Example 4: Simulating Monovalent Ion Distributions around Nucleic Acid Duplexes

Contents

[Solute all-atom coordinate files and electrostatic potentials 1](#_Toc479411189)

[Example file folders 2](#_Toc479411190)

[Example 4A: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl , using the Unrestricted Primitive Model (UPM) 3](#_Toc479411191)

[Simulation steps 3](#_Toc479411192)

[Example 4B: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl, using the SPM model with Lennard-Jones ion-water interactions 8](#_Toc479411193)

[Simulation steps 8](#_Toc479411194)

[RDF plots 9](#_Toc479411195)

[Example 4C: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using UPM 10](#_Toc479411196)

[Example 4D: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using the SPM model with Lennard-Jones ion-water interactions 11](#_Toc479411197)

[RDF plots 11](#_Toc479411198)

# Solute all-atom coordinate files and electrostatic potentials

Uncompress the dna\_rna\_allatom\_coordinates\_potentials.tar.gz to extract the all-atom coordinate files (.pqr) and electrostatic potential maps (.dx) of four 25 base pair nucleic acid duplexes.

$ tar –zxvf dna\_rna\_allatom\_coordinates\_potentials.tar.gz

The folder names and files are listed in the table below. The PQR files were generated using PDB2PQR software, from PDB files provided by Alexey Onufriev and Igor Tolokh (from Virginia Tech). The electrostatic potentials (in .dx files) were approximated by solving the Poisson equation with the Adaptive Poisson Boltzman software (APBS) package. The input file used for the APBS calculations are also provided in the file folders.

|  |  |  |
| --- | --- | --- |
| File folder | 25 bp nucleic acid duplex | .pqr and .dx files |
| polyA\_polyT\_DNA\_model | poly(dA):poly(dT) | polyAT25.pqr, pot\_polyAT25\_npbe\_pdie2.dx |
| A-RNA\_model | A-RNA | rGCAU25\_cc.pqr, pot\_rGCAU25\_cc\_npbe\_pdie2.dx |
| mixed\_sequence\_DNA\_model | Mixed sequence DNA | dGCAT25\_cc.pqr, pot\_dGCAT25\_cc\_npbe\_pdie2.dx |
| DNA\_RNA\_hybrid\_model | DNA:RNA hybrid | hyb\_DNA\_RNA\_cc.pqr, pot\_hyb\_DNA\_RNA\_cc\_npbe\_pdie2.dx |

# Example file folders

Uncompress the files listed in the table below, using the command

$ tar –zxvf {filename}.tar.gz

|  |  |
| --- | --- |
| Simulation Examples | Compressed files |
| Example 4A: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl , using the Unrestricted Primitive Model (UPM) | ./gibs/Simulation\_Examples/excess\_chempot\_vs\_conc\_NaCl\_PID\_UPM/CONC\_100mM.tar.gz  ./gibs/Simulation\_Examples/NaCl\_UPM\_PolyAT25.tar.gz |
| Example 4B: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl, using the SPM model with Lennard-Jones ion-water interactions | ./gibs/Simulation\_Examples/excess\_chempot\_nacl\_100mM\_PID\_SPM+LJ.tar.gz  ./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25.tar.gz |
| Example 4C: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using UPM | ./gibs/Simulation\_Examples/excess\_chempot\_rbcl\_100mM\_PID\_UPM.tar.gz  ./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25.tar.gz |
| Example 4D: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using the SPM model with Lennard-Jones ion-water interactions | ./gibs/Simulation\_Examples/excess\_chempot\_rbcl\_100mM\_PID\_SPM+LJ.tar.gz  ./gibs/Simulation\_Examples/RbCl\_SPM\_LJ+IW\_PolyAT25.tar.gz |

Note: All input files are provided. For the solute simulations, due to the large file sizes associated with particle counts and acceptance rates, only the RDF output files and final simulation states are provided. However, for the equilibration runs, all output files are provided. In some cases (e.g, the excess potential calculations), only the input files may be available but the final chemical potential values can be checked by comparing the values to those listed in this document.

# Example 4A: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl , using the Unrestricted Primitive Model (UPM)

In this example, we compute the cylindrical radial distribution of ions around a 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl, using the UPM model.

## Simulation steps

**Step 1:** First compute the excess chemical potential of Na+ and Cl- in 100 mM NaCl.

We can use either the A-GCMC method (Example 1) or the PID method (Examples 2 and 3A). PID is recommended for simulations with water hard spheres. See Examples 1 to 3 for guidance in setting up and running simulations for calculating the excess chemical potential of ions in bulk electrolyte concentrations.

