HPC in Molecular Modeling in INCDTIM

Attila Bende

Molecular and Biomolecular Physics Department
National Institute for Research and Development of Isotopic and Molecular Technologies,
Cluj-Napoca, ROMANIA

5^{th} Romania Tier 2 Federation Conference "Grid, Cloud & High Performance Computing Science"

Outline

The first theoretical works in Molecular Modeling at ITIM, Cluj-Napoca

– The first theoretical calculations in the field of molecular physics were performed in the early 80's

– Vibrational normal mode analysis and normal mode frequency calculations

– Fourier analysis of EXAFS and XANES crystallographic data

– Simulation of multiphoton excitation and absorption spectra

They were "Home Made" programs written in Fortran 77 language or "Friendship" programs obtained from collaborators.

The first computer system dedicated for Molecular Modeling calculations

Hardware

- Supermicro Motherboard
- 2xOpteron dual-core 265
- 4GB RAM DDR 333 MHz
- HDD 200GB

Operating system

- Opensuse 9.3
- Network File System (NFS)
- Network Information System (NIS)
- No any batch system for the jobs

Mol. Mod. Software

- Gaussian03
- SAPT for intermolecular interaction
- NWChem 5.0
- DFTB+
- WASP
- Spinevolution
- YACAS

Mol. Graph. Software

- Gabedit
- MOLDEN
- XCrysDen

2005-2008

- Three 2xOpteron dual-core 265
- Two 2xOpteron dual-core 275
- One 2xOpteron dual-core 2224
- Three HP Intel Xeon quad-core E5430

2005-2008

- Molpro 2008.1
- QuantumEspresso
- Amber 9.0

2005-2008

- 12 active users

Compilers

- Intel Fortran Compiler
- GNU Fortran
- Perl

Communication Protocols for Parallel Computing

- MPICH2 – http://www.mcs.anl.gov/mpi/mpich
- Open MPI – www.open-mpi.org
- OpenMP – http://openmp.org
- Global Arrays Toolkit – http://www.emsl.pnl.gov/docs/global

Network Communications between Nodes

- 10 MB Ethernet Switch – Slow network communication between the nodes
- Number of processor limit due this restriction

INCDTIM Data Center

– It was built in 2007

Project POS-CCEA 2009

Title: "Îmbunătățirea capacității și fiabilității centrului GRID al INCDTIM în vederea integrării acestuia în rețele internaționale de specialitate."

EULER System

Hewlett Packard Blade C7000 with 16 Proliant BL280c G6:

- 2 Intel Quad-core Xeon x5570 2.93 GHz, 16 GB RAM DDR2, 500 GB HDD) – 128 cores
- 10 GB Ethernet Switch
- TORQUE, MAUI, GANGLIA (http://hpc.itim-cj.ro), NAGIOS, configured from scratch – Scientific Linux 5.3 (Boron)

Quantum Chemistry

- Gaussian03
- Gaussian09
- Molpro 2012.1
- NWChem 6.1
- Crystal 09
- Orca 2.9
- GAMESS
- DFTB+
- Siesta
- VASP
- Accelrys Materials Studio
- Mopac 2012
- Spinevolution

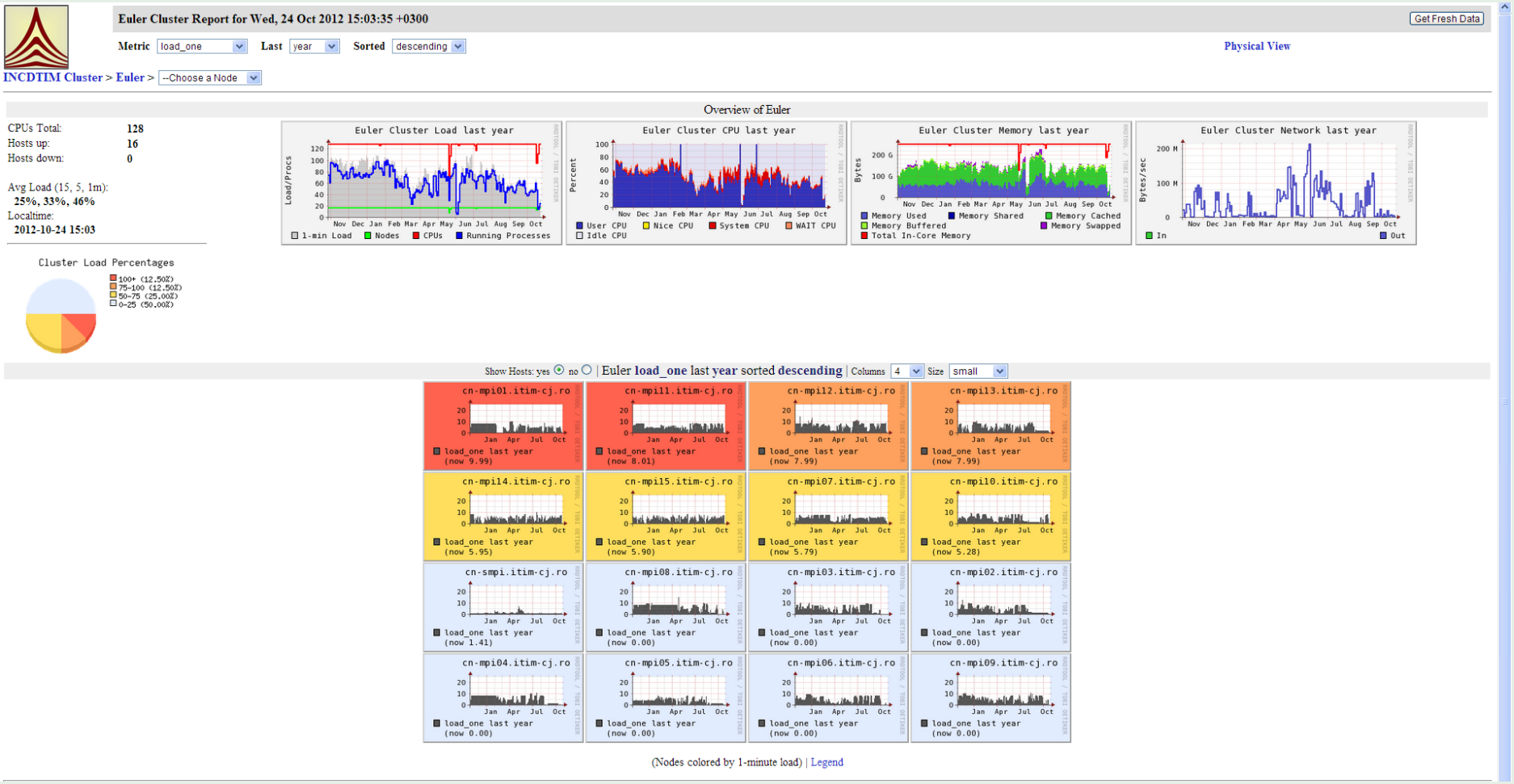
Molecular Dynamics

- Amber 09
- Amber 12
- GROMACS
- NAMD
- LAMMPS
- Car-Parinello MD
- CP2K

Molecular Graphics

- Gabedit
- GaussView
- MOLDEN
- XCrysDen

Job Control - Ganglia



Users

> 20 who used at least one time, 10 users use regularly.

Future: Project PN-II CAPACITĂȚII – MDFMOLBIO

IBM System x iDataPlex DX360 M4:

- 26 Nodes: Two Intel Xeon Sandy Bridge E5-2665 8-Core 2,4 GHz, 20 MB L3 cache, DDR3 64 GB 1600 MHz ECC (416 cores)
- 2 Nodes: *Idem.* + NVIDIA Tesla M2090 GPU (32 cores)
- Head Node Servers: Two CISCx86 6-core, with 2.93 GHz
- Storage: 20TB with SAS 2.0 disks with 10000 rpm
- File Server for Storage: General Parallel File System (GPFS)
- FDR Infinite band with 56 GB bandwidth

Why we need more and more powerful computer facilities?

