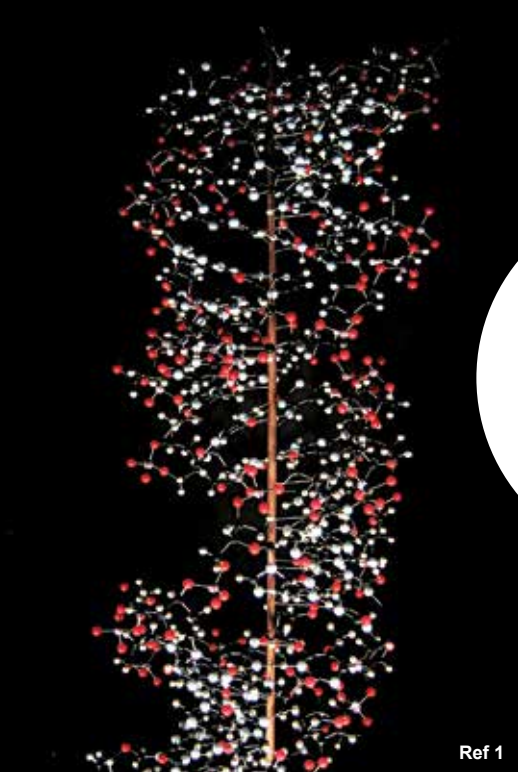

Ref 1

archer

LAMMPS: ENABLING NEW DNA SIMULATIONS ON ARCHER

The code

LAMMPS is a widely used molecular dynamics code developed by Sandia National Laboratories, USA. It contains several features which make it useful for both atomistic and coarse-grained simulation.

