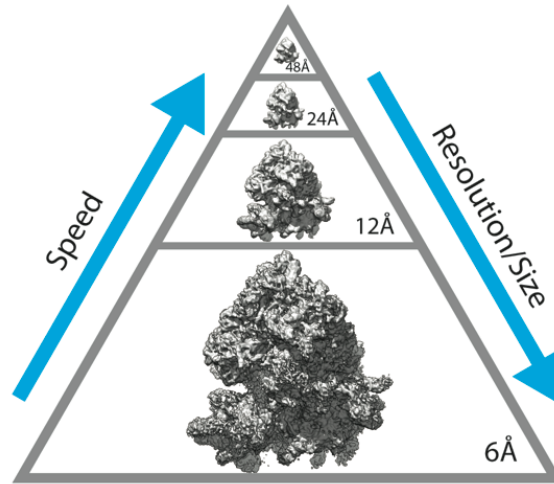




**DisVis and PowerFit:
Explorative and Integrative Modeling
of Biomolecular Complexes
harvesting EGI GPGPU resources**

**Alexandre Bonvin
Utrecht University**

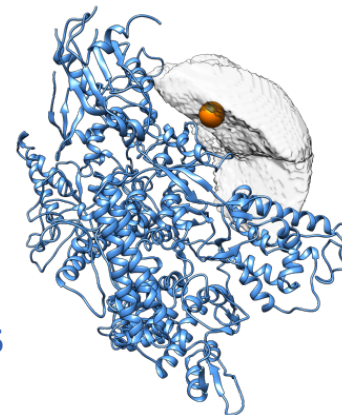


PowerFit: fast and sensitive rigid body fitting in lower-resolution densities

Van Zundert & Bonvin, J. Proteome. Res. 2015

DisVis: quantification and visualization of the information content of distance restraints

Van Zundert & Bonvin, Bioinformatics 2015



Cryo-electron microscopy: The rising star in structural biology

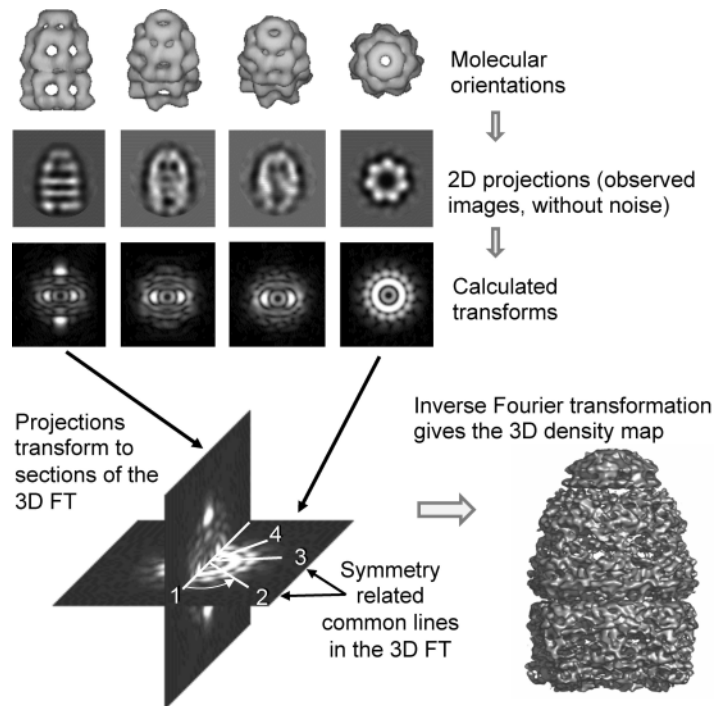
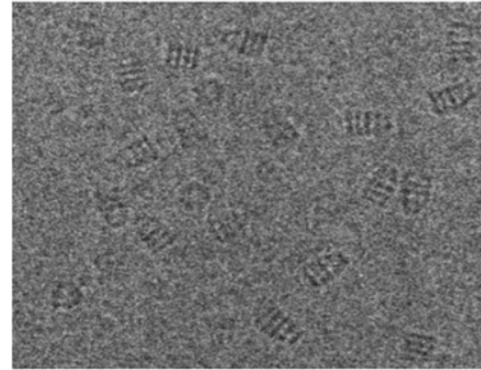
**Transmission electron
microscopy**

**Sample is vitrified in
liquid ethane**

**Images noisy 2D
projections**

**Class averages are
constructed for each
orientation**

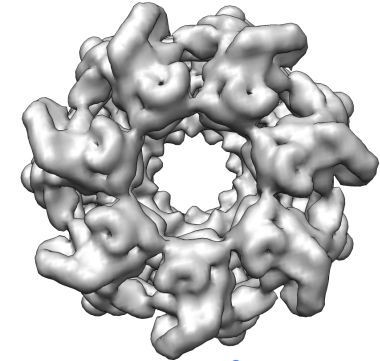
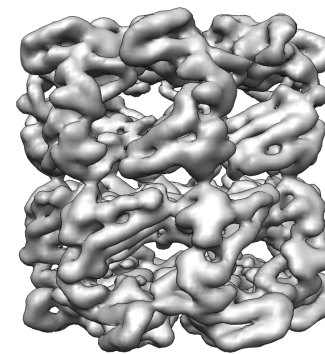
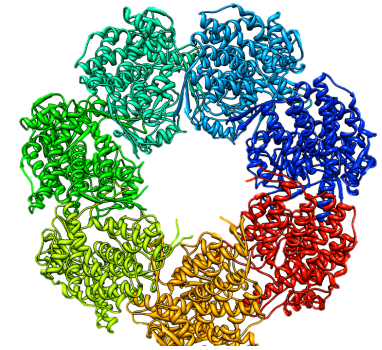
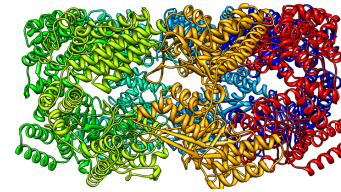
**Ultimately results in
3D density maps**



