# EMAN2 and EMEN2: Flexible Python-based platforms for electron microscopy

SciPy 2013
June 27, 2012

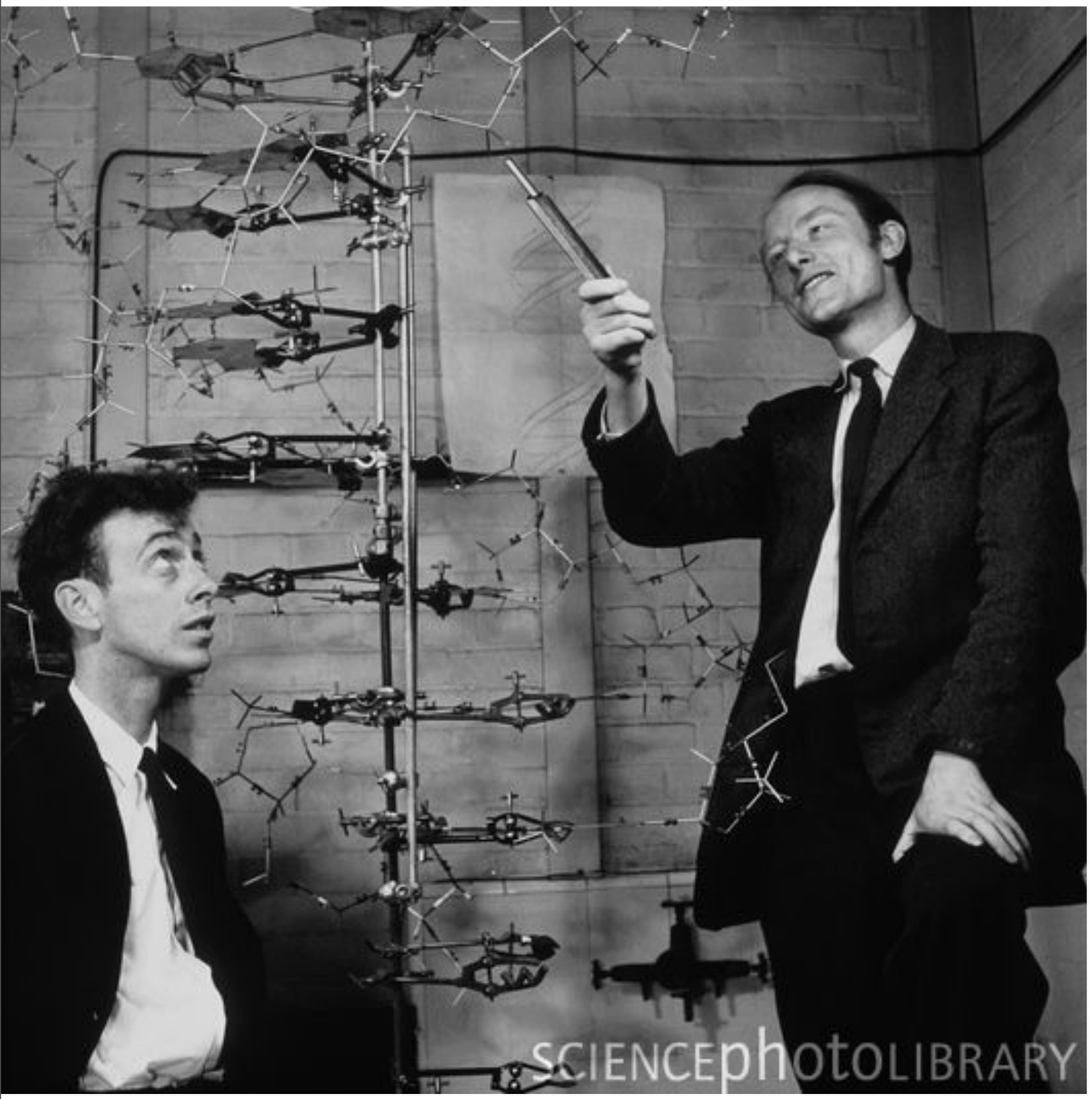

Ian Rees
National Center for Macromolecular Imaging
Baylor College of Medicine
Houston, TX
ian.rees@bcm.edu
ian@ianrees.net


