

Integrating Globus into a Science Gateway for Cryo-EM

Michael Cianfrocco

Life Sciences Institute

Department of Biological Chemistry

University of Michigan

Globus World 2018

Executive summary

Biologists need atomic-detailed structures of proteins in order to understand healthy and diseased states of organisms at the molecular level.

A **revolution** has occurred in structural biology
- these questions can now be answered with
cryo-electron microscopy (cryo-EM)

But...

Cryo-EM is a big data technology, generating
20+TB per person per project

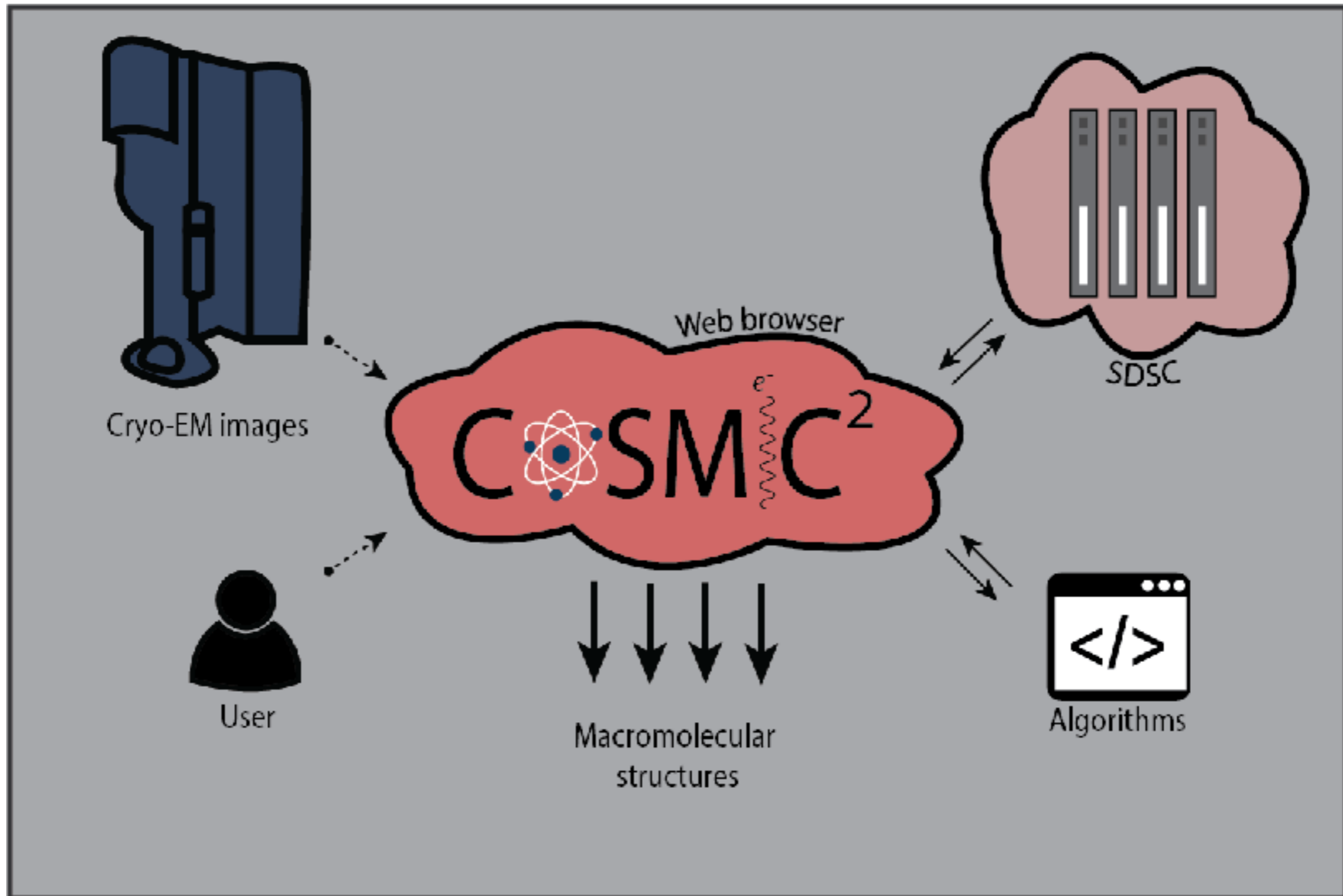
So...

We **incorporated Globus into a web-platform** to allow point-and-click data upload & analysis on HPC resources

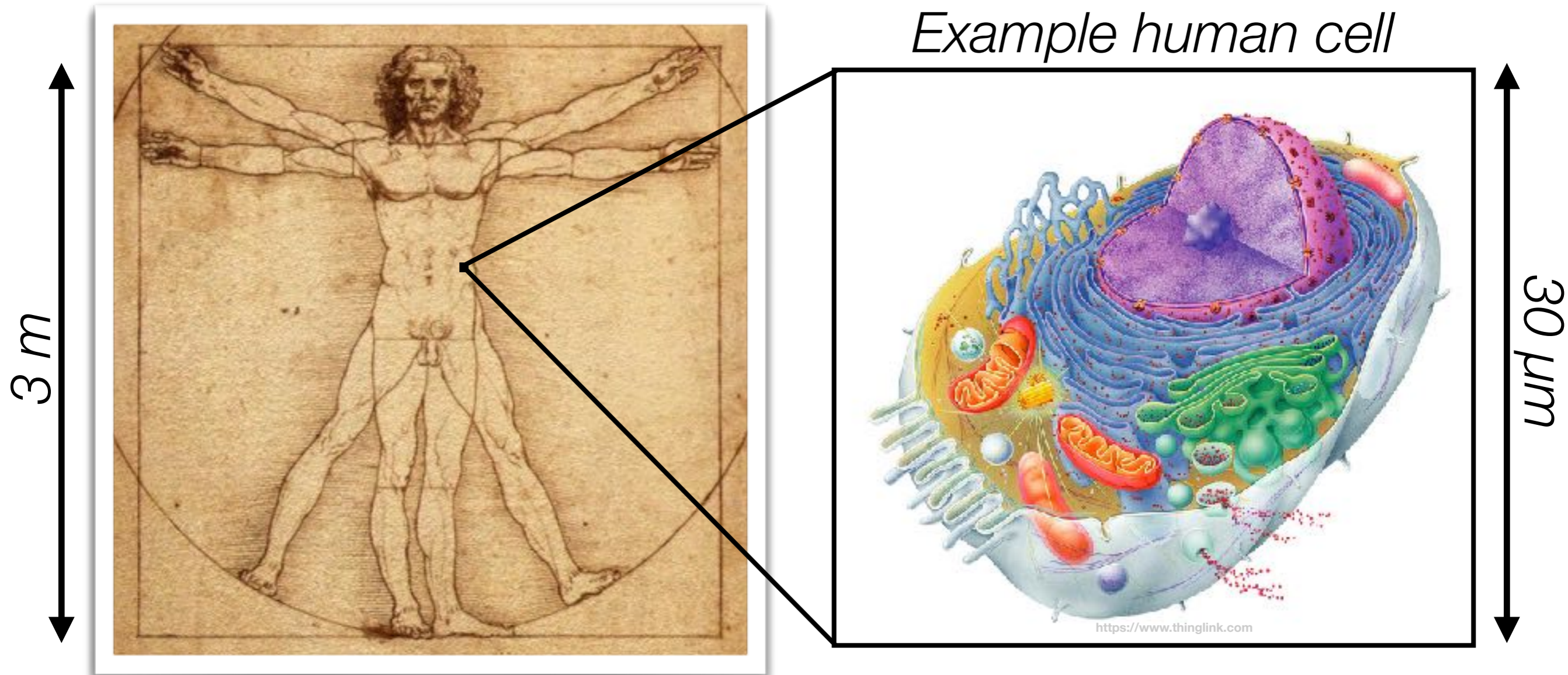


COSMIC²: a platform for determine cryo-EM protein structures

Cryo-EM **O**pen **S**ource **M**ultiplatform **I**nfrastructure for **C**loud **C**omputing



