Efficient Explorer of Conformational Space of Biomolecular Systems

grid-CICADA

Jakub Štěpán,[1] Petr Kulhánek,[1] Zdeněk Kříž,[1] Jan Kmuníček,[2] Jaroslav Koča,[1] Luděk Matyska[2]

1) National Centre for Biochemical Research, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic, xstepan3@chemi.muni.cz
2) CESNET, z. s. p. o., Zikova 4, 160 00 Praha 6, Czech Republic, kmunicek@ics.muni.cz

Introduction

Knowledge of conformational space is essential in assessment of **dynamical behaviour of biomolecules**. Such information plays crucial role in recent computer aided drug design and protein folding studies. To explore the conformational space, deep insight into **potential energy surface** (PES) is necessary [1]. The PES expresses energy dependency on systems degrees of freedom. Unfortunately, biomolecular systems exhibit large number of degrees of freedom that makes the exploration of PES extremely complicated. To circumvent this problem, we have developed the program grid-CICADA that tries to **rationalize the conformational search** in such a space. We demonstrate the grid-CICADA utilization in distinct grid environments by the study of selected middle size biomolecules and subsequently we show analysis of multidimensional conformational space results.

grid-CICADA

Exploration of PES is targeted by driving of selected degrees of freedom. In grid-CICADA, only torsion angles are being drived because torsions are the most essential factors determining overall molecule geometry (**Figure 1**). Such driving is performed by optimising structure with selected torsion restrained to various angle values.

Current Challenges:

* Time demands of graph generation
* Results analysis and visualization

grid-CICADA Solutions:

* Parallel implementation usable in grid
* Full conformational space reduced to torsions space
* Sophisticated torsion abstraction

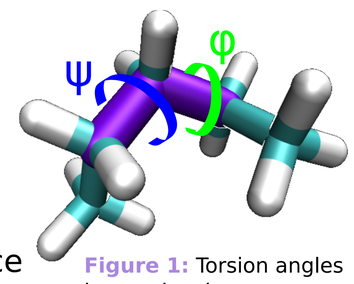

Figure 1: Torsion angles in a molecule

The grid-CICADA program represents PES as a graph *G*, with structures as vertices *V* and structure conversion paths between them as edges *E*. Calculation starts with single initial structure in the graph. As program proceeds the exploration of PES by torsion drivings, new structures and paths are added to the graph.

In proposed algorithm, the processing of one degree of freedom is completely independent on results of other degrees of freedom, which makes parallel processing a straightforward solution.

Parallel run of grid-CICADA consists of server process and a set of clients. The server holds the graph and provides two services: a) **Selector** and b) **Inserter**. The Selector selects a structure and a torsion that is subsequently processed by one of clients. As a result, the Inserter receives newly calculated vertices and edges from a client and incorporates them into the graph. In such a way, the server explores the whole graph. The scheme of client/server parallel grid-CICADA is in **Figure 2**.

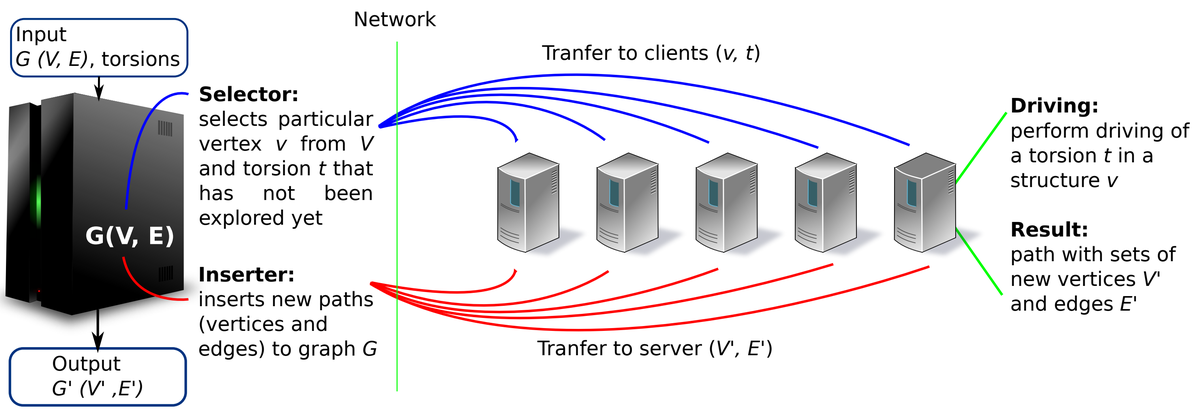

Figure 2: Parallel processing using client/server model. The server manages clients and results, which are obtained by individual drivings of particular torsions on selected structures.

To further reduce number of degrees of freedom, grid-CICADA introduces "coarse grained" torsion definition (**Figure 3**). Using this approach, larger biomolecular systems can be explored in resonable time as well.

