

# Acceleration of a Molecular Modelling Code for the Analysis and Visualization of Weak Interactions between Molecules

A. Roussel, J-C.Boisson, H. Deleau,
M. Krajecki and E. Hénon

ICMR
Institut de Chimie Moléculaire de Reims
UMR CNRS 7312

MAISON DE LA SIMULATION
de Champagne-Ardenne

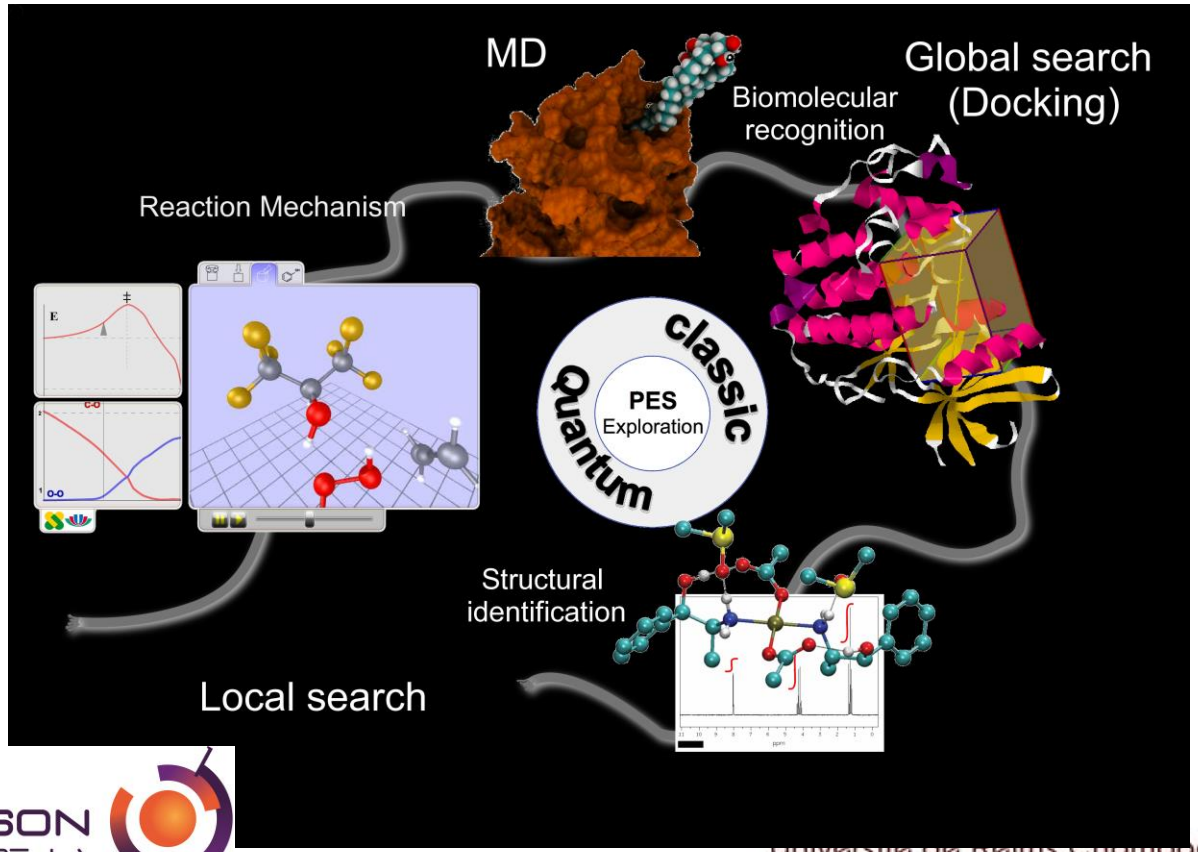
CReSTIC

Université de Reims Champagne-Ardenne

# Modeling Activities : ICMR Lab

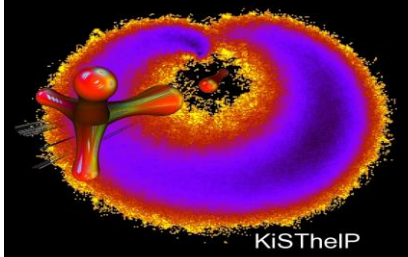
- ICMR = Experimental laboratory « augmented » by theoretical calculations