Understanding biology requires an understanding of the 'microcosmos'



100,000 times smaller

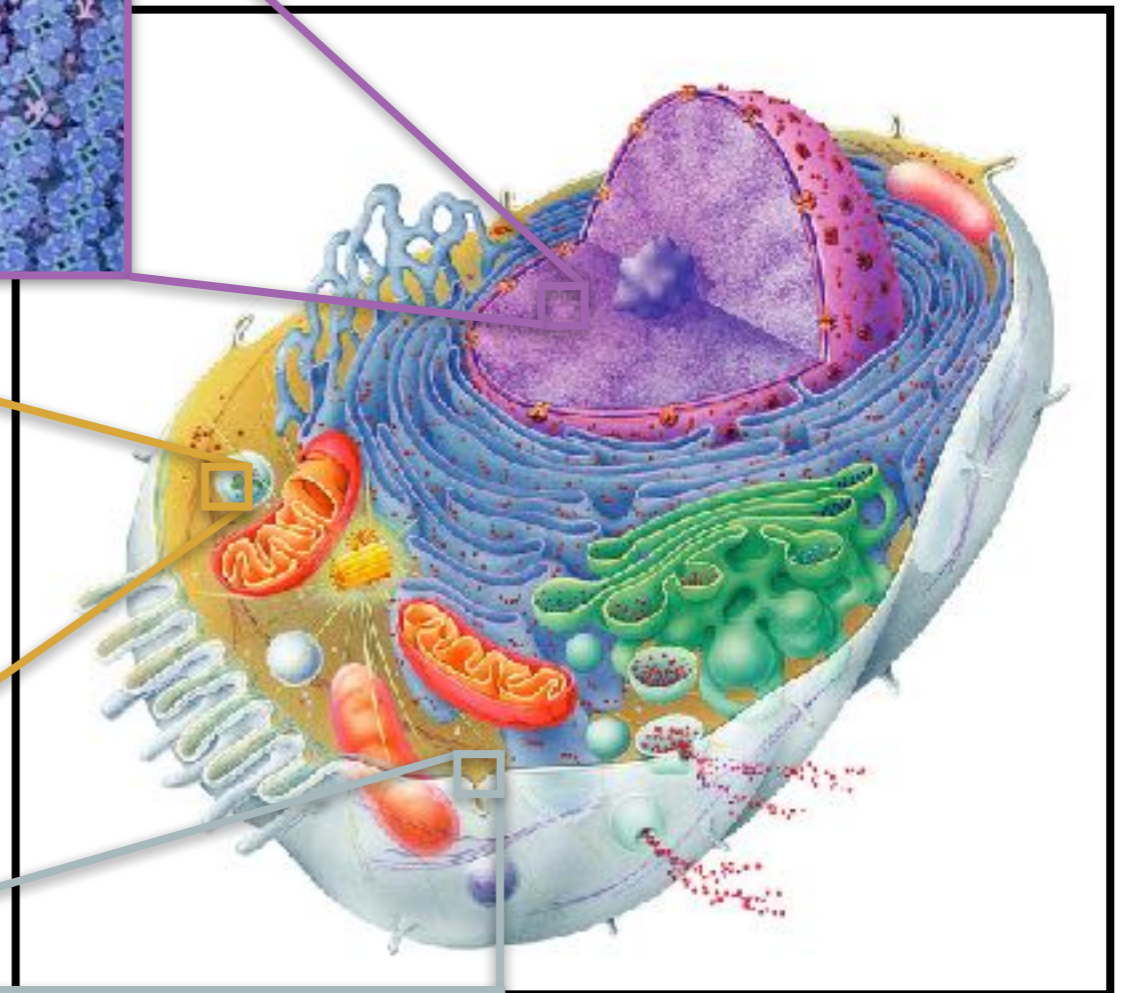
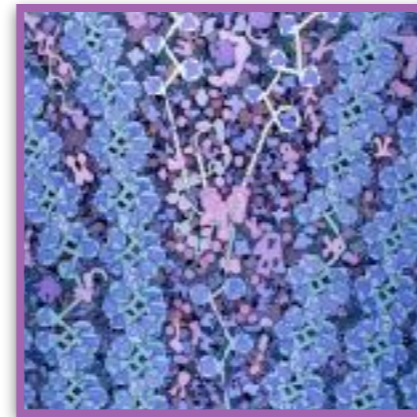
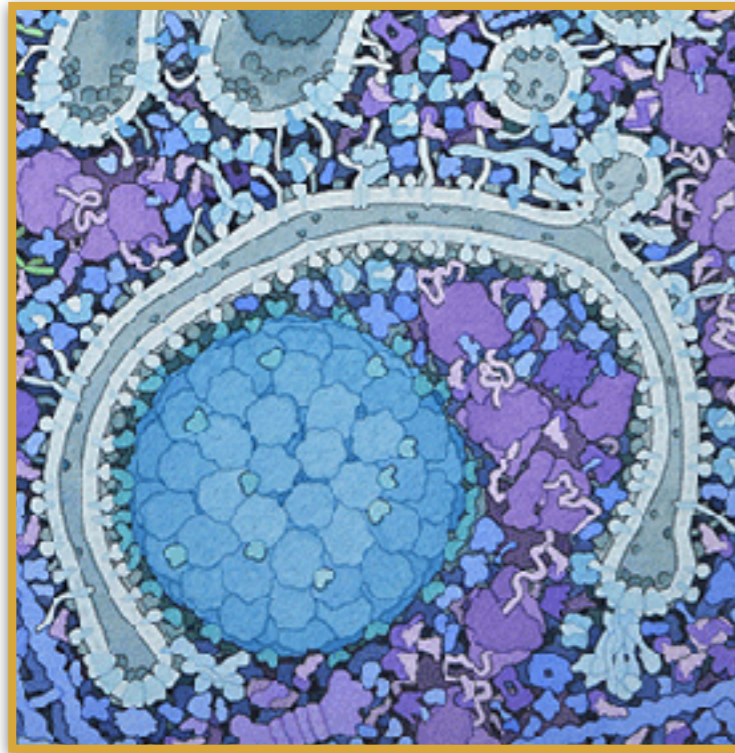
Understanding biology requires an understanding of the 'microcosmos'

