



## gUSE in CancerGrid project

József Kovács  
[smith@sztaki.hu](mailto:smith@sztaki.hu)  
MTA SZTAKI

Joint EGEE-EDGeS Summer School, 03/07/2009,  
Budapest, Hungary

# The CancerGrid infrastructure



- The general framework components
  - WS-PGRADE/gUSE portal
  - SZTAKI Private/Local Desktop Grid
  - 3GBridge
- Customisation
  - algorithms
  - workflows
  - molecule database
  - resources
- Scenario
  - Handling molecules in the database
  - Workflow configuration/submission & monitoring

# The CancerGrid project

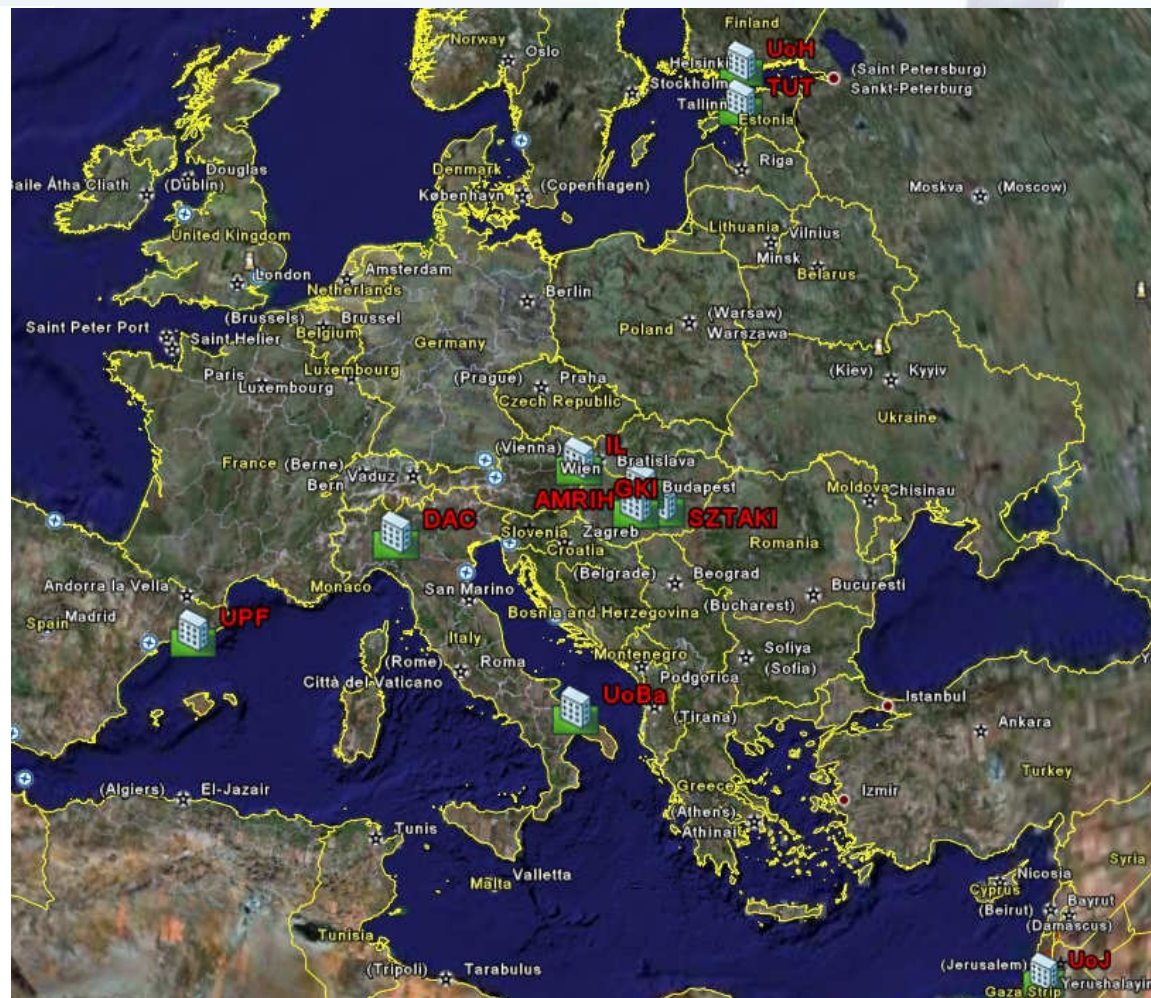


- EU Framework Program 6
- Title: Grid Aided Computer System For Rapid Anti-Cancer Drug Design
- Project period
  - January 1, 2007 – December 31, 2009 (June 30, 2010)
- Goals:
  - Developing *focused libraries* with a high content of anti-cancer leads, building *models* for predicting various molecule properties
  - Developing a *computer system* based on grid technology, which helps to accelerate and automate the *in silico design* of libraries for drug discovery processes



# The CancerGrid consortium

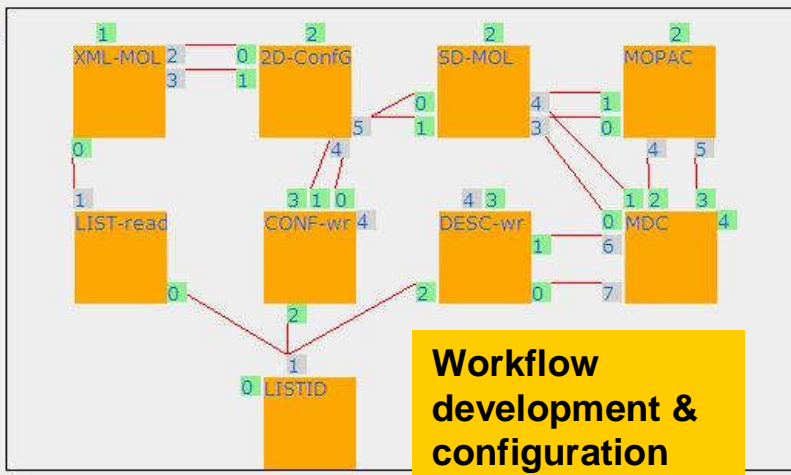
- TargetEx, Inc., Hungary (coordinator) [chemistry]
- AMRI Hungary, Inc., Hungary [chemistry]
- Inte:Ligand, Austria [chemoinformatics]
- Tallinn University of Technology, Estonia [chemoinformatics]
- University of Helsinki, Finland [biotechnology]
- GKI Economic Research, Hungary [economics]
- SZTAKI, Hungary [computer science]
- University of Jerusalem, Israel [chemoinformatics]
- DAC, Italy [biotechnology]
- University of Bari, Italy [chemistry]
- University of Pompeu Fabra, Spain [chemoinformatics]



# The CancerGrid portal (gUSE & SZTAKI DG)



<b>Workflow name:</b>	CG-wf2-DB
<b>Note :</b>	2008-7-9
<b>Workflow Graph :</b>	CG-wf2-DB-graph -- -- Optional selection of a new
<b>Workflow Template :</b>	-- -- Optional selection of a new Template : CG



2008-7-10 14:29 - 2x100[id337] finished [Details](#) [Delete](#) [Visualize](#)

**Selected WF Instance:**  
2008-7-10 14:29 - 2x100[id337]

Job	Status	Instances	[ Actions ]
MDC	<span style="background-color: green; color: white;">finished</span>	41	<a href="#">View content(s)</a>
LISTID	<span style="background-color: green; color: white;">finished</span>	1	<a href="#">View content(s)</a>
2D-ConfGen	<span style="background-color: green; color: white;">finished</span> no input	1	<a href="#">View content(s)</a>
MOPAC	<span style="background-color: green; color: white;">finished</span> no input	41 59	<a href="#">View content(s)</a>
LIST-read	<span style="background-color: green; color: white;">finished</span>	1	<a href="#">View content(s)</a>
SD-MOL	<span style="background-color: green; color: white;">finished</span>	1	
CONF-wr	<span style="background-color: green; color: white;">finished</span>	1	
DESC-wr	<span style="background-color: green; color: white;">finished</span>	41	
XML-MOL	<span style="background-color: green; color: white;">finished</span>	1	