**Applied theoretical chemistry**

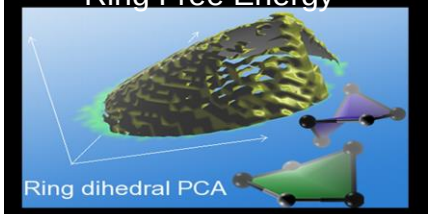



**Models & Prog.**

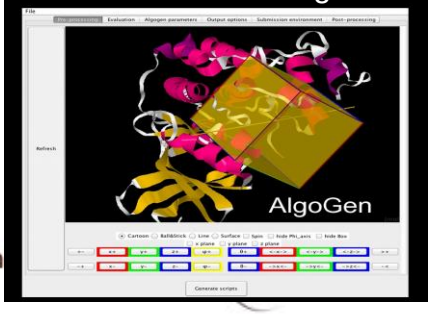
Kinetics, Thermodynamics




Ring Free Energy




Molecular Docking




MAISON DE LA SIMULATION de Champagne-Ardenne

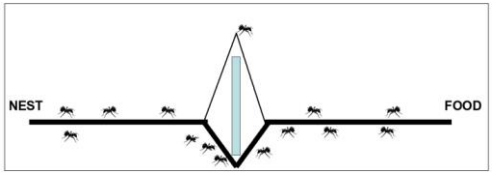
Université de Reims Champagn

# Modeling Activities : CReSTIC Lab

- CReSTIC = computer science laboratory

**Parallel and distributed algorithms**

→ Combinatorial optimisation (genetic algorithm, ant/bee colony)




