



# BOINC extensions in the SZTAKI DesktopGrid system

József Kovács  
smith@sztaki.hu



*BOINC Workshop, Grenoble, 10/09/2008*

# SZTAKI Desktop Grid: BOINC project



SZTAKI Desktop Grid - Microsoft Internet Explorer

Address: <http://szdg.lpds.sztaki.hu/szdg/>

MTA SZTAKI SZTAKI DESKTOP GRID BOINC

Number of users: **22947** Number of hosts: **31004**

Active hosts in last 48 hours: **1479**

Estimated performance of last 48 hours<sup>1,2</sup>: **755.531 GFlop/s** Peak performance<sup>1,2</sup>: **1.5 TFlop/s**

Workunits processed in last 48 hours<sup>3</sup>: **N/A**

Join SZTAKI Desktop Grid

Rules and policies [read this first]  
Getting started  
More information on desktopgrid.hu  
Wikipedia article about SZTAKI Desktop Grid  
Frequently Asked Questions(FAQ)  
Create account  
Applications

Returning participants

Your account - view stats, modify preferences

Application currens

SZTAKI Desktop G  
number systems.  
Description on the ap

User of the day

Arisa

BOINC Manager

Choose a project

To choose a project, click its name or type its URL below.

- World Community Grid
- Malariacontrol.net
- Proteins@Home
- Predictor@home
- GPU Grid - P53 Grid
- Superlink@Technion
- The Lattice Project
- Rosetta@home
- Chess@Home
- SZTAKI Desktop Grid**
- PrimeGrid
- SHA-1 Collision Search Graz

Projekt URL:

< Vissza Előre > Mégsem

<http://szdg.lpds.sztaki.hu/szdg>



# SZTAKI Desktop Grid: BOINC extensions



SZTAKI Desktop Grid is a collection of various developments towards Desktop Grid direction, based on BOINC:

- **Debian package** of the BOINC server
- application programming interface: **DC-API**
- integration with various backends: e.g: **Condor**
- supporting various application types on the client side: e.g.: **Java, MPI**
- aggregating the power of different BOINC projects: **hierarchically** connected DGs
- ease the application porting under BOINC client: **genwrapper**
- improving security: introducing **certificates** in BOINC, **sandboxing** under BOINC client
- generalise job creation/handling: **queuemanager** on BOINC server

Most of them can be downloaded from [www.desktopgrid.hu](http://www.desktopgrid.hu), others are under development, ⇒ **desktopgrid@lpds.sztaki.hu**

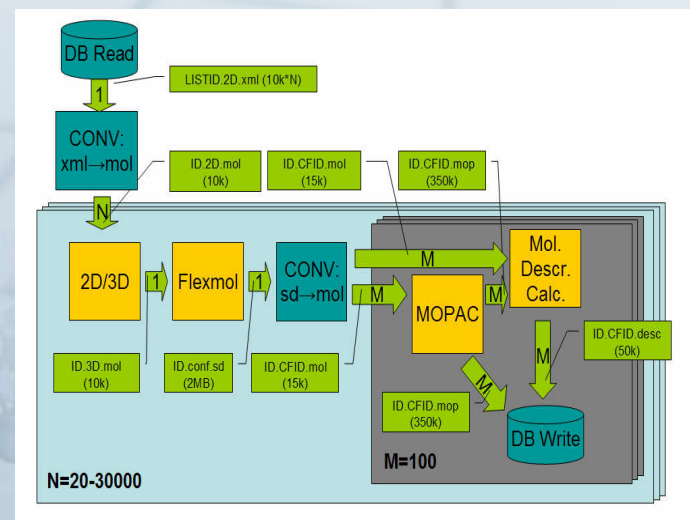
# CancerGrid EU FP6 project

- Grid Aided Computer System For Rapid Anti-Cancer Drug Design
  - January 1, 2007 – December 31, 2009
  - 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

