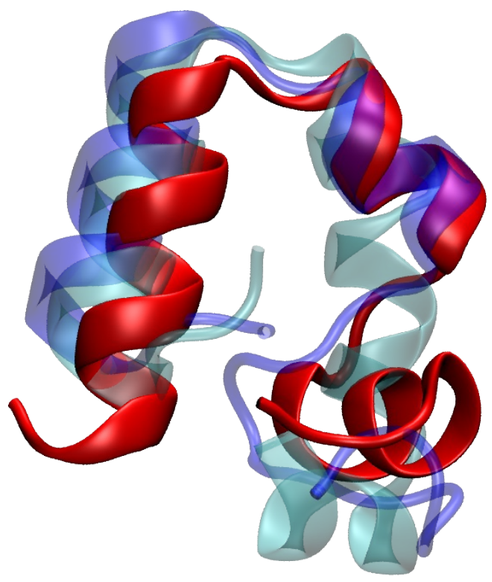
Approaching protein folding using the EU-IndiaGRID infrastructure

Riccardo Di Meo[1], Fabio Pietrucci[2], Alessandro Laio[2], Stefano Cozzini[1,3]

[1]ICTP EU-IndiaGRID, [2]SISSA-ISAS Trieste, [3]Democritos National Simulation Center

The Bias exchange algorithm

Bias-Exchange Metadynamics allows reconstructing the free energy of complex systems as a function of a large number of reaction coordinates:
several replicas of the system are evolved in parallel by molecular dynamics, biasing a different reaction coordinate in each replica, and allowing the replicas to exchange the biasing potential from time to time.



This technique, implemented in the Gromacs code, has proved successful in simulating the folding in explicit solvent of the Trp cage (20 residues) and Advillin (36 residues).

Our implementation.

1. A server receives the connections from the clients and appoints them a task to do; then the client retrieves the data files and starts the execution

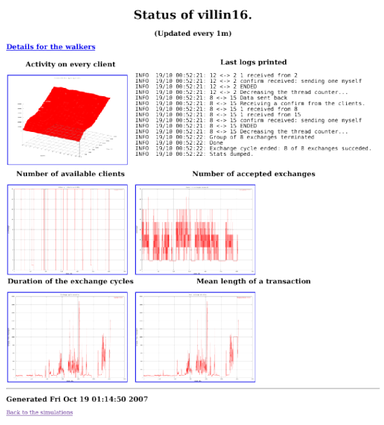
2. The server synchronizes the simulation triggering the bias exchange between the available clients and bridges the data between different CEs.

3. The server is designed to handle the disconnection of a client and the availability of new ones and to recruit them dynamically.

Extra features.

A logging service integrated in the clients sends in **real time** information about the simulation from the WNs to the UI

A custom userfriedy interface and a small web-server Interface ease the task of setting up and running a simulation.



Bibliography.

•S. Piana and A. Laio, J. Phys. Chem. B 111, 4553 (2007).

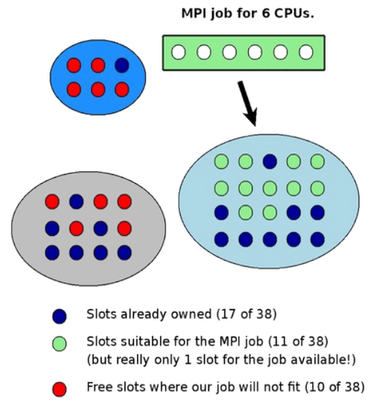
•S. Piana, A. Laio, F. Marinelli, M. Van Troys, D. Bourry, C. Ampe, and J. C. Martins, J. Mol. Biol. (in press).

•The EU-IndiaGRID project site at ICTP: www.euindia.uctp.it/bemuse

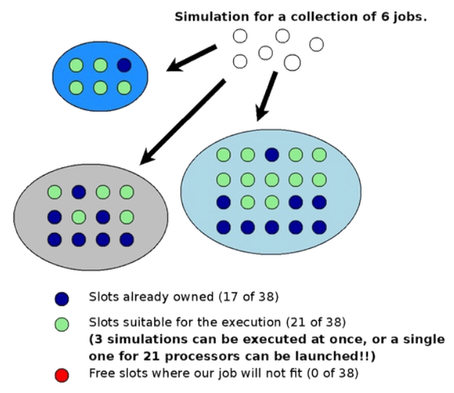
•The homepage of Alessandro Laio at SISSA http://people.sissa.it/~laio/metadynamics.htm

Going beyond MPI.