→ Parallel algorithms for GPU acceleration

URCA = the first CUDA Research Center in France

**High-Performance Computing**

↓

**High-Performance Molecular Modeling**

# Outline

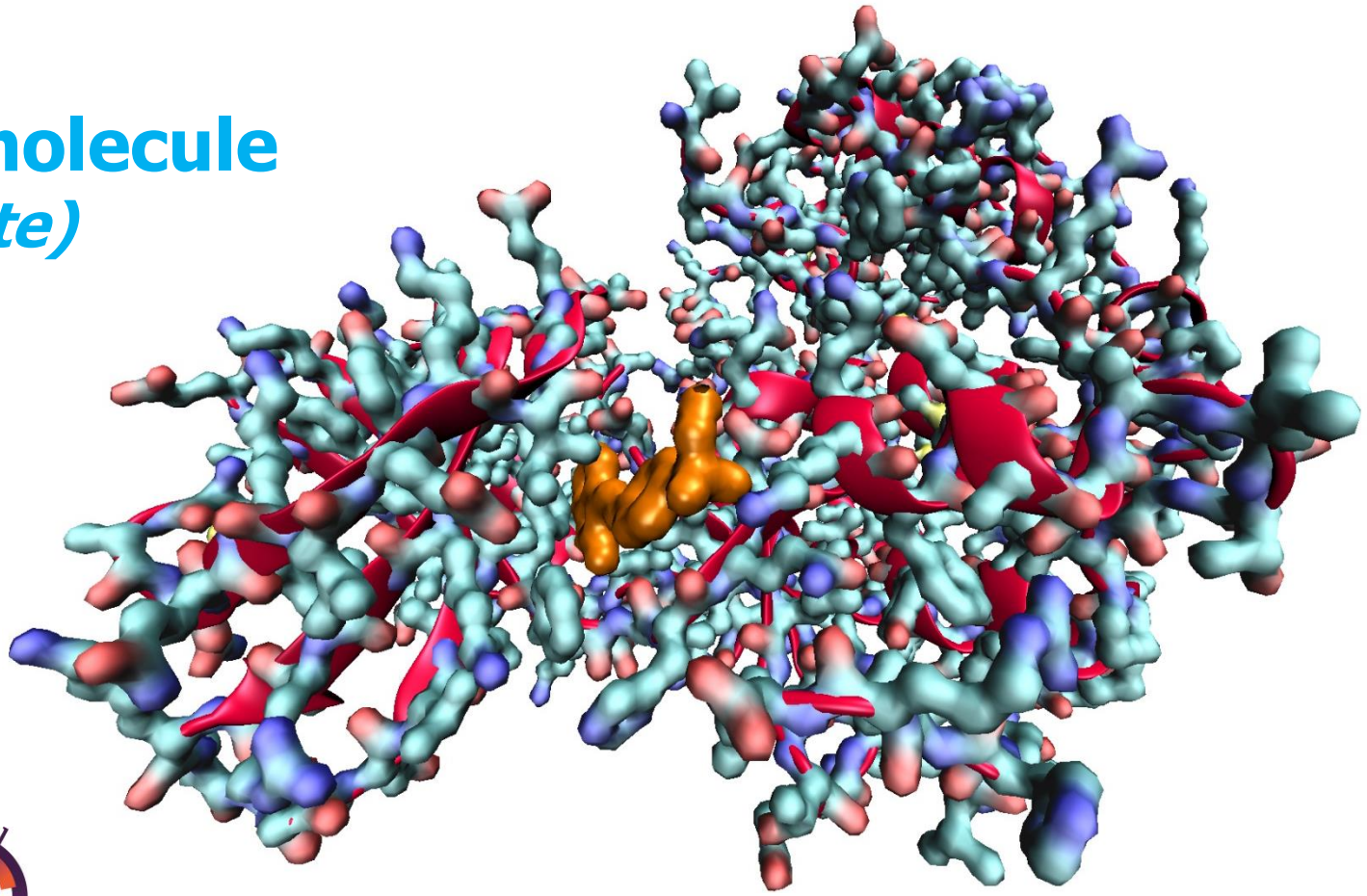
- Context:  docking and scoring functions
- Methods: AlgoGen, NCI
- NCI scoring function on GPU
- Conclusions and perspectives

# Docking

**Macro-molecule**
***(site)***

**+**

**Ligand**





MAISON DE LA SIMULATION de Champagne-Ardenne

Université de Reims Champagne-Ardenne

# Docking tools

- Combination of:
  - A solution representation
    - ⇔quaternion, torsion, …
  - An associated search space according to data flexibility



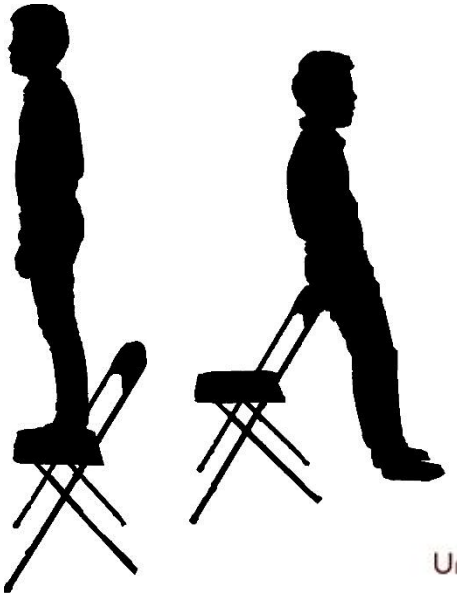
MAISON DE LA SIMULATION de Champagne-Ardenne

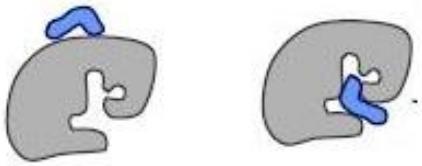
Université de Reims Champagne-Ardenne

# Docking tools

– An associated search space according to data flexibility:

- No flexibility ⇔ rigid docking:
  – Key / lock paradigm
  – Basic good interaction information





