Protein Molecular Dynamics using Quaternions and Monte Carlo Methods

Claudia Caudai
ISTI - CNR Pisa

Bioinformatiha 5
Firenze, 28 Ottobre 2016

Research funded under the flagship Project "InterOmics" (PB.P05)

InterOmics
CNR

Outline

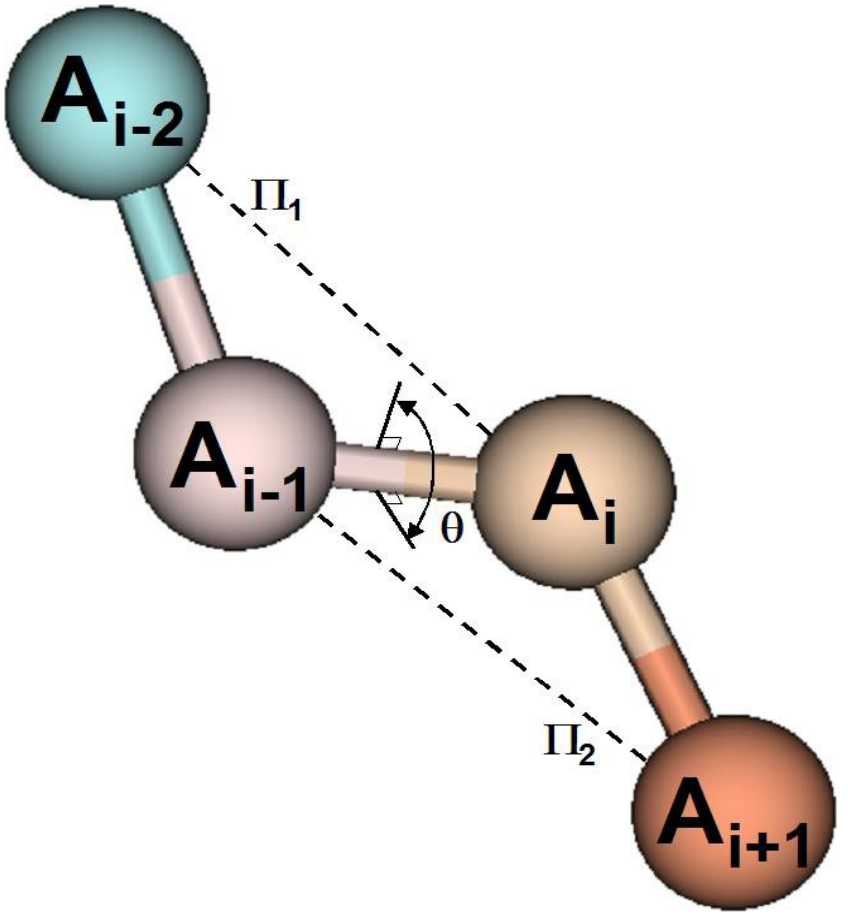
- **Introduction**
  - Classical MD
  - Atoms movements in proteins
- **Method**
  - Dynamical models of proteins with quaternions
  - Sampling conformations with Monte Carlo method
  - Program Flow
- **Results**
  - Experiments on Calmodulin (1cfc)

Classical Molecular Dynamics (MD)

- Computer simulations used for studying the physical movements of atoms and molecules.
- The trajectories of atoms and molecules are determined for a system of interacting particles.
- Molecular (protein) systems are typically very complex:
  - large number of particles (impossible to solve them analytically)
  - numerical methods take very long calculation time
  - All atoms / Coarse grained
  - Force fields

Atoms Movements in Proteins

- Atoms are connected to one another through **covalent bonds**. Their separation distance is called the **bond length**.
- A set of three atoms bonded in sequence defines the angle between the two adjacent bonds, called **bond angle**.
- The most important internal degree of freedom is rotation about **dihedral angles**, defined by four consecutively bonded atoms.