**Workflow execution**

42 items 1 / 1

admin back insert filter

DELETE	INSERT AS NEW	1 RECORD MODE	SELECT MORE	REPORTS	MENU	MOLECULE	LISTITEM ID	LIST ID	MOLECULE ID	CANCERGRID CODE
						ASC DESC	ASC DESC	ASC DESC	ASC DESC	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186691	337	266720	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186690	337	266719	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186689	337	266718	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186688	337	266717	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186687	337	266716	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186686	337	266715	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186685	337	266714	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186684	337	266713	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186683	337	266712	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186682	337	266711	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186681	337	266710	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186680	337	266709	
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	1186679	337	266708	

**Molecule database browser**

view property window

**Cmol3D (conformational search)**

Conformational search:

Using 2D molecules:

Population size:

Number of population initializers:

Energy window for saving structures (kcal/mol):

Maximum number of structures to save:

Maximum distance between atoms in equal structures:

Maximum number of perturbations for a single conformation:

Step size for low-mode move:

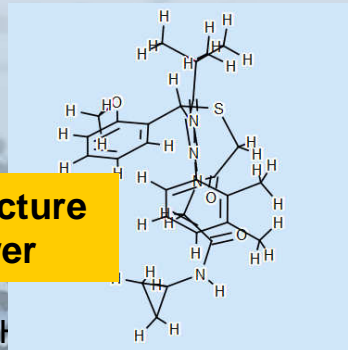
Maximum distance for low-mode move:

Number of lowest eigenvectors to use:

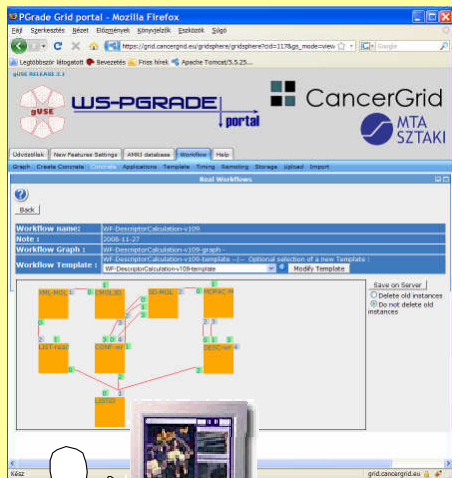
**Algorithms configuration**

**Integrated components of CancerGrid portal**

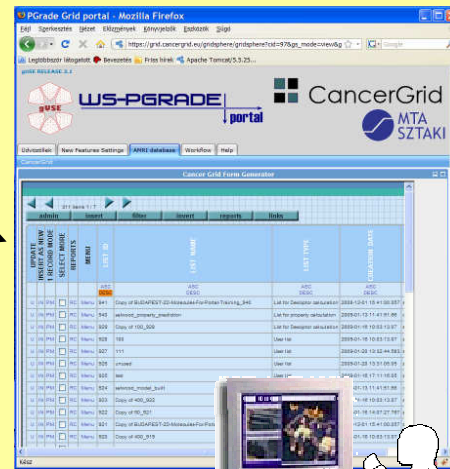
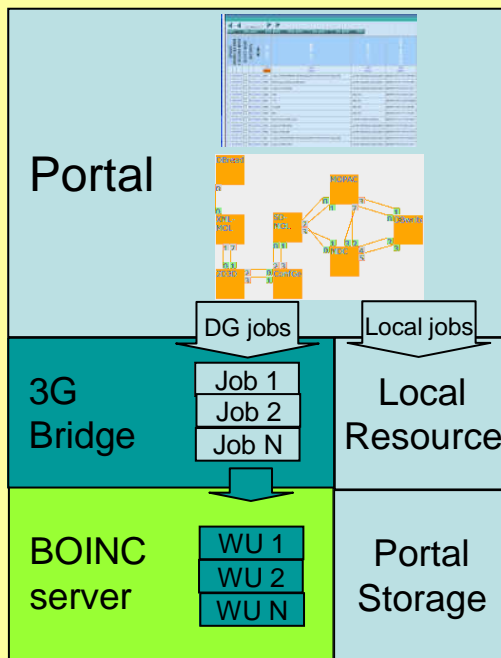
**Structure viewer**



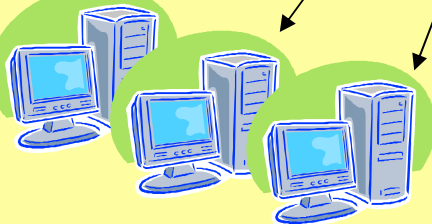
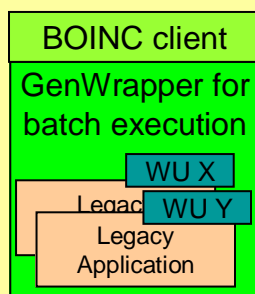