# Docking tools

– An associated search space according to data flexibility:
  - No flexibility ⇔ rigid docking:
    – Key / lock paradigm
    – Basic good interaction information
  - Ligand flexibility ⇔ semi-flexible docking:
    – Conformation adaptation of the ligand to fit the site

# Docking tools

– An associated search space according to data flexibility:
  - No flexibility ⇔ rigid docking:
    – Key / lock paradigm
    – Basic good interaction information
  - Ligand flexibility ⇔ semi-flexible docking:
    – Conformation adaptation of the ligand to fit the site
  - Ligand and site flexible ⇔ full-flexible docking.
    – Case of unapproachable site.
    – Depending of the molecule size: from conformation adaptation of the lateral chains to backbone folding

# Docking tools

– An optimization procedure:
  - Only one method:
    – genetic algorithm, ant/bee colony, …
  - cooperative approaches:
    – Lamarckian algorithm, …

– A scoring function

⇔evaluation of the ligand/site complex quality

⇔ energy (main objective)

# Scoring functions

- Parameterized force field:
  - Empirical definition of molecular interactions
  - Pros:
    - Very fast ⇔ only few seconds on big systems
    - Well integrated in tool suite: Autodock, Glide, …
    - Enables full-flexible docking

# Scoring functions

- Parameterized force field:
  - Empirical definition of molecular interactions
  - Cons:
    - Each molecular family ⇔ specific parameters
    - Not able to describe all realistic interactions
    - Substantial input preparation needed

# Scoring functions

- Quantum mechanics:
  - Strict exploitation of electronic information
  - Pros:
    - No need of (empirical) parameters
    - All the interactions can be described
    - No specific input preparation

# Scoring functions

- Quantum mechanics:
  - Strict exploitation of electronic information
  - Cons:
    - Very (very) slow: several hours to days for small systems
    - Not (yet) dedicated for docking analysis:

      ⇔Rigid docking only

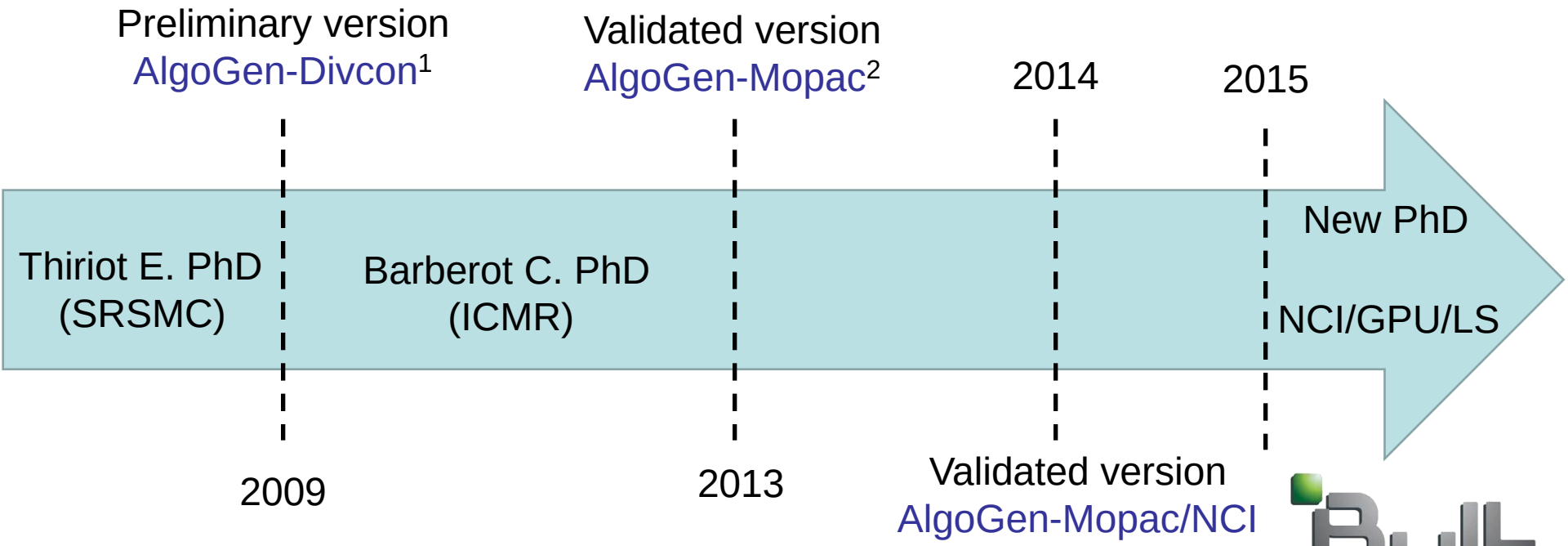
# Outline

- Context:  docking and scoring functions
- Methods: AlgoGen, NCI
- NCI scoring function on GPU
- Conclusions and perspectives

# AlgoGen

- Framework for rigid quantum docking based on:
  - A genetic algorithm as optimization method
  - No specific evaluation scoring:
    - Divcon, Mopac, …
    - Gaussian, …
  - A master/slave parallel model

# Algogen




[1]Thiriot, E.; Monard, G. *THEOCHEM.* **2009**, *898*, 31–41.

[2]Barberot and al., Comp.Theor. Chem. **2014**, *1028*, 7-18.



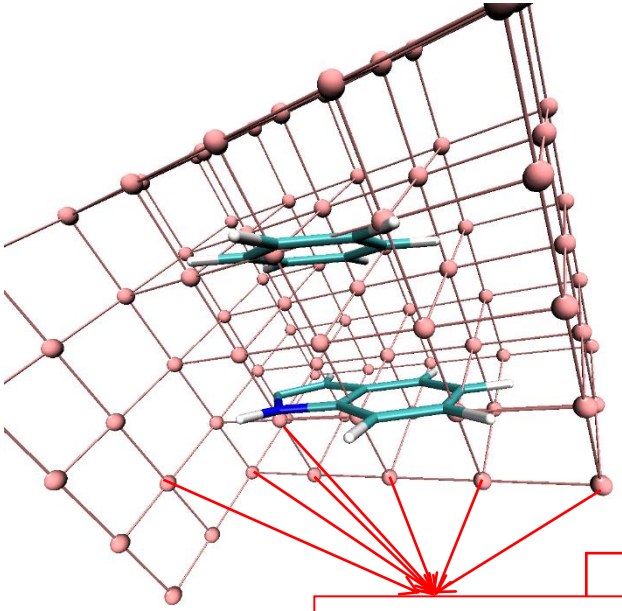
MAISON DE LA SIMULATION de Champagne-Ardenne

Université de Reims Champagne-Ardenne

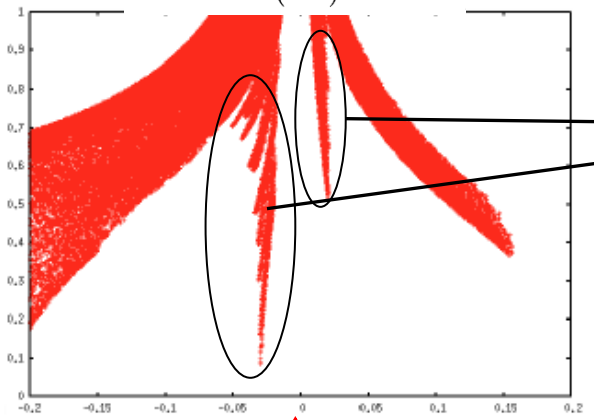
# NCI

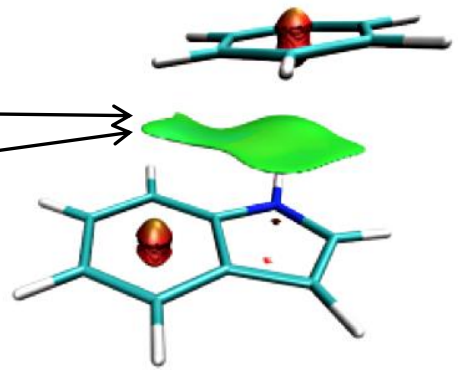
- New method to predict, visualize and interprete Contreras-Garcia, J. and al, *J. Phys. Chem. A*. **2011**,115, 12983.

