

XtalOpt User Guide

Release 9.0

Generated by Doxygen 1.8.9.1

Tue Sep 8 2015 09:41:13

Contents

1	XtalOpt Tutorial	1
1.1	Contents	1
1.2	Launch XtalOpt	1
1.3	Enter composition and restraints	2
1.4	Optimizer setup	2
1.4.1	VASP	3
1.4.2	GULP	4
1.4.3	PWscf	5
1.4.4	CASTEP	6
1.4.5	SIESTA	7
1.5	Queue setup	8
1.5.1	Using a remote PBS cluster	9
1.5.2	Using a remote SGE cluster	11
1.5.3	Using a remote SLURM cluster	12
1.5.4	Using a remote LSF cluster	13
1.5.5	Using a remote LoadLeveler cluster	14
1.5.6	Running optimizations locally	15
1.6	What is written to the local directory?	16
1.7	Search Settings	16
1.8	"Begin"	17
1.9	Monitor progress	18
1.9.1	View trends	19
2	Saving and Resuming Sessions in XtalOpt	20
2.1	Contents	20
2.2	How to save your session	20
2.3	How to resume your session	20
3	Optimization Schemes	21
3.1	Contents	21
3.2	Overview: What are optimization schemes, and why use them?	21
3.2.1	In a nutshell..	21
3.2.2	More details	22
3.3	Optimization scheme user interface	23
3.3.1	Optimization step list	23
3.3.2	Add new optimization step	23
3.3.3	Remove current optimization step	23
3.3.4	Select template	24
3.3.5	Template editor	24

3.3.6	Save scheme	24
3.3.7	Resume scheme	24
3.4	How to build an optimization scheme?	24
3.5	How to save an optimization scheme for later?	25
3.6	How to load an optimization scheme?	25
3.7	What is saved?	25
3.8	Suggestions for optimization schemes	25
3.8.1	Crystals (XtalOpt)	25
4	RandomDock Tutorial	25

1 XtalOpt Tutorial

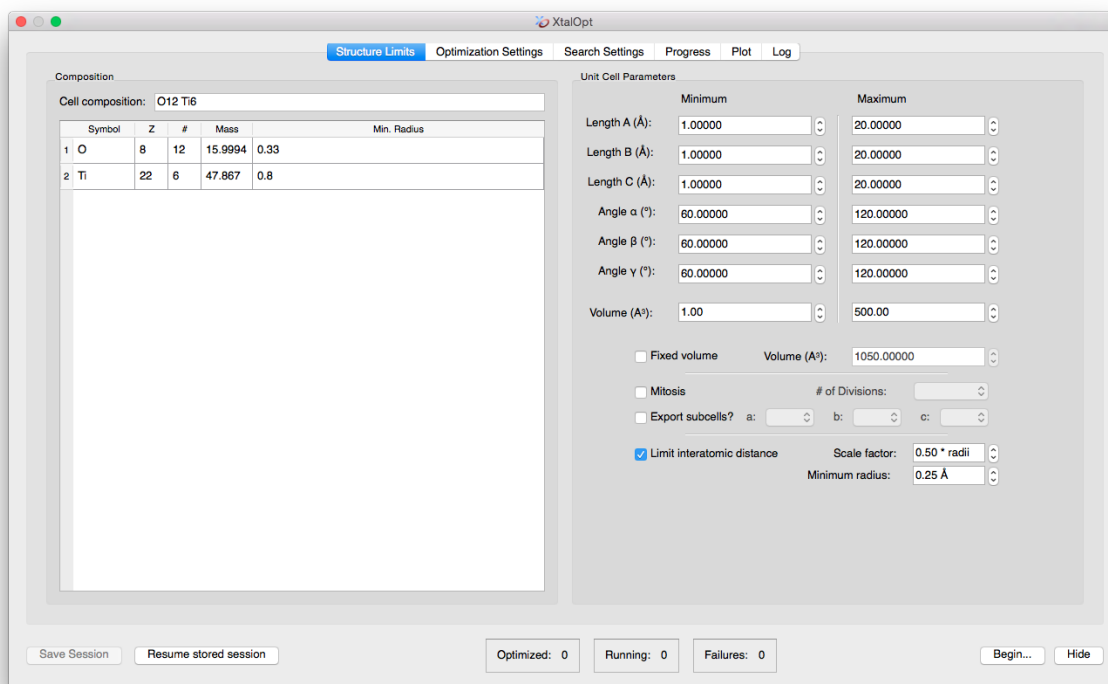
1.1 Contents

- [Launch XtalOpt](#)
- [Enter composition and restraints](#)
- [Optimizer setup](#)
 - [VASP](#)
 - [GULP](#)
 - [PWscf](#)
 - [CASTEP](#)
 - [SIESTA](#)
- [Queue setup](#)
 - [Using a remote PBS cluster](#)
 - [Using a remote SGE cluster](#)
 - [Using a remote SLURM cluster](#)
 - [Using a remote LSF cluster](#)
 - [Using a remote LoadLeveler cluster](#)
 - [Running optimizations locally](#)
- [What is written to the local directory?](#)
- [Search Settings](#)
- ["Begin"](#)
- [Monitor progress](#)
 - [View trends](#)

1.2 Launch XtalOpt

Open Avogadro, go to the "Extensions" menu and select "XtalOpt".

1.3 Enter composition and restraints

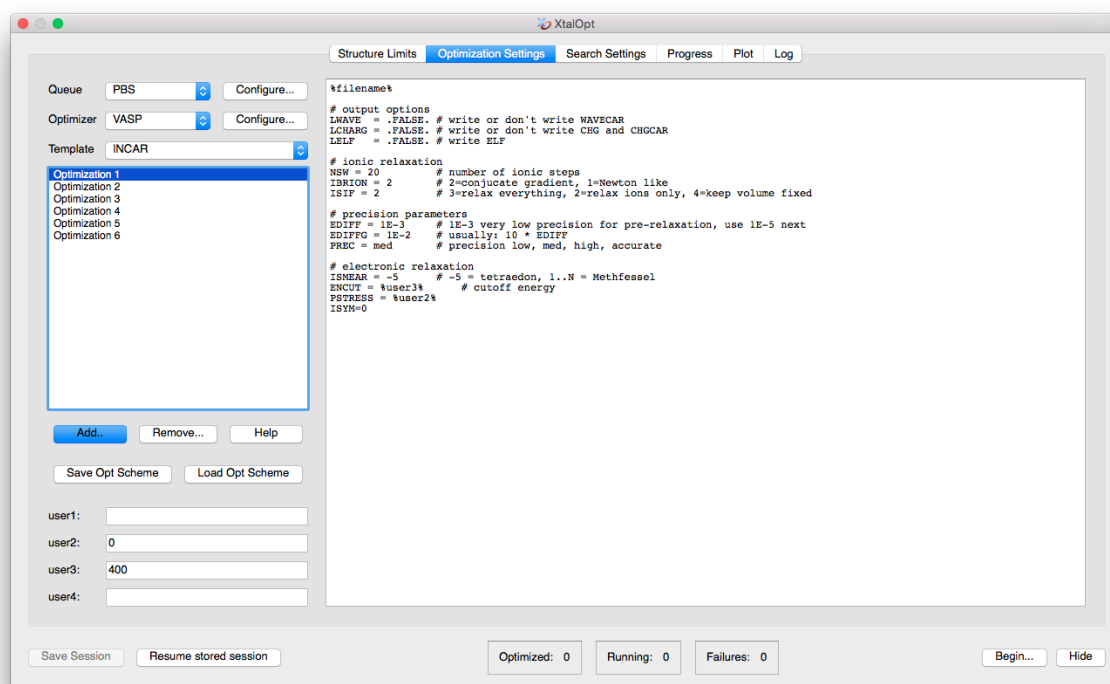


The interface opens to the "Structure Limits" tab, shown above. We will use a 6 formula unit supercell of titanium dioxide for this tutorial, so enter "Ti6 O12" for the cell composition. We will assume that we know nothing about the system and use very loose restraints (however, note that a search is much more effective if chemically reasonable restraints are used). Set all cell length minima to 1 angstrom and maxima to 20 angstroms. Constrain the angles to be between 60 and 120 degrees, and the volume from between 1 and 500 cubic angstroms. Specify a minimum interatomic distance of 0.5 angstrom. (Note that due to the angle adjustment described in CPC, 2011, 182, 372-387, 60-120 degrees is the largest range of cell angles that XtalOpt will generate.)

1.4 Optimizer setup

XtalOpt currently supports the [VASP](#), [GULP](#), [PWscf](#), [CASTEP](#), and [SIESTA](#) codes for performing geometry optimizations. Each is detailed in its own section below.

1.4.1 VASP



On the next tab, load the optimization scheme by clicking the "Load Opt Scheme" button and selecting the "samples/vasp-xtalopt.scheme" file that is distributed with the source code. If you do not have a copy of the source code, the scheme file can be obtained by clicking [here](#).

For more details on optimization schemes, see [Optimization Schemes](#).

After loading the optimization scheme, XtalOpt will prompt for the POTCAR files to use. Select files appropriate for the prompted atom. XtalOpt will construct the POTCAR files on the local computer, and then copy them over to the cluster when the calculation is submitted. It is necessary to have the VASP POTCAR files for each atomic species located somewhere on the local computer. See the VASP manual for information on obtaining the POTCAR files.

