Distributed Way of Biomolecular Conformational Space Exploration - Experience of grid-CICADA Tool Utilization

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

kulhanek@chemi.muni.cz

1) National Centre for Biomolecular Research, Masaryk University
Faculty of Science, Kotlářská 2, CZ-61137 Brno,
Czech Republic

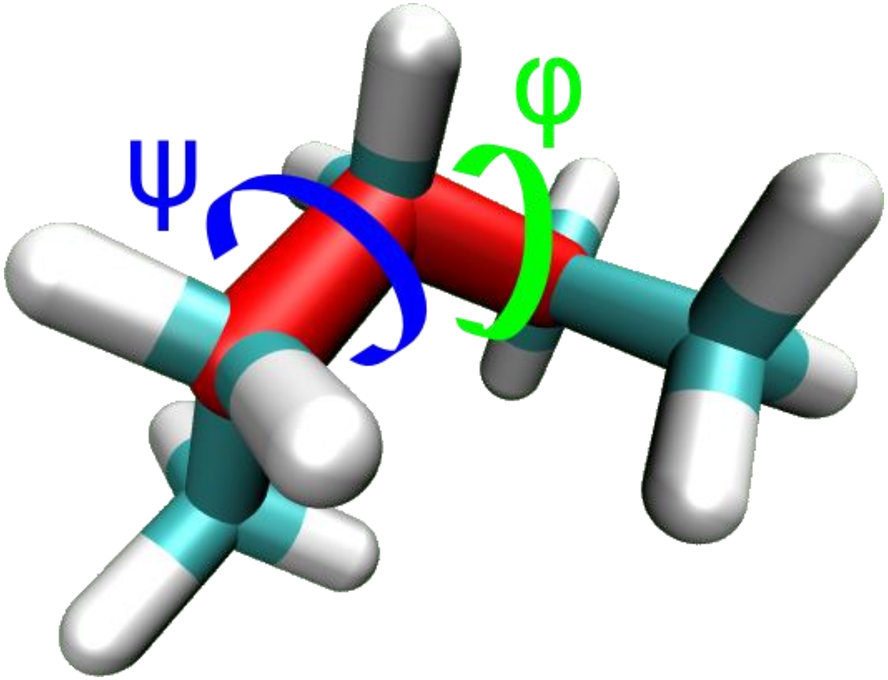
2) CESNET z.s.p.o., Zikova 4, CZ-160 00 Praha 6, Czech Republic

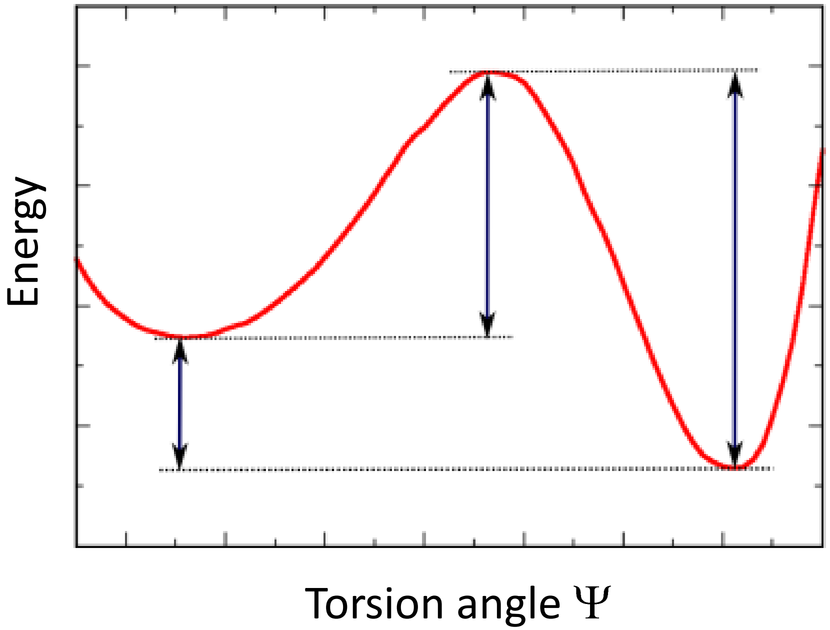
Outline

- Molecular Conformational Space
- CICADA and grid-CICADA
- Experience (VOCE, EUAsia, MetaCentrum)
- Summary
- Future Directions

