

Structure Solving with ASDP

What Is ASDP

The Use of ASDP

<http://asdp.bnl.gov>

Jiansheng Jiang
Biology Department, BNL

Automated Structure Determination Platform

What Is ASDP?

- Needed for high-throughput structural genomics.
- Rapid structure solving at the synchrotron facilities.
- Enable users to solve structures remotely over Internet.
- Make structure determination as automated as possible.
- Improve phasing methods and capture meta data.
- Consist of crystallographic packages & web interfaces.
- Utilize databases and provide data-exchange tools.
- A “pipeline” from data collection to PDB submission.

ASDP: the Methods

Phasing Methods

MAD: Multi-wavelength Anomalous Dispersion

SAD: Single-wavelength Anomalous Dispersion

MIR: Multiple Isomorphous Replacement

SIR: (Single) Isomorphous Replacement

MR: Molecular Replacement

Direct Methods

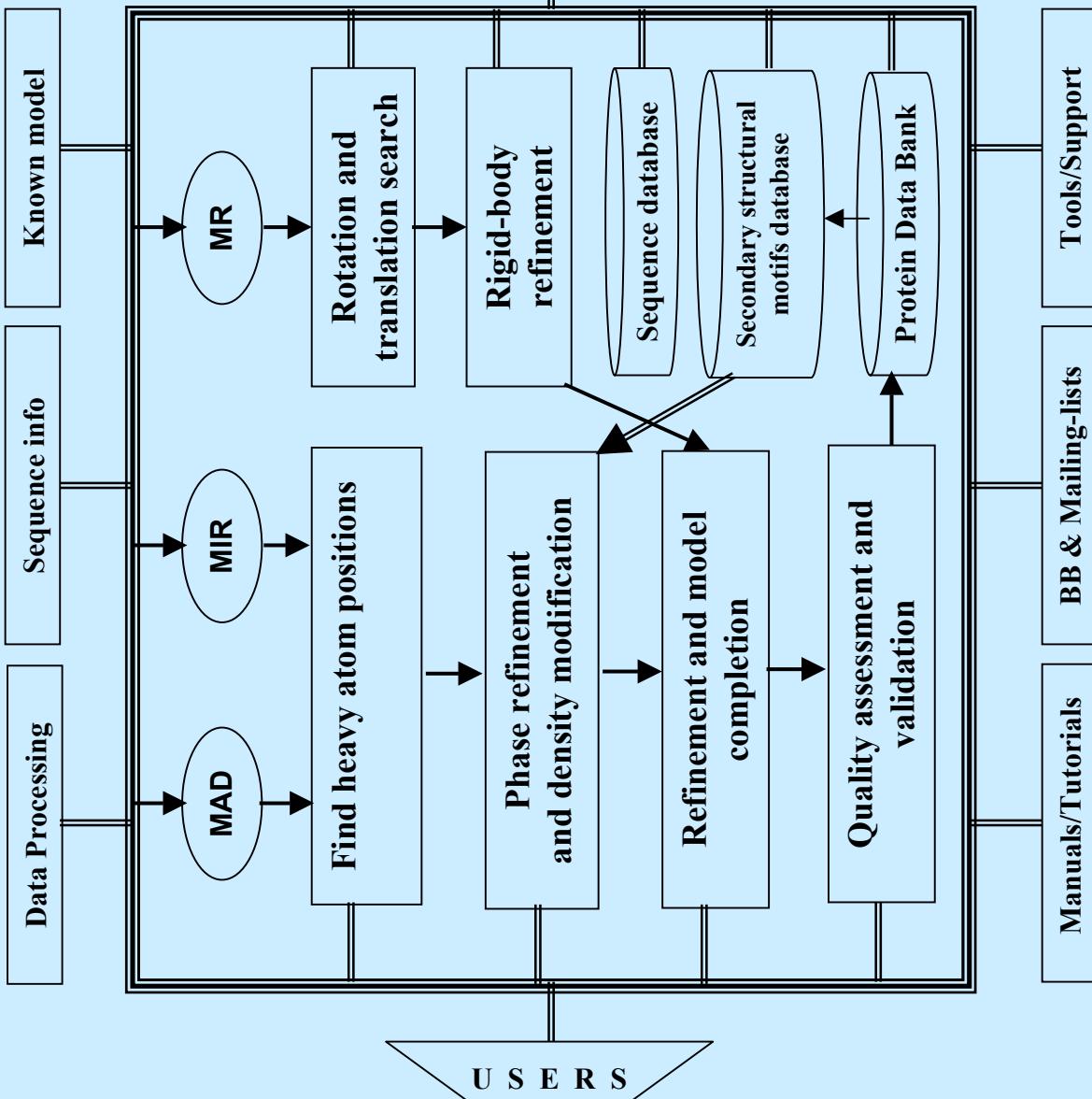
Combinations: **MIRAS, SIRAS, MADIR, MRMIR**