NCMI National Center for Macromolecular Imaging

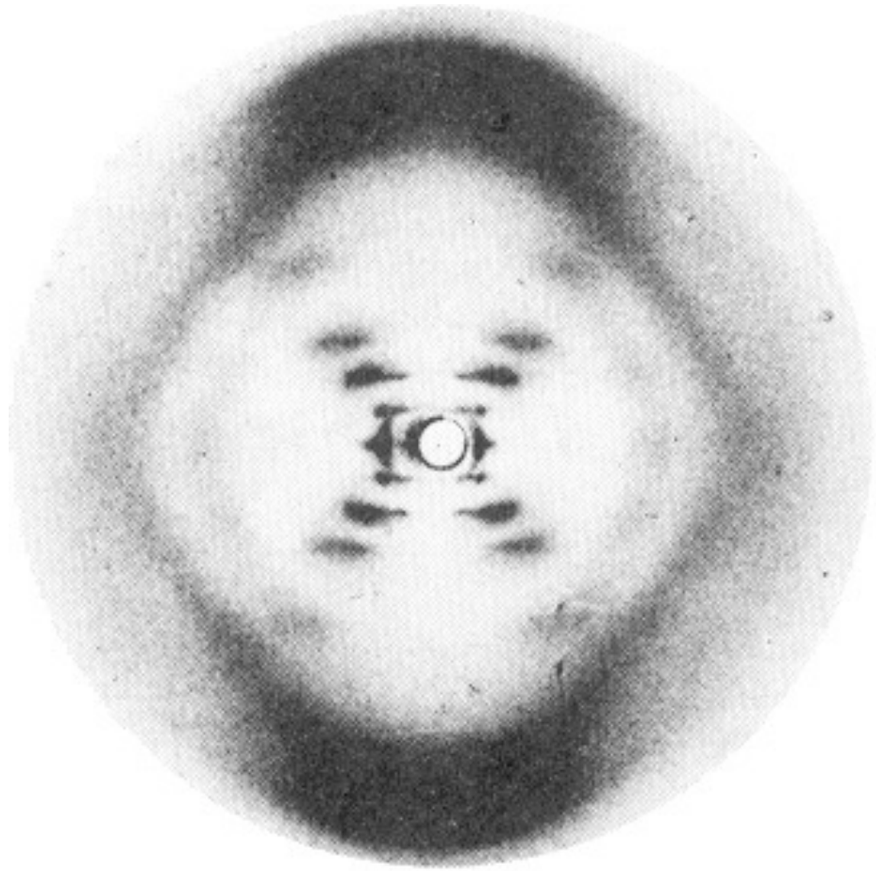
# Overview

- National Center for Macromolecular Imaging
- EMEN2 Database

# Structure!




http://www.sciencephoto.com/media/222783/enlarge



http://en.wikipedia.org/wiki/John_Kendrew

Franklin & Gosling 1953

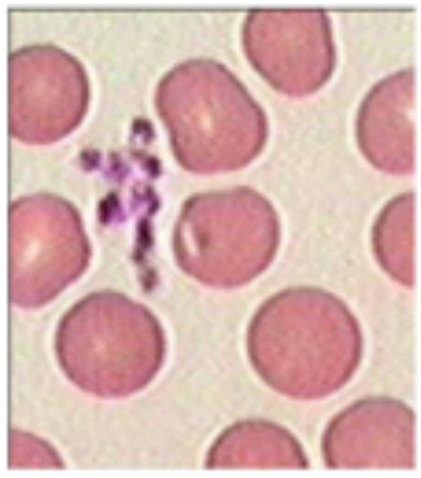
# Imaging macromolecules

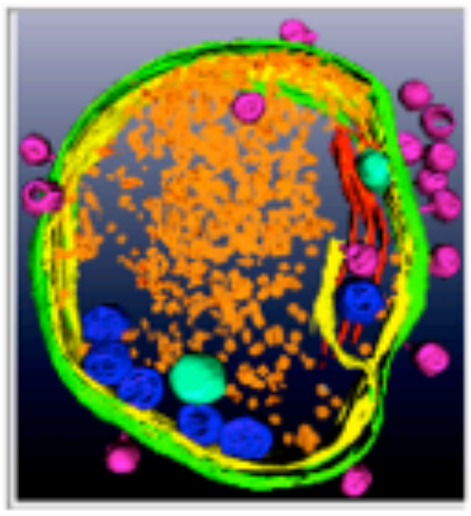
Optical microscopy

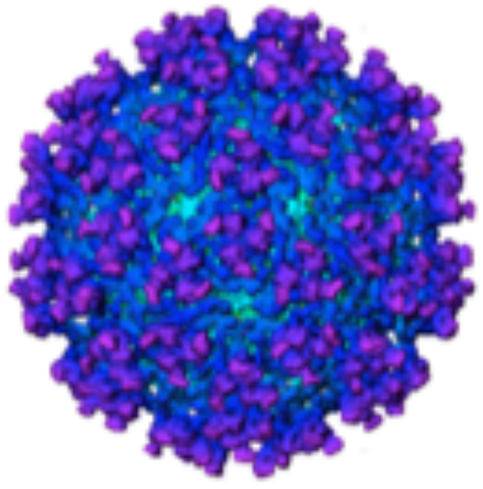
Electron tomography

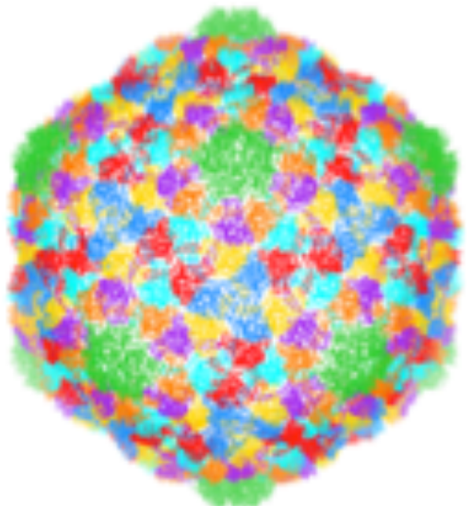
Single particle reconstruction

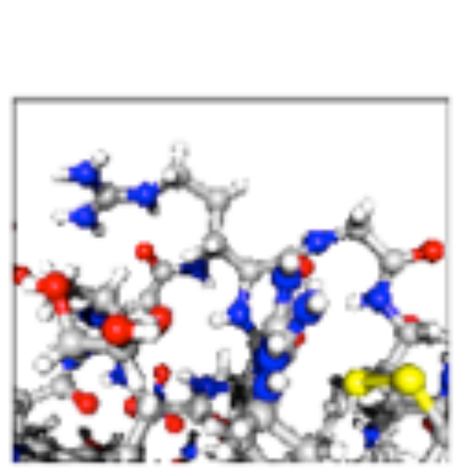
Crystallography NMR



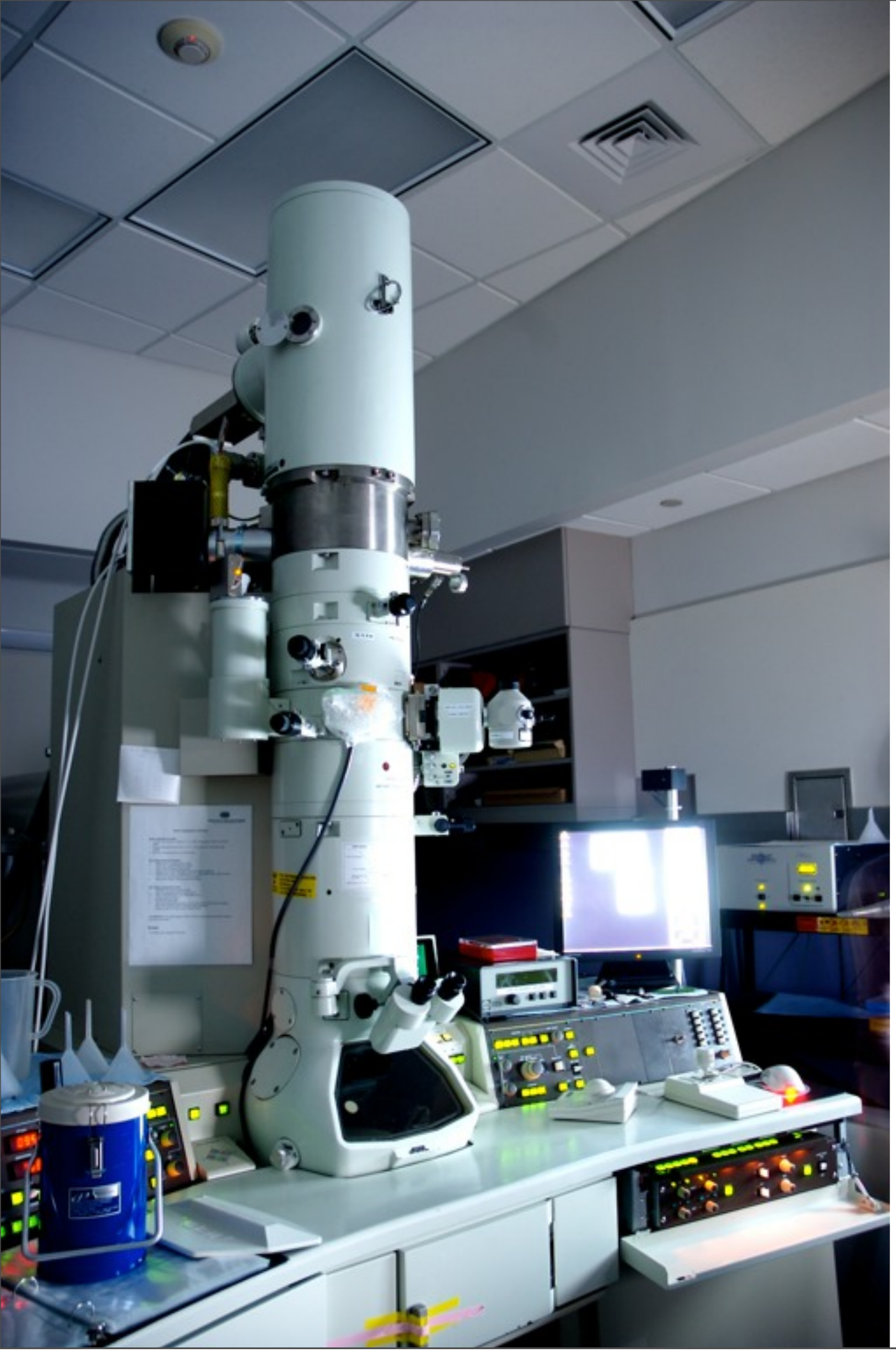








2000 Å 60 Å 10 Å 4 Å 1 Å



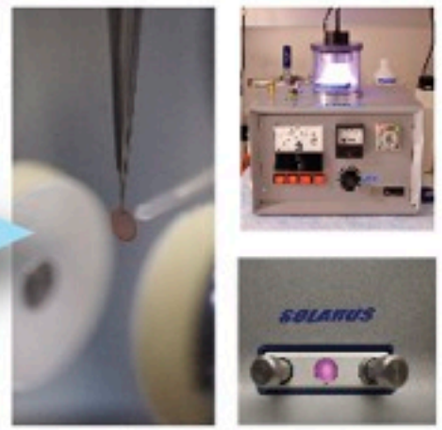
# Cryo-EM

- No crystals
- Near-native solvent environment
- Large complexes
- Tomography: 20-40 Å
- Single-particle: 3-10 Å

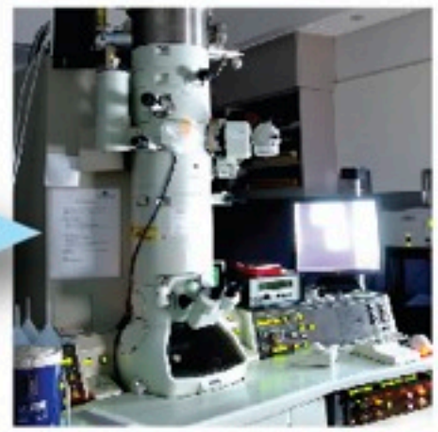
# Cryo-EM pipeline

Biochemical Prep

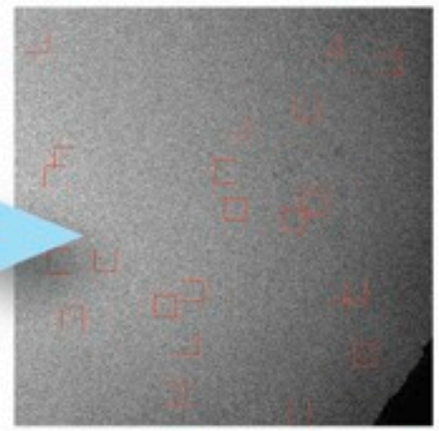
Cryo-EM Prep




Screening



Data Collection



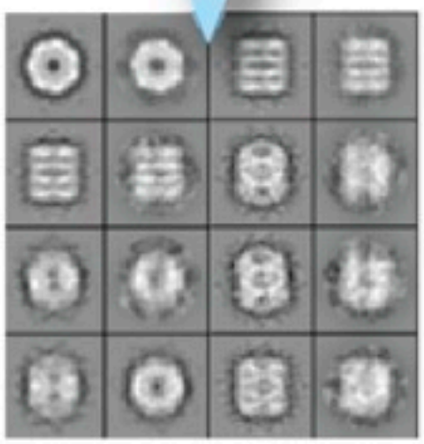
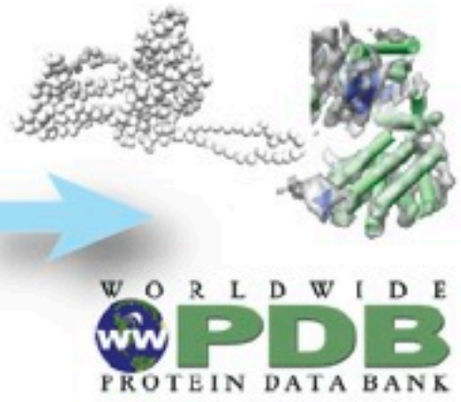


