mg2+ sites in RNA structures was designed and applied to the benchmark dataset, yielding a set of 41 types of inner-sphere and 95 types of outer-sphere coordinating patterns. This classification system has also been applied to describe six previously reported Mg2+-binding motifs and detect them in new RNA structures. Investigation of the most populous site types resulted in the identification of seven novel Mg2+-binding motifs, and all RNA structures in the PDB were screened for the presence of these motifs.

INTRODUCTION

Metal ions are indispensable for proper RNA folding, stability and function in various biological processes (1). The positive charge of metal cations is needed to compensate for the negative charge of RNA’s highly acidic phosphate backbone, permitting RNA to form and retain compact and specific three-dimensional structures (2). The resulting structural complexity and wide repertoire of structural arrangements allows RNA to effectively perform a multitude of key cellular functions. In addition to their ubiquitous role as counter ions, metal ions are also crucial for some RNA molecules to recognize binding partners (3,4). In some ribozymes, metal ions have been found to directly mediate catalytic processes (5).

Mg2+ is generally accepted as the most important ion for RNA stabilization (1,6) and is the most frequently identified metal in RNA structures. Magnesium ions are nearly ubiquitous in RNA structures and many different types of coordination architectures have been observed for Mg2+ in RNA. A comprehensive survey of Mg2+ binding sites in RNA should be particularly useful for the prediction and annotation of RNA structure, function and the recognition of binding partners. Recent advancements in macromolecule crystallography have led to the determination of many structurally diverse metal-containing RNA crystal structures, offering a unique opportunity for such a survey of Mg2+ binding sites.

Most previous studies of Mg2+-binding architectures in RNA were not performed on a variety of crystal structures of different RNA families but were limited to the analysis of a single structure (2,7–12). However, two databases specialized for the investigation of metal ions in multiple RNA structures are available: MeRNA (13) and MINAS (14). MeRNA focuses on eight previously reported metal-binding motifs and is based on 389 structures deposited in the Protein Data Bank (PDB) (15) before February 2007.
MINAS offers a multitude of search functions for metal ligands defined by element, functional group or residue through all RNA structures in the PDB. Nevertheless, neither MeRNA nor MINAS offers a readily interpretable, systematic classification of Mg$^{2+}$ in RNA structures.

A few classification schemes of metal ion binding sites in RNA crystal structures have been proposed (10,11). Based on the analysis of one large ribosomal subunit, Klein et al. provided a simplified model to classify magnesium binding sites into six types based on the number and geometric isomerism of non-water ligands in the inner sphere (10). Lippert et al. evaluated possible inner-sphere interactions by phosphate group, sugar entity, nucleobase and their theoretical combinations and generalized more than 50 binding patterns of a metal ion by a single nucleotide (11). Unfortunately, outer-sphere ligands were not described systematically in either of these studies. Yet in many cases the inner sphere around a magnesium ion is composed of only (or mostly) water molecules, and the specific structural features of these sites can only be differentiated by outer-sphere interactions with RNA. Additionally, no comprehensive analysis has been performed on the likelihood of different nucleotides to bind Mg$^{2+}$ or on the relative abundance of each particular Mg$^{2+}$-binding architecture.

A potential pitfall in classifying and surveying Mg$^{2+}$ binding sites is the widespread misidentification of Mg$^{2+}$ in RNA crystal structures. The misinterpretation of small molecules and ions bound to macromolecules, including metals, has been spotted in many macromolecule structures (16,17) and Mg$^{2+}$ is not an exception (18). Mg$^{2+}$ has the same number of electrons as water and Na$^+$, neither of which can be distinguished from Mg$^{2+}$ by difference electron density maps alone (19). Moreover, the X-ray absorption K-edge (9.5 Å) of Mg$^{2+}$ is outside of the wavelength range producible by a typical synchrotron beamline, which makes it difficult to localize and differentiate it from water or Na$^+$ by anomalous scattering. A significant number of incorrectly identified metal sites can impose a strong bias on metal binding analysis, especially for large-scale studies where it is infeasible to manually analyze and verify each site individually. Therefore it is critical to retrieve only trustworthy sites (i.e. sites in agreement with experimental data and known bioinorganic chemistry) for analysis.

In the present study, a systematic analysis of Mg$^{2+}$ binding by RNA was performed with the following five objectives: (i) validate Mg$^{2+}$ binding sites in RNA crystal structures deposited to the PDB and discard poorly modeled or misidentified sites; (ii) statistically analyze preferences of nucleotides and their individual atoms for Mg$^{2+}$ coordination; (iii) craft and apply a hierarchical Mg$^{2+}$ site classification system which takes into account both inner- and outer-sphere ligands; (iv) employ the classification system for describing and detecting Mg$^{2+}$-binding motifs in RNA structures and (v) discover new Mg$^{2+}$-binding motifs within populous site types.

**Dataset used for analysis**

The initial screening set comprises all structures in the PDB (15) deposited on or before 30 September 2014 that satisfied three criteria: (i) determined by X-ray crystallography, (ii) contains at least three common ribonucleotide residues (A, G, C or U) covalently linked by phosphodiester bonds and (iii) contains at least one modeled Mg$^{2+}$ ion. This dataset of all Mg$^{2+}$ sites is herein referred to as the ‘full dataset.’ Mg$^{2+}$ sites were analyzed using the NEIGHBORHOOD database (20), which takes into account crystallographic symmetry and stores information on all identified atoms and residues together with their interactions with neighboring atoms and residues. Structures containing at least one ribosomal subunit were assigned to the ‘ribosome’ subset, while all other structures were placed in the ‘non-ribosome’ subset.

**Types of coordinating atoms from ribonucleotides**

The atoms from common ribonucleotides that can potentially coordinate Mg$^{2+}$ were classified into four different types: (i) O$_{\text{ph}}$, phosphate oxygen (OP1/OP2); (ii) O$_{\text{n}}$, oxygens in ribose (O2′/O4′) or oxygens bridging phosphate and ribose (O3′/O5′); (iii) O$_{\text{n}}$, nucleobase oxygen and (iv) N$_{\text{n}}$, nucleobase nitrogen.