Molecular Conformational Space

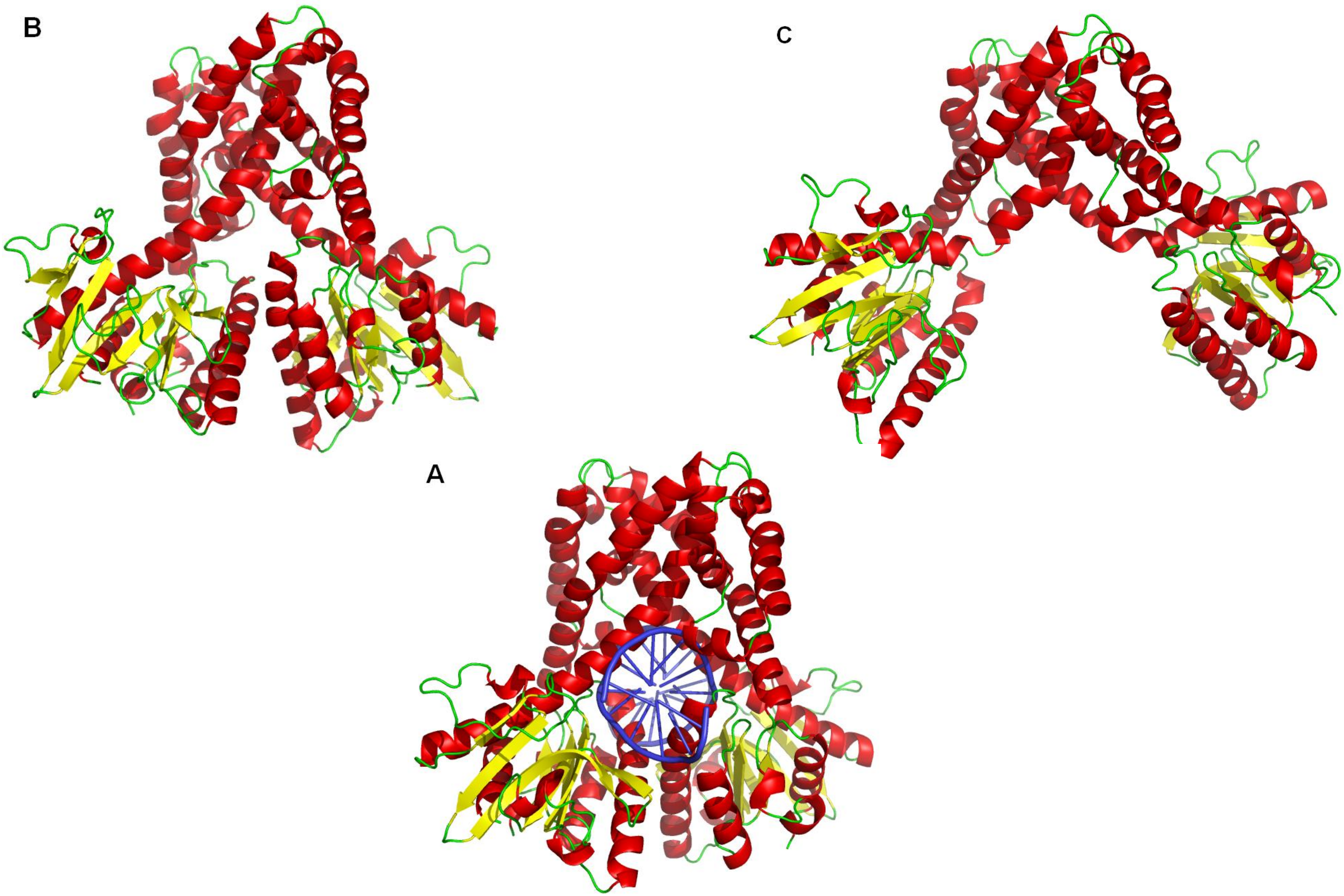
Conformational Space





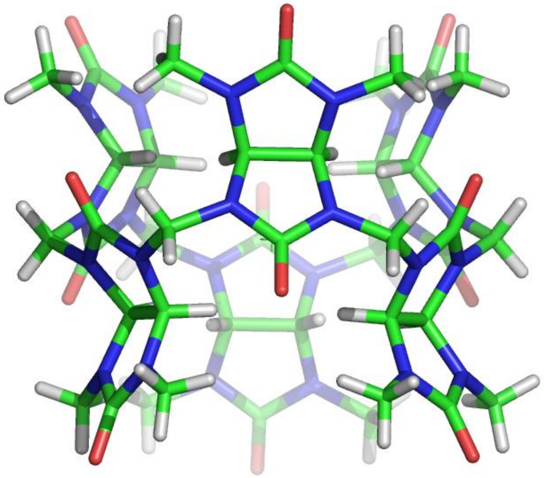
- **conformational space is very large**
- **only some parts are important for chemical processes**