Image Processing

Reconstruction

Structural Analysis

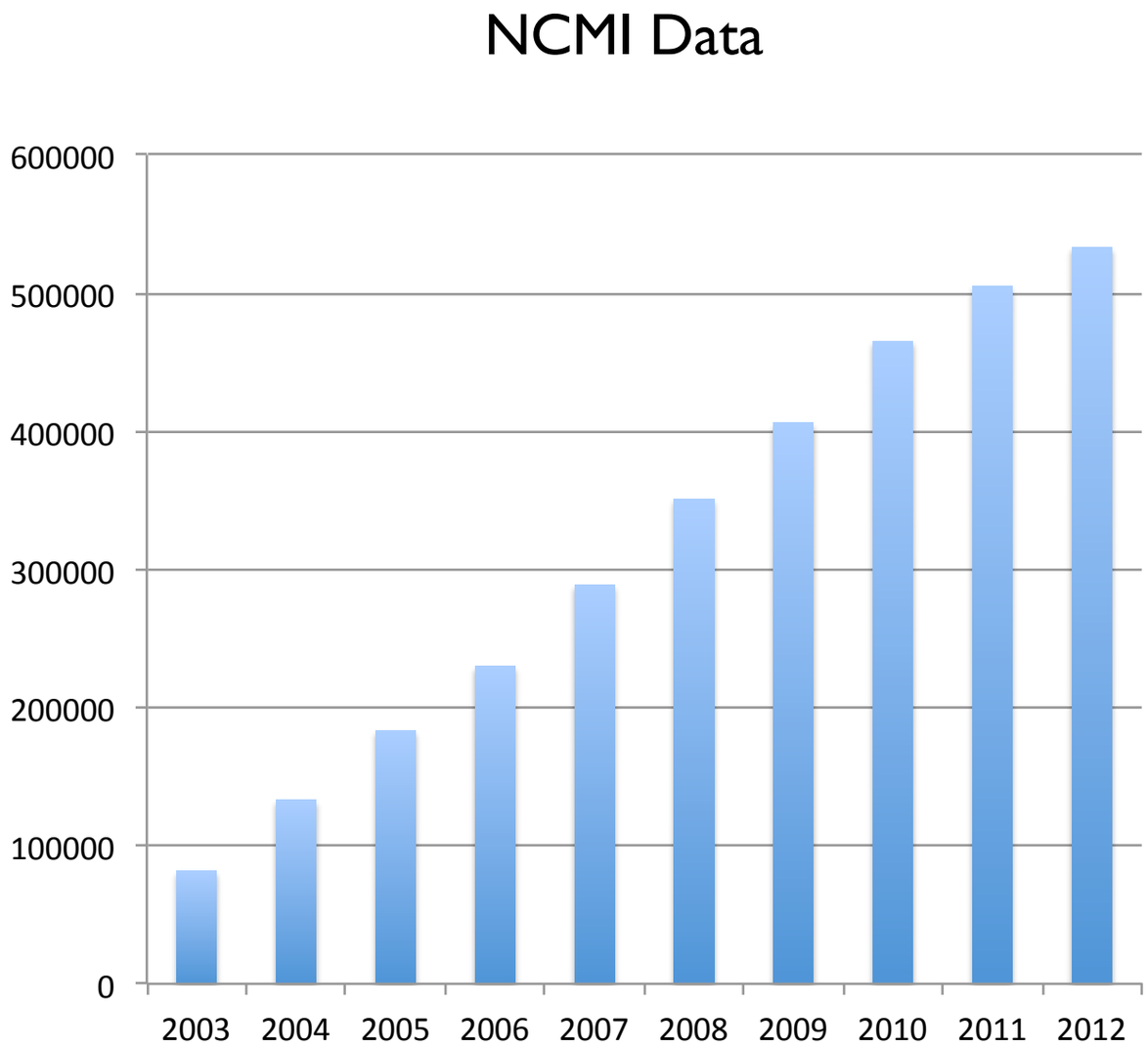



Annotation & Deposition

# Cryo-EM pipeline

- "Garbage in, garbage out"
- Optimize each stage of the pipeline
- A Big Investment!
  - Time
  - Microscopes and other equipment
  - Computational resources
  - Skilled microscopists (and students...?)

# Large and growing data




- 300+ projects
- 800+ users
- ~5TB / month
- Variety of hardware and software

# Beyond just archival...

- Encourage good record keeping (automation)
- Find things after people leave the lab
- Reconstruct an experimental protocol refined over years
- Share with collaborators
- Instrument usage accounting

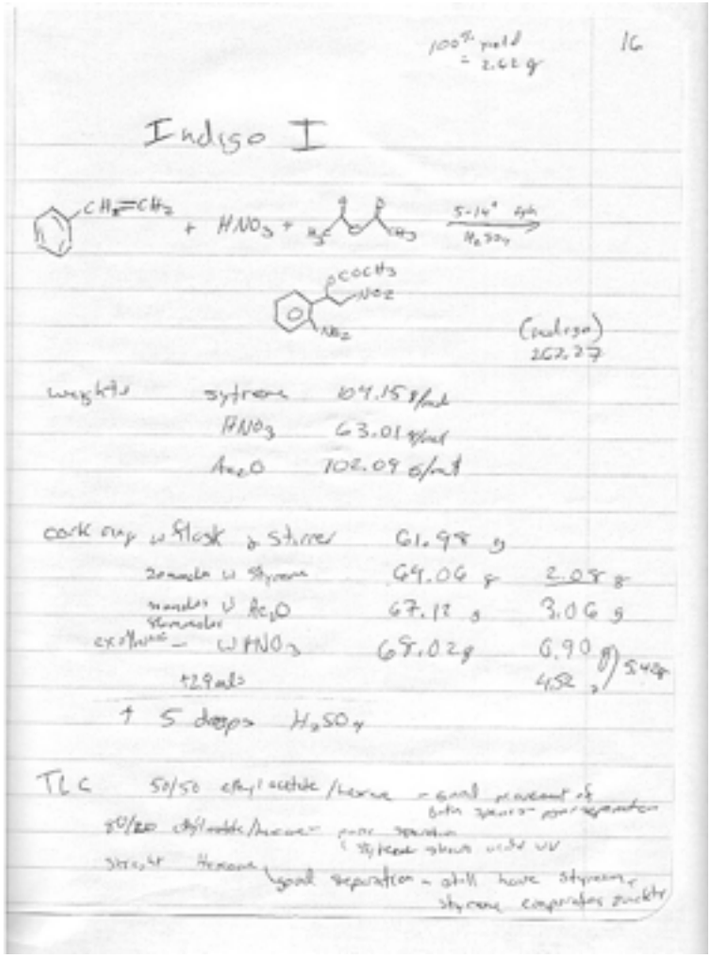
# Experimental protocols?

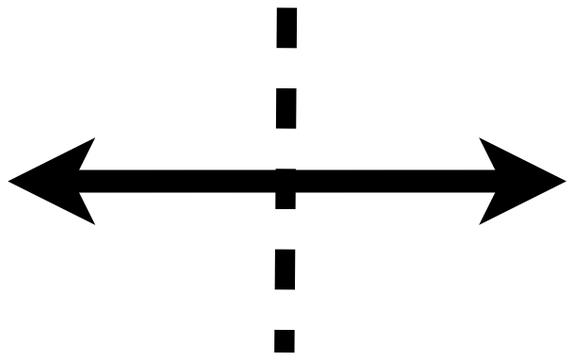
- Hard to establish standards...
- Differences between investigators and labs
- Difficult to anticipate future needs

# Protocols evolve...

- Protocol variations and customizations
- Sometimes a dead end!
- History of a protocol is itself important

# How to store?



Our Needs

| Film/Frame Id | Defocus | Mag | Total Dose | ExpTime |
|---|---|---|---|---|
| ccd_20040314 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_20040314 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_search1 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_search2 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_search3 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_search4 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |
| ccd_search5 #1 | 0.8 um | 138.44k | 20.0 | 1.0 s |
| ccd_search6 #2 | 2.5 um | 138.44k | 20.0 | 1.0 s |