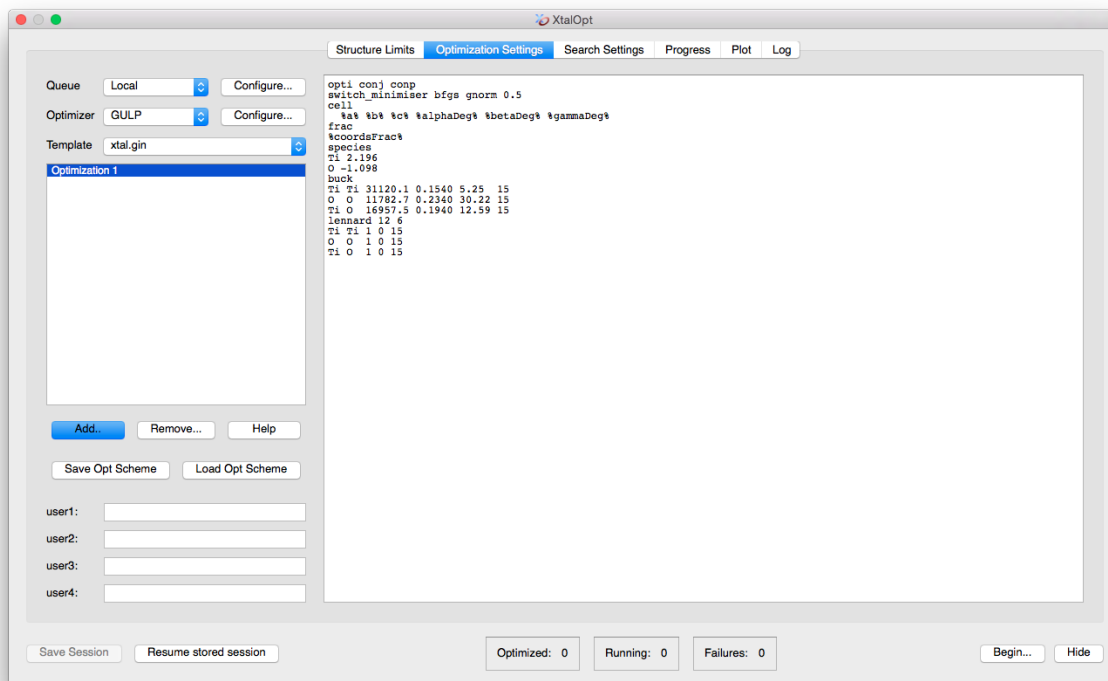
Take a moment to look through each file for each optimization step. Notice that the INCAR template includes two user-specified values, %user2% and %user3% for the external pressure and the energy cutoff, respectively. By entering appropriate values in the "user2:" and "user3:" fields on the left, it is easy to update these values for all optimization steps.

Notice the other %keyword% values in the job.pbs templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

XtalOpt expects VASP to use the default filenames, mainly POSCAR, CONTCAR, and OUTCAR.

[Skip to next section.](#)

1.4.2 GULP



On the next tab we choose GULP for the local optimizer and enter a template for GULP to use. Select "GULP" as the "Optimizer" and "xtal.gin" as "Template". Next, fill out the text field on the right with the following template:

```
opti conj comp
switch_minimiser bfgs gnorm 0.5
cell
  %a% %b% %c% %alphaDeg% %betaDeg% %gammaDeg%
frac
%coordsFrac%
species
Ti 2.196
O -1.098
buck
Ti Ti 31120.1 0.1540 5.25 15
O O 11782.7 0.2340 30.22 15
Ti O 16957.5 0.1940 12.59 15
lennard 12 6
Ti Ti 1 0 15
O O 1 0 15
Ti O 1 0 15
```

Alternatively, one can load the scheme file distributed with the source code under `samples/gulp-TiO-xtalopt.scheme`. If the source code is not available, the scheme file can be obtained by clicking [here](#).

For more details on optimization schemes, see [Optimization Schemes](#).

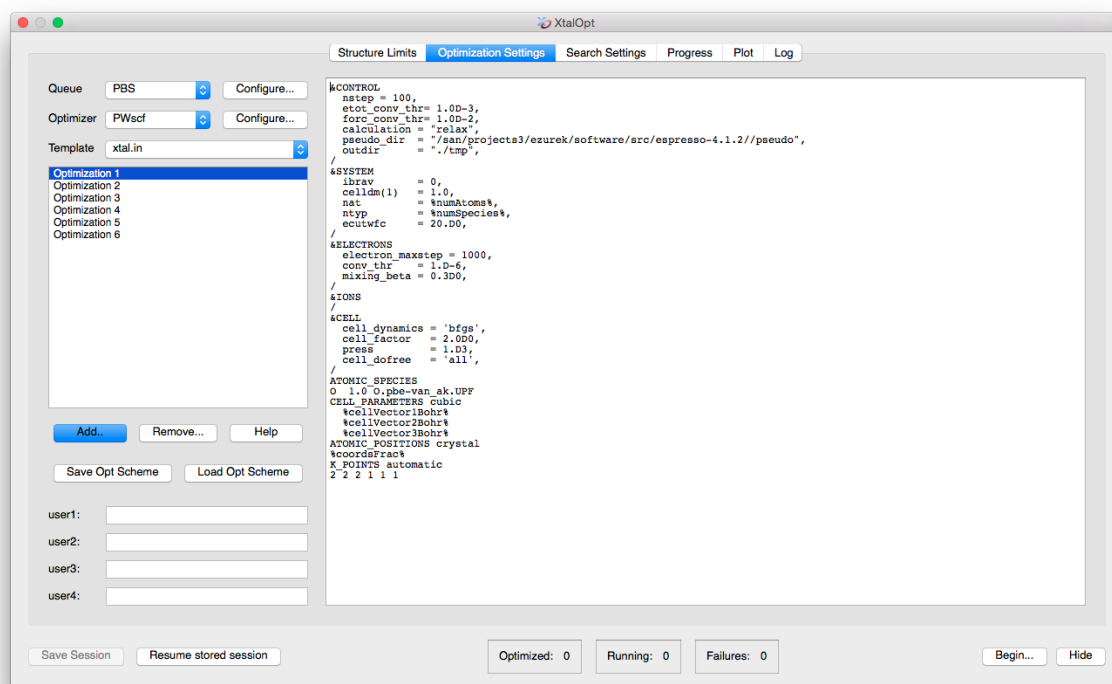
Note the "%" surrounding various keywords. These will be replaced by the structure-specific data when the optimizer is invoked for each structure. Click "Help" to view all of the keywords available. The number of optimization steps can be modified with the "Add/Resume" buttons. The "user" fields in the lower left corner allow users to specify their own keyword/value pairs, which is useful for making changes to multiple optimization steps at once. We will only be using one optimization step in this tutorial.

XtalOpt expects GULP to use the following filenames:

```
gulp < xtal.gin > xtal.got
```

[Skip to next section.](#)

1.4.3 PWscf



On the next tab, load the optimization scheme that is distributed with the source code under the samples/ directory. The scheme that we want is named "pwsf-xtalopt.scheme". If the source code is not available, the scheme file can be obtained by clicking the "Original Format" link at the bottom of the page [here](#).

For more details on optimization schemes, see [Optimization Schemes](#).

Each PWscf input file will need to be edited to specify:

1. The pseudo_dir containing the pseudopotential files on the remote cluster, and
2. The pseudopotentials for each atom (under ATOMIC_SPECIES)

Take a moment to look through each file for each optimization step.

Notice the %keyword% values in the job.pbs templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

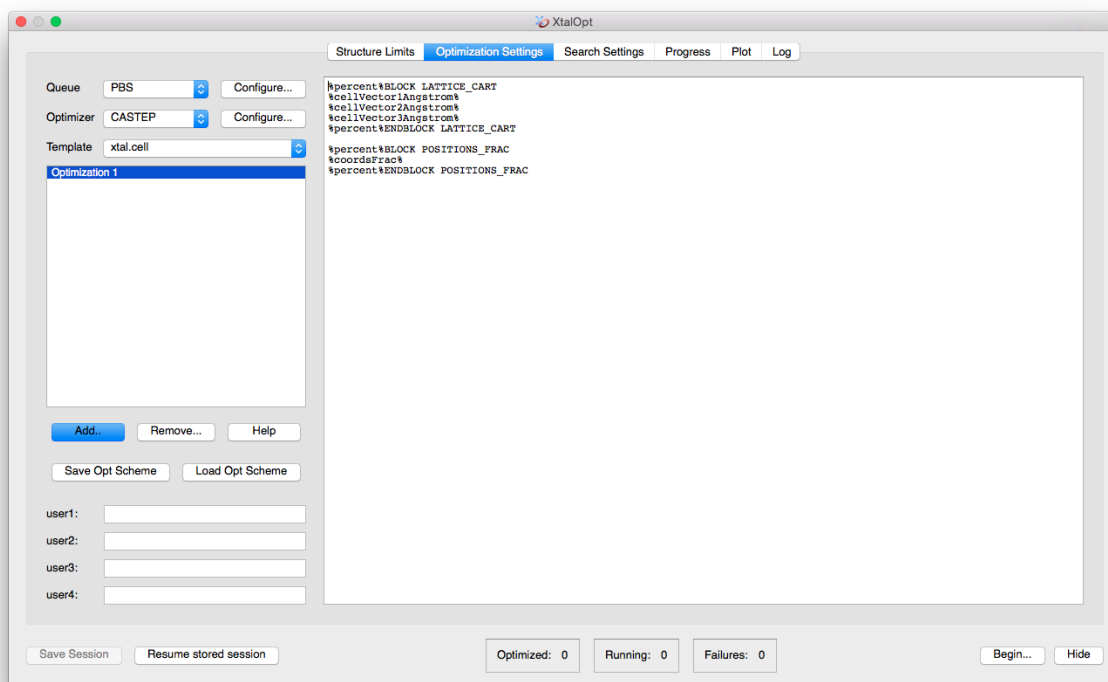
Be aware that every PWscf/CASTEP installation is different, and it is almost certain that the job.pbs file included with this scheme will not work on any cluster other than the Zurek group's "parity" cluster at SUNY Buffalo's Center for Computational Resources. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance for information about MPI, executable locations, etc. Perhaps the easiest method to find the correct PBS script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

XtalOpt expects PWscf to use the following filenames:

```
pw.x < xtal.in > xtal.out
```

[Skip to next section.](#)

1.4.4 CASTEP



On the next tab, load the optimization scheme that is distributed with the source code under the `samples/` directory. The scheme that we want is named "castep-xtalopt.scheme". If the source code is not available, the scheme file can be obtained by clicking the "Original Format" link at the bottom of the page [here](#).

For more details on optimization schemes, see [Optimization Schemes](#).

It is important to note that CASTEP input files require the "%" character to define blocks. The percent character is special in the XtalOpt input template parser to define keywords (see below). To insert a literal "%" into the input, use `percent%`.

E.g. Specification of the fractional coordinate block in the `.cell` template should look like:

```
%percent%BLOCK POSITIONS_FRAC
%coordsFrac%
%percent%ENDBLOCK POSITIONS_FRAC
```

Take a moment to look through each file for each optimization step.