***Model** illustrations*

Nucleus

Example human cell

Cytoplasm



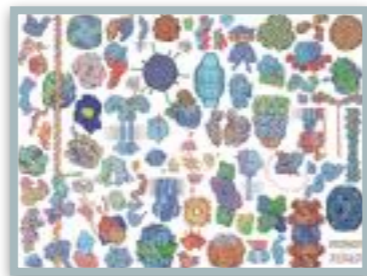
3 μ m

Plasma membrane

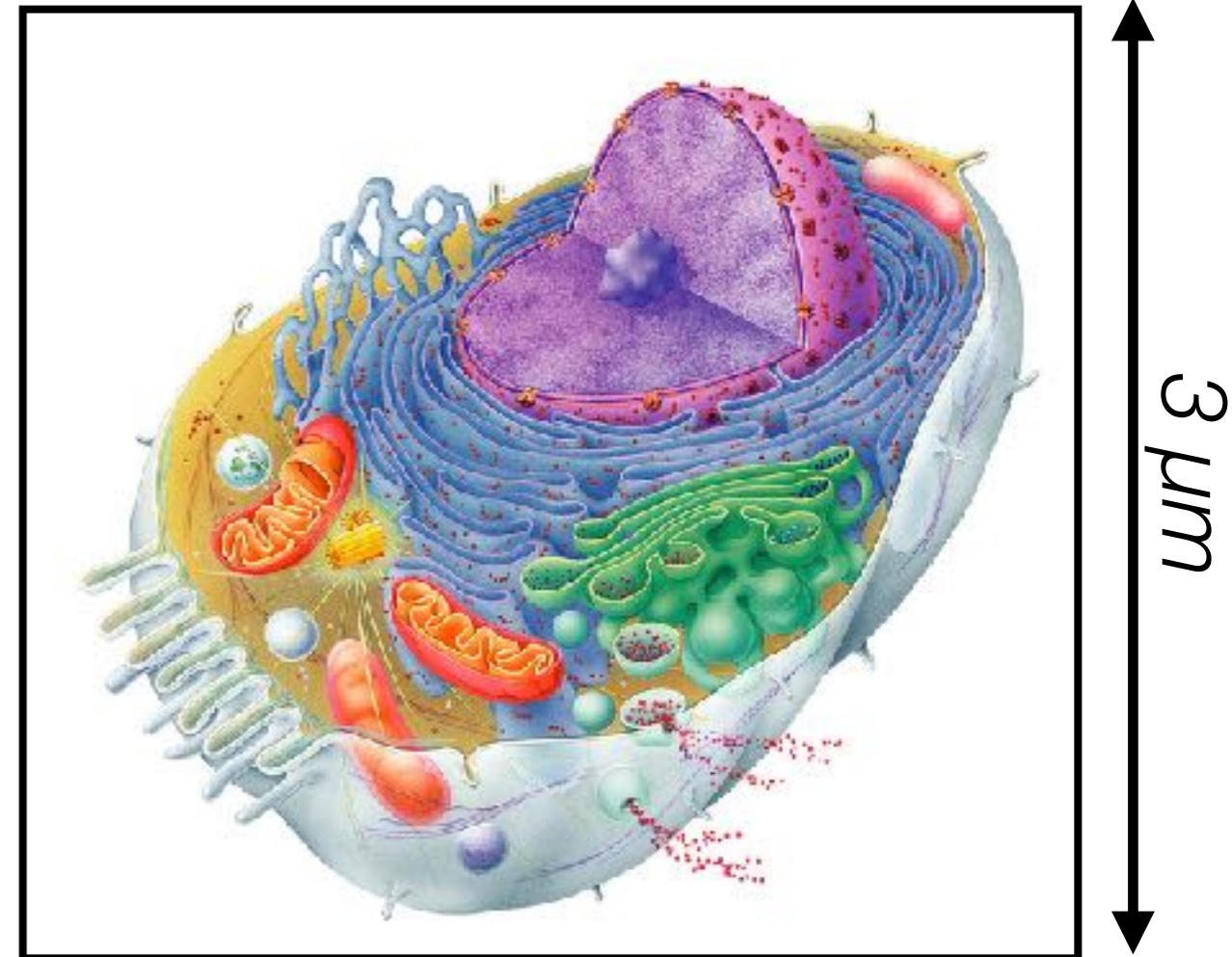


Understanding biology requires an understanding of the 'microcosmos'

Example subset of proteins (~100)



Example human cell

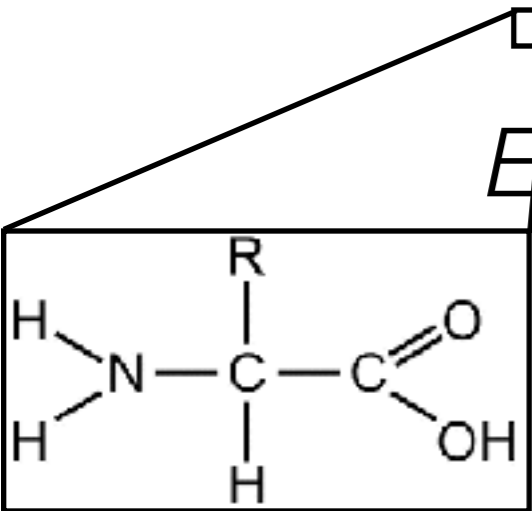


Understanding biology requires an understanding of the 'microcosmos'

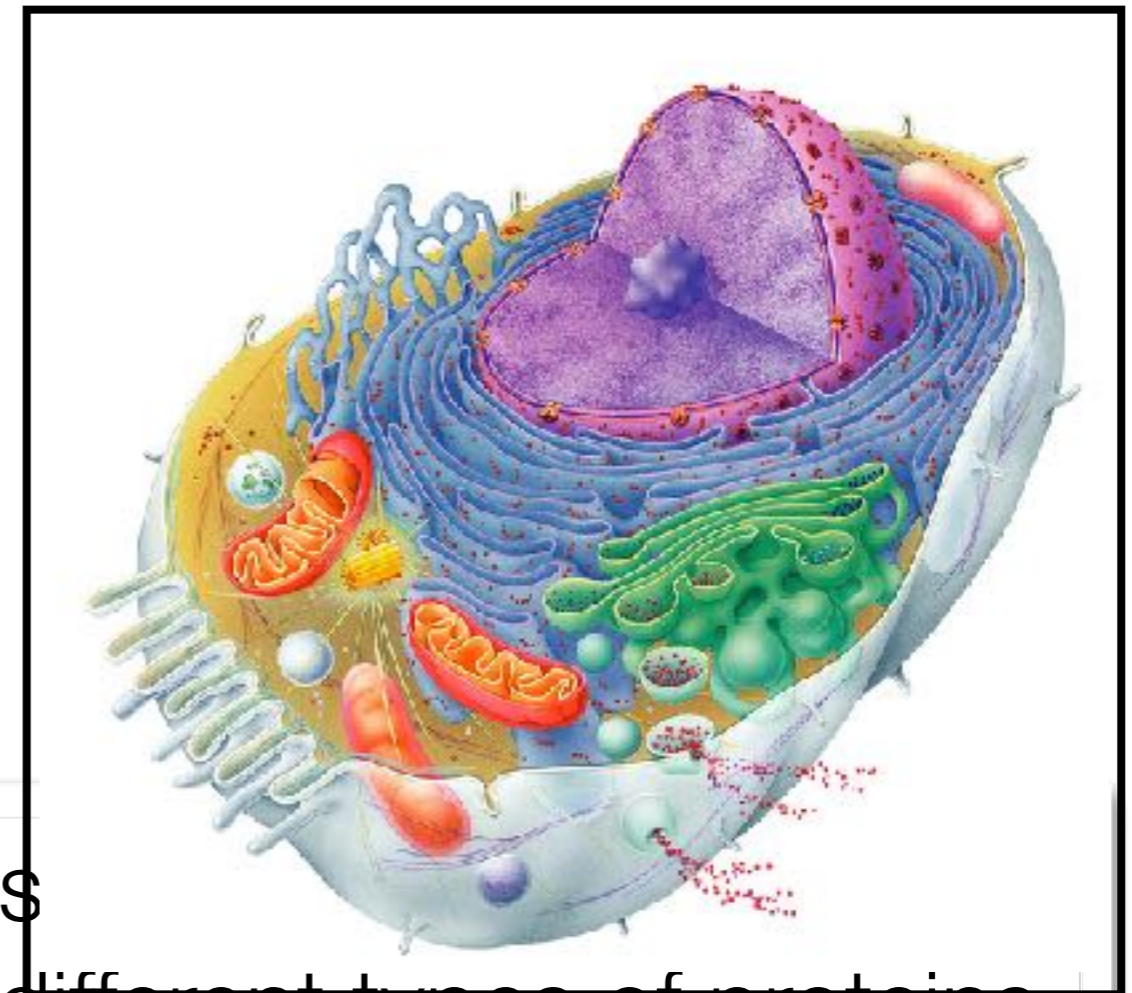
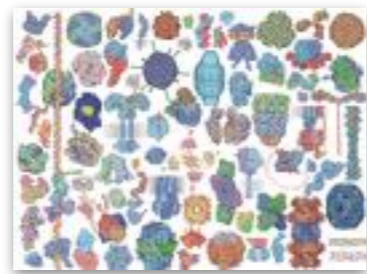
Diagram of ~100 proteins / molecules

Example human cell

Example subset of proteins (~100)