Human readable
Unstructured
Easy to change

Machine readable
Structured
Hard to change

# Electronic Microscopy Electronic Notebook

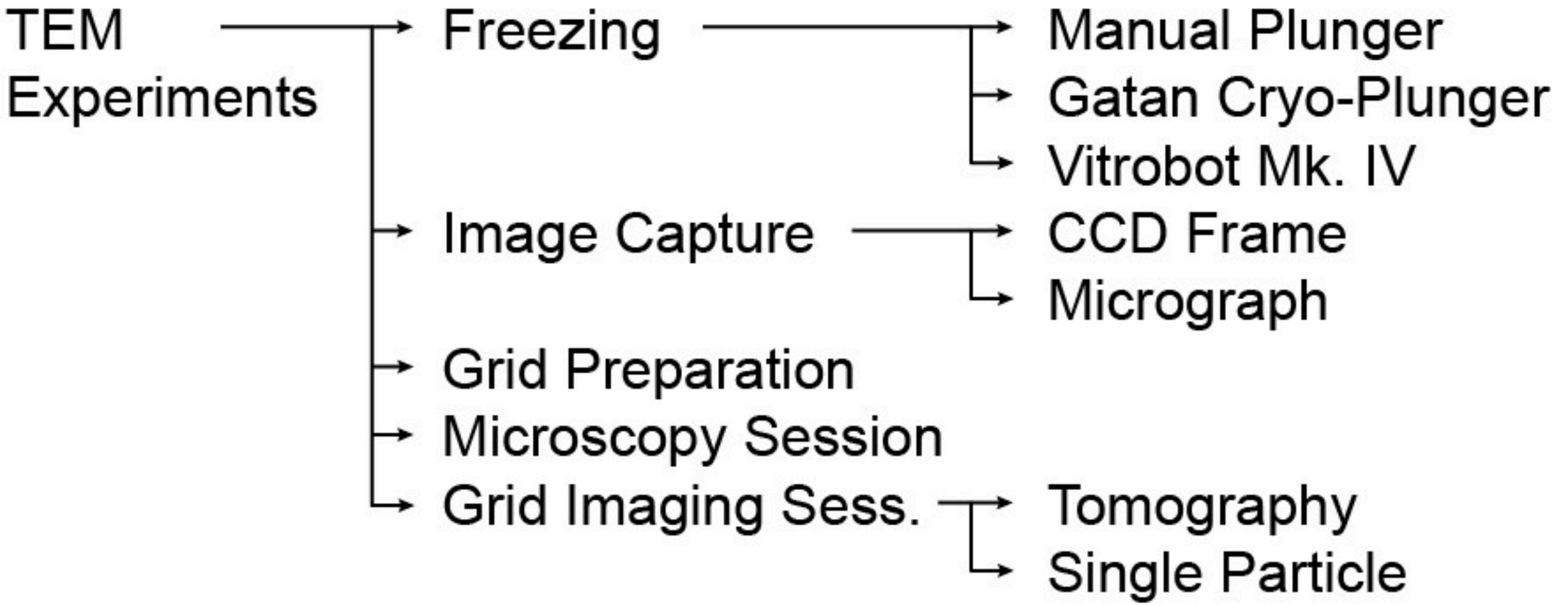
- BerkeleyDB key-value backend
- Flexible, end-user defined schema
- Originally designed for cryo-EM, but adaptable to other domains
- Extensions and custom applications

# Protocols

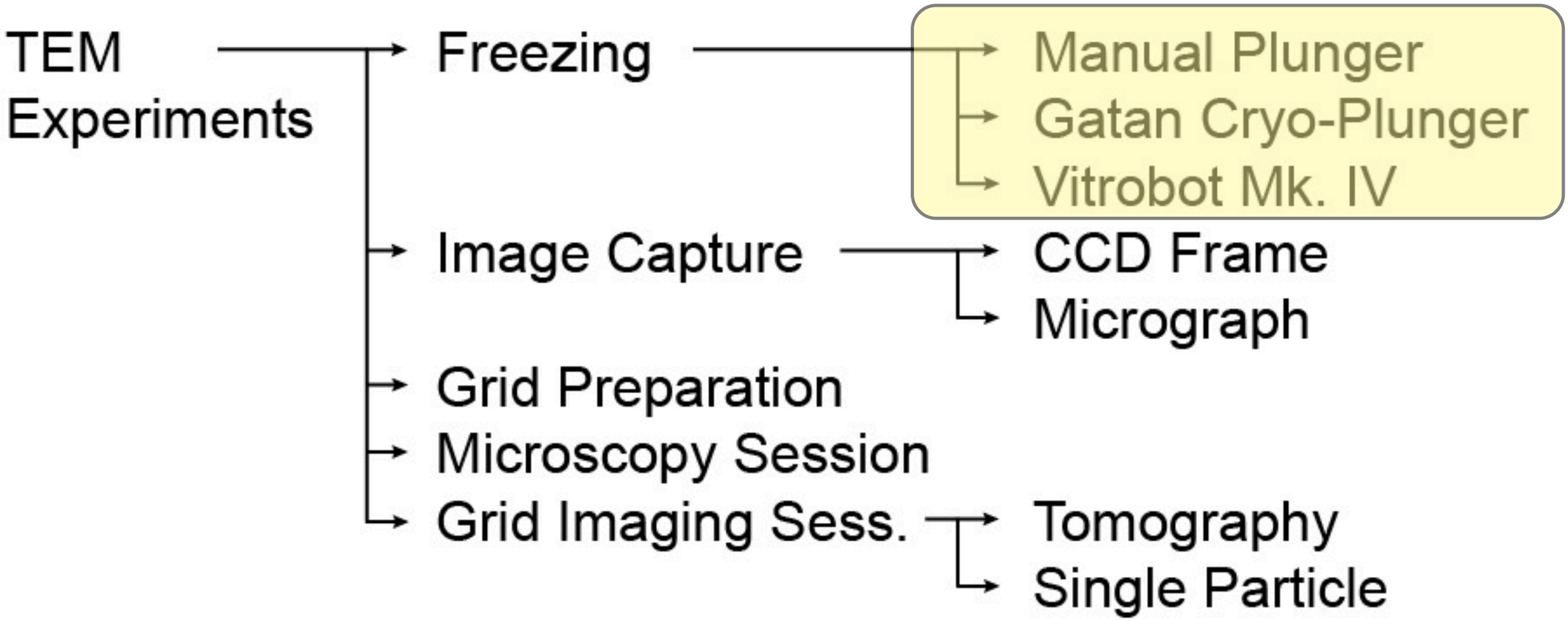
- Plain text description of experiment
- Embedded parameters

*A preprepared grid from batch $grid_batch_id was placed in a pair of forceps and loaded into the plunger. $cryogen was preprepared below the plunger.....*