Cryo-EM data: “high” resolution modelling

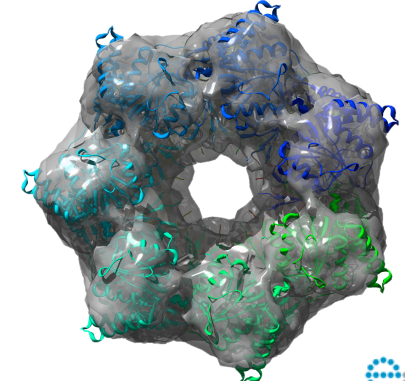
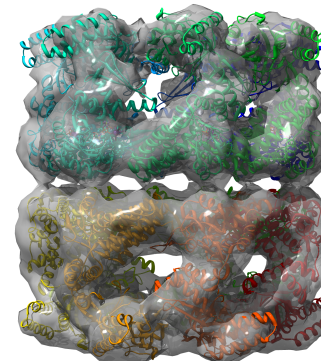
- **Rigid body fitting**

- Manual fitting (UCSF Chimera)
- Automatic fitting software (CoLoRes, PowerFit, Mod-EM)
- In most cases, does not take into account the **flexibility and energetics of the interface**



- **Flexible fitting**

- Requires an **unambiguous fit** of the subunits
- Various approaches (e.g. normal modes, flexible refinement (Flex-EM))
- The applicable resolution extend is debatable
- **Overfitting** is an issue
- Often does not take into account **other sources of data** (mutagenesis, etc.)



PowerFit:

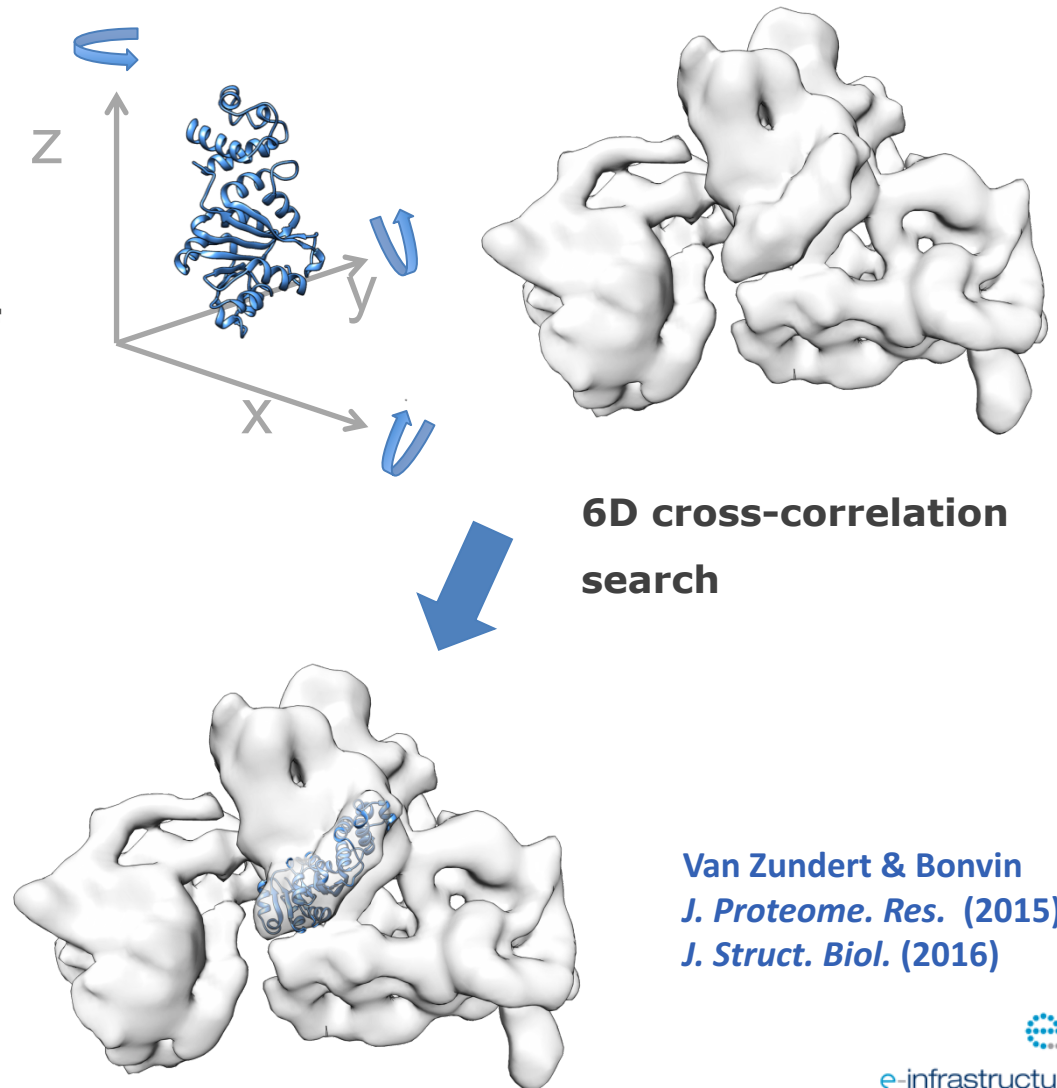
Cross-correlation based rigid body fitting

Automatic rigid body fitting of biomolecular structures in cryo-EM densities

Exhaustive 6D search of 3 translational and 3 rotational degrees of freedom

Calculate cross-correlation at every scanned position

Issues in the field:
Sensitivity
Speed



6D cross-correlation search

Van Zundert & Bonvin
J. Proteome. Res. (2015)
J. Struct. Biol. (2016)

DisVis: Defining the information content and consistency of distance restraints

Given 2 interacting structures and a set of distance restraints between them, are there any solutions that satisfy N restraints?