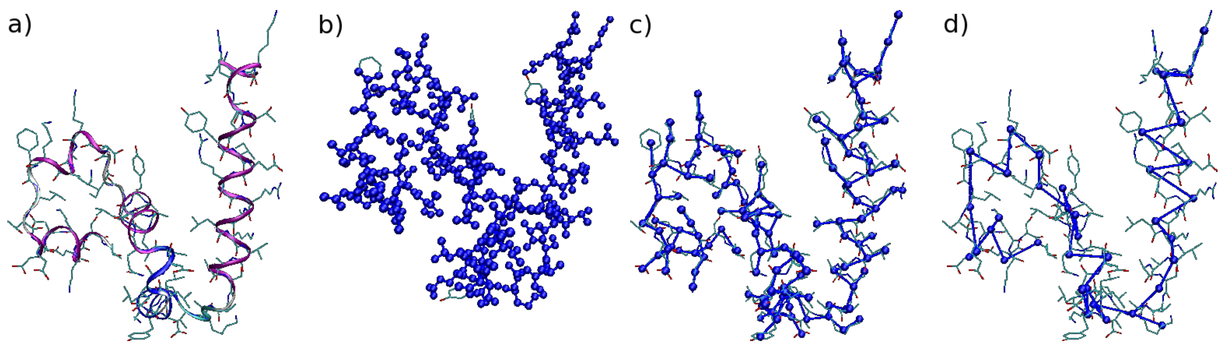

Figure 3: Protein structure a) studied biomolecule; torsion definition for b) all torsions, c) coarse grained abstraction employing 3 atoms per group, d) coarse grained abstraction employing 10 atoms per group.

Computational Resources

Computational resources for testing of grid-CICADA were provided by METACentrum, VOCE and EUAsia Virtual Organisations (VOs).

Project METACentrum (http://meta.cesnet.cz) covers majority of activities concerning cluster, grid and high performance computing in general in the Czech Republic. MetaCentrum operates and manages distributed computing infrastructure consisting of computing and storage resources owned by CESNET as well as resources of co-operative academic centres.

Virtual Organization for Central Europe – VOCE – (http://egee.cesnet.cz/en/voce) consists of computational resources and storage capacities provided by the Central European resource owners. VOCE serves for all academic researchers within the Central Europe region as defined by EGEE project.

EUAsia VO (http://www.euasiagrid.org) is build as catch-all VO service provided by resource owners within the Asian Pacific region. EUAsia VO is generic, application neutral VO that is not tightly bound to any specific application area and thus supports testing of new, promising applications before their large-scale deployment in the grid environment.

Results

In our study, we focused on conformational study of **Casein kinase I**, one of proteins that ensures cell signal transduction. The protein passes a signal by phosphorylation of serine and threonine amino acids.

grid-CICADA calculation produced the graph representation of explored PES, this graph was analysed by CLANS [2] cluster analysis program. Resulted PES visualisation is shown in **Figure 4**. Calculated PES points are clustered according to torsion values similarity. Clusters with high density of vertices are local minima.

