RNA and its ionic cloud:
Solution scattering experiments and atomically detailed simulations
(Supporting Material)

Serdal Kirmizialtin, Suzette A. Pabit, Steve P. Meisburger,
Lois Pollack and Ron Elber

Simulation Detail and the choice of Force Field

We used the OPLS-AA force field for the mobile ions and water molecules and kept the RNA molecule rigid. We focus on the study of the ions and rely on the experimental observation that the A form is indeed quite rigid with a persistence length of ~500 basepairs. Any significant "dynamics" observed for A-form of RNA is therefore questionable and may be an issue with the force field. We first tried the AMBER-99 force field since it is more widely used for nucleic acids. However simulations with the Aqvist parameters as implemented in AMBER-99 (before they were corrected by Joung and Cheatham (57)) show salt crystals even in aqueous solutions. The same phenomenon was noted and published by Chen et al. (58). The force field of OPLS-AA used the same Aqvist parameters. However, the combination rules of Van der Waals parameters are different. The Van der Waals radius of particles *i* and *j* in OPLS-AA is defined as $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$ and in AMBER as $\sigma_{ij} = 1/2\left(\sigma_i + \sigma_j\right)$. The parameters were designed for geometric averages and it is therefore not surprising that they performed better in the OPLS-AA force field. Control simulation in aqueous solution and near the RNA confirmed these observations. We finally note that a detailed study with comparisons of ionic solution force fields was published by Patra and Kartunnen(59) in 2004 and the satisfactory behavior of the OPLS-AA parameters is well documented in ref. (58).

In studies of the duplex and monovalent ions, the volume of the periodic box is 78.8×78.8×96.0Å^3. It includes 78 cations and 30 Cl^- ions with a total of 57,614 particles. The box size in the simulation with the divalent ions is 118×118×88.8Å^3 with 101 Sr^{2+} and 154 Cl^- ions and a total of 119,207 atoms shown in Figure S1. We used a larger box for strontium to examine convergence with box size. The larger box for divalent ions makes it possible to search for subtle long-range correlations, which at the end of the study were not found. Hence, it was possible to use significantly smaller box for the Sr^{2+} simulations. The actual simulated concentrations are the asymptotic converged concentrations, far from the RNA, and are found to be 0.1 ± 0.04 M for Sr^{2+} and 0.11 ± 0.04 M for monovalent solutions.

Starting with a random distribution of ions in the periodic box we equilibrated the system by Replica Exchange Molecular Dynamics (60). Replica exchange simulations were carried out for each of the aqueous solutions. For the duplex-Sr^{2+} system 88 replicas between the temperatures 300.0 to 380.6 K are used while for duplex-monovalent solutions 64 replicas were enough to cover the temperature range of 296.7 to 346.6 K. The temperatures were distributed to give an average acceptance probability of ~0.2 for exchanges between replicas. Simulation protocols for 0.4 M NaCl and 0.2 M $MgCl_2$ with A-form of RNA were described previously (39).

We simulated the properties of the solutions keeping the RNA molecule frozen. A matrix variant of SHAKE (61) is used to keep the water molecules rigid and smooth Particle Mesh Ewald (62) is used to compute long-range electrostatic interactions with a real-space cutoff of 9 Å for electrostatics and 8 Å for the Lennard Jones interactions. We used a time step of 1.5 femtoseconds. The total length of a simulation for each ionic concentration was about 20 nanoseconds. We check the statistical convergence of observables computed in the simulation as described in Kirmizialtin and Elber (39).

Pseudoknot simulations are conducted at T = 300K. Simulation details that are different from above are summarized below. The crystal structure of 437D was solvated with TIP3P water molecules, 40 Na^+ and 13 Cl^- ions. In a periodic box of $(63.5\ Å)^3$ the asymptotic concentration of Na^+ is found to be 0.17 ± 0.01 M. We run a 15 ns equilibration simulation in which the RNA molecule was kept rigid in the conformation of the crystal while the ions and water molecules were allowed to relax. After the equilibration period, a production run was conducted for another 15 ns saving configurations every 2.5 ns. The configurations saved are used to initiate six runs of ~40 ns from each structure. The simulations were conducted in the canonical ensemble using velocity scaling. We saved configurations every 2 ps and used these data for clustering and SAXS calculation.