InterOmics CNR

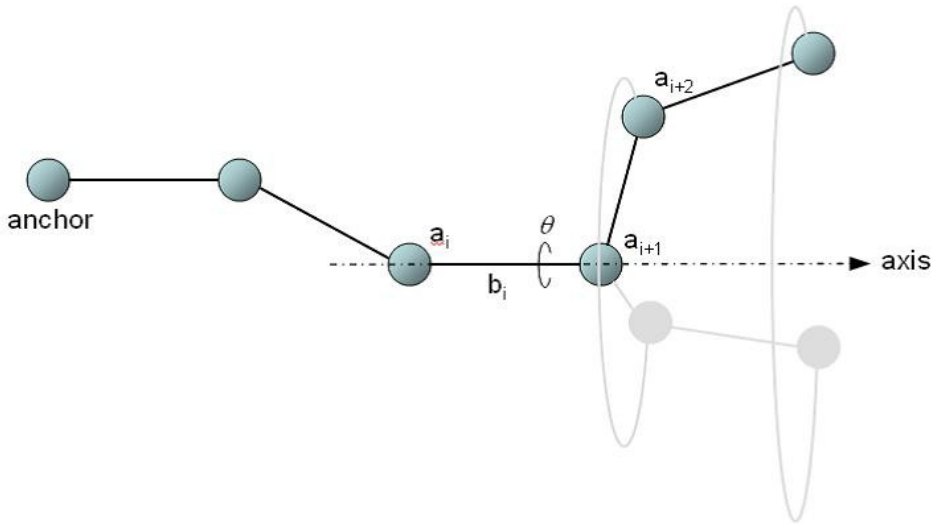
Dynamical model of Proteins with Quaternions (1)

Different structures of a protein are generated by varying dihedral angles.

In our modelization only dihedral angles are free to vary for kinematic manipulations.

For every atom a precise, circular trajectory is allowed, around the axis identified by the two previous atoms. To do this we temporarily assume as fixed:

- bond lengths;
- planar angles;
- omega angles.

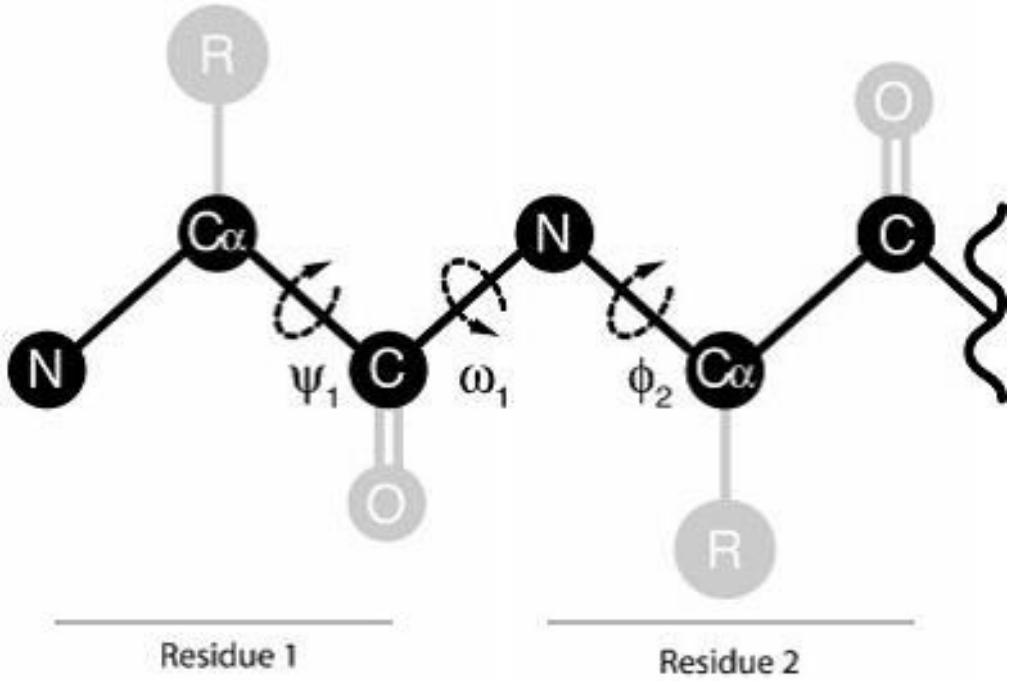


InterOmics CNR

Dynamical model of Proteins with Quaternions (2)

Amino acids have a nitrogen **N**, carbon **C** and carbon $\mathbf{C}_{\alpha}$, which make up the backbone of the protein.

The structure of the backbone is completely defined by the coordinates of these 3 atoms which define the related angles ⌒, ♥, ϕ.