Amino acid



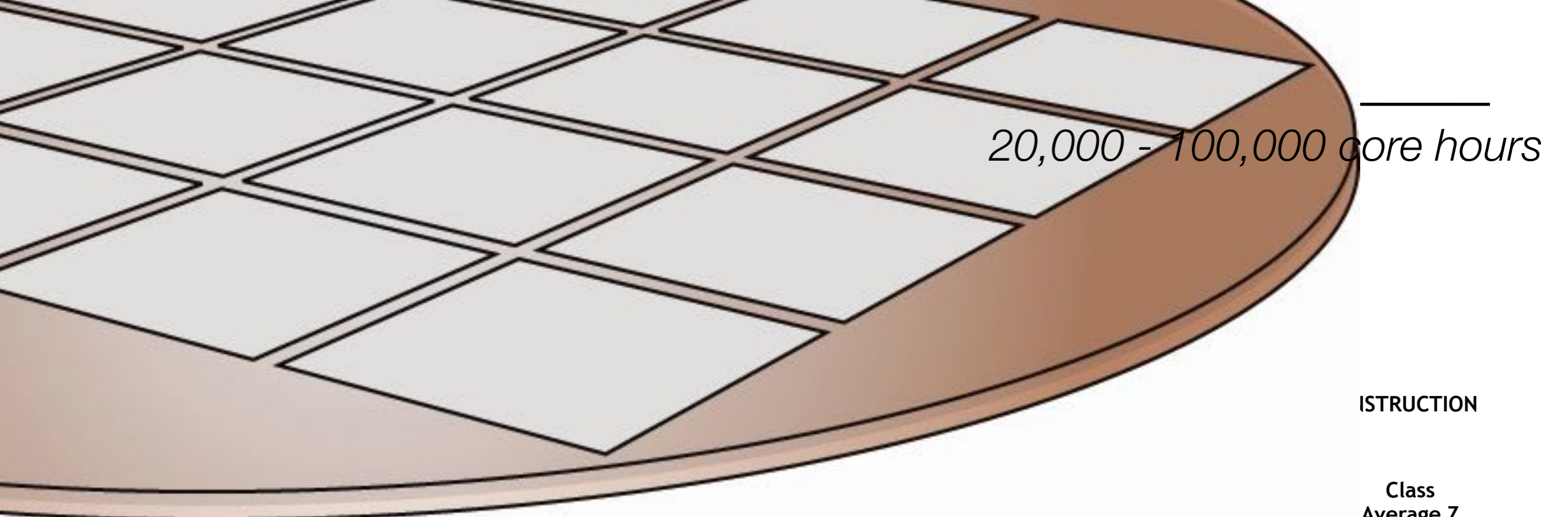
- ▶ Humans have 21,000 genes
- ▶ Predicted to form 100,000 different types of proteins

Proteins are linear chains of amino acids strung together (Usually > 300)

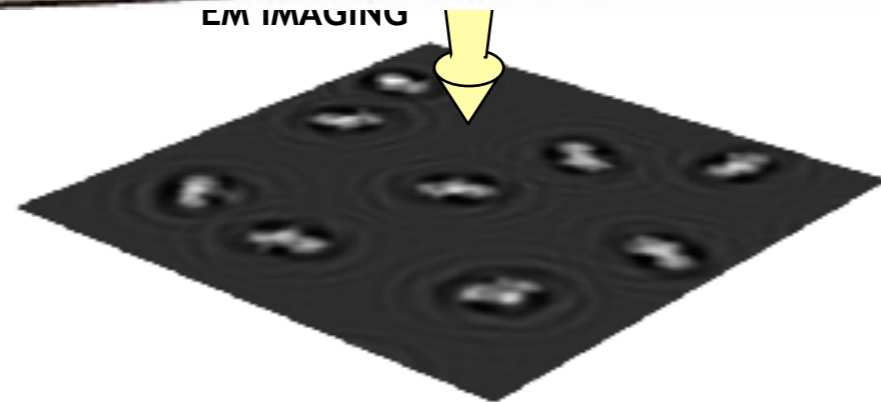
Cryo-EM utilizes transmission electron microscopes to take images of proteins



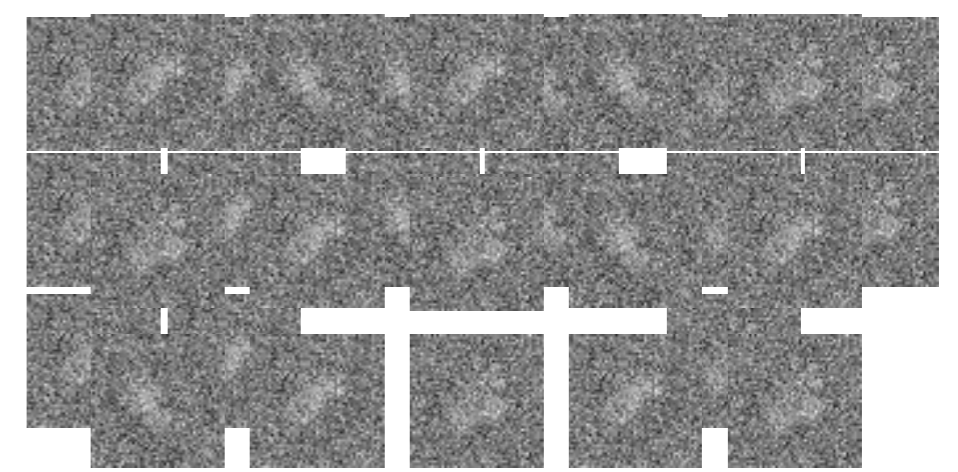
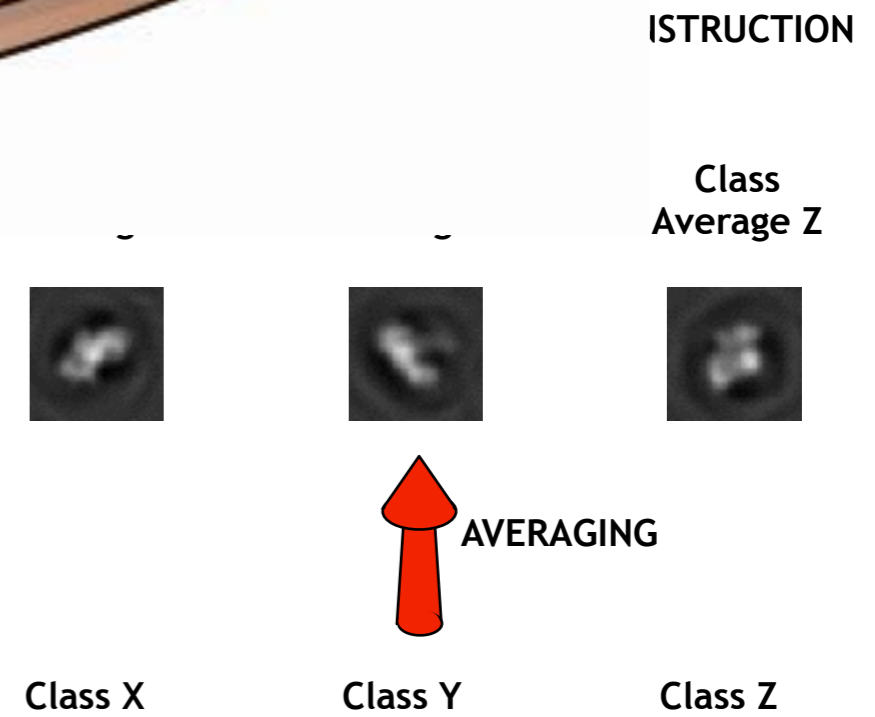
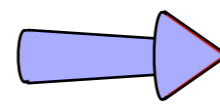
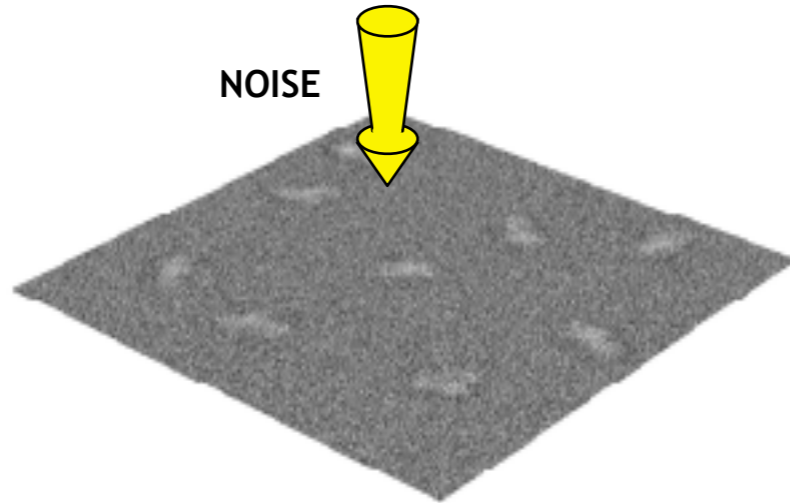
10th generation