In Example 3A, we computed the average excess chemical potential values of Na+ and Cl- in 100 mM NaCl, using the PID method: = -0.16483 and = -0.16333 kcal/mol. For the current example, we had used values from a previous version, -0.16506 and -0.16326 kcal/mol for Na+ and Cl-, respectively. These values are within the standard deviation of 0.03, reported in Example 3A, and they don’t affect the results of the solute GCMC simulations Therefore, it should be noted that average values can slightly vary depending on the number of GCMC steps and iterations used. Nevertheless, the number of simulation steps should be chosen large enough for the excess chemical potential and the number of each particle type (ions or solvent) to reach a steady oscillatory state, in order to calculate steady state (average) values.

**Step 2:** Create a new simulation run directory for the solute GCMC simulations, e.g.,

mkdir EQUILIBRATION\_RUN

In this directory, we perform a solute-ion simulation for 108 GCMC steps, starting with a random configuration (state) of particle species (Na+, Cl-) in the simulation box.

Input files are provided in the folder: ./gibs/Simulation\_Examples/NaCl\_UPM\_PolyAT25/EQUILIBRATION\_RUN/inputfiles

**Step 3:** Set the parameters in the inputfiles/inputparameters.in file are

Specify the type of simulation to run for simulations using all-atom solute models.

SIMULATION\_TYPE GCMC\_WITH\_SOLUTE\_ALLATOM\_MODEL

Specify the type of solute model. In this example, we use an all-atom model.

SOLUTE\_MODEL\_TYPE all\_atom

Since we are using an all-atom solute model, we need to specify the name and format of the structure file.

SOLUTE\_ALLATOM\_COORDINATE\_FILETYPE PQR

SOLUTE\_ALLATOM\_COORDINATE\_FILENAME ../../../polyA\_polyT\_DNA\_model/polyAT25.pqr

For UPM simulations, turn off the use of SPM and MWater models.

USE\_SPM NO

USE\_MWATER NO

Set the maximum allowed number of particles in the box, to allocate memory for arrays and vectors. (Note: The value used was too large for this system. About 1000 is sufficient).

MAXIMUM\_NUMBER\_OF\_PARTICLES 40000

Set the number of iterations to 1 (always) and the number of GCMC steps to 108.

NUM\_ITER 1

NUM\_STEPS 100000000

Set the number of insertion/deletion cycles (NUM\_GCMCCYCL) and the single-particle displacement cycles (NUM\_MOVCYCL). The NUM\_GCMCCYCL and NUM\_MOVCYCL values are normally set such that 50-70% of the random moves (NUM\_STEPS) are single-particle displacements. The NUM\_GCMCCYCL value is normally set equal to or less than the expected (average) number of total ions in the simulation box (See book by D. Frenkel and B. Smit. Understanding Molecular Simulation. Academic Press, San Diego (1996)). For example, during a GCMC step, if a random number generated between 0 and (NUM\_GCMCCYCL+NUM\_MOVCYCL) is greater than the particles in the box, then an insertion/deletion is attempted; otherwise, a single-particle displacement is attempted.

NUM\_GCMCCYCL 600

NUM\_MOVCYCL 600

Set the number of steps for equilibration. This keyword is used to record the states after the system has reached equilibrium (the number of particles of each type in the system has reached a steady (stationary) state)

NUM\_EQSTEPS 10000000

Since the initial state is a random configuration, it is not read from the state\_0 file. Therefore, set START\_STATE to zero.

START\_STATE 0

To record the simulation outputs

RECORD\_PARTICLETYPE\_COUNT YES

RECORD\_MOVE\_ACCEPTANCE\_RATE YES

RECORD\_SYSTEM\_ENERGY YES

RECORD\_EVERY\_N\_STEPS 100

In this simulation, we record 20 states from the equilibrated phase (stationary part) of the simulation

NUM\_RECORD\_STATES 20

In this simulation, we do not calculate RDFs

CALCULATE\_SOLUTE\_PARTICLE\_RDF NO

CALCULATE\_PARTICLE\_PAIR\_RDF NO

Ion type labels, charge, radius, concentration and excess chemical potentials are specified as below

ION Na 1.0 1.02 0.1 -0.16506

ION Cl -1.0 1.81 0.1 -0.16326

Box dimensions and grid spacing should match the values used for the APBS electrostatic potential calculation (recorded in .dx file).