Electronic Structure Methods:

Method	Accuracy	Max atoms
Semiempirical	Low	∼2000
Hartree–Fock or DFT	Medium	∼500
Perturbation or Variation Meth.	High	∼50
Coupled-Cluster Meth.	Very High	∼20

Accuracy of Hartree-Fock:

Property	Accuracy
Bond lengths	±0.02 Å
Bond angles	$\pm 2^o$
Vib. Freq.	±11 %
Dipole moments	±3 D
Relative energy	±25-40 kcal/mol

Accuracy of Coupled-Cluster:

Property	Accuracy
Bond lengths	±0.004 Å
Bond angles	$\pm 0.03^o$
Vib. Freq.	±2 %
Dipole moments	±0.05 D
Relative energy	±1.5 kcal/mol

How to choose the proper HPC system?

- To have fast, last generation processors (Intel Sandy Bridge)
- To have huge memory amount per processor, DDR3
- To have huge temporary scratch disk – with high RPM
- To have huge permanent storage for stoking data
- To have fast network transfer - Infinite Band at QDR or FDR levels
- To have fast communication protocols between processors and RAM memory

– Not every parallel environment can take advantage of all these facilities!!!

Molecular Modeling with HPC

- Molecular and Biomolecular Structure Prediction
- Molecular Cluster Structure Prediction
- Molecular Crystal Structure Prediction
- Nature of Intermolecular Interaction
- IR, Raman, UV and NMR Spectra Prediction
- Molecular Self-Assembling
- Molecular Excited State
- Photochemical and Photophysical Relaxation Pathways
- Electron Conduction and Charge Transport through Molecular Chains
- Molecular Electronics
- Molecular Absorption on Metal Surface
- Molecular Dynamics of unusual Liquid – Solid Phase Transition

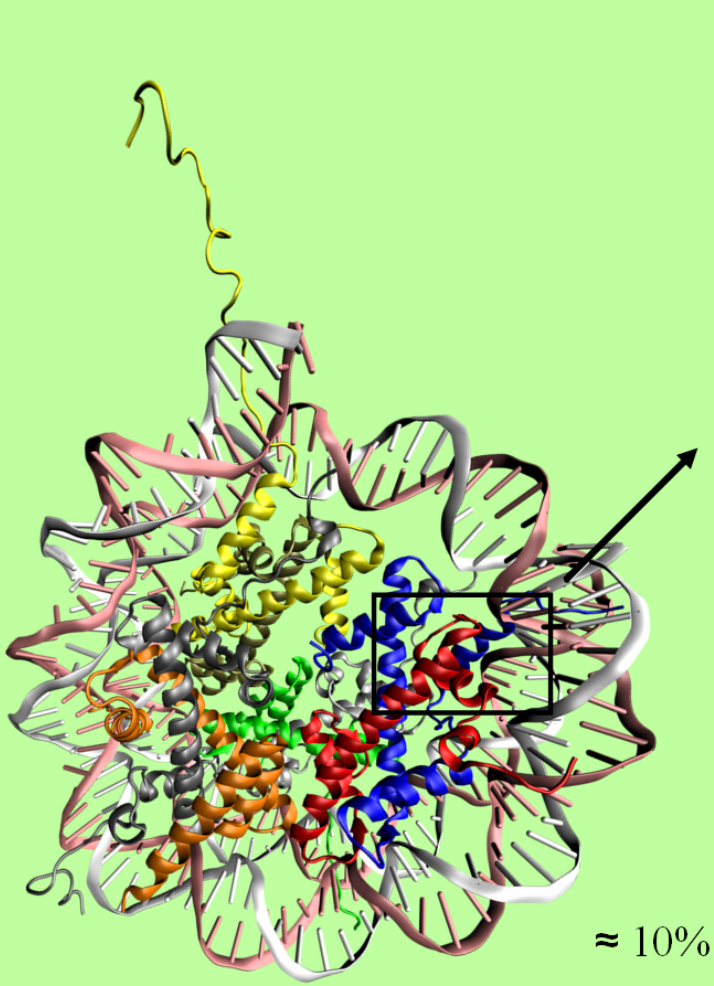
Scientific Publications

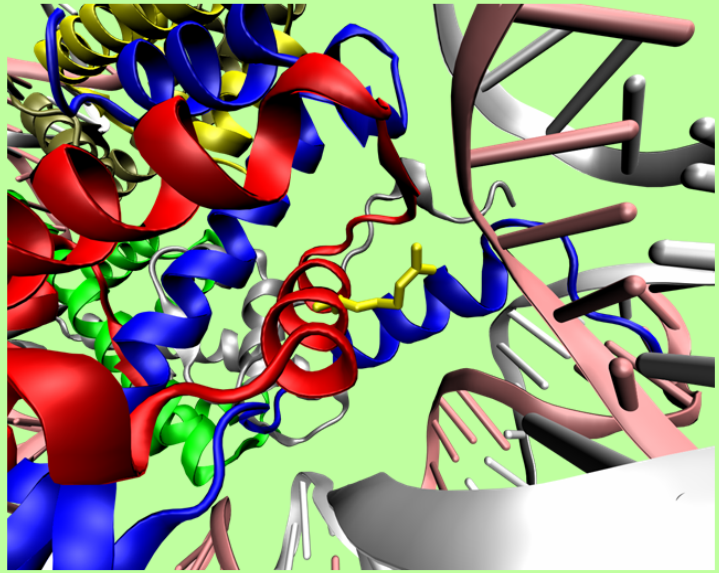
Journals:	Impact Factor:
– ACS Nano	– 10.774
– Journal of Physical Chemistry C	– 4.805
– Physical Chemistry Chemical Physics	– 3.573
– Journal of Chemical Physics	– 3.333
– Physical Review B	– 2.946
– Journal of Physical Chemistry A	– 2.691
– Chemical Physics Letters	– 2.337
– Acta Crystallographica B	– 2.286
– Physical Review E	– 2.255
– Theoretical Chemistry Accounts	– 2.162
– Supramolecular Chemistry	– 2.145
– Spectrochimica Acta A	– 2.098
– Chemical Physics	– 1.896
– Solid State Communications	– 1.649
– Physics Letters A	– 1.632

Around 40 ISI scientific articles have been published in the period of 2007 – 2012 with the contribution of the INCDTIM Data Center

Molecular and Biomolecular Structure Prediction

Detailed molecular structure:





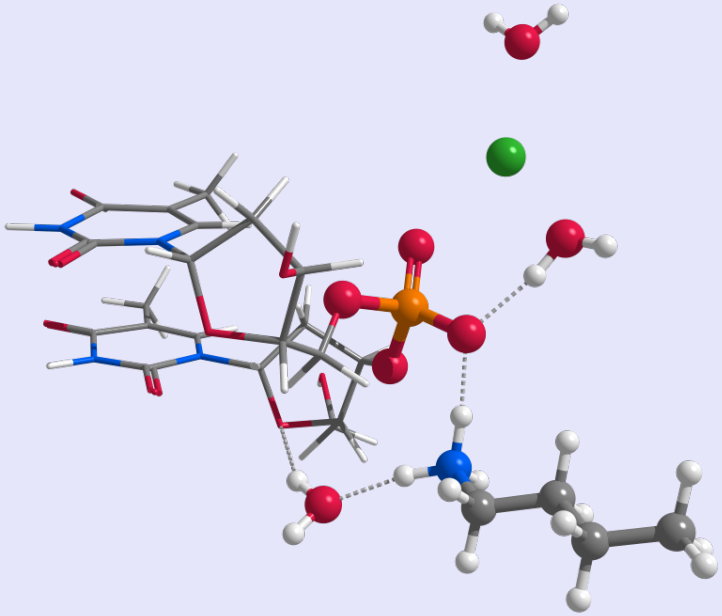
≈ 40% of amino acids are:

