# Molecular dynamics applications on Grid

Presented by

Vinod Kasam

EGEE07, October 4th 2007, Budapest


# Outline

- Molecular dynamics
- Computational grids
- Amber software
- Production environment
- Deployment strategy
- Initial results
- Issues
- Future steps

# Real Data to perform molecular dynamics: WISDOM project

**Targets**

Different targets from Protein database and homology models

**Chemical Compounds**

ZINC compounds

**Docking tools**

FlexX

**Grid Infrastructure**

EGEE, EELA,EUChinaGrid

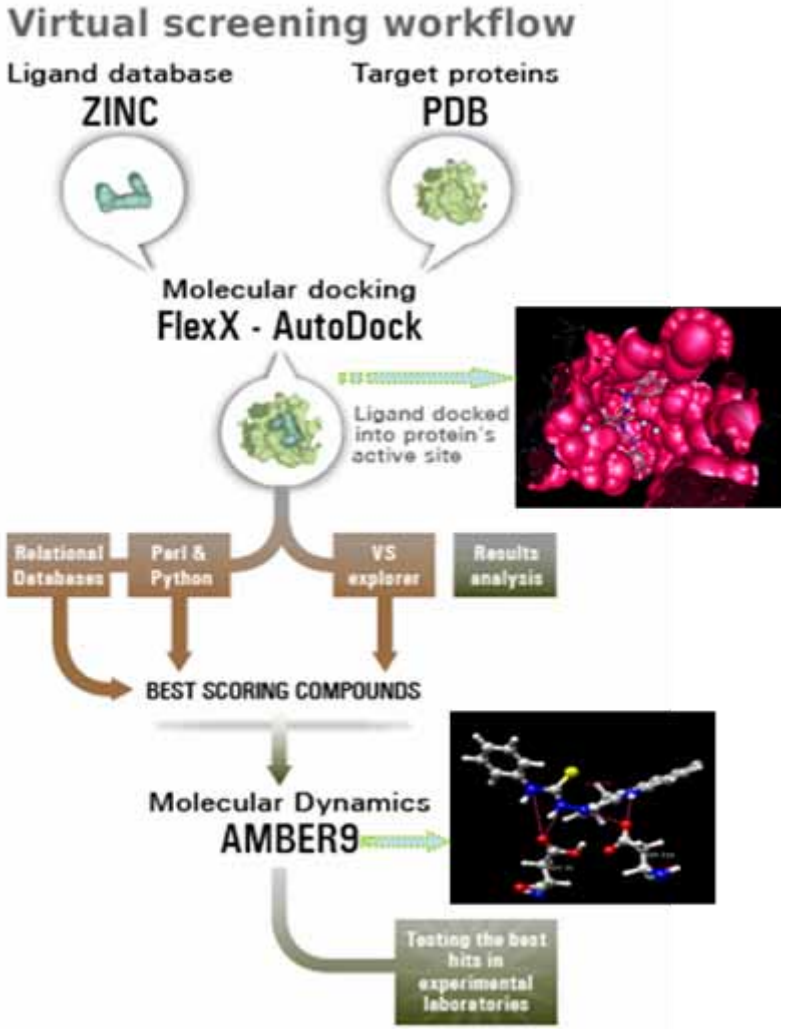
**Results**

Python and Perl scripts
VS explorer
MySQL databases

**Rescoring**

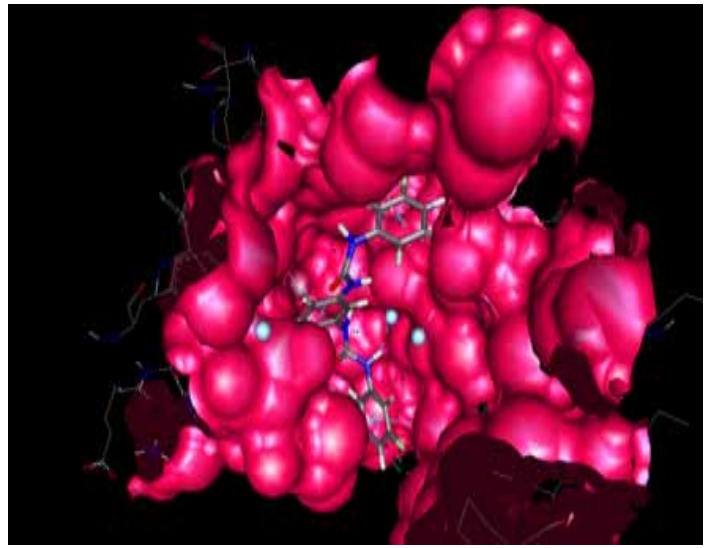
AMBER

MMPBSA and MMPGBSA



# Molecular dynamics and Grids

- Molecular Dynamics = computation of the motion of atoms within a molecular system using molecular mechanics



- Molecular Dynamics is commonly used for drug design and drug discovery
  - Molecular modelling of drugs
  - Measurement of binding energies between ligands and biological targets
  - Addresses the issues of
    - electrostatic solvation parameters
    - Protein flexibility
    - Calculates accurate binding energies

Grids offer promising perspectives for *in silico* drug discovery
Identification of drug candidates using computing tools