A solution is a complex that satisfies all N distance restraints

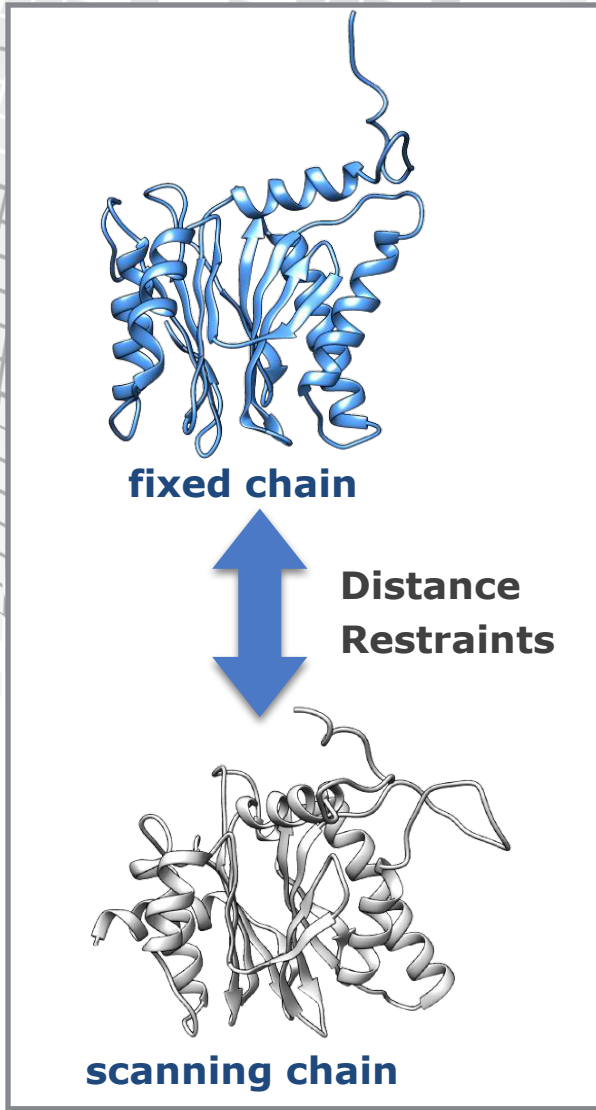
A complex is a conformation where:

The subunits are interacting

The subunits are not clashing

*The **accessible interaction space** is the set of all solutions satisfying at least N restraints*

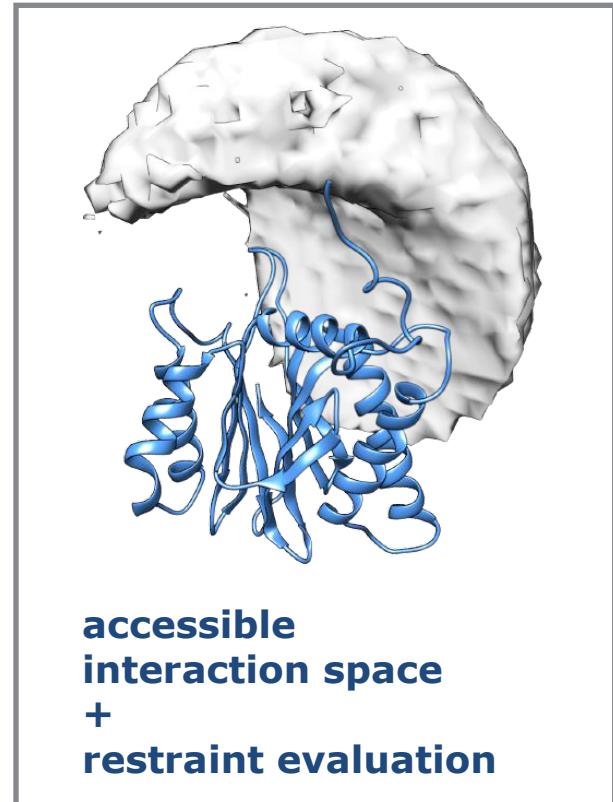
DisVis



6D search of
accessible
interaction
space



At every grid
position, save
the maximum
number of
consistent
restraints found
during the 6D
search



Van Zundert & Bonvin
Bioinformatics 2015

RNA-polymerase II: Accessible interaction space

Number of consistent restraints (N)	Number of accessible complexes consistent with at least N restraints	Fraction of accessible complexes consistent with at least N restraints
0	18940752204	1.0000
1	2370295166	0.1251
2	977410985	0.0516
3	298922038	0.0158
4	92651659	0.0049
5	17687776	0.0009
6	5172437	0.0003
7	9716	0.0000
8	0	0.0000



DisVis 6D systematic search with a 1Å grid size and 5.27° interval

Speeding up the search

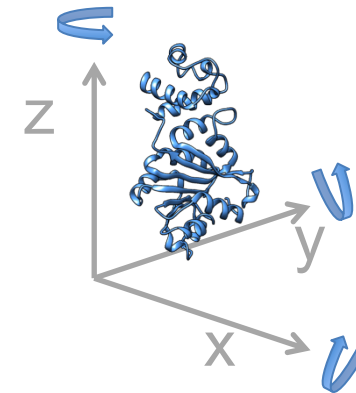
$$\text{CW-LCC} = \frac{1}{N} \frac{\sum_i^N \rho_c^n \cdot w_i \rho_o}{\sqrt{(\overline{\rho_o^w})^2 - (\rho_o^w)^2}}$$

$$\text{CW-GCC} = \mathcal{F}^{-1} [\mathcal{F}(w \rho_c^n)^* \times \mathcal{F}(\rho_o)]$$

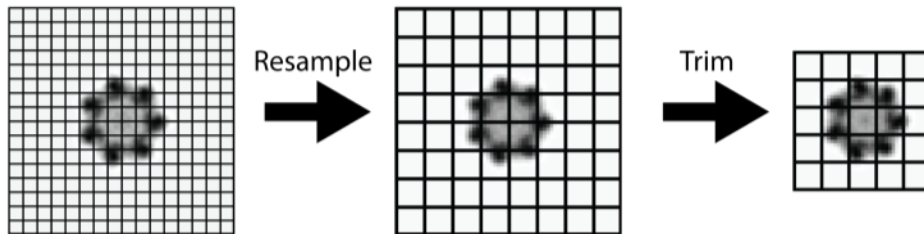
$$(\overline{\rho_o^w})^2 = \mathcal{F}^{-1} [\mathcal{F}(w)^* \times \mathcal{F}(\rho_o)]^2$$

$$(\rho_o^w)^2 = \mathcal{F}^{-1} [\mathcal{F}(w^2)^* \times \mathcal{F}(\rho_o^2)]$$

**Fast Fourier Transform
for fast translational
scans**



Optimized rotation sets



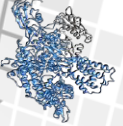
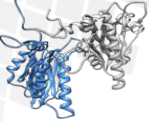
Resampling and trimming target




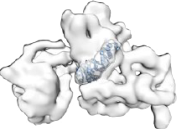
GPU acceleration

GPU-acceleration

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

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

(in house GTX680 card vs single core)

GRID-enabled web portals

Home >> DISVIS >> Index

DISVIS

GRID-enabled web portal @BonvinLab

HADDOCK CPOR **DISVIS** POWERFIT PRODIGY SPOTON 3D-DART BONVIN LAB

About Submit Register Examples Help/Manual Support Forum

WELCOME TO THE GRID-ENABLED DISVIS WEBSERVER! >>

DisVis visualizes the accessible interaction space!