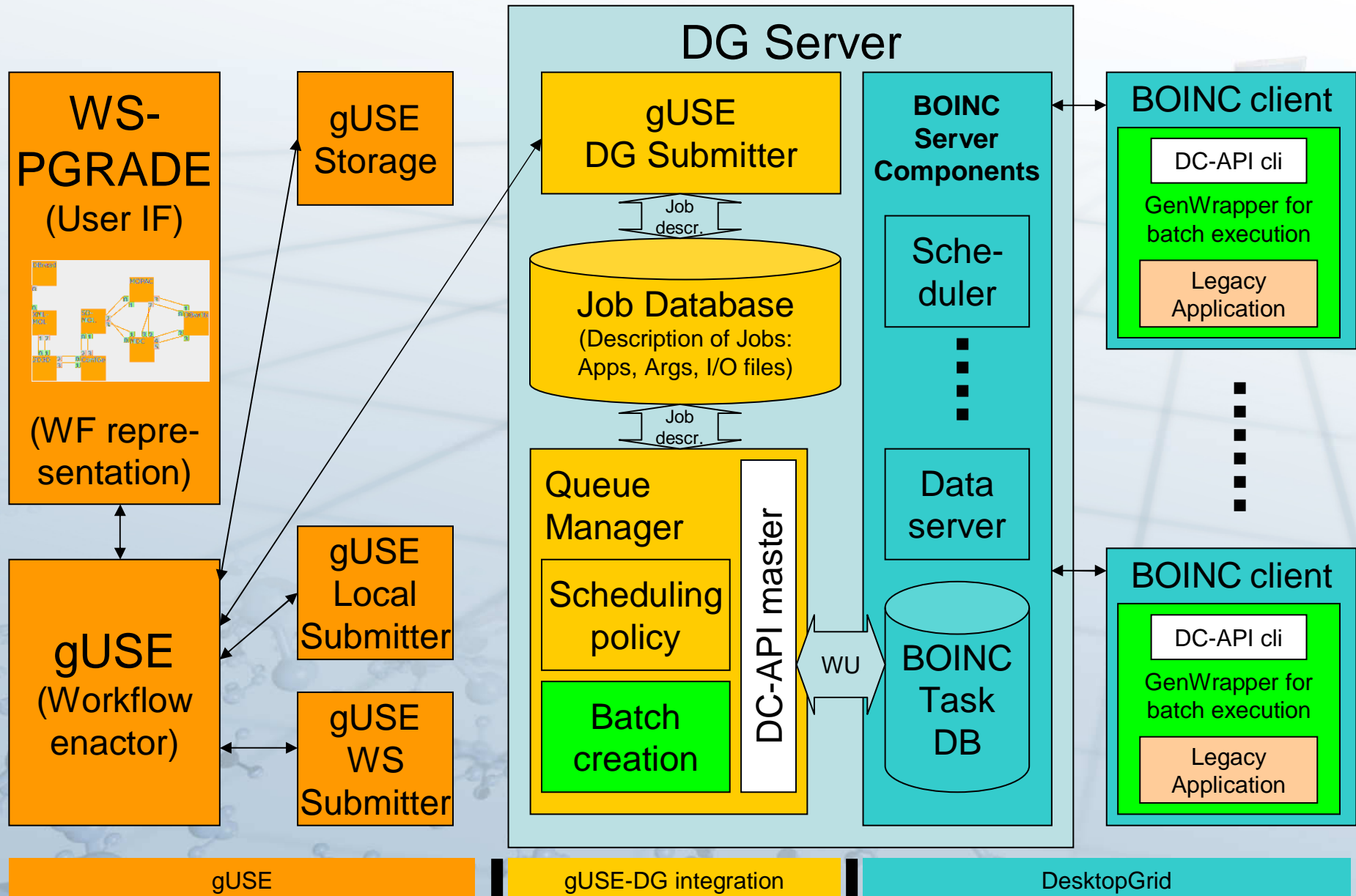


Bag of tasks: cmol3d, mopac, mdc, fmt, fma, etc.

- Fortran, C, C++
- Processing/Memory requirements
- Multi-binary applications, Libraries
- For some apps source is not available
- Config file preparation before execution
- Pure logging/debugging information



# The CancerGrid architecture



# Generic Wrapper (GenWrapper)

- **Why did we developed?**
  - The features of **BOINC wrapper is not enough** (e.g. patching config files on client machines, generating extra messages, independent jobs in a WU, etc.)
  - Wanted to **be prepared for unknown requirements** might be raised by future Cancergrid applications
  - We did not want to extend the BOINC wrapper to make it an XML-based programming language, we choose to **BOINCify an existing language** -> Bourne shell
- **How does it work?**
  - a shell interpreter (gitbox - a variant of busybox) is started instead of the real application
  - it executes an application script, that
    - realizes boincification through script commands
    - may run legacy applications in any way (e.g.: multiple input process)
    - may perform any preparation on input-, output files, environment, etc.
    - may do **whatever you can do by a script**