Normal MPI submission:
only one CE available for every job!

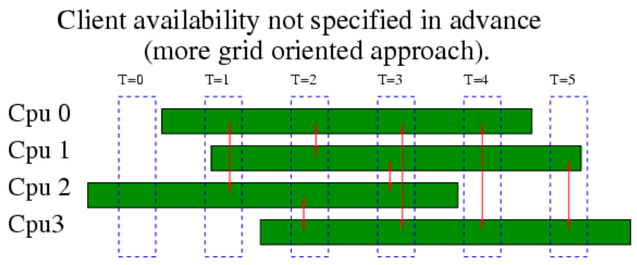


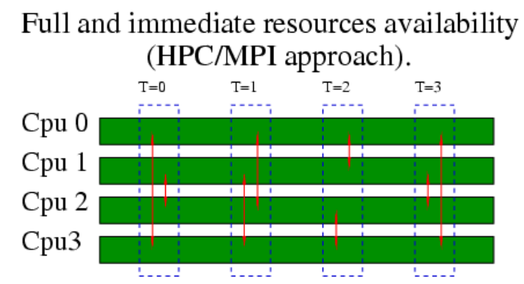
Custom networking:
all free resources are available!



We chose to extend the original MPI implementation in order to exploit the opportunity offered by the GRID

In the original MPI impl.
all CPUs start at the same time, have the same hardware and similar input files.



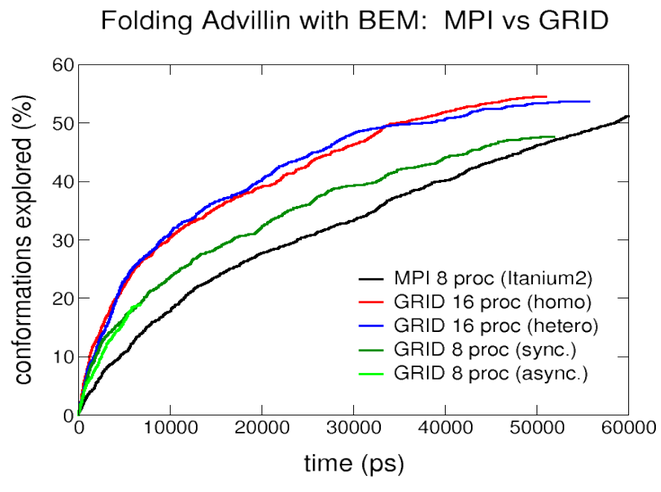


With our approach, exchanges are still synchronous, but the CPUs arrive at different time, have different speed and they may be mixed with other which ran for **very different** intervals!

Results.

•We test the approach on the advillin protein at 300 K using the AMBER all-atom force field and TIP3P explicit water, biasing 5 general reaction coordinates (the helical content, the number of hydrogen bonds,etc.)
•We perform the following runs:

• two simulations (on 8 and 16 homogeneous CPU), scheduled in synchronous cluster like way.
• a short asyncronous/homogeneous simulation on 8 CPU, where every process starts at intervals of 2 hours .
• a 16 CPU simulation on heterogeneous hardware, scheduled asynchronously, submitted automatically using the grid.
Our simulations were able to **explore the conformations of the protein at a speed comparable to MPI ! (see figure)**
Starting from an extended conformation, within 50 ns /CPU the protein RMSD from the experimental folded state is lower than 3.5 A.



Within our approach we have been able to solve the problem of folding a 36-aa protein at a computational cost (< 1 year of single CPU time) far lower than in other GRID-based techniques: Folding@Home [V. S. Pande, Biopolymers 68, 91, 2003] required 1000 years of single-PC time at a lower level of accuracy (implicit solvent).