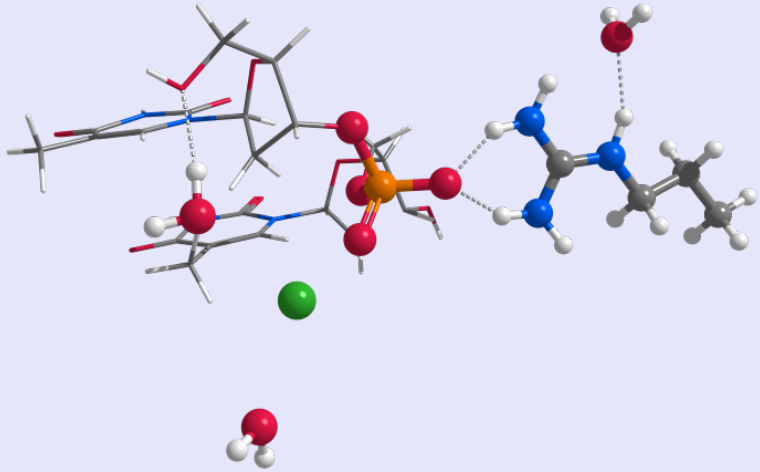
positively charged
Lysine or **Arginine**

≈ 10% of amino acids are:

negatively charged **Aspartate** and **Glutamate**

Molecular and Biomolecular Structure Prediction

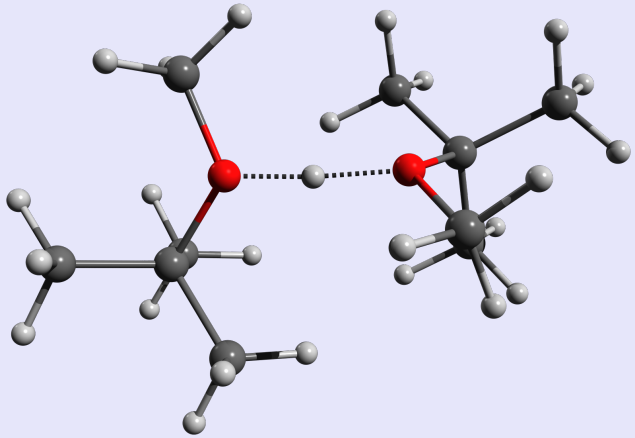


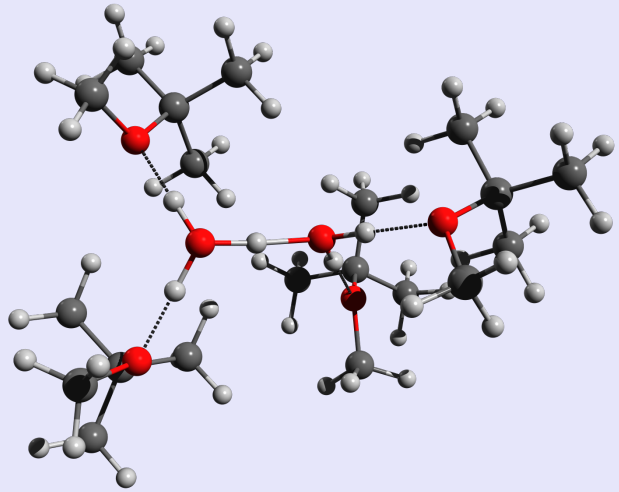


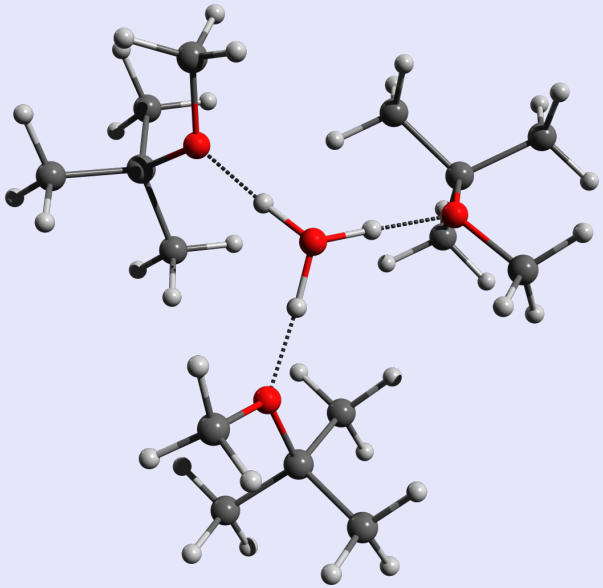
The DNA – Histon protein stability is given by the charge transfer between the positively charged $Lysine^{+}$ and $Arginine^{+}$ amino acids and the PO_4^{-} group of the DNA backbone chain.

Molecular Cluster Structure Prediction

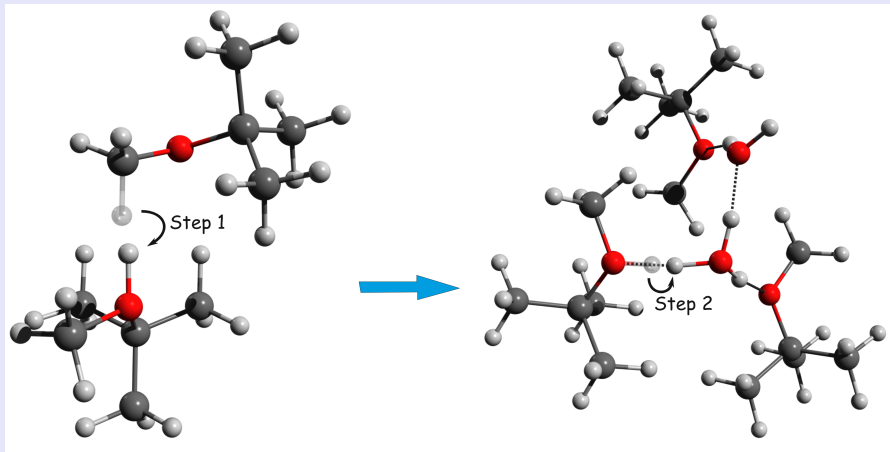
– The methyl tert-butyl ether – Water Ionic Cluster