EM IMAGING



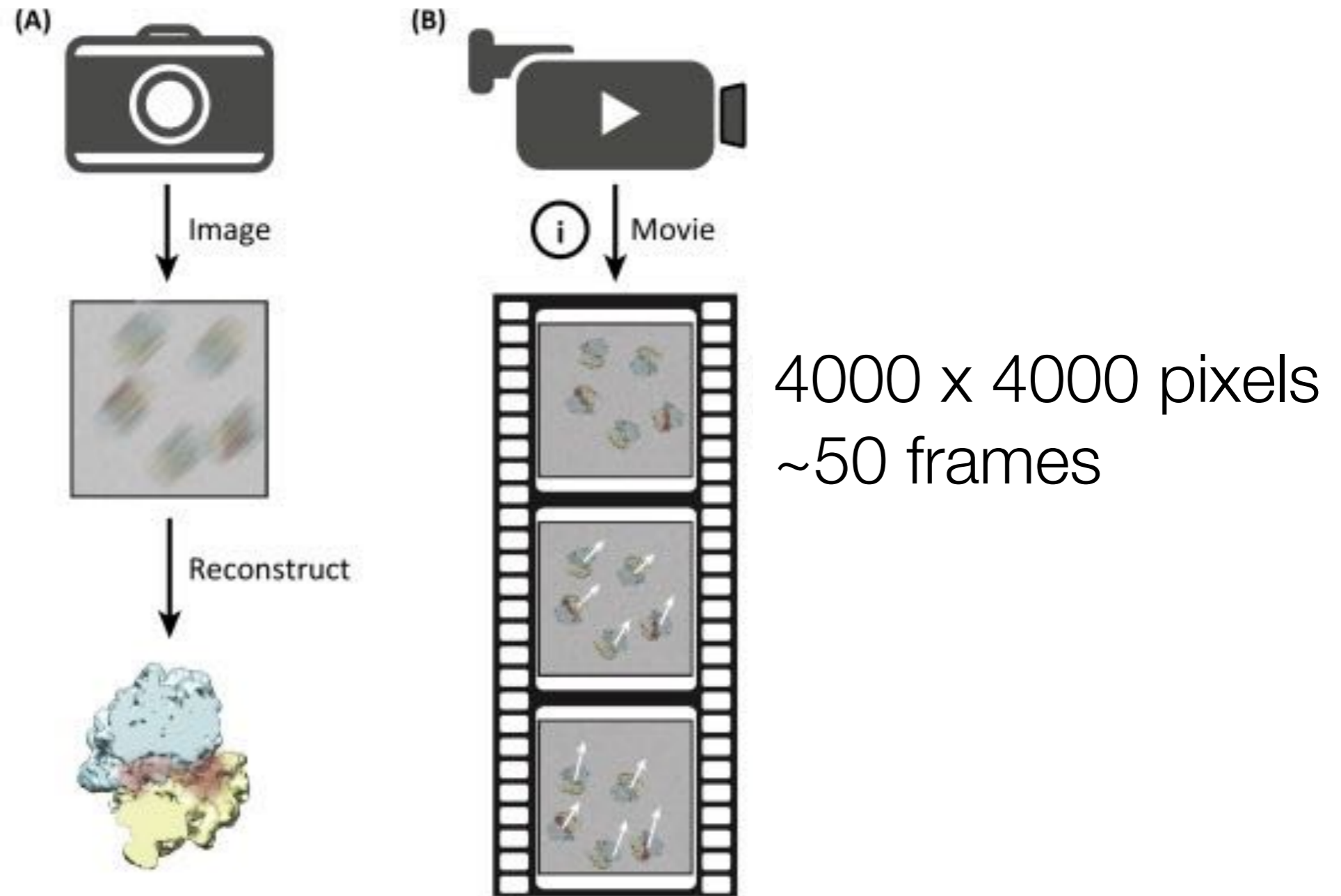
NOISE



CLASSIFICATION



Cryo-EM relies on movies instead of images of protein samples

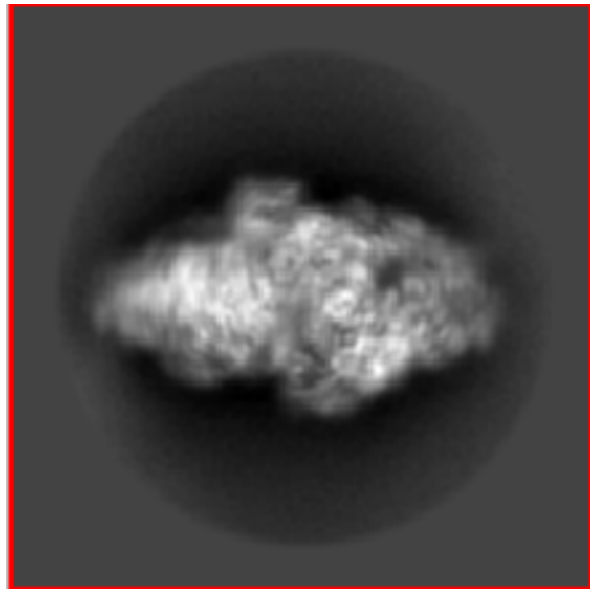


1-5 GB / movie

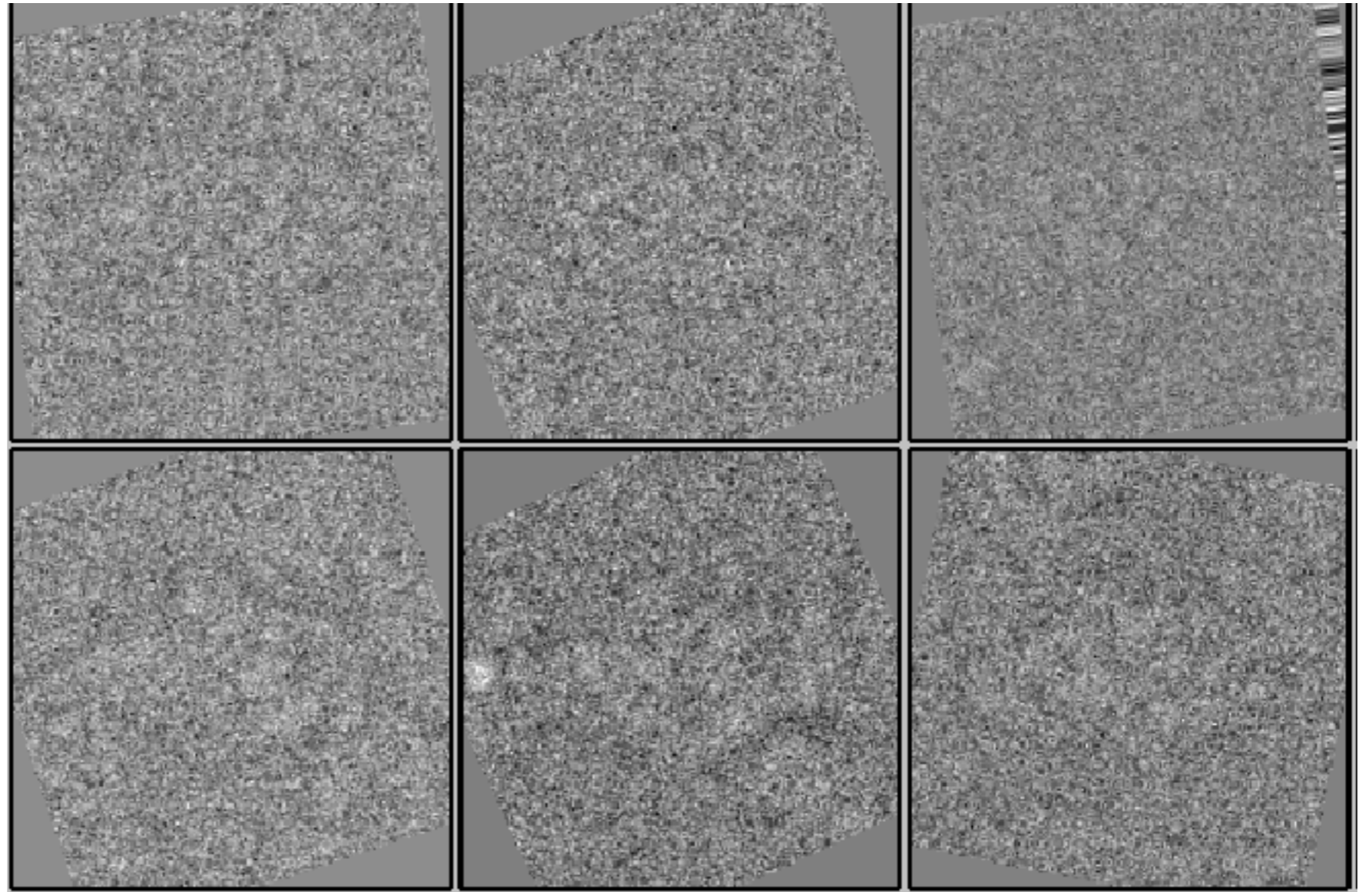