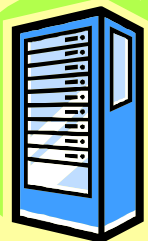
executing workflows



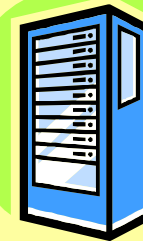
browsing molecules



DG clients from all partners



Portal and DesktopGrid server



molecule database

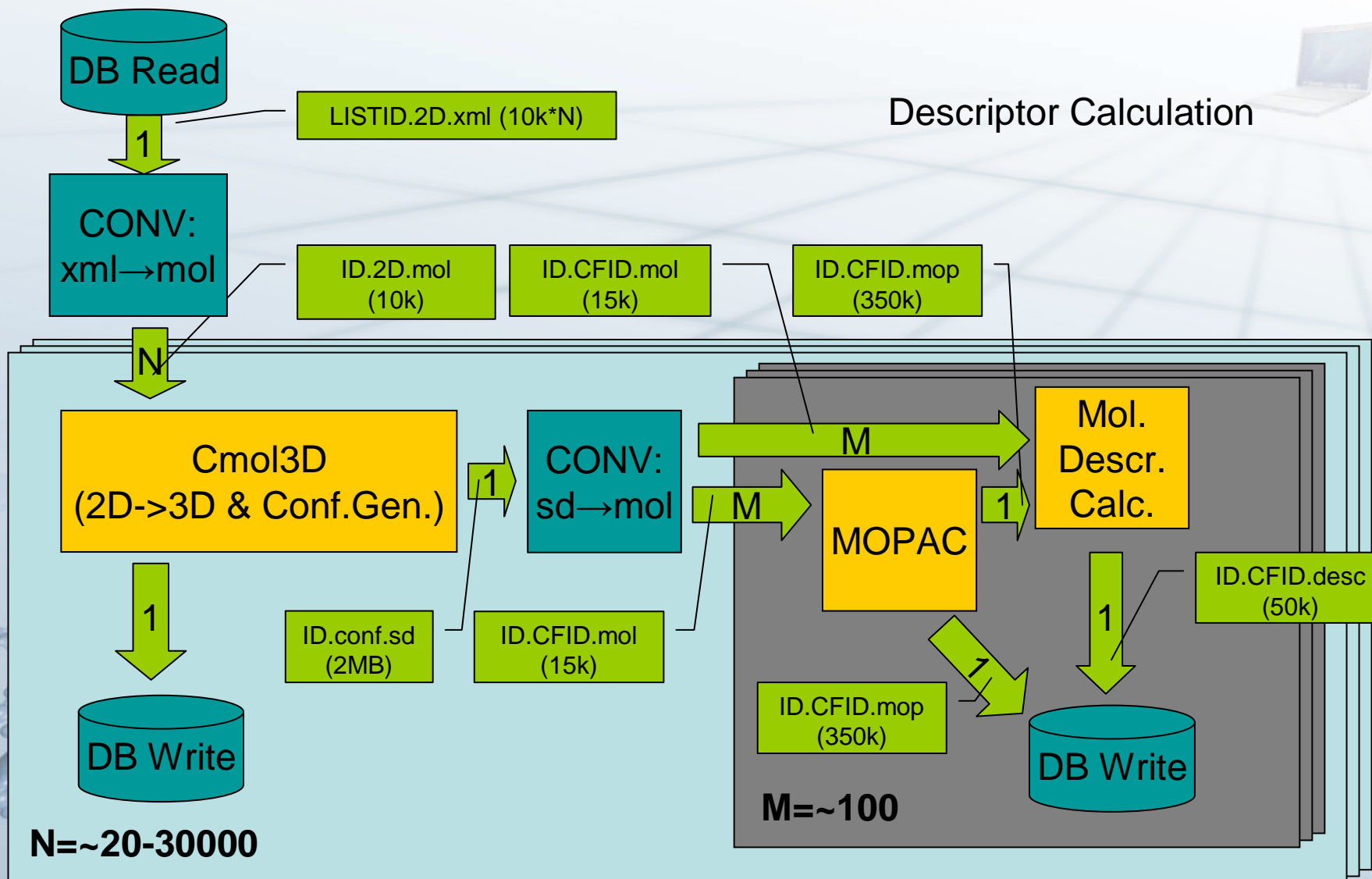
Molecule database server

# Applications in CancerGrid



- Flexmol is an XML-based molecular language
- Molecule 2D/3D converter (Cmol3D)
- Molecule 3D conformation generator (Cmol3D)
- MOPAC (Molecular Orbital PACKage) is a semiempirical quantum chemistry program based on Dewar and Thiel's NDDO approximation
- **Codessa Pro (Comprehensive Descriptors for Structural and Statistical Analysis)** is a software suite for developing quantitative structure-activity/property relationships
- Matrix former
- QSAR Model builder **Quantitative structure-activity relationship (QSAR)** is the process by which chemical structure is quantitatively correlated with a well defined process, such as biological activity or chemical reactivity.
- (Chemical) Property Predictor
- File format converters (to integrate the previous tools into a workflow)