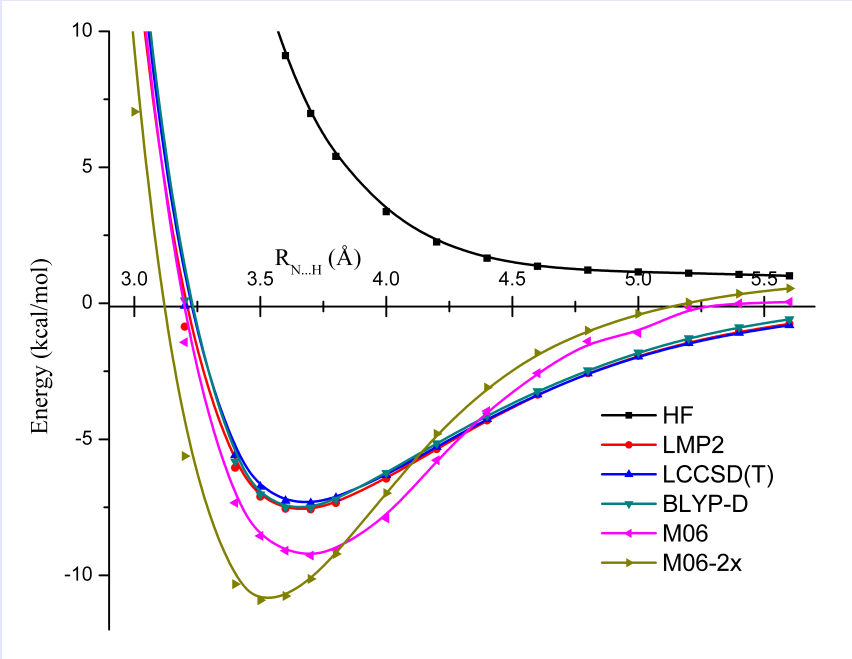


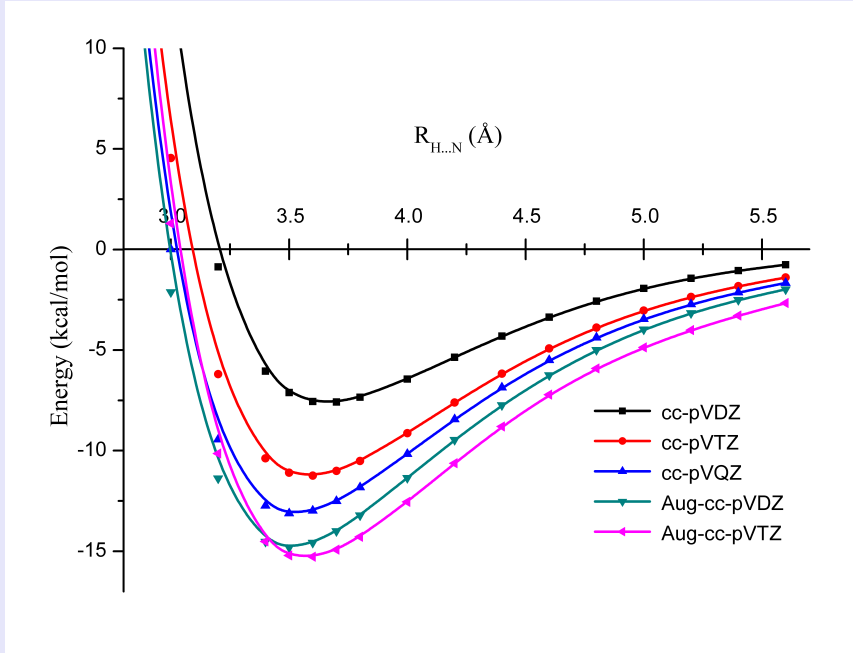


$(MTBE)_m(H_2O)_nH^+$ with n=m – 2



Nature of Intermolecular Interactions

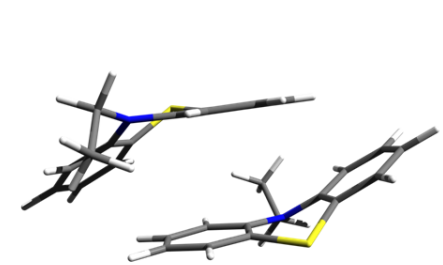




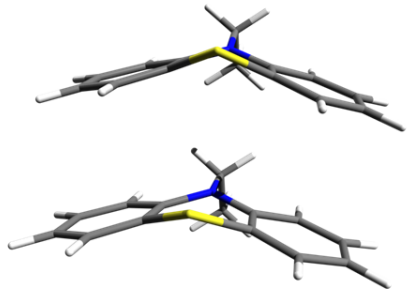
We can explain the role of the electron correlation effects covered by different theoretical models: Hartree–Fock, Many-Body Perturbation Theory, Density Functional Theory, Coupled–Cluster as well as the role of the basis sets.

Molecular Self-Assembling

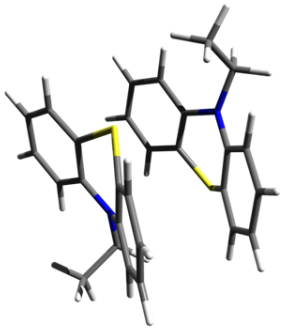
Molecular Self-Assembling of Phenothiazine Dimers



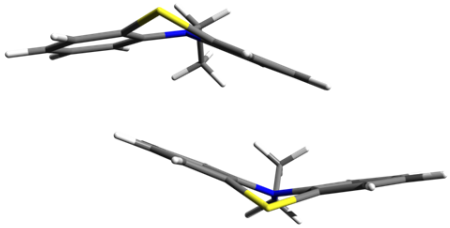
EPT-C1



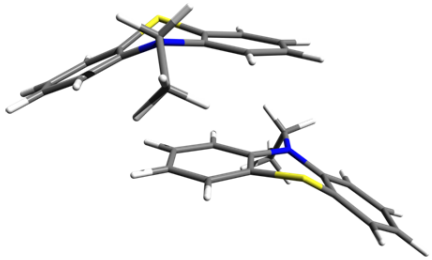
EPT-C2



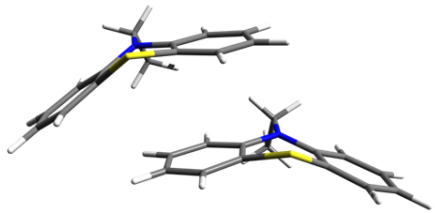
EPT-C3



EPT-C4



EPT-C5



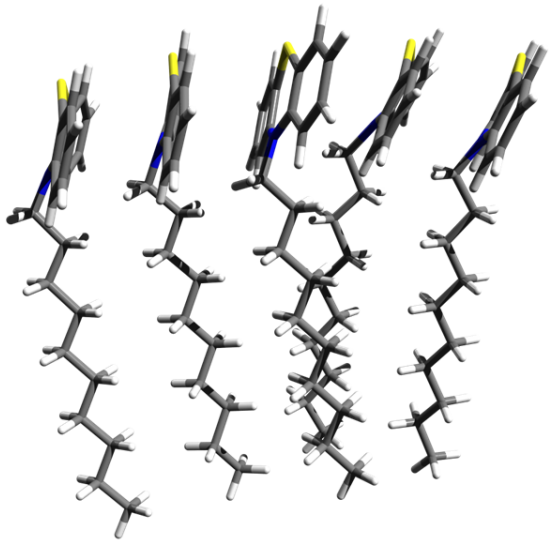
EPT-C6

$$E^{C1} \xrightarrow{+0.497\ \text{kcal/mol}} E^{C2} \xrightarrow{+0.611\ \text{kcal/mol}} E^{C3} \xrightarrow{+0.980\ \text{kcal/mol}} E^{C4} \xrightarrow{+0.788\ \text{kcal/mol}} E^{C5} \xrightarrow{+0.085\ \text{kcal/mol}} E^{C6}$$

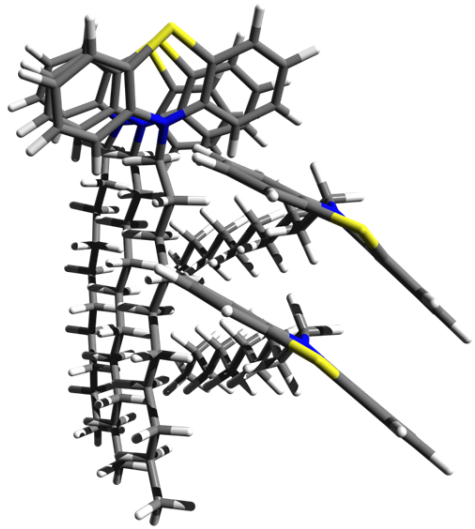
We are able to predict the most probable scenario of the self-assembling process.