**Definition of inner-sphere ligand atoms and coordination number**

Only oxygen and nitrogen were considered as potential inner-sphere ligand atoms. The search for inner-sphere ligand atoms was performed in two steps to optimally account for the potentially large metal-ligand distance deviations in RNA structures determined at medium to low resolution. The ideal distances ($d_{\text{ideal}}$) for Mg$^{2+}$–O (2.08 Å) and Mg$^{2+}$–N (2.20 Å) bonds were defined as the mean distances observed in the Cambridge Structural Database (21). In the first step, all oxygen and nitrogen atoms with a distance $d$ to a magnesium ion where $d \leq d_{\text{ideal}} + 0.5$ Å were identified as inner-sphere ligand atoms.

The second step, which was performed only if the number of ligand atoms found in the first step was less than six, included ligands with distances up to 0.5 Å longer ($d_{\text{ideal}} + 0.5$ Å $< d \leq d_{\text{ideal}} + 1.0$ Å). This second step identified additional nearby oxygen and nitrogen atoms which could potentially complete the octahedral geometry of the inner sphere. Three additional rules were applied in this second step to preclude chemically unfavorable interactions: (i) a second oxygen from the same phosphate group was not accepted since a phosphate group cannot form a bidentate interaction with Mg$^{2+}$ (22); (ii) the only allowed coordinating nitrogen atoms from a nucleobase were endocyclic nitrogen with a lone electron pair in the plane of the aromatic ring ($N = \text{-}$) and (iii) ligand atoms were accepted only if they have a planar ligand-Mg-ligand angle greater than 50° with all previously found ligands. The coordination number (CN) of a magnesium ion was defined as the number of inner-sphere ligand atoms identified by the procedure outlined above.

**MATERIALS AND METHODS**

Terminology, abbreviations and definitions used in this manuscript are provided as Supplementary Lexicon 1.
Definition of outer-sphere atoms

Outer-sphere coordinating atoms were identified based on the presence of hydrogen bonds to any inner-sphere water molecules. Hydrogen bonds were identified by the Probe program (23) from the MolProbity suite (24). For crystallographic symmetry related interactions, hydrogen bonds were identified by the CONTACT program from the CCP4 suite (25).

Outer-sphere moieties

For annotation of outer-sphere interactions, each individual RNA moiety (phosphate, ribose or nucleobase) was counted only once, and only if it did not form inner-sphere interactions with a given Mg$^{2+}$ ion. If two or more moieties contribute hydrogen bonds to the same water molecule, these were still counted as separate moieties. O3’ and O5’ atoms were assigned to the ribose moiety; except when the connected phosphate moiety contributes to the inner or outer sphere of the same magnesium ion, the O3’ and O5’ atoms were considered to be part of the phosphate group. Outer-sphere moieties were labeled as Pout (phosphate), Rout (ribose) and Bout (nucleobase).

Validation parameters

Three customized parameters $Q_i$, $Q_o$ and $Q_e$ were used to quantitatively identify the ‘quality’ of Mg$^{2+}$ sites (in addition to other criteria). The value of each parameter has a maximum value of 1, with a lower value indicating poorer reliability and a higher value indicating better reliability.

$Q_i$, measures the agreement of the bond valence summation ($\sum V_i$) of the inner-sphere interactions with the oxidation state of magnesium (+2) as defined by $Q_i = 1 - \frac{\sum V_i - 2}{\sum V_i}$, in which $V_i$ is the bond valence derived from Mg$^{2+}$-ligand distance of coordinating ligand $i$ (26). A table of the relationships between distance and valence is provided for both Mg$^{2+}$-O and Mg$^{2+}$-N separately (Supplementary Table S1).

$Q_o$, measures the geometrical symmetry of the ligands distribution around the Mg$^{2+}$ required for octahedral geometry by calculating the amplitude of the vector sum of the bond valence vectors $v_i$ (27). The sum should be of magnitude 0 for a perfectly symmetrical set of ligands. $Q_o$ is defined as $Q_o = \frac{\text{max}_i |v_i|}{\sum |v_i|}$.

$Q_e$, measures the agreement of the isotropic atomic displacement parameter (B-factor) of the Mg$^{2+}$ ($B_m$) and its occupancy ($O_m$) compared to those of all atoms in its environment ($B_o$, $O_o$). The environmental B-factor ($B_o$) and occupancy ($O_o$) were calculated as the valence weighted sum of those parameters for all non-hydrogen atoms within 4 Å of the Mg$^{2+}$ divided by the overall valence ($B_o = \frac{\sum V_i B_i}{\sum V_i}$, $O_o = \frac{\sum V_i O_i}{\sum V_i}$). In the majority of cases, both the Mg$^{2+}$ and the atoms in its environment have full occupancy ($O_m = O_o = 1$) and $Q_e$ is defined as the smaller B-factor $\min(B_m, B_o)$ divided by the larger B-factor $\max(B_m, B_o)$. When partial occupancy was encountered for a magnesium ion or its environment, $Q_e$ was weighted by the occupancy using the formula:

$$\min(O_m, O_o) \times \min\left(\frac{B_m}{O_m}, \frac{B_o}{O_o}\right).$$

Normalized interaction frequencies of coordinating atoms and moieties

Normalized interaction frequencies of atoms ($F_{\text{atom}}$) were calculated in a similar manner as has been reported previously (20). For an atom of type $X$, $F_{\text{atom}}(X)$ was calculated by the formula

$$F_{\text{atom}}(X) = \frac{p(X) - p(Mg-X)}{p(Mg-X)}.$$

$p(X)$ is defined as the percentage of atoms of type $X$ out of all coordinating atoms in the full dataset. Hence, if $F_{\text{atom}}(X) > 1$ and thus $p(Mg-X) < p(X)$, atoms of type $X$ interact with Mg$^{2+}$ with relatively high frequency. Conversely, if $F_{\text{atom}}(X) < 1$ and thus $p(Mg-X) > p(X)$, Mg-X interactions occur with relatively low frequency.

Geometric isomerism

Mg$^{2+}$ sites coordinated more than one phosphate group (Oph) in the inner sphere were differentiated by geometrical arrangement as either cis- or trans- isoforms (two or four Oph) or fac- or mer- isoforms (three Oph). The trans- or mer-isoforms were defined by the presence of two Oph atoms opposite to each other in a trans- conformation, defined as a ligand-Mg-ligand angle larger than 135°. In the absence of such pair of opposite Oph ligands, the site was defined as cis- (two Oph) or fac- (three Oph).