X\_LEN 150

Y\_LEN 150

Z\_LEN 180

HX 1.171875

HY 1.171875

HZ 1.40625

X\_MIN -74.395

Y\_MIN -74.144

Z\_MIN -88.061

Set the temperature and solvent dielectric constant

TEMPERATURE 298.0

SOLVENT\_DIELECTRIC 78.5

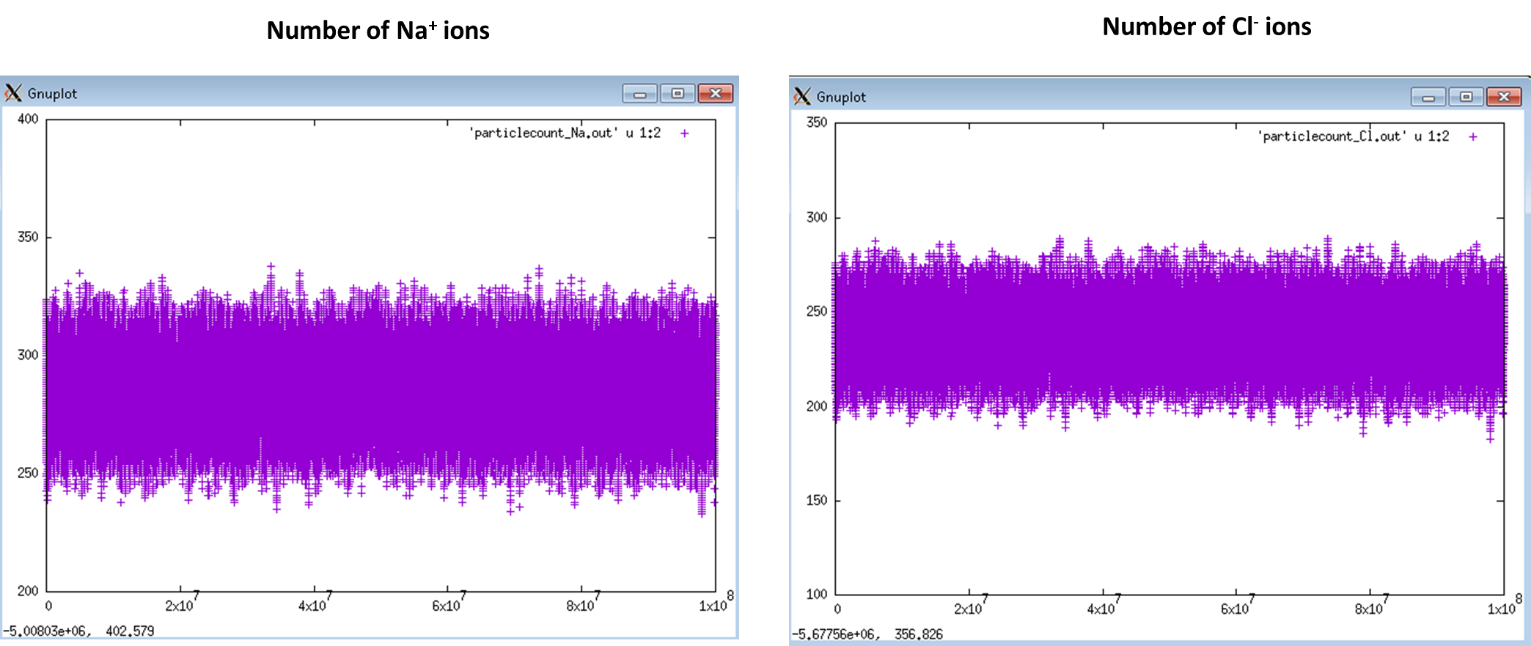
Specify the path and name of the .dx file to use for the electrostatic potential

ELECTROSTATIC\_POTENTIAL\_DX\_MAP ../../../polyA\_polyT\_DNA\_model/pot\_polyAT25\_npbe\_pdie2.dx

**Step 4**: Run the simulation. For example,

$ ./gibs.exe > run\_1\_29\_2017\_r1 &

Check if the system has reached equilibrium by plotting the number of particles in the system.