Molecular Self-Assembling

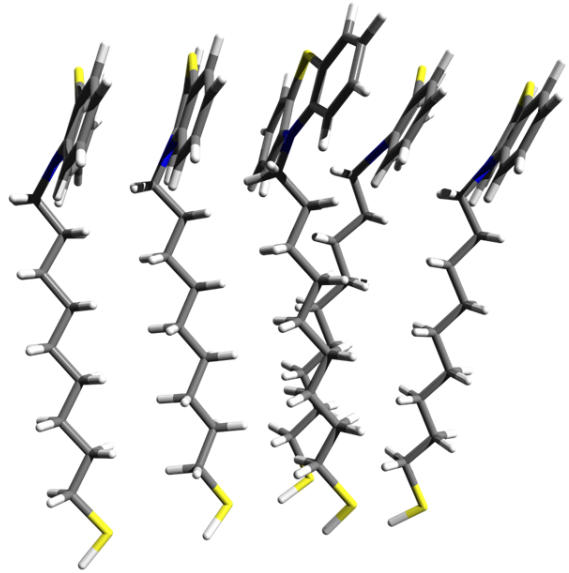
Molecular Self-Assembling of Phenothiazine Dimers



A: decyl-PTZ (a)



B: decyl-PTZ (b)

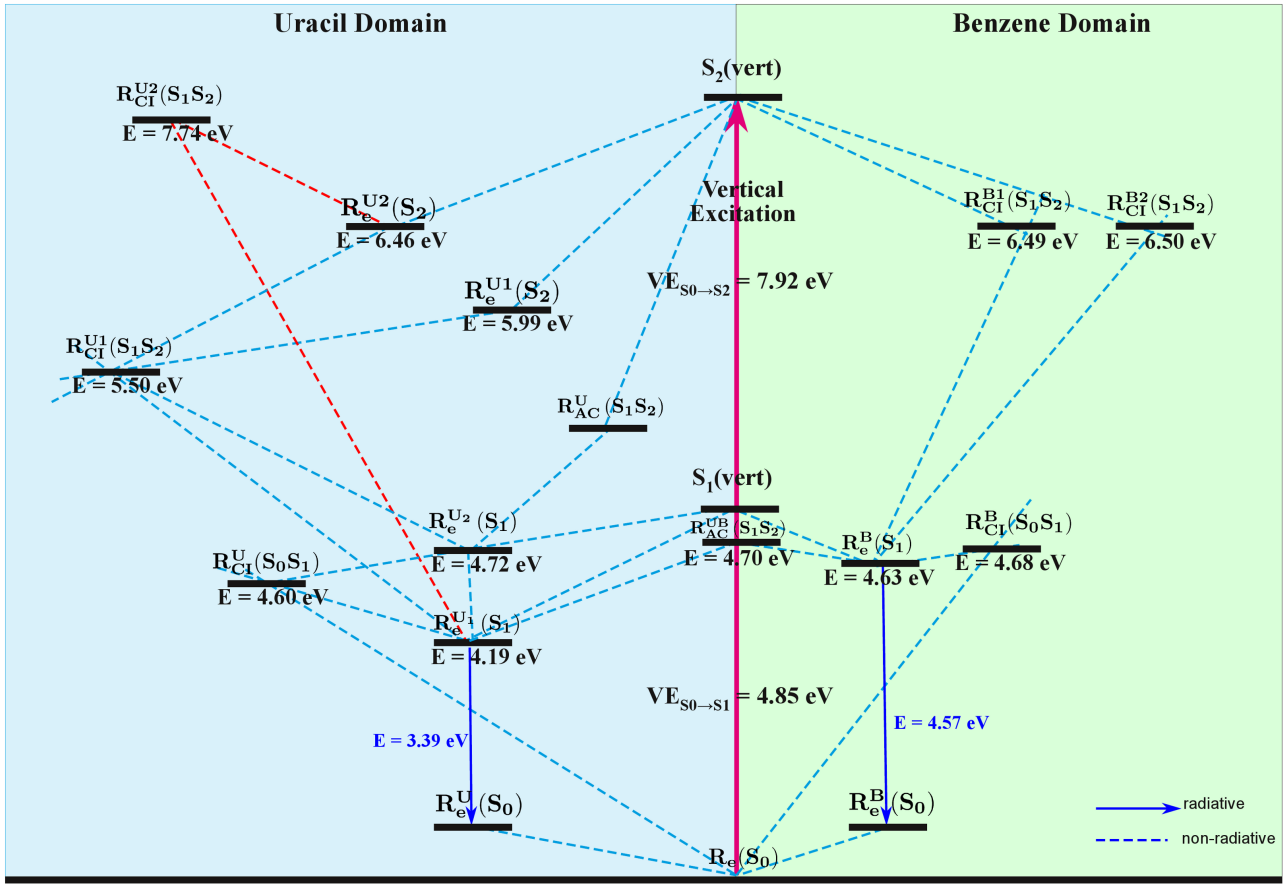


C: thio-nonyl-PTZ

BUT, we are able to predict also the possible defects during the self-assembling process.

Photochemical and Photophysical Relaxation Pathways

The Excited State Relaxation Pathways of 5BU



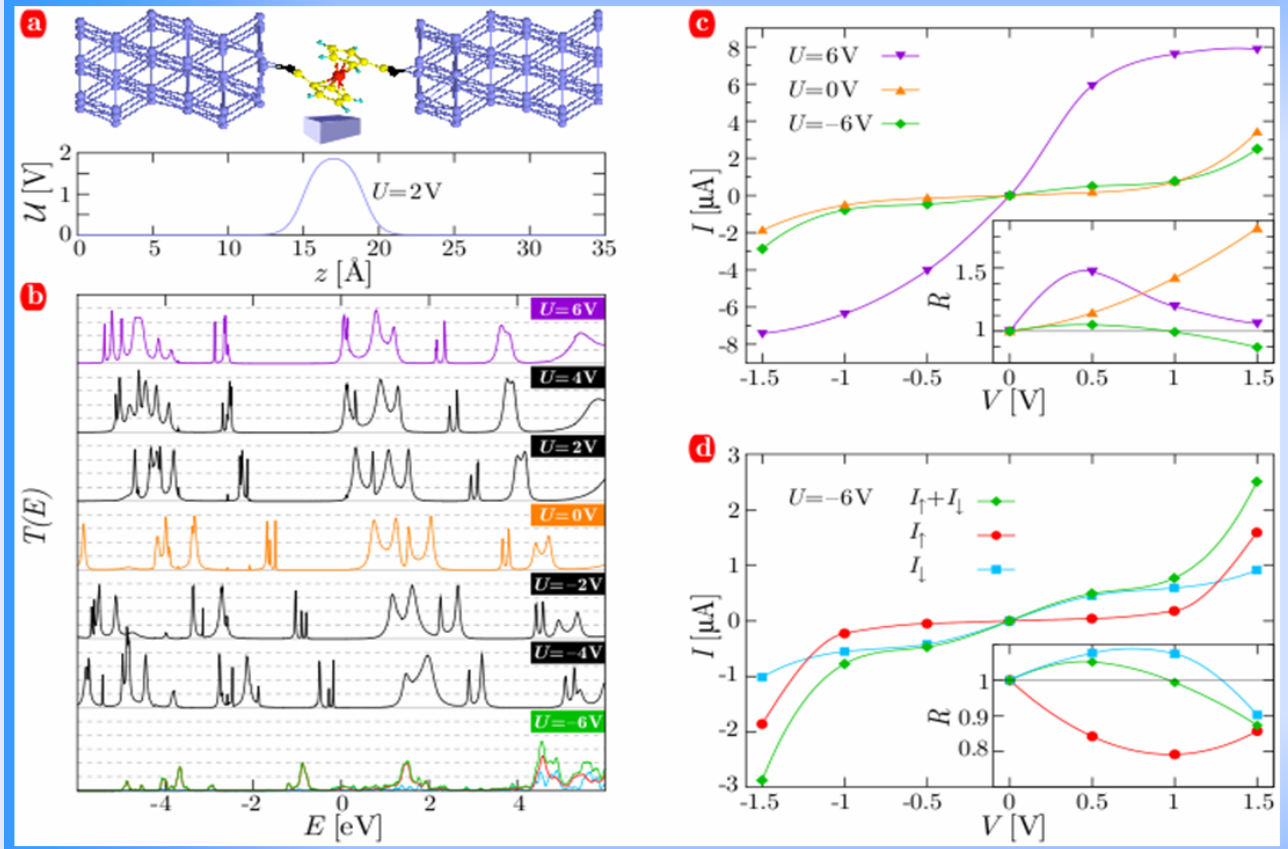
We can locate equilibrium geometries, conical intersections and transition barriers on the potential energy surfaces of different excited state levels.