RESULTS

Prevalence of poorly coordinated Mg$^{2+}$ sites in RNA crystal structures from the PDB

The full dataset contains 99260 Mg$^{2+}$ sites from 1036 structures, consisting of 95406 sites from 494 ribosome structures and 3854 sites from 542 non-ribosome structures. Most sites in the full dataset are from low-resolution structures; only 2.5% of the sites (2508 sites) are from structures determined at a resolution better than 2.4 Å (Figure 1A). All structures are of resolution better than 4.5 Å, and most sites found in structures of resolution worse than 2.1 Å come from ribosome (Supplementary Figure S1). Most ribosomal structures in the full dataset, including both large subunit (~3000 nucleotides) and small subunit (~1500 nucleotides) structures, contain 100–1000 Mg$^{2+}$, though a few contain five or fewer Mg$^{2+}$ sites. Notably, 42% of the ribosome structures in the PDB do not contain a single Mg$^{2+}$ site. The majority of the non-ribosome structures contain fewer than 10 Mg$^{2+}$ sites each (Figure 2). This trend indicates that the number of modeled Mg$^{2+}$ sites located by X-ray crystallography is often insufficient to neutralize the negative charge of RNA due to diffusely bound Mg$^{2+}$, presence of other cations, limited resolution of the crystal structure and/or difficulty in Mg$^{2+}$ identification.
More than half of the Mg$^{2+}$ sites in the full dataset exhibit a highly incomplete inner sphere with coordination number (CN) in the range of 0–3 (Figure 1B), even when a very generous distance cutoff (1 A above the ideal distance) is used in the search for inner-sphere atoms. Though sites in structures of higher resolution (≤ 2.0 Å) have a higher average CN, Mg$^{2+}$ sites with a CN of three or less are still commonly observed (Figure 1A). In the full dataset, Mg$^{2+}$ sites with a (relatively) complete inner sphere (CN = 4–6) compose around 30% of all sites in ribosome and around 50% of all sites in non-ribosome structures. A small number of sites (53, <0.1%) had CN > 6. Manual inspection of the sites with CN > 6 revealed modeling errors with severe clashes between the inner-sphere ligands and/or unlikely bidentate coordination by phosphate (22).

Mg$^{2+}$ sites from structures determined at 2.9–3.7 Å resolution show the most incomplete inner spheres. For resolutions better than 2.9 Å, the completeness of coordination appears to be correlated with resolution (higher mean CN at better resolutions and lower mean CN at worse resolutions). This trend reverses at resolution worse than 3.7 Å, and well-coordinated sites become more abundant in worse resolution structures, likely due to the common practice of using a restrained hexahydrated Mg$^{2+}$ during refinement instead of a single Mg$^{2+}$ at lower resolutions with very poor electron density maps (28,29).

**Benchmark dataset**

The benchmark dataset, which excludes highly questionable Mg$^{2+}$ sites, was used for all further analyses. Mg$^{2+}$ sites were included in the benchmark dataset only if all of the following criteria were met: (i) the site is RNA-bound (through inner and/or outer sphere); (ii) the site CN = 4–6; (iii) all three validation parameters are higher or equal to the threshold values ($Q_v \geq 0.5$, $Q_s \geq 0.6$ and $Q_e \geq 0.5$) as determined in the preliminary research described in the supplementary data (Supplementary Text 1, Supplementary Figures S2/S3); (iv) sites should not be coordinated by nucleobase nitrogen other than an endocyclic nitrogen with a lone electron pair in the plane of the aromatic ring ($-N=$) in the inner sphere.

The benchmark dataset consists of 15334 Mg$^{2+}$ binding sites (489 structures), which constitutes only 15% of the full dataset. The size of the benchmark dataset is significantly smaller than the size of the full dataset, mostly due to the abundance of sites with a highly incomplete Mg$^{2+}$ inner
Frequencies of nucleotide atoms to coordinate magnesium ions

The frequencies of RNA atoms to coordinate Mg$^{2+}$ were evaluated using the normalized interaction frequency ($F_{\text{atom}}$), which measures the frequency a particular kind of atom served as a ligand normalized by the frequency of that atom type in RNA structures overall. The most commonly observed inner-sphere ligand atom type is oxygen from phosphate ($O_{\text{ph}}$) (Figure 3, Supplementary Table S2), which are the negatively charged RNA atoms that are often compensated for by the positive charge of Mg$^{2+}$. The only other types of nucleotide oxygen atoms with inner-sphere $F_{\text{atom}} \geq 1$ are two types of keto-oxygens (U-O4, G-O6) from nucleobase (O₅). The $F_{\text{atom}}$ value of the most frequently observed O₅ ligand atom (U-O4) is just two times smaller than the value for O₅ atoms, even though U-O4 lacks the ability to significantly compensate for the positive charge of Mg$^{2+}$. Not all O₅ atoms exhibited similar normalized frequency to serve as ligands for Mg$^{2+}$ binding. The $F_{\text{atom}}$ values for O₅ atoms adjacent to the ribose bond (C-O2 and U-O2) were up to 290 times lower than that for O₅ atoms located opposite to the sugar edge (G-O6 and U-O4) (Figure 3). The reason is that in most binding site configurations, the ribose blocks the adjacent oxygen atom (C-O2 or U-O2) from binding Mg$^{2+}$ by steric clashes between Mg$^{2+}$ inner-sphere waters and the ribose. Unlike O₅ and some O₅ atoms, ribose oxygen atoms (O₅) are rarely found in Mg$^{2+}$ inner sphere.

Besides oxygen atoms, nitrogen atoms from nucleobases (N₅) are the only other RNA atoms to serve as inner-sphere ligands, but generally they were observed in the inner sphere less frequently (Figure 3). Out of all six types of N₅ atoms with a lone electron pair in the plane of the aromatic ring (−N=), only two (G-N7 and A-N7) have an inner-sphere $F_{\text{atom}}$ value higher than one. All other −N= atoms were almost never observed to coordinate Mg$^{2+}$ in the inner sphere.

