TOLEDO: a new software that is able to run and analyze Desmond Molecular Dynamics simulations using parallel programming

Miguel Carmena-Bargueño, Carlos Martínez-Cortés, Antonio Jesús Banegas Luna, Horacio Pérez-Sánchez

Universidad Católica de Murcia (UCAM)

BIO-HPC
Structural Bioinformatics and High Performance Computing Research Group

II Symposium of Chemial and Physical Sciences for Young Researches
Murcia 22-24 March

STRUCTURAL BIOINFORMATICS AND HIGH PERFORMANCE COMPUTING (BIO-HPC) RESEARCH GROUP (2022 presentation)

- 4 research full-time UCAM members, 3 part-time UCAM, 3 external
- 4 Main objectives/pillars
  - Method/tool developments
  - Implementations in supercomputers
  - Web-access to tools
  - Application to pharma/biotech/food contexts
- International collaborations and companies
- Scientific output since 2013
  - More than 140 JCR publications (Q1, Q2)
  - 10 patents filed (3 licensed), 2 patents in process
  - 8 PhD students supervised, 4 PhD in progress
  - 13 competitive research projects as PI (regional, national and international calls)
  - Contracts with companies
  - Several software and web tools developed

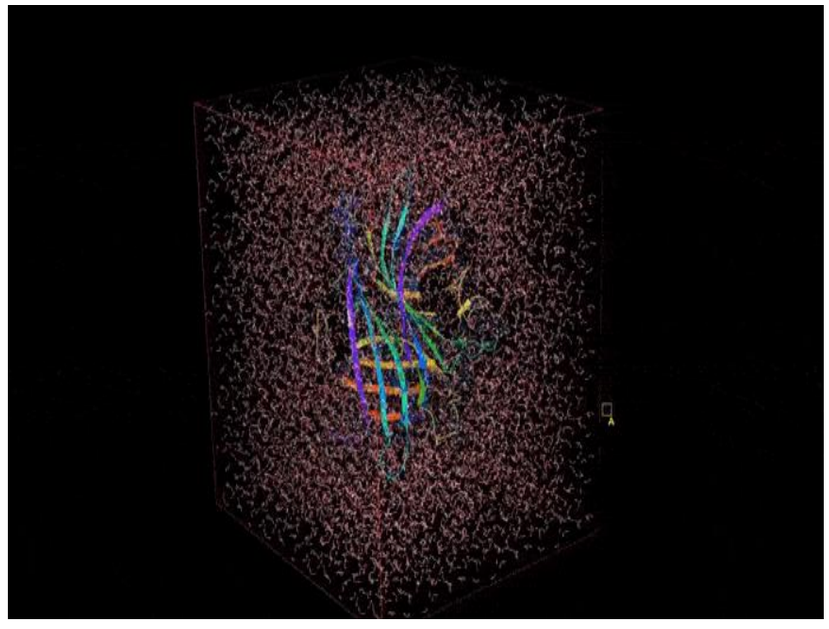
INTRODUCTION:

WHAT IS MOLECULAR DYNAMICS?

Molecular Dynamics Simulations (MD)

Allow to simulate evolution of protein-ligand complex along short time (100ns)

Software use GPUs



Cheaper and less time-consuming that *in vitro/ in vivo* tools

More accuracy that other *in silico* tools (Docking)

COMPLEXITY OF RUN MDs TO NON-EXPERT USERS

PROBLEM:

High perfomance was obtained using **rCUDA** and **Gromacs** by our **group (BIO-HPC)**

Table 1. Performance achieved by several GROMACS configurations.

Configuration	Label	Performance (ns/day)
CPU 20 threads	A	121.33
CPU 3 threads	B	22.59
CUDA 10 threads	C	305.46
rCUDA 2 nodes: eight 3-thread instances	D	542.53
rCUDA 1 node: five 3-thread instances	E	452.64

rCUDA: remote CUDA; GROMACS: GROningen MAchine for Chemical Simulations.

