

Docking and Molecular Dynamics Simulations on a Grid Computing Platform

1. What research is being done by the researcher(s)

Computer programs which perform docking simulations have significant potential to contribute to biomedical research if the results of the simulation can prove consistent with the outcome of conventional wet laboratory experiments.

Carbohydrate recognition is a phenomenon critical to a number of biological functions in humans including responses of the immune system to pathogens and hyper-acute rejection of tissue transplants from non-human sources.

Unlike polypeptides and proteins, oligosaccharides do not form well organised tertiary structures in solution and they lack secondary structure motifs. In comparison to peptides, oligosaccharides are dynamic entities which occupy different conformations in space and over time. Molecular dynamics simulation packages such as AMBER and CHARMM can provide quantified data over time and conformational information in solvents. AutoDock which uses a search algorithm (e.g. Genetic Algorithm and Monte Carlo Stimulated Annealing) in combination with a free energy desolvation function, provides a quick method for identification of putative active/binding sites in recognition. AMBER and CHARMM are then used to dissect the mechanism of recognition and aid in the design of carbohydrate-based vaccines and other therapeutic agents.

Computer programs which perform docking and molecular dynamics simulations have significant potential to contribute to biomedical research if the results of the simulation can prove consistent with the outcome of conventional wet laboratory experiments. However, these simulations are computationally intensive and could take days or even weeks on a single PC to complete.

At present Babak Afrough is using Autogrid to generate the input files for Autodeck offline. Autodock is then run and the large number of output files are returned separately and merged later on with the help of Python scripts provided for the Autodock package. Hans Heindl uses the results from Autodock in AMBER to try different problems. Babak and Hans are the primary users of computing infrastructure in their group, although several members are involved in the same research.

They are currently using the Westminster NGS cluster resource along with the RAL resource, to manually run many of the steps in isolation. There are currently configuration and licensing issues which mean they cannot use the Leeds and Manchester nodes. This is a computationally significant problem in this domain and their initial allocation of 1000 hours was used up in under two weeks.

2. What are the issues that ENGAGE could address?

- 1) They currently access the applications on the NGS (Westminster and RAL). Providing a way of overcoming configuration issues would allow them to access more resources across different systems.
- 2) They would like to increase the number of people in their group and collaborating groups who can use the tools for their research. By accessing resources and software in a more tailored fashion novice users unfamiliar with grid or cluster technology can benefit. The NGS Application Repository provides some of this functionality, but the interface and paradigm it uses assumes familiarity with job

submission technology (e.g. JSDL). What is required is something which presents a different customized interface. Initial requirements analysis suggests that this could be done through:

- a. User interface designed for novice users (based on P-Grade, NGS Portal, ...)
 - b. Tailored application provision – creating predefined scenarios based on user requirements which allow variation of key parameters to achieve common tasks.
- 3) Create automation between components to improve the efficiency of the process
- a. Babak wanted data fed automatically from Autogrid to Autodock to python analysis scripts.
 - b. Hans would like to control the process himself, and let him try different things out.
- 4) Some kind of automation to run parameter sweeps. Most experiments they run are parameter sweeps where they will run batches of 2000 experiments (find how one molecular structure fits with another) – each Autodock run will say if it's a good guess or a bad guess. You need to analyse many runs before you know if a guess is good or not.
- 5) Provide feedback via the user interface on the current set of simulations, e.g. researcher will want to check outputs of how one Autodock simulation in an ensemble is going to understand if ensemble is likely to provide a good guess. Currently, only way to do this is using a GSI-SSH and tailing the files.

3. Future benefit to user community

In the short term, it will make common tasks easier to run on more resources. Additionally Westminster are organising a workshop for Autodock/Amber users from the UK which could be used to disseminate the results of the project.

In the longer term, it should help uptake of computing tools amongst novice users, and also increase the number of different scenarios which can be run by them.