Estimates of the Errors

There are two sources of errors in estimating averages using Markov chain sampling of a distribution (MD is one way of generating Markov chains). The first source, poor statistics, can be tested with a so-called "standard statistical test" in which the complete set is considered. The second source of errors is due to the fact that Markov chains are not always ergodic and therefore the sampling from the distribution is not uniform. This is a subtle effect that is harder to test and sometimes ignored. By dividing the Markov chain into sequential segments we can examine the possibility of a drift in the average and incomplete ergodicity. In principle the average should be the same (up to the standard errors) in each segment. But the second source of errors is usually larger. To address that we plot the two estimates of the errors in Figure S2. Indeed the "standard test" (first method) produces smaller error bars, but we believe the second method (with larger error bars) is more meaningful.

We have developed an ergodic measure that tests how a particular observable deviates from an average expected from an ergodic, uniform sampling. Consider an observable D with a mean $\langle D\rangle_N$ obtained from N sampling points. The standard deviation of the N points is σ_N. We define the measure $\chi(N) \equiv N^{1/2}\sigma_N / \langle D\rangle_N$. If the sampling is made from a uniform distribution the measure should approach a constant value. A plot of $\chi(N)$ as a function of N is shown for all cations studied in Figure S3. The observable we chose is the number of ions within 5Å from the RNA. Since the volume near the RNA is small this observable usually suffers from small statistics and significant errors. It is therefore a useful target to assess the accuracy of the calculation.

Calculating the ASAXS profile from MD simulations:

The anomalous scattering signal from ions is the difference between the scattering intensities when the X-ray energy is below the ions atomic resonance ($E_{off,}$ for off-edge) and near it (E_{on}, for on-edge)

$$I_{anom}(q) = I(q,E_{off}) - I(q,E_{on}) \qquad (1)$$

where $|q| = \frac{4\pi \sin(\theta)}{\lambda}$ is the amplitude of momentum transfer, λ is wavelength and 2θ is the scattering angle.

The energy-dependent scattering intensity $I(q,E)$ can be calculated from the Debye formula (47) as

$$\sum_{i,j}^{N} f_i(q,E) f_j(q,E) \frac{\sin(qr_{ij})}{qr_{ij}} + \Delta\rho_b \sum_{i}^{N} \sum_{j}^{M} f_i(q,E) f_j(q,E) \frac{\sin(qr_{ij})}{qr_{ij}} \qquad (2)$$

Here, the first term sums over *N* atoms of solute (RNA and ions) and is the scattering intensity of the solute.

$f_i(q,E)$ is the energy dependent relative form factor of atom *i* in water and is given by

$$f_i(q,E) = g_i(q,E) - \nu_i \rho_s \exp[-\pi \nu_i^{2/3} q^2] \qquad (3)$$

where $g_i(q,E)$ the atomic form factor, is equal to the number of electrons for atom *i* at *q=0* and, ν_i is the volume of atom *i*. The values of these two parameters for all atom types are taken from Ref. (52, 63). ρ_s is electron density of water, 0.334e/Å^3 and r_{ij} is the interatomic distance between pairs *i*, *j*. The change in the form factor of Sr^{2+} at two different energies is given as $g(q,E_{on}) = g(q,E_{off}) - 6.5$; the change in form factor between the two Rb^+ energies is 4 electrons. The energy change does not affect the form factors of the rest of the terms.

The second summation in Eq. 2 is over the *N* solute atoms and *M* water molecules that are close to the RNA with a distance cutoff $d \le 3\text{Å}$. A weight factor $\Delta\rho_b$ is used to account for the excess electron density of the hydration shell which is taken to be 0.1 in this study (*d* and $\Delta\rho$ are CRYSOL (52) defaults). Ions far from the RNA have no effect on the anomalous scattering. Thus we choose an optimal value for the cutoff according to the distribution shown in Figure 1. The beginning of the constant ion distribution marks the bulk. Ions that are further than 18 Å away from the center of the long axis are considered bulk in Sr^{2+} and ignored in ASAXS calculations. For Rb^+ the influence radius is larger thus the cutoff value is chosen to be at 28 Å. An alternative approach would be to include all ions and subtract the calculated distribution from the distributions of bulk solvent but this process is computationally demanding since extensive statistics is required to make bulk subtractions equal zero. Empirically, we find this alternative approach hard to converge.