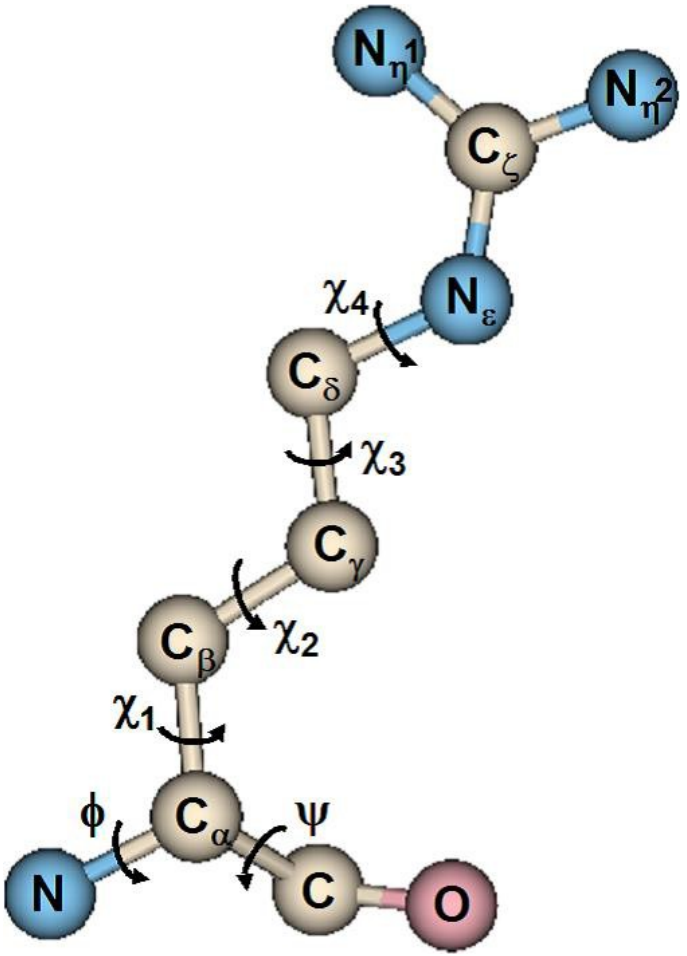


InterOmics CNR

Dynamical model of Proteins with Quaternions (3)

Every aminoacid contains a different number of side chain dihedrals angles, called χ.

For example Arginine has 4.

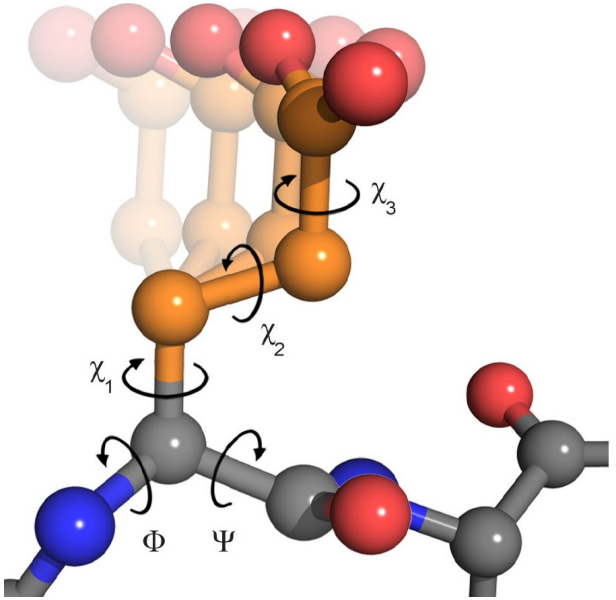


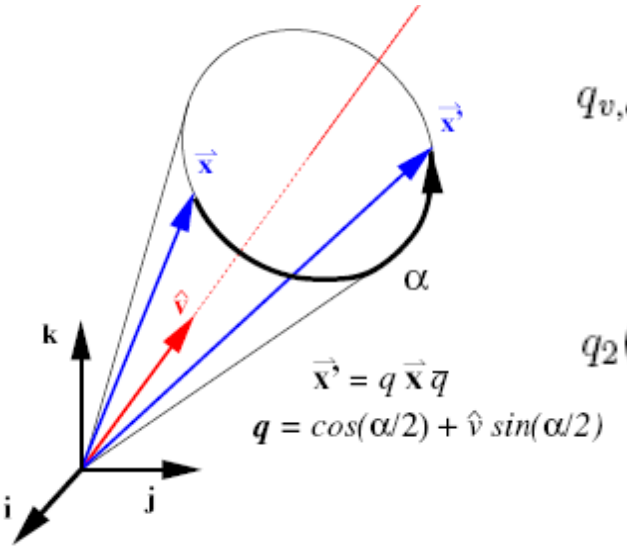
InterOmics CNR

Dynamical model of Proteins with Quaternions (4)

Considering that atoms move on circular trajectories, **unitary quaternions** are appropriate to model their movements.