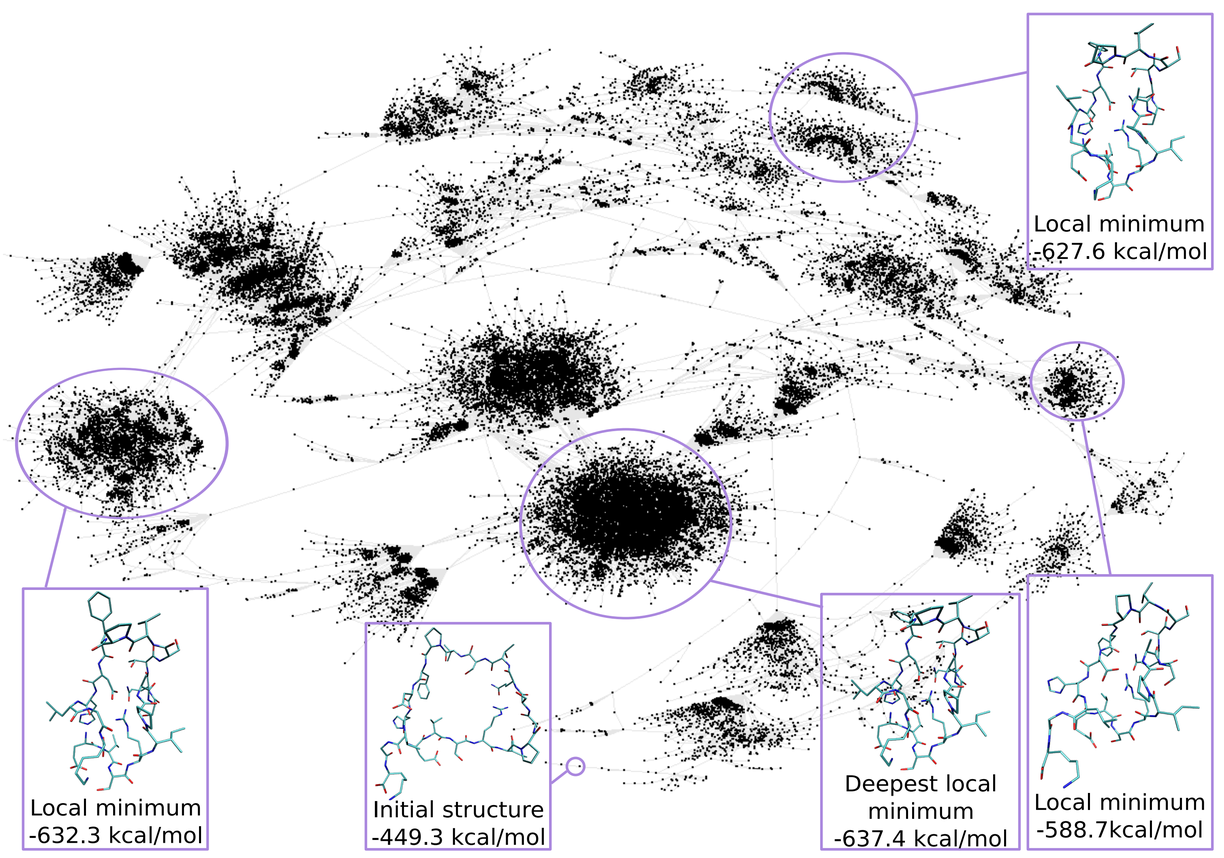

Figure 4: Result of grid-CICADA conformational search of Casein kinase I processed by cluster analysis by program CLANS. Selected structures are shown with energies.

grid-CICADA was executed in different environments: MetaCentrum, VOCE, EUAsia VO. Moreover, simultaneous run was performed using VOCE and EUAsia VO too. Significant acceleration of conformational search was achieved (**Figure 5** and **Figure 6**).

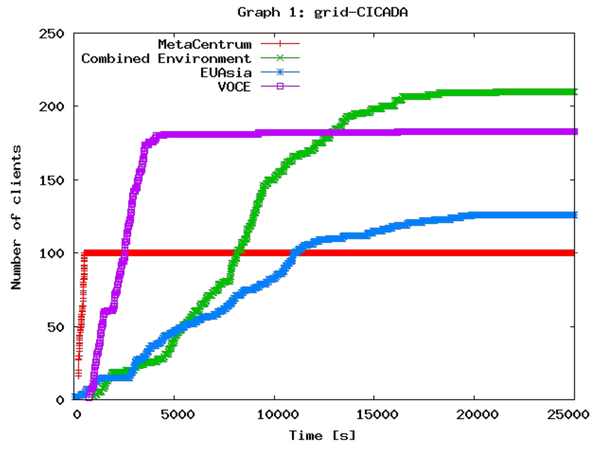

Figure 5: Number of grid-CICADA clients registered by the server over time.

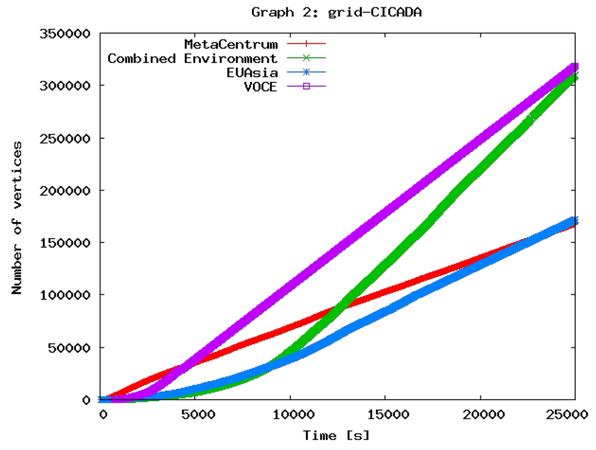

Figure 6: Number of explored vertices by grid-CICADA over time.

Conclusions

- ✓ **development and implementation of grid-CICADA program**
- ✓ **fast PES exploration of selected biomolecular system with analysis of achieved results**
- ✓ **efficient utilisation of heterogeneous computational resources from three different VOs**

Acknowledgements

This work has been supported by Ministry of Education of the Czech Republic (MSM0021622413, LC06030) and by EGEE III project funded by European Commision (contract number INFSO-RI-222667). The access to the METACentrum supercomputing facilities provided under the research intent MSM6383917201 is highly appreciated. The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement no. 205872.

1. Koča, J. Computer-Program CICADA - Traveling Along Conformational Potential-Energy Hypersurface. Theochem-Journal of Molecular Structure **1994**, *114*, 13-24.
2. Frickey, T.; Lupas, A. CLANS: A Java Application for Visualizing Protein Families Based on Pairwise Similarity. Bioinformatics **2004**, *20*, 3702-3704.