**Virtual screening: Screening large compound libraries to find novel hits against a particular target or a disease**

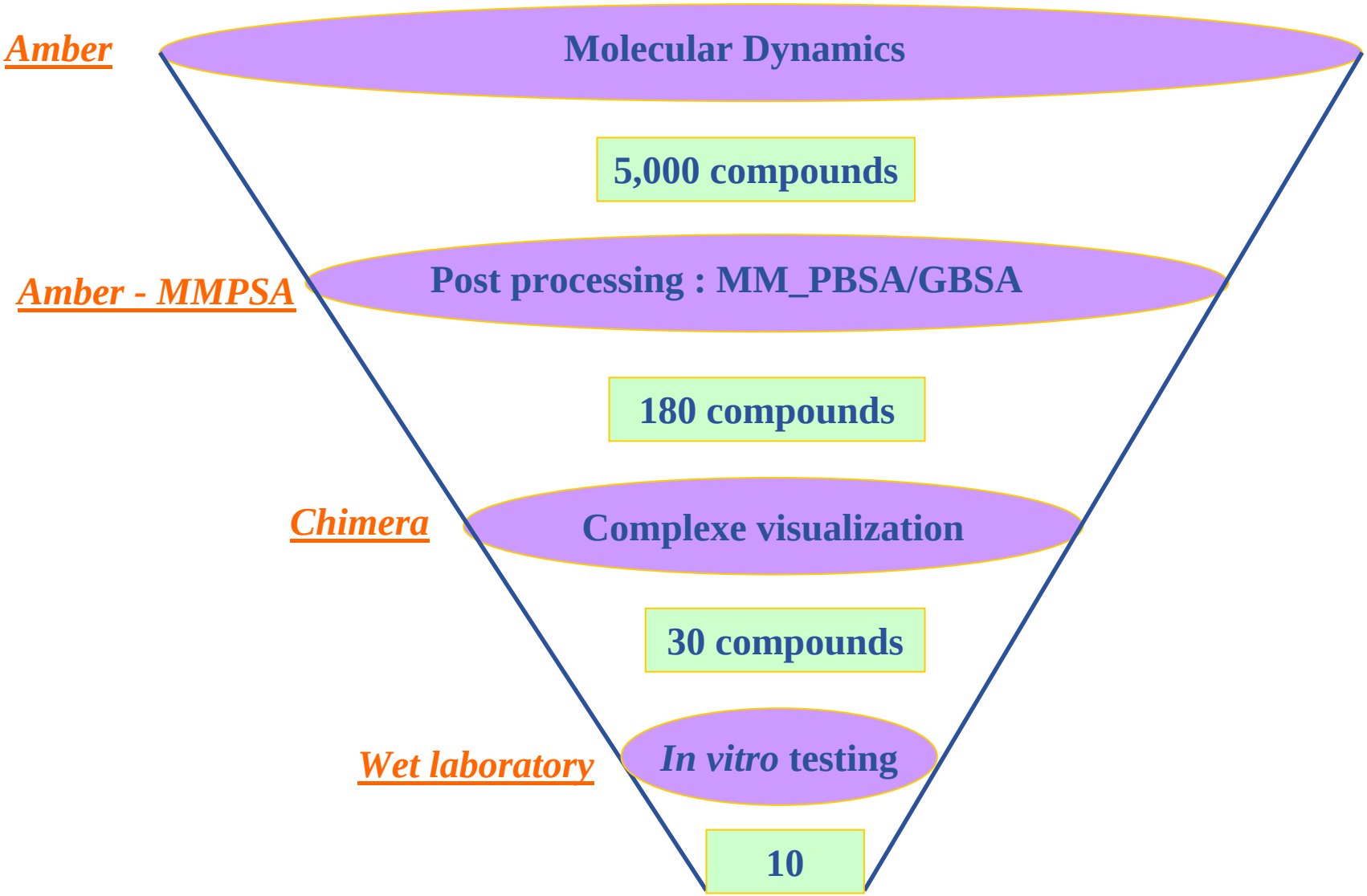
# Criteria for MD software selection

- Criteria for MD software selection
  - Local expertise
  - Type of simulations to perform (protein folding or re-ranking after virtual screening, etc)
  - Licensing
  - Choice of infrastructure (Linux/Unix/Windows)
  - Robustness, reliability and available literature
- Motivations for the choice of Amber9
  - Very well suited for protein ligand complex simulations
  - Deployed on many operating systems including linux
  - Very well known and fully documented MD software
  - Running collaboration with groups at Univ. Modena and SCAI Fraunhofer experienced with Amber
  - Issue: licensed software
- Other MD software considered: Gromacs