POWERED BY

West-Life
bioexcel
INDIGO - DataCloud
MoBrain
esi

DISVIS allows you to visualize and quantify the information content of distance restraints between macromolecular complexes. It performs a full and systematic 6 dimensional search of the three translational and rotational degrees of freedom to determine the number of complexes consistent with the restraints. In addition, it outputs the percentage of restraints being violated and a density that represents the center-of-mass position of the scanning chain corresponding to the highest number of consistent restraints at every position in space.

DISVIS WEBSERVER

REGISTRATION: To use the DisVis server you must have registered for an account. If you do not have an account yet you can [register here](#)

Submit your job to:

- DISVIS GPU accelerated grid-enabled server
- DISVIS server

Home >> POWERFIT >> Index

POWERFIT

GRID-enabled web portal @BonvinLab

HADDOCK CPOR DISVIS **POWERFIT** PRODIGY SPOTON 3D-DART BONVIN LAB

About Submit Register Examples Help/Manual Support Forum

WELCOME TO THE GRID-ENABLED POWERFIT WEBSERVER! >>

PowerFit fits your 3D structures in any map!

POWERED BY

West-Life
bioexcel
INDIGO - DataCloud
MoBrain
esi

PowerFit automatically fits high-resolution atomic structures into cryo-EM densities. To this end it performs a full-exhaustive 6-dimensional cross-correlation search between the atomic structure and the density. It takes as input an atomic structure in PDB- or mmCIF-format and a cryo-EM density with its resolution; and outputs positions and rotations of the atomic structure corresponding to high correlation values. PowerFit uses the local cross-correlation function as its base score. The score is by default enhanced with an optional Laplace pre-filter and a core-weighted version to minimize overlapping densities from neighboring subunits.

POWERFIT WEBSERVER

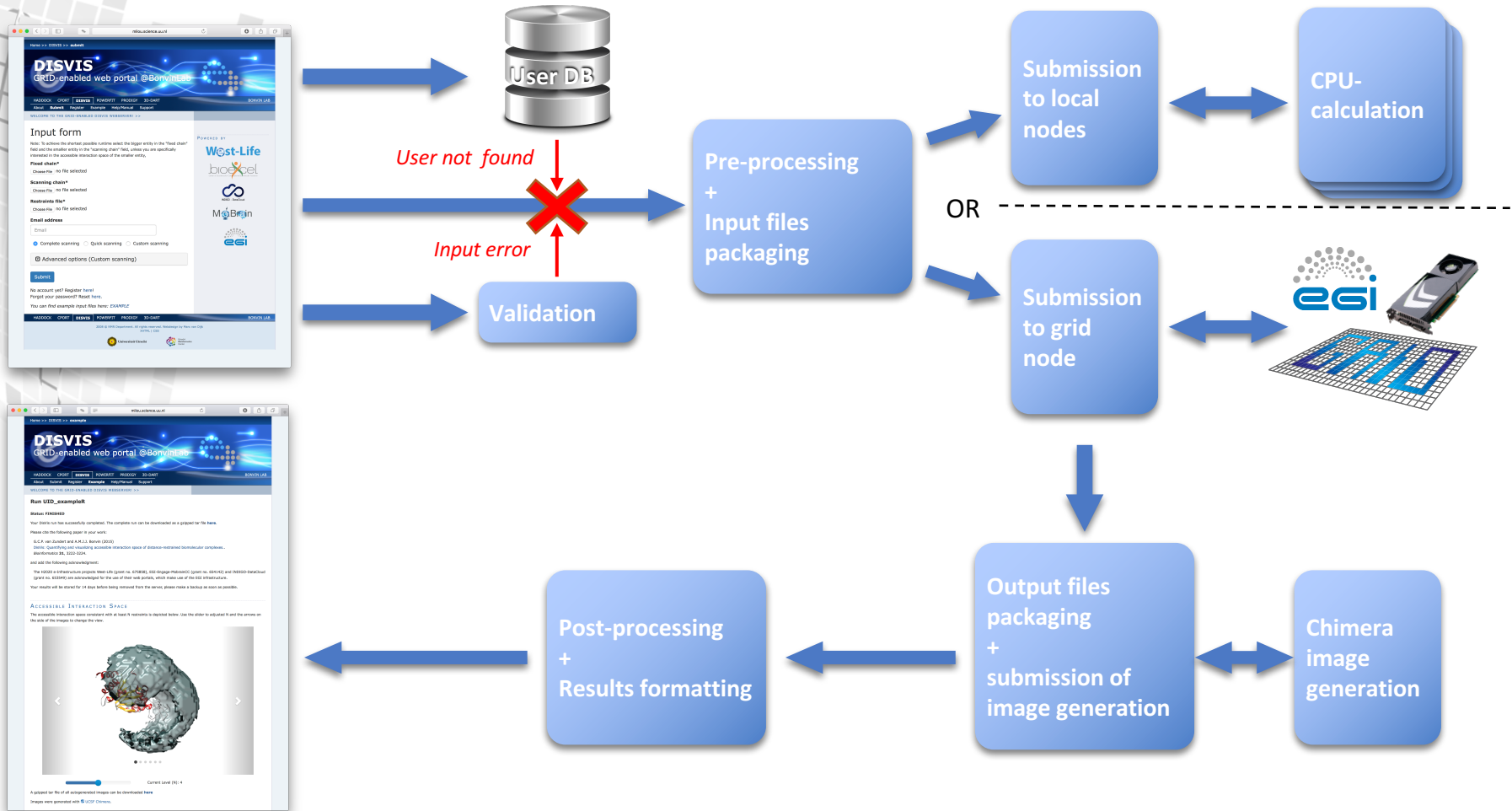
REGISTRATION: To use the PowerFit server you must have registered for an account. If you do not have an account yet you can [register here](#)

Submit your job to:

<http://milou.science.uu.nl/enmr/services/DISVIS/>

<http://milou.science.uu.nl/enmr/services/POWERFIT/>

Architecture behind the portals



WEB CLIENT

WEB SERVER

MASTER NODE

WORKING NODE

Software Provisioning

Because of complex software dependencies we use docker containers

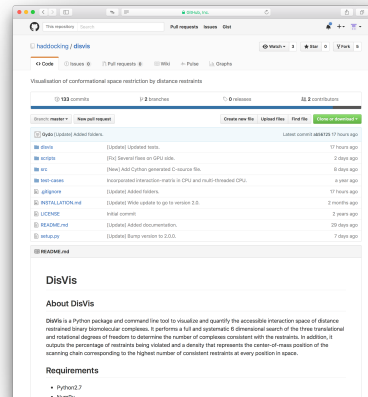
- Python2.7
- NumPy 1.8+
- SciPy
- FFTW3
- pyFFTW 0.10+
- OpenCL1.1+
- pyopencil
- cIFFT
- gpyfft



GitHub

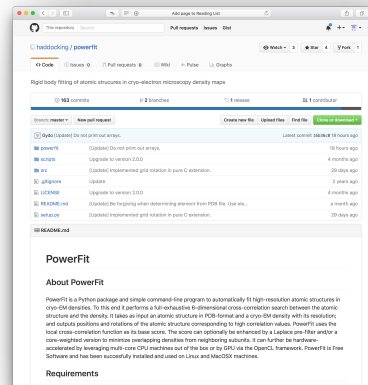


INDIGO - DataCloud



docker

indigodatacloudapps/disvis



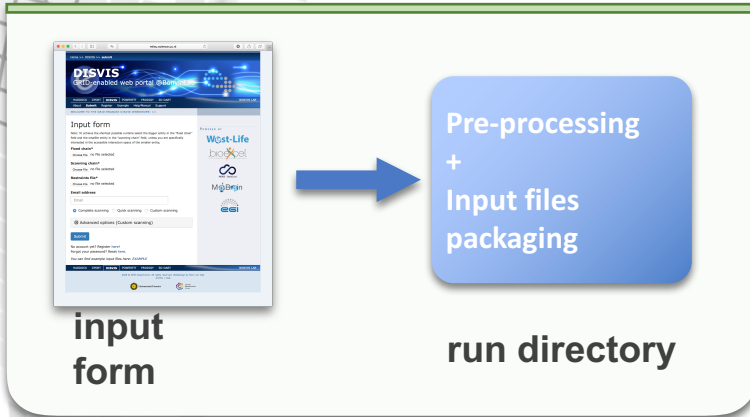
docker

indigodatacloudapps/powerfit

And to avoid security issues on the grid side, **udocker** from INDIGO

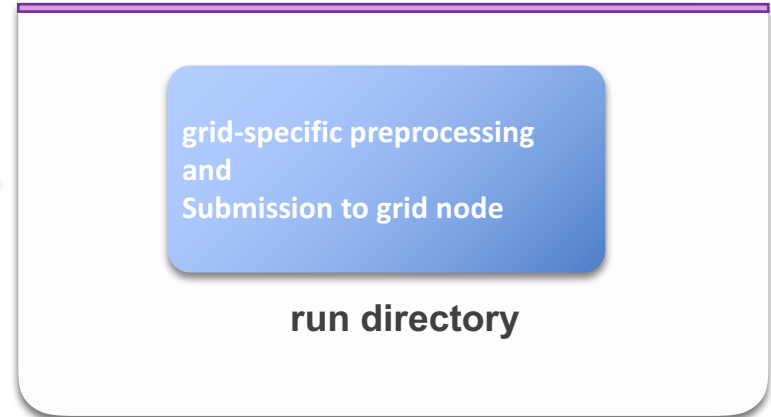
Grid submission

milou.science.uu.nl



scp
run files

ui-enmr.lsg.bcbr.uu.nl



`glite-ce-job-submit`

`disvis.jdl`
`disvis.sh`
`input files`



cegpu.cerm.unifi.it:8443/cream-pbs-batch

- Submission with robot proxy
- Submission to specific site
(cegpu.cerm.unifi.it)

• Command:

```
glite-ce-job-submit -n -o XXX.id -a -r\  
cegpu.cerm.unifi.it:8443/cream-pbs-batch\  
disvis.jdl
```

JDL-File

```
[
executable = "disvis.sh";
inputSandbox = { "aac_3mizzqnQ/disvis.sh",
                 "aac_3mizzqnQ/fixed_chain.pdb",
                 "aac_3mizzqnQ/scanning_chain.pdb",
                 "aac_3mizzqnQ/restraints.txt" };

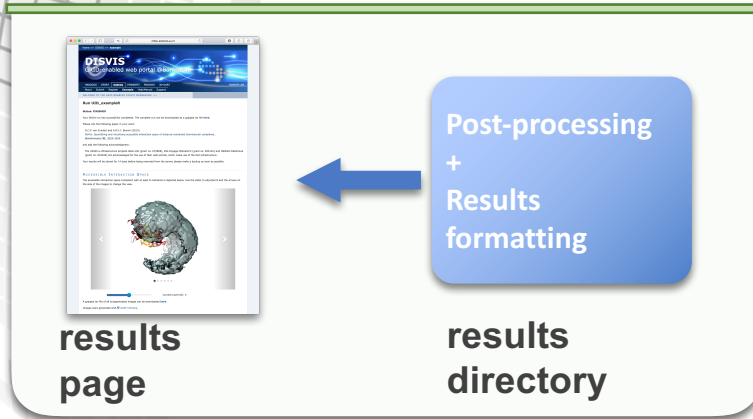
stdout = "disvis.out";
outputsandboxbasedesturi = "gsiftp://localhost";
stderr = "disvis.err";
outputsandbox = { "disvis.out" , "disvis.err" , "aac_3mizzqnQ.tgz" };
GPUNumber=1;
]
```