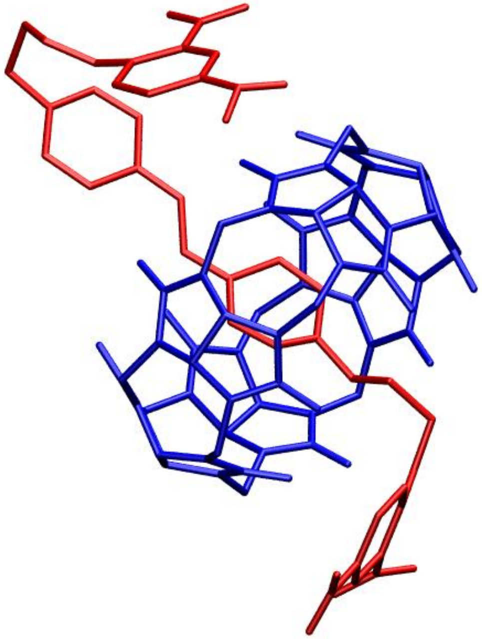
Conformational Space



Our goals

- to efficiently explore conformational space
- only small or middle size molecules (supramolecular chemistry)

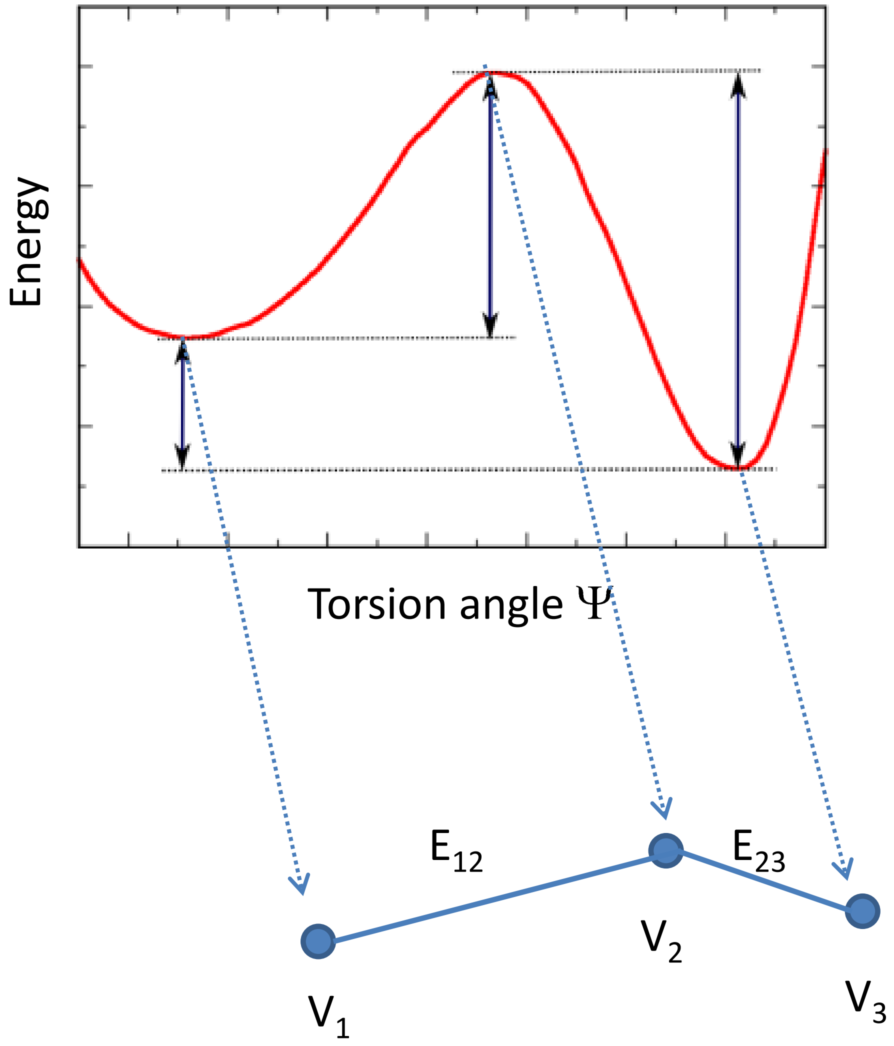