1 movie / minute (24/7)

Cryo-EM data is low SNR...

Grouped into classes and averaged to create averaged images:



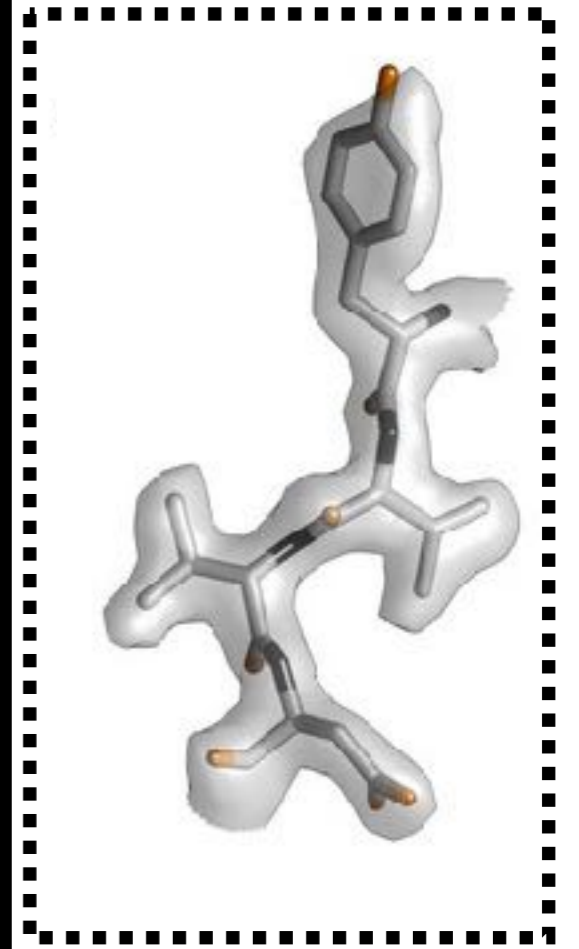
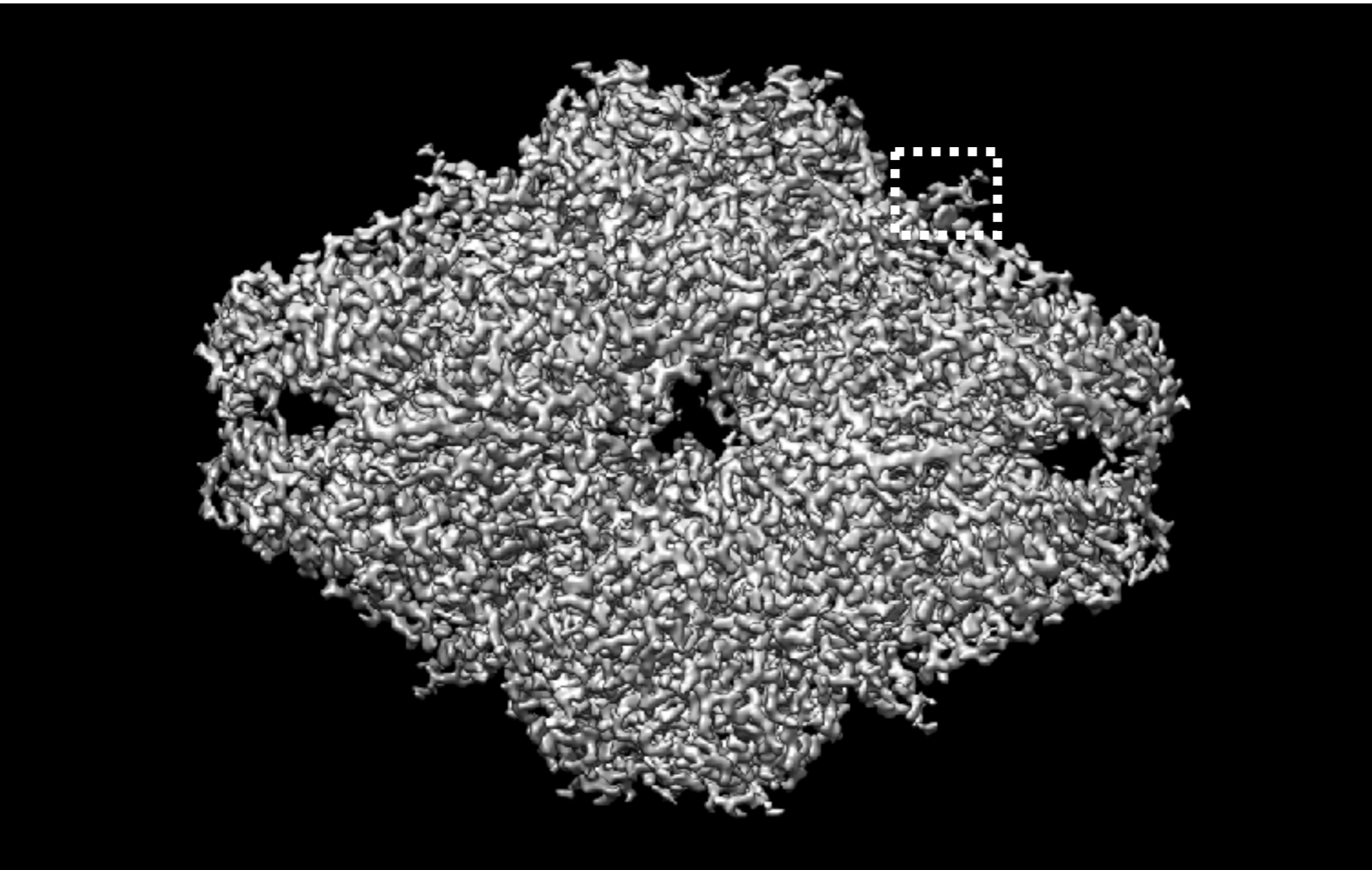
Averaged image