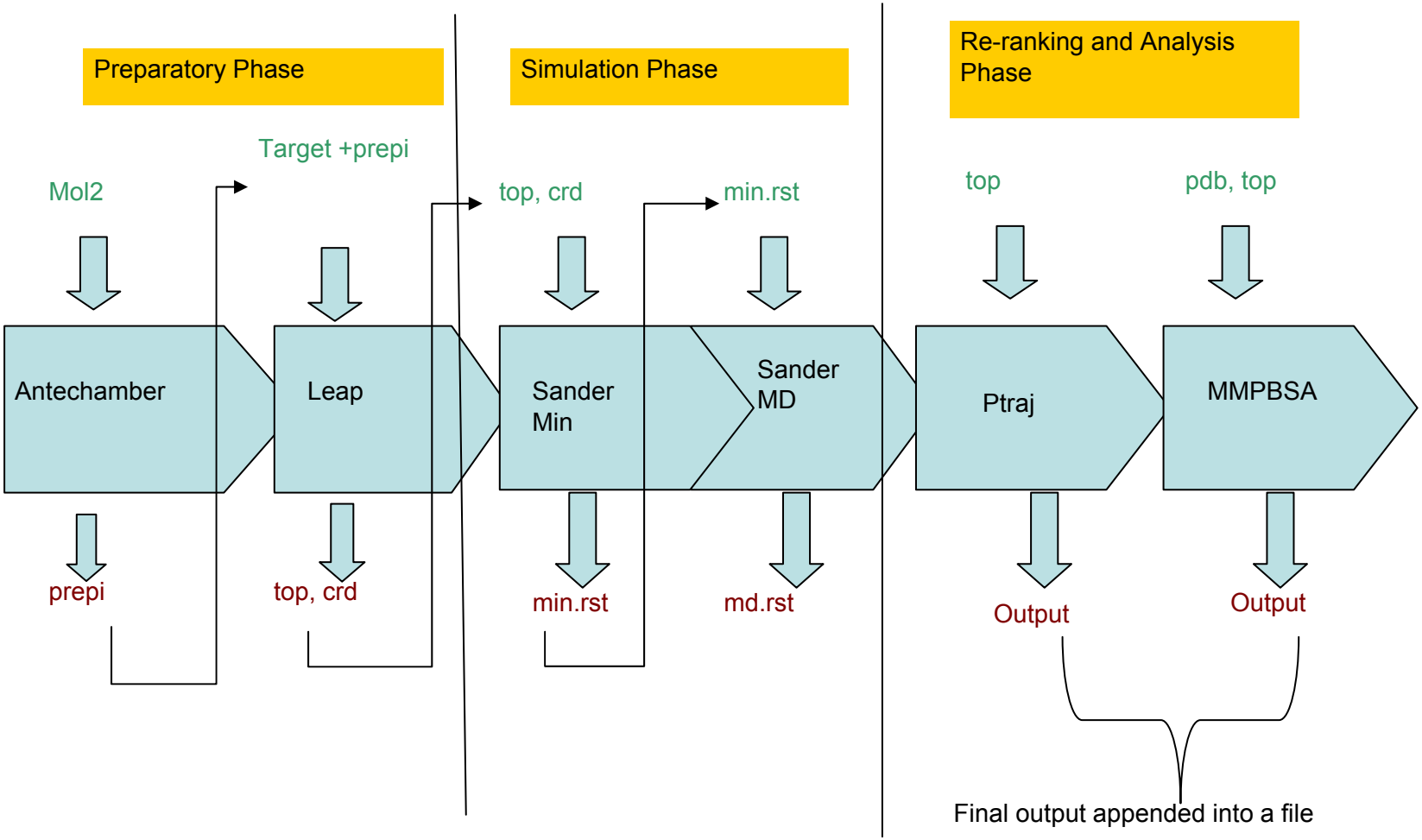
# Criteria for grid infrastructure selection

- Grids of supercomputers and grids of PCs are relevant environment for MD computations
  - Grids of PCs have limited memory (2GB today on EGEE): limitation for large molecules and memory greedy software
  - Supercomputers are harder to access
- Criteria for selection
  - Local expertise
  - Type of simulations to perform (memory requirements, CPU time)
  - Accessibility
  - Production environment
- Infrastructure selected: EGEE
  - Easy access to resources
  - No CPU constrainsts
  - Previous experience and existing production environment