## Non Convalent molecular Interactions



$$RDG(r) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho(r)|}{\rho(r)^{4/3}}$$



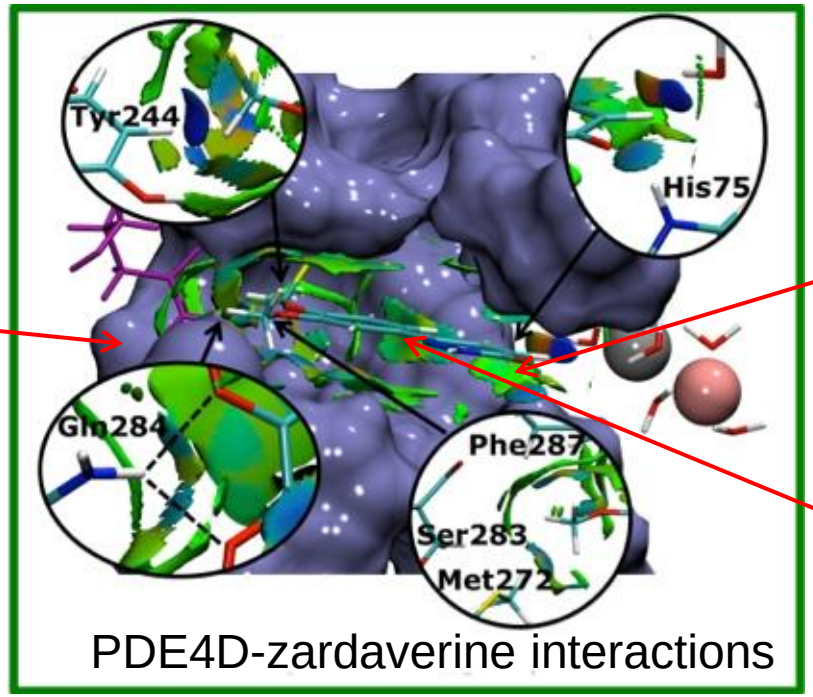
Electron density ρ(r)
Electron density gradient ∇ρ(r)
Electron density hessien

# NCI Post-treatment

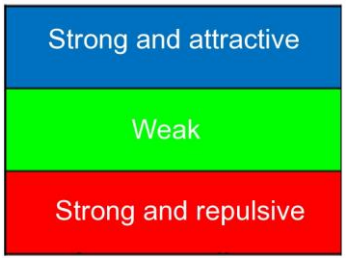
PhosphoDiesterase 4D

NCI interaction surfaces

Zardaverine inhibitor




PDE4D-zardaverine interactions



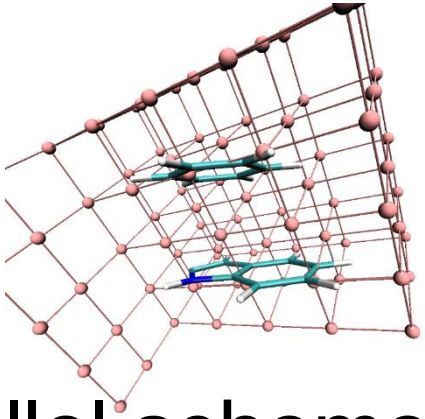



MAISON DE LA SIMULATION de Champagne-Ardenne

Université de Reims Champagne-Ardenne

# NCI as a score

- NCI: based on a grid of atom interactions describing attraction/repulsion forces
- Each point can be computed individually



- Natural parallel scheme:

➔ from NCI grid to GPU grid

# Outline

- Context:  docking and scoring functions
- Methods: AlgoGen, NCI
- NCI scoring function on GPU
- Conclusions and perspectives