Notice the `%keyword%` values in the `job.pbs` templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

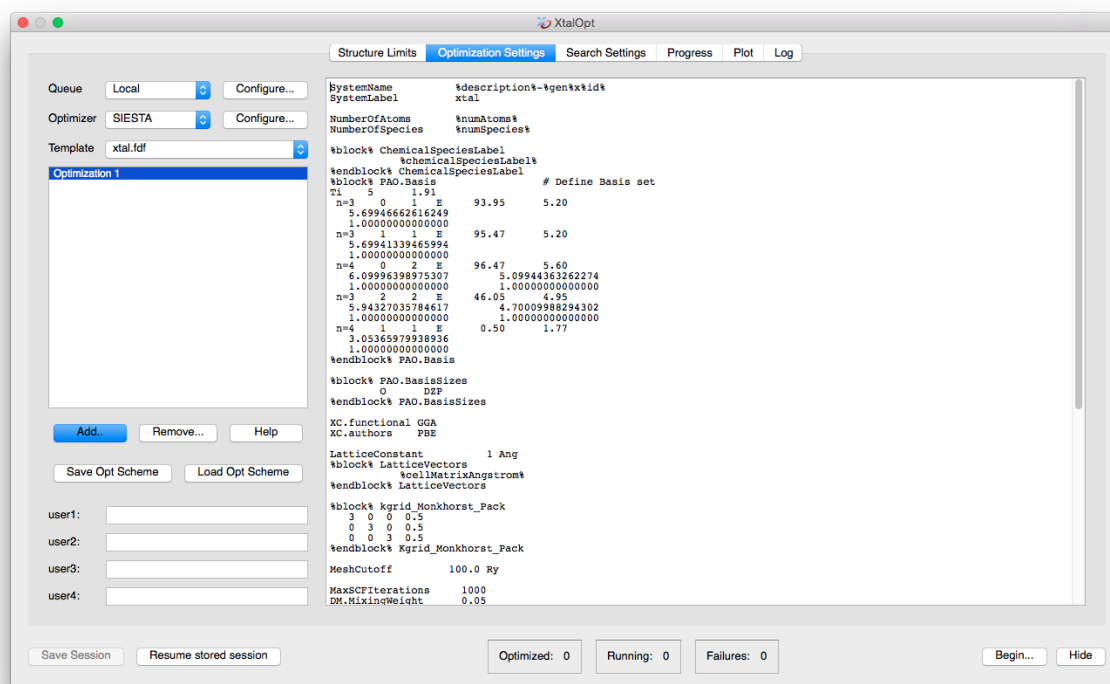
Be aware that every PWscf/CASTEP installation is different, and it is almost certain that the `job.pbs` file included with this scheme will not work on any cluster other than the Zurek group's "parity" cluster at SUNY Buffalo's Center for Computational Resources. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance for information about MPI, executable locations, etc. Perhaps the easiest method to find the correct PBS script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

XtalOpt expects CASTEP to use the following filenames:

```
# XtalOpt will write xtal.cell, xtal.param
castep xtal
# CASTEP will create xtal.castep
```

[Skip to next section.](#)

1.4.5 SIESTA



On the next tab we choose SIESTA for the local optimizer and enter a template for SIESTA to use. Select "SIESTA" as the "Optimizer" and "xtal.fdf" as "Template".

Next, fill out the text field on the right with the following template:

```
SystemName          %description%-%gen%x%id%
SystemLabel         xtal

NumberOfAtoms       %numAtoms%
NumberOfSpecies     %numSpecies%

%block% ChemicalSpeciesLabel
%chemicalSpeciesLabel%
%endblock% ChemicalSpeciesLabel

%block% PAO.Basis          # Define Basis set
Ti      5      1.91
n=3     0      1 E      93.95      5.20
5.69946662616249
1.0000000000000000
n=3     1      1 E      95.47      5.20
5.69941339465994
1.0000000000000000
n=4     0      2 E      96.47      5.60
6.09996398975307      5.09944363262274
1.0000000000000000
n=3     2      2 E      46.05      4.95
5.94327035784617      4.70009988294302
1.0000000000000000
n=4     1      1 E      0.50      1.77
3.05365979938936
1.0000000000000000
%endblock% PAO.Basis

%block% PAO.BasisSizes
O      DZP
%endblock% PAO.BasisSizes

XC.functional GGA
XC.authors    PBE
```

```

LatticeConstant      1 Ang
%block% LatticeVectors
%cellMatrixAngstrom%
%endblock% LatticeVectors

%block% kgrid_Monkhorst_Pack
  3  0  0  0.5
  0  3  0  0.5
  0  0  3  0.5
%endblock% Kgrid_Monkhorst_Pack

MeshCutoff           100.0 Ry

MaxSCFIterations      1000
DM.MixingWeight        0.05
DM.NumberPulay         3
DM.Tolerance           1.d-4

SolutionMethod        diagonal

SpinPolarized         true
LongOutput             true

MD.TypeOfRun           cg
MD.NumCGSteps          1000
MD.VariableCell         true
MD.MaxForceTol         0.01 eV/Ang #0.005 eV/Ang
WriteForces             true
WriteCoorCeriux         true
WriteCoorXmol           false
WriteDenchar            true
WriteMullikenPop        1

UseSaveData            true

#Diag.ParallelOverK true

AtomicCoordinatesFormat Fractional
%block% AtomicCoordinatesAndAtomicSpecies
%atomicCoordsAndAtomicSpecies%
%endblock% AtomicCoordinatesAndAtomicSpecies

```

Or load the optimization scheme by clicking the "Load Opt Scheme" button and selecting the "samples/siesta-TiO-xtalopt.scheme" file that is distributed with the source code. If the source code is not available, the scheme file can be obtained by clicking [here](#).

For more details on optimization schemes, see [Optimization Schemes](#).

After loading the optimization scheme, XtalOpt will prompt for the xtal.psf files to use. Select files appropriate for the prompted atom. XtalOpt will copy the individual files over to each structure directory when the calculation is submitted. See the SIESTA manual for information on obtaining the .psf files.

Notice the other keyword% values in the xtal.fdf templates. These are used to enter information that is specific to a search or structure when the actual input files are written prior to job submission. Click the "Help" button for a full listing of the available keywords.

Note: that in the current implementation XtalOpt uses the Total Final Energy printed in the output to determine the fitness of a structure. If the user would like to use a different thermodynamic quantity for the fitness, please contact the XtalOpt developers.

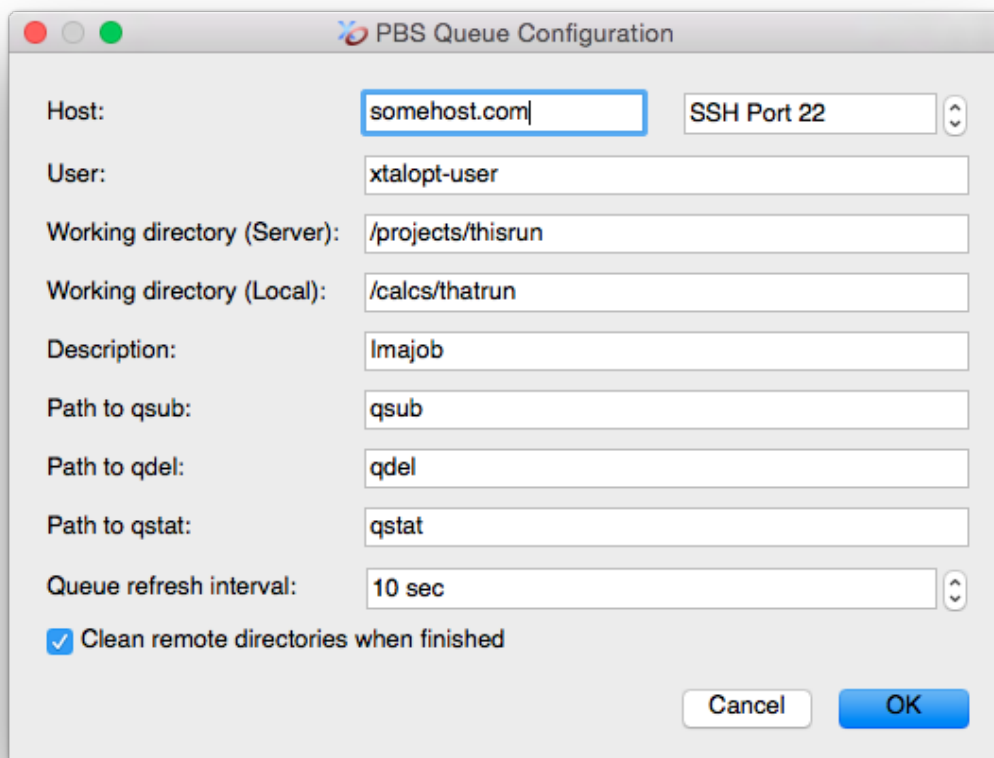
XtalOpt expects SIESTA to use the following filenames:

```
siesta < xtal.fdf > xtal.out
```

1.5 Queue setup

XtalOpt currently supports using the [PBS](#), [SGE](#), [SLURM](#), [LSF](#), and [LoadLeveler](#) queuing systems on remote SSH-accessible clusters, as well as an internal [local](#) queue that manages calculations on the user's workstation. Each queuing interface is detailed in its own section below.