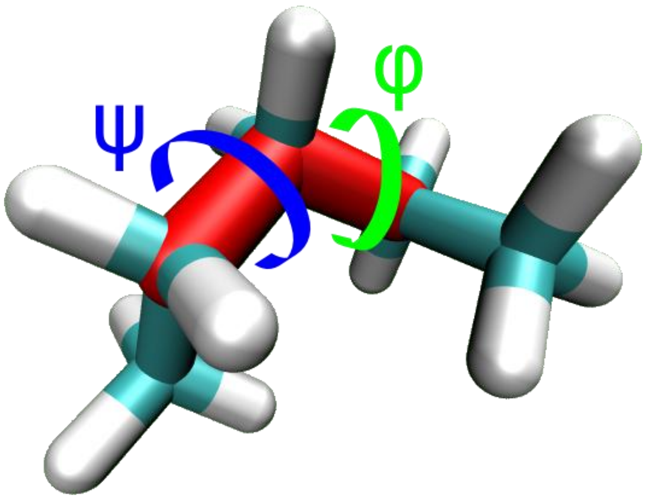


CICADA

wikipedia.org

Single coordinate driving

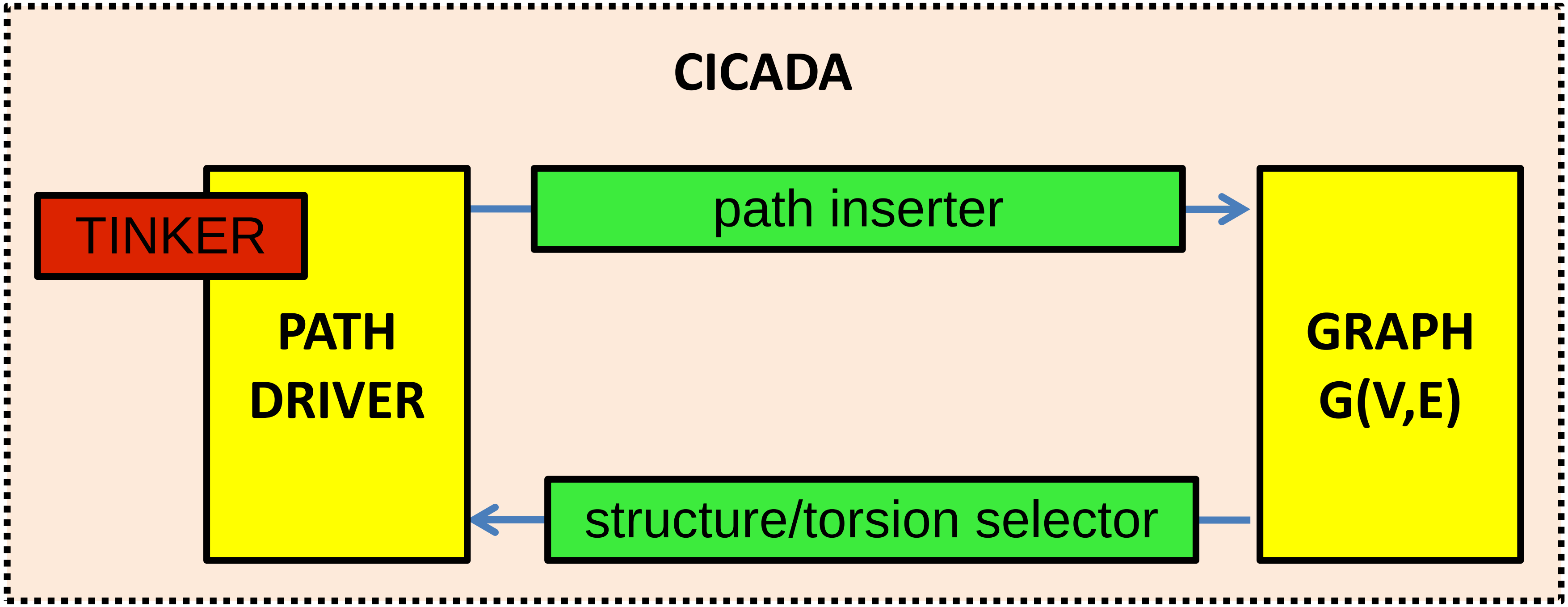




“soft” driving by imposed restraint

structure + status of used torsions