# Molecular dynamics Workflow



# Amber software architecture




**Management of the Interdependent steps is challenging**

# MMPBSA approach

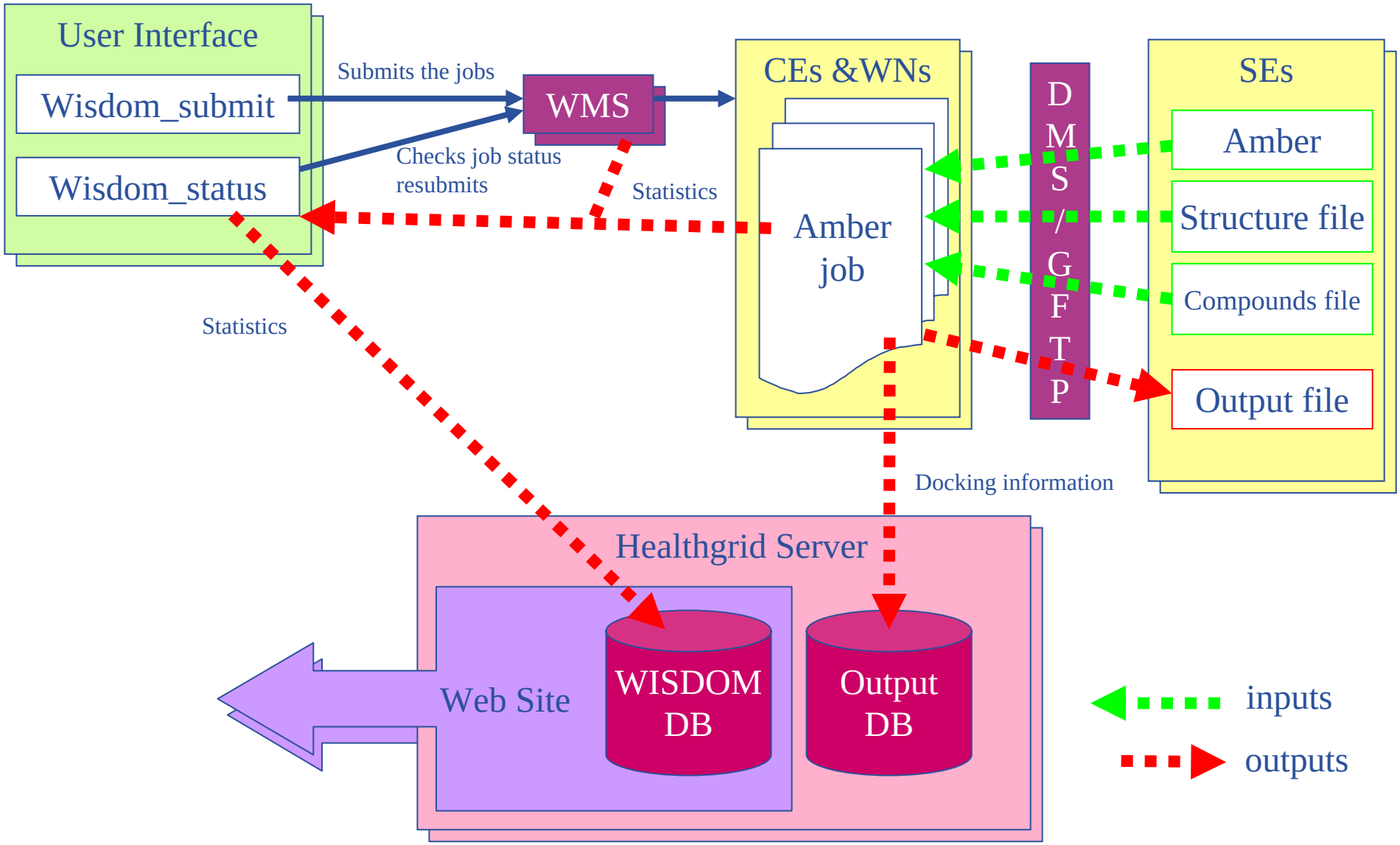
- MM-PBSA is a program which is used to estimate energies and entropies from the snapshots contained within trajectory files.
- The calculations are organized and spawned by a Perl script
- mm-pbsa, collects statistics and formats the output in tabular form.
- The analysis is primarily based on continuum solvation models.
  - MM-PBSA
  - MM-GBSA

*A. Ferrari, G. Degliesposti, M. Sgobba, G. Rastelli. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. Bioorganic & Medicinal Chemistry. 2007. In Press.*

# Molecular dynamics procedure

- Protocol: Distance-dependent dielectric energy minimization followed by molecular dynamics on the active site as well as ligand atoms and final re-minimization.
- The ΔGbind value of each inhibitor can be calculated according to the equation
- $\Delta Gbind = \Delta E_{MM} + \Delta Gsolv - T\Delta Ssolute$,
  - $E_{MM}$ is the molecular mechanics contribution expressed as the sum of the internal, electrostatics and van der Waals contributions to binding *in vacuo*,
  - Gsolv is the solvation free energy contribution to binding expressed as the sum of polar and nonpolar solvation free energies ($\Delta Gsolv = \Delta Gpsolv + \Delta Gnpsolv$, respectively),
  - $T\Delta Ssolute$ is the contribution of solute entropy to binding.

# Production environment



# Deployment strategy

- Deployment is segmented into 3 phases
  - Store all the necessary input on the Storage Element of the grid
    - 5000 chemical compound multifile made into 100 fragments
    - Target (1LEE: Protein database ID)
    - Bash Script to the run the Amber job
  - Run the jobs
    - Running the jobs on Clermont cluster by using WISDOM production environment
  - Checking the status of the jobs
    - Done automatically (developed for WISDOM data challenges)

# Licensing and issues

- Licensing
  - Amber is commercial software with an academic license for computers, like a cluster, of an institute.
  - The license is integrated in the software package (so no license server required)
  - Amber is easily used on the EGEE grid
- Issues
  - One license per grid site deploying Amber
  - Use of Amber restricted to grid users coming from institutes owning a license
- Issue addressed
  - Access granted to all the nodes where Amber is installed
    - Amber is deployed on SCAI Fraunhofer