1.5.1 Using a remote PBS cluster



The image shows a 'PBS Queue Configuration' dialog box with the following fields and values:

Field	Value
Host:	somehost.com
SSH Port	22
User:	xtalopt-user
Working directory (Server):	/projects/thisrun
Working directory (Local):	/calcs/thatrun
Description:	Imajob
Path to qsub:	qsub
Path to qdel:	qdel
Path to qstat:	qstat
Queue refresh interval:	10 sec
Clean remote directories when finished	<input checked="" type="checkbox"/>

Buttons: Cancel, OK

Select "PBS" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- host: The hostname of the PBS cluster's head node
- user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- Description: Used for the %description% keyword in input templates.
- Path to qsub: Where to find the qsub executable on the remote cluster. Note that if qsub is in the cluster's \$PATH, setting this to just 'qsub' will work.
- Path to qdel: Where to find the qdel executable on the remote cluster. Note that if qdel is in the cluster's \$PATH, setting this to just 'qdel' will work.
- Path to qstat: Where to find the qstat executable on the remote cluster. Note that if qstat is in the cluster's \$PATH, setting this to just 'qstat' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.pbs" is added to the list of available templates. This is the job submission script for PBS. This script should roughly follow this design:

```

#!/bin/bash
#PBS -l nodes=1:ppn=8
#PBS -o ../%gen%x%id%-%optstep%.out
#PBS -e ../%gen%x%id%-%optstep%.err
#PBS -N %description%-%gen%x%id%-%optstep%

###Include this for XtalOpt scripts!###
export PBS_O_WORKDIR=%rempath%

# Change to structure's working directory, copy input files to node's scratch dirs:
for node in `cat $PBS_NODEFILE | sort | uniq`; do
rsh $node "cp $PBS_O_WORKDIR/* $PBSTMPDIR/";
done

# Move to the scratch directory
cd $PBSTMPDIR
echo "running in directory $PBSTMPDIR"

# Set any environment variables needed for the optimizer/MPI here:

# Run optimizer, be sure to use the filenames that XtalOpt expects.
# See the template menu in XtalOpt and the example templates in the
# samples/ directory of the XtalOpt sources.

# Don't forget to clean up after MPI if needed!

// Print files from each node
for node in `cat $PBS_NODEFILE | sort | uniq`; do
echo "$node:"
rsh $node "ls -l $PBSTMPDIR"
done
# Copy back results from master node's scratch directory
cp $PBSTMPDIR/* $PBS_O_WORKDIR/

```

For more details on optimization schemes, see [Optimization Schemes](#).

Be aware that every installation is different, and it is almost certain that the job.pbs file included with this scheme will not work on any cluster other than the Zurek group's "parity" cluster at SUNY Buffalo's Center for Computational Resources. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct PBS script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

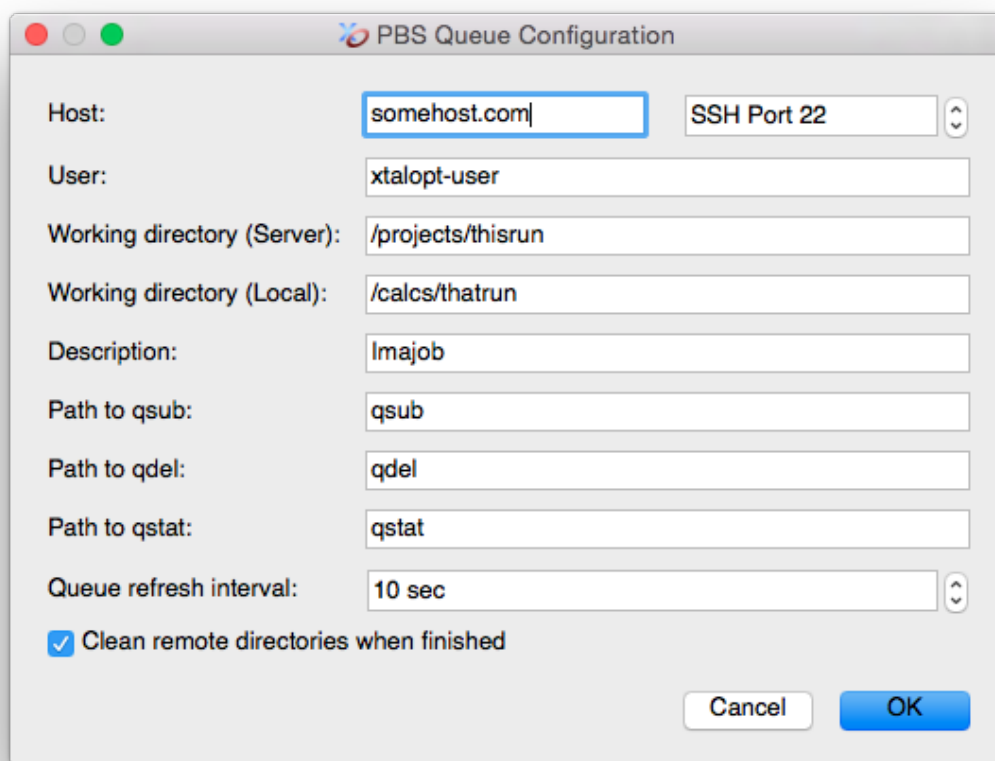
A handy trick for monitoring jobs outside of XtalOpt is to include the following line in job.pbs:

```
#PBS -N %description%-%gen%x%id%-%optstep%
```

This will name each job, for example, xtalSearch-3x4-2, where xtalSearch is a user-specified description of the search, and 3x4-2 means that it is the fourth structure in the third generation running its second optimization step.

[Skip to next section.](#)

1.5.2 Using a remote SGE cluster



The image shows a 'PBS Queue Configuration' dialog box. It contains several input fields and a checkbox. The fields are: Host (somehost.com), SSH Port (22), User (xtalopt-user), Working directory (Server) (/projects/thisrun), Working directory (Local) (/calcs/thatrun), Description (Imajob), Path to qsub (qsub), Path to qdel (qdel), Path to qstat (qstat), and Queue refresh interval (10 sec). There is a checkbox labeled 'Clean remote directories when finished' which is checked. At the bottom right are 'Cancel' and 'OK' buttons.

Host:	somehost.com	SSH Port	22
User:	xtalopt-user		
Working directory (Server):	/projects/thisrun		
Working directory (Local):	/calcs/thatrun		
Description:	Imajob		
Path to qsub:	qsub		
Path to qdel:	qdel		
Path to qstat:	qstat		
Queue refresh interval:	10 sec		
<input checked="" type="checkbox"/> Clean remote directories when finished			
		Cancel	OK

Select "SGE" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

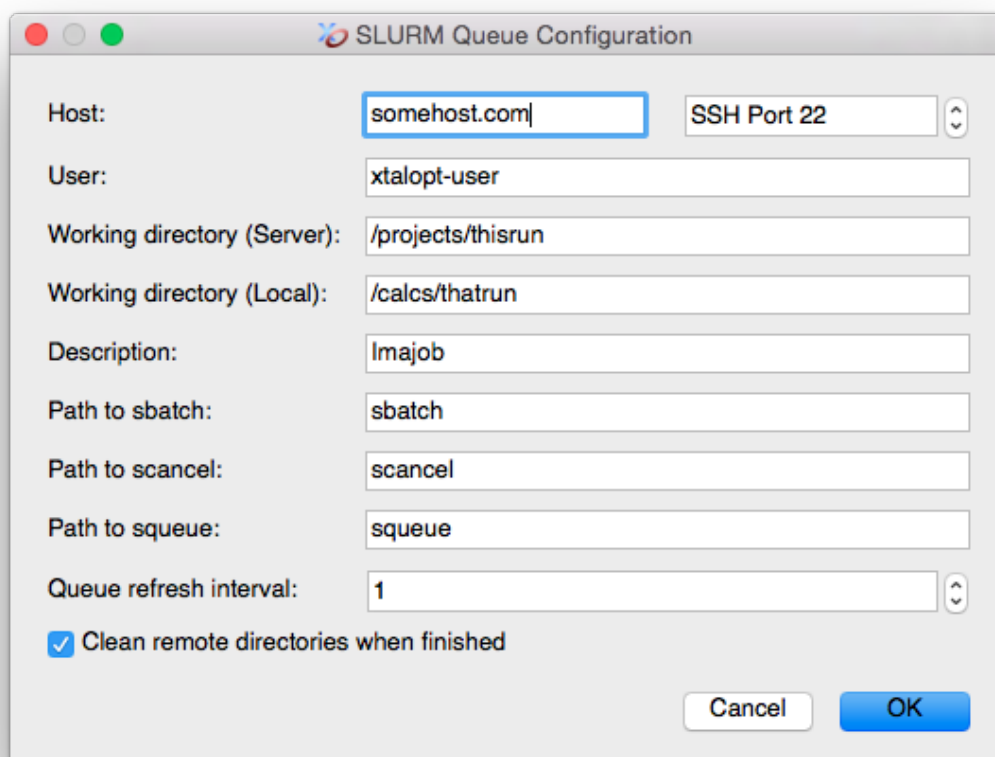
- host: The hostname of the SGE cluster's head node
- user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- Description: Used for the %description% keyword in input templates.
- Path to qsub: Where to find the qsub executable on the remote cluster. Note that if qsub is in the cluster's \$PATH, setting this to just 'qsub' will work.
- Path to qdel: Where to find the qdel executable on the remote cluster. Note that if qdel is in the cluster's \$PATH, setting this to just 'qdel' will work.
- Path to qstat: Where to find the qstat executable on the remote cluster. Note that if qstat is in the cluster's \$PATH, setting this to just 'qstat' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.sge" is added to the list of available templates. This is the job submission script for SGE. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct SGE script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see [Optimization Schemes](#).

[Skip to next section.](#)

1.5.3 Using a remote SLURM cluster



The image shows a macOS-style dialog box titled "SLURM Queue Configuration". It contains several input fields and a checkbox. The fields are: Host (with a blue selection box around "somehost.com"), SSH Port (set to 22), User (set to "xtalopt-user"), Working directory (Server) (set to "/projects/thisrun"), Working directory (Local) (set to "/calcs/thatrun"), Description (set to "Imajob"), Path to sbatch (set to "sbatch"), Path to scancel (set to "scancel"), Path to squeue (set to "squeue"), and Queue refresh interval (set to 1). At the bottom, there is a checked checkbox labeled "Clean remote directories when finished". Two buttons, "Cancel" and "OK", are at the bottom right.

Select "SLURM" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- host: The hostname of the SGE cluster's head node
- user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- Description: Used for the %description% keyword in input templates.
- Path to sbatch: Where to find the sbatch executable on the remote cluster. Note that if sbatch is in the cluster's \$PATH, setting this to just 'sbatch' will work.

- Path to scancel: Where to find the scancel executable on the remote cluster. Note that if scancel is in the cluster's \$PATH, setting this to just 'scancel' will work.
- Path to squeue: Where to find the squeue executable on the remote cluster. Note that if squeue is in the cluster's \$PATH, setting this to just 'squeue' will work.
- "Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.slurm" is added to the list of available templates. This is the job submission script for SLURM. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct SLURM script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see [Optimization Schemes](#).

[Skip to next section.](#)

1.5.4 Using a remote LSF cluster

Select "LSF" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- host: The hostname of the LSF cluster's head node
- user: The username used to log into the cluster
- Working directory (Server): A directory that is readable/writable by "user" on the cluster, used when performing optimizations.