CICADA



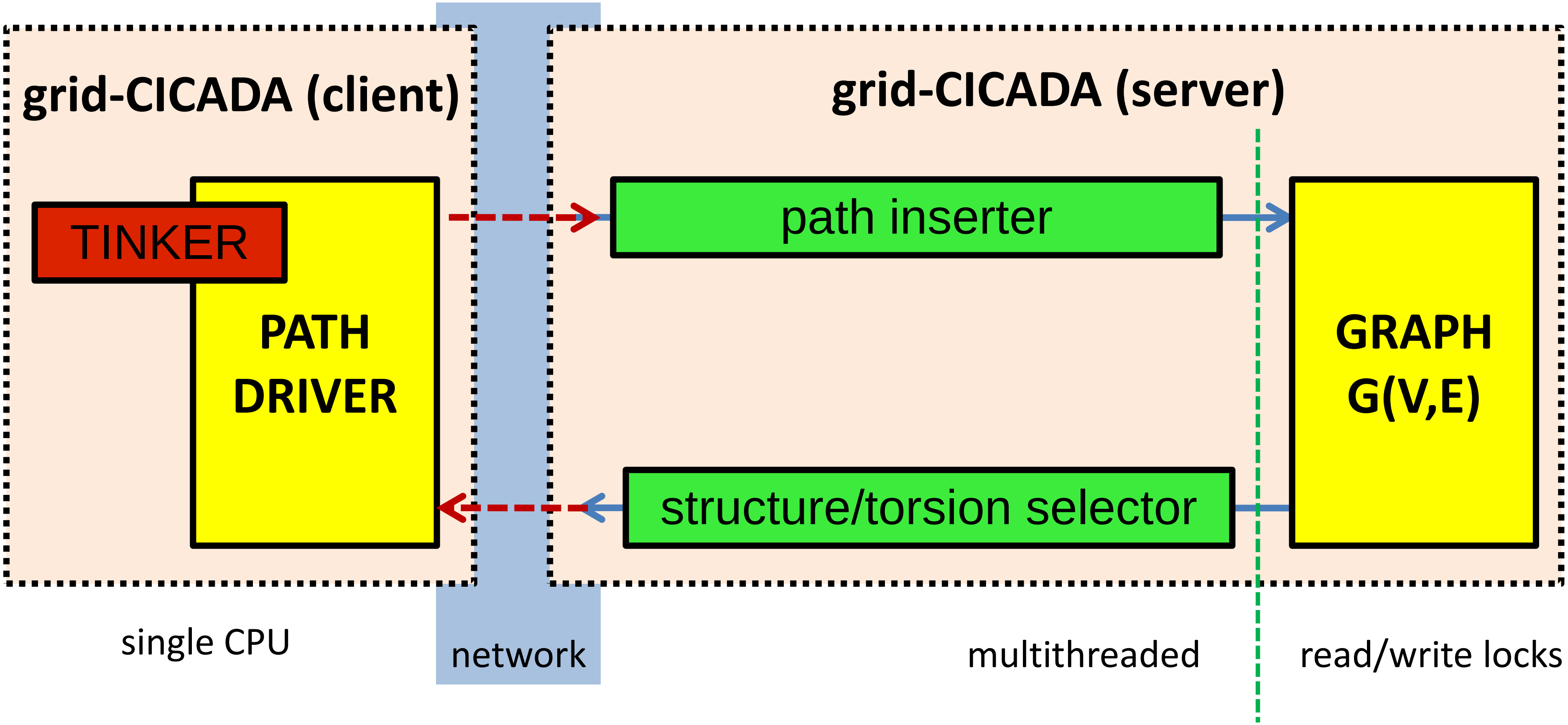
Structure/torsion selector:

- systematic
- lowest energy
- random

Path inserters:

- simple
- match similar (in torsion space)

grid-CICADA


grid-CICADA (client)