Though outer-sphere hydrogen bonds are chemically different from the inner-sphere coordination bonds, the most frequent inner-sphere atoms (O₅, U-O4, G-O6, G-N7 and A-N7) are also among the most frequent outer-sphere atoms of Mg$^{2+}$ (Figure 3). However, the order of the relative frequencies differs and O₅ atoms are no longer the most frequently observed. The most frequent outer-sphere atoms are G-O6 and G-N7, followed by O₅, U-O4 and A-N7. Similar to the inner sphere, the frequencies of outer-sphere interactions for O₅ atoms located opposite to the sugar edge (G-O6 and U-O4) are much greater than those for O₅ atoms adjacent to the ribose bond (C-O2 and U-O2). In spite of the similar high frequencies of certain atoms to coordinate Mg$^{2+}$ both through the outer and inner sphere, several types of atoms rarely found in the inner sphere were found to form outer-sphere hydrogen bonds quite frequently (Figure 3);

![Figure 3](https://academic.oup.com/nar/article-abstract/43/7/3793/2414646)
most notably the O3` and O5` atoms from ribose and A-N6 and C-N4 from the exocyclic amino groups of nucleobases (–NH2). A more extensive discussion of the differences in atomic Mg2+ coordination frequencies is presented in Supplementary Text 2.

Guanine is the most frequently observed nucleobase to coordinate Mg2+ in the inner sphere due to the combined effect of two atoms that have high F_atom values (G-O6 and G-N7). The inner-sphere F_atom value for U-O4 is higher than that for A-N7, which makes uracil the second most frequently observed nucleobase. As a net result, the frequency for nucleobases to coordinate Mg2+ in the inner sphere is (in descending order) G > U > A > C. In the outer sphere, the guanine moiety is also the most frequently observed nucleobase, followed by adenine and uracil, exhibiting a slightly different trend (G > A > U > C) than is observed for the inner sphere.

Overview of Mg2+ site classification

All Mg2+ sites in the benchmark dataset were divided according to their binding environment into four mutually exclusive classes in the following order (Table 1): (i) sites with additional metal ion(s) within 4 Å from Mg2+; (ii) sites with non-RNA, non-water atoms in the inner sphere; (iii) sites with only RNA atoms as non-water ligands in the inner sphere (RNA-inner) and (iv) sites with only water molecules in the inner sphere and at least one RNA moiety in the outer sphere (RNA-outs).

Classes (i) and (ii) comprise only a small fraction of sites in the benchmark dataset (2.7% and 1.7%, respectively; Table 1). The RNA-inner and RNA-outer classes represent the majority (96%) of sites in the benchmark dataset (Table 1). Therefore only these two classes were further classified into 136 types based on the structural arrangements of their binding environments (Supplementary Tables S3/S4) in this manuscript. The detailed classification of the RNA-inner and RNA-outer classes is accessible at http://www.csgid.org/metalnas/ (Figure 5).

Table 1. Numbers of Mg2+ sites included in the benchmark dataset for the four site classes described in the text

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total</th>
<th>Metal atom within 4 Å of Mg2⁺</th>
<th>Protein or small molecule in Mg2⁺ inner sphere</th>
<th>RNA-inner</th>
<th>RNA-outer</th>
</tr>
</thead>
<tbody>
<tr>
<td>All data</td>
<td>15334</td>
<td>412 [2.7%]</td>
<td>268 [1.7%]</td>
<td>9628 [63%]</td>
<td>5026 [33%]</td>
</tr>
<tr>
<td>Ribosome</td>
<td>14682</td>
<td>373 [2.5%]</td>
<td>195 [1.3%]</td>
<td>9382 [64%]</td>
<td>4732 [32%]</td>
</tr>
<tr>
<td>Non-ribosome</td>
<td>652</td>
<td>39 [6.0%]</td>
<td>73 [11%]</td>
<td>246 [38%]</td>
<td>294 [45%]</td>
</tr>
</tbody>
</table>

The percentage of sites in each site class is shown in brackets.
RNA-inner sites classification

The RNA-inner class is the most populous class, representing 63% of the benchmark dataset (Table 1). Of the four types of inner-sphere atoms ($O_{ph}$, $O_r$, $O_b$, and $N_b$), $O_{ph}$ is most abundant (Figure 3) and is likely to contribute most to the energy of Mg$^{2+}$ binding via its distinct ability to compensate for the ion’s charge. Hence, the number of inner-sphere $O_{ph}$ atoms was the first criterion chosen to classify RNA-inner site types (Supplementary Table S3), while the total number of other inner-sphere atoms ($O_r, O_b, N_b$) was chosen to be the second criterion. The maximum number...
of RNA atoms serving as inner-sphere ligands is 4 per site (Figure 4C), so combining the first and second criteria resulted in 14 possible subclasses (Supplementary Table S3). These 14 combinations were further divided into 41 types of Mg\(^{2+}\) binding sites based on specific types of \(O_I|O_B|N_b\) atoms and the geometrical isomerism of \(O_{ph}\) atoms (cis/-trans/-mer/-fac-) if more than one \(O_{ph}\) is present. For the most populated branch (\(#O_{ph} = 1\), \(#O_I|O_B|N_b = 0\), the number of outer-sphere phosphates was also taken as an additional criterion to subdivide sites.

The general trend seen in the inner-sphere \(F_{atom}\) values for nucleobase and ribose atoms (\(O_{ph}|N_b>_{Ribose}, Figure 3B\)) is also observed in the population of different site types within the same subclass (i.e. with the same number of inner-sphere \(O_{ph}\) and \(O_I|O_B|N_b\) atoms). The only exception to this trend is the \(2N_b\) type in the \(#O_{ph} = 0\), \(#O_I|O_B|N_b\) = 2 subclass, which contains a disproportionately high population (158 sites). All other types with more than one \(O_I|O_B|N_b\) inner-sphere ligand, regardless of subclass, contain 28 sites or less. Both fac- and mer-conformations have similar abundances for sites with \(#O_{ph} = 3,\) while the cis-conformation is more frequently observed than the trans-conformation for all \(#O_{ph} = 2\) subclasses. Eighty-six percent of the sites in the subclass \(#O_{ph} = 2,\) \(#(O_I|O_B|N_b) = 0\) are in the cis-conformation. The cis-2O_{ph} conformation is even more predominant (283 out of 286 sites) in the sites in which non-\(O_{ph}\) inner-sphere ligands are also present (subclasses \(#O_{ph} = 2,\) \(#O_I|O_B|N_b\) = 1, 2).