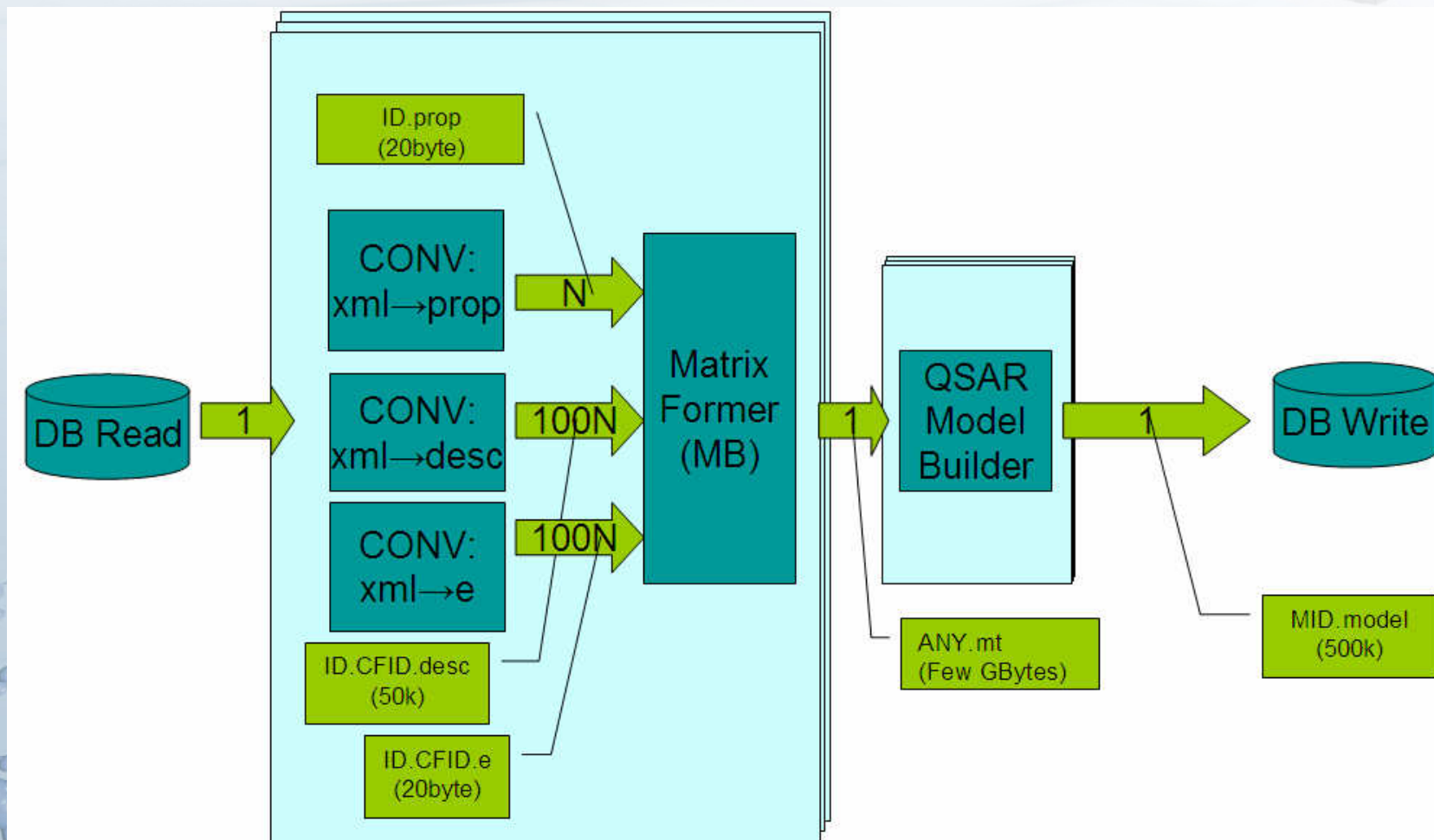
# Supported workflow 1/4





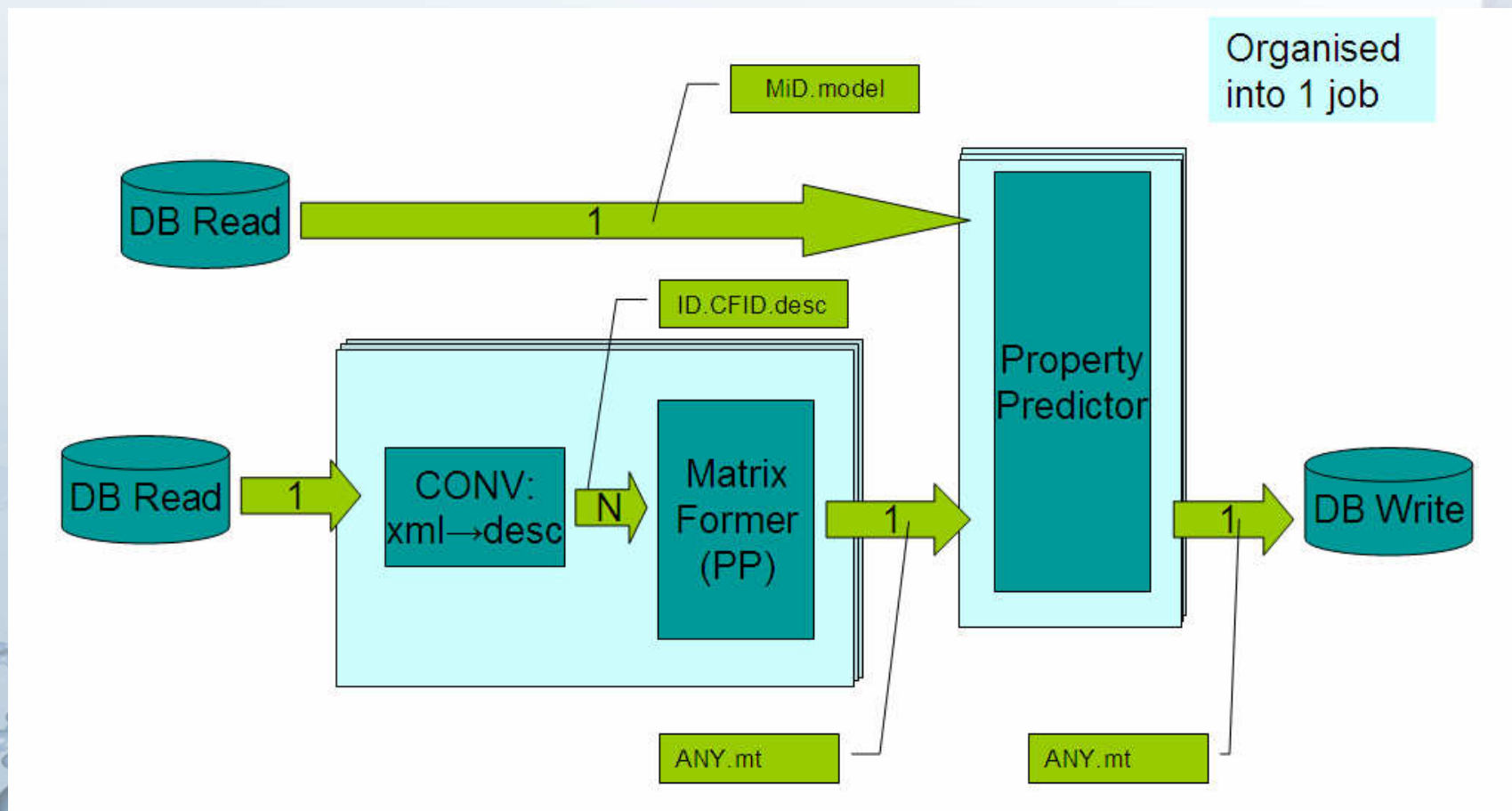
# Supported workflow 2/4

Model building

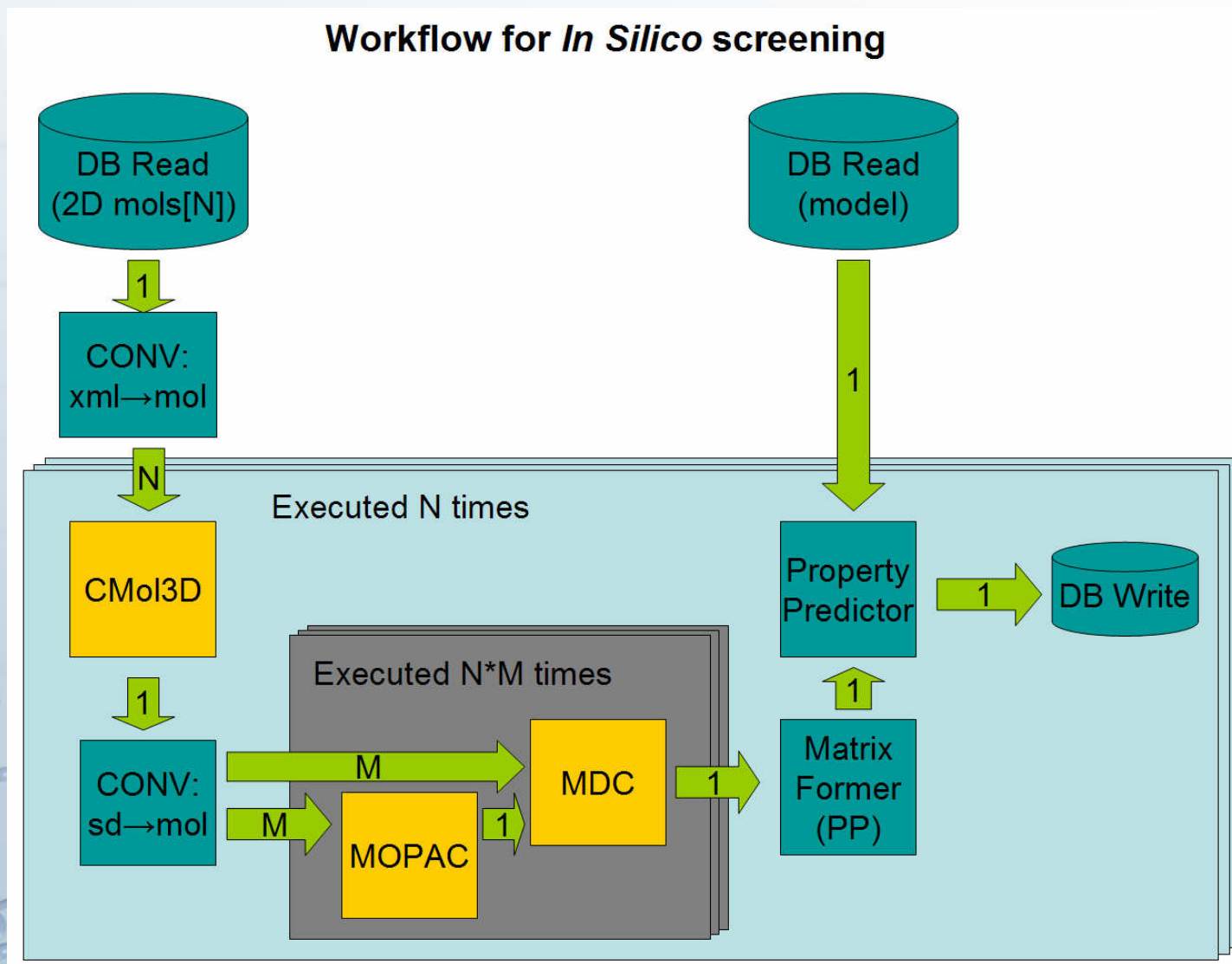


# Supported workflow 3/4

## Property Prediction

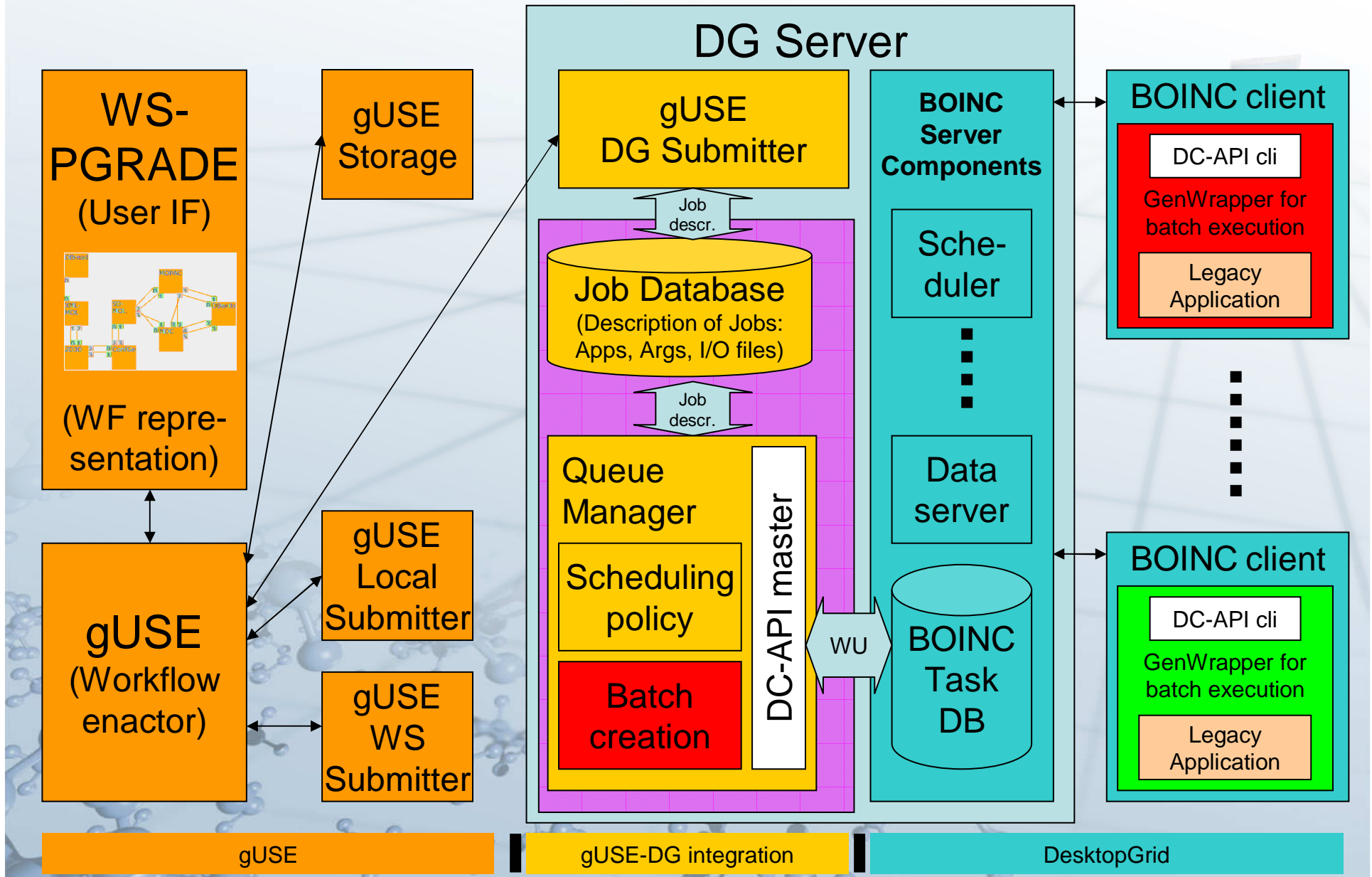


# Supported workflow 4/4





# gUSE & SZTAKI integration through 3GBridge

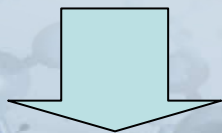


# Job handling

Applications: **cmol3d**, **mopac**, **mdc**, **fmt**, **fma**,  
etc.