TINKER

PATH DRIVER

grid-CICADA (server)

path inserter

structure/torsion selector

GRAPH G(V,E)

single CPU

network

multithreaded

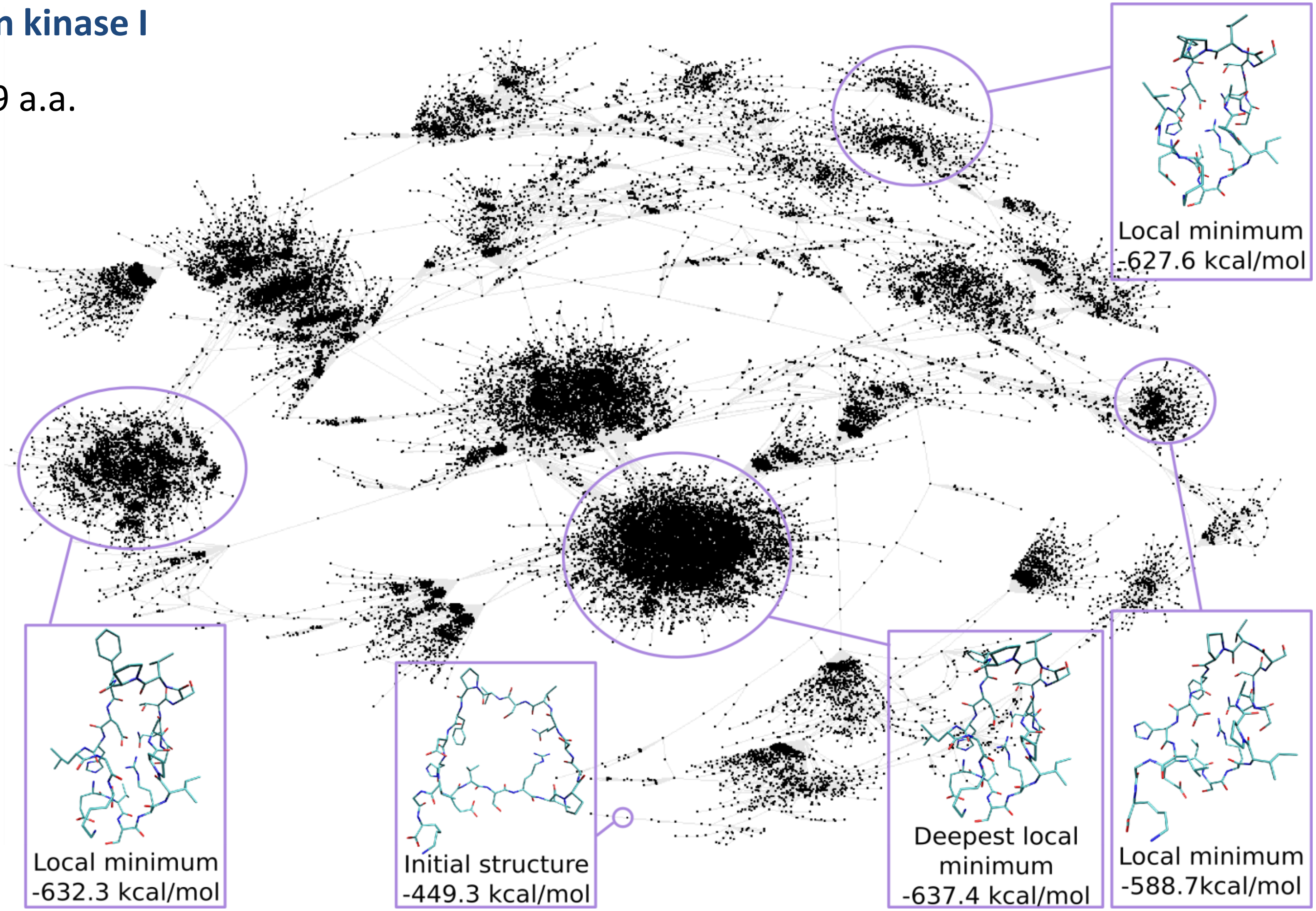
read/write locks

Practical Experience

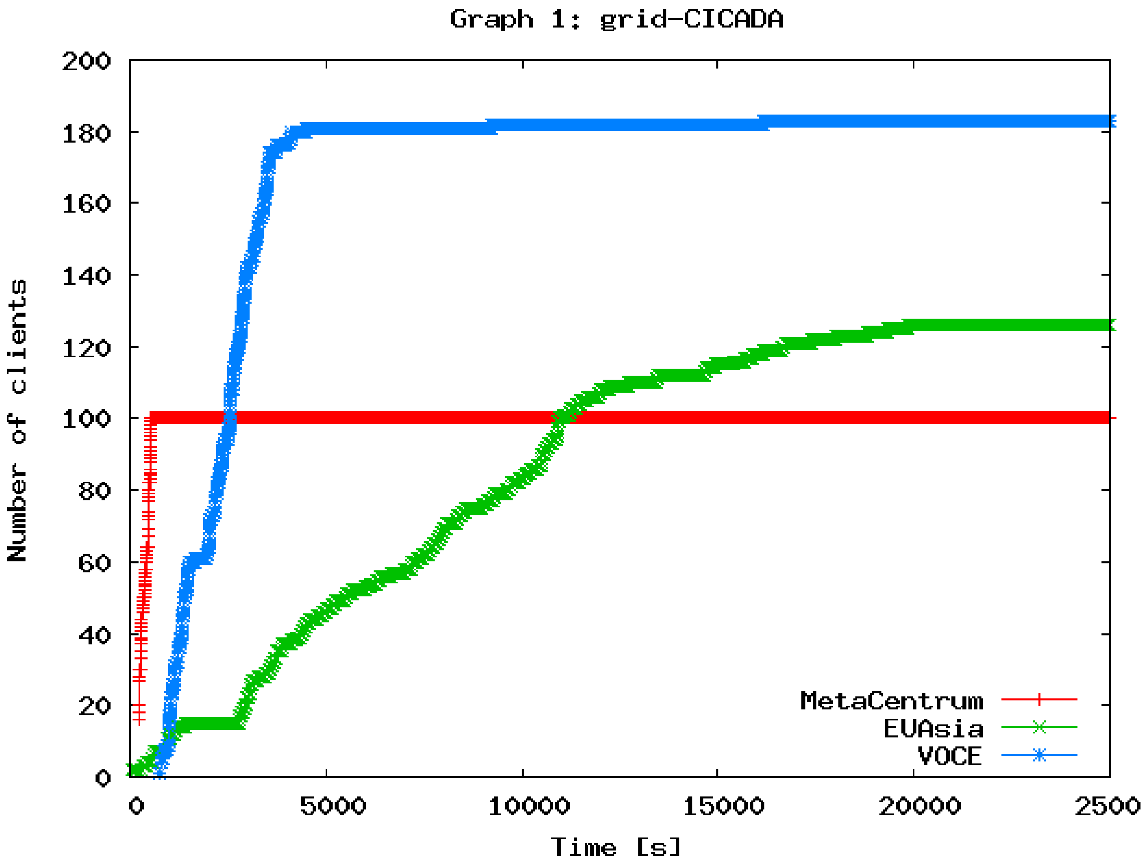
Experience

Casein kinase I