# Protocol ontology



# Protocol ontology



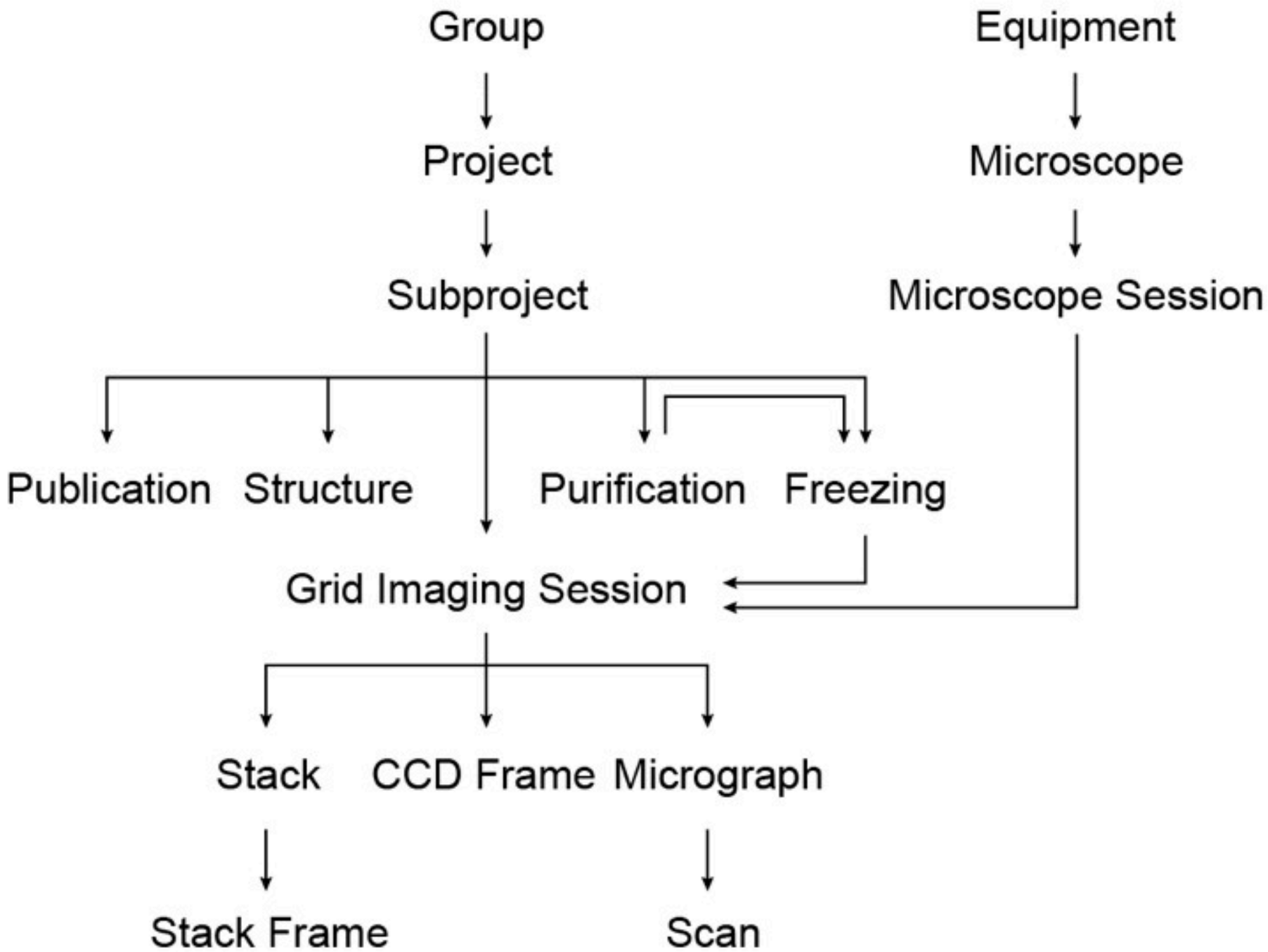
# Parameters

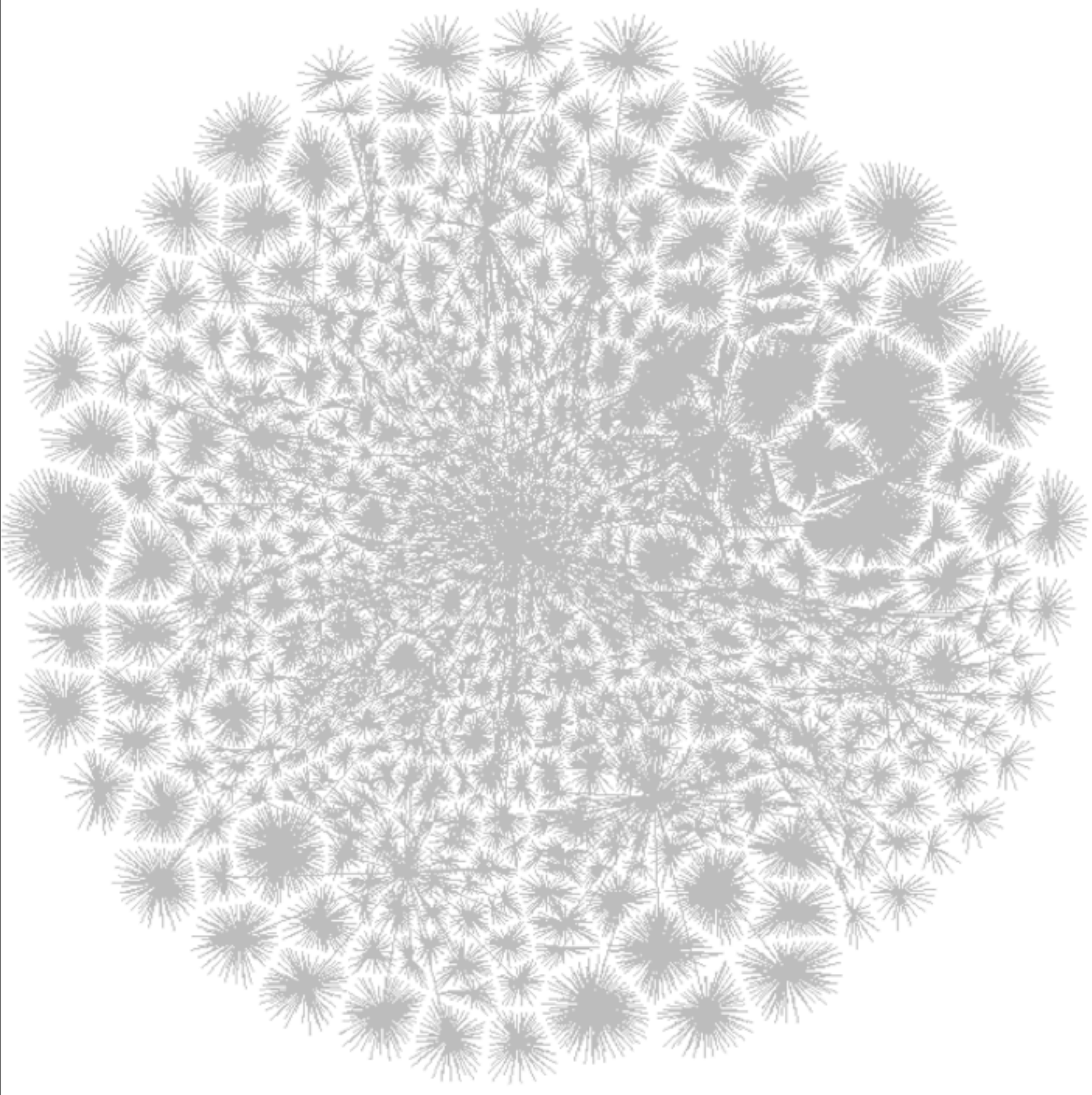
- Description of the parameter and intended use
- Data type (string, integer, list, etc.) and units
- Constraints and validation

*$aperture_objective*

*The aperture of the objective lens*
*Integer, default unit is micrometer*
*0µm, 20µm, 50µm, 70µm, 150µm.*

# Flexible record organization





# GroEL

10 years

22 users

15,500 records

- 18 subprojects
- 6,000 micrographs
- 5,000 ccd frames
- 14 purifications
- 34 freezings
- 170 grids
- 102 notes

```
>>> print db.record.get(‘46605’)
{
 'rectype': 'project',
 'name_project': 'High resolution GroEL',
 'name_specimen': 'GroEL',
 'description_goals': 'We have found previously that during the assembly....',
 'mass_specimen': 800.0,
 'project_status': 'Complete',
 'project_type': 'Core',
 'project_investigators': ['jj', ‘chrisb’, ‘matt’, ...],
 'name_pi': ['davidchuang'],
 'file_binary': ['bdo:20091112400000'],
 'name': '46605',
 'children': set(['268417', '294994', '294995', ...]),
 'parents': set(['136']),
 'keytype': 'record',
 'creator': 'root',
 'creationtime': '2001-10-03T05:00:00+00:00',
 'modifyuser': 'mikeschmid',
 'modifytime': '2010-09-21T18:21:35+00:00',
 'groups': set('public'),
 'permissions': ((',), (), ('jj', 'matt', 'chrisb'), (‘davidchuang’)),
 'comments': [],
 'history': [['mikeschmid', '2010-10-21T20:08:48+00:00', 'project_status', None]],
 'uri': 'http://ncmidb.bcm.edu/record/46605/'
}
```

# Security & logging

- Per-record, with four levels of privilege:
  - Read, read and comment, edit values, change permissions
- Groups and roles
- All changes are logged
- History log is integral to record

# Open-source

- Dependencies:
  - Python, BerkeleyDB, bsddb3, dateutil
    - Twisted, Mako Templates, Markdown, JSONRPC
- EMAN2, ImageMagick, pyOpenSSL

NCMI Database: Home

ncmidb.bcm.edu

NCMI National Center for Macromolecular Imaging

Home Query Bookmarks Admin Ian Rees

Home

Groups

- Workshops
- Public Data
- Collaboration & Data Exchange
- Tests
- NCMI
- CCBC
- JEOL Demo Projects
- CryoEM Training
- Prasad Lab
- Test Group
- Nanomedicine
- Test Group
- Visitors
- Shared Data

