# Porting the AMBER forcefield to LAMMPS–massively parallel molecular dynamics simulations of DNA

C.R. Grindon*, S.A. Harris*, T.J. Evans$, K.E. Novik$, P.V. Coveney$, C.A. Laughton*.

*Cancer Research Laboratories, School of Pharmaceutical Sciences, University of Nottingham, Nottingham NG7 2Rdleiver a.

$Centre for Computational Science, Department of Chemistry, Queen Mary, University of London, Mile End Road, London E1 4NS.

Molecular dynamics (MD) is proving to be a valuable method for analysing the structure and flexibility of DNA (1). Atomistic MD simulations of DNA are computationally very expensive and, using conventional algorithms, are limited in practice to around 10ns (2). There are important motions within DNA which are predicted to occur over longer timescales than this (Table 1).

Table 1:– Timescales for typical DNA motions.

| Time scale | Main types of internal motion |
|---|---|
| Picosecond | Short living motions and oscillations of atoms. |
| Nanosecond | Oscillations of small groups of atoms: sugars, phosphates, bases; bending and twisting of the double helix. |
| Microsecond | Winding and unwinding of the double helix; opening of base pairs. |
| Millisecond | Dissociation of the double helix; super helicity; overall rotation. |
| Second | Writhing; isomerisation; division of bacteria. |

## Porting of the Forcefield

LAMMPS (Large Atomic/Molecular Massively Parallel Simulator) (3) is a parallel MD code with accurate treatment of long-range electrostatic interactions (a particularly important consideration in DNA simulations) based on the PPPM (particle-particle/particle mesh) algorithm and Ewald summation to handle the periodic boundary conditions. The advantage of LAMMPS over other codes is its ability to run calculations on parallel machines with large numbers of processors without great loss of efficiency (4), increasing the size and complexity of systems one is able to study.

We have ported the AMBER forcefield (5), which is well established for the simulation of DNA dynamics, into LAMMPS. The aim of this study has been to test this porting by comparing the dynamics of the DNA dodecamer d(CTTTTGCAAAAG) as predicted by LAMMPS with dynamics data from our previous extensive analysis of this sequence using AMBER (6).

Minor changes to the LAMMPS code were required to cope with AMBER's treatment of 1-4 non-bonded interactions. Static energy calculations then showed excellent agreement between AMBER and LAMMPS (table 2).

Table 2: Static energy analysis of a representative structure of $d(CTTTTGCAAAAG)_2$.

| ENERGY TYPE | AMBER (Kcal/mol) | LAMMPS (Kcal/mol) |
|---|---|---|
| Bond | 0.0239 | 0.0239 |
| Angle | 399.8833 | 399.8833 |
| Dihedral | 438.7989 | 438.7989 |
| Total VDW | 2532.0560 | 2532.2143 |
| Total Electrostatic | -27167.3825 | -27167.0726 |
| Total Energy | -23796.6204 | -23796.1522 |

## Temperature Control

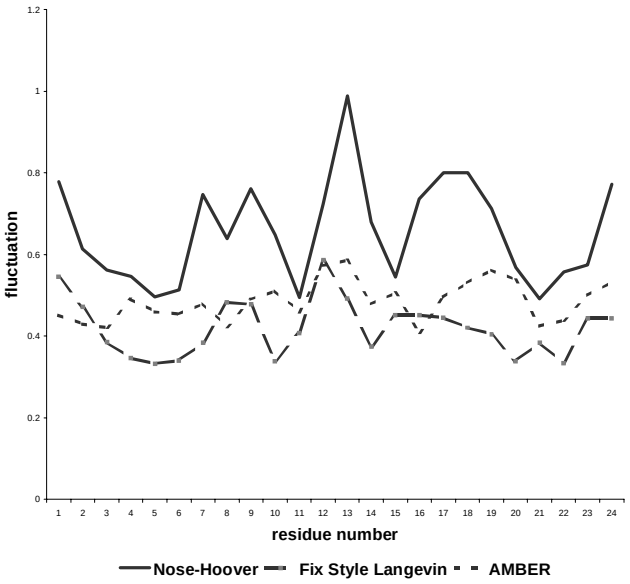
AMBER implements the Berendsen algorithm to control temperature but this is not available in LAMMPS. Two alternative temperature controls are available in LAMMPS, Langevin and Nose-Hoover. The latter approach was abandoned due to a hot solute/cold solvent problem. We have adopted the "Fix Style Langevin" temperature control where the solvent and solute temperatures are scaled separately.

Table 3: Average temperatures during 10ps runs using Nose-Hoover and Fix Style Langevin temperature couplings.

| | Average Temperature (K) | Standard Deviation |
|---|---|---|
| Nose-Hoover Solute | 320.4484 | 11.5720 |
| Nose-Hoover Solvent | 296.7024 | 5.8680 |
| Nose-Hoover Whole system | 299.7040 | 5.0824 |
| Fix Style Langevin Solute | 300.7899 | 8.9964 |
| Fix Style Langevin Solvent | 299.4550 | 3.9176 |
| Fix Style Langevin Whole system | 299.6238 | 3.5603 |

The hyperflexibilty of the DNA, particularly at the termini of the double helix, when Nose-Hoover temperature coupling was used are apparent from the analysis of atomic fluctuations (Graph 1).

