Bringing Scalable Molecular Dynamics to the CSAR Community

Neil Stringfellow
HPC Team, Manchester Computing, University of Manchester

Molecular dynamics (MD) is a computational method which calculates the time dependent behaviour of a molecular system. These systems can be of varying sizes from a few hundred atoms up to large scale systems using tens or hundreds of thousands of atoms, and the time scales for which results are required can be from a few picoseconds (10^{-12} s) to nanoseconds (10^{-9} s) and beyond.

There are several MD codes available on the CSAR systems such as DL_POLY, but the standard package for molecular dynamics simulations involving biological systems which is provided on the CSAR systems is Amber and the latest version of Amber which we have installed is Amber 7.

Amber is implemented using a replicated data model and this produces good performance and allows scalability up to around 16 or 32 processors. However, for large molecular systems which require the memory and computational power of more processors, Amber is limited in its scope.

With the expansion in computing power which has taken place in the last decade, several groups of computational scientists have looked at ways of designing new codes which can solve large problems and are scalable to large numbers of processors. In order to appeal to the wider biological molecular simulation community, these codes have been developed to use many features of some of the older popular codes.

Two new codes from the United States have been provided on the CSAR Origin machines, LAMMPS which was developed by a consortium headed by Sandia National Laboratories and NAMD which was developed at the University of Illinois. The standard input files to these codes consists of Protein Data Bank files and in addition both codes are able to provide compatibility with some Amber input files. The two codes are able to implement the Amber force field, and in addition NAMD can implement the CHARMM and Gromacs force fields and can read the files produced by the related codes.

Testing of the LAMMPS and NAMD codes has shown that they can scale very well, and in the case of NAMD this has been up to the full 512 processor capacity of Green. The scalability of these codes also applies to the number of atoms involved in the simulation and for NAMD tests were carried out on a variety of systems including one of over 300,000 atoms. Timings produced by inputs to NAMD have shown that 1 nanosecond (a million femtosecond steps) of simulation of a 92,000 atom system can be performed in about one and a half days real time, using 256 CPUs on Green. This brings real, large scale simulation of biological molecules within the grasp of the CSAR user community.

For more information about accessing these packages see the CSAR website ~ http://www.csar.cfs.ac.uk/software ~ or contact the CSAR Helpdesk.

Acknowledgements: Professor Peter Coveney, Queen Mary, University of London.