Monte Carlo methods are used to perform incremental rotations and to control energy values.





$$q_{v,\alpha} \stackrel{def}{=} (\cos\frac{\alpha}{2}, \hat{v} \cdot \sin\frac{\alpha}{2})$$

$$\vec{x}' = q\vec{x}\bar{q}$$

$$q_2(q_1\vec{x}\,\overline{q_1})\overline{q_2} = (q_2 q_1)\vec{x}\,(\overline{q_1}\,\overline{q_2})$$

InterOmics CNR

Sampling conformations with Monte Carlo method (1)

Random incremental rotations can be differentiated for each amino acid, following the biological propensity to motion of various parts of the molecule.

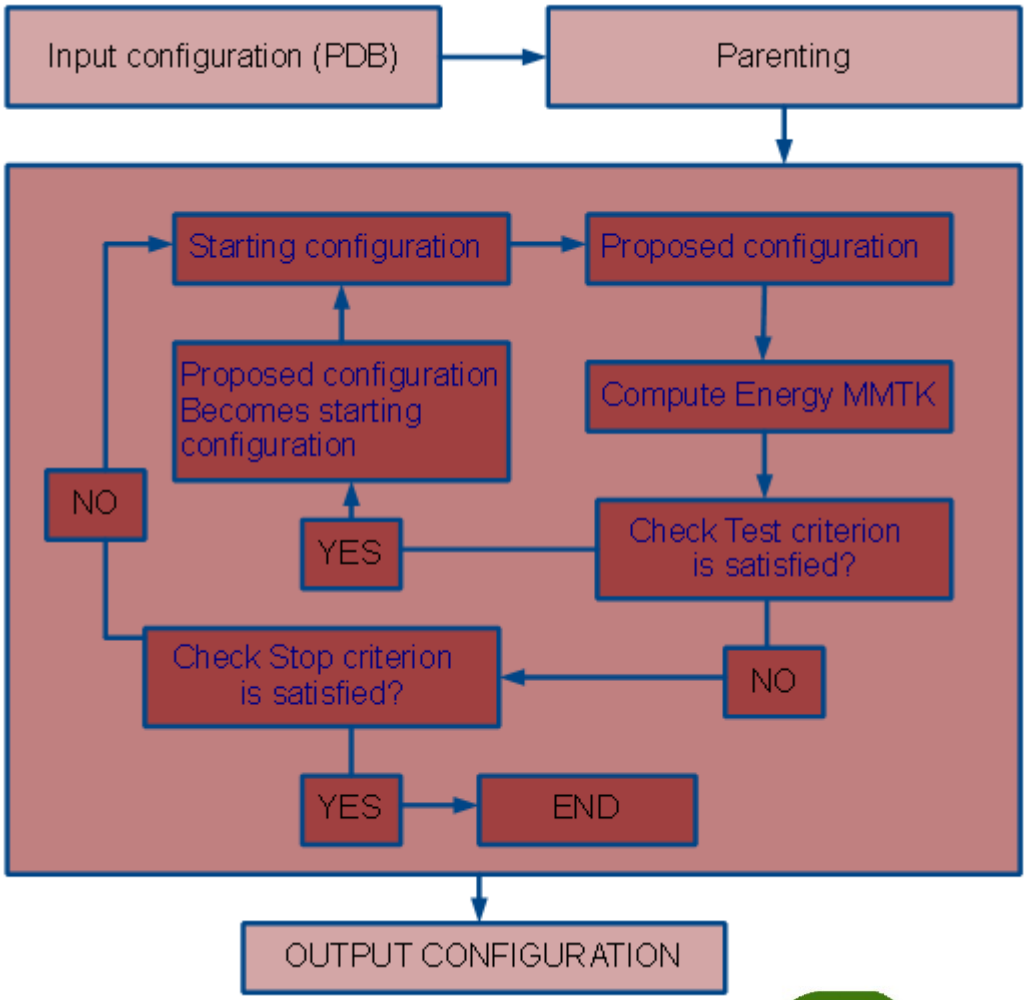
For example, terminal atoms of every aminoacid can rotate much more than backbone carbons (see table).

As suggested by Karney (2007), modeling molecules with quaternions allows a very handy application of Monte Carlo methods, because random incremental rotations become very easy.