19 a.a.



Experience with various GRIDs



Every job was submitted individually to the grid environment.

Experience with various GRIDs

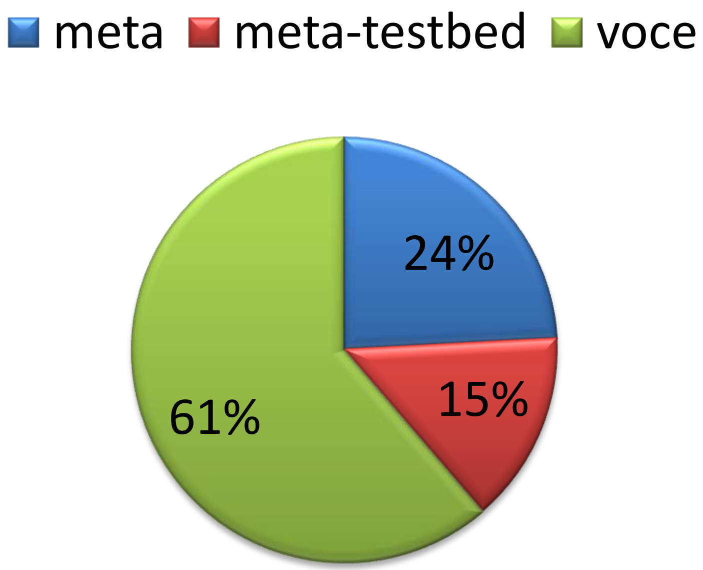
- clients only need to know where the server is running ...
- parametric jobs are now used to lunch all clients