Calculating the number of ions using ASAXS measurements:

Anomalous Small-Angle X-ray Scattering (ASAXS) experiments were used to characterize the ion atmosphere around RNA molecules. We measured SAXS signals at several x-ray energies, close to but below the absorption edge of the ions of interest. The measured absorption edge for Rb^+ and Sr^{2+} ions were 15.200 keV and 16.113 keV, respectively. Near an ion's absorption edge, the atomic scattering factor (in units of electrons) is denoted by:

$$f_{ion}(E) = f_0 + f'(E) + i f''(E) \qquad (4)$$

where f_0 is the energy-independent solvent-corrected scattering factor (atomic number Z in vacuum) of the resonant element, f' and f'' are the energy-dependent anomalous scattering factors, and E is the x-ray energy. The scattering intensity from the nucleic acid and counterion cloud system is a function both of energy E and momentum transfer q, ($q = (4\pi/\lambda)\sin\theta$, where λ is the x-ray wavelength and 2θ is the scattering angle) and is given by Eq. 2 in the manuscript.

$$I(q,E) = |f_{NA} F_{NA}(q) + f_{ion}(E) N_{ions} F_{ion}(q)|^2. \qquad (5)$$

Terms described by F's reflect the spatial arrangement of the scattering particles (treated as unity at $q = 0$), and those represented by f_{NA} describe the effective number of electrons from a nucleic acid duplex. N_{ions} equals the number of excess cations around the nucleic acid (greater than the ion number in the bulk solution).

To effectively measure the number of ions, N_{ions} , around the nucleic acid, we follow the following steps (19):

1. We measure the scattering factors, f' and f'' using x-ray fluorescence measurements from a buffer solution containing the energy-dependent scatterer, e.g. 30 mM RbAcetate or SrAcetate. X -ray fluorescence was collected 90° from the incident beam using an Xflash detector (Rontec, Carlisle, MA). CHOOCH (40), a program commonly used for heavy-atom refinement, was applied to extract f' and f'' from x-ray fluorescence data.

2. We calibrate the x-ray scattering intensity of the nucleic acids using water as a calibration standard following Eq. 6 below (19, 41):

$$I_{RNA}(q) = \frac{c_{H_2O}}{c_{RNA}} \cdot \frac{P_{H_2O}(0) \cdot S_{H_2O}(0)}{I^N_{H_2O}(0)} I^N_{RNA}(0) \qquad (6)$$

The SAXS intensities, I_N, are normalized using the transmitted signal from the direct beam reflected by an amorphous beam stop to an X-flash counter (20). $P_{H2O}(0)$ is the scattering form factor of water, $P_{H2O}(0)$ = 100 electrons squared. $S_{H2O}(0)$ is the

structure factor of water at $q = 0$ which is 0.062 at temperature of 23.4°C (41, 64) . Concentrations of the water and RNA samples are indicated by c.

3. We acquire the small-angle scattering profiles of the RNA-ion system at multiple energies, $I(q,E)$, tabulated below:

Near Rb edge		**Near Sr edge**	
Energy (keV)	**Rb *f'*(electrons)**	**Energy (keV)**	**Sr *f'*(electrons)**
15.020	-3.58	15.930	-3.61
15.093	-4.09	16.003	-4.11
15.160	-5.09	16.070	-5.13
15.180	-5.82	16.092	-5.95
15.193	-7.02	16.103	-6.99

4. We decompose $I(q,E)$ in terms of $f'(E)$. To do this, we extract the q dependent functions, $a(q)$, $b(q)$ and $c(q)$ from measurement of $I(q)$ at several different energies by plotting the measured $I(f'(E))$ at each q value (or over a small range in q to improve statistics), and carrying out a quadratic fit. This procedure is repeated for all q, reconstructing the functions of interest, point by point (65) as shown in Eq. 7 below:

$$I(q,E) = a\cdot(f'(E)^2) + b\cdot f'(E) + c \quad \text{where} \qquad (7)$$
$$a(q) = N_{ions}{}^2 F_{ion}{}^2$$
$$b(q) = N_{ions}\cdot[\, 2f_{NA}F_{NA}F_{ion} + 2f_o N_{ions}F_{ion}{}^2 \,]$$
$$c(q) = (f_{NA}F_{NA})^2 + 2f_{NA}f_o N_{ions}F_{NA}F_{ion} + f_o{}^2 N_{ions}{}^2 F_{ion}{}^2$$