# MM_PBSA / GBSA ranking

BioinfoGRID

## Procedure : Prof. Giulio Rastelli, University of Modena, Italy

| NAME | RMS | GAS | PBSOL | PBTOT | GBSOL | GBTOT |
|---|---|---|---|---|---|---|
| test2_3558.mol2 | 3.27030 | 174.46 | -341.7 | **-167.25** | -199.69 | -25.23 |
| test2_1427.mol2 | 1.58468 | -85.89 | -31.21 | **-117.09** | 28.20 | -57.69 |
| test2_4359.mol2 | 2.17294 | -319.54 | 212.27 | **-107.27** | 274.48 | -45.06 |
| test2_0772.mol2 | 2.26663 | -78.69 | -3.81 | **-82.50** | 39.25 | -39.44 |
| test2_1852.mol2 | 1.36253 | -116.85 | 75.45 | **-41.40** | 46.25 | -70.59 |
| test2_3217.mol2 | 1.91890 | -95.37 | 54.03 | **-41.34** | 37.71 | -57.66 |
| test2_3968.mol2 | 1.82006 | -337.27 | 295.96 | **-41.31** | 275.73 | -61.54 |
| test2_0890.mol2 | 2.83841 | -108.20 | 69.07 | **-39.13** | 50.78 | -57.42 |
| test2_2685.mol2 | 1.26087 | -86.07 | 47.61 | **-38.46** | 37.45 | -48.62 |
| test2_0406.mol2 | 1.19697 | -81.45 | 43.15 | **-38.31** | 28.88 | -52.57 |
| test2_2007.mol2 | 1.25467 | -83.54 | 46.40 | **-37.15** | 27.51 | -56.04 |
| test2_0674.mol2 | 1.56722 | -119.16 | 82.03 | **-37.13** | 59.65 | -59.51 |
| test2_4083.mol2 | 1.28737 | -340.55 | 303.50 | **-37.05** | 285.44 | -55.11 |
| test2_4085.mol2 | 1.96007 | -346.54 | 309.74 | **-36.79** | 287.85 | -58.69 |
| test2_2027.mol2 | 1.19531 | -83.49 | 46.79 | **-36.70** | 27.26 | -56.22 |
| test2_3918.mol2 | 1.04100 | -388.52 | 351.98 | **-36.53** | 325.87 | -62.64 |
| test2_0174.mol2 | 1.37354 | -101.15 | 65.09 | **-36.07** | 43.61 | -57.54 |
| test2_3996.mol2 | 3.56394 | -328.43 | 292.61 | **-35.81** | 274.20 | -54.22 |
| test2_1052.mol2 | 1.31059 | -81.38 | 45.74 | **-35.64** | 30.26 | -51.12 |
| test2_4572.mol2 | 3.87161 | -335.24 | 299.61 | **-35.63** | 273.70 | -61.54 |
| test2_3833.mol2 | 1.82142 | -384.89 | 349.36 | **-35.54** | 325.93 | -58.96 |

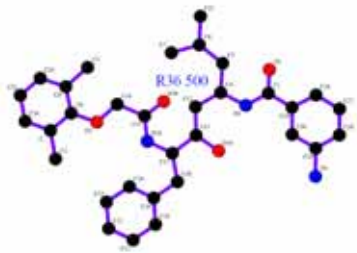
**And so on ...**

| NAME | RMS | GAS | PBSOL | PBTOT | GBSOL | GBTOT |
|---|---|---|---|---|---|---|
| test2_1852.mol2 | 1.36253 | -116.85 | 75.45 | -41.40 | 46.25 | **-70.59** |
| test2_4378.mol2 | 1.19950 | -340.36 | 306.22 | -34.14 | 273.48 | **-66.88** |
| test2_3759.mol2 | 1.97719 | -353.43 | 324.43 | -29.00 | 287.06 | **-66.37** |
| test2_4727.mol2 | 1.89557 | -569.94 | 537.53 | -32.41 | 504.08 | **-65.86** |
| test2_4835.mol2 | 1.03069 | -602.13 | 583.63 | -18.49 | 538.52 | **-63.60** |
| test2_4362.mol2 | 1.39490 | -333.77 | 299.48 | -34.29 | 270.49 | **-63.28** |
| test2_3918.mol2 | 1.04100 | -388.52 | 351.98 | -36.53 | 325.87 | **-62.64** |
| test2_3765.mol2 | 1.32300 | -338.18 | 315.68 | -22.49 | 275.95 | **-62.22** |
| test2_4878.mol2 | 1.58222 | -558.25 | 556.27 | -1.98 | 496.43 | **-61.82** |
| test2_4849.mol2 | 1.19517 | -566.30 | 560.14 | -6.16 | 504.51 | **-61.79** |
| test2_4802.mol2 | 1.58881 | -571.23 | 555.46 | -15.77 | 509.47 | **-61.76** |
| test2_3968.mol2 | 1.82006 | -337.27 | 295.96 | -41.31 | 275.73 | **-61.54** |
| test2_4572.mol2 | 3.87161 | -335.24 | 299.61 | -35.63 | 273.70 | **-61.54** |
| test2_4568.mol2 | 1.72708 | -350.38 | 325.43 | -24.95 | 288.87 | **-61.51** |
| test2_4556.mol2 | 2.23472 | -348.02 | 318.30 | -29.71 | 286.55 | **-61.47** |
| test2_4338.mol2 | 2.42596 | -339.52 | 310.54 | -28.99 | 278.14 | **-61.39** |
| test2_0222.mol2 | 2.26272 | -86.83 | 59.84 | -26.99 | 25.55 | **-61.28** |
| test2_4065.mol2 | 3.10552 | -327.10 | 295.45 | -31.65 | 265.83 | **-61.27** |
| test2_4094.mol2 | 2.27359 | -339.42 | 306.96 | -32.46 | 278.36 | **-61.07** |
| test2_4037.mol2 | 1.74182 | -357.54 | 320.12 | -37.42 | 296.49 | **-61.05** |
| test2_4458.mol2 | 1.53125 | -331.02 | 307.86 | -23.17 | 270.04 | **-60.99** |