**Step 5**: Create simulation run directories for the production runs, and the sub-directories, inputfiles and outputfiles. In this example, there are 4 production runs.

mkdir PRODUCTION\_RUN1

mkdir PRODUCTION\_RUN2

mkdir PRODUCTION\_RUN3

mkdir PRODUCTION\_RUN4

**Step 6**: Copy all the input files from the simulation folder, /EQUILIBRATION\_RUN/inputfiles to the inputfiles folder of each production run directory.

**Step 7**: Copy one of the recorded states from the EQUILIBRATION\_RUN as initial state for each production run

$ cp ./EQUILIBRATION\_RUN /outputfiles/state\_100000000 ./PRODUCTION\_RUN1/inputfiles/state\_0

$ cp ./EQUILIBRATION\_RUN/outputfiles/state\_95500000 ./PRODUCTION\_RUN2/inputfiles/state\_0

$ cp ./EQUILIBRATION\_RUN/outputfiles/state\_91000000 ./PRODUCTION\_RUN3/inputfiles/state\_0

$ cp ./EQUILIBRATION\_RUN/outputfiles/state\_86500000 ./PRODUCTION\_RUN4/inputfiles/state\_0

**Step 8**: Change the following parameter values (as shown) in the inputparameters.in file of each production run

To read the initial state from state\_0 file, set

START\_STATE 1

Since we are starting from a state at equilibrium, we don’t need to wait for the system to equilibrate. So, use a small non-zero value for the number of equilibrium steps. Here we use 100.

NUM\_EQSTEPS 100

To compute the cylindrical RDF of each ion, the region surrounding the nucleic acid duplex is divided into cylindrical shells of width 0.5 A (RDF\_BIN\_SIZE), and length equal to 80% (RDF\_REF\_AXIS\_LENGTH\_FRACTION) of the NA duplex length (to minimize Coulombic end effects).

CALCULATE\_SOLUTE\_PARTICLE\_RDF YES

CALCULATE\_PARTICLE\_PAIR\_RDF NO

RDF\_MAX\_DISTANCE 70

RDF\_BIN\_SIZE 0.5

NUM\_RDF\_BINS 140

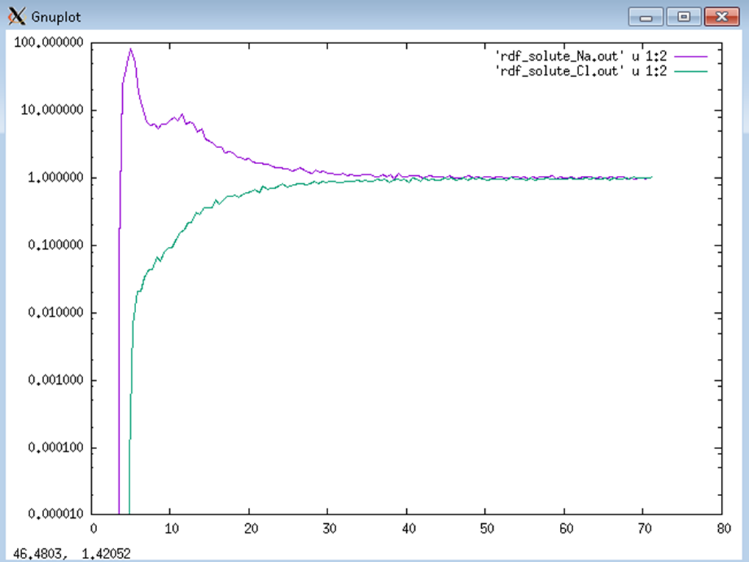
RDF\_REF AXIS\_OF\_CYLINDER

RDF\_REF\_AXIS\_LENGTH\_FRACTION 0.8

**Step 9**: Start the production runs.

The RDFs will be written out to rdf\_solute\_Na.out and rdf\_solute\_Cl.out files.

The figure below shows how the RDF plot looks like from the simulation in PRODUCTION\_RUN1 (y-axis is in log scale)



**Step 10**: Average the RDFs from the 4 production runs to get the final RDFs. Only use the data in columns 1 and 2 of the RDF output files. Ignore the other columns.

# Example 4B: GCMC Simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl, using the SPM model with Lennard-Jones ion-water interactions

In this example, we compute the cylindrical radial distribution of ions and water around a 25 bp poly(dA):poly(dT) DNA duplex in 100 mM NaCl and water (solvent packing fraction 0.3 in 100 Angstrom cube box), using the SPM model with ion-water Lennard-Jones interactions. This example is different from example 4A in that, it includes Lennard-Jones interactions between the ion and water hard spheres.