5. For the RNA-ion system taken into consideration, the contribution to the scattering profile from the nucleic acid is much greater than that from the ion cloud. For a 1-bp RNA molecule with 2 Rb^+ ions: $c(0)/b(0) \approx 50$ and $b(0)/a(0) \approx 200$. The $b(0)/a(0)$ ratio is even larger when considering Sr^{2+} ions. Therefore, the '*a'* term is negligible relative to the others and can be ignored. We can then apply a linear fit:

$$I(q,f'(E)) = b(q)\cdot f' + c(q) \qquad (8)$$

6. We use GNOM (66) or a Guinier analysis to extrapolate $b(q)$ and $c(q)$ to $q = 0$ and use these values to calculate the number of ions:

$$N_{ions} = \frac{b(0)}{2\sqrt{c(0)}} \qquad (9)$$

7. This approach can further be simplified for two-energy ASAXS as long as steps (1) and (2) are properly implemented

$$N_{ions} = \frac{\sqrt{I(0,E_1)} - \sqrt{I(0,E_2)}}{f'(E_1) - f'(E_2)} \qquad (10)$$

8. If there are intermolecular interactions in solution (as shown in Figure 3b where the RNA duplexes exhibit end-to-end stacking), a structure factor correction needs to be implemented. Note that solution scattering for a system with intermolecular interactions is given by:

$$I(q) = \frac{N}{V} I_0(q) \cdot S(q) \qquad (11)$$

$I_0(q)$ is the scattering from a single molecule in an N molecule ensemble within a volume V. $S(q)$ is the structure factor from intermolecular interactions. For non-interacting systems, $S(q) = 1$. For weakly interacting systems, $S(0) > 1$ indicates intermolecular attraction and $S(0) < 1$ indicates intermolecular repulsion. In this case, the number of ions will be:

$$N_{ions} = \frac{1}{\sqrt{S(0)}} \cdot \frac{\sqrt{I(0,E_1)} - \sqrt{I(0,E_2)}}{f'(E_1) - f'(E_2)} \qquad (12)$$

$S(0)$ can be determined by careful measurements of the second virial coefficients as shown in ref. (13).

ASAXS also reports the spatial distribution of the ions with respect to the nucleic acid. For a conventional ASAXS experiment using 2 x-ray energies (13, 20), the anomalous signal is given by (see Eq. 1, above) :

$$I_{anom}(q) = I(q, E_1) - I(q, E_2) \approx 2\alpha N_{ions} f_{NA} F_{NA}(q) F_{ion}(q) \, (f'(E_1) - f'(E_2)) \qquad (13)$$

E_1 is an x-ray energy far away from the edge (e.g. 15.093 keV for Rb) and E_2 is an x-ray energy close to the absorption edge (e.g. 15.193 keV for Rb). This is equivalent to the $b(q)$ term derived above and the data are shown in Figure 3.

References appear in the main text.