Backbone					
ϕ	[-0.05,0.05]				
ψ	[-0.05,0.05]				
Side Chains	χ_1	χ_2	χ_3	χ_4	χ_5
ARG	[-0.1,0.1]	[-0.1,0.1]	[-0.5,0.5]	[-1,1]	[-5,5]
ASN	[-0.1,0.1]	[-5,5]			
ASP	[-0.1,0.1]	[-5,5]			
CYS	[-1,1]				
GLN	[-0.1,0.1]	[-0.5,0.5]	[-5,5]		
GLU	[-0.1,0.1]	[-0.5,0.5]	[-5,5]		
HIS	[-0.1,0.1]	[-0.5,0.5]			
ILE	[-0.1,0.1]	[-1,1]			
LEU	[-0.1,0.1]	[-5,5]			
LYS	[-0.1,0.1]	[-0.1,0.1]	[-0.5,0.5]	[-1,1]	
MET	[-0.1,0.1]	[-0.5,0.5]	[-1,1]		
PHE	[-0.1,0.1]	[-0.5,0.5]			
PRO	[-0.01,0.01]				
SER	[-0.5,0.5]				
THR	[-0.5,0.5]				
TRP	[-0.1,0.1]	[-0.5,0.5]			
TYR	[-0.1,0.1]	[-0.5,0.5]			
VAL	[-0.5,0.5]				

Sampling conformations with Monte Carlo method (2)

- Geometrical movements are performed using exclusively **unitary quaternions** (norm=1).
- Perturbations are managed by **Simulated Annealing**, taking as data-fit function the potential energy of the molecule, calculated by **MMTK**.
- Given the current configuration of the protein, an altered one is proposed and accepted or not according to a probabilistic test, following Simulated Annealing procedure.
- Program is written in **Python**.



InterOmics CNR

Sampling conformations with Monte Carlo method (3)

Starting from a PDB file, different runs normally produce different configurations, which are compared with other stable configurations stored in the PDB data bank.

Variation of bond lengths, planar angles and omega angles are very small.

We initially consider them fixed, but these approximations could produce errors. To reduce these errors conformations are adjusted by minimization every *n* steps (in our experiments *n* = 10).

Currently, given a configuration C the data-fit function only considers its potential energy:

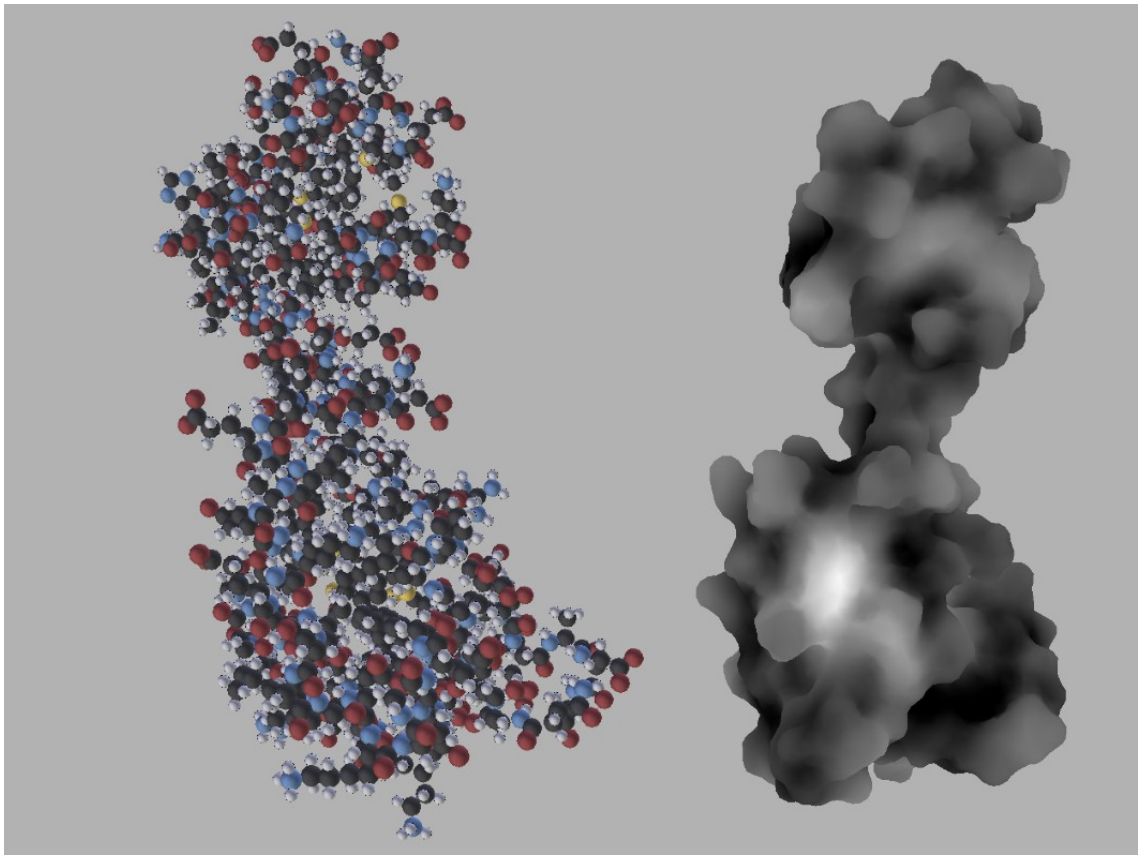
$$\Phi(\mathscr{C}) = energy(\mathscr{C})$$

by introducing a factor of distance to another conformation,it will be possible to study paths which transit proteins among different conformations:

$$\Phi(\mathscr{C}) = energy(\mathscr{C}) + RMSD(\mathscr{C}, \hat{\mathscr{C}})$$

