3SPN.2 LAMMPS Implementation

Daniel M. Hinckley[1] and Juan J. de Pablo[2]

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Maidon
[2]Institute for Molecular Engineering, University of Chicago

October 12, 2015

1. Introduction

3SPN.2 is an improvement on the previous version[5] of the **3-S**ite-**P**er-**Nucleotide** (3SPN) coarse-grained (CG) DNA model. Key improvements include the replacement of Gō-like interactions with angle-dependent potentials, a reduction in the magnitude of the explicit charge on the phosphate sites, and modification of the dihedral potential to increase the flexibility of ssDNA. These improvements remedy a number of limitations that were identified by the members of the de Pablo group and others. The resulting model can capture the persistence lengths of both ssDNA and dsDNA, proper melting temperatures for duplexes and hairpins, and has stable major and minor grooves. For further details please see Hinckley, D. H. et al.[4].

The 3SPN.2 CG model has been implemented in LAMMPS as a user package. This `USER-3SPN2` package contains this documentation file, examples of a number of simulations, a folder `DSIM_ICNF` that contains a configuration generator, and the source file to be added to the LAMMPS source. The following sections explain the particulars of the LAMMPS implementation, how to generate initial configurations for B–DNA, steps for running and visualizing trajectories, and limitations to this implementation. Extra features beyond the original 3SPN.2 publication are also discussed. Lastly, instructions are provided for compiling the serial version of LAMMPS with 3SPN.2. If you have any questions or problems, contact Dan Hinckley (dhinckley@wisc.edu).

2. Implementation Details

3SPN.2 consists of bond, angle, and dihedral bonded interactions as well as numerous nonbonded interactions. For functional forms of these potentials, we refer the user to Ref. [**?**]. The bonded interactions are implemented using existing LAMMPS potentials or slight modifications thereof. Bonds and angles are modeled using the `class2` and `harmonic` bond and angle styles, respectively. The Gaussian well dihedral potential is modeled using a new `3spn2` dihedral style (`dihedral_3spn2.cpp`). The base-stacking nonbonded interactions currently only occur between predetermined sets of sites. As such, they are implemented as a modified angle potential in `angle_3spn2_stacking.cpp`. Standard harmonic angles, as well as these "stacking" angles are both applied through angle style `hybrid`. The energies of both interactions are combined and output as `Eangle` in the `thermo` output.

All remaining interactions are captured by the `pair_3spn2.cpp`. When initialized, this pair style sets the value of equilibrium angles and distances. It also populates the arrays specifying the strength of interactions and any modulating parameters. Lastly, it also creates a base pairing array that assigns flags used to determine whether or not cross-stacking interactions are to be calculated. This is because a base that is base pairing with a base at the end of a DNA strand often has no base with which to cross-stack. In order to determine whether or not a base is at the 5' or 3' end of DNA, special types (types 7-14) are assigned to the bases at the 5' and 3' ends. This is similar to the notation used in all-atom force fields, with the obvious difference that in our coarse-grained representation the bases are topologically identical.

In the `3spn2` pair style, it is first determined whether or not the pair of sites (i,j) are on different molecules, separated by more than 3 nucleotides for the purposes of base pairing, and

if the base pairing flag is set. It is also determined whether or not the bases are separated by at least 5 sites. If base pairing is to be performed, the absolute indices of the sites i and j are used to get the indices of the neighboring sites that participate in the angle-dependent potentials. Then a instance of the BasePair object is created and populated with the instantaneous angles and distances. That done, member functions are called to calculate the cross stacking and base pairing interactions. If base pairing interactions are not present, excluded volume interactions are calculated. Electrostatics interactions are then calculated and the resulting force and that from excluded volume (if applicable), are then applied to sites i and j.

The energies are saved to a vector that allows for extraction of these energies using a compute. The `3spn2.in` input file displays file to screen as follows:

```
<step> <num. bp> <Ebond> <Eangle(harmonic and stacking)> <Edihedral> ...
       <Ebp> <Ecstk> <Eelectro> <T>
```

3. Coefficients for New Styles

The new styles have the following coefficients:

- `dihedral_3spn2.cpp` -
  `dihedral_coeff [dihedral number]` [K_ϕ] [ϕ_o] [$\sigma_{\phi,o}$]
- `angle_3spn2_stacking.cpp` -
  `angle_coeff [angle number]` [ϵ] [r_o] [θ_o] [α] [K]
- `pair_3spn2.cpp` -
  `pair_style [DNA type]` [T `(Kelvin)`] [I `(mM)`] `[Short Range Cutoff] [Coulombic Cutoff]`
  `pair_coeff 1 1` [ϵ][σ]

where `DNA type` can be `bdna`, `bdna/curv`, or `adna`. These correspond to B-DNA, B-DNA with intrinsic curvature, and A-DNA. Please see Section 8 for additional details. See the `examples/` directory for examples of the use of these styles.

4. Generating an Initial Configuration

A configuration generator for generating DNA according to the coordinates of Arnott et al. [1] is provided in `DSIM_ICNF/`. Navigate into the directory and type `make` to build `icnf.exe`, which generates all of the needed files for simulation and visualization. It takes the following arguments:

```
./icnf.exe <sequence file> <dna type> <complemenarity flag> <output directory> ...
    <ions flag> <ions file>
```

The sequence file is formatted as follows:

```
<NBPS>
<sense sequence (5'->3')>
<antisense sequence (3'->5')>
```

for example,

```
32
ATACAAAGGTGCGAGGTTTCTATGCTCCCACG
TATGTTTCCACGCTCCAAAGATACGAGGGTGC
```

If the antisense strand is not specified, it is assumed that you want completely complementary DNA. The `dna type` argument should be 0 for B-DNA, 1 for intrinsically curved B-DNA, and 2 for A-DNA (see Section 8 for additional details regarding the latter two topologies). The complementarity flag should be 0 for ssDNA and 1 for dsdNA. If the ions flag is 0, no explicit ions will be added. If the ions flag is 1 or 2, the ions file will be read. It is formatted as follows:

```
<Concentration of NaCl (mM)>
<Concentration of MgCl2 (mM)>
<Box Scaling factor (default 1.0) (mM)>
<Cubic box side length for ions only>
```

The concentrations are self-explanatory. The box scaling factor allows the user to add more or less ions padding around the DNA. The last line corresponds to the dimensions of the box if only ions are desired (ions flag = 2). Note that the DNA sequence file is needed even if a box containing only ions is desired.