Supporting Figures in the Supplementary Material

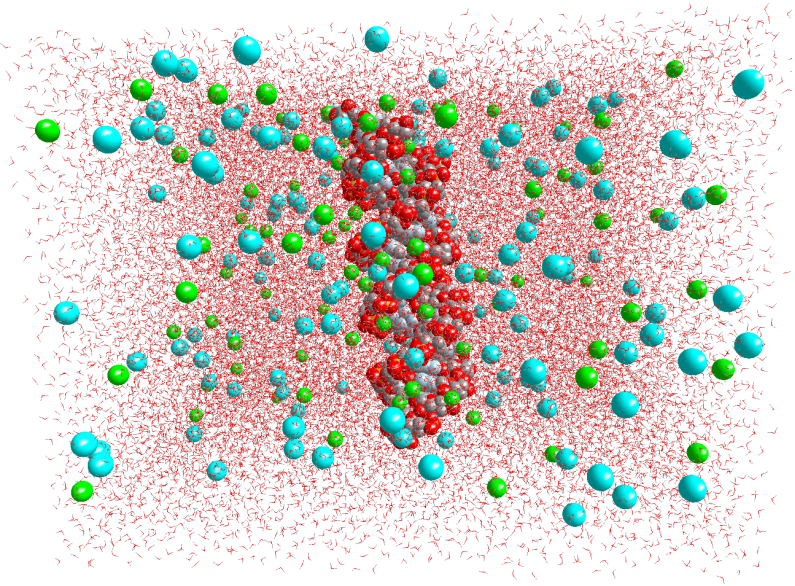

Figure S1. An equilibrium configuration of ions and explicit water molecules around an A-form 25 base pair RNA duplex of length 74 Å is shown. Strontium ions are depicted with green spheres, chloride ions with cyan spheres and waters as red and white sticks. This image was generated by the graphic module of MOIL -- Zmoil (http://clsb.ices.utexas.edu/prebuilt/).

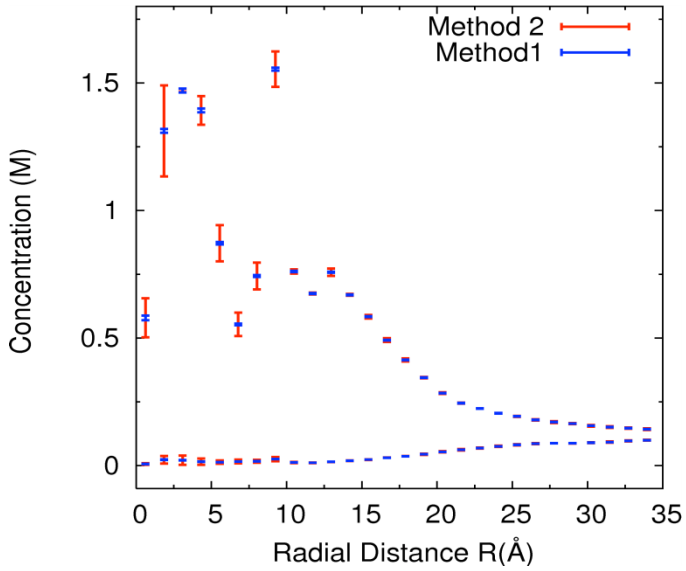

Figure S2. Two methods of calculating the error bars are shown. In method 1 (blue) we compute the error bars from all data sets. In method 2 (red) we divide the data into three blocks and compute the concentration profiles separately. We used the results of three sets to compute the error bars.