Experiment (1)

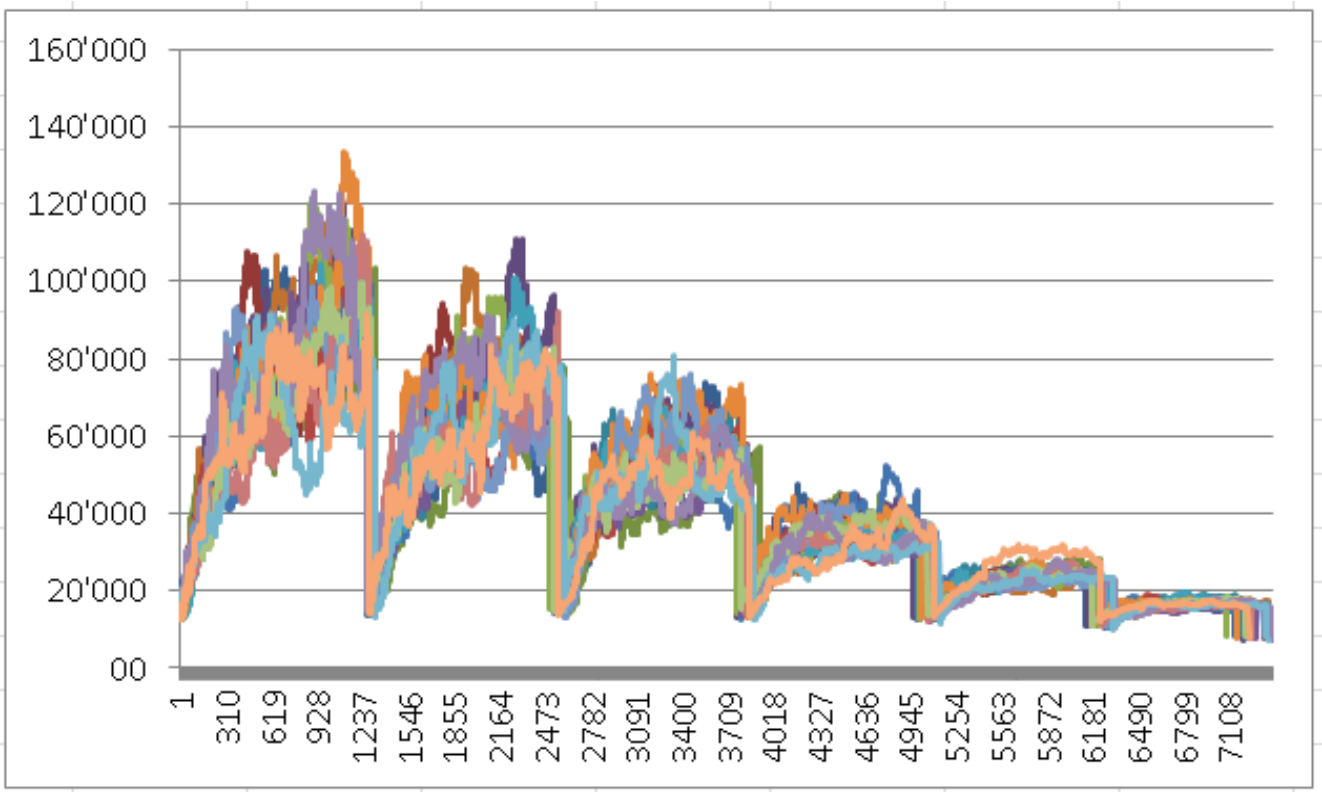
- For our experiments we use *Calmodulin* (CaM, pdb file 1cfc)
- CaM has 148 aminoacids (1165 heavy atoms).
- 1cfc contains 25 conformations with low potential energy.
- CPU time ~10.000 sec.
- Starting from conformation 1, 100 runs provide different output configurations.
- Final configurations are aligned with all the 25 configurations of 1cfc and RMSD are calculated.



