Contribution ID: **84** Type: **Oral**

Grid assisted structure calculation of large protein systems in solid-state NMR context

Tuesday, 13 April 2010 11:15 (15 minutes)

Due to the present experimental limitations of solid-state NMR (ssNMR), 3D structure calculations of proteins using these NMR spectra as a source for structural constraints are demanding in terms of computing power. The application ARIA solid-state NMR is dealing with the automated assignment and structure calculation of large protein
systems in ssNMR context. We are currently adapting and evaluating that structural-biology application to grid computing within the EGEE infrastructure, and to cloud computing infrastructure using the HIPCAL and Eucalyptus systems.

Detailed analysis

The solid state NMR (ssNMR) is a valuable alternative to X-ray crystallography and solution NMR for structural analysis of insoluble proteins. Iterative assignment methods like ARIA (Ambiguous Restraints for Iterative Assignment), based on successive simulated annealing (SA) procedures, reliably assign NMR cross-peaks and calculate protein structures by using ambiguous distance restraints derived from the NMR cross-peaks intensities. The application of ARIA to ssNMR spectra is demanding in term of computing resources. Indeed, because of the low spectral resolution, a large number of possible assignments has to be processed. Also, it is difficult to derive precise distances from the peak intensities and that impairs the convergence. In order to face the higher complexity due to assignment ambiguity and the poor intensity quantification, several parameters have to be substantially increased, as the number of integration steps in the SA procedure, the number of protein conformations generated, and the number of possible assignments explored. These increases rely on a more intensive use of molecular dynamics softwares performing the SA, as CNS (Crystallography & NMR System).

Conclusions and Future Work

Solid-state NMR (ssNMR) spectroscopy can address questions on structure, dynamics and interactions of insoluble proteins. The large CPU resources of the grid computing allows a straightforward parallelisation of the conformer generations at each step of the iterative procedure. Adapting ARIA to grid computing infrastructure will thus help structural biologists to process highly complex NMR experiments. Future work includes the porting and integration of ARIA on both grid and cloud infrastructures, in collaboration between CNRS IBCP and Institut Pasteur Paris, and also through a collaboration with the eNMR EU project.

Impact

Adapting the ARIA method to the grid computing infrastructure will help structural biologists to study and assign large biological systems, as membrane proteins and protein fibrils using ssNMR, with more efficiency and reliability. Indeed, with the help of the grid it will be possible to enhance the convergence capability and the assignment reliability of the software by substantially increasing the complexity through the number of integration steps in the SA procedure, the number of protein conformations generated, and the number

of possible assignments explored. A typical iterative procedure requires 8 steps. Each step consists of the calculation of several conformations, ranging from 20 to 100 instances, using the CNS software. Between two steps, the ARIA software analyses the calculated
structures from the previous step and defines the new restraints used in the conformation calculations during the next step. Having access to a large scale grid platform such as EGEE opens the way to run in parallel several structure determination by several users, and/or increases the capabilities of structure and assignment procedures on large systems.

Keywords

Applications porting; Structural Biology; Grid Computing

URL for further information

http://aria.pasteur.fr

Primary authors: Mr JOSEPH, Alexandre (CNRS IBCP); Dr BOCKMANN, Anja (CNRS IBCP); Dr LOQUET, Antoine (CNRS IBCP); Dr BLANCHET, Christophe (CNRS IBCP); Dr MAREUIL, Fabien (Unite de Bioinformatique Structurale, URA CNRS 2185 et Institut Pasteur); Dr NILGES, Michael (Institut Pasteur); Dr MALLIAVIN, Therese E. (Institut Pasteur)

Presenter: Dr BLANCHET, Christophe (CNRS IBCP)

Session Classification: Bioinformatics

Track Classification: Scientific results obtained using distributed computing technologies