New group

## Ian Rees

Edit Profile


Department: Structural and Computational Biology and Molecular Biophysics

Institution: Baylor College of Medicine

Address: 1 Baylor Plaza
N420
Houston, TX 77030
United States

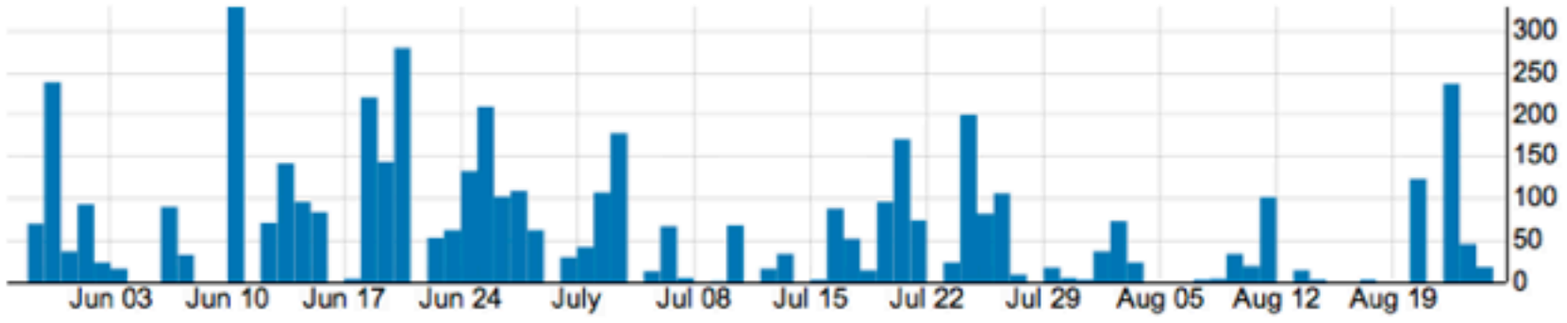
Email: ian@ianrees.net

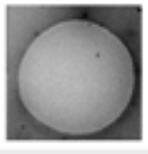
Phone: None

Web: http://ncmi.bcm.edu

## Activity and recent records

All records | Record tree




| recname | thumbnail | Protocol | Object Name | Creation time | Created by |
|---|---|---|---|---|---|
| Lab notebook: Current status 8/24/2012 | | labnotebook | 524968 | 2012/08/24 18:16 | Chang, Juan |
| CCD |  | ccd | 524967 | 2012/08/24 14:59 | Dillard, Rebecca |

NCMI Database: Record: Sub

ncmidb.bcm.edu/record/46605/children/grid_imaging/

NCMI National Center for Macromolecular Imaging

Home Query Bookmarks Admin Ian Rees

NCMI — Project GroEL (GroEL) — Subproject High resolution GroEL

Record: Subproject High resolution GroEL | Visualization (1) | Publication (2) | Reference (10) | Purification (7)

Grid imaging session (105) | Lab notebook entry (34) | Structure (1)

Record: Subproject High resolution GroEL

- Edit
- New
- Permissions
- Attachments
- Relationships
- Views
- 1 Comments
- 4 Edits

Steven J Ludtke
2010/09/21 13:21

Children

| | |
|---|---|
| Visualization | 1 |
| Publication | 2 |
| Reference | 10 |
| Purification | 7 |
| Grid imaging session | 105 |
| Lab notebook entry | 34 |
| Structure | 1 |

105 records (0.88s) Modify query — New grid_imaging — Rows — 1 / 2 ›

| Set Magnification | Specimen temp | Purpose | parentvalue title_freezing | # ccd | # ddd | # micrograph | # scan | Creation time | Created by |
|---|---|---|---|---|---|---|---|---|---|
| 80 | 3.7 K | collecting data | | 49 | 0 | 0 | 0 | 2007/03/28 00:00 | Admin |
| 80 | 3.7 K | collecting data | | 108 | 0 | 0 | 0 | 2007/03/27 00:00 | Admin |
| 80 | 3.71 K | collecting data | GroEL G8 G9 | 205 | 0 | 0 | 0 | 2007/03/27 00:00 | Admin |
| 80 | 3.71 K | collecting data | GroEL G8 G9 | 133 | 0 | 0 | 0 | 2007/03/27 00:00 | Admin |
| 80 | 3.71 K | collecting data | GroEL G8 G9 | 182 | 0 | 0 | 0 | 2007/03/27 00:00 | Admin |
| 80 | 4.18 K | collecting data | HR GroEL 2005 | 202 | 0 | 0 | 0 | 2006/03/24 00:00 | Admin |
| 80 | 4.17 K | collecting data | HR GroEL 2005 | 112 | 0 | 0 | 0 | 2006/03/23 00:00 | Admin |
| 80 | 4.17 K | collecting data | HR GroEL 2005 | 92 | 0 | 0 | 0 | 2006/03/22 00:00 | Admin |
| 80 | 4.17 K | collecting data | HR GroEL 2005 | 208 | 0 | 0 | 0 | 2006/03/21 00:00 | Admin |
| 80 | 4.16 K | collecting data | HR GroEL 2005 | 101 | 0 | 0 | 0 | 2006/03/20 00:00 | Admin |
| 60 | 4.18 K | collecting data | GroEL G8 G9 | 0 | 0 | 100 | 100 | 2005/07/21 00:00 | Admin |
| 60 | 4.17 K | collecting data | GroEL G8 G9 | 0 | 0 | 80 | 80 | 2005/07/19 00:00 | Admin |
| 60 | 4.17 K | collecting data | GroEL g1 g2 | 0 | 0 | 100 | 100 | 2005/07/18 | Admin |

NCMI Database: Record: CCD

ncmidb.bcm.edu/record/56969/

NCMI National Center for Macromolecular Imaging

Home Query Bookmarks Admin Ian Rees

Record marked as published data

— Microscopy: JEOL 3000SFF by Donghua Chen @ : collecting data
— Subproject High resolution GroEL
08 00:00 — Vitrobot: GroEL G8 G9 by Donghua Chen @ 2005/03/11 00:00

Imaging: on JEOL 3000SFF by Admin @ 2007/03/27 00:00: 205 images — CCD ccd_3160

## Record: CCD ccd_3160

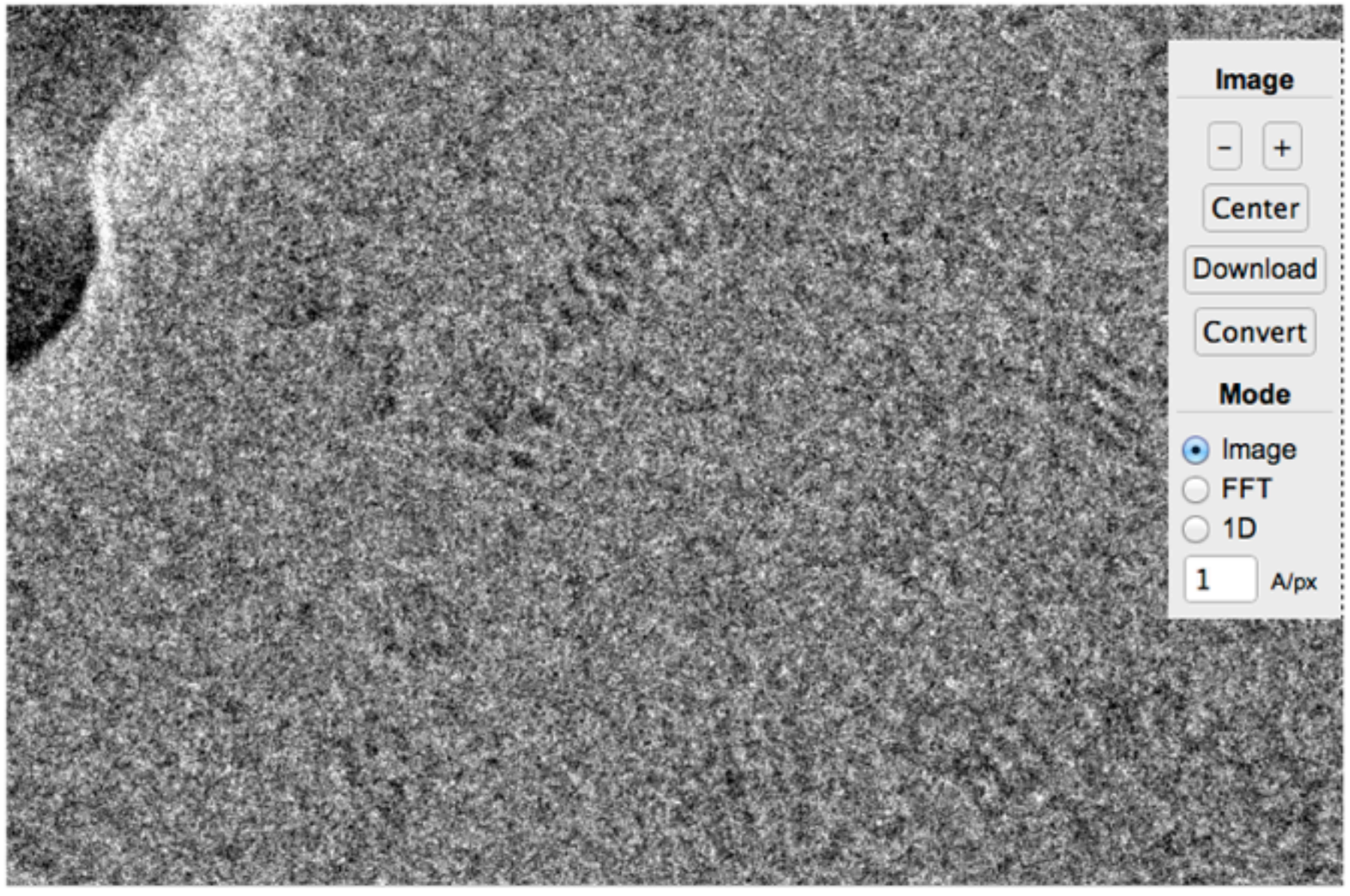
Record: CCD ccd_3160

- Edit
- New
- Permissions
- 1 Attachments
- Relationships
- Views
- 0 Comments
- 3 Edits

