m03.p04

Remote services for model building in macromolecular crystallography

Parthasarathy Venkataraman[a], Gerrit Langer[a],
Frank Schmitz[a], Anastasis Perrakis[b], Victor Lamzin[a]

[a]EMBL-Hamburg, Building 25A, Notkestrasse 85, 22603 Hamburg, Germany [b]NKI, Plesmanlaan 12A, 1006 CX Amsterdam, The Netherlands

Keywords: ARPwARP, remote services, model building

A server to run remote ARP/wARP jobs for automated model building in macromolecular crystallography was set up at EMBL-Hamburg, based on a 16 processor Linux Cluster. The project aimed to provide a user-friendly service for macromolecular crystallographers, with the following intentions: 1) to allow users to go far beyond routine data acqusition. 2) to obtain data for software development and project tracking 3) to provide users with the latest executables 4). to provide users with CPU power. The server was opened to the community in July 2004. This paper will describe the way remote job submission is carried out. An example will be given to lead the reader through the process. Statistics on the number of jobs run at the cluster, type and frequency of software and hardware errors, and a variety of other lessons learnt will be presented. Plans for scaling up and improvement of the services will also be described.

---

m03.p05

DNA & e-HTPX: High Throughput and Remote Access Protein Crystallography

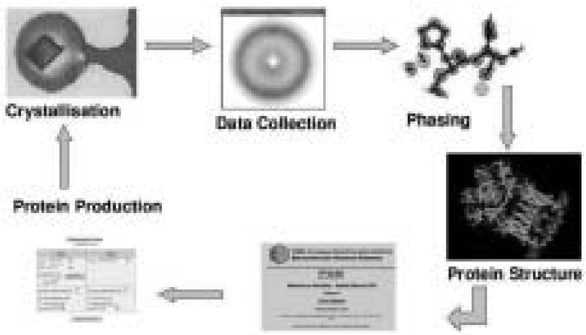
Graeme Winter[a], DNA Developers[abcdg], e-HTPX Developers[adefh]

[a]Daresbury Laboratory, Keckwick Lane, Warrington, UK. [b]ESRF, Grenoble, France, [c]MRC LMB, Cambridge, UK, [d]EMBL Grenoble, Grenoble, France, [e]EBI, Hinxton, UK, [f]York University, York, UK, [g]EMBL Hamburg, Hamburg, Germany, [h]Wellcome Trust Centre for Human Genetics, Oxford
E-mail: g.winter@dl.ac.uk

Keywords: high-throughput, Remote Access, automation

The DNA project has developed reliable automation for data collection and processing of macromolecular diffraction data. DNA consists of a number of separate modules, to abstract the beamline control, data processing, analysis and decision making, unified by a common communication standard. This has lead to a highly portable system, which is currently operating at the ESRF, SRS, EMBL Hamburg and NSLS.

The e-HTPX project aims to provide remote access to the entire protein crystallography pipeline through a standard web interface. This has been achieved by contribution to existing efforts, including the DNA project [1] and CCP4 [2], as well as development of new tools, including XIA [3] and Mr. BUMP. Web service (SOAP) interfaces to all of the key stages have been provided, as well as a single portal to unify the workflow into a ”1-stop shop” for PX.



To provide this kind of services, there are three main areas of development work needed. Firstly, it is essential to provide a simple but powerful and intuitive interface to the flow. Secondly, it is helpful to instill ”expertise” in the services, so that the user need only specify their requirements rather than micro-managing the entire process. Finally, development of the technologies, both at the end points to provide sufficient computing resources and in the distributed transactions (flexible data model, robust communications), is imperative, since the end user will anticipate 100% reliability. The challenges and solutions we have found will be described.

The development audience for e-HTPX has included academic and commercial groups, from low and high throughput laboratories.

e-HTPX is an e-Science Pilot Project funded by the BBSRC.

---

[1] http://www.dna.ac.uk
[2] http://www.ccp4.ac.uk
[3] Winter, G., ”Application of Automation to Data Processing & Analysis”, CCP4 Winter Newsletter 2005