ASDP: Stages in Structure Solving

- Data collection and processing
- Finding heavy atom positions (MAD/SAD/MIR/SIR)
- Density modification and phase refinement
- Initial map interpreting and model building
- Structure refinement and model completion
- Quality assessment and submission

ASDP: the Platform

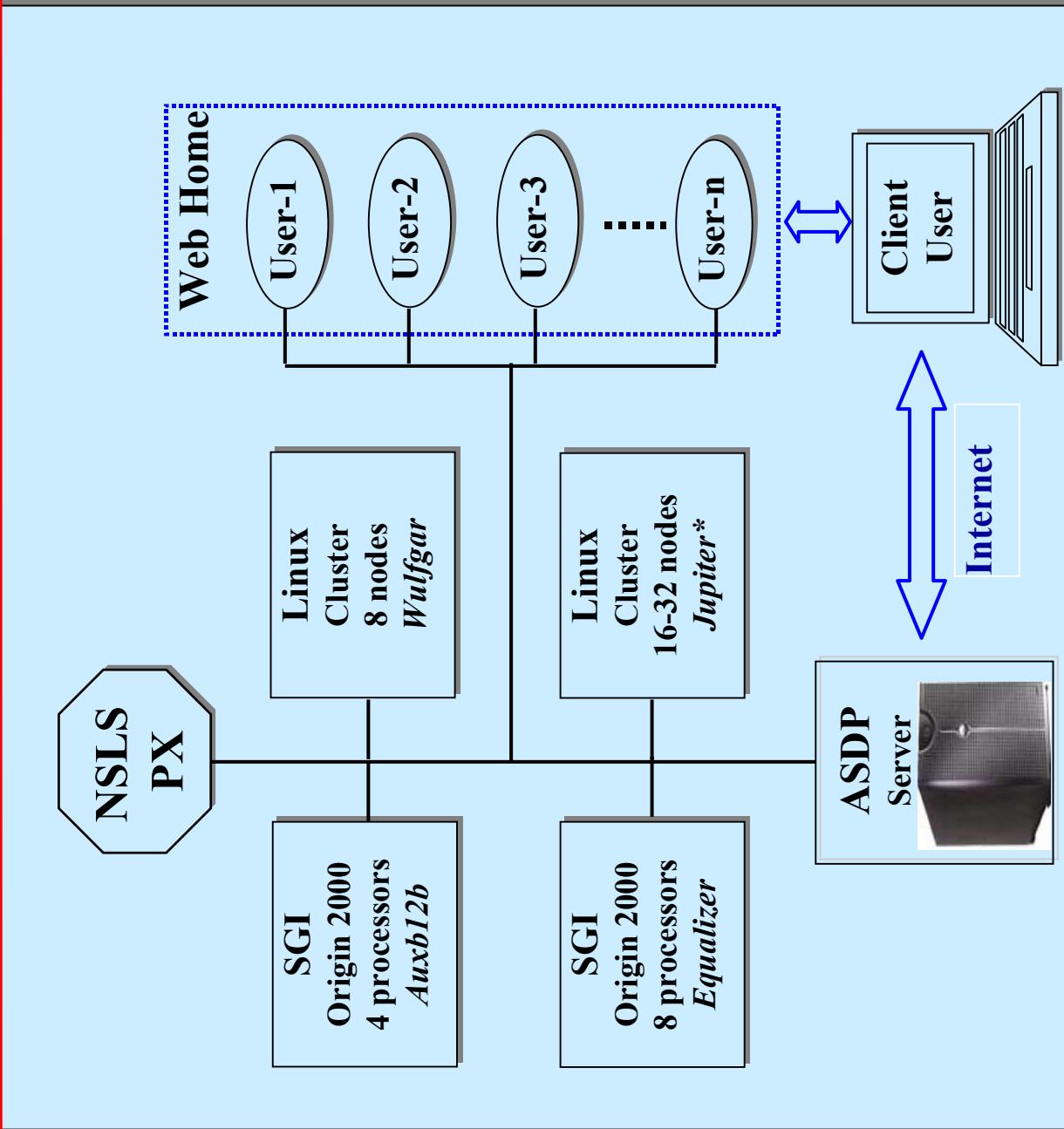
→ Procedure flow
== Data exchange
===== Web interface