Experiment (2)

- Energy of configuration 1, calculated by MMTK is 12377.9 kj/mol
- Energies of our final conformations are 10000 -15000 kj/mol

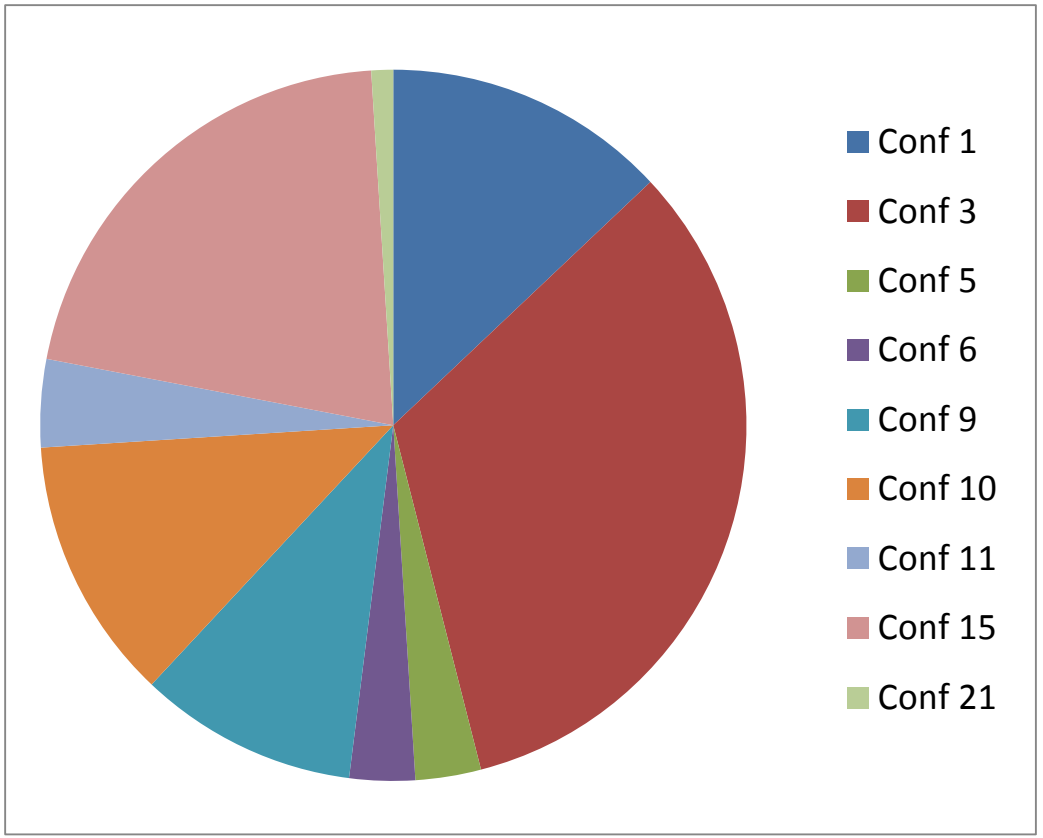
	Energy (kj/mol)
Conf 1	13748.63608
Conf 2	12346.68016
Conf 3	13065.73218
Conf 4	11964.94925
Conf 5	13187.04004
Conf 6	13412.16238
Conf 7	13155.0501
Conf 8	13589.59844
Conf 9	13516.91566
Conf 10	12328.88234
Conf 11	14548.09825
Conf 12	14286.80988
Conf 13	13525.29639
Conf 14	12786.08585
Conf 15	14770.12598
Conf 16	12406.55434
Conf 17	14575.93313
Conf 18	11546.4750
Conf 19	12840.35095
Conf 20	16949.96806
Conf 21	15952.42768
Conf 22	12253.50514
Conf 23	12541.7688
Conf 24	11608.5729
Conf 25	14441.19508



Experiment (3)

	RMSD from 1
Conf 1	0
Conf 2	8.6
Conf 3	2.84
Conf 4	5.16
Conf 5	4.45
Conf 6	4.54
Conf 7	11.8
Conf 8	8.09
Conf 9	3.25
Conf 10	4.82
Conf 11	5.17
Conf 12	5.46
Conf 13	5.68
Conf 14	8.58
Conf 15	2.42
Conf 16	9.82
Conf 17	7.35
Conf 18	4.3
Conf 19	9.73
Conf 20	4.13
Conf 21	2.93
Conf 22	8.46
Conf 23	6.29
Conf 24	8.2
Conf 25	4.01

	Reached	Av. RMSD
Conf 1	13	2.62
Conf 3	33	2.99
Conf 5	3	2.35
Conf 6	3	3.08
Conf 9	10	2.82
Conf 10	12	3.41
Conf 11	4	3.42
Conf 15	21	2.78
Conf 21	1	5.03