Imperial College London

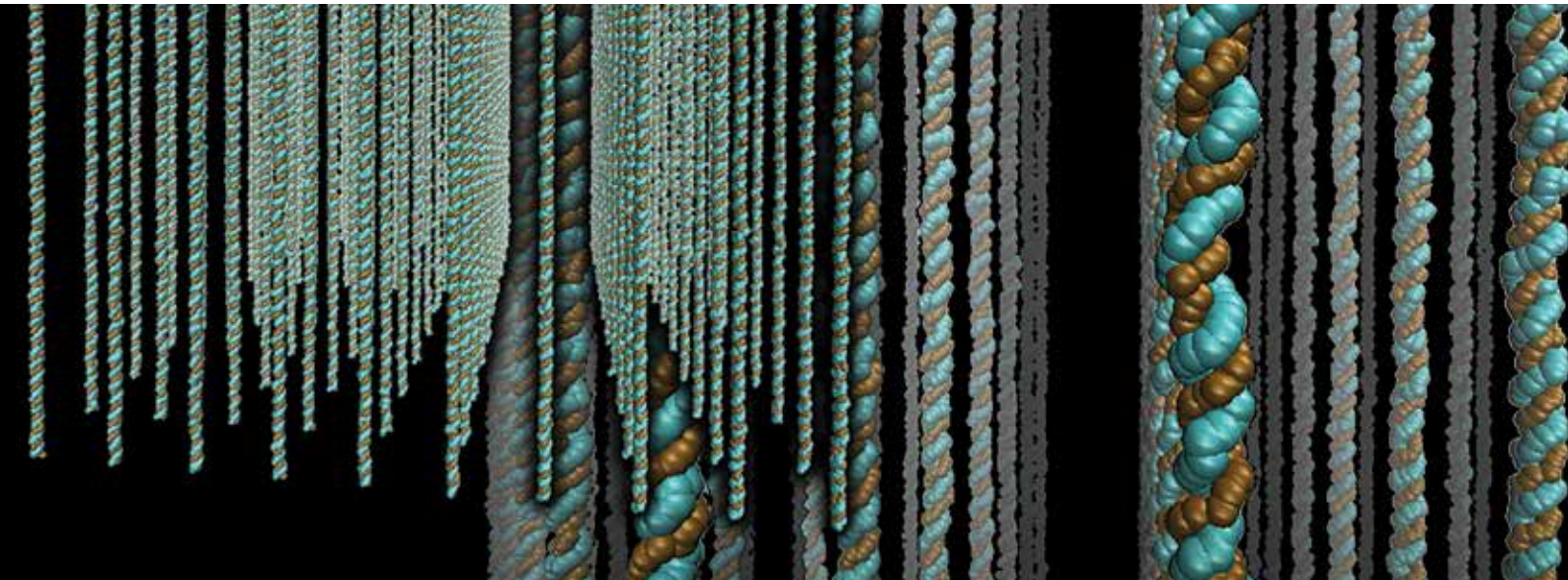
University of Strathclyde Glasgow

THE UNIVERSITY OF EDINBURGH

LAMMPS is specifically designed to be easily modifiable and extendable, which gives it a vital advantage over other popular biomolecular simulation software. The core of LAMMPS runs efficiently on both single processor machines and high-performance computing (HPC) facilities, which makes it ideal for many scientific areas of research including biology. Owing to its global user community new features are regularly implemented. EPCC has already worked with the code in the past, developing a Forward Flux Sampling module (EPSRC Grant No. EP/I030298/1) that can be linked to LAMMPS, as well as other codes.

There are several options available for atomistic simulations of DNA, but far fewer for coarse-grained simulation, particularly for DNA. Coarse-grained (CG) simulations have significantly increased in popularity in recent years, however. This type of simulation is used for DNA modelling on timescales in the microsecond range and beyond, or in modelling many or very long DNA strands. There is an increasing need for codes which can handle this type of simulation.

oxDNA is one of a small number of CG DNA models that has been developed in the past few years. It was previously only available as standalone software, which means it had a considerable entry barrier for new users. However, by implementing oxDNA into the popular LAMMPS code, researchers can now easily access one of the most powerful CG models for their simulations. It is available as USER-CGDNA package and distributed via the central LAMMPS repository at Sandia.

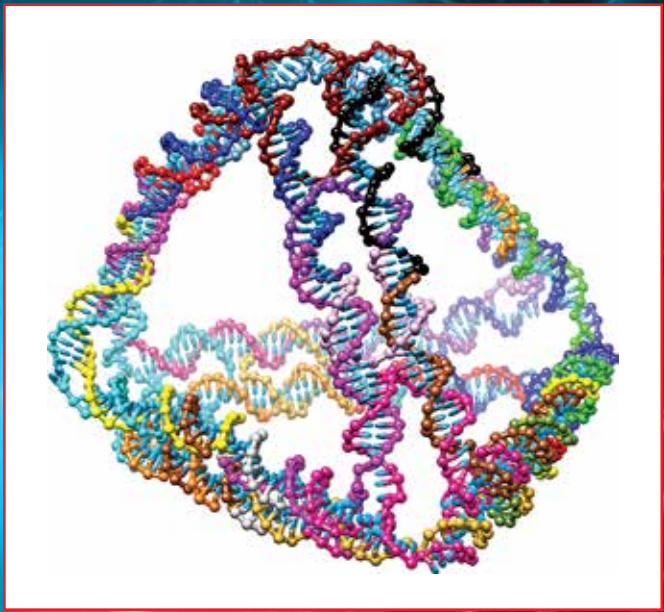


The justification

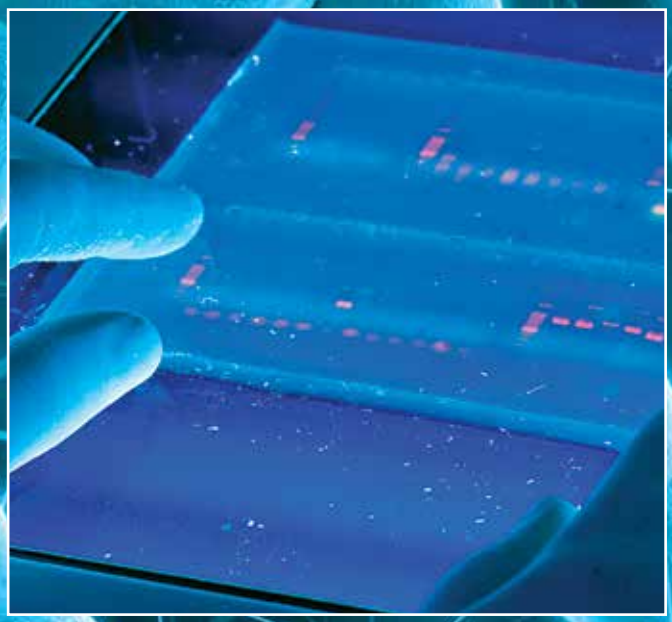
The world of genetics and genomics has exploded in the past decade, as new tools make research and analysis much more practical than it had previously been. Simulation-based approaches to DNA modelling are used across biomolecular research, to further understanding, discover how disease develops, and predict how changes in DNA will affect an organism.