ASDP: the Performers

- Heavy atom (MAD/SAD/MIR/SIR/DMe)
 - ◆ SOLVE, CNS, SnB, SHELLX
- Phase refinement and density modification
 - ◆ DMo, SHARP, RESOLVE, CNS, MLPHARE, PHASES
- Molecular replacement
 - ◆ CNS, AMoRe, MERLOT, MolRep
- Model building (auto or not auto)
 - ◆ ARP/wARP, O, XtalView, SOLVE
- Refinement and completion
 - ◆ CNS, REFMAC/CCP4, SHELLX
- Analysis and validation
 - ◆ PROCHECK, WHATIF, SFCHECK

ASDP: the Server



ASDP: the Features

- Provide users a web home directory (**disk space**).
- Provide ample computing power.
- A login/password system.
- A unix-like operating system on the web.
- Job submissions from the web home.
- A project-oriented setup interface.
- Validation tools and PDB submission.
- User-friendly, easy-use, flexible and extendable.

ASDP: Home Directories

Index of /data/whome/rd2001/demo

Name	Action	Last Modified	Kb
Parent Directory		-	
ha.pdb	vedcan	Apr 22 01:35	-
mad_fbar.scl	edcan	Apr 20 23:4	
mad_fpfm.scl	edcan	Apr 20 22:4	
omac	edcan	Apr 22 02:0	
omap	edcan	Apr 22 02:0	
p2.pdb	vedcan	Apr 21 23:4	
p9_se.pdb	vedcan	Apr 21 23:4	
p9_se_w2.sca	edcan	Apr 20 22:4	
phases-hl.drg	edcan	Apr 20 23:4	
phases-hl.export	edcan	Apr 20 23:4	
phases-hl.script	fedcan	Apr 20 23:4	
post.resolve.csh	edcan	Apr 22 02:4	
post_solve.csh	edcan	Apr 22 02:4	
resolve.log	edcan	Apr 21 00:1	
resolve.mtz	vedcan	Apr 21 00:1	
resolve_all.map	edcan	Apr 22 02:4	
resolve_lc.map	edcan	Apr 22 02:4	
sad.logfile	edcan	Apr 20 23:4	