- Working directory (Local): A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
 - Description: Used for the description% keyword in input templates.
 - Path to bsub: Where to find the bsub executable on the remote cluster. Note that if bsub is in the cluster's \$PATH, setting this to just 'bsub' will work.
 - Path to bkill: Where to find the bkill executable on the remote cluster. Note that if bkill is in the cluster's \$PATH, setting this to just 'bkill' will work.
 - Path to bjobs: Where to find the bjobs executable on the remote cluster. Note that if bjobs is in the cluster's \$PATH, setting this to just 'bjobs' will work.

"Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.lsf" is added to the list of available templates. This is the job submission script for LSF. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct LSF script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see [Optimization Schemes](#).

[Skip to next section.](#)

1.5.5 Using a remote LoadLeveler cluster

LoadLeveler Queue Configuration

Host: somehost.com SSH Port 22

User: xtalopt-user

Working directory (Server): /projects/thisrun

Working directory (Local): /calcs/thatrun

Description: Imajob

Path to lsubmit: lsubmit

Path to llcancel: llcancel

Path to llq: llq

Queue refresh interval: 20 sec

☐ Clean remote directories when finished

Cancel OK

Select "LoadLeveler" from the list of Queues, and then click the "Configure..." button. A new window will prompt for:

- **host:** The hostname of the LoadLeveler cluster's head node
- **user:** The username used to log into the cluster
- **Working directory (Server):** A directory that is readable/writable by "user" on the cluster, used when performing optimizations.
- **Working directory (Local):** A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.
- **Description:** Used for the description% keyword in input templates.
- **Path to llsbmit:** Where to find the llsbmit executable on the remote cluster. Note that if llsbmit is in the cluster's \$PATH, setting this to just 'llsbmit' will work.
- **Path to llcancel:** Where to find the llcancel executable on the remote cluster. Note that if llcancel is in the cluster's \$PATH, setting this to just 'llcancel' will work.
- **Path to llq:** Where to find the llq executable on the remote cluster. Note that if llq is in the cluster's \$PATH, setting this to just 'llq' will work.

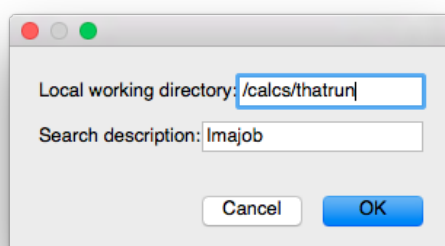
"Clean remote directories when finished": will remove all of the generated files from the cluster. Only the files on the local computer will be kept. If you do not want this to occur, make sure to uncheck this option.

A new template, "job.ll" is added to the list of available templates. This is the job submission script for LoadLeveler. It may take some experimentation to get jobs to submit successfully, and you may need to contact the system administrators of the cluster for assistance or information about MPI, executable locations, etc. Perhaps the easiest method to find the correct LoadLeveler script is to run some trial submissions by hand, and then replace the structure/search specific information with the appropriate keywords once a working script has been generated.

For more details on optimization schemes, see [Optimization Schemes](#).

[Skip to next section.](#)

1.5.6 Running optimizations locally



Select "Local" from the list of Queues, and then click the configure button. A new window will prompt for:

- **Local working directory:** A directory that is readable/writable by the current user on the local computer. This is where the final structures and resume files are written.

If the optimizer's executable (vasp, gulp, pw.x, castep, etc) is not in your system path, you will need to specify the location of the executable by clicking the "Configure..." button next to the optimizer selection menu.

For more details on optimization schemes, see [Optimization Schemes](#).

1.6 What is written to the local directory?

A directory for each structure is created at

```
[Local working directory]/<gen#>x<id#>
```

that will contain input, output, and data files specific to each structure. Two additional files are also written to the local filesystem:

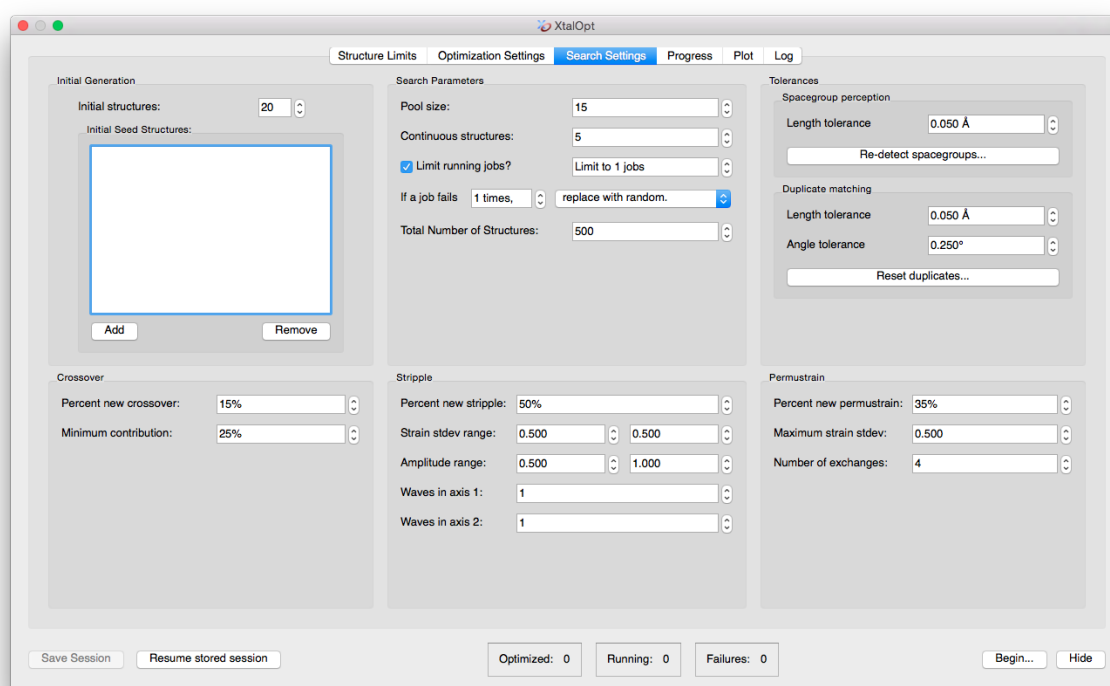
```
[Local working directory]/xtalopt.state
```

which contains save/resume information to continue a session that has been stopped, and

```
[Local working directory]/results.txt
```

which stores a list of all structures sorted by increasing enthalpy. The latter file is handy for offline analysis, since there is no need to open XtalOpt to find the most stable structures of a previous search.

1.7 Search Settings



In the "Search Settings" tab, most of the default settings should suffice (See CPC, 2011, 182, 372-387). We arbitrarily set the initial structures to 20 and the continuous structures to 5, although these may need to be adjusted based on available resources. We will not specify initial seeds, but the option to do so exists on this screen.

It is not necessary to limit the number of running jobs unless running locally, as the PBS queue on the cluster will manage job control for us. If running locally, set the job limit no higher than [number of available processor cores] - 1 (e.g. for a quadcore processor, allow three jobs to run simultaneously). This allows one core to remain free for the system to run.

There is now a termination criteria called "Total Number of Structures" that will end the run once a certain number of structures have been produced by XtalOpt.

1.8 "Begin"

Gen	Mol	Job ID	Status	Time Elapsed	Enthalpy (eV)	Volume	Space Group	Ancestry
1	1	N/A	Waiting for Optimization (1 of 1)	0:00:00	N/A	281.59	1: P1	Randomly generated
2	1	2	Waiting for Optimization (1 of 1)	0:00:00	N/A	324.85	1: P1	Randomly generated
3	1	3	Waiting for Optimization (1 of 1)	0:00:00	N/A	436.01	1: P1	Randomly generated
4	1	4	Waiting for Optimization (1 of 1)	0:00:00	N/A	384.94	1: P1	Randomly generated
5	1	5	Waiting for Optimization (1 of 1)	0:00:00	N/A	83.60	1: P1	Randomly generated
6	1	6	Waiting for Optimization (1 of 1)	0:00:00	N/A	76.25	1: P1	Randomly generated
7	1	7	Waiting for Optimization (1 of 1)	0:00:00	N/A	325.54	1: P1	Randomly generated
8	1	8	Waiting for Optimization (1 of 1)	0:00:00	N/A	298.76	1: P1	Randomly generated
9	1	9	Waiting for Optimization (1 of 1)	0:00:00	N/A	226.71	1: P1	Randomly generated
10	1	10	Waiting for Optimization (1 of 1)	0:00:00	N/A	359.45	1: P1	Randomly generated
11	1	11	Waiting for Optimization (1 of 1)	0:00:00	N/A	246.84	1: P1	Randomly generated
12	1	12	Waiting for Optimization (1 of 1)	0:00:00	N/A	245.57	1: P1	Randomly generated
13	1	13	Waiting for Optimization (1 of 1)	0:00:00	N/A	238.33	1: P1	Randomly generated
14	1	14	Waiting for Optimization (1 of 1)	0:00:00	N/A	492.63	1: P1	Randomly generated
15	1	15	Waiting for Optimization (1 of 1)	0:00:00	N/A	437.93	1: P1	Randomly generated
16	1	16	Waiting for Optimization (1 of 1)	0:00:00	N/A	156.03	1: P1	Randomly generated
17	1	17	Waiting for Optimization (1 of 1)	0:00:00	N/A	493.89	1: P1	Randomly generated
18	1	18	Waiting for Optimization (1 of 1)	0:00:00	N/A	193.47	1: P1	Randomly generated

Figure 1: The "Progress" tab immediately after starting a search

XtalOpt has everything it needs to start its search at this point; click the "Begin" button in the lower right corner of the application to tell it to start the search algorithm. A progress bar appears as the random first generation is created. Switch to the "Progress" tab and 20 entries will appear, all with a status of "Waiting for Optimization". Click "Refresh" on this tab to begin the local optimizations. From here, XtalOpt will continue to run without user input, starting new optimizations and generating new structures until it is stopped by the user.