# Methodology

- Direct use of Fortran code to CUDA
- Isolation of specific structures and transformation to one-dimension arrays
- Thread repartition with redundant calculi

# Input data

- Test on 3 quantum instances +1 molecular docking instance (CCDC Astex dataset)

| Instance Name | Number of atoms in the NCI Grid |
|---|---|
| 3bench2 | 313 |
| 4bench3 | 326 |
| 5bench4 | 497 |
| 6rsa | 1666 |

# Romeo HPC Tesla Cluster

## Computing

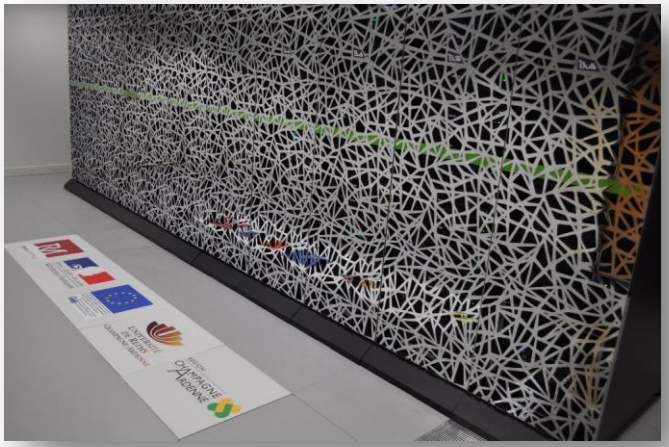
**5th** **3131 MFLOPS/W**
Bull **Cool Cabinet Door**

**151th** **254.9 Tflops**
Linpack

260 NVIDIA Tesla
**K20X** accelerators

**130 Bull servers**
bullx R421 E3 – Bull AE & MPI

**260 INTEL Ivy Bridge** E5-2650 v2 Processor, non-blocking **Mellanox Infiniband,** Slurm, 88 To Lustre (NetApp), 57 To home, 100 To Storage



## Displaying

**Big Data, on-demand and remote**

**VirtualGL** technology servers
Quadro 6000 & 5800

**NVIDIA GRID** + Citrix Virtualisation
**NVIDIA VGX K2**

**Scalable Graphics** 3D cloud solution
**NVIDIA K6000**

# GPU Accelerator

- Nvidia Tesla K20X (Kepler):
  - 2688 processor cores
  - 6 GB GDDR5
  - Peak performance:
    - 1.31 Tflops (double-precision floating point)
    - 3.95 Tflops (single-precision floating point)

Centre de Calcul ROMEO
MAISON DE LA SIMULATION de Champagne-Ardenne

# Proof of concept results

- CPU Intel Ivy Bridge (8 cores) vs Tesla K20X:
  - Equivalent purchase and exploitation price
- Sequential CPU vs :
  - OpenMP (8): computation time / 4
  - Tesla K20X: computation time / 300
- OpenMP (8) vs Tesla K20X
  - Computation time / 75

# AlgoGen NCI GPU

- Extrapolated results:
  - AlgoGen NCI (on a small system)
    - CPU version ⇔ 16000 evaluations * 2min
      ➔ 22 days
    - GPU version ⇔ 16000 evaluations * 0.4 s
      ➔ < 2h

# Outline

- Context:  docking and scoring functions
- Methods: AlgoGen, NCI
- NCI scoring function on GPU
- Conclusions and perspectives

# Conclusions and perspectives

- The proof of concept is valid
- Next steps:
  - Production phase
  - Pipeline of evaluations
  - NCIPLOT code extraction and optimization

# Conclusions and perspectives

- Application of NCI to docking
  - submitted French ANR project by NCI authors (E-NERGY).
- New PhD:
  - New scoring methods
    - Including collaboration with the authors of DFTB codes (CSC group, Brême, Germany; LCPQ Toulouse, France, LCT group, Paris, France)
  - Flexibility management
    - Including collaboration with Marie Brut (LAAS Toulouse)