Netscape: LSDIR

File Edit View Go Communicator Help

File Upload Download View Edit Delete Copy Rename Save Open Get Set Run Help

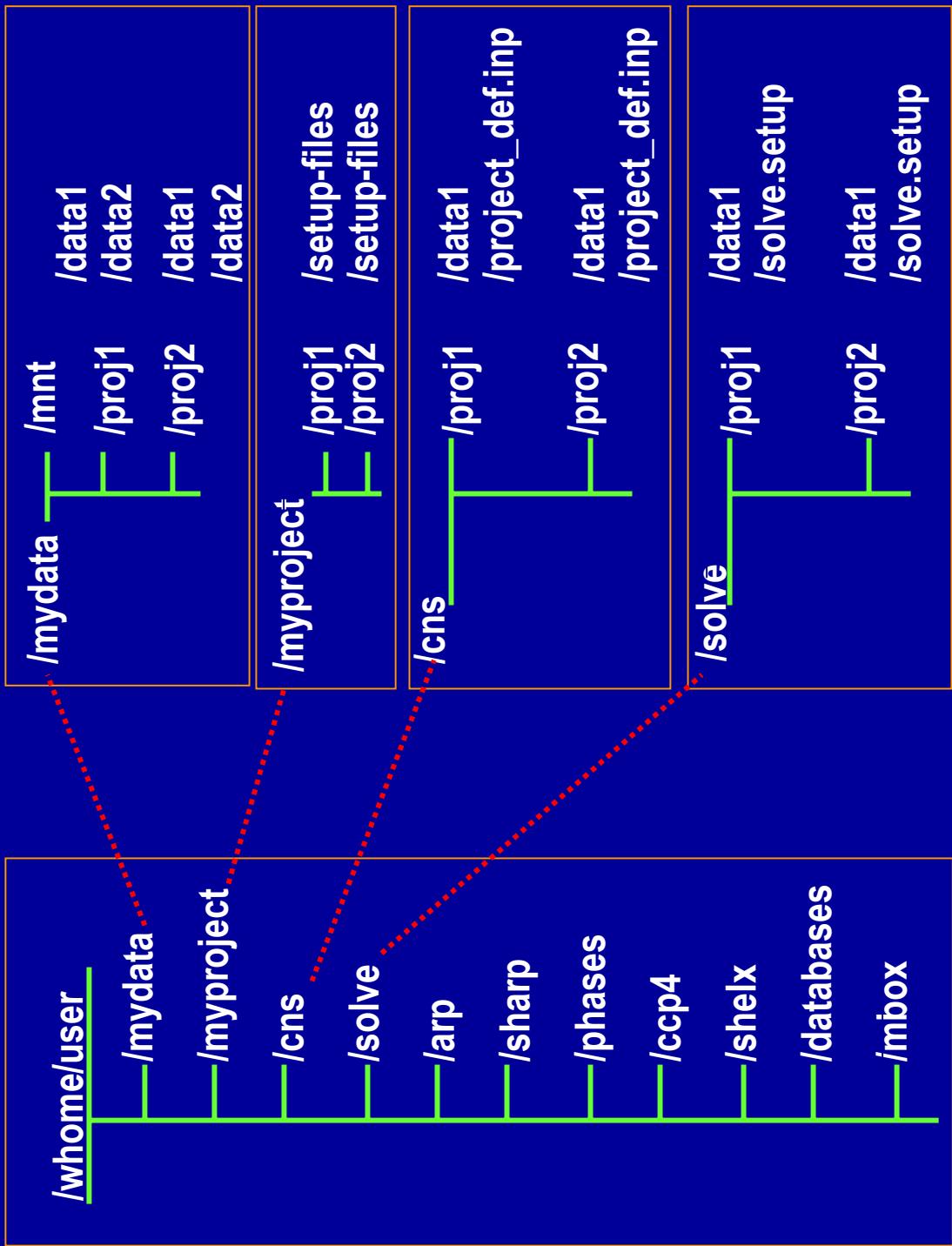
WHome Upload Download View Edit Delete Copy Rename Save Open Get Set Run Help

Index of /data/whome/rd2001/solve/p9_sad

javascipt:isFileOpen('/cgi-bin/sc/file_main.cgi?&at=MDL')

4/19/2002

ASDP: Web Home (whome)