1.9 Monitor progress

Gen	Mol	Job ID	Status	Time Elapsed	Enthalpy (eV)	Volume	Space Group	Ancestry
20	1	20	Optimized	0:00:16	-230.119	286.73	1: P1	Randomly generated
21	2	1	Optimized	0:00:06	-238.802	182.20	136: P4_2/mnm	Stripple: 1x12 stdev=0.50000 amp=0.02264 waves=1,1
22	2	2	Duplicate of 1x2	0:00:04	-236.612	184.22	6: Pm	Stripple: 1x2 stdev=0.50000 amp=0.11003 waves=1,1
23	2	3	Optimized	0:00:11	-234.285	205.06	1: P1	Stripple: 1x17 stdev=0.22894 amp=0.53867 waves=1,1
24	2	4	Optimized	0:00:07	-235.109	174.88	6: Pm	Stripple: 1x12 stdev=0.18816 amp=0.97071 waves=1,1
25	3	1	Optimized	0:00:08	-233.592	185.11	4: P2_1	Stripple: 2x1 stdev=0.35948 amp=0.60497 waves=1,1
26	2	5	Duplicate of 1x12	0:00:20	-238.802	182.20	136: P4_2/mnm	Permustrain: 1x1 stdev=0.16277 exch=4
27	3	2	Optimized	0:00:26	-235.758	190.59	5: C2	Stripple: 2x1 stdev=0.50000 amp=0.49720 waves=1,1
28	3	3	Duplicate of 2x1	0:00:12	-238.802	182.20	136: P4_2/mnm	Stripple: 2x1 stdev=0.50000 amp=0.18255 waves=1,1
29	2	6	Optimized	0:00:04	-238.802	182.20	136: P4_2/mnm	Stripple: 1x12 stdev=0.12458 amp=0.75044 waves=1,1
30	2	7	Optimized	0:00:07	-234.446	182.16	6: Pm	Crossover: 1x2 (60%) + 1x19 (40%)
31	2	8	Optimized	0:00:14	-235.327	166.53	225: Fm-3m	Stripple: 1x12 stdev=0.41797 amp=0.51789 waves=1,1
32	4	1	Optimized	0:00:09	-238.59	180.75	14: P2_1/c	Stripple: 3x3 stdev=0.12166 amp=0.59499 waves=1,1
33	2	9	Optimized	0:00:14	-235.173	201.34	1: P1	Crossover: 1x19 (48%) + 1x4 (51%)
34	3	4	Running (Opt Step 1 of 1, 0 failures)	0:00:25	N/A	182.20	1: P1	Permustrain: 2x6 stdev=0.38651 exch=4
35	4	2	Checking status...	0:00:18	-232.786	214.89	1: P1	Stripple: 3x2 stdev=0.44728 amp=0.82453 waves=1,1
36	2	10	Waiting for Optimization (1 of 1)	0:00:00	N/A	185.81	1: P1	Permustrain: 1x19 stdev=0.17759 exch=4
37	5	1	Waiting for Optimization (1 of 1)	0:00:00	N/A	180.75	1: P1	Permustrain: 4x1 stdev=0.03405 exch=4

Figure 2: The "Progress" tab mid-run

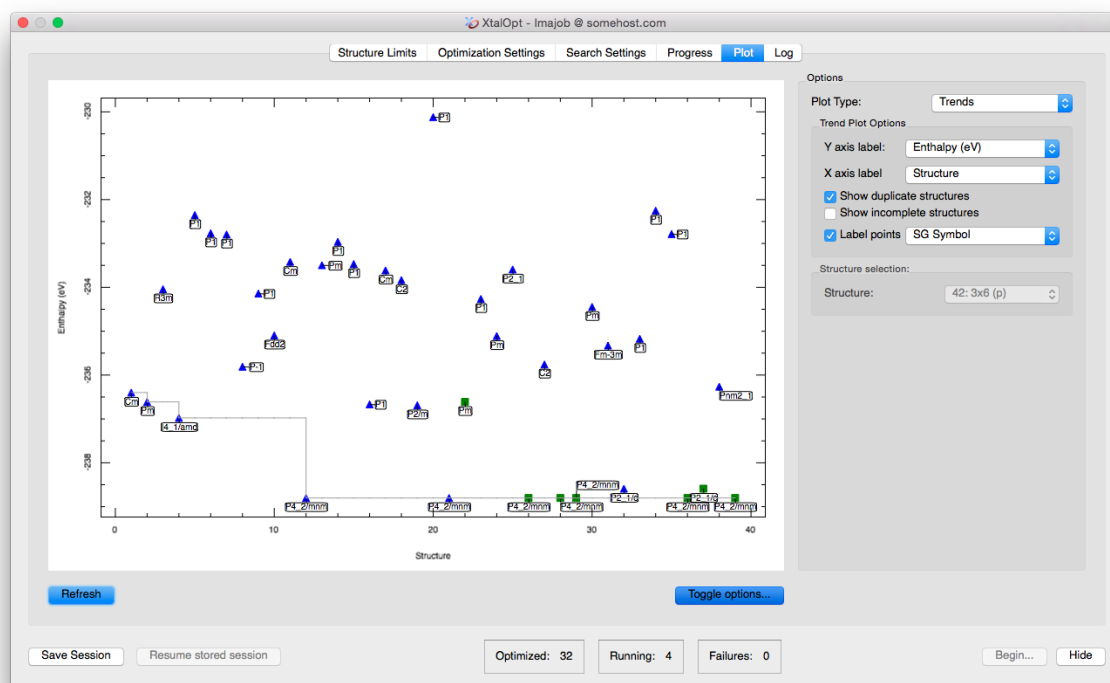
As XtalOpt performs the search, the progress table continuously updates, providing information about each structure. We see individuals in various stages of completion: most are optimized (in blue), structure 2x72 has been automatically marked as a duplicate (dark green) of structure 1x2 and removed from the breeding pool, structure 3x4 is currently undergoing a local optimization (light green), while structure 2x10 is waiting to be optimized (light blue).

Other useful information is displayed about each structure, such as the time spent in optimization, the optimized enthalpy, the cell volume, spacegroup, and each structure's ancestry (i.e. parent(s) and parameters for the genetic operator that generated it). A status bar on the bottom of the window shows the number of structures that are optimized, running, and failing at any given time. This information is visible regardless of which tab is currently being viewed.

An additional feature of the progress table is the ability to immediately visualize any of the individuals in the Avogadro main window – simply clicking on a row in this table will display the three-dimensional structure in Avogadro, where it can be visualized, modified, or exported. If the user would like to add a bit of "intelligent design" to the evolutionary process, a structure can be modified and then resubmitted using a context (right-click) menu from the progress table. The context menu provides tools to (un)kill a structure, resubmit for local optimization at an arbitrary optimization step, or replace a problematic structure with a new, random individual.

Three additional buttons found near the "Refresh All" button in this tab are also available. The "Print Results File" button generates a run-results.txt file that lists all of the information about each structure in order of generation and structure number (As compared to the results.txt file which ranks the structures). The "Remove Extra Files" button is used for VASP runs. It removes any extraneous files in each local subdirectory in order to reduce disk usage. Files kept are structure.state, POTCAR, CONTCAR, OSZICAR, job.slurm and OUTCAR. Finally, the "Rank All" button ranks all currently optimized structures and exports them to a subdirectory (Ranked) in two forms (depending on the optimizer): POSCAR/.got and .cif. Each can be found in separate directories. (Only works for GULP and VASP runs currently).

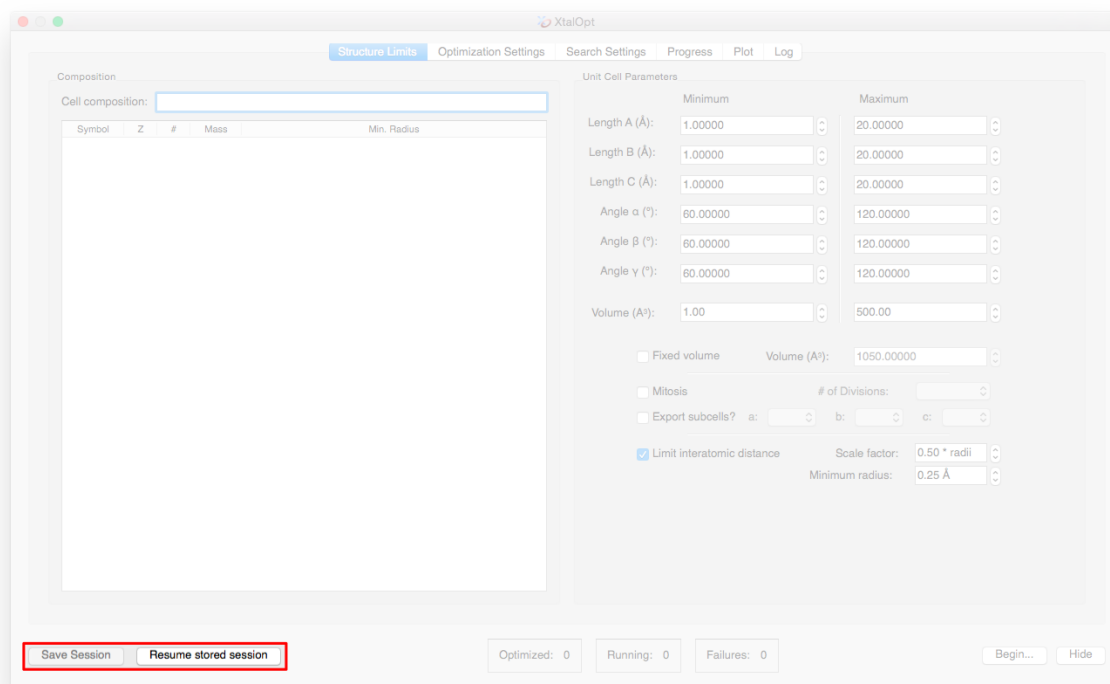
1.9.1 View trends



2 Saving and Resuming Sessions in XtalOpt

2.1 Contents

- [How to save your session](#)
- [How to resume your session](#)



2.2 How to save your session

XtalOpt will write a small file named `xtalopt.state` to its working directory that contains the information necessary to resume the session at a later time. The file can be rewritten manually by clicking the "Save Session" button highlighted above, and XtalOpt will automatically save the session every time a structure is updated.

XtalOpt will also write a file "structure.state" in each candidate structure's directory. This file stores XtalOpt-specific information about the structure.

2.3 How to resume your session

To resume a session, simply click "Resume stored session" (highlighted above) and select the `xtalopt.state` file in the working directory of the session you would like to resume. XtalOpt will then begin to load the structures and search parameters. You can monitor the progress with the progress bar that appears at the bottom of the window.