**RNA-outer site classification**

The RNA-outer class represents a significant fraction (33\%) of all sites in the benchmark dataset (Table 1). The outer-sphere composition was used to subdivide the class due to the absence of inner-sphere RNA atoms (Supplementary Table S4). Instead of employing individual interactions as was done for the inner sphere, individual outer-sphere moieties were used to subdivide RNA-outer sites. We believe that the number of moieties represents the uniqueness of a given structural arrangement of RNA, while the actual number of interactions (hydrogen bonds) of the moiety with inner-sphere water molecules just slightly contributes to the energy of binding and may easily vary between very similar sites. Similar to RNA-inner site classification, the number of outer-sphere phosphate moieties \(#P_{out}\) was chosen as the first criterion and the total number of ribose and nucleobase moieties \(#(R_{out}/B_{out})\) was used as the second criterion to subdivide the RNA-outer class, resulting in 39 subclasses. The specific combinations of ribose \(#R_{out}\) and nucleobase \(#B_{out}\) moieties in the outer sphere produced a list of 95 types of RNA-outer sites (Supplementary Table S4).

The most populous RNA-outer site type is \(2B_{out}\) with only two nucleobase moieties in its outer sphere (555 sites). This site type is often found in RNA helices, as base stacking results in very close positions of two consecutive bases coordinating one Mg\(^{2+}\). Site types containing only phosphate moieties are less populous yet still abundant, represented by \(P_{out}(94 sites), 2P_{out}(191 sites), 3P_{out}(207 sites)\) and \(4P_{out}(174 sites)\). The rarity of the ribose moiety in the outer sphere (Figure 4B) was reflected in the low populations of \(R_{out}\)-containing types, not exceeding 60 instances per type.

**Detection of previously reported Mg\(^{2+}\)-binding motifs**

In the current study, we sought to identify and analyze ‘validated motifs,’ which we considered to be a specific structural arrangement provided by RNA for Mg\(^{2+}\) binding, which is found in structures of multiple RNA molecules. Most Mg\(^{2+}\)-binding motifs reported previously, including some of those annotated in MeRNA (13), were based on the analysis of a single RNA structure (2.7-12). Therefore, it remains to be verified if these reported architectures could be found in other structures or whether they were a unique feature in a given structure.

The systematic site classification presented in this paper was used to describe six validated Mg\(^{2+}\)-binding motifs previously reported in the literature (Table 2, Supplementary Figure S5). A site type was defined for each literature-derived motif, and a few additional criteria were applied to specify the motif precisely within the site types. For example, the two \(O_p\) ligands that define the ‘magnesium clamp’ (7,30) must be from different chains or from nucleotides separated by more than seven residues from one another (Table 2). Those additional criteria were implemented in the form of database queries which retrieved all the sites from the benchmark dataset that fulfill the requirements. The validity of our motif definitions was further verified by the capability of these customized queries to locate the Mg\(^{2+}\)-binding sites reported in the original literature. Certain previously reported motifs involving specific RNA structures with a given type of Mg\(^{2+}\) binding site were rarely observed and therefore were not considered a ‘validated motif’ herein, such as G-A pair (5,9), sheared G-A pairs, A-rich bulge and the three helix junction (2).

All six validated motifs are reasonably abundant in the benchmark set, with the number of instances ranging between 34 for the Triple-G motif (2) and 1030 for the magnesium clamp (7,30). The 10-member-ring motif (12), with two inner-sphere \(O_{ph}\) atoms from consecutive nucleotides forming the core of the motif, is found in four different variants. The G-phosphate motif (10) and metal ion zipper (8) were found only in rRNA.

**Identification of novel Mg\(^{2+}\)-binding motifs**

The classification clusters Mg\(^{2+}\) sites with similar RNA interaction patterns, which allows a search for novel validated motifs. Selected sites from some of the more populated site types were investigated visually for the presence of a recurring specific pattern in a variety of structures. For each found pattern a respective database query was created to screen the benchmark dataset and retrieve all sites with that specific pattern. This motif-discovery approach resulted in the detection of seven validated motifs, five in the RNA-inner class and two in the RNA-outer class (Table 3).

The ‘Y-clamp’ motif (Figure 6A), which is observed in both ribosomal subunits, riboswitch and ribozyme, stabilizes RNA structures by anchoring two different strands or two distant parts of the same strand together in a similar manner as the magnesium clamp (7). The letter ‘Y’ in
Table 2. Three RNA-inner and three RNA-outer validated Mg²⁺-binding motifs previously reported in the literature (I–VI)

<table>
<thead>
<tr>
<th>ID</th>
<th>Motif name</th>
<th>Reference</th>
<th>Systematic classification</th>
<th>Detailed features of the motif</th>
<th>Number of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Magnesium Clamp</td>
<td>Ennifar et al. (1999) (7) Petrov et al. (2011) (30)</td>
<td>RNA-inner</td>
<td>The two Oph atoms are from distant phosphates</td>
<td>823 (cis)207 (trans)</td>
</tr>
<tr>
<td>I</td>
<td>10-member ring</td>
<td>Hsiao et al. (2009) (12)</td>
<td>RNA-inner</td>
<td>The two Oph atoms are from sequentially consecutive phosphates</td>
<td>802</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The three Oph atoms are from sequentially consecutive phosphates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10-member ring with additional Oph atom coming from phosphate which is separated by one residue from either 10-member ring phosphates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>One Mg²⁺ coordinated by two unrelated 10-member rings</td>
<td>37</td>
</tr>
<tr>
<td>IV</td>
<td>G⋅G metal-binding site</td>
<td>Correll et al. (1997) (8)</td>
<td>RNA-inner</td>
<td>Mg²⁺ is bound through the outer sphere by two sequentially consecutive guanines and by two phosphate moieties from residues n and n-1 where n is residue number of the lower guanine in the sequence</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>Triple G Motif</td>
<td>Tinoco &amp; Keift (1997) (2)</td>
<td>RNA-outer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>Metal ion zipper</td>
<td>Correll et al. (1997) (8)</td>
<td>RNA-outer</td>
<td>Mg²⁺ binds two distant phosphates through the outer sphere</td>
<td>97</td>
</tr>
</tbody>
</table>

The site type(s) and additional features used to define each motif are tabulated. The term ‘distant’ phosphates/residues is used to specify two phosphates or residues coming from different RNA chains or separated by more than seven residues in the same chain. The terms ‘upstream’/‘downstream’ residues are used to specify a residue lower/higher in the sequence of the same chain. Representative examples of each motif are depicted in Supplementary Figure S5.