Executable

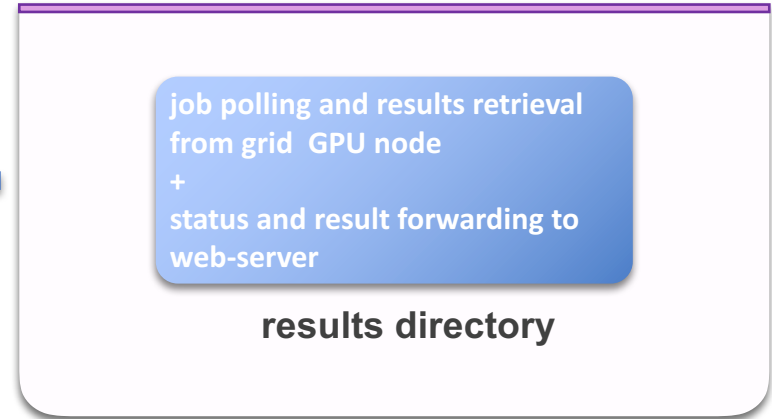
```
#!/bin/sh
driver=$(nvidia-smi | awk '/Driver Version/ {print $6}')
...
# Untar input files archive
tar xzf $jobid.tgz
...
# Get udocker from github
git clone https://github.com/indigo-dc/udocker
cd udocker
# Download docker image from indigodatacloud
./udocker.py pull indigodatacloudapps/disvis:nvdrv_$driver
rnd=$RANDOM
# Create temporary instance of docker image
./udocker.py create --name=disvis-$rnd indigodatacloudapps/disvis:nvdrv_$driver
#
mkdir $WDIR/results_$jobid
# Run disvis within the docker image
./udocker.py run -v /dev --volume=$WDIR:/home disvis-$rnd disvis /home/$pdb_A /home/$pdb_B \
                /home/restraints.txt -g -a $a -vs $vs -ir $ir -cv $cv -iv $iv $is $oa \
                -d /home/results_$jobid
...
# Archive results
tar cfz $jobid.tgz results_$jobid
```


Results retrieval

milou.science.uu.nl



ui-enmr.lsg.bcbr.uu.nl

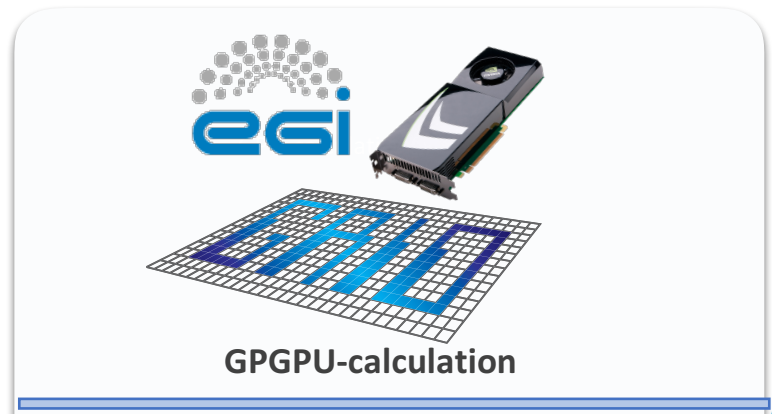


rsync
results files

glite-ce-job-status/
glite-ce-job-output

↑ result and log files

- polling of status and retrieval of results by cron-jobs with respective *glite-ce* command
- Forwarding of results to web server via *rsync*



cegpu.cerm.unifi.it:8443/cream-pbs-batch

Rich result pages

Home >> DISVIS >> example

DISVIS

GRID-enabled web portal @BonvinLab

HADDOCK CPORT **DISVIS** POWERFIT PRODIGY 3D-DART BONVIN LAB

About Submit Register **Example** Help/Manual Support

WELCOME TO THE GRID-ENABLED DISVIS WEBSERVER! >>

Run UID_exampler

Status: FINISHED

Your DisVis run has successfully completed. The complete run can be downloaded as a gzipped tar file [here](#).

Please cite the following paper in your work:

G.C.P. van Zundert and A.M.J.J. Bonvin (2015)
DisVis: Quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes.
Bioinformatics **31**, 3222-3224.

and add the following acknowledgment:

The H2020 e-Infrastructure projects West-Life (grant no. 675858), EGI-Engage-MobrainCC (grant no. 654142) and INDIGO-DataCloud (grant no. 653549) are acknowledged for the use of their web portals, which make use of the EGI infrastructure.

Your results will be stored for 14 days before being removed from the server, please make a backup as soon as possible.

ACCESSIBLE INTERACTION SPACE

The accessible interaction space consistent with at least N restraints is depicted below. Use the slider to adjusted N and the arrows on the side of the images to change the view.

Current Level (N): 4

A gzipped tar file of all autogenerated images can be downloaded [here](#)

Images were generated with [UCSF Chimera](#).

Home >> POWERFIT >> example

POWERFIT

GRID-enabled web portal @BonvinLab

HADDOCK CPORT DISVIS **POWERFIT** PRODIGY 3D-DART BONVIN LAB

About Submit Register **Example** Help/Manual Support

WELCOME TO THE GRID-ENABLED POWERFIT WEBSERVER! >>

Run UID_exampler

Status: FINISHED

Your PowerFit run has successfully completed. The complete run can be downloaded as a gzipped tar file [here](#).

Please cite the following paper in your work:

G.C.P. van Zundert and A.M.J.J. Bonvin (2015)
PowerFit: Quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes.
Bioinformatics **31**, 3222-3224.

and add the following acknowledgment:

The H2020 e-Infrastructure projects West-Life (grant no. 675858), EGI-Engage-MobrainCC (grant no. 654142) and INDIGO-DataCloud (grant no. 653549) are acknowledged for the use of their web portals, which make use of the EGI infrastructure.

Your results will be stored for 14 days before being removed from the server, please make a backup as soon as possible.

FIT 1

Rank	1
Cross Correlation Score	0.759
Fisher z-score	0.993
Number of standard deviations	39.239
PDB	Download

Current Level (N): 4

Guided interpretation of results

Z-SCORE

The table below features the z-Score for each restraint. The higher the score, the more likely the restraint is a false-positive. Z-scores above 1.0 are explicitly mentioned in the output of DisVis.

#	Restraint	Average violated fraction	Standard deviation	Z-score
8	A1092(CB)-E152(CB)	1.00	0.01	2.05
7	A180(CB)-E122(CB)	0.80	0.33	1.29
4	A15(CB)-E171(CB)	0.39	0.30	-0.29
5	A934(CB)-E201(CB)	0.38	0.29	-0.35
6	A938(CB)-E201(CB)	0.3		
2	A129(CB)-E161(CB)	0.2		
1	A1003(CB)-E166(CB)	0.2		
3	A129(CB)-E171(CB)	0.2		

SOLUTIONS

The table below lists the 15 best non-redundant solutions found by correlation score. The first column shows the rank, column 2 the correlation score, column 3 the Fisher z-score column 4 the zscore as factor of standard deviations (z/σ), and column 5 the sigma difference to the best fit. (see N. Volkmann 2009, and Van Zundert and Bonvin 2016).