-GROMACS needs expert users to be configured
-High number of parameters

→

```
61  ; Output frequency for coords (x), velocities (v) and forces (f)
62  nstxout                 = 0
63  nstvout                 = 0
64  nstfout                 = 0
65  ; Output frequency for energies to log file and energy file
66  nstlog                  = 1000
67  nstcalcenergy           = 1000
68  nstenergy               = 1000
69  ; Output frequency and precision for .xtc file
70  nstxout-compressed      = 100
71  compressed-x-precision  = 1000
72  ; This selects the subset of atoms for the compressed
73  ; trajectory file. You can select multiple groups. By
74  ; default, all atoms will be written.
75  compressed-x-grps       = Protein
76  ; Selection of energy groups
77  energygrps              = System
78
79  ; NEIGHBORSEARCHING PARAMETERS
80  ; cut-off scheme (Verlet: particle based cut-offs, group: using charge groups)
81  cutoff-scheme           = Verlet
82  ; nblist update frequency
83  nstlist                 = 10
84  ; Periodic boundary conditions: xyz, no, xy
85  pbc                     = xyz
86  periodic-molecules      = no
87  ; Allowed energy error due to the Verlet buffer in kJ/mol/ps per atom,
88  ; a value of -1 means: use rlist
89  verlet-buffer-tolerance  = 0.005
90  ; nblist cut-off
91  rlist                   = 1
92  ; long-range cut-off for switched potentials
93
94  ; OPTIONS FOR ELECTROSTATICS AND VDW
95  ; Method for doing electrostatics
96  coulombtype             = pme
97  coulomb-modifier        = Potential-shift-Verlet
98  rcoulomb-switch         = 0
99  rcoulomb                = 1.0
100 ; Relative dielectric constant for the medium and the reaction field
101 epsilon-r               = 1
102 epsilon-rf              = 0
103 ; Method for doing Van der Waals
104 vdw-type                = Cut-off
105 vdw-modifier            = Potential-shift
106 ; cut-off lengths
107 rvdw-switch             = 0
108 rvdw                    = 1.0
109 ; Apply long range dispersion corrections for Energy and Pressure
```

SOLUTION:

RUN MD USING TOLEDO

(Throughput Optimization of Ligand-Protein system Exploration through Dynamics simulation in Optimized HPC systems)

Run classical MDs (cMDs) using **Maestro-Desmond**

- **Can be used by non-expert users**
- **Low number of parameters**

-Only one MD at the same time

Run multiple cMDs in parallel using **TOLEDO** in **clusters**

- **Allow Run Multiple cMDs at the same time**
- **Allow Run long cMDs (>1000 ns)**

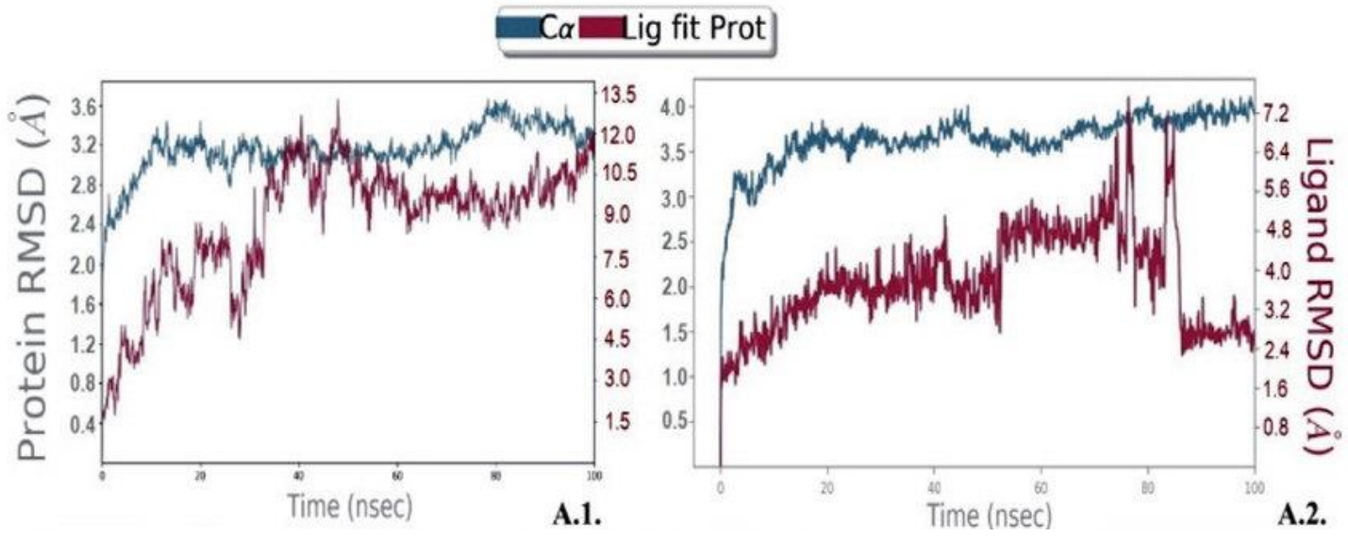
-Needs GPU systems
-Maestro-Desmond is not free (Purchase license)