**And so on ...**

**Ranking of best 75 complexes according to PB score**

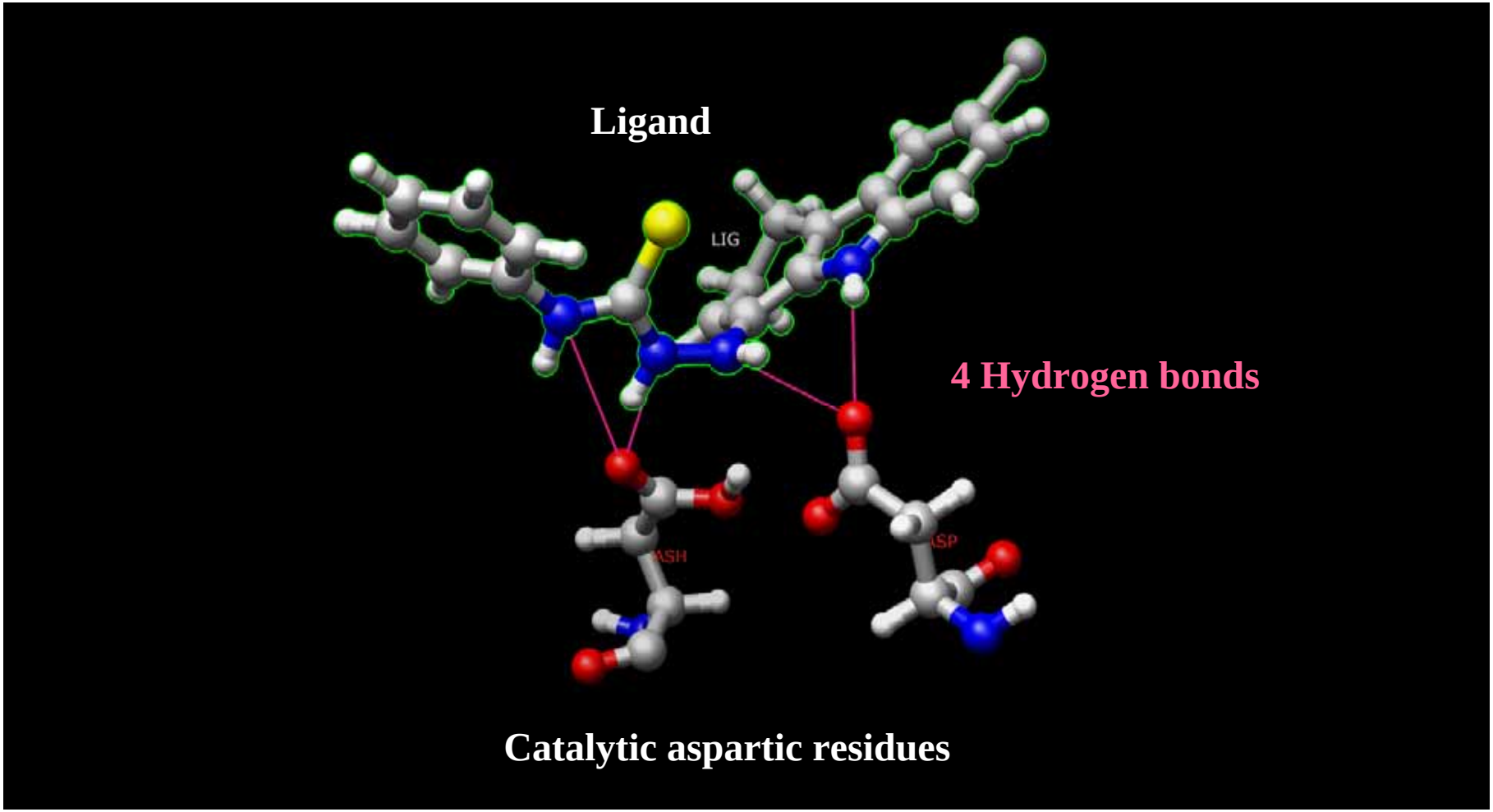
*PBtot R36 = -42.76 kcal/mol*




**Ranking of best 75 complexes according to GB score**

*GBtot R36 = -66,53 kcal/mol*

# Interactions visualized with Chimera




Result_2559

# Results MM_PBSA / GBSA

BioinfoGRID

From Pbtot and Gbtot
From Pbtot
From Gbtot

| | File Name | RMS | PbTot | GbTot | H-bonds Interactions | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | ASH 34 | ASP 214 | VAL 78 | SER 79 | SER 218 | PHE 294 | TYR 192 |
| 1 | test2_0162 | 1.47 | -32.11 | -54.51 | No | LIG N3-ASP OD1<br>LIG N5-ASP OD1 | Yes | Yes | Yes | No | Yes |
| 2 | test2_1511 | 1,18 | -31,68 | -49,72 | LIG N1-ASH OD2 | LIG N1-ASP OD1<br>LIG N2-ASP OD1 | Yes | Yes | Yes | No | No |
| 3 | test2_1866 | 1,6 | -32,47 | -50,41 | LIG N2 ASHOD1 | LIG N1 ASP OD1<br>LIG N1 ASP OD2 | Yes | Yes | Yes | No | No |
| 4 | test2_1941 | 1,75 | -33,17 | -52,2 | LIG N3-ASH OD1<br>LIG N3-ASH OD2 | LIG N4-ASP OD1<br>LIG N4-ASP OD2<br>LIG N1-ASP OD1 | Yes | Yes | Yes | No | No |
| 5 | test2_2020 | 2,01 | -33,57 | -57,99 | LIG N1-ASH OD1<br>LIG N2-ASH OD2 | LIG N2-ASP OD1 | Yes | Yes | Yes | No | No |
| 6 | test2_2062 | 1,36 | -31,64 | -57,6 | LIG N2-ASH OD1 | LIG N1-ASP OD1 | Yes | Yes | No | No | No |
| 7 | test2_2339 | 2,03 | -31,6 | -47,36 | LIG N1-ASH OD1<br>LIG N2-ASH OD1 | LIG N3-ASP OD1<br>LIG N4-ASP OD1 | Yes | Yes | No | No | No |
| 8 | test2_3293 | 1,76 | -32,73 | -51,57 | ASH OD2-LIG N2 | LIG N3-ASP OD1<br>LIG N3-ASP OD2 | Yes | Yes | Yes | No | No |
| 9 | test2_3759 | 1,97 | -29 | -66,37 | LIG N5-ASH OD1 | LIG N2-ASP OD1<br>LIG N3-ASP OD1 | Yes | Yes | No | No | No |
| 10 | test2_3762 | 1,62 | -22,19 | -59,4 | LIG N5-ASH OD1 | LIG N3-ASP OD1 | No | Yes | Yes | No | No |