While the structures are loading, you may encounter errors that say:

```
Error, no (or not appropriate for [OPTIMIZER]) xtal data in [DIRECTORY].
```

This could be a result of resuming a structure that has not yet done any local optimizations. If so, safely ignore this message.

As mentioned in the message, these can typically be ignored if it only happens for a handful of structures. This occurs when a structure has been generated in XtalOpt, but it has not completed any geometry optimization so

there are no output files from which to load the geometry. If it happens for a significant number of structures (or structures that are known to have completed at least one geometry optimization step), the output files from the optimizer may be missing or corrupt.

After resuming a session, XtalOpt will ask if you would like to continue the search or enter read-only mode. Read-only mode will not generate new structures or submit geometry optimizations.

Note

If you are considering resuming a read-only session, take a look at the results.txt file in the working directory. It contains a list of all structures, sorted by enthalpy, with additional useful information. This can save some time when trying to locate the most stable structure of an old search.

The working directories for XtalOpt are relocatable, meaning that the directory containing xtalo.state and the [gen]x[id] structure folders may be moved, tarred, zipped, etc. and still be resumed at a later time from a different location on the filesystem, or even a different computer.

3 Optimization Schemes

3.1 Contents

- [Overview: What are optimization schemes, and why use them?](#)
 - [In a nutshell..](#)
 - [More details](#)
- [Optimization scheme user interface](#)
 - [Optimization step list](#)
 - [Add new optimization step](#)
 - [Remove current optimization step](#)
 - [Select template](#)
 - [Template editor](#)
 - [Save scheme](#)
 - [Resume scheme](#)
- [How to build an optimization scheme?](#)
- [How to save an optimization scheme for later?](#)
- [How to load an optimization scheme?](#)
- [What is saved?](#)
- [Suggestions for optimization schemes](#)
 - [Crystals \(XtalOpt\)](#)

3.2 Overview: What are optimization schemes, and why use them?

3.2.1 In a nutshell..

An optimization scheme is a series of optimization steps ("optsteps") that are to be performed in sequence on a structure. Each optimization step consists of a set of input file templates for the queuing system and optimizer to be used, and the structure is updated after each completes. So if an optimization scheme contains three optimization steps, a structure's lifecycle is:

1. Generation of initial structure
2. Perform optstep 1 on initial structure
3. Update structure from the results of optstep 1
4. Perform optstep 2 on current structure (result of optstep 1)
5. Update structure from the results of optstep 2
6. Perform optstep 3 on current structure (result of optstep 2)
7. Update structure from the results of optstep 3
8. Current structure (result of optstep 3) is either accepted into the breeding pool or discarded, depending on its enthalpy relative to the other optimized structures.

3.2.2 More details

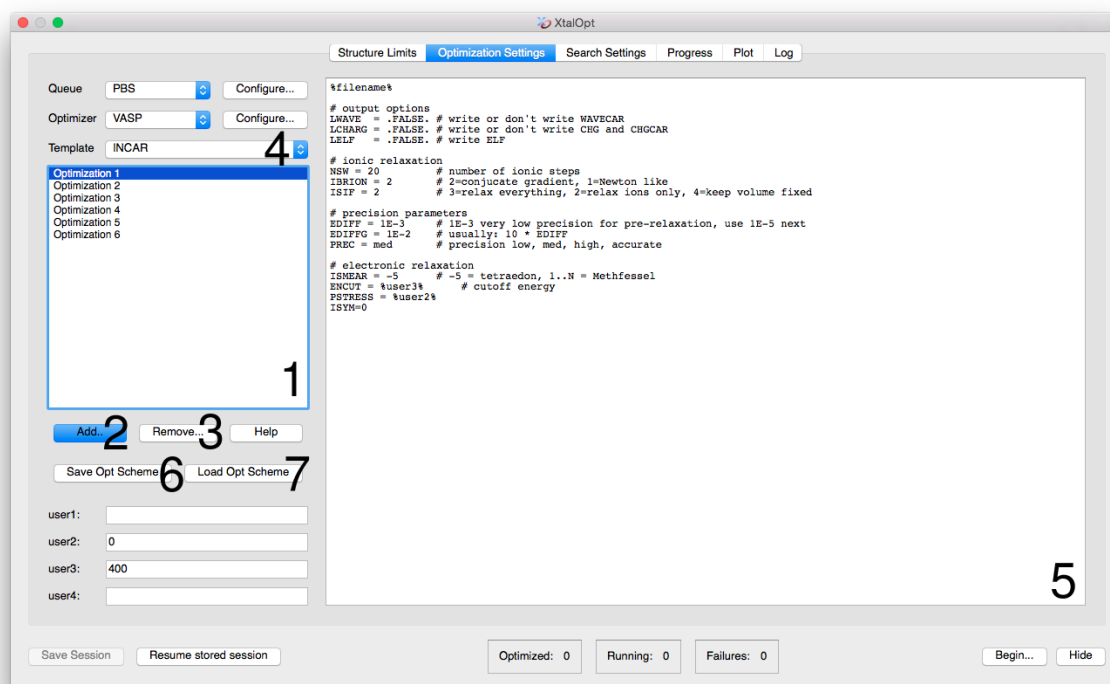
The efficiency of searching a potential energy surface for a global minimum can be significantly improved by moving each candidate structure to the nearest local minimum, i.e. performing a geometry optimization. The differences between searching with and without carrying out these local optimizations are explored in detail in Woodley SM, Catlow CRA. *Comp. Mat. Sci.* 2009;45(1):84-95 (Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0927025608003030>).

Why not just perform a single geometry optimization on each structure? Stochastic search techniques, such as XtalOpt, will often need to perform geometry optimizations on structures that are far from a stationary point on the potential energy surface. For example, the randomly generated structures in the first generation of an evolutionary search are often highly disordered with unrealistic atomic separations. If these structures were to be optimized in a single step with accurately small convergence criteria, it would be quite expensive. Also, it is more than likely that most of the optimizations would not finish successfully before reaching the maximum number of geometry steps allowed by the optimizer or specified in the input. A second issue is that complex structures (periodic crystals, for example) often have so many degrees of freedom that convergence in a single step is difficult from a poor starting point (consider the effect on atomic coordinates when a unit cell's translation vector is modified).

The first problem (effectively optimizing to small convergence) can be solved by implementing an optimization scheme that optimizes to successively smaller convergence cutoffs.

The second problem can be addressed by reducing the degrees of freedom in the early optsteps and only optimizing everything once each component has individually converged to a reasonable parameterization. See [Suggestions for optimization schemes](#) for examples.

3.3 Optimization scheme user interface



We will use the above screenshot as we describe the process of creating, saving, and loading optimization schemes. The numbers indicate:

1. List of optimization steps
2. Button to add new optimization step
3. Button to remove current optimization step
4. Template selection menu
5. Template editor
6. Button to save current optimization scheme to file
7. Button to load optimization scheme from file

3.3.1 Optimization step list

This list shows the currently available optimization steps in the order that they will be performed. The optstep that is currently selected for editing is highlighted, and the editable optstep can be selected by clicking the appropriate entry.

3.3.2 Add new optimization step

Clicking this button will append a new optimization step to the optstep list. The new optstep's templates will be copies of the currently selected optstep's templates.

3.3.3 Remove current optimization step

Click this button to delete the currently selected optimization step.

3.3.4 Select template

This menu contains the filenames of the templates that are required by the currently selected queuing system (e.g. PBS, SGE, local...) and optimizer. The currently selected template is displayed in the template editor, and selecting a different template will update the editor.

3.3.5 Template editor

This text editor is used to view and edit the currently selected template for the current optstep.

3.3.6 Save scheme

This button will prompt for a location to save a .scheme file containing the current optimization step.

3.3.7 Resume scheme

This button will prompt for an existing .scheme file to load.

3.4 How to build an optimization scheme?

Creating a working scheme from scratch may take some time. We recommend checking the samples/ directory of the source code to obtain sample scheme for each optimizer (see [How to load an optimization scheme?](#)) and verifying that they are appropriate for the system under consideration before starting a search.

If there is not an appropriate sample, the following prescription may be used to generate your own:

1. Generate a random structure of the system under consideration. This may be done by hand, or by running a search just long enough to create the first random generation and saving one of the structures.
2. Create a starting optstep with the desired convergence criteria
3. Manually submit the optimization
4. If the optimization fails:
 - (a) First determine why – if the maximum iterations were exceeded or the optimization was aborted due to a badly performing minimizer, try one of the ideas below. Other optimization problems are beyond the scope of this document.
 - (b) Reduce the convergence criteria of the current trial optstep
 - (c) Remove degrees of freedom, e.g. by fixing cell parameters, atomic positions, etc
 - (d) Reduce the accuracy of the calculation in other ways (use a coarser integration grid, etc).
 - (e) Change the minimizer (e.g. tell the optimizer to use conjugate gradients rather than BFGS, etc)
5. Once the optimization succeeds, create another set of input files with the desired convergence criteria for all degrees of freedom.
6. Manually submit the new optimization step. If it fails, try the ideas above until it converges.
7. Once the structure has converged to the desired level of accuracy, try to optimize another randomly generated structure using the optsteps that succeeded previously. Refine them if needed.
8. Once you have successfully optimized enough random structures that you are confident in your method, gather all of the inputs used and write your scheme from them.

The scheme may be written by copying each input file into the template editor (with the appropriate optstep and template selected, of course) and replacing the structure-specific information with the appropriate keywords. Click the "Help" button for the complete list of keywords.

We have found that the optimization schemes are surprisingly transferable within an optimizer, so once you have a working optimization scheme for a given optimization code only minor tweaks (usually to the energy cutoffs, etc) are necessary to use it on a different chemical system.

3.5 How to save an optimization scheme for later?

Once you have written your optimization scheme, you will want to save it for fast retrieval later (otherwise you will need to copy/paste and edit all of the templates again!). To save, simply click the "Save Opt Scheme" button and enter an appropriate filename with an extension of .scheme.

3.6 How to load an optimization scheme?

Loading an optimization is quite simple – just click the "Load Opt Scheme" button and select the .scheme file you wish to load. This will also update the current queuing system and optimizer to those specified by the scheme.

3.7 What is saved?

The optimization scheme files contain more than just the templates for each optstep. They also store queue and optimizer specific settings. This is useful for storing configuration options for different clusters along with the scheme. Note that although XtalOpt will prompt for an SSH password if needed, it is **NOT** stored in the scheme file.