Table 3. Seven newly discovered and validated Mg²⁺-binding motifs

<table>
<thead>
<tr>
<th>ID</th>
<th>Motif name</th>
<th>Representative site</th>
<th>Systematic classification</th>
<th>Detailed features of the motif</th>
<th>Number of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Y-clamp</td>
<td>Z75-B301</td>
<td>RNA-inner</td>
<td>10-member ring and one distant Oph atom in mer-conformation</td>
<td>279</td>
</tr>
<tr>
<td>B</td>
<td>U-phosphate 12-member ring</td>
<td>YIE-Z1116</td>
<td>RNA-inner</td>
<td>Oph atom is from uracil</td>
<td>199</td>
</tr>
<tr>
<td>C</td>
<td>Purine-N7 seat</td>
<td>AVY-A1566</td>
<td>RNA-inner</td>
<td>Mg²⁺ is bound through the outer sphere by two sequentially consecutive phosphates forming a ring with 12 non-hydrogen atoms</td>
<td>82</td>
</tr>
<tr>
<td>D</td>
<td>G-N7 macro-chelate I</td>
<td>V29-A3359</td>
<td>RNA-inner</td>
<td>Mg²⁺ is bound by two guanine or adenine N7 atoms in inner sphere and capped by 3 or 4 downstream phosphates through the outer sphere</td>
<td>141</td>
</tr>
<tr>
<td>E</td>
<td>G-N7 macro-chelate II</td>
<td>AW7-A1569</td>
<td>RNA-outer</td>
<td>Mg²⁺ is bound by guanine N7 atom in inner sphere and phosphate of the same residue in the outer sphere</td>
<td>60</td>
</tr>
<tr>
<td>F</td>
<td>10-member ring with purine-N7</td>
<td>V23-A3540</td>
<td>RNA-inner</td>
<td>Mg²⁺ is bound through the outer sphere by guanine N7 atom and phosphate of the same residue</td>
<td>18</td>
</tr>
</tbody>
</table>

The PDB accession codes with the residue number for the magnesium ion are shown for representative sites of each motif (as depicted in Figure 6). Site type and additional features used to define each motif are tabulated.
\textbf{Figure 6.} Newly discovered and validated Mg\textsuperscript{2+}-binding motifs. Magnesium ions and inner-sphere interactions are shown in green. Outer-sphere hydrogen bonds are shown as gray dashed lines. The nucleotides forming the motifs are labeled. The figure was prepared using PyMol. The representative sites defined by PDB codes, Mg\textsuperscript{2+} residue number, structure resolution, and validation Q-values (Q\textsubscript{v} / Q\textsubscript{s} / Q\textsubscript{e}) are as follows: (A) 2Z75-B301, 1.7 Å (0.8 / 0.9 / 0.9); (B) 2YIE-Z1116, 2.94 Å (0.7 / 0.8 / 1.0); (C) 2AVY-A1566, 3.46 Å (1.0 / 1.0 / 1.0); (D) 2QBA-B3321, 3.54 Å (1.0 / 0.9 / 0.9); (E) 3V29-A3359, 3.1 Å (0.8 / 0.9 / 0.8); (F) 2AW7-A1569, 3.46 Å (1.0 / 1.0 / 0.5); (G) 3V23-A3540, 3 Å (0.6 / 0.9 / 0.9).

the name of the motif was chosen to resemble the three-way configuration of inner-sphere O\textsubscript{P} atoms, while ‘clamp’ refers to the bridging capability of phosphates by Mg\textsuperscript{2+}. This unique feature for maintaining an RNA fold, which resembles the disulfide bridge linkage in proteins, is very common in RNA structures: 814 magnesium clamps and 238 Y-clamps were found. The ‘U-phosphate’ motif (Figure 6B) resembles the previously reported G-phosphate motif (10), save that uracil is substituted for guanine. Similar to the G-phosphate motif, this motif was found mostly in ribosome. The ‘12-member ring’ RNA-outter motif involves two outer-sphere phosphate moieties from consecutive residues in the RNA backbone (Figure 6C). This motif is similar to the ‘10-member ring’ RNA-inner motif, save that it has outer-sphere instead of inner-sphere interactions. Even though not further explored, multiple variations of the ‘12-member ring’ motifs are expected similarly to the multiple variations reported for 10-member ring motifs (12).

Four more validated motifs were found exclusively in ribosome. The ‘purine N7-seat’ motif contains a very characteristic coordination pattern formed by inner-sphere nucleobases (2N\textsubscript{b}) and outer-sphere phosphate moieties (Figure 6D). The high population of the 2N\textsubscript{b} type as observed in the benchmark dataset (Supplementary Table S3). The N7 atoms serving as ligands in motif are usually found in two guanine bases, but sometimes an adenine-guanine pair was found. Two single-nucleotide and guanine-specific ‘macrochelate’ motifs were identified with G-N7 serving as a ligand in either inner sphere or outer sphere, and ‘macrochelated’ with an outer-sphere phosphate moiety (Figure 6E and F). Similar ‘macrochelate’ patterns have been previously reported for structures of mononucleotides (11,31). A specific motif named ‘10-member ring with Purine-N7’ was found to have a 10-member ring together with an additional ligand formed by the N7 atom of a purine base, which is separated by one residue from either of the 10-member ring phosphates (Figure 6G).

\textbf{DISCUSSION}

Validation strategy

Due to the poor quality of many RNA crystal structures deposited in the PDB, validation is a critical step for any structural data mining study to have biological relevance. Unfortunately, the common practice of using resolution as the main selection criteria to define a ‘good’ dataset is infeasible for our analysis for two reasons. First, the high flexibility of RNA and large unit cell dimensions of RNA crystals means that many structures of RNA are of relatively poor resolution—the majority of magnesium ions in the PDB are found in structures of 2.5 Å resolution or worse (Figure 1A). Second, even in high-resolution structures, a substantial fraction of Mg\textsuperscript{2+} are still significantly undercoordinated (Figure 1A) and therefore questionable. To account for the significant errors in atomic positions present at low resolution (32), a unique two-step search of inner-sphere ligands was employed to conditionally allow a generous Mg\textsuperscript{2+}-ligand distance deviation of up to 1.0 Å for selected favorable ligands, yet limit the inclusion of interactions unlikely to be specific. Tailored for a whole range of resolutions, the inner-sphere definition used herein recognizes ligands more specifically than do the algorithms for other metal databases which use a simple distance cutoff (13,14,33–35).