## Application 1

application1\_1.01\_windows\_intelx86.exe

application1\_1.01\_windows\_intelx86.zip

## slot dir

uncompress

GenWrapper

Legacy Executable 1..n (LE)

Legacy Executable Dependencies

Profile Script

## Workunit

Input Files

command line parameters

Application Script

Output Files

1. unzips `%BASENAME%.zip` to slot dir
2. executes `dc_init()` or `boinc_init()` [ and `dc_finish()` or `boinc_finish()` at the end ]

3. creates a script that:
  - sources the Profile Script
  - starts the Application Script
4. starts GenWrapper

5. executes the generated script
  - compound BOINC application

- handles architecture/ platform dependent pre-run tasks for the application

- architecture independent tasks
- handles batch-run
- prepares Input Files for *LE(s)*
- executes *LE(s)*
- moves Output Files to predefined location

6. exit status becomes the exit status of the WU

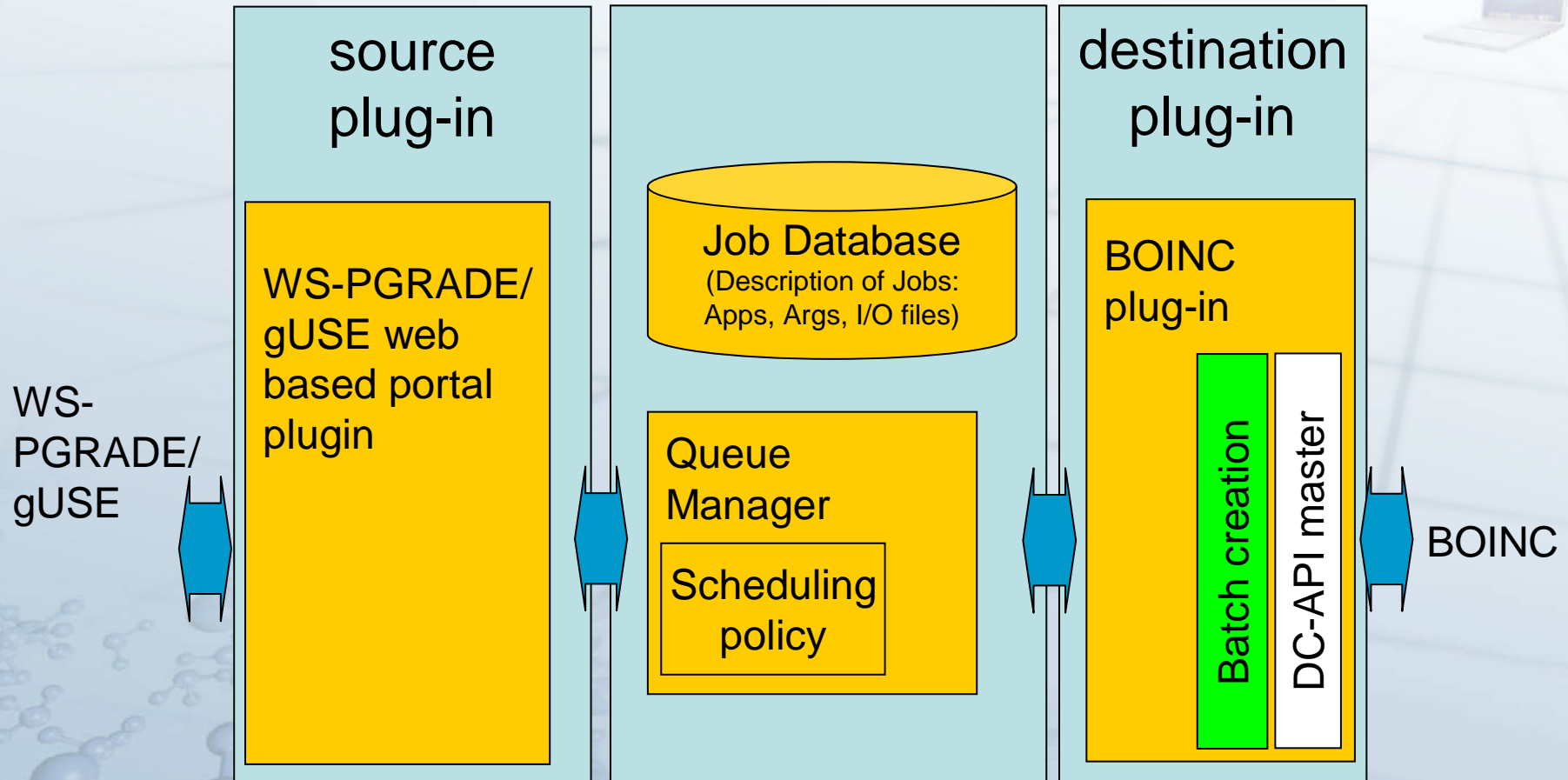
# Sample GenWrapper Script

```
1. IN=`boinc resolve_filename in`
2. OUT=`boinc resolve_filename out`
3. NUM=`cat ${IN}`
4. PERCENT_PER_ITER=$((100000 / NUM))
5. for i in `seq $NUM`; do
6.     PERCENT_COMPLETE=$((PERCENT_PER_ITER * i / 1000))
7.     boinc fraction_done_percent ${PERCENT_COMPLETE}
8.     echo -e "I am ${PERCENT_COMPLETE}% complete." >> ${OUT}
9.     sleep 1;
10.done
```