## Simulation steps

**Step 1**: Run a GCMC simulation to calculate the excess chemical potential of the ions and water, using the PID method.

This simulation files are provided in the folder: ./gibs/Simulation\_Examples/excess\_chempot\_nacl\_100mM\_PID\_SPM+LJ

The particle\_interactions.in file are set up as shown below

ParticleType ParticleType HARD\_SPHERE\_REPULSION COULOMB\_POTENTIAL LENNARD\_JONES\_POTENTIAL SQUARE\_WELL\_POTENTIAL LOOKUP\_TABLE

Na Na 1 1 0 0 0

Cl Cl 1 1 0 0 0

Water Water 1 0 0 0 0

Na Cl 1 1 0 0 0

Na Water 1 0 1 0 0

Cl Water 1 0 1 0 0

Set up the inputparameters.in file for SPM simulations as described in Example 2A.

Set the Lennard-Jones cut-off distance (typically 12.0 Angstroms) in the inputparameters.in file:

LENNARD\_JONES\_CUTOFF 12.0

The excess chemical potential values for Na+, Cl-, and water are -0.1890 ± 0.04, 0.9778 ± 0.04, and 1.048 ± 0.03 kcal/mol, respectively. The current version of GIBS calibrates the excess chemical potential of each particle type in the system based on error in the target mass density (target concentration × molar mass).

In a previous version of GIBS, the excess chemical potentials were estimated using the PID method based on error in the target concentration (100 mM for Na+ and Cl-, 22.693 M for water). The average values for Na+, Cl-, and water were -0.1739, 1.009, and 1.0650 kcal/mol, respectively. These values are very close to the values obtained with target mass density. It is important to note these minor differences can occur in the excess chemical potential calculations depending on the target conditions, and the GCMC simulation steps, but these differences are so minor that they hardly affect the results of the solute GCMC simulations and the RDF calculations.

**Step 2**: Run the solute GCMC simulations by following the same steps described in Example 4A.

Perform an initial run to equilibrate the system. The input files are provided in the folder: ./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/EQUILIBRATION\_RUN/inputfiles

**Step 3:** Perform six production runs to compute the RDFs. In the example folders, there are six simulation folders

./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN1

./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN2

./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN3

./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN4

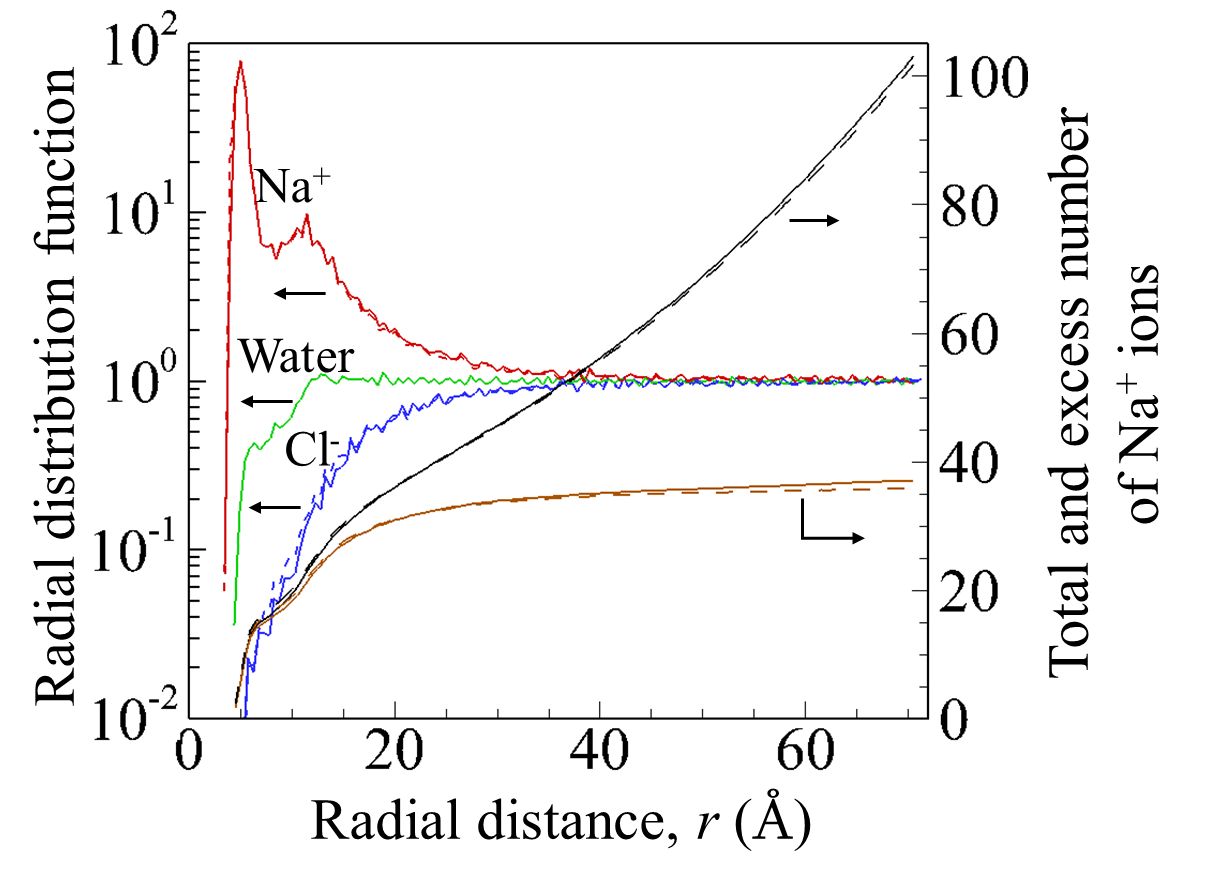
./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN5