Molecular modelling and simulation is the working horse within biomolecular communities; there are several community codes in use, including popular packages such as GROMACS, HADDOCK, and the software used here, LAMMPS.

The wider scientific benefit of this project lies beyond enhanced capabilities of oxDNA and the science that can be directly addressed through it. CG DNA modelling currently suffers from a lack of systematic comparison between the different CG models that have emerged to data and as well as between CG and atomistic models. This forms a formidable roadblock in current DNA-research. Without a common basis on which these models can be compared, further progress in the field of DNA modelling is likely to come to a halt. This project can form a starting point for multiscale modelling of DNA by combining different CG DNA models in one single simulation with the LAMMPS code as underpinning computational engine. This will be of great value for the large community of theoretical and experimental physicists, chemists, engineers and biologists interested in DNA and chromatin modelling, genetics and nanotechnology.



oxDNA representation of a nanostructured DNA tetrahedron that is formed from four 3-arm star tiles.



"There is a growing need for new ways to simulate DNA. Thanks to work done on ARCHER, a powerful new way of simulating DNA with LAMMPS is now possible."

The impact

With the work done within this eCSE project, users of the LAMMPS molecular dynamics code now have access to a powerful tool for coarse grained simulations of DNA.

There are three major research areas across biological research that the work of this eCSE will have an impact on.

Firstly, DNA nanotechnology often involves large systems containing many thousands of nucleotides. This is an emerging field which includes DNA origami, DNA bricks or compound tiles. With the previous implementation of oxDNA, only single DNA origami structures could be studied. However, the new code could be used to study the structure and mechanics of larger multi-origami nanostructures.

A second important field impacted is biological systems. This includes applications like the study of protein-DNA systems, DNA supercoiling and RNA hybridisation, as well as DNA-RNA interactions. In all these cases the DNA strands that need to be studied are often tens or hundreds of kilo base pairs long. The new code will allow the first ever study of the dynamics of supercoiling DNA domains, each of which is about 100,000 base pairs in size.

Another potential benefit of this work is for those who need to simulate smaller systems, up to a few thousand nucleotides. In these system, slow transitions opposed by large free-energy barriers are hard to measure. Sampling can be enhanced by techniques such as Forward Flux Sampling (FFS). One of the current bottlenecks of this method is that for more complicated processes, each individual simulation can take a long time to run. This will be much improved with the new code, as each of the simulations can be run as a parallel rather than a serial job. Example studies might include the assembly of DNA polyhedra in DNA-origami or the detailed investigation on DNA hairpin formation at intermediate temperatures or RNA self-hybridisation.

Contact:

Dr Oliver Henrich, University of Strathclyde
oliver.henrich@strath.ac.uk

Reference:

O. Henrich, Y. A. Gutierrez-Fosado, T. Curk, T.E. Ouldridge,
Coarse-Grained Simulation of DNA using LAMMPS, Eur. Phys. J. E 41, 57 (2018)

Some of the work described here was also carried out under an EPSRC Research Software Engineer Fellowship, EP/N019180/2

www.archer.ac.uk

About ARCHER

ARCHER is the UK National Supercomputing Service. The service is provided to the UK research community by EPSRC, UoE HPCx Ltd and its subcontractors: EPCC and STFC's Daresbury Laboratory, and by Cray Inc. The Computational Science and Engineering (CSE) partners provide expertise to support the UK research community in the use of ARCHER. The ARCHER CSE partners are EPSRC and EPCC at the University of Edinburgh.

The eCSE Programme

The Embedded CSE (eCSE) programme provides funding to the ARCHER user community to develop software in a sustainable manner to run on ARCHER. Funding enables the employment of a researcher or code developer to work specifically on the relevant software to enable new features or improve the performance of the code.

The Case Study Series

This case study has been produced as part of the ARCHER CSE service, supported by EPSRC research grant number EP/N006321/1.

Ref 1

Atomistic representation of about 1 1/2 turns of double stranded DNA