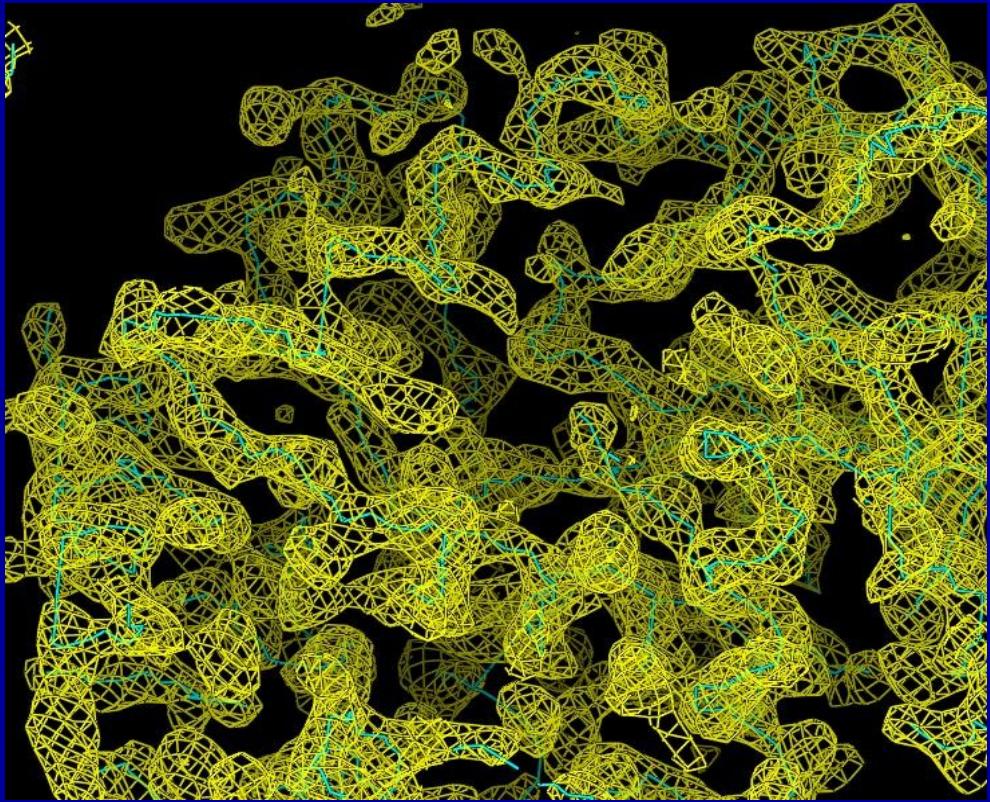
ASDP: File Type/Extension

- * .pdb (pdb)
- * .sca (scalepack)
- * .mtz (ccp4/arp/sharp)
- * .hkl (cns/snb)
- * .cif (iucr)
- * .sh, * .csh (shell/ccp4/arp)
- * .inp (cns)
- * .sin (sharp)
- * .com (ccp4/unix)
- * .script (solve)
- * .par (arp)
- * .d (phases)
- * .ins (shelx)

Example of ASDP: P097

Structure was solved within two hours right after data collection.

A YEAST PROTEIN IN SPX19-GCR2 INTERGENIC REGION



P097 Monomer

Example of ASDP: P097

DATA COLLECTION:

3 MAD data sets at
X12C/NSLS (20 hours).

MW: 27.5KD, 246 aa/mole
Unit Cell: 57.75 68.68 125.2
Space Group: P212121
Two moles in AU, 6 Se atoms.

STRUCTURE SOLVING

scripts ("recipe") were setup by using ASDP.

1. run SOLVE with "mad_solve.script", cut at 2.8A resolution.
found 6 Se atom sites with FOM=0.65 Z-score=41 (30 mins)
2. run RESOLVE, FOM=0.77, map clearly show helices (90 mins).
3. use DM for phase extension from 2.8A to 1.9A, FOM=0.86 (30 mins)
4. run ARP/WARP for chain tracing, cycles 30/200, connectivity=0.93,
and run "side_dock.sh" to map sequence and build side-chains.
about 85% of model were built automatically (15 hours).
5. run CNS refinement with no water R/R-free=0.26/0.29 (4 hours).

Toward Automation of Structure Determination - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://asdp.bnl.gov/

About ASDP wHome BnML FAQ Contact Search Utility Run Login Register Calendar
DOE/BNL Privacy and Security Notice

Register & Login

help

Register

User * rd2002a

Password *

Confirm *

Real Name RapiData2002 Group A

Group Protein Crystallography Users at NSLS/BNL

E-mail * x12cuser@x12wulfgar10.nsls.bnl.g*i*

Telephone

Postal Address

login help

User rd2002a

Password

logout change_password

login

submit update reset

*required info

Unix-like Commands

By clicking the single code in "Action" column on the line where the file is located, it will take the command instantly. Clicking the filename itself, it will show the file contents.

Command	Code	Action Description
Upload	u	upload a file from client machine to the server.
Download	n	download a file from the server to the client machine.
View	v	view a file in the plan text format (pagelized).
Edit	e	edit a file in the plan text format (pagelized).
Fedit	f	edit a file in the web form format (depends on file extension name).
Delete	d	delete a file or a directory.
Copy	c	copy a file to another.
Rename	a	rename a file to a new name.
Save	as	save a template (editing or viewing) file.
Open	o	open a file or a new file, or create a new directory.
Set	s	setup a project, a data file or location.
Get	g	get a file from a database.
Run	r	run a command, a job, or submit a job or a script file to the server.
Help	h	help - (instruction or documentation).