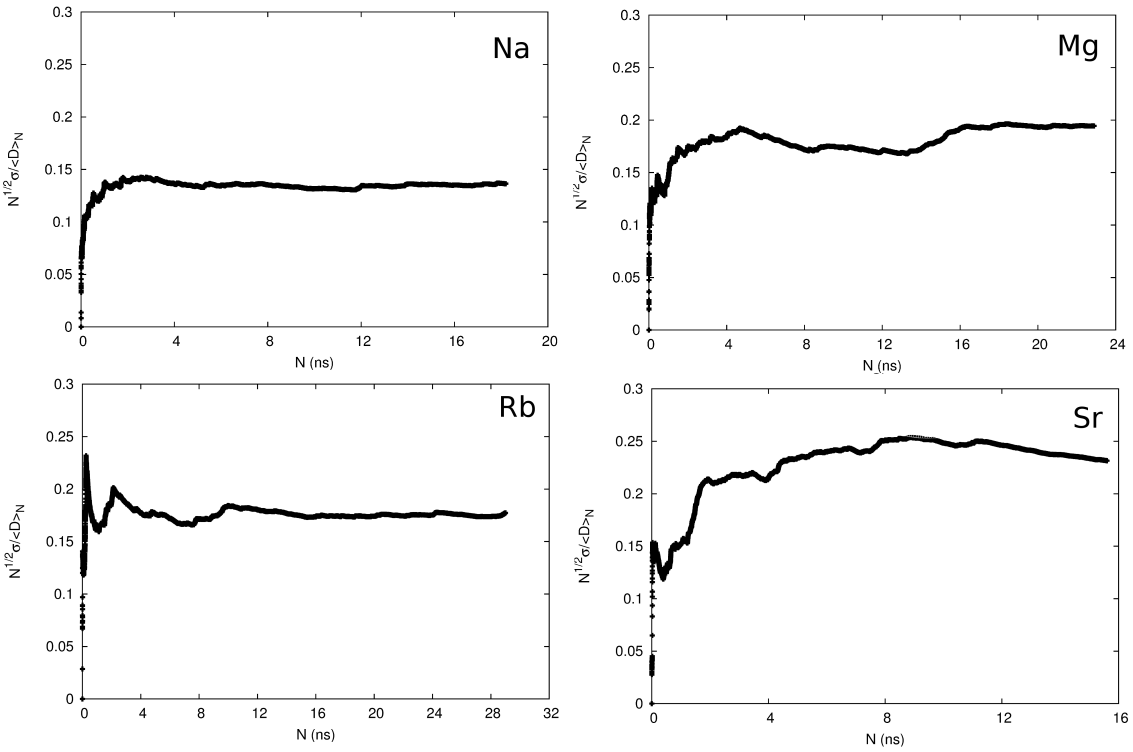

Figure S3. The measure of ergodicity, $\chi(N) \equiv N^{1/2}\sigma_N / D_N$, of the equilibrium distributions shown for all ion types. The observable *D* is total number of ions within 5Å distance from the closest RNA atom.

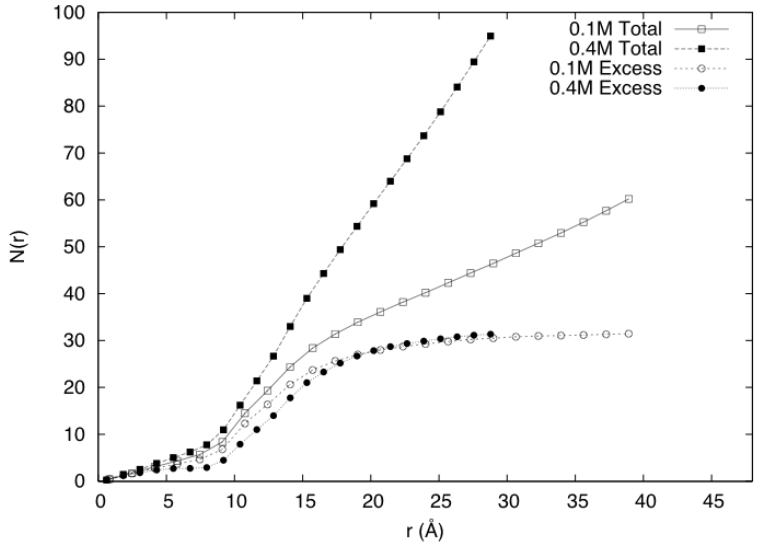

Figure S4. The number of cations in a cylindrical volume of radius *r* is shown for different concentrations of the same ion. Filled symbols are from 0.4 M NaCl while open symbols are from 0.1 M NaCl solution. Total number of ions in a cylinder of radius r is shown as squares. The number of cations excluding the bulk contribution is shown as circles. Overlap in the number of excess ions at $r > 18$ Å suggests that the number of condensed ions is independent of monovalent salt concentration.