Rank (N)	Cross Correlation Score	Fisher z-score	z-score/ σ	Sigma difference (z_1-z_N)/ σ
1	0.560	0.633	31.0	0.00
2	0.517	0.573	28.0	2.98
3	0.432	0.463	22.7	8.35
4	0.301	0.311	15.2	15.79
5	0.298	0.307	15.1	15.97
6	0.298	0.307	15.0	15.99
7	0.293	0.302	14.8	16.24
8	0.290	0.298	14.6	16.42
9	0.288	0.297	14.5	16.49
10	0.288	0.296	14.5	16.53

Some usage stats

Operational since Aug. 2016

DISVIS WEBSERVER STATS

Number of users: 70

Total number of served requests as of August 1st 2016: 287 of which 113 on GPGPU-enabled EGI grid sites

POWERFIT WEBSERVER STATS

Number of users: 59

Total number of served requests as of August 1st 2016: 103 of which 53 on GPGPU-enabled grid sites

Published Dec. 2016



The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes

G.C.P. van Zundert¹, M. Trellet¹, J. Schaarschmidt¹, Z. Kurkcuoglu¹, M. David²,
M. Verlato³, A. Rosato⁴ and A.M.J.J. Bonvin¹


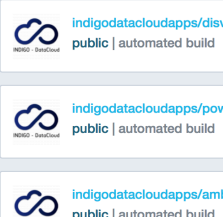
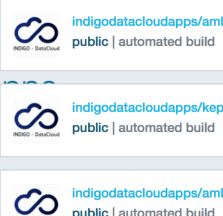
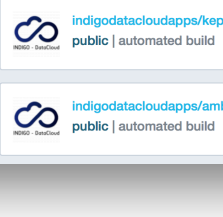

¹ - Bijvoet Center for Biomolecular Research, Faculty of Science – Chemistry, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

² - LIP - Laboratório de Instrumentação e Física Experimental de Partículas, Avenida Elias Garcia 14, 1000 Lisbon, Portugal

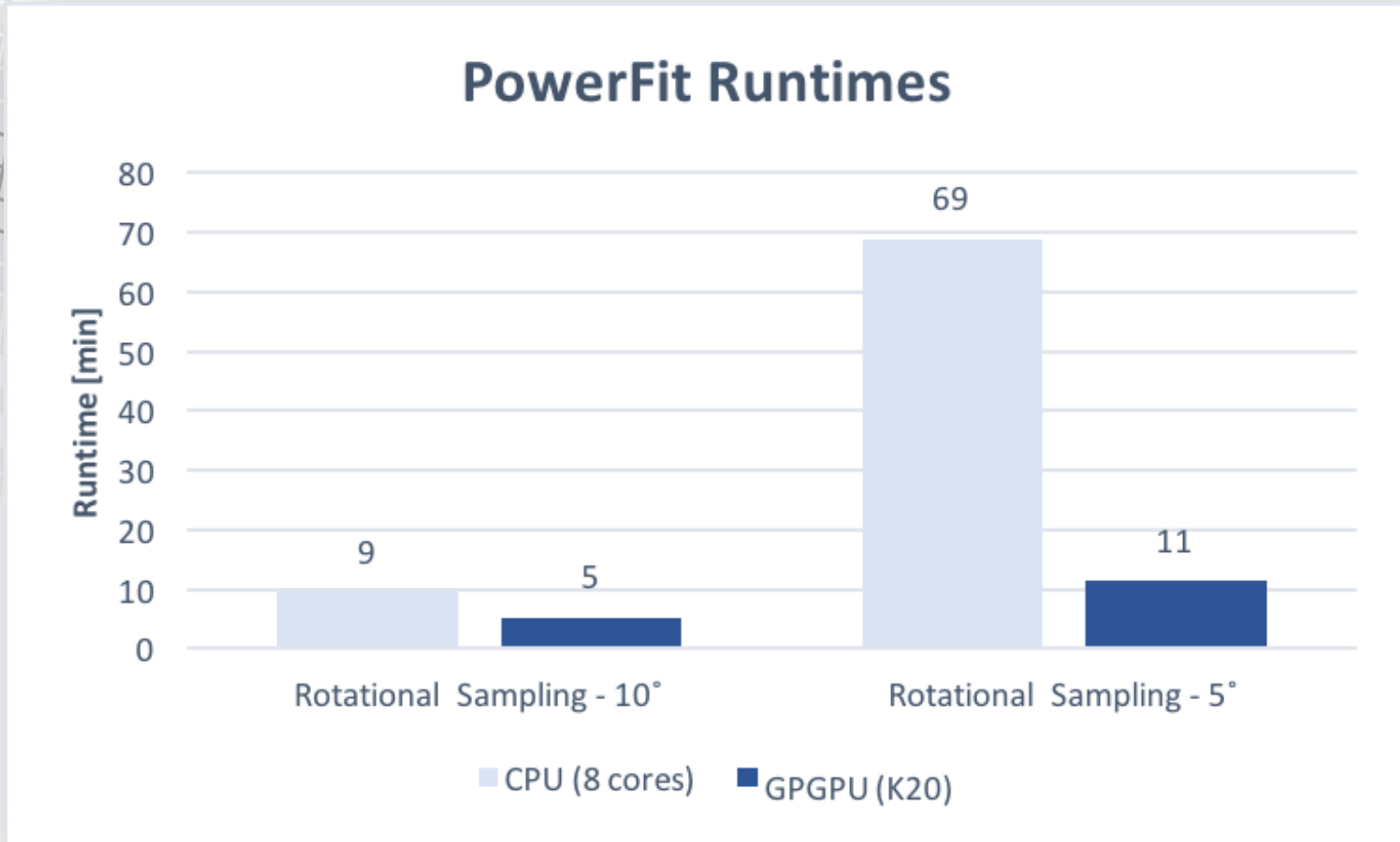
³ - Istituto Nazionale di Fisica Nucleare - Sezione di Padova, Via Marzolo 8, 35131 Padova, Italy

⁴ - Magnetic Resonance Center and Department of Chemistry "Ugo Schiff", University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