Individual particles in averaged image

...but can solve atomic protein structures

Combine data together in 3D dimensions



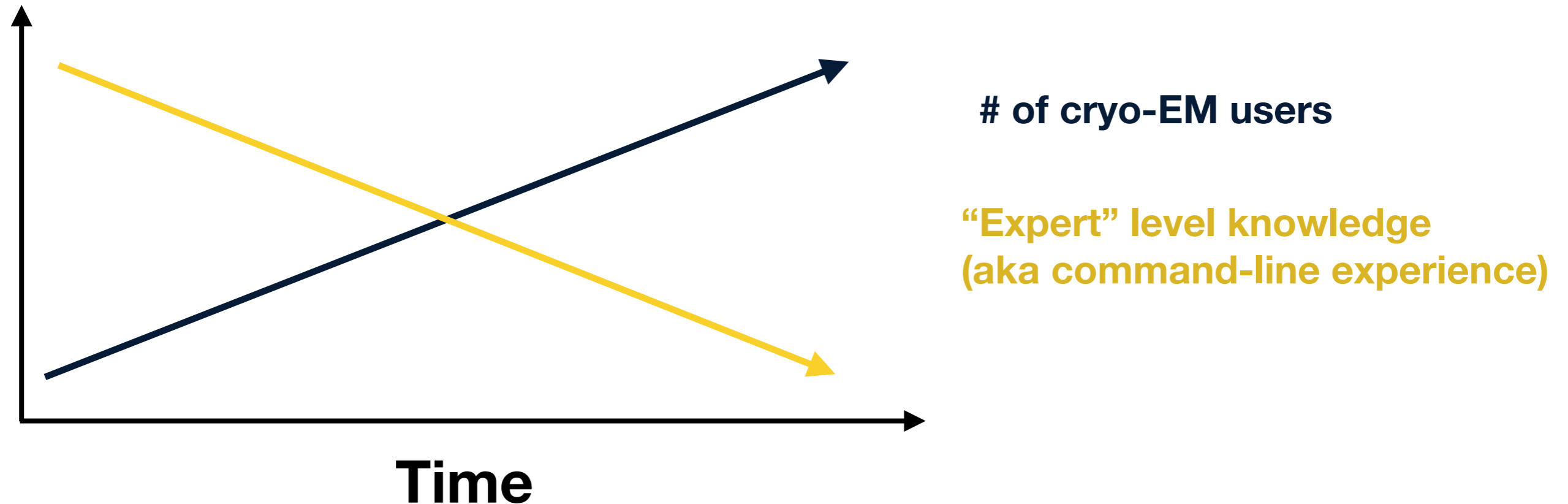
Atomic protein structure from ~400,000 images
(12 TB)

Allows atomic
model to be built

2017 Nobel Prize in Chemistry



Cryo-EM is undergoing a rapid expansion



Instrument collects ~2 TB / day (24/7)

-> Each individual requires collecting ~10 - 20 TB

~60 instruments in the US, growing quickly

Building a science gateway for cryo-EM

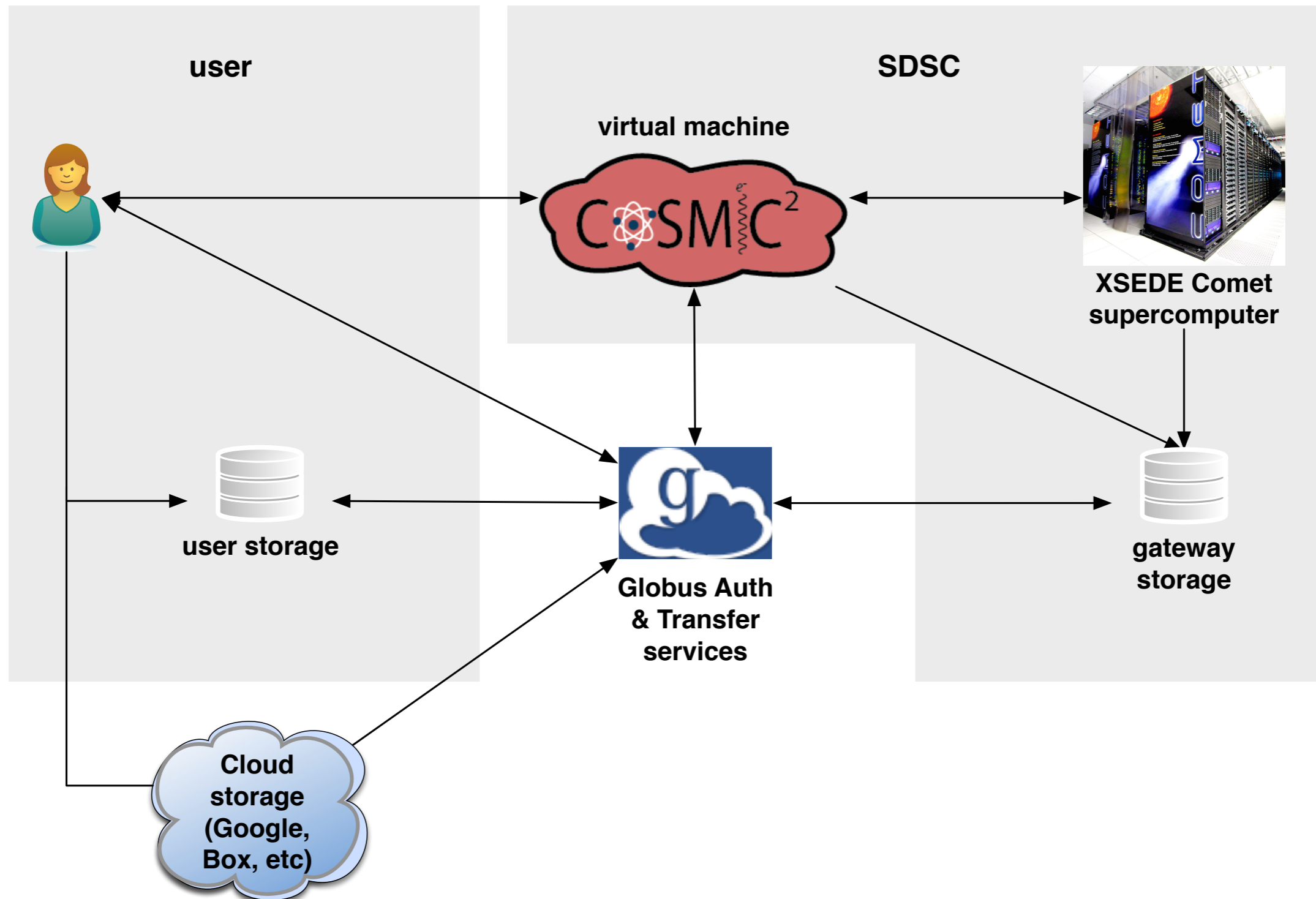
Short term:

1. Remove command-line interface for cryo-EM job submission
2. Remove decisions regarding HPC job environment
3. Create centralized location for data analysis software

Long term:

1. Connect users to cloud storage
2. Become platform for training and implementing advanced algorithm development
 - e.g. Machine learning / neural networks
3. Integrate educational materials to train next generation of structural biologists

Globus as a central hub for COSMIC²



Comet supercomputer at San Diego Supercomputer Center



46,656 CPUs

288 GPUs

XSEDE

Extreme Science and Engineering
Discovery Environment



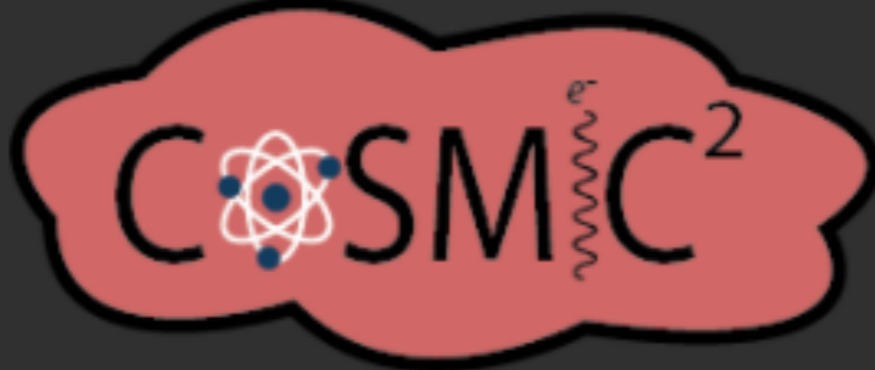
Computing allocations

Current allocation on COSMIC²: ~25,000 GPU-hours

All users: 500 GPU hours (free, no questions asked)

If you run out:

- Apply for supplement from COSMIC²
- Receive guidance on how to submit XSEDE computing allocation from COSMIC²



Science Gateway

Home

Toolkit

Help

How to Cite Us

Welcome to COSMIC²!