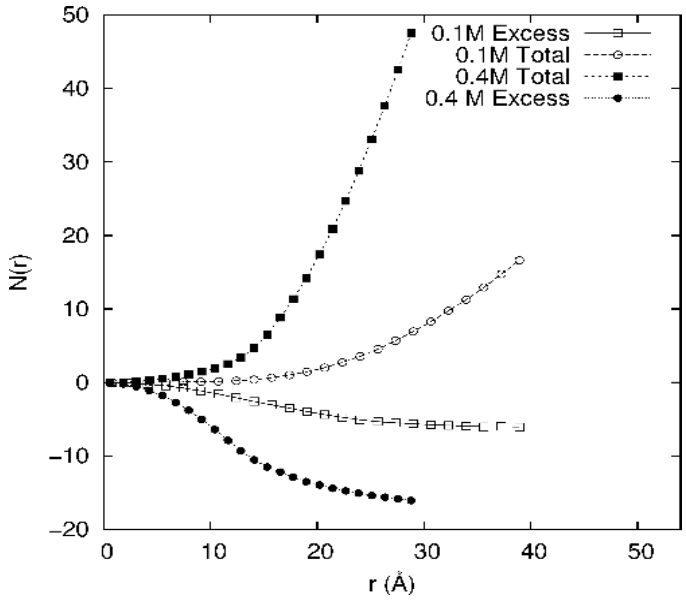

Figure S5. The number of co-ions in a cylindrical volume of radius *r* is shown for two different concentrations of the same cation. Filled symbols are from 0.4 M NaCl while open symbols are from 0.1 M NaCl solution. Total number of ions in a cylinder of radius r is shown as squares. The number of co-ions excluding the bulk contribution is shown as circles.

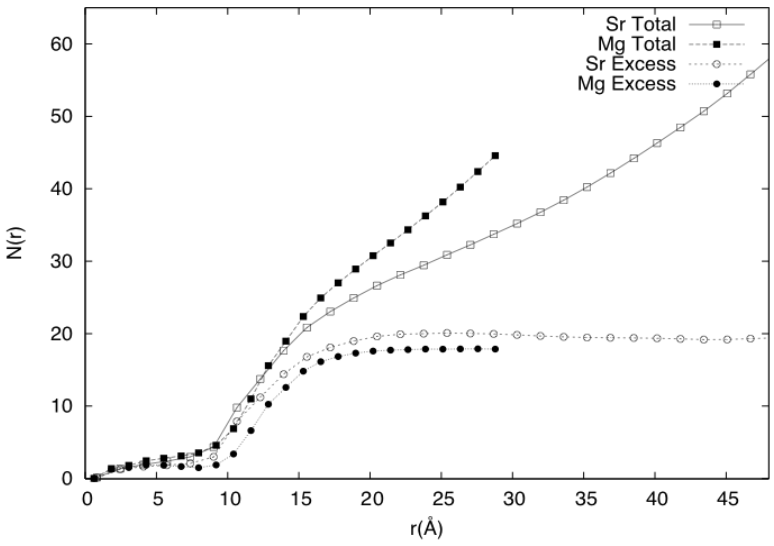

Figure S6. The number of cations in a cylindrical volume of radius *r* is shown for different ions of the same valence. Filled symbols are from 0.2 M $MgCl_2$ solution while open symbols are from 0.1 M $SrCl_2$. The total number of ions in a cylinder of radius *r* is shown as squares while the number of cations excluding the bulk contribution is shown as circles.

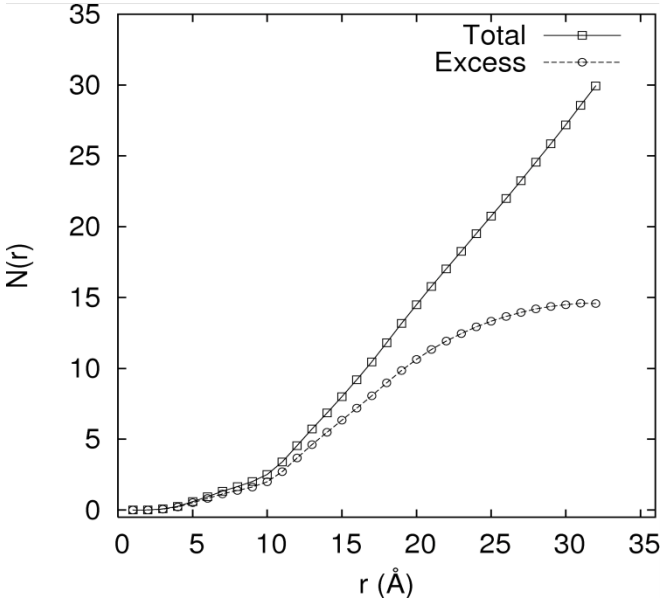

Figure S7. The sodium ion distribution around the BWYV pseudoknot is shown as derived from MD simulations. The total number of ions contained in a spherical volume of radius *r*, the distance from the center of the pseudoknot, is shown. The open squares represent the total number of sodium ions; the open circles are the excess ions computed by subtracting the bulk contribution from the total number.