## Various requirements

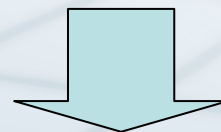
- Fortran, C, C++
- processing/ memory requirements
- multi-binary applications (already contain some wrappers), libraries
- legacy binaries for Linux and Windows
- config file preparation before execution
- pure logging/ debugging information
- variable number of output files



GenWrapper tool  
on the client machine

## Low granularity

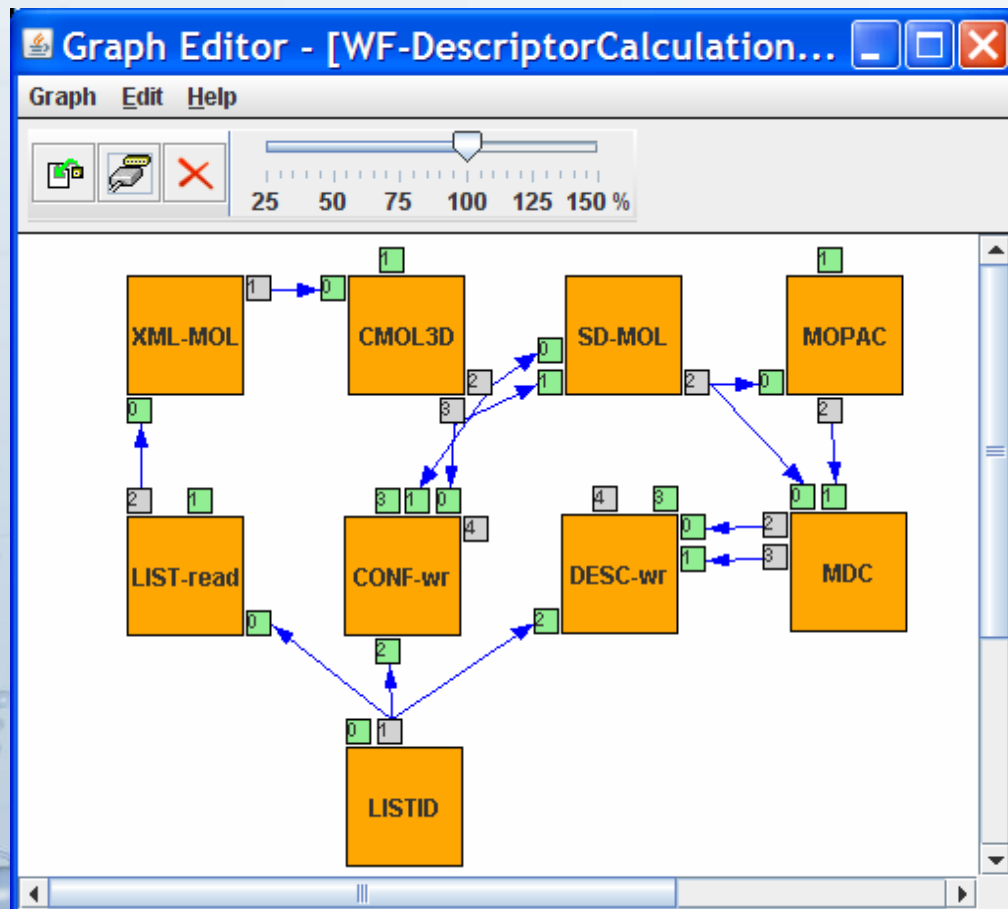
- Execution time is very small (sometimes only a few minutes)



## Batch creation plugin in 3GBridge

- The workunit contains the shell script that is created to manage the execution of the batch
  - assembled from **head**, **body** and **tail** fragments
  - body part is repeated for each job in the batch
  - may contain macros like `%{name}`

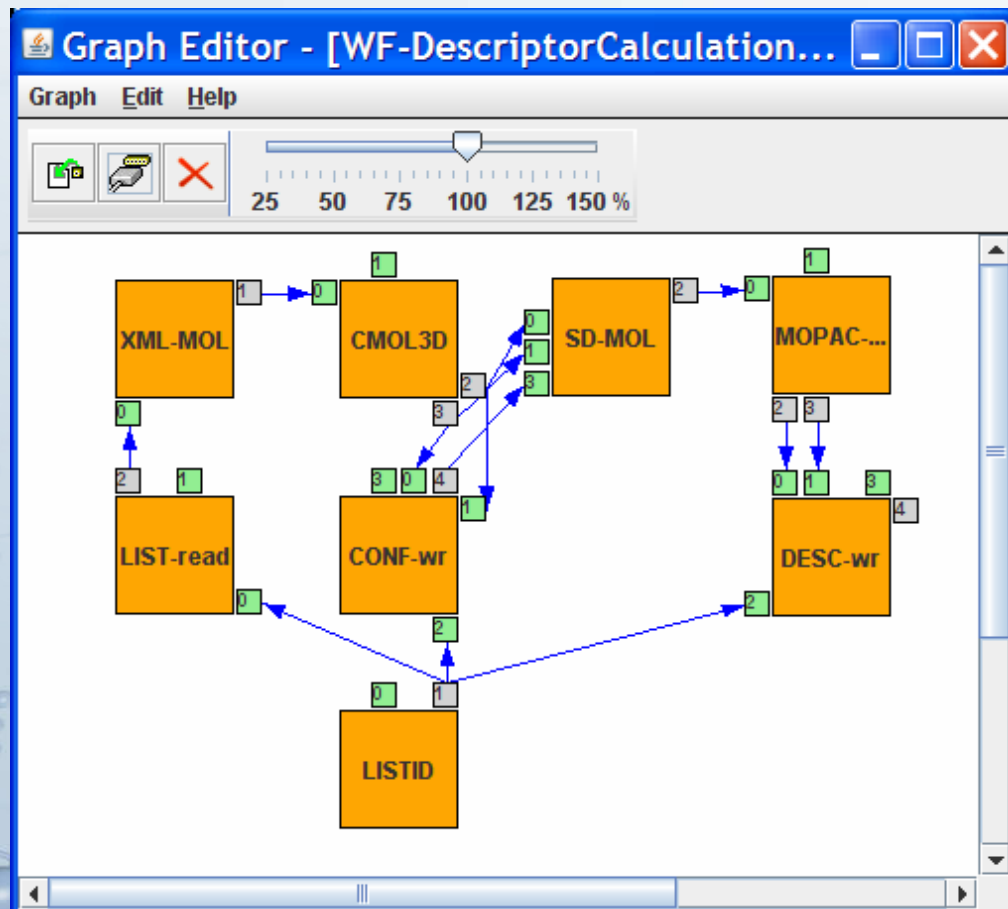
# Descriptor calculation workflow in gUSE



- DG jobs:
  - CMOL3D
  - MOPAC
  - MDC
- WS jobs:
  - LIST-read
  - CONF-wr
  - DESC-wr
- Local jobs:
  - LISTID
  - XML-MOL
  - SD-MOL



# Optimised descriptor calculation workflow in gUSE



- DG jobs:

- CMOL3D

- MOPAC

- MDC

**built into  
one job**

- WS jobs:

- LIST-read

- CONF-wr

- DESC-wr

- Local jobs:

- LISTID

- XML-MOL

- SD-MOL

## High-level scenario

- Initial state: molecules/structures stored in **DB**,  
**organised into lists**
- User selects **list of molecules/structures**
- User selects/downloads a **workflow** from repository
- User **configures** the workflow to take the **list as input**
- User optionally updates **parameters** of the modules
- **Submits** workflow
- Optionally monitors the status
- When workflow finished, results are stored in the **DB**