3.8 Suggestions for optimization schemes

3.8.1 Crystals (XtalOpt)

The following list describes the optimization steps used in the samples/vasp-xtalopt.scheme file distributed with the XtalOpt source code:

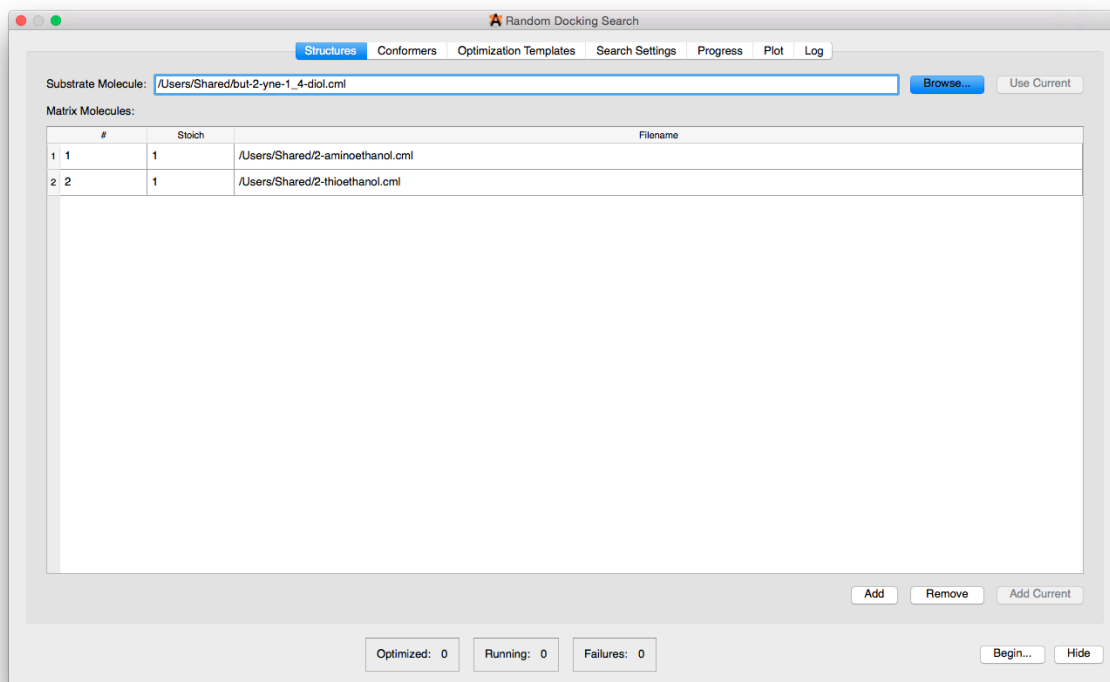
1. Fix unit cell, only optimize atomic coordinates. A very loose convergence criterion is used, and the number of KPOINTS is kept small.
2. The cell volume is fixed, but atomic positions and cell parameters are allowed to vary. The convergence criteria is the same as before, as is the KPOINT grid.
3. All degrees of freedom are considered using the same convergence criteria as before, but with a finer KPOINT grid.
4. Same as before, but with a stricter convergence criteria.
5. Same as before, but with a stricter convergence criteria and more KPOINTS.
6. Same as before, but with more KPOINTS.

This is only one of many possible optimization schemes that may work for crystals. It may need to be modified to work for your particular system.

4 RandomDock Tutorial

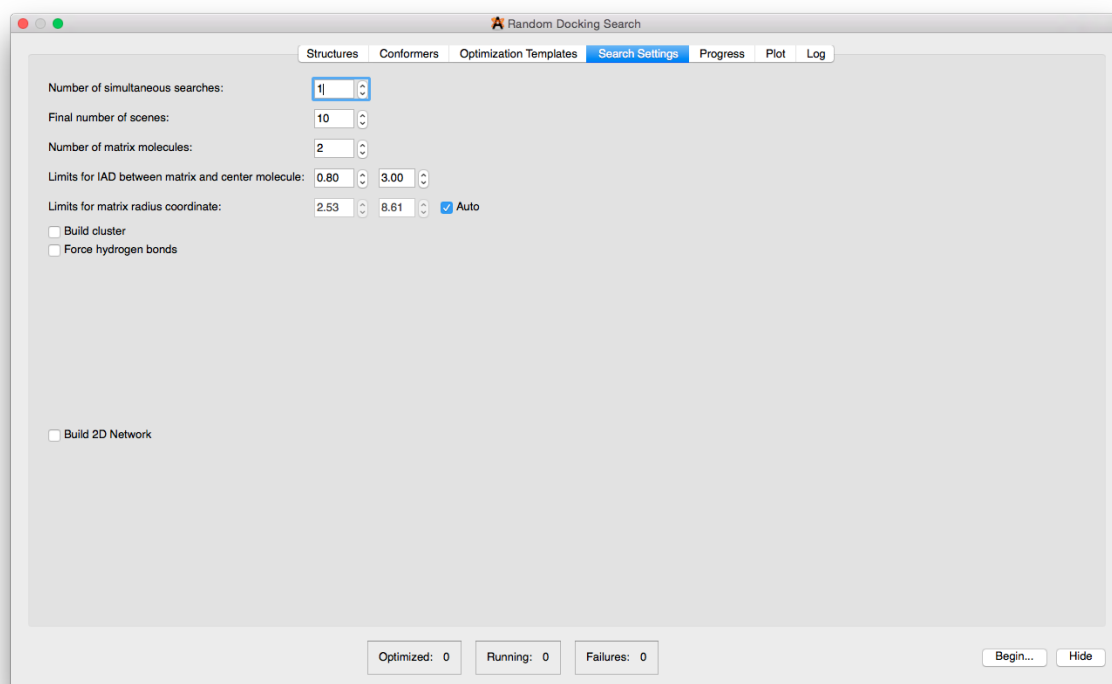
RandomDock generates random structures of ensembles of molecules by arranging them in 3D space subject to several constraints:

- One molecule is specified as the substrate
- One or more matrix molecules are added. These molecules will be used to form a "cage" around the substrate. Their stoichiometries may be adjusted to control the relative quantities of each.



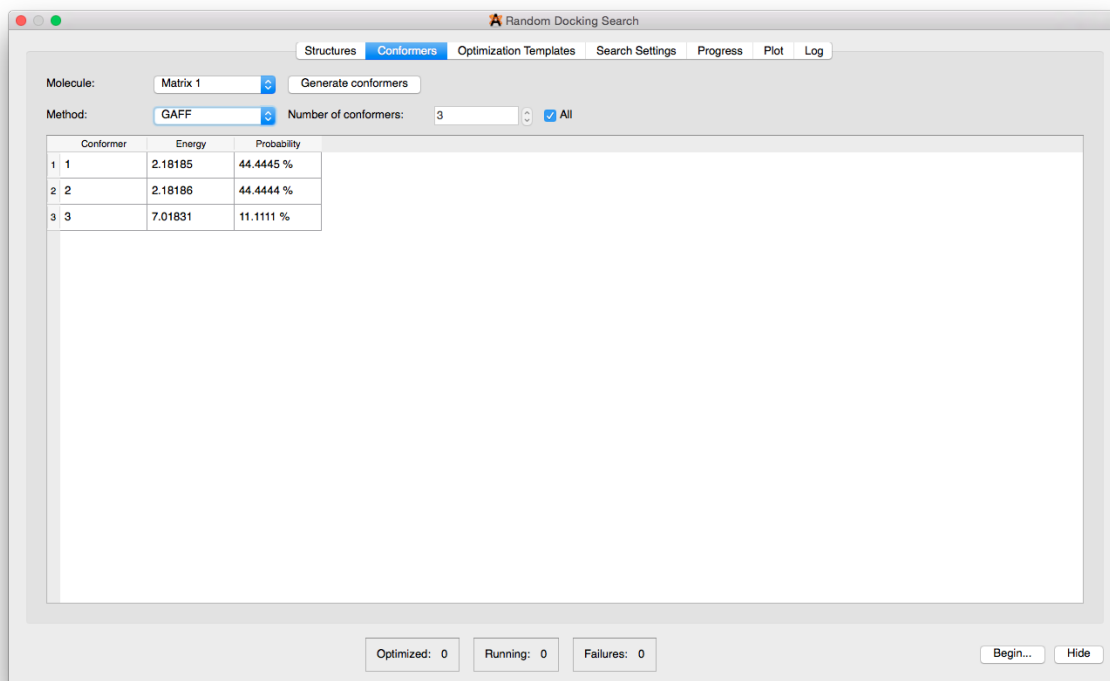
- The conformation and orientation of each matrix and substrate molecule is randomized in each scene.
- Matrix molecules are placed at random positions within a specified distance of the substrate's geometric center.
- One atom of each matrix molecule must be within a specified distance from any atom in the substrate.
- No two atoms from distinct molecules in the scene may be closer than a specified minimum distance.
- Optionally, it may be required that at least one possible hydrogen bond exists between each matrix molecule and any other molecule in the scene.

Several of the constraints are parametrized, and may be modified via the user interface shown below.



After the scene has been generated, the atomic positions are optimized using either ADF, GAMESS, GAUSSIAN or MOPAC. The calculations may be performed on a remote supercomputing cluster by specifying SSH login details, or locally on the user's workstation. Remote calculations are automatically staged, copied, submitted, monitored, retrieved, and analyzed by RandomDock, making searches fully automatic. Configuration of remote computational resources and optimizer input files is performed using the same procedure as XtalOpt.

The default behavior of RandomDock is to generate a single, possibly partial, shell of matrix molecules around the substrate, suitable for producing simple solvation models or studying binding interactions. Alternatively, it can be configured to generate molecular clusters. This mode produces clusters of molecules with randomized positions and orientations of molecules. A RandomDock search will continue until a specified number of scenes have been generated and optimized.



Before any scenes are generated, a cached set of conformations for each molecule is generated. Conformers are relaxed and ranked using the forcefield implementations in the OpenBabel library. As shown above, the user specifies a forcefield implementation and a number of conformers for each molecule. By checking the “All” option, an exhaustive search is performed for all possible conformations of the molecule. Each conformer is assigned a probability according to the probability distribution. This probability is used during the generation of new scenes to favor usage of stable conformers in both the substrate and matrix molecules.