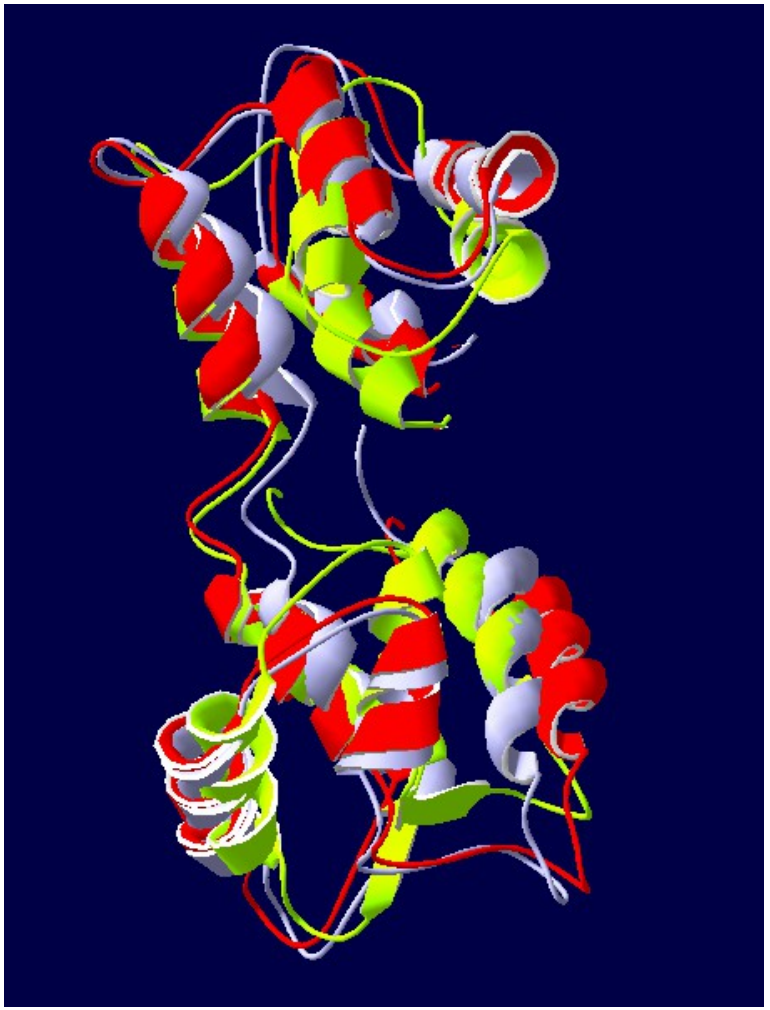


InterOmics
CNR

Experiment (4)

- Conformations:
  - Yellow: input
  - Red: obtained
  - Violet: conformation 10
- RMSD
  - red-yellow 4.7 Å
  - red-violet 2.4 Å
  - violet-yellow 4.8 Å

Starting from conformation 1 (yellow), we obtain a new model very similar to conformation 10 (violet).



Conclusions

- **Novelties**
  - Quaternions.
  - Large set of solutions in a fast way.

- **Future Improvements**
  - Change the data-fit function in order to obtain a driven transition (steered MD)
  - Program optimization to speed up calculations

Authors & References

Authors

- **Claudia Caudai**, CNR- ISTI, Pisa Italy **(claudia.caudai@isti.cnr.it)**
- **Monica Zoppè**, CNR- IFC, Pisa Italy
- **Maria Antonietta Pascali**, CNR- ISTI, Pisa Italy
- **Emanuele Salerno**, CNR- ISTI, Pisa Italy
- **Anna Tonazzini**, CNR- ISTI, Pisa Italy

References

- **Alder, B.J. (1959)** Studies in Molecular Dynamics, *J. Chem. Physics.*, 51(2)
- **Karney,C.F. (2007)** Quaternions in molecular modeling, *J. Mol. Graph. Model.*, 25, 595-604
- **Seok, C. *et al.* (2007)** Efficiency of Rotational Operators for Geometric Manipulation of Chain Molecules. *Bull. Korean Chem. Soc.*, 28(10), 1705–1708.
- **Hanson,A.J., *et al.* (2012)** Quaternion Maps of Global Protein Structure, *J. Mol. Graph. Model*., 38, 256–278.
- **Hinsen, K. (2000)** The Molecular Modeling Toolkit: A New Approach to Molecular Simulations, *J. Comp. Chem.* 21, 79-85.

Any Questions?

Research funded under the flagship Project “InterOmics” (PB.P05)

InterOmics
CNR