# Molecule lists

314 items 1 / 7

admin insert filter invert reports links

UPDATE	INSERT AS NEW	RECORD MODE	SELECT MORE	REPORTS	MENU	LIST ID	LIST NAME	LIST TYPE	CREATION DATE	OWNER	NOTES	MODIFICATION DATE	PROPERTY_CODE
U	I	N	P	M	R	C	ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC	A D
U	I	N	P	M	R	C	Menu	1082	Copy of 2-mols-for-quick-functional-test_1081	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1081	Copy of 2-mols-for-quick-functional-test_1080	Screening	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1080	Copy of 2-mols-for-quick-functional-test_1079	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1079	Copy of Test_Descriptor_Calculation_002_5000_1078	List for Descriptor calculation	2009-05-28 13:36:24.16	gpocz	
U	I	N	P	M	R	C	Menu	1078	Copy of Test_Descriptor_Calculation_001_5000_1077	List for Descriptor calculation	2009-05-28 13:36:17.363	gpocz	
U	I	N	P	M	R	C	Menu	1077	Copy of 2-mols-for-quick-functional-test_1076	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1076	2-mols-for-quick-functional-test	User list	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1075	Copy of BUDAPEST-2D-Molecules-For-Portal-Training_1074	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1074	Copy of Copy of BUDAPEST-2D-Molecules-For-Portal-Training_1072_1073	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1073	Copy of BUDAPEST-2D-Molecules-For-Portal-Training_1072	List for Descriptor calculation	2008-12-01 15:41:00.057	smith	
U	I	N	P	M	R	C	Menu	1072	Copy of Test_Descriptor_Calculation_006_5000_1071	List for Descriptor calculation	2009-05-28 13:36:51.33	gpocz	
U	I	N	P	M	R	C	Menu	1071	Copy of Test_Descriptor_Calculation_005_5000_1070	List for Descriptor calculation	2009-05-28 13:36:43.503	gpocz	
U	I	N	P	M	R	C	Menu	1070	Copy of Test_Descriptor_Calculation_004_5000_1069	List for Descriptor calculation	2009-05-28 13:36:36.597	gpocz	
U	I	N	P	M	R	C	Menu	1069	Copy of Test_Descriptor_Calculation_003_5000_1068	List for Descriptor calculation	2009-05-28 13:36:30.317	gpocz	



# Listitems view

53510 items 1 / 1071

admin back insert filter invert reports main menu

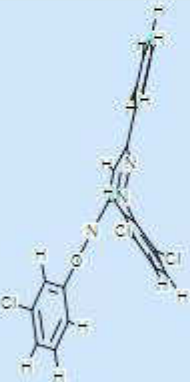
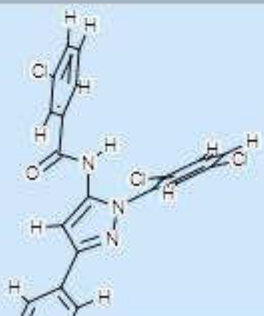
DELETE	INSERT AS NEW	RECORD MODE	SELECT MORE	REPORTS	MENU	MOLECULE	LISTITEM ID	LIST ID	MOLECULE ID	CANCERGRID CODE	PARENT MOLECULE ID
							ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985042	1072	1286246		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985041	1072	1286245		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985040	1072	1286244		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985039	1072	1286243		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985038	1072	1286242		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985037	1072	1286241		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985036	1072	1286240		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985035	1072	1286239		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985034	1072	1286238		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985033	1072	1286237		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985032	1072	1286236		248615
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985031	1072	1286235		248615
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985030	1072	1286234		248615
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985029	1072	1286233		248615
D	IN	PM	<input type="checkbox"/>	RC	Menu	-	5985028	1072	1286232		248615



# Molecule viewer

53510 items 1 / 1071

admin back insert filter invert reports main menu

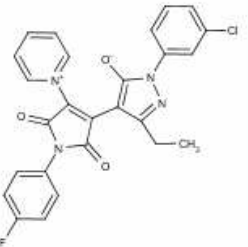
DELETE	INSERT AS NEW	1 RECORD MODE	SELECT MORE	REPORTS	MENU	MOLECULE	LISTITEM ID	LIST ID	MOLECULE ID	CANCERGRID CODE	PARENT MOLECULE ID
							ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC
D	IN	PM	<input type="checkbox"/>	RC	Menu		5985042	1072	1286246		258336
D	IN	PM	<input type="checkbox"/>	RC	Menu		5985041	1072	1286245		258336



53144 items 53143 / 53144

admin back insert filter main menu  
delete insert as new page mode reports menu

**MOLECULE**



PARENT MOLECULE ID

LISTITEM ID	5670985
LIST_ID	1067
MOLECULE ID	67265
CANCERGRID CODE	CGX-03037568


Descriptors

33 items 1 / 3

admin filter invert reports

1 RECORD MODE	SELECT MORE	REPORTS	LISTITEM ID	DESCRIPTORITE M ID	CD ID	DESCRIPTOR ID	DESCRIPTORTY PE NAME	DESCRIPTOR_V ALUE	DESCRIPTOR_TE XT VALUE
			ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC	ASC DESC
PM	<input type="checkbox"/>	RC	5670985	988408	67265	lipinski_score	ADMETox descriptors	0.0	
PM	<input type="checkbox"/>	RC	5670985	988407	67265	HBDC	ADMETox descriptors	0.0	
PM	<input type="checkbox"/>	RC	5670985	988406	67265	HBAC	ADMETox descriptors	7.0	
PM	<input type="checkbox"/>	RC	5670985	988405	67265	rotatable_bonds	ADMETox descriptors	5.0	
PM	<input type="checkbox"/>	RC	5670985	988404	67265	HBD	ADMETox descriptors	0.0	
PM	<input type="checkbox"/>	RC	5670985	988403	67265	HBA	ADMETox descriptors	3.0	
PM	<input type="checkbox"/>	RC	5670985	988402	67265	pKa_bas2	ADMETox descriptors		
PM	<input type="checkbox"/>	RC	5670985	988401	67265	pKa_bas1	ADMETox descriptors	0.53	
PM	<input type="checkbox"/>	RC	5670985	988400	67265	tpsa	ADMETox descriptors	79.31	

Inspecting molecules one by one...





# Manipulating the lists


U	IN	PM	<input type="checkbox"/>	RC	Menu	1074	Copy of Copy of BUDAPEST-2D-Molecules-For-Portal-Training_1072_1073	List for Descriptor calculation	2008-12-01 15:41:00.057	smith
U	IN	PM	<input type="checkbox"/>	RC	Menu	1073	Copy of BUDAPEST-2D-Molecules-For-Portal-Training_1072	List for Descriptor calculation	2008-12-01 15:41:00.057	smith
U	IN	PM	<input type="checkbox"/>	RC	Menu	1072	Copy of Test_Descriptor_Calculation_005_5000_1071	List for Descriptor calculation	2009-05-28 13:36:51.33	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1071	Copy of Test_Descriptor_Calculation_005_5000_1070	List for Descriptor calculation	2009-05-28 13:36:43.503	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1070	Copy of Test_Descriptor_Calculation_004_5000_1069	List for Descriptor calculation	2009-05-28 13:36:36.597	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1069	Copy of Test_Descriptor_Calculation_003_5000_1068	List for Descriptor calculation	2009-05-28 13:36:30.317	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1068	Copy of Test_Descriptor_Calculation_002_5000_1067	List for Descriptor calculation	2009-05-28 13:36:24.16	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1067	Copy of Test_Descriptor_Calculation_001_5000_1066	List for Descriptor calculation	2009-05-28 13:36:17.363	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1066	Test_Descriptor_Calculation_006_5000	User list	2009-05-28 13:36:51.33	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1065	Test_Descriptor_Calculation_005_5000	User list	2009-05-28 13:36:43.503	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1064	Test_Descriptor_Calculation_004_5000	User list	2009-05-28 13:36:36.597	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1063	Search CG code _Calculation_003_5000	User list	2009-05-28 13:36:30.317	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1062	Test_Descriptor_Calculation_002_5000	User list	2009-05-28 13:36:24.16	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1061	Test_Descriptor_Calculation_001_5000	User list	2009-05-28 13:36:17.363	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1060	Model building (type=property)	User list	2009-05-28 10:03:15.177	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1059	Property prediction (type=descriptors)	User list	2009-05-28 10:01:46.397	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1058	Test_Screening_003_5_50000_Duplicated_List_Not_work_property	Screening	2009-05-25 15:20:33.38	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1057	Copy of Test_Screening_003_5_50000_1056	Screening	2009-05-25 15:20:33.38	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1056	Test_Screening_003_4_50000_Duplicated_List_Not_work_property	Screening	2009-05-25 15:20:16.037	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1055	Copy of Test_Screening_003_4_50000_1053	Screening	2009-05-25 15:20:16.037	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1053	Copy of Test_Screening_003_3_50000_1051	Screening	2009-05-25 15:20:09.16	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1051	Copy of Test_Screening_003_2_50000_1044	Screening	2009-05-25 15:20:02.973	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1044	Copy of Test_Screening_003_1_50000_1041	Screening	2009-05-25 15:19:56.253	gpocz
U	IN	PM	<input type="checkbox"/>	RC	Menu	1041	Copy of Test_Screening_002_10000_1040	Screening	2009-05-25 15:19:42.083	gpocz