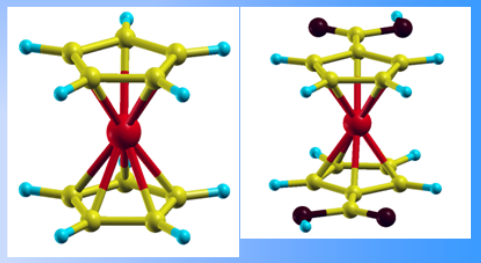
Electron Conduction and Charge Transport through Molecular Chains

INCDTIM

Gate potential effect for ferrocene derivatives with Al(111) electrodes

2012 Evaluation
E08 – Numerical Modeling





Ferrocene FDCA

Electronic transport in a single FDCA molecule is investigated base on DFT and Non-Equilibrium Green functions

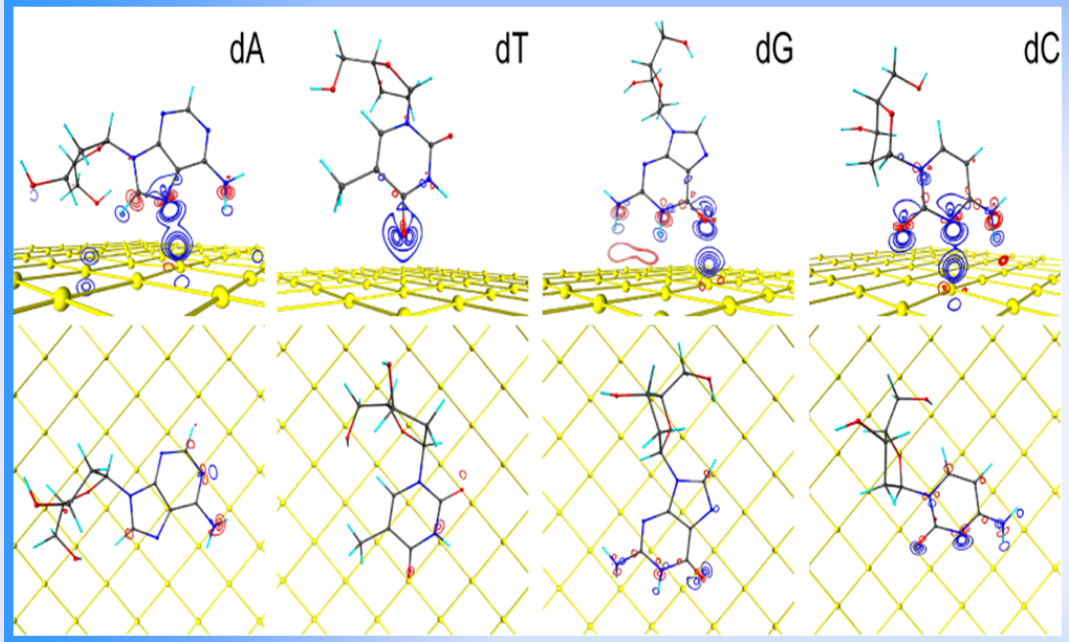
C. Morari et. al. *ACS Nano* **3(12)**, 4137-4143 (2009)

Molecular Absorption on Metal Surface

INCDTIM

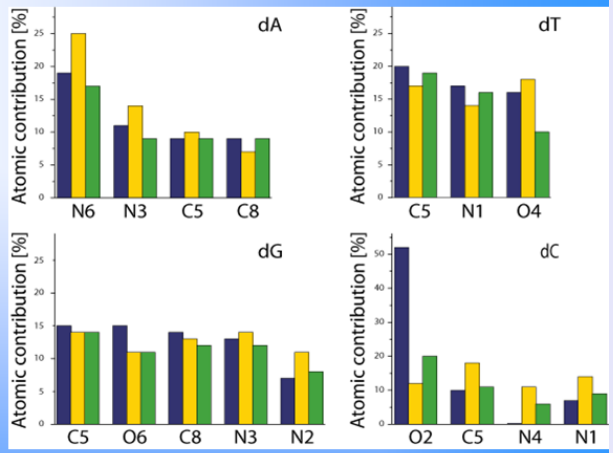
DNA nucleosides adsorbed on Au(100) surface

2012 Evaluation
E08 – Numerical Modeling



UP: contour plots in step of 0.003 e/Bohr^3 for the charge metal-molecule charge transfer. Red: negative, blue: positive values.

DOWN: atomic contributions to the HOMO for free (dark blue), perpendicular (yellow) and parallel (green) adsorbed nucleosides



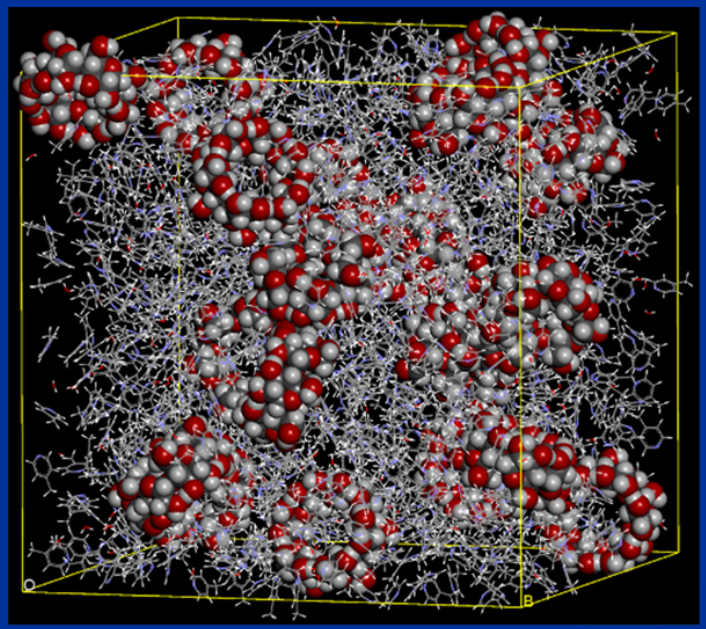
D. Bogdan and C. Morari, *J. Phys. Chem. C* **116**, 7351-7359 (2012)
DOI: http://dx.doi.org/10/1021/jp210229e

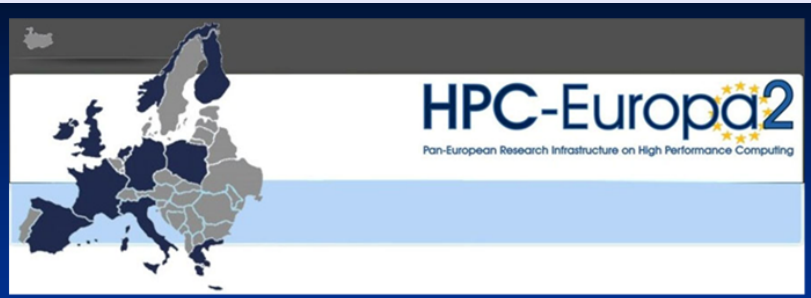
Molecular Dynamics of unusual Liquid – Solid Phase Transition

- 20 α-CD molecules
- 1120 molecules of 4MP
- 240 water molecules
- NPT ensemble MD using AMBER9
- 60 A^3 box

⇒ **18920 atoms**

- speed of 0.22ns/day (1 core), 0.39ns/day (2 cores), 0.69ns (4 cores) and more than 20 ns on a GPU





