Structural Biology Group
www.bonvinlab.org
z.kurkcuoglusoner@uu.nl

Science in the Clouds: Virtualizing HADDOCK, PowerFit and DisVis using INDIGO-DataCloud Solutions

Zeynep Kurkcuoglu and A.M.J.J. Bonvin

Faculty of Science – Chemistry, Bijvoet Center, Utrecht University (The Netherlands)

Introduction

Reliable bioinformatics and computational methods are needed for complementing experimental techniques, to study structure-dynamics-interactions of biomolecular machines at atomistic level.

Accordingly, we have developed HADDOCK [1], our integrative, information-driven approach for modelling biomolecular complexes. Access to HADDOCK is offered to more than 6500 users worldwide through a user-friendly web interface [2], offered both on local sources and worldwide distributed grid resources within WeNMR Virtual Research Community [3]. PowerFit [4] and DisVis [5] are two other softwares recently developed by our group, for automatic rigid body fitting of biomolecular structures in Cryo-EM densities and for visualization and quantification of the accessible interaction space of distance restrained binary biomolecular complexes, respectively.

Within the European H2020 INDIGO-DataCloud project (INtegrating Distributed data Infrastructures for Global ExplOitation) we aim at virtualizing the HADDOCK portal, together with PowerFit and DisVis. The use of eScience CLOUD solutions will decrease the dependency on local hardware and also facilitate the deployment, configuration and customization of these softwares. This will in turn allow the end users to conduct scientific research without the distraction of technical details at software-hardware level.

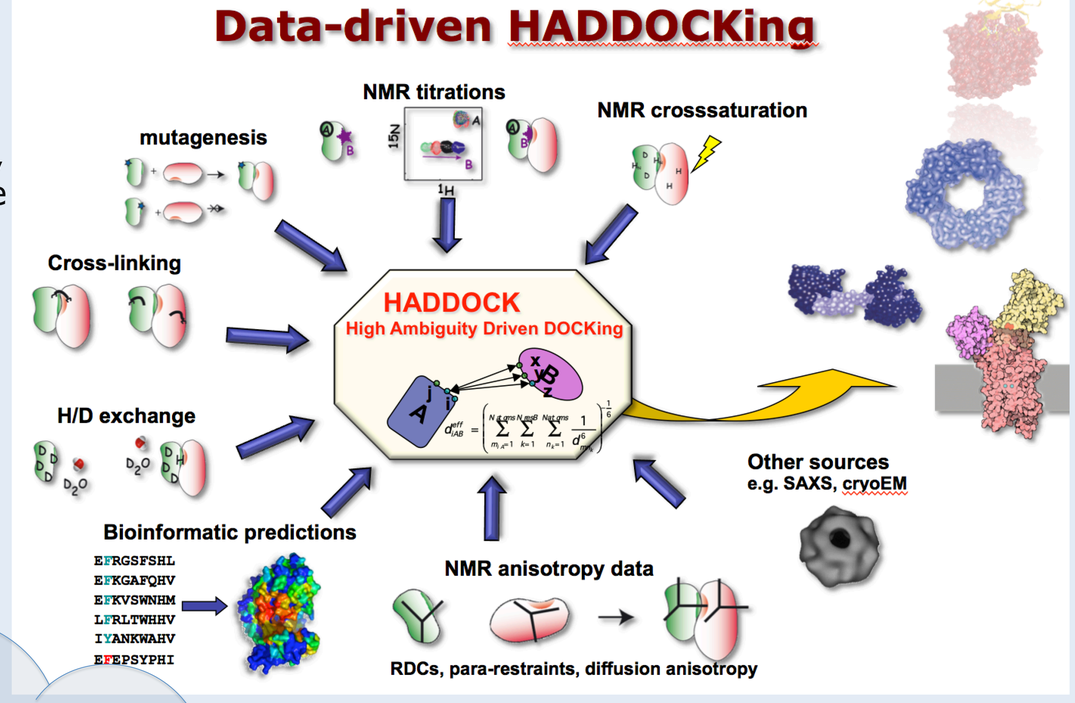
HADDOCK Portal on Cloud

So far…

Local version of HADDOCK, tested on a virtual machine provided by

EC3: Elastic Cloud Computing Cluster

deployed on
EGI Federated Cloud.