# Results MM_PBSA / GBSA

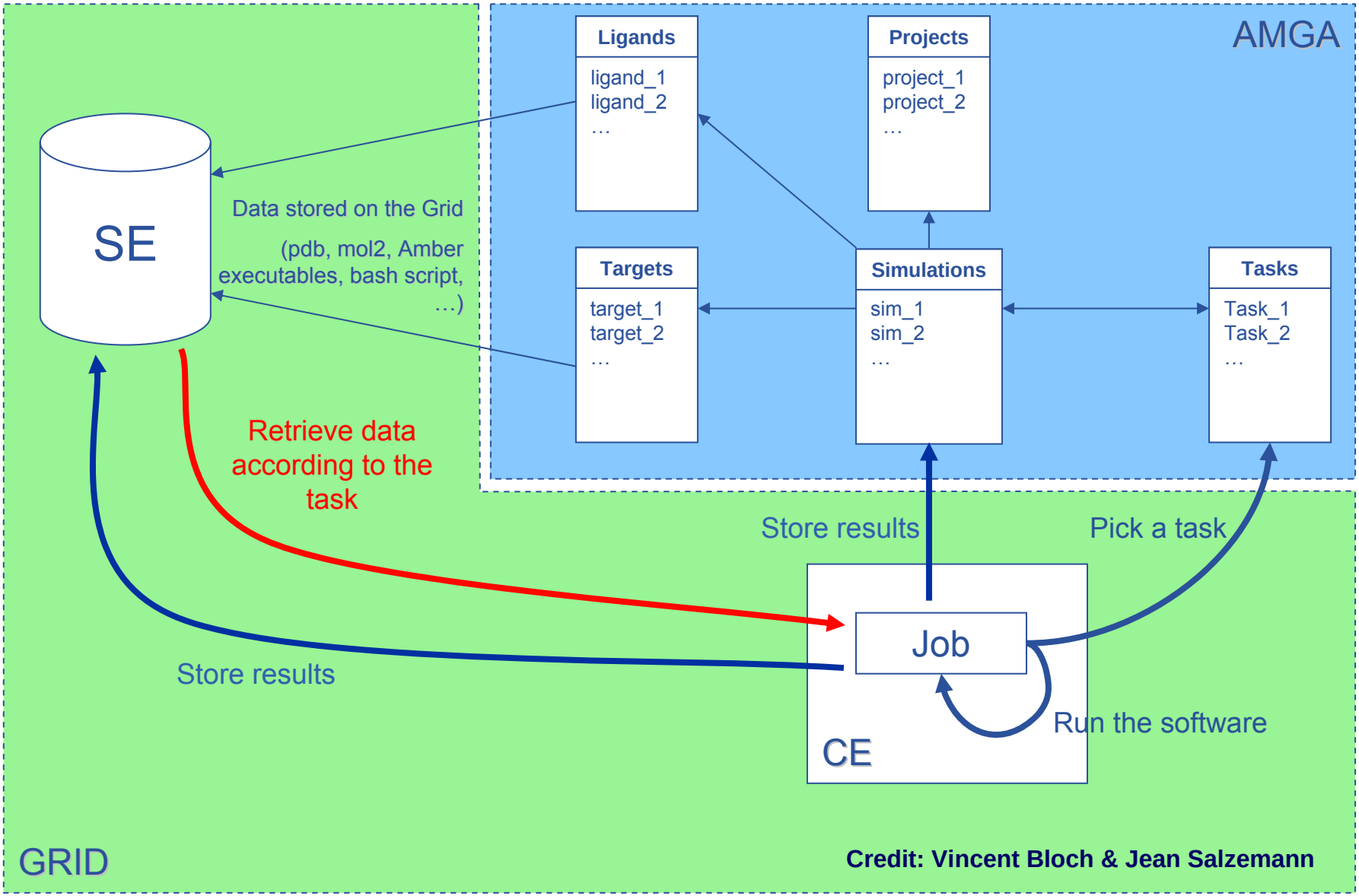
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | test2_3781 | 1,18 | -11,02 | -57,66 | LIG N1-ASH OD1<br>LIGN1-ASH OD2<br>ASH OD2-LIG O5 | LIG N1-ASP OD1 | Yes | Yes | No | No | Yes |
| 12 | test2_3918 | 1,04 | -36,5 | -62,6 | ASH OD2-LIG O3 | LIG N3-ASP OD1<br>LIG N4-ASP OD1 | Yes | Yes | No | No | No |
| 13 | test2_3920 | 1,62 | -27,32 | -57,87 | ASH OD2-LIG N2<br>LIG N2-ASH OD1 | LIG N5-ASP OD1 | Yes | Yes | No | No | Yes |
| 14 | test2_3957 | 1,26 | -31,81 | -54,21 | ASH OD2-LIG N2 | LIG N1-ASP OD1<br>LIG N1-ASP OD2<br>LIG N2-ASP OD2 | No | Yes | No | No | No |
| 15 | test2_3968 | 1,82 | -41,31 | -61,54 | LIG N2-ASHOD1<br>LIG N3-ASHOD1 | LIG N1-ASP OD1 | Yes | Yes | No | No | No |
| 16 | test2_3996 | 3,56 | -35,81 | -54,22 | LIG N3-ASH OD1 | LIG N3-ASP OD1 | Yes | No | No | No | No |
| 17 | test2_4094 | 2,27 | -32,46 | -61,07 | ASH OD2-LIG O3 | LIG N1-ASP OD1<br>LIG N1-ASP OD2<br>LIG N2-ASP OD1<br>LIG N2-ASP OD2 | Yes | No | Yes | No | No |
| 18 | test2_4321 | 1,15 | -32,9 | -55,36 | LIG N3-ASH OD1 | LIG N3-ASP OD1 | Yes | Yes | No | No | Yes |
| 19 | test2_4362 | 1,39 | -34,29 | -63,28 | ASH OD2-LIG N2 | LIG N1-ASP OD1<br>LIG N2-ASP OD1<br>LIG N2-ASP OD2 | No | Yes | No | No | No |
| 20 | test2_4378 | 1,19 | -34,14 | -66,88 | LIGN5-ASH OD1<br>LIGN6-ASH OD1<br>LIGN6-ASH OD2 | LIG N6-ASP OD1<br>LIG N6-ASP OD2 | Yes | Yes | No | No | No |