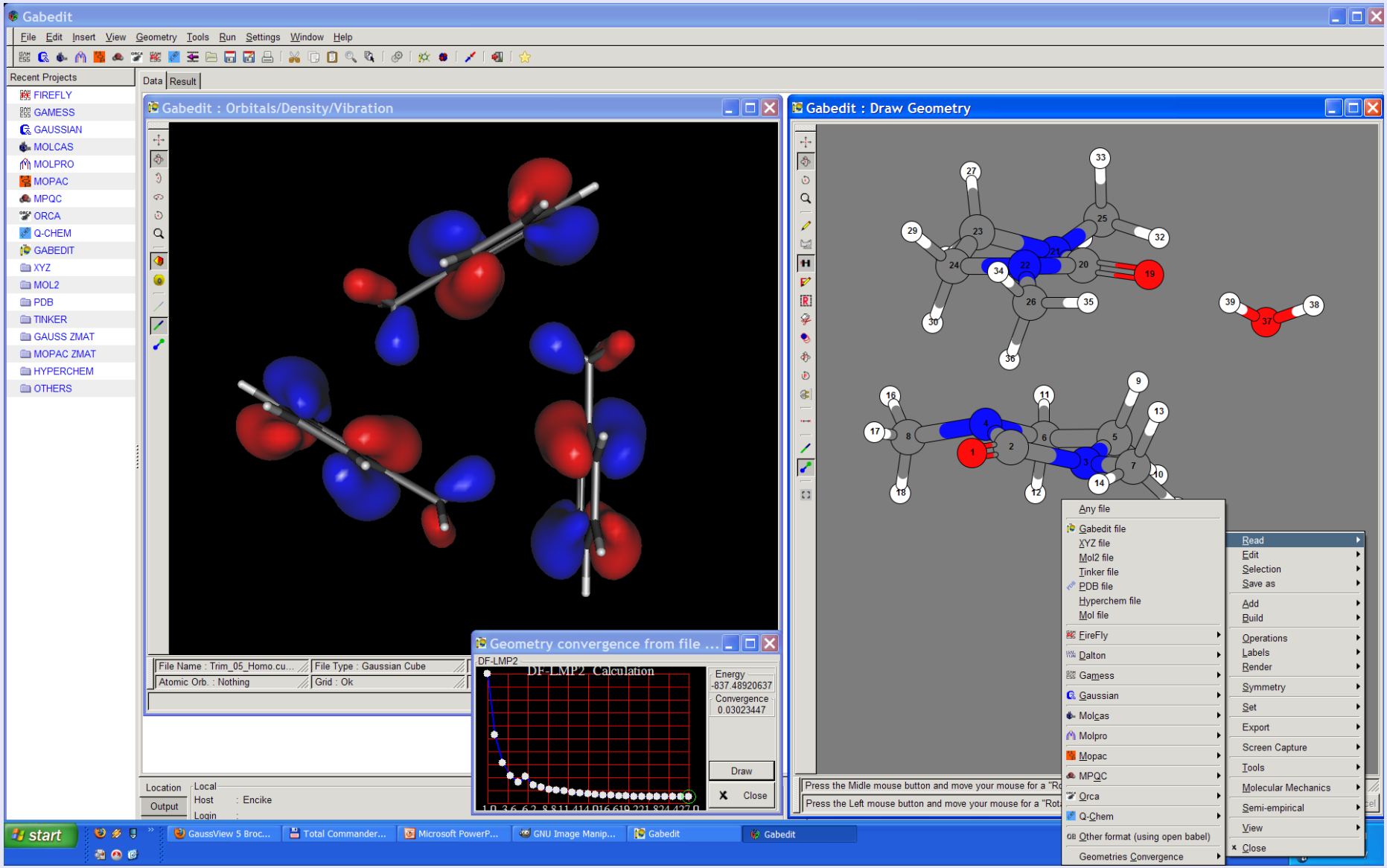
- Force field optimization using the *force-matching method*
- 100 ns long trajectories at different temperatures has been calculated for good statistics
- Hydrogen-bond dynamics and cluster analysis
- Correlation coefficients

This system was studied in collaboration with CINECA, Italy, through the project financed by HPC-Europa2 program on GPUs

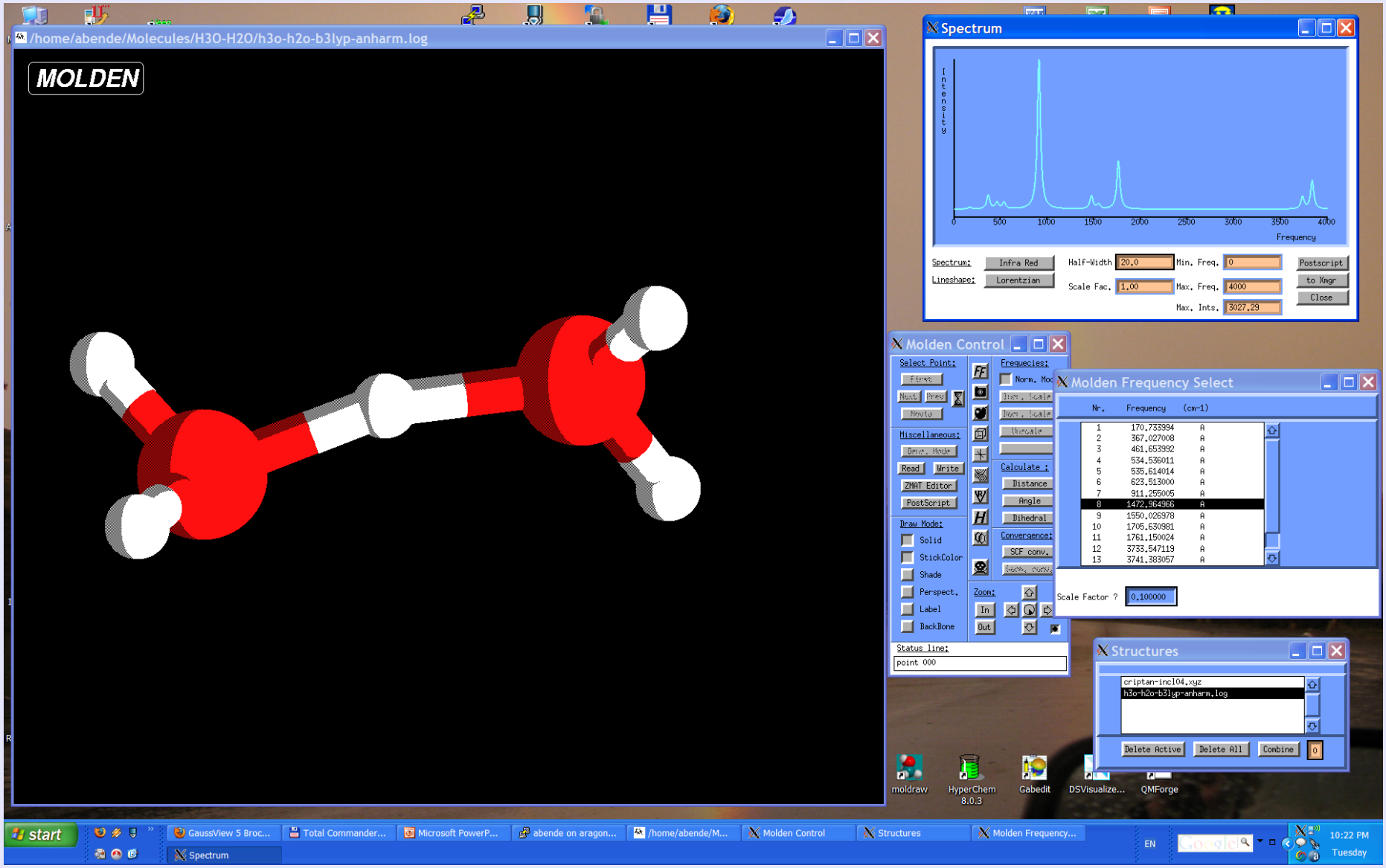
Molecular Graphics

- Create input geometries for molecular modeling softwares
- Visualize 3D molecular structures
- Measure bond distances or bond angles
- Draw orbital shapes
- Visualize molecular dynamics trajectories
- Check your intermediate data
- Draw different theoretical spectra (IR, Raman, EPR, UV or 1D-NMR)
- Create high resolution images for publications