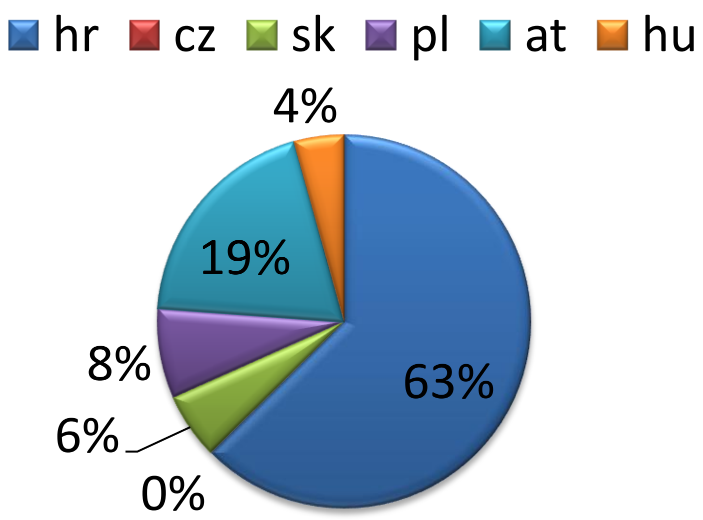
```
# parametric job
[
JobType = "Parametric";
Executable = "run-all";
StdOutput = "output_PARAM_.txt";
StdError = "error_PARAM_.txt";
Parameters = 1000;
ParameterStart = 1;
ParameterStep = 1;
InputSandbox = {"run-all","run-infinity",
                "run-client", "pa2l.config", "pa2l.pdbqt",
                 "server.key", "infinity.key", "extract.awk"};
OutputSandbox = {"output_PARAM_.txt", "error_PARAM_.txt"};
MyProxyServer = "myproxy1.egee.cesnet.cz";
]
```

Experience with various GRIDs

Similar problem / same infrastructure (virtual screening, two days test)

- autodock vina
- single docking ca 1-10 minuts/ligand
- single client job length ca 22 hours
  (myproxy is required)
- 2500 jobs in total
- ca 900 concurrently running jobs
- docking speed about 250 000 ligands per day



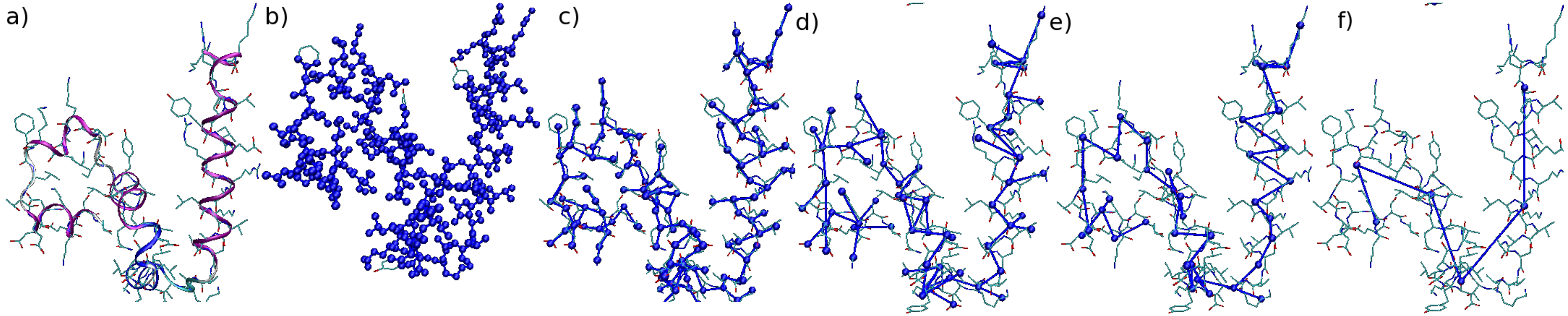


Summary

- We re-implemented CICADA in client/server fashion.
- We successfully show its utilizations in various grid and cluster environments.
- Current implementation is able to handle up to 1000 concurrently running clients.

Future Directions

- better selectors and inserters (with some heuristic graph analysis)
- stochastic driving (Monte Carlo)
- dynamics and coarse-grained torsion space



Acknowledgements

- **prof. Jaroslav Koča, prof. Luděk Matyska**
- **Jakub Štěpán, Zdeněk Kříž, Jan Kmuníček**
- **POSTBIOMIN (financial support)**
- **METACentrum , VOCE, and EUAsiaVO (computational resources)**