./gibs/Simulation\_Examples/NaCl\_SPM\_LJ+IW\_PolyAT25/PRODUCTION\_RUN6

The input files needed for the simulations are provided in inputfiles folders. Run each simulation and plot the RDFs. Compare them to the RDFs values provided in each production run folder (rdf\_solute\_Na.out, rdf\_solute\_Cl.out, rdf\_solute\_Water.out). Only use the data in columns 1 and 2 of the RDF output files. Ignore the other columns. The final RDF is obtained by averaging the RDFs from the six production runs.

## RDF plots

The figure below shows RDF plots from the UPM (from Example 4A) and SPM+LJ simulations for 100 mM NaCl. The total and excess number of Na+ ions are also plotted.



# Example 4C: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using UPM

In this example, we compute the cylindrical radial distribution of ions around a 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using the UPM model.

The simulations are performed in the same way as in Example 4A.

Input files for the excess chemical potential calculations are provided in the folder: ./gibs/Simulation\_Examples/excess\_chempot\_vs\_conc\_RbCl\_PID\_UPM

Input files for the solute simulations are provided in the folders:

./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25/EQUILIBRATION\_RUN

./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25/PRODUCTION\_RUN1

./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25/PRODUCTION\_RUN2

./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25/PRODUCTION\_RUN3

./gibs/Simulation\_Examples/RbCl\_UPM\_PolyAT25/PRODUCTION\_RUN4

Run each simulation. Compare the RDFs to the data provided (rdf\_solute\_Rb.out, rdf\_solute\_Cl.out) in the production folder. Only use the data in columns 1 and 2 of the RDF output files. Ignore the other columns. The final RDF is obtained by averaging the RDFs from the six production runs.

# Example 4D: GCMC simulation with 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl, using the SPM model with Lennard-Jones ion-water interactions

In this example, we compute the cylindrical radial distribution of ions and water around a 25 bp poly(dA):poly(dT) DNA duplex in 100 mM RbCl and water (solvent packing fraction 0.3 in 100 Angstrom cube box), using the SPM model with ion-water Lennard-Jones interactions. This example is different from example 4C in that, it includes Lennard-Jones interactions between the ion and water hard spheres.

The simulations are performed in the same way as in Example 4B.

Input and output files for the excess chemical potential calculations are provided in the folder: ./gibs/Simulation\_Examples/excess\_chempot\_rbcl\_100mM\_PID\_SPM+LJ

Input files for the solute simulations (equilibration and production) are provided in the folders:

./gibs/Simulation\_Examples/RbCl\_SPM\_LJ+IW\_PolyAT25

Run each simulation and plot the RDFs. Compare them to the RDFs values provided in each production run folder (rdf\_solute\_Na.out, rdf\_solute\_Cl.out, rdf\_solute\_Water.out). Only use the data in columns 1 and 2 of the RDF output files. Ignore the other columns. The final RDF is obtained by averaging the RDFs from the six production runs.

## RDF plots

The figure below shows plots of RDFs from the UPM (from Example 4C) and SPM+LJ simulations for 100 mM RbCl. The total and excess number of Rb+ ions are also plotted.