| 21 | test2_4425 | 1,84 | -29,8 | -56,77 | LIG N3-ASH OD1 | LIG N2-ASP OD1<br>LIG N6-ASP OD1 | Yes | Yes | No | No | No |
| 22 | test2_4568 | 1,72 | -24,95 | -61,51 | LIG N3-ASH OD1<br>LIG N4-ASH OD1 | LIG N1-ASP OD1 | Yes | Yes | No | No | No |

# Achievements so far and Future steps

- Achievements
  - Successfully deployed Amber software on grid infrastructure
  - The best 5000 compounds comming from WISDOM-I are tested
  - Absolutely useful in rational drug discovery

- Future steps
  - Deployment of the same procedure at other grid nodes
    - kalkan1.ulakbim.gov.tr
    - University of Modena
  - Fine tuning of the parameters
  - Simulations with full protein flexibility
  - The bests compounds from WISDOM-II will be refined by using the same procedure:
    - Best 15000 compounds against PfDHFR are identified for molecular dynamics simulations
    - Compounds against GST is under process
  - Finally selected compounds will be tested in the experimental laboratories

# New production environment




**Credit: Vincent Bloch & Jean Salzemann**

# Acknowledgments


- *BIOINFOGRID* http://www.bioinfogrid.eu

- Giulio Rastelli (Univ. Modena)
- Jean Salzemann
- Ana Da Costa
- Vincent Bloch
- Astrid Maass (Scai-Fraunhofer)
- Nicolas Jacq
- Vincent Breton
- WISDOM collaboration