Graph 1 – Atomic fluctuation data for 10ps runs of Nose-Hoover and Fix Style Langevin temperature couplings compared to AMBER.




## Scaling and Efficiency

We are using the 816 node Cray T3E-1200E supercomputer at CSAR, the UK's national supercomputing facility, to run dynamics. Short (10ps) simulations on up to 64 processors indicate that the code is relatively efficient (equation 1) and scales well, although not linearly, in comparison to AMBER which performs much more poorly in parallel situations (see graph 2).

Efficiency = np*t/l (Equation 1)
Where np = number of processors
t = time taken (s)
l = length of simulation (fs)

Graph 2 - Efficiency of LAMMPS compared to AMBER, normalised to the efficiency of LAMMPS on 1 processor.

## ANALYSIS OF LONGER SIMULATIONS

Three longer simulations (3ns+) have been run using 64 processors and analysis has been carried out on 2ns equilibrated portions of these. The three simulations differ only in temperature rescaling parameters, LAMMPS 1 = 0.01, LAMMPS 2 = 0.001 and LAMMPS 3 = 0.0001 (rescaling parameter in inverse time units ($fs^{-1}$) therefore LAMMPS 3 has most relaxed temperature rescaling).

Three analysis techniques were used to compare the LAMMPS simulations to an AMBER simulation previously carried out:

**1. RMSD.** RMSD's have been calculated between the time-averaged structure from AMBER and corresponding time-averaged structures from the three LAMMPS simulations.

**2. ENTROPY.** Configurational entropies (7) have been calculated to obtain an overall representation of the flexibility of these systems compared to AMBER.

**3. PRINCIPAL COMPONENT ANALYSIS (PCA).** PCA has been used to identify the major modes of motion in each trajectory. The similarity between these modes in the AMBER simulation and in each LAMMPS simulation was characterised by calculating the overlap of the top 10 modes (eigenvectors) (8).

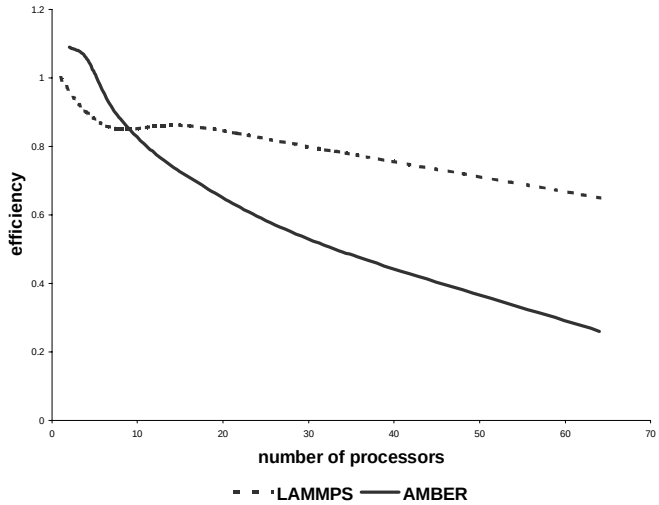


Table 4 – Characteristics of LAMMPS simulations compared to the original AMBER simulation.

| SYSTEM | RMSD (Å) | ENTROPY Kcal/mol | PCA OVERLAP |
|---|---|---|---|
| AMBER | n/a | 693.67 | n/a |
| LAMMPS 1 | 0.9483 | 589.79 | 0.4214 |
| LAMMPS2 | 0.6486 | 637.56 | 0.6309 |
| LAMMPS3 | 0.6297 | 646.17 | 0.6678 |

## Summary

The AMBER forcefield for DNA has been successfully ported into LAMMPS, this is confirmed by excellent agreement of static energies.

We have determined (some of) the optimal parameters for stable simulations of DNA in LAMMPS. We have shown that, given the same problem, LAMMPS scales better than AMBER on a T3E. We have shown that the time-averaged behaviour of the DNA is well preserved between the two simulation techniques and that dynamical characteristics are also well preserved. It is clear from this data that LAMMPS has great potential in this field. Given large scale access to the T3E at CSAR, we would expect to be able to probe hitherto unprecedented regimes of dynamical behaviour.

## Acknowledgements

We would like to thank CSAR for access to the Cray T3E, Steve Plimpton for his assistance and the BBSRC for funding of a Class 3 grant.

## References

(1) Sherer E.C., Harris S.A., Soliva R., Orozco M., Laughton C.A, *J.Am.Chem.Soc.,* 1999, **121**, p5981.

(2) Cubero E., Sherer E.C, Luque F.J., Orozco M., Laughton C.A., *J.Am.Chem.Soc.,* 1999, **121**, p8653.

(3) Plimpton S.J., 1997, LAMMPS version5, CRADA Collaboration, Sandia National Laboratory, USA.

(4) Maillet J.-B., Lachet V., Coveney P.V, *Phys.Chem.Chem.Phys,* 1999, **1**, p5277.

(5) Case D.A. *et al.* 1999, AMBER 6, University of California, San Francisco.

(6) Gavathiotis E., Sharman G.J., Searle M.S., *N.A.R.* 2000, **28**, (3) p728.

(7) Schlitter J., *Chem.Phys.Lett.,* 1993, **215**, (6) p617.

(8) Hess B., *Phys.Rev.E,* 2000, **62**, (6) p8438.