Prepare the Protein-Ligand complexes

Run cMDs

./Toledo.sh –f complex.txt –t 10000 –p gpu

Obtain Results



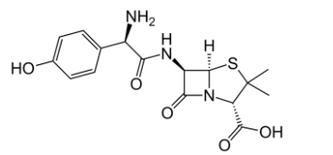
CASE STUDIES

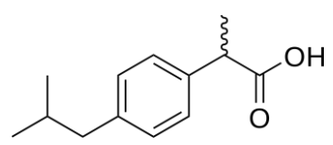
DOCKING RESULTS OBTAINED BY

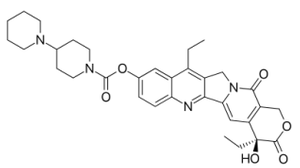
Virtual Screening
Poster 22

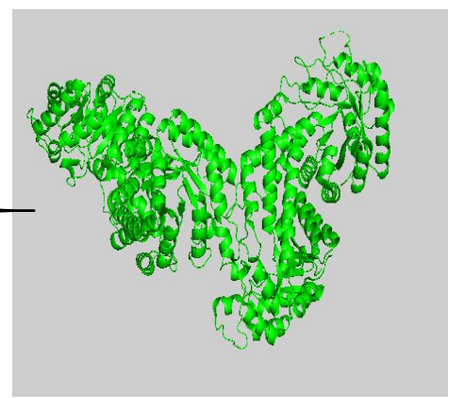
(https://github.com/bio-hpc/metascreener)

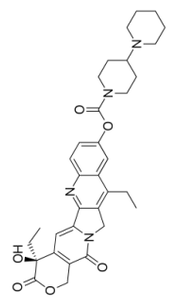
BIO-HPC
Structural Bioinformatics and High Performance Computing Research Group





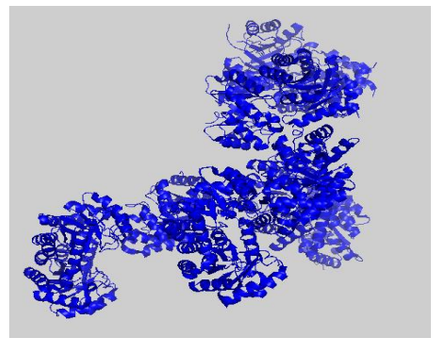


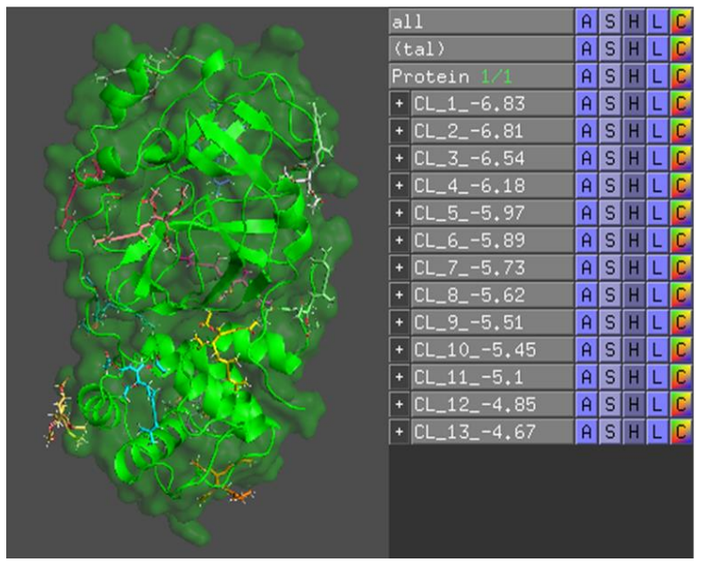




Blind Docking
(https://bio-hpc.ucam.edu/achilles/)