Index of /data/vhome/p097/solve/p97_85dl_mad/tl

Name	Last Modified	Action
Parent Directory	Jul 20 00:21	
ha.pdb	Jun 17 01:53	vedcan
mad_solve_script	Jun 17 01:09	fedcanr
solve_log	Jun 17 01:53	edcan
solve_mtz	Jun 17 01:53	vedcan
solve_prt	Jun 17 01:53	edcan
solve_setup	Jun 17 00:33	fedcanr
solve_status	Jun 17 01:53	edcan

send to esdp-help@vivaldi.bio.bnl.gov for further helps.

Set functions

The screenshot shows the LSDIR software interface. On the left, a red box highlights the 'Set functions' section. This section contains a table with ten rows, each with a 'go' button and a descriptive text. The rows are:

- Set the default project p97_85d1_mad_3
- Set a new project ID and title, and its links
- Set a file system (nfs) to your data directory
- Set datafiles to be used for a project
- Set crystallographic parameters
- Set phasing methods and procedures
- Set heavy atom and molecule information
- Set sequence and structural information
- Set preferences (personal settings)
- Set a macro procedure for automation

On the right, a red box highlights the 'File Manager' window. It shows a list of files in a table format:

Action	Last Modified	Kb
	Jul 20 00:21	-
Y ed c an	Jun 17 01:53	0
f ed c an r	Jun 17 01:09	3
Y ed c an	Jun 17 01:53	302
Y ed c an	Jun 17 01:53	560
Y ed c an	Jun 17 01:53	35
f ed c an r	Jun 17 00:33	0
Y ed c an	Jun 17 01:53	29

The 'File Manager' window has a toolbar at the top with buttons for Home, Upload, Download, View, Edit, Delete, Copy, Rename, Save, Open, Get, Set, Run, and Help. The 'Set' button is circled in red.

Set project

Set functions

help

Set the default project `p97_85d1_mad_3`

Set a new project ID and title, and its links

Set a file system (nfs) to your data directory

Set datafiles to be used for a project

Set crystallographic parameters

~~Set phasing methods and procedures~~

Set project

Your project directory: /data/whome/p097/myproject

Define a new project ID:

997 8581 Mad

Title of this project:

Hypothetical Yeast Protein - P0997 - Crystal [85cl]

next
cancel
reset
ok

Show my project list

Title of this project: Hypothetical Yeast Protein - P0997 - Crystal [85d1]

Set nfs

help

Your mounted data directory: /data/whome/p097/mydata/mnt

Set nfs

Local and NLS file systems (nfs) are mounted on the ASDP server. (Ask your system administrator to help if your system is not here)

Select a mounted file system below and click "ok",
a symbolic link will be set into your mounted data directory.

```
lsx12n.nsls bnl.gov/x12c_hotswap1 ▶
```

lsx12n.nsls bnl.gov/x12c_hotswap1 is mounted
on /lsx12n/x12c_hotswap1.
You have set /lsx12n/x12c_hotswap1 as a link in your data directory
/data/whome/p097/mydata/mnt
You can access it now
/data/whome/p097/mydata/mnt/lsx12n/x12c_hotswap1

Current directory: /data/whome/p097

Set datafiles

Set datafile

help

Datafile to be installed for this project:

p97_47_mad_sg19

Datafile will be copied from the original (or a mounted) location to /data/whome/jiang1/mydata and under the project directory.

Original datafile location (full path):

/data/whome/jiang1/mydata/mnt/w07/data2/x12cuser

Datafile name: p97_47h_sg19.sca

Datafile type: anomalous

Data collected at: remote Wavelength: 0.9600

Data processed by: scalepack Resolution: 3.0

Extract crystallographic parameters? yes no

Filename	Type	At	Wavel'th	Processed	Resol
p97_47p_sg19.sca	anomalous	peak	0.97886	scalepack	3.0
p97_47i_sg19.sca	anomalous	inflection	0.97912	scalepack	3.0
p97_47h_sg19.sca	anomalous	remote	0.96000	scalepack	3.0

Set cryst. parameters

help

Set crystallographic parameters

Crystallographic parameters for project:

p97_47_mad_sg19

Click "go" if re-select a project

Extract from the datafile: **p97_47h_sg19.sca**

Click "go" to take parameters from the select datafile

Unit cell dimensions:

a:	57.702	b:	68.747	c:	125.537
alpha:	90.000	beta:	90.000	gamma:	90.000

Space group: **p212121**

Resolution limits: dmin: **3.0** dmax: **50.0**

Set method

help

Set phasing methods

Phasing method to be used for this project:

p97_47_mad_sg19



Click "go" if re-select a project



Phasing method: MAD combine with: null

Select the phasing method that you intend to use in structure solving. ASDP will recommend procedures to you and install the necessary script files for you.