# Downloading workflow from repository




gUSE RELEASE 3.1



**WS-PGRADE** portal



CancerGrid



Welcome | New Features Settings | AMRI database | **Workflow** | Help

Graph | Create Concrete | Concrete | Applications | Template | Timing | Remoting | Storage | Upload | Import

**Import**

Select type : Application

Application list of selectable

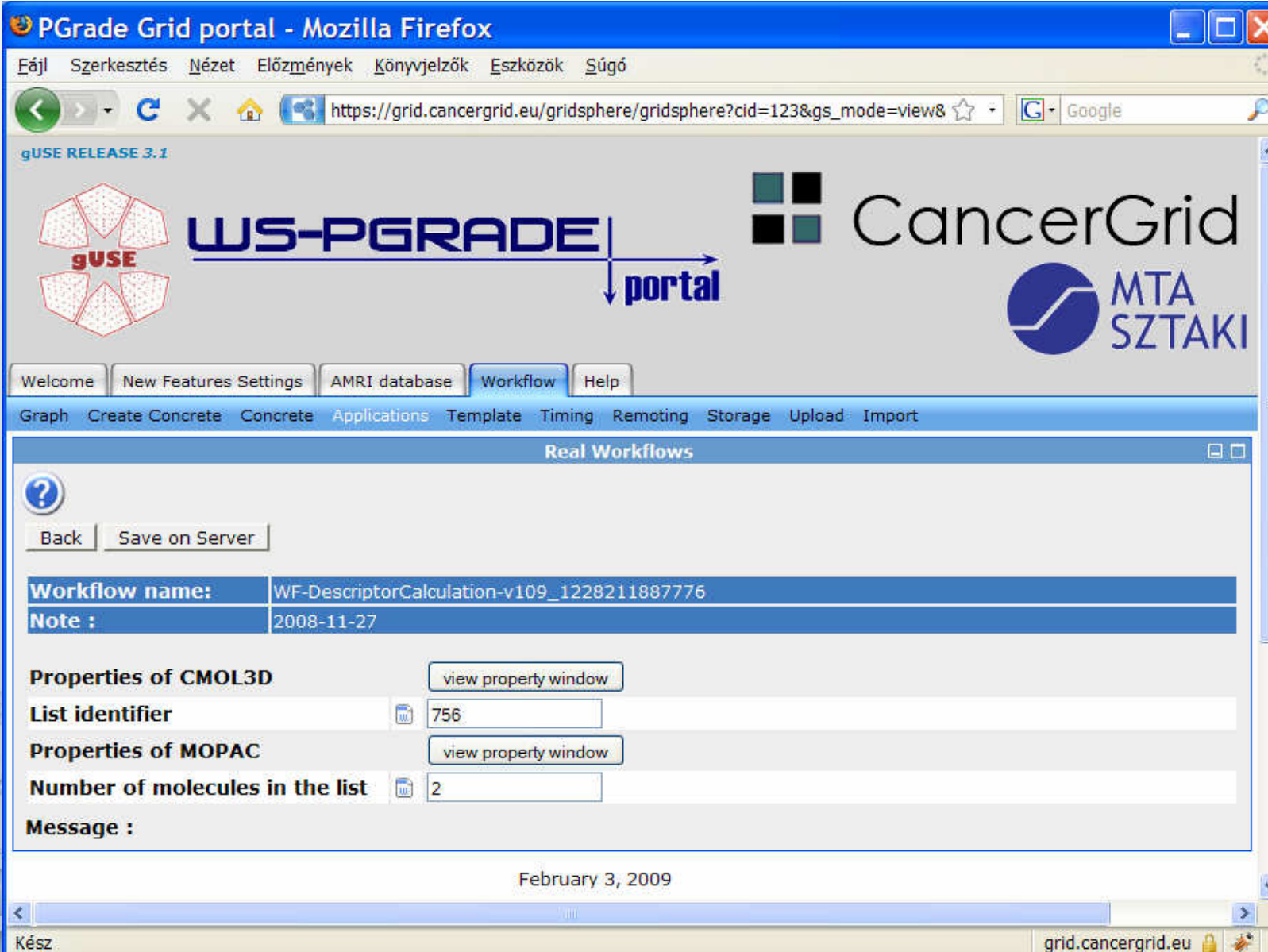
Name	Notes
<input checked="" type="radio"/> WF-DescriptorCalculation-v109	This workflow application takes a list of 2D molecules and performs conformational search for them. Quantum chemical and descriptor calculations are then carried out for each resulting conformer.
<input type="radio"/> WF-ModelBuilding-v102	This workflow application takes the values of descriptors, the values of the selected property for each molecule in the selected list of molecules and forms the matrix for model development. The matrix is then used to build a set of models.
<input type="radio"/> WF-PropertyPrediction-v102	This workflow application takes the values of descriptors for each molecule in the selected list of molecules and forms a matrix for property prediction. The matrix is then used to predict property values for the selected list of structures.

Select action type : Import

Kész

grid.cancergrid.eu

# Workflow configuration




PGrade Grid portal - Mozilla Firefox

Éjl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Sűgó

https://grid.cancergrid.eu/gridsphere/gridsphere?cid=123&gs\_mode=view&

gUSE RELEASE 3.1

 **WS-PGRADE** portal

CancerGrid  
MTA SZTAKI

Welcome New Features Settings AMRI database **Workflow** Help

Graph Create Concrete Concrete Applications Template Timing Remoting Storage Upload Import

**Real Workflows**

Back Save on Server

**Workflow name:** WF-DescriptorCalculation-v109\_1228211887776

**Note :** 2008-11-27

**Properties of CMOL3D**

**List identifier**

**Properties of MOPAC**

**Number of molecules in the list**

**Message :**

February 3, 2009

Kész grid.cancergrid.eu

# Cmol3D property settings

view property window

Cmol3D (conformational search)

Conformational search:



Using 2D molecules:



Population size:

Number of population initializers:

Energy window for saving structures  
(kcal/mol):

Maximum number of structures to save:

Maximum distance between atoms in equal  
structures:

Maximum number of perturbations for a single  
conformation:

Step size for low-mode move:

Maximum distance for low-mode move:

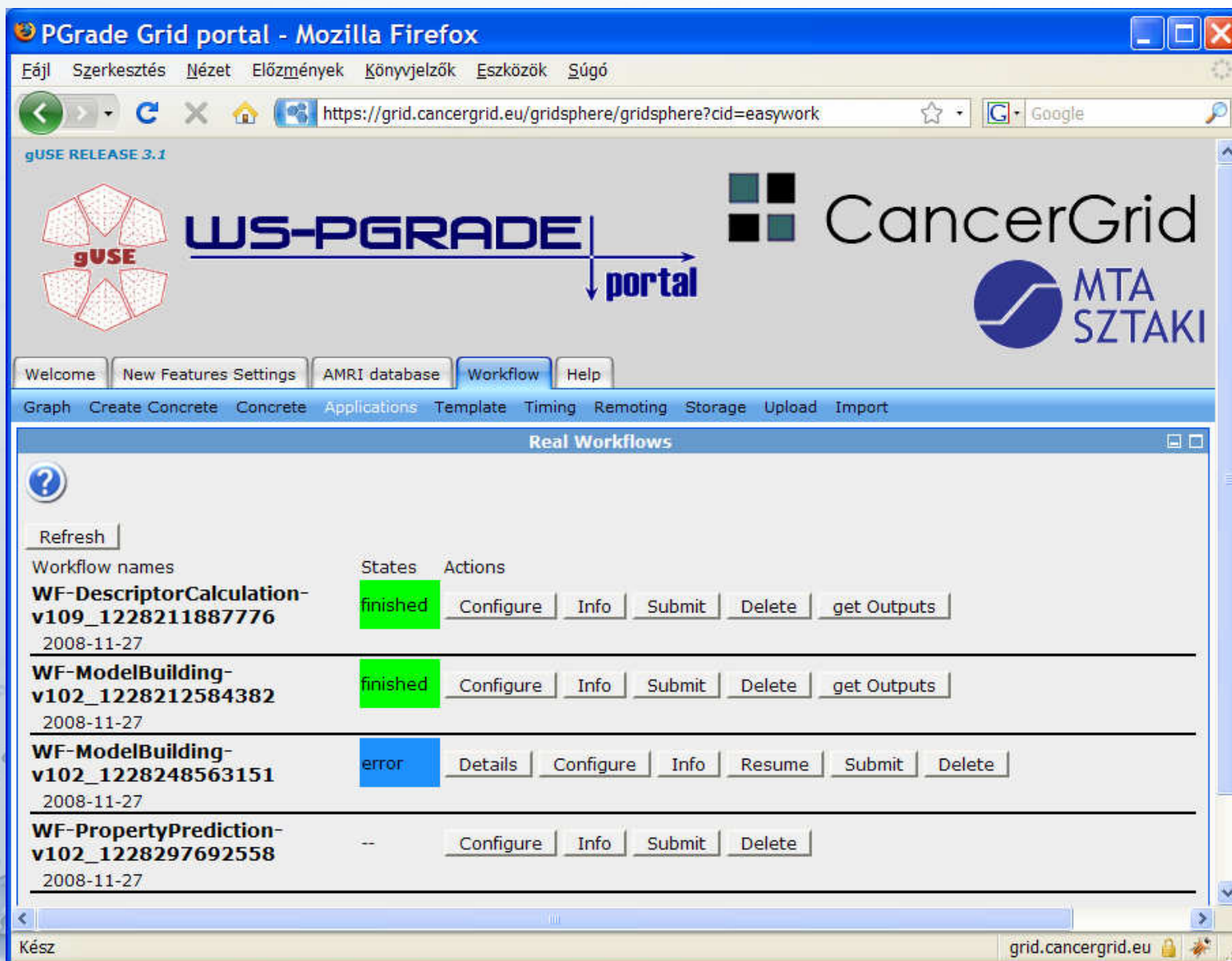
Number of lowest eigenvectors to use:

Ok

Cancel

Defaults

# List of workflows (Novice user view)






PGrade Grid portal - Mozilla Firefox

Eájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

https://grid.cancergrid.eu/gridsphere/gridsphere?cid=easywork

gUSE RELEASE 3.1

 **WS-PGRADE** portal  CancerGrid 

Welcome New Features Settings AMRI database Workflow Help

Graph Create Concrete Concrete Applications Template Timing Remoting Storage Upload Import

**Real Workflows**

Refresh

Workflow names	States	Actions
<b>WF-DescriptorCalculation-v109_1228211887776</b> 2008-11-27	finished	Configure Info Submit Delete get Outputs
<b>WF-ModelBuilding-v102_1228212584382</b> 2008-11-27	finished	Configure Info Submit Delete get Outputs
<b>WF-ModelBuilding-v102_1228248563151</b> 2008-11-27	error	Details Configure Info Resume Submit Delete
<b>WF-PropertyPrediction-v102_1228297692558</b> 2008-11-27	--	Configure Info Submit Delete

Kész grid.cancergrid.eu




# Status monitor

PGrade Grid portal - Mozilla Firefox


Fájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

https://grid.cancergrid.eu/gridsphere/gridsphere?cid=123&gs\_mode=view& Google


gUSE RELEASE 3.1



**WS-PGRADE** portal



CancerGrid



Üdvözöllek Adminisztráció New Features Settings AMRI database Workflow Help

Graph Create Concrete Concrete Applications Template Timing Remoting Storage Upload Import

### Real Workflows

Back Refresh

Workflow name : WF-DC-970-pocze\_1233662682826  
Note : 2008-11-27

Workflow status: **running**

Status	Instances
init	0
<b>running:</b>	22
done:	4734
error:	31
sum:	4787

[portal.easywrk.statusbar]

Kész grid.cancergrid.eu

# Job statuses of a workflow (Advanced view)



PGrade Grid portal - Mozilla Firefox

Fájl Szerkesztés Nézet Előzmények Deliculous Könyvjelzők Eszközök Súgó

https://grid.cancergrid.eu/gridsphere/gridsphere?cid=123&gs\_mode=view&gs

PGrade Grid portal

Workflow Template : WF-DescriptorCalculation-v110-template\_1243944971171

einstance **running** Details Suspend Visualize

Selected WF Instance: einstance

Job	Status	Instances	[ Actions ]
XML-MOL	finished	1	View finished View all content(s)
CONF-wr	finished	277	View finished
	submitted	1	View submitted View all content(s)
LIST-read	finished	1	View finished View all content(s)
CMOL3D	finished	3754	View finished
	waiting	12	View waiting
	submitted	1234	View submitted View all content(s)
MOPAC-MDC	finished	731	View finished
	no input	16	View no input
	waiting	19	View waiting
	running/error	1	View running/error
	submitted	1510	View submitted View all content(s)
SD-MOL	finished	250	View finished View all content(s)
LISTID	finished	1	View finished View all content(s)

Kész grid.cancergrid.eu

# Screenshots



gUSE RELEASE 3.1



**WS-PGRADE** portal



CancerGrid



Logout  
Welcome, Kovacs  
Jozsef

- Welcome
- Administration
- New Features Settings
- AMRI database
- Workflow
- Help
- System load

WFI

## WFI information

Workflow name	Number of jobs	State	State of jobs	
DC-30-4-1070_1244123597990[einstance]	115003	running/error	init	1751
			running	3
			finished	111501
			error	223
			no input	1525
DC-30-6-1072_1244231190059[einstance]	115003	running/error	init	1611
			running	23
			finished	111779
			error	100
			no input	1490

June 15, 2009



# Monitoring the clients





# The system in numbers...

## Statistics

- Currently 70 machines are integrated into the system, the goal is to have about 200 machines
- Requires only 1 central server to be maintained (clients needs no maintenance!)
- 13 algorithms has been integrated
- 4 workflows are available
- 5-6 partners from the consortium have used it for different purposes

## Peak performance

- Description calculation for 30.000 compounds (with 10 conformers) takes 5 days (executing the combination of CMOL3D, MOPAC and MDC tools from CODESSA using 3D descriptors). With 1 PC it could be a year...

Note: there are faster algorithms, but Grid is worth applying only for computationally intensive algorithms (like the ones above)

- By decreasing the number of conformers and by increasing the number of machines, the calculation can be even 100 times faster i.e. millions of compounds can be processed within 1 week. This is our current direction...

# Summary & Future plan

- **Any community** that has several computational intensive algorithms can easily execute them on a SZTAKI DesktopGrid (SZDG) system:
  - the community can create its **own institutional SZDG** project
  - can simply **attach** the available **PC computers** as resources
  - can **insert the algorithms** in their SZDG project
  - can easily **combine these components** into more complex workflow applications, if necessary
- Such a system
  - has been prototyped for the Cancer Research community within the CancerGrid projects
  - can be **specialized** for any community requirements
- SZTAKI is ready to **support other communities** with this technology
- In the future
  - we plan to continue extending the CancerGrid portal with more algorithms, workflows and PC resources
  - we plan to create an open portal where various algorithms from the field of computational chemistry and bioinformatics will be supported, which will be publicly available
  - we plan to integrate ZINC database

If you need more detailed (technical) information,  
email to [desktopgrid@lpds.sztaki.hu](mailto:desktopgrid@lpds.sztaki.hu) or  
visit [www.desktopgrid.hu](http://www.desktopgrid.hu)



Thank you for your attention!  
Questions?



Acknowledgement:

CancerGrid EU FP6 project (FP6-2005-LIFESCTHTALTH-7)

<http://www.cancergrid.eu>