Three files are generated by `icnf.exe`. The first is `in00_cvmd.psf`, a topology file that can be used in VMD to visualize the initial configuration. The second is `in00_conf.xyz`, an .xyz file containing the coordinates of each coarse-grained site. The last file is `conf_lammps.in`, the topology file used in LAMMPS.

After running `incf.exe`, it is possible to visualize the configuration in VMD. This is done with the following command:

```
vmd in00_cvmd.psf in00_conf.xyz
```

To capture the correct excluded volume of each site, go to Extensions>TkConsole and type `source /path/to/spheres.vmd`. The `spheres.vmd` file is also found inside `USER-3SPN2`. Then, if you set the visualization style to VDW, you will have the correct excluded volume of the sites.

5. Examples

Included in `USER-3SPN2` is an `example` directory containing examples for simulations of B-DNA. It also includes examples for simulating B-DNA with intrinsic curvature, A-DNA, and B-DNA with ions. For further details regarding these non-standard models, please Section 8.

6. Generating a Movie

After performing a simulation, it is easy to visualize its trajectory in VMD with the following command:

```
vmd in00_cvmd.psf traj.xyz
```

As before, source the `spheres.vmd` script to use the correct excluded volume. If you would like to render a particular frame of the trajectory, it can be done by going to File>Render. To make to a movie, go to Extensions>Visualization>Movie Maker. Most of the time it is desirable to only render a fraction of the total frames. The other frames can be dropped by right clicking on the trajectory in the main window.

7. Limitation to the 3SPN.2 LAMMPS Implementation

There are a number of limitations to the LAMMPS implementation that the user should be aware of. These are enumerated below:

(1) 3SPN.2 assumes that the 14 DNA types are listed first in your data file. Listing these types after other site types or neglecting the special site types for bases located at the ends of each ssDNA will lead to failure of the simulation.
(2) 3SPN.2 is performed in an implicit solvent. Consequently, the simulation box is mostly empty. This makes parallelization via LAMMPS's spatial decomposition horribly ineffective. 3SPN.2 will not scale well on multiple processors unless dense DNA arrays or explicit ions are being simulated.
(3) The virial (fdotr) is not currently being calculated properly.

(4) The Langevin integrator is specified with a damping constant (inverse viscosity) that has not been rigorously examined. It should not affect thermodynamics but it will influence dynamics.

8. Beyond 3SPN.2

The latest version of `USER-3SPN2` contains additional features that were not part of the original 3SPN.2 paper. These include explicit ions, intrinsic curvature, and A-DNA. The explicit ions are currently modeled using the CG representation presented by Freeman *et al.*[2], originally developed for 3SPN.1. The 3SPN team is actively working on an improvement to this ions model which incorporates improved data from all-atom simulations. As the present explicit ions model was not developed specifically for 3SPN.2, it should be viewed as experimental; use at your own risk. The next ions model should be available by July 2014.

Intrinsic curvature has been added to 3SPN.2 in order to simulate dsDNA interacting with proteins. For additional details, please refer to 3SPN_2C.pdf and Ref. [3].

Lastly, at the request of one of our users, a version of 3SPN.2 has been created for simulating A-DNA. The equilibrium distances and angles in the bonded and nonbonded interactions have been adjusted to represent A-DNA according to [1]. The energies of all interactions are unchanged from 3SPN.2's B-DNA. It is unknown how well these energies capture the physical behavior of A-DNA, in part due to the lack of experimental data of persistence lengths and melting temperatures for such systems. Therefore, use this model at your own risk.

9. Compiling the source

The following instructions should be sufficient to compile the serial version of LAMMPS with 3SPN.2 for simulations without ions.

```
svn co svn://svn.icms.temple.edu/lammps-ro/trunk mylammps
cp -r USER-3SPN2 mylammps/src
cd mylammps/src
cd STUBS
make
cd ..
make yes-MOLECULE
make yes-CLASS2
make yes-USER-3SPN2
make serial
```

For simulations with ions, the `USER-MISC` and `KSPACE` packages should also be added before compiling. The Makefile may need to be modified in order to include an FFTW library.

References

[1] Struther Arnott, P. J. Campbell Smith, and R. Chandrasekaran. *CRC Handbook of Biochemistry and Molecular Biology*, 2:411–422, 1976.

[2] Gordon S. Freeman, Daniel M. Hinckley, and Juan J. de Pablo. *J. Chem. Phys.*, 135(16):165104, 2011.

[3] Gordon S. Freeman, Daniel M. Hinckley, Joshua P. Lequieu, Jonathan K. Whitmer, and Juan J. de Pablo. *arXiv preprint arXiv:1404.7726*, 2014.

[4] Daniel M. Hinckley, Gordon S. Freeman, Jonathan K. Whitmer, and Juan J. de Pablo. *J. Chem. Phys.*, 139:144903, 2013.

[5] Edward J. Sambriski, David C. Schwartz, and Juan J. De Pablo. *Biophysical journal*, 96(5):1675–1690, 2009.