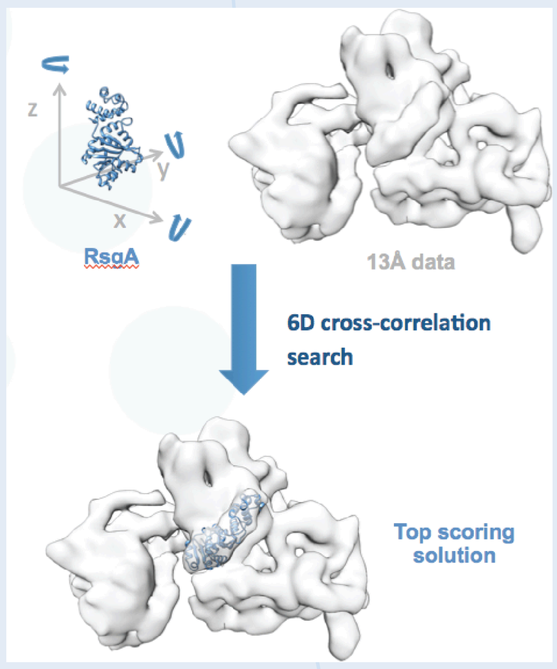


Next steps

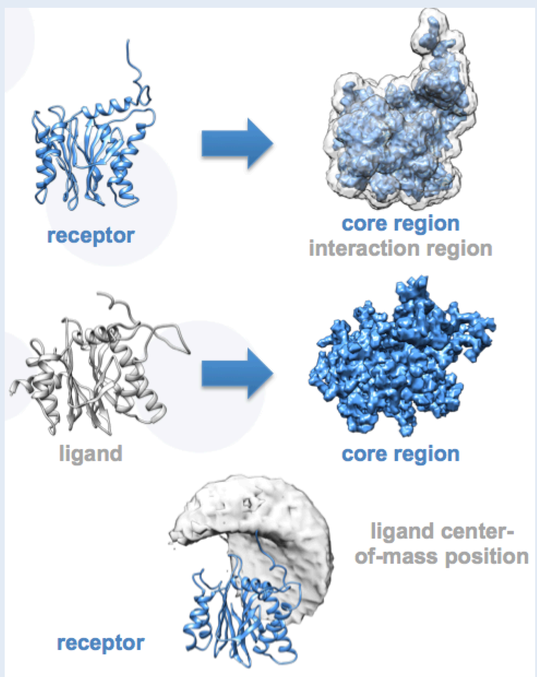
Adapting "Scientific Computational Portal" as a service:

- A VM with a web portal front end
- Exploit of a batch system to run the applications
- Monitor job status
- Temporary storage of simulation data

PowerFit

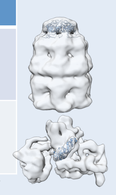


DisVis



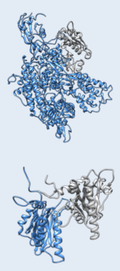
Performance CPU vs GPGPU (in house): PowerFit

System	Map size (voxels)	Rotations sampled	Time CPU	Time GPU	Speedup
GroEL-GroES	90 x 72 x 72	70728	1h 29m	4m 9s	21x
RsgA into ribosome	72 x 80 x 72	70728	1h 16m	4m 2s	19x



Performance CPU vs GPGPU (in house): DisVis

System	Number of complexes sampled	Time CPU	Time GPU	Speedup
RNA-polymerase II	19×10^9	19h 44m	56m	21x
PRE5-PUP2	7×10^9	7h 12m	15m	29x



CPU: AMD Opteron 6344 using FFTW3
GPU: NVIDIA GeForce GTX 680 using clFFT

So far…

- Docker containers are built for both PowerFit and DisVis, with multiple CPUs and GPGPU flavours
- GPGPU version is site specific
- Preliminary test results for GPGPU version show no loss of performance

Next steps

- Web portal for PowerFit and DisVis with docker containers running at the background

References

1. Dominguez C. et al. J.Am.Chem.Soc. 125:1731-37 (2003)
2. van Zundert G.C.P. et al. J.Mol.Biol.online (2016)
3. Wassenaar et al. J.Grid.Comp.10:743-67 (2012)
4. van Zundert G.C.P. and A.M.J.J. Bonvin, AIMS Biophysics, 2:73-87 (2015)
5. van Zundert G.C.P. and A.M.J.J. Bonvin, Bioinformatics, 31:3222-24 (2015)
6. INDIGO DataCloud Deliverables D2.3 and D2.4 (https://www.indigo-datacloud.eu/documents-deliverables)