ACHILLES
BLIND DOCKING SERVER





CASE STUDIES

DOCKING RESULTS OBTAINED BY

Virtual Screening Poster 22
(https://github.com/bio-hpc/metascreener)

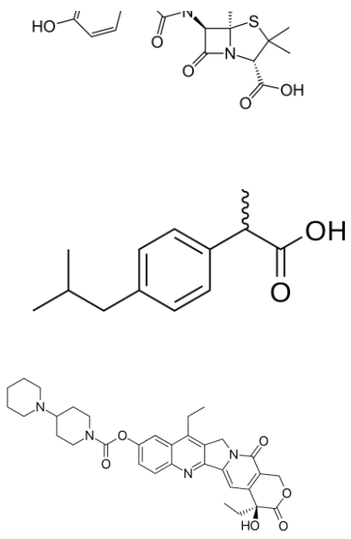
BIO-HPC
Structural Bioinformatics and High Performance Computing Research Group

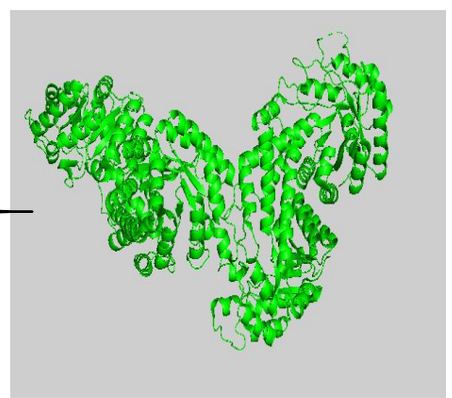
Blind Docking
(https://bio-hpc.ucam.edu/achilles/)

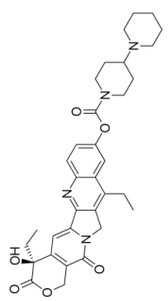
ACHILLES

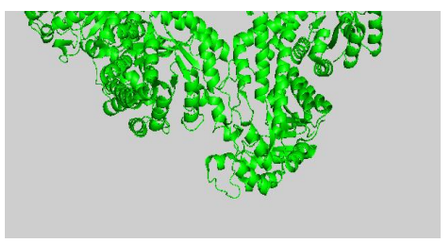
Please if you can run MDs please contact with us:

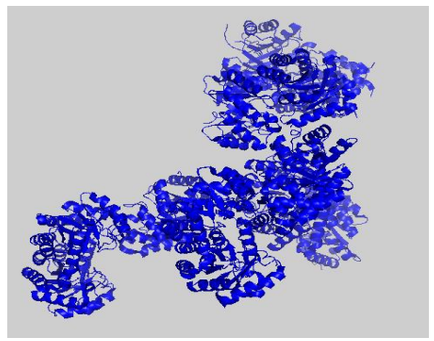
Miguel Carmena-Bargueño: mcarmena@ucam.edu
Horacio Pérez-Sánchez: hperez@ucam.edu

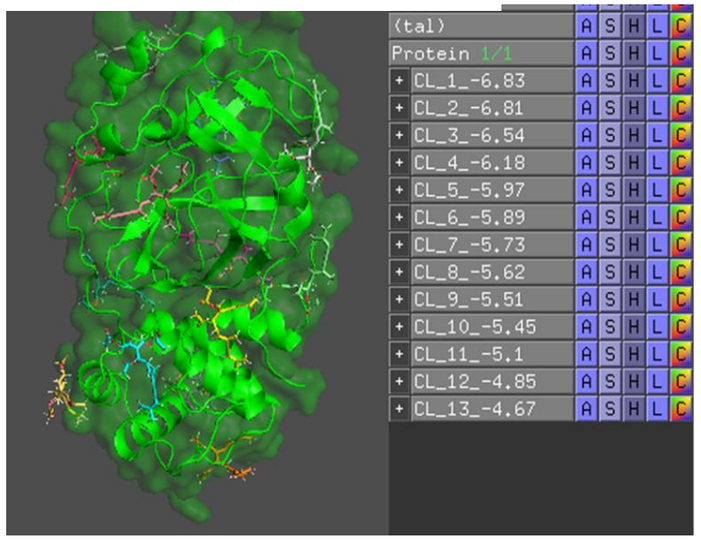












Acknowledgments

UCAM
UNIVERSIDAD CATÓLICA SAN ANTONIO

Barcelona Supercomputing Center
Centro Nacional de Supercomputación