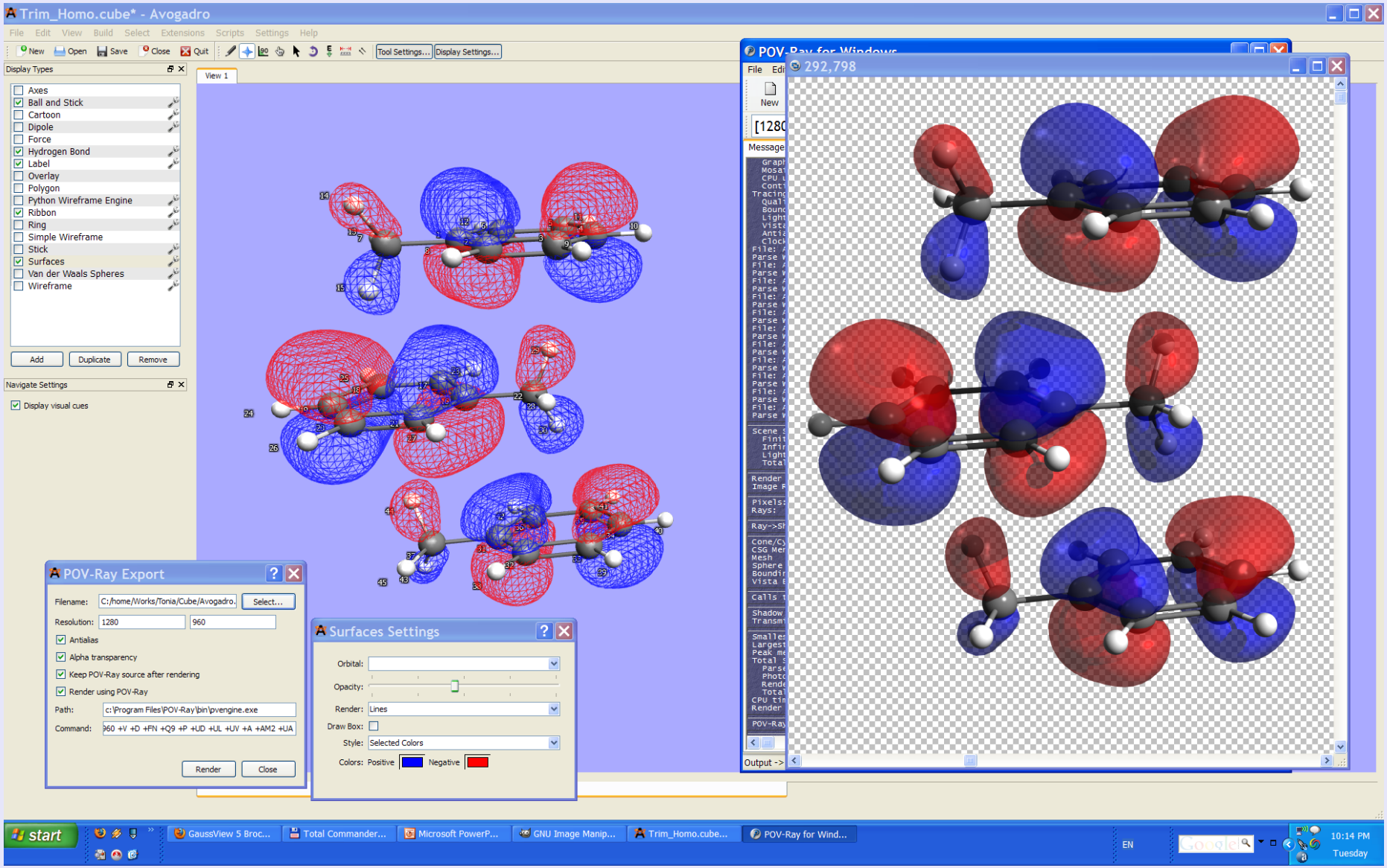
GABEDIT


Gabedit
File Edit Insert View Geometry Tools Run Settings Window Help
Recent Projects
FIREFLY
GAMESS
GAUSSIAN
MOLCAS
MOLPRO
MOPAC
MPQC
ORCA
Q-CHEM
GABEDIT
XYZ
MOL2
PDB
TINKER
GAUSS ZMAT
MOPAC ZMAT
HYPERCHEM
OTHERS
Data Result
Gabedit : Orbitals/Density/Vibration
File Name : Trim_05_Homo.cu...
File Type : Gaussian Cube
Atomic Orb. : Nothing
Grid : Ok
Geometry convergence from file ...
DF-LMP2
DF-LMP2 Calculation
Energy
-837.48920637
Convergence
0.03023447
Draw
Close
Location
Output
Local
Host : Encike
Login :
Gabedit : Draw Geometry
Any file
Gabedit file
XYZ file
Mol2 file
Tinker file
PDB file
Hyperchem file
Mol file
FireFly
Dalton
Gamess
Gaussian
Molcas
Molpro
Mopac
MPQC
Orca
Q-Chem
Other format (using open babel)
Geometries Convergence
Read
Edit
Selection
Save as
Add
Build
Operations
Labels
Render
Symmetry
Set
Export
Screen Capture
Tools
Molecular Mechanics
Semi-empirical
View
Close
Press the Midle mouse button and move your mouse for a "Ro
Press the Left mouse button and move your mouse for a "Rot
start
GaussView 5 Broc...
Total Commander...
Microsoft PowerP...
GNU Image Manip...
Gabedit
Gabedit

MOLDEN


/home/abende/Molecules/H3O-H2O/h3o-h2o-b3lyp-anharm.log
MOLDEN
Spectrum
Intensity
Frequency
Molden Control
Molden Frequency Select
Structures
criptan-incl04.xyz
h3o-h2o-b3lyp-anharm.log
moldraw
HyperChem 8.0.3
Gabedit
DSVisualize...
QMForge
start

AVOGADRO


Trim_Homo.cube* - Avogadro
File Edit View Build Select Extensions Scripts Settings Help
New Open Save Close Quit
Tool Settings... Display Settings...
Display Types
Axes
Ball and Stick
Cartoon
Dipole
Force
Hydrogen Bond
Label
Overlay
Polygon
Python Wireframe Engine
Ribbon
Ring
Simple Wireframe
Stick
Surfaces
Van der Waals Spheres
Wireframe
Add Duplicate Remove
Navigate Settings
Display visual cues
View 1
POV-Ray Export
Filename: C:/home/Works/Tonia/Cube/Avogadro.
Select...
Resolution: 1280 960
Antialias
Alpha transparency
Keep POV-Ray source after rendering
Render using POV-Ray
Path: c:\Program Files\POV-Ray\bin\pvengine.exe
Command: 960 +V +D +FN +Q9 +P +UD +UL +UV +A +AM2 +UA
Render Close
Surfaces Settings
Orbital:
Opacity:
Render: Lines
Draw Box:
Style: Selected Colors
Colors: Positive Negative
POV-Ray for Windows
292,798
start
GaussView 5 Broc...
Total Commander...
Microsoft PowerP...
GNU Image Manip...
Trim_Homo.cube...
POV-Ray for Wind...
EN
10:14 PM
Tuesday

Conclusions

Molecular modeling can't take place of the experiment, BUT

- Can help to better understand the details where the experimental measurements are not possible
- Can validate models built with the help of the experimental measurements
- Can give information about the further experimental setup
- Can give a motivation for new experimental investigations
- Can validate new theories or theoretical approximations
- Can Save Money

Thank You for Your Attentions