Datafiles to be used: (click button in "y/n" column)

y/n	Filename	Type	At	Wavelength
<input checked="" type="checkbox"/>	p97_47p_sg19.sca	anomalous <input type="checkbox"/>	peak <input type="checkbox"/>	0.97886
<input checked="" type="checkbox"/>	p97_47i_sg19.sca	anomalous <input type="checkbox"/>	inflection <input type="checkbox"/>	0.97912
<input checked="" type="checkbox"/>	p97_47h_sg19.sca	anomalous <input type="checkbox"/>	remote <input type="checkbox"/>	0.96000
<input type="checkbox"/>		null <input type="checkbox"/>	null <input type="checkbox"/>	
<input type="checkbox"/>		null <input type="checkbox"/>	null <input type="checkbox"/>	

Set heavy atom & mole. info

Set heavy atom and molecule info

Heavy atom and molecule info for this project:

Click "go" if re-select a project

Experimental X-ray method (data collection):
Heavy atom information is required if MAD or MIR is selected.

Heavy atom (HA):

Associated with this datafile:

Expected number of heavy atoms per molecule:

Number of molecules in asymmetry unit (AU):

Datafile	HA	Sites/Mole	Sites/AU	Sites/UC*
p97_47p_sg19.sca	Se	3	6	24
p97_47i_sg19.sca	Se	3	6	24
p97_47h_sg19.sca	Se	3	6	24

* Unit Cell - space group is "P212121", number of symmetry operators is 4

Number of amino acids of the molecule:

Molecular weight (MW in Dalton):
Click "est" to estimate MW by using 110 Dalton per amino acid.

Solvent content (in fraction):
Click "est" to estimate solvent content.

Matthews coefficient:
Click "est" to estimate Matthews coefficient.

Possible oligomeric state:

Submit jobs

LSDIR - Microsoft Internet Explorer

File Edit View Favorites Tools Help

whome Upload Download View Edit Delete Copy Rename Save Open Get Set Run Help

Index of /data/whome/p097/solve/p97 85d1 mad/tl

Name	Action	Last Modified	Kb
Parent Directory		Jul 20 00:21	-
ha.pdb	yedcan	Jun 17 01:53	0
#mad_solve_script	fedcanf	Jun 17 01:09	3
solve.log	edcan	Jun 17 01:53	302
solve.mtz		Jun 17 01:52	500
solve.prt			
solve.setup			
solve.status			

Submit a job

Current directory: /data/whome/rd2001/solve/p9_sad

Please enter a command or submit a job.

sad_solve . script

Submit to: asdp.bnl.gov

Shortcut to javascript:fileOpen()

"nice +2 csh sad_solve.script > sad_solve.script_log21591 &"

Job 21591 is submitted to equalizer.bio.bnl.gov on April 22, 2001 at 06:07:53
You may monitor the log file now at [sad_solve.script_log21591](#)
You will be notified by email when this job is completed.

Current directory:
/data/whome/rd2001/solve/p9_sad

Upload & Download

Upload

C:\My Documents\yang\p097\1gdn.pdb

Download

Which file do you want to be downloaded?

From: /data/whome/rd2001/solve/p9_sad/to_o

/data/whome/rd2001/solve/p9_sad/to_o/bones.odb

Index of /data/whome/rd2001/solve/p9_sad/

Name

	Apr 22 02:40 -	
edcan	Apr 22 02:40 1062	
	Apr 22 02:40 0	
vedcan	Apr 22 02:40 0	
edcan	Apr 22 02:40 0	
edcan	Apr 22 02:40 0	
vedcan	Apr 22 02:28 3613	

Exercises with ASDP:

From any machine, access the ASDP web site via netscape
or Internet explorer

<http://asdp.bnl.gov>

Login: rd2002? ("?" is the group id, in lower case)

Password:

Web home: <http://asdp.bnl.gov/whome/rd2002?>

Physical disk: </jupiter/data/whome/rd2002?>

Exercise-1: P9

Mad data collected at 3 wavelengths

/data/whome/rd2002?/solve/examples/p9

p9_se_w1.sca, 0.9790, inflection

p9_se_w2.sca, 0.9788, peak

p9_se_w4.sca, 0.9600, remote

Four Selenium atoms per molecule

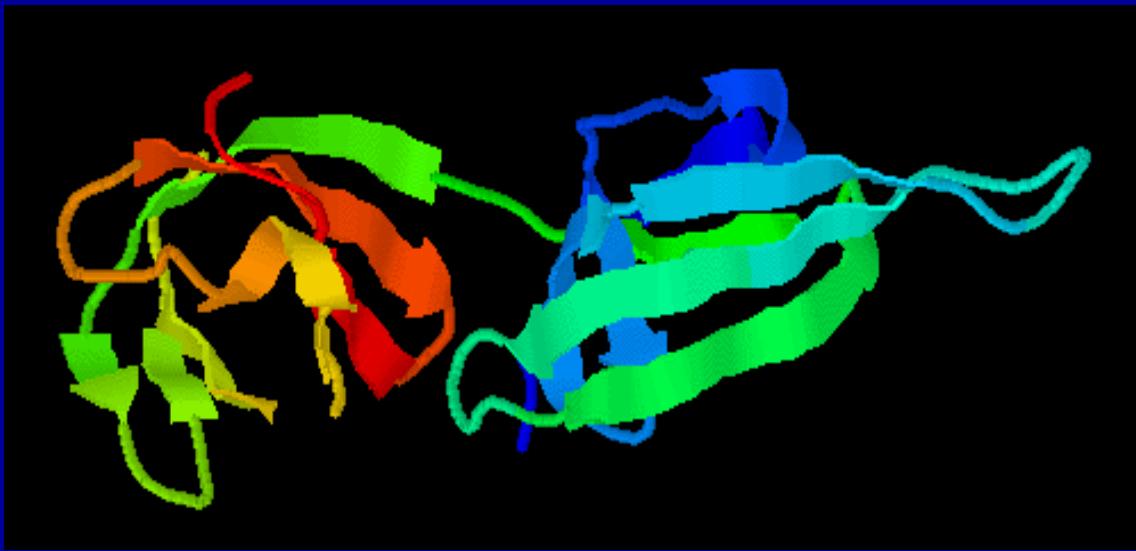
Molecule has 147 residues

Space group is I4

Unit cell: $a=113.95$ $b=113.95$ $c=32.47$

One molecule per asymmetry unit

Resolution 20Å to 2.1Å



Exercise-1: SAD by SOLVE

- Click “Set” (link at the top of the whome directory page).
- Set a new project id: “p9_sad_test” and use the peak data.
- Click “solve”, “p9_sad_test”, and “sad_solve.script”. Edit it if need.
- Submit “sad_solve.script” on “equalizer” (takes 20-30 mins).
- Run “post_resolve.csh” if you see “ha.pdb” and “resolve.mtz” .
- Download all files in the subdirectory “to_o” to local SGI machine.
- Type “ono” and hit return keys until O’s window appears. Read in “omac” and “menu omap”, and play.

Exercise-2 P097

Mad data collected at 3 wavelengths

/data/whome/rd2002?/solve/examples/p97

p97_85i_ano.sca, 0.9788, inflection

p97_85p_ano.sca, 0.9785, peak

p97_85r_ano.sca, 0.9400, remote

4 Se atoms per molecule (expect 3)

Molecule has 246 residues

Space group is P212121

Unit cell: $a=57.75$ $b=68.68$ $c=125.2$

2 molecules per asymmetry unit

Resolution 50Å to 1.9Å



Exercise-2: MAD by SOLVE

- Click “Set” (at the top of the web home directory page).
- Set a new project id: “p97_mad_test” and the related info.
 - Click “Solve”, “p97_mad_test”, and “mad_solve.script”. Edit it if need.
 - Run “mad_solve.script” on “equalizer” (40-60 mins).
 - Run “post_resolve.csh” if you see “ha.pdb” and “resolve.mtz”.
- Download all files in the subdirectory “to_0” to the local SGI machine.
- Type “ono” and hit return keys until O’s window appears. Read in “omac” and “menu omap”, and play.