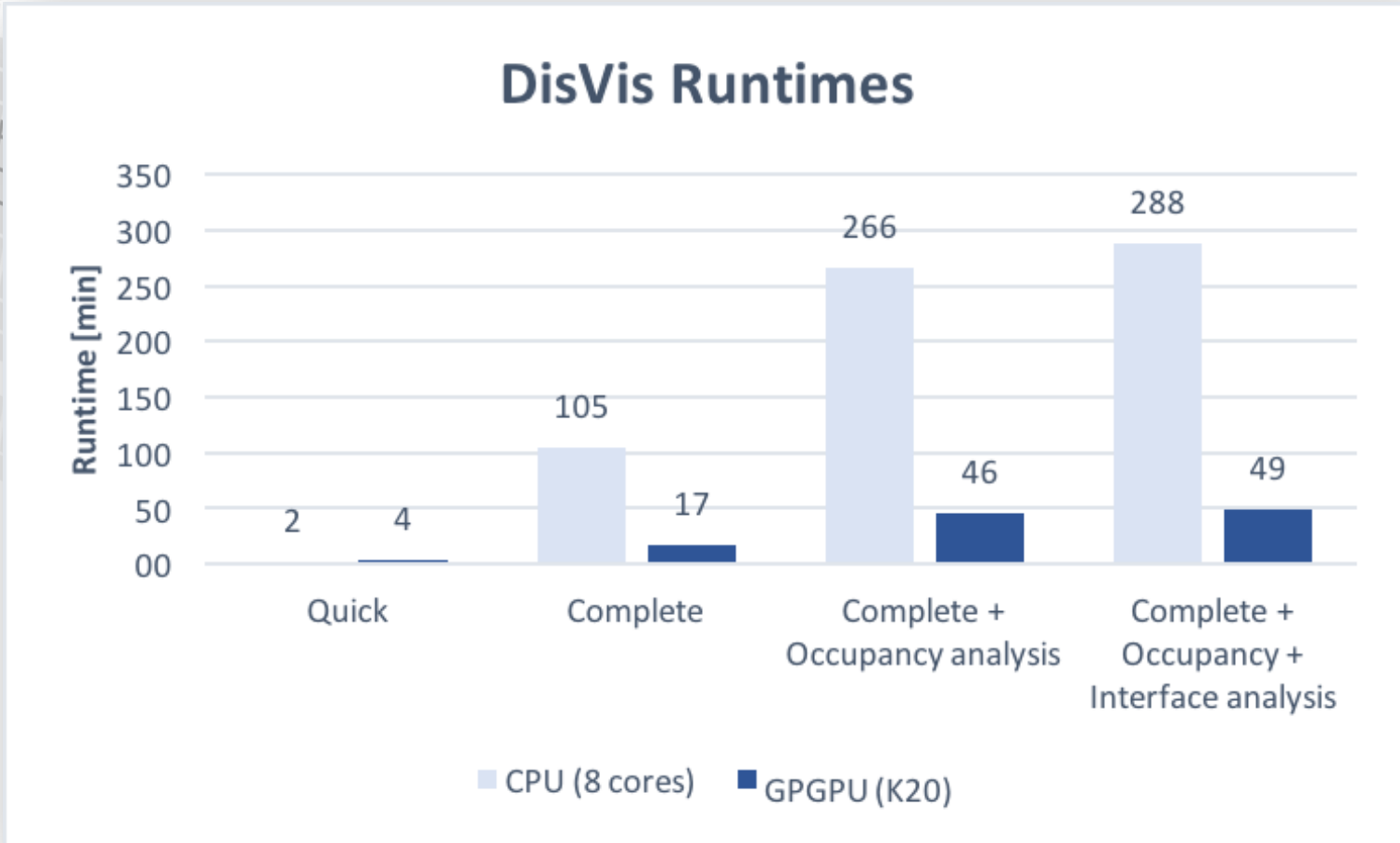
Top pulls in INDIGO docker hub

Repo	Stars	Pulls	Details
 indigodatacloudapps/disvis public automated build	1	1.0K	DETAILS
 indigodatacloudapps/powerfit public automated build	0	462	DETAILS
 indigodatacloudapps/ambertools-oneclient public automated build	0	207	DETAILS
 indigodatacloudapps/kepler public automated build	0	102	DETAILS
 indigodatacloudapps/ambertools public automated build	0	76	DETAILS

Some timing stats



Some timing stats



Listed in EGI applications

The screenshot shows the EGI website interface. At the top, there is a navigation bar with a hamburger menu icon on the left, the EGI logo in the center, and a search icon on the right. Below the navigation bar is a blue banner with the text "EGI / USE CASES / SCIENTIFIC APPLICATIONS AND TOOLS". The main content area features the heading "Scientific applications and tools" and a sub-heading "See also". A paragraph states: "The scientific applications and tools presented here are provided by members and partners of the EGI Community." To the right of this paragraph is a blue button labeled "EGI SERVICE CATALOGUE". Below this is a grey bar with the text "For Life Sciences". Underneath, there are four columns, each representing a different tool: HADDOCK, PowerFit, DisVis, and Virtual Imaging Platform. Each column includes a brief description of the tool's function.

EGI / USE CASES / SCIENTIFIC APPLICATIONS AND TOOLS

Scientific applications and tools

See also

The scientific applications and tools presented here are provided by members and partners of the EGI Community.

[EGI SERVICE CATALOGUE](#)

For Life Sciences

<p>HADDOCK</p> <p>Computational tools to model complexes of proteins and other biomolecules</p>	<p>PowerFit</p> <p>A tool for rigid body fitting of atomic structures into cryo-EM density maps</p>	<p>DisVis</p> <p>Visualise and quantify the accessible interaction space in macromolecular complexes</p>	<p>Virtual Imaging Platform</p> <p>Web portal for medical simulation and image data analysis</p>
--	--	---	---

Conclusions

- **Two new, user-friendly web portals in operation**
- **Harvest grid GPGPU resources**
- **Complex software provisioning via Docker**
- **Build on INDIGO udocker solution**
- **Current issues**
 - Limited resources (only one site in production)
 - Automatic updates of nvidia drivers on sites must be coupled to automatic update of docker containers
 - Speed-up depends on graphic card (might thus require code optimization for new graphic cards).

Acknowledgements



CSB group @UU botanical garden, May 2016

Gydo van Zundert, Jörg Schaarschmidt, Mikael Trellet

Antonio Rosato (CERM), Marco Verlato (INFN), Mario David (LIP)



WeNMR
West-Life
EGI-Engage
INDIGO-
Datacloud
BioExcel CoE



**Thank you for your
attention!**