- shell script contains the BOINC commands
- every filename needs to be resolved
- status: on-going development, still missing some features (CPU time calculation, signal handling, background process, checkpointing, etc.)



# Integration of the web-portal to DG



# Batching in QM

- Substrings like “`%{<word>}`” are substituted with the appropriate value. Unknown substitutions are left alone and copied as-is
- 3 template scripts must be prepared for every app
- Head template: extracts `%{inputs}`
- Per-job templates:
  - All input files are under `%{input_dir}` (relative to the directory where the script is started)
    - Moves the input files to appropriate location if necessary
  - Calls “*application* `%{args}`”
  - Moves all output files to `%{output_dir}` (relative to the directory where the script is started)
- Tail template: packs the directory `%{output_dir}` as `%{outputs}`



# Example templates

- Example head template:

```
set +e  
tar xzf %{inputs}  
BASEDIR=`pwd`
```

- Example per-job template:

```
cd $BASEDIR/%{input_dir}  
$BASEDIR/app %{args} >stdout 2>stderr  
mv out_file stdout stderr $BASEDIR/%{output_dir}  
cd -
```

- Example tail template:

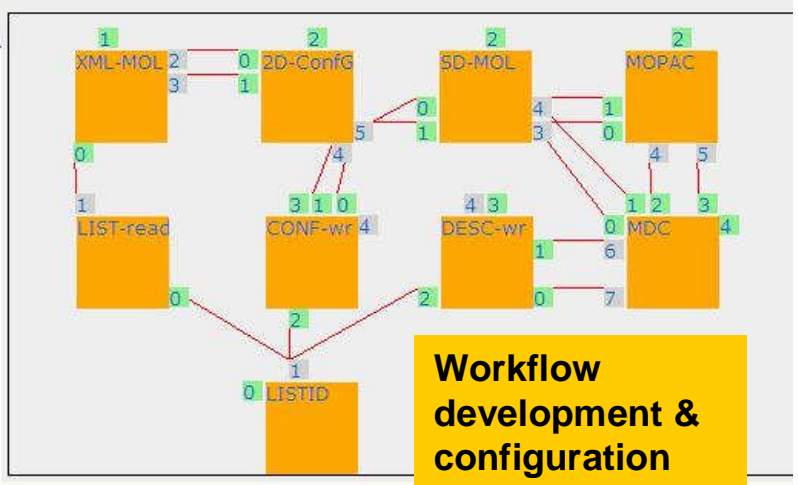
```
cd $BASEDIR/outputs  
tar czf $BASEDIR/%{outputs} *
```



# 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><br>no input | 1<br>1    | <a href="#">View content(s)</a> |
| MOPAC      | <span style="background-color: green; color: white;">finished</span><br>no input | 41<br>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<br>DESC | ASC<br>DESC | ASC<br>DESC | ASC<br>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**

# Conclusion

- Any community that has a class of workflow type applications requiring bag of task type of components can easily use a BOINC system:
  - the community can create its own non-public BOINC project
  - can easily map the bag of task components into BOINC applications
  - can easily combine these components into more complex workflow applications
- Such a system
  - has been prototyped for the Cancer Research community within the CancerGrid projects
  - will be available as production system in Q4 of 2008
- Within the EDGeS project we would like to support other communities with this technology



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>