Since the scope of this study was Mg\textsuperscript{2+} sites in RNA structures, which informed the selection of the benchmark dataset, only RNA-bound (through either inner or outer sphere) sites were accepted in order to exclude those Mg\textsuperscript{2+} sites which are bound only to protein in RNA–protein complexes or do not have any clear connection to RNA. The Mg\textsuperscript{2+} validation procedure was largely based on the similarity of each site to the rigid octahedral arrangement of inner-sphere ligands expected for Mg\textsuperscript{2+} and the characteristically short Mg\textsuperscript{2+}-ligand distances. Harnessing those intrinsic properties for Mg\textsuperscript{2+} identification required that all sites have a relatively complete coordination sphere; therefore only sites with CN = 6 were accepted. The three validation parameters used to select sites for the benchmark dataset evaluated Mg\textsuperscript{2+}-ligand distances (Q\textsubscript{s}), the symmetry of the inner-sphere ligands arrangement (Q\textsubscript{e}) and the agreement of Mg\textsuperscript{2+} B-factors with the surrounding atoms (Q\textsubscript{v}). Using the combination of multiple parameters as filtering criteria effectively removed the majority of poorly modeled Mg\textsuperscript{2+} sites. However, a few Mg\textsuperscript{2+} sites with chemically infeasible interactions were still not caught by the criteria, for example, those with amino nitrogen atoms in the inner sphere. To exclude those cases, we introduced an additional criterion for sites with nitrogen atoms from nucleobases in the
inner sphere: namely, accepting only those sites which have an endocyclic nitrogen with a lone electron pair in the plane of the aromatic ring (−N=) since only this type of nitrogen from nucleobase can feasibly coordinate Mg$^{2+}$ (Supplementary Text 2).

Some potential Mg$^{2+}$ binding sites in RNA are absent from the benchmark dataset because the electron density for Mg$^{2+}$ was not observed, or density was present but was incorrectly modeled as another metal ion or as a water molecule. Other sites might have had a true magnesium ion, but an incompletely modeled inner sphere resulted in some of the validation parameters being below the benchmark dataset thresholds. In the latter case, the inner sphere can often be completed by placing additional water molecules during crystallographic re-refinement of problematic structures. In this way, the benchmark dataset can be extended by revealing additional ‘trustworthy’ Mg$^{2+}$ sites, but would require manual inspection of each site and is beyond the scope of the current study. Nevertheless, our algorithms enabled extraction of many existing trustworthy Mg$^{2+}$ sites which can be used for further statistical and classification studies. For example, the benchmark dataset is particularly valuable to be used as a training dataset for tools to predict the positions of Mg$^{2+}$ ions in both experimental and theoretical models of RNA structures and/or increase the accuracy of Mg$^{2+}$ prediction. Thus far, we have implemented the benchmark dataset developed in this work as one of the alternative reference datasets in MetalionRNA predictor at http://metalionrna.genesilico.pl/ (36).

Crystallographic model-building artifacts

The reliability of a Mg$^{2+}$ binding site is highly dependent on modeling strategy and whether restraints were properly used during crystallographic refinement. The distance distribution of Mg$^{2+}$-water distances reveals the presence of two large sharp peaks which originate from strongly restrained sites at correct (2.08 Å) or incorrect (2.18 Å) Mg$^{2+}$-water distances used in some refinement programs by default (Supplementary Figure S4A). (We use ‘correct’ or ‘incorrect’ in the sense that the values do or do not agree with the mean distances observed in atomic resolution small molecular crystal structures.) Around 200 strongly restrained Mg$^{2+}$ sites at another incorrect (1.83 Å) Mg$^{2+}$-water distance were also identified (Supplementary Figure S4A) in a 40S ribosomal subunit (PDB codes 2XZM, 2XZN) (29). On the contrary, the absence of prominent peaks in the distributions of Mg$^{2+}$−O (non-water) and Mg$^{2+}$−N distances, even in sites that satisfy the validation criteria, suggests that those interactions are loosely restrained during crystallographic refinement (Supplementary Figure S4B and C).

The proper use of Mg$^{2+}$-ligand distance restraints is essential for correct modeling of a Mg$^{2+}$ inner sphere, especially of low-resolution structures. However, improperly strict or weak restraints may result in misinterpretation of experimental data, and thus in turn impair subsequent research. Given the prevalence of poorly modeled Mg$^{2+}$ sites in RNA structures and improper use of crystallographic restraints (either too strict or too loose), we propose that the crystallographic refinement software should include easy-to-use restraints to enforce correct CN and geometrical arrangement of ligands around metal ions.

Benchmark dataset redundancy

The main objective in the construction of the benchmark dataset was ensuring the reliability of the Mg$^{2+}$ binding sites. Redundant structures (i.e. having even 100% identical nucleotide sequences) were not excluded. It is beneficial to include sites that satisfy validation criteria from all structures in the analysis because different structures of the same macromolecule may carry different sets of reliable Mg$^{2+}$ sites due to differences in diffraction data quality, refinement strategies, crystallization conditions, bound ligands and/or macromolecule conformation. Moreover, it is essential to preserve as much variety of the different coordination patterns that may be present for otherwise equivalent Mg$^{2+}$ sites observed in different PDB deposits of homologous RNA molecules.

Even though a certain level of redundancy was observed in the benchmark dataset, all further analyses were carefully designed to control for the presence of redundancy. The statistics used for estimation of the preferences of atoms and nucleotides to bind Mg$^{2+}$ are minimally affected by dataset redundancy, because the frequencies of individual interactions are normalized by the frequency of atoms or nucleotides in the dataset. Dataset redundancy is beneficial for the classification because it ensures that the architecture of each reasonable Mg$^{2+}$ site was accounted for. The numbers of instances of each validated Mg$^{2+}$ motif do indeed contain redundant sites, but each motif was manually confirmed to represent a variety of sites in non-redundant structures.