Admin
2010/09/21 23:15

Children

No children



EMEN2: Query

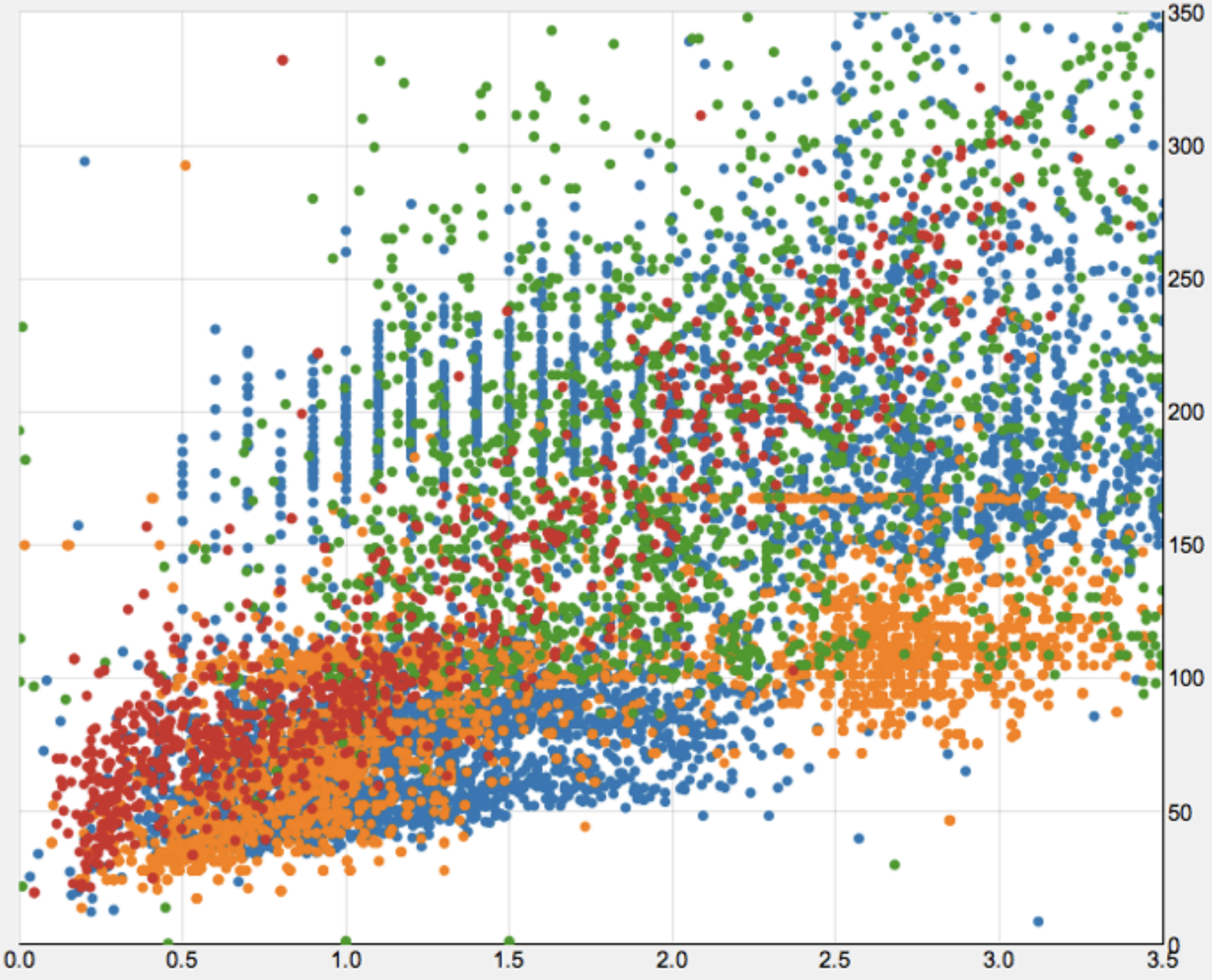
ncmidb.bcm.edu/query/rectype.is.image_capture*/?x.key=ctf_defocus_measured&y.key=ctf_bfactor&z.key=$@parentvalue(tem_name...

NCMI National Center for Macromolecular Imaging

Home Query Bookmarks Admin Ian Rees Search

Query

12199 records, 8 protocols Query Edit Plot Rows 1 / 122 › »




**X**

Param: ctf_defocus_measu

Range: 0 - 3.5

Bin:

**Y**

Param: ctf_bfactor

Range: 0 - 350

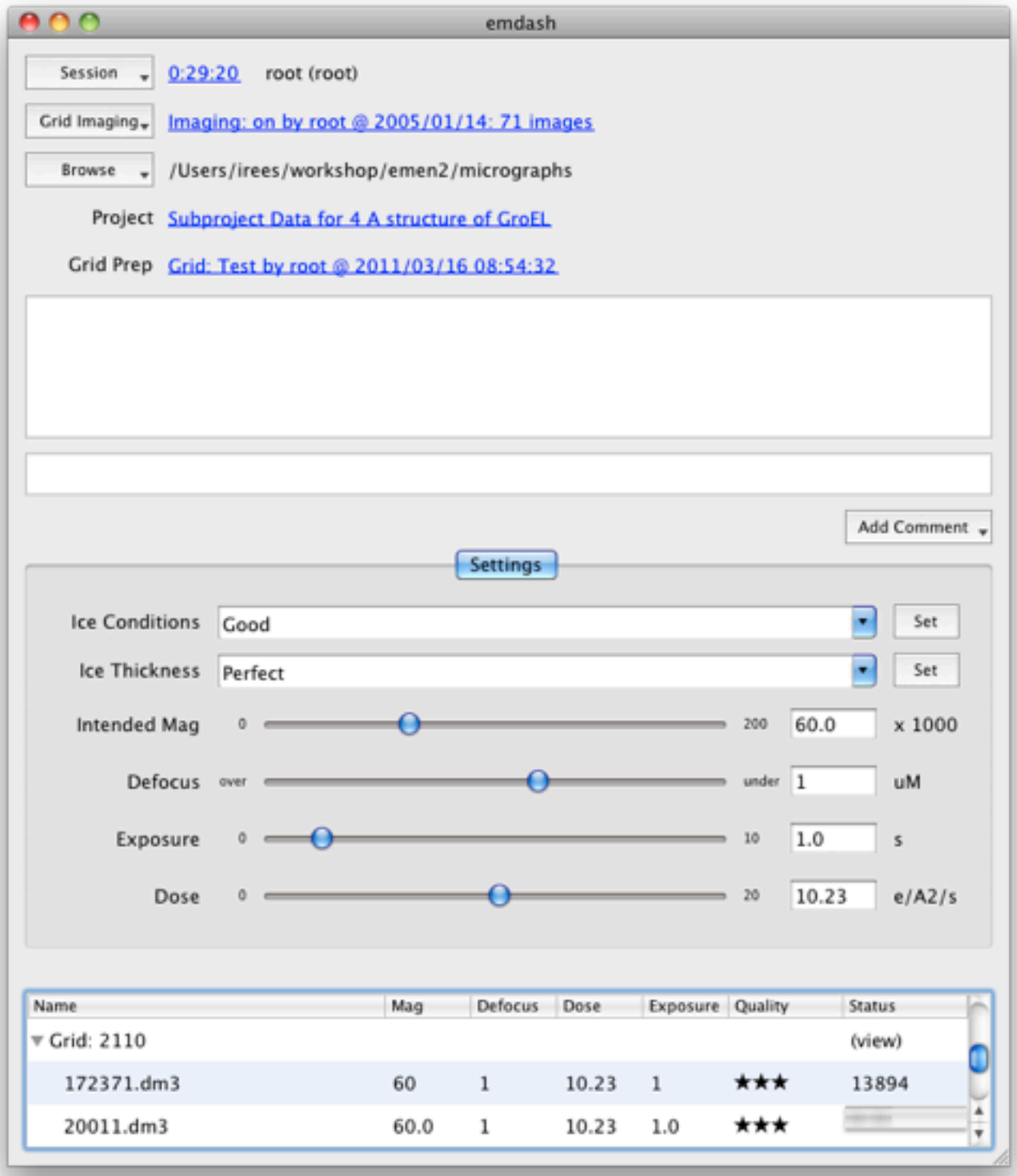
**Z**

Param: $@parentvalue(tem

| | | |
|---|---|---|
| ☑ JEOL 3200FSC | 5185 | |
| ☑ JEOL 3000SFF | 3116 | |
| ☑ JEOL 2010F | 2568 | |
| ☑ JEOL 4000 | 989 | |
| ☐ | 341 | |
| Total: | 12199 | |