Magnesium ion binding preferences

The $F_{atom}$ values produced by statistical analysis of the benchmark dataset indicate the frequency at which certain types of RNA atoms served as ligands for Mg$^{2+}$ in the inner or outer coordination sphere, and are normalized so that values for different atoms can be directly compared to one another. These values are consistent with the steric accessibility and chemical properties of each atom type (Supplementary Text 2), and we propose that these frequencies are reasonable estimates of the ‘propensity’ or ‘preference’ of these RNA atom types to coordinate Mg$^{2+}$. However, we cannot formally rule out the possibility of sampling bias (i.e. the binding sites in the benchmark set may not be wholly representative of Mg$^{2+}$ binding in RNA universally). For example, the analyzed set was necessarily limited to RNA molecules which form diffraction-quality crystals and, as mentioned above, may exclude sites that were incompletely or incorrectly modeled. The values of $F_{atom}$ (and the preferences they imply) can be used as prior knowledge for predicting ‘probable’ versus ‘improbable’ Mg$^{2+}$ sites in both computational modeling and crystallographic refinement.

Even if the main role of Mg$^{2+}$ is believed to be neutralization of the negative charge of phosphate moieties in RNA (2), nucleobase moieties were shown to be relatively abundant in the inner sphere and are almost as abundant as phosphate moieties in Mg$^{2+}$ outer spheres (Figure 4).
Therefore the coordination of Mg$^{2+}$ by nucleobases should be considered as a significant factor in the stabilization of RNA structure. Our data suggests (Figure 3B) that guanosine nucleotide has a more pronounced effect on RNA structure stabilization than any other nucleotide due to its predominance in Mg$^{2+}$ binding, which is supplemented by its ability to form a greater number of hydrogen bonds in base pairing.

**Mg$^{2+}$ site classification and motif description**

The classification system is based on a simple dendrogram-like hierarchy of Mg$^{2+}$ interactions, with the number of ligands and their chemical differences serving as the main criteria to define branches. This classification system was designed to be readily understandable and easy to automate, to make it an efficient tool for investigation of the diversity of Mg$^{2+}$ binding sites. The number of site types resulting from the classification strategy is within a reasonable range for practical usage; i.e. too many site types would render the system difficult to use, whereas too few types would be not enough to distinguish motifs and specific sites. The 136 site types in the RNA-inner and RNA-outer classes observed in the benchmark dataset used herein is not an exhaustive list; as additional RNA structures are determined, new site types are likely to be discovered. The hierarchical classification system offers different levels of abstraction depending on the particular application. For example, the use of just the number of inner-sphere phosphates as a determinant of site family yields only five families in the RNA-inner class. The naming convention of site types explicitly spells out the types and number of all ligands (or outer-sphere moieties in the case of the RNA-outer class) present in the site type. The name of each site type is unique and contains enough information to infer the subclasses at broader levels of classification.

The practical use of the classification system has been shown by the precise definition of previously reported Mg$^{2+}$-binding motifs, which (along with just a few additional criteria) is sufficient to identify new instances of these motifs (Table 2) and by the discovery of entirely novel motifs (Table 3). Both the previously reported and the newly found Mg$^{2+}$-binding motifs highlighted the ubiquitous presence of outer-sphere interactions, which play vital roles in inducing and maintaining the proper folding and formation of the Mg$^{2+}$-binding pocket. Three out of the six literature-reported motifs are defined by outer-sphere interactions exclusively (Table 2). As for the seven newly discovered motifs, two of them involve only outer-sphere interactions, while two others involve both characteristic inner- and outer-sphere interactions (Table 3). Therefore, comprehensive handling of outer-sphere interactions is of indispensable importance for the description of Mg$^{2+}$-binding motifs.

**Future applications of Mg$^{2+}$ site classification**

Although a lot of Mg$^{2+}$-binding motifs are expected to play mostly a universal structural role by charge compensation and fold stabilization, the presence of some Mg$^{2+}$ structural motifs may also indicate functional implications. A preliminary study has been carried out toward this direction, by correlating RNA functional families with site type as defined in our classification system. We noticed that inner-sphere interactions are more abundant in the large ribosomal subunit than in the small ribosomal subunit. We also noticed that base stacking is more frequent in the outer sphere of Mg$^{2+}$ sites in ribozyme. However, further study will be necessary to find more detailed correlations.

We believe our classification system has potential to detect unique structural sites responsible for specific functional roles. Mg$^{2+}$ binding sites with unique and complicated structural arrangements, especially those that are rarely observed, may be good candidates for investigation. Our tables of the populations of site types (Supplementary Tables S3/S4) and the server that describes them may be used by researchers to determine the extent of uniqueness of a particular coordinating pattern, allowing them to highlight candidate Mg$^{2+}$ sites for specific consideration.

With the availability of more data in the future, the method used may provide evidence of new validated Mg$^{2+}$ motifs, i.e. some specific coordinating patterns will prove to be sufficiently populated. Moreover, increasing the number of sites could allow a detailed classification of more complicated sites with additional metal or non-RNA ligand in the Mg$^{2+}$ coordination sphere. The growth of structural data will also permit the systematic investigation of other metals less commonly observed in RNA.

**Online access**

The classification of Mg$^{2+}$ sites in the benchmark dataset can be accessed via URL http://www.csgid.org/metalnas/. The main page of the server lists all site types with schematic drawings and the number of sites found for each site type. Detailed information for each site type may be accessed, which includes an image of a representative site and a list of all PDB entries containing a site of that type with chain and residue ids. Each Mg$^{2+}$ site may be visualized in Jmol (37). Users can specify a particular PDB ID or upload a RNA structure in the PDB format for analysis. A simple REST API is provided for users to download the whole benchmark dataset, as well as results of each particular search or analysis, in various formats (Figure 5A).

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR online.

**ACKNOWLEDGEMENT**

The authors would like to thank David R. Cooper, Rachel Vigour and Matthew D. Zimmerman for critically reading the manuscript. The authors would also like to thank Janusz Petkowski, Mahendra D. Chordia, Jing Hou and Igor A. Shumilin for valuable comments and discussions.

**FUNDING**

This work was supported, in whole or in part, by Federal funds from the National Institute of Allergy and Infectious Diseases, US National Institutes of Health, Department of Health and Human Services, under contract...
number HHSN27220120026C, as well as NIH grants GM094585, GM094662, and GM093342. J.M.B. is funded by the Foundation for Polish Science (FNP, TEAM/2009-4/2 to J.M.B.). Funding for open access charge: National Institutes of Health (US).

Conflict of interest statement. None declared.

REFERENCES