Apply

# Data upload




- EMDash
- Runs on microscope computer
- Manages sessions, uploads data

# Extensible architecture

- EMEN2 Public API using JSON-RPC
- Extensions:
  - Data types, units, macros
  - File format handlers
  - Web interface
- Extend existing functionality, or entirely new applications

# EMEN2 API

```
>>> import jsonrpc.proxy
>>> db = jsonrpc.proxy.JSONRPCProxy("http://localhost:8080")
>>> db.login("ianrees", getpass.getpass())
Password:

>>> children = db.rel.children('201364', recurse=-1)
>>> len(children)
63627

>>> q = db.query([['children','==','201364*'],['rectype','==','image_capture*'],
['creationtime','>','2012']])
>>> len(q['names'])
5015

>>> r = db.record.new(rectype='folder', inherit=['201364'])
>>> r['name_folder'] = "Presentations"
>>> db.record.put(r)
{u'name': u'563581', u'creator': u'ianrees', u'keytype': u'record', u'creationtime':
u'2013-06-27T20:47:43+00:00', u'comments': [], u'rectype': u'folder', u'name_folder':
u'Presentations', ....}
```

# EMEN2 Extensions

```python
from emen2.web.view import View

@View.register
class TestView(View):
    @View.add_matcher(r'^/test/view/(?P<name>[^/]*)/$')
    def main(self, name, **kwargs):
        self.title = 'Test view'
        self.template = '/test/testview'
        # do stuff here.
        self.ctxt['name'] = name
        self.ctxt['value'] = 'Test!'
```

```
<%inherit file="/page" />

<h1>Test page</h1>

This uses Mako-style templates. The page name was ${name}
and the value was: ${value}
```

# Ext.: Mark public data

NCMI database: Record: Sub

ncmidb.bcm.edu/record/46605/publish/

NCMI National Center for Macromolecular Imaging

Home Query Bookmarks Admin Ian Rees

Record: Subproject High resolution GroEL | Visualization (1) | Publication (2) | Reference (10) | Purification (7) | Grid imaging session (105) | Lab notebook entry (34) | Structure (1)

Record: Subproject High resolution GroEL

- Edit
- New
- Permissions
- Attachments
- Relationships
- Views
- 1 Comments
- 4 Edits

Steven J Ludtke
2010/09/21 13:21

Children

| | |
|---|---|
| Visualization | 1 |
| Publication | 2 |
| Reference | 10 |
| Purification | 7 |
| Grid imaging session | 105 |
| Lab notebook entry | 34 |
| Structure | 1 |

## Manage public data — 7581 records selected

Save

Records marked in "orange" will be marked as public data when this form is saved. The number of selected records is shown above.

Click on a record to select or deselect it. You can also hold down "shift" while clicking to select or deselect all the children of that record.

- Subproject High resolution GroEL
  - Imaging: on JEOL 2010F by Admin @ 2002/04/22 00:00: 36 images
  - Imaging: on JEOL 2010F by Admin @ 2002/04/26 00:00: 36 images
  - Imaging: on JEOL 2010F by Admin @ 2002/04/26 00:00: 76 images
  - Imaging: on JEOL 2010F by Admin @ 2002/05/01 00:00: 73 images
  - Imaging: on JEOL 2010F by Admin @ 2002/05/07 00:00: 93 images
  - Imaging: on JEOL 2010F by Admin @ 2002/05/09 00:00: 137 images
  - Imaging: on JEOL 2010F by Admin @ 2002/05/11 00:00: 67 images
  - Imaging: on JEOL 2010F by Admin @ 2002/06/12 00:00: 62 images
  - Imaging: on JEOL 2010F by Admin @ 2002/07/18 00:00: 10 images
  - Imaging: on JEOL 2010F by Admin @ 2002/10/15 00:00: 12 images
  - Imaging: on JEOL 2010F by Admin @ 2002/11/04 00:00: 0 images

# Ext.: Grant reporting

```
<?xml version="1.0"?>
<P41_Progress_Report xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://sis.ncrr.nih.gov"
xsi:schemaLocation="http://sis.ncrr.nih.gov http://APRSIS.NCRR.NIH.GOV/XML/p41_progress_report.xsd">
    <Grant_Info>
        <Serial_Number>002250</Serial_Number>
        <Fiscal_Year>2013</Fiscal_Year>
        <Reporting_To_Date>2012-10-31</Reporting_To_Date>
        <Reporting_From_Date>2011-11-01</Reporting_From_Date>
        <Project_Title>3-DIMENSIONAL ELECTRON MICROSCOPY OF MACROMOLECULES</Project_Title>
        <Recipient_Institution_Name>Baylor College of Medicine</Recipient_Institution_Name>
        <Activity_Code>P41</Activity_Code>
        <Director_Person_ID>271657</Director_Person_ID>
        <Patent_Or_Copyright_Award>N</Patent_Or_Copyright_Award>
    </Grant_Info>
    <Person Person_ID="271657">
        <Last_Name>Chiu</Last_Name>
        <First_Name>Wah</First_Name>
        <Full_Name>Wah Chiu</Full_Name>
        <Academic_Degree>PHD</Academic_Degree>
        <Phone_Number>+17137986985</Phone_Number>
        <Fax_Number>+17137988682</Fax_Number>
        <Email_Address>wah@bcm.edu</Email_Address>
        <Department>the National Center for Macromolecular Imaging</Department>
        <Non_Federal_Funding>
            <Source_Type>FDN</Source_Type>
            <Sub_ID>273871</Sub_ID>
            <Organization_Name>NSF</Organization_Name>
            <Grant_Or_Contract_Number>IIS-0705474</Grant_Or_Contract_Number>
            <Total_Support_Funds>60000</Total_Support_Funds>
        </Non_Federal_Funding>
        <Non_Federal_Funding>
            <Source_Type>FDN</Source_Type>
            <Sub_ID>273871</Sub_ID>
            <Organization_Name>NSF</Organization_Name>
            <Grant_Or_Contract_Number>EIA-0325004 </Grant_Or_Contract_Number>
            <Total_Support_Funds>0</Total_Support_Funds>
        </Non_Federal_Funding>
        <Federal_PHS_Funding>
            <Organization>NIH</Organization>
            <Sub_ID>273835</Sub_ID>
            <Sub_ID>273871</Sub_ID>
            <Grant_Or_Contract_Number>5PN2EY016525</Grant_Or_Contract_Number>
            <Total_Support_Funds>5110000</Total_Support_Funds>
        </Federal_PHS_Funding>
```

# Simple ext:: Event registration

BMB/Pharm Retreat: Registra

blake.grid.bcm.edu/bmbretreat/

**The Verna & Marrs McLean Department of Biochemistry & Molecular Biology**
**and**
**The Department of Pharmacology**
**Research Conference 2012**

October 11 & 12, 2012
The Tremont House, Galveston TX

## Ian Rees Edit

| | |
|---|---|
| Email: | ian@ianrees.net |
| Institution: | Baylor College of Medicine |
| Department: | BCM - Biochemistry and Molecular Biology |

## Registration Edit

| | |
|---|---|
| I am: | Student |
| Primary PI: | Ludtke, Steve |
| Secondary PI: | Chiu, Wah |
| I plan to attend: | 2012-10-11, 2012-10-12 |
| Presentation: | I would like to present a poster. |

| | |
|---|---|
| Accommodations: | I am a faculty member and will have my own hotel a room. |
| Meal preference: | |
| T-shirt: | Men's XL |

# Install

- Install dependencies, or EMAN2 binary package
- easy_install emen2
- python -m emen2.db.create -h test.db
- python -m emen2.web -h test.db
- http://localhost:8080

# EMEN2 next steps

- Protocol sharing system
- Additional documentation:
  - tutorials, sphinx, example extensions
- New instruments and software

# Acknowledgments


- Steven Ludtke
- Wah Chiu
- Mike Schmid
- Irina Serysheva
- Kendal Hirschi
- NCMI staff

- P41 GM103832
- R01 GM080139
- R01 GM079429
- PN1EY016525

NIH Training Grant:

- T15LM007093

# NCMI

- http://ncmi.bcm.edu
- http://blake.grid.bcm.edu/emanwiki
- Feel free to contact me to discuss:
  - ian@ianrees.net