This is a *freely available*, science gateway for cryo-EM structure determination.

Please login below with your university credentials and then you are ready to go!

Login

Questions about this login process? [Please read more here.](#)



Log in to use COSMIC2

Use your existing organizational login

e.g., university, national lab, facility, project

University of Michigan

Didn't find your organization? Then use [Globus ID to sign in](#). ([What's this?](#))

Continue



Globus uses CI Logon to enable you to Log In from this organization. By clicking Continue, you agree to the [CI Logon privacy policy](#) and you agree to share your username, email address, and affiliation with CI Logon and Globus. You also agree for CI Logon to issue a certificate that allows Globus to act on your behalf.

Or







Sign in with Google



Sign in with ORCID iD

Enter your Login ID and Password

	<input type="text" value="mcianfro"/>	
	<input type="password" value="....."/>	
<input type="button" value="Log In"/>		
Need help?		


By your use of these resources, you agree to abide by [Responsible Use of Information Resources \(SPG 601.07\)](#), in addition to all relevant state and federal laws.




Folders

Total Storage: 0 bytes

Gateways 2017

 Data (2)

 Tasks (1)

Welcome

[Create New Folder](#)

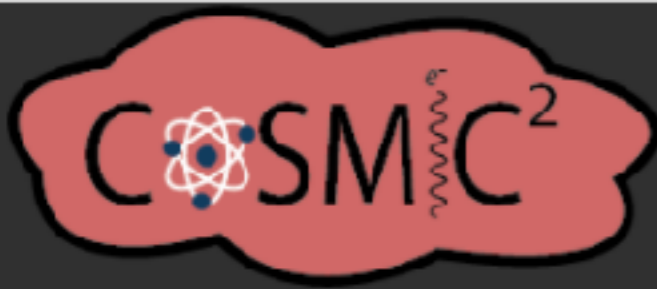
Current Folder Details

Label	Gateways 2017
Description	Demo project for Gateways 2017

[Create Subfolder](#)

[Edit Folder](#)

[Delete Folder](#)



Folders

Total Storage: 0 bytes

- First test
 - Data (1)
 - Tasks (1)

Enter name of
Globus endpoint



Globus Transfer Service

Endpoint Search:

Add

My Endpoints

Select your endpoint

Add

My Endpoints

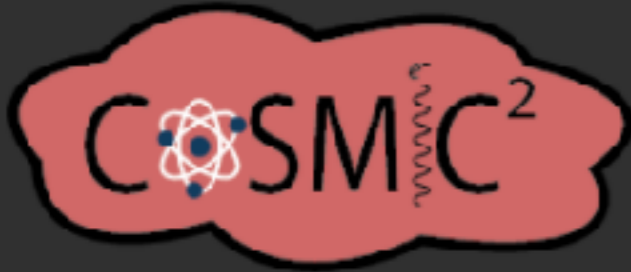
Please register your Globus Connect Personal Endpoints.

XSEDE Endpoint

Endpoint	List Files
COSMIC2 Science Gateway storage on XSEDE Comet	List

Transfer

Please register your Globus Endpoints.



Folders

Total Storage: 0 bytes

↳ **First test**

📁 **Data (1)**

📁 **Tasks (1)**

Find endpoint
in list



Globus Transfer Service

Endpoint Search:

LSI-CryoEM

LSI-XTAL

LSI-CryoEM

M Life Sciences Institute - Crystallography Endpoint

PI Life Sciences Institute - CryoEM Endpoint

My Endpoints

Select your endpoint

Add

XSEDE Endpoint

Endpoint

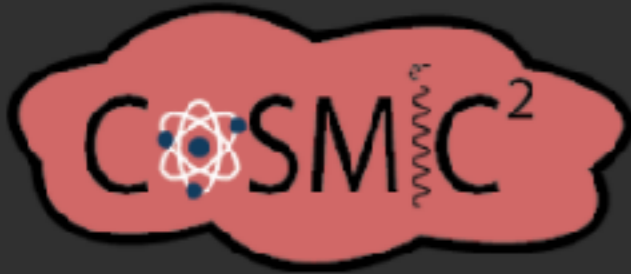
List Files

COSMIC2 Science Gateway storage on XSEDE Comet

List

Transfer

Please register your Globus Endpoints.



Folders

Total Storage: 0 bytes

➤ **First test**

📁 **Data (1)**

📁 **Tasks (1)**

Globus Transfer Service

Endpoint Search: <input type="text" value="LSI-CryoEM"/> Add	My Endpoints <input type="text" value="Select your endpoint"/> Add
---	---

My Endpoints

Endpoint	Path	List Files
LSI-CryoEM (Source)	<input type="text" value="/~/"/>	List Delete

XSEDE Endpoint

Endpoint	List Files
COSMIC2 Science Gateway storage on XSEDE Comet	List

↖ Enter directory name here

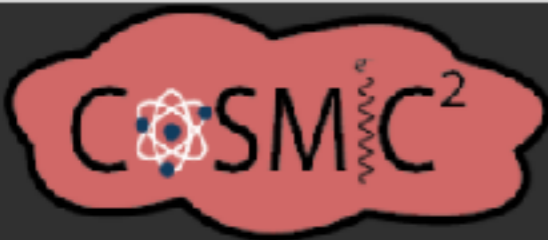
Transfer

[Switch Source and Destination](#)

Destination Endpoint: COSMIC2 Science Gateway storage on XSEDE Comet

Source Endpoint: LSI-CryoEM





Folders

Total Storage: 0 bytes

First test

Data (1)

Tasks (1)

Globus Transfer Service

Endpoint Search: <input type="text"/> Add	My Endpoints Select your endpoint <input type="text"/> Add
--	--

My Endpoints

Endpoint	Path	List Files
LSI-CryoEM (Source)	<input type="text" value="/ls/groups/mcianfro/colab/mcianfro/cosmic/"/>	List Delete

1. List directory by clicking 'List'

XSEDE Endpoint

Endpoint	List Files
COSMIC2 Science Gateway storage on XSEDE Comet	List

Transfer

[Switch Source and Destination](#)

Destination Endpoint: COSMIC2 Science Gateway storage on XSEDE Comet

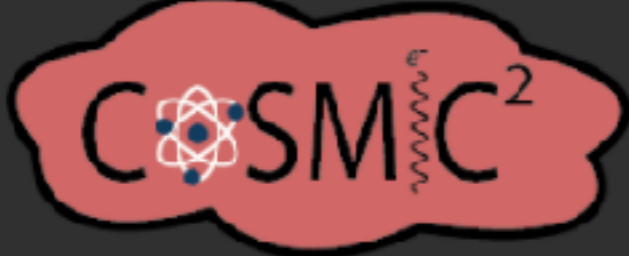
Source Endpoint: LSI-CryoEM

2. Select directory to be transferred

File / Folder	Size	Select
<input type="checkbox"/> Particles	60	<input checked="" type="checkbox"/>

[Transfer](#) ← 3. Transfer





Folders

Total Storage: 0 bytes

↳ **First test**

📁 **Data (1)**

📁 **Tasks (1)**

Transfer Status

Task ID: 59cccf76-b6c7-11e7-b0e8-22000a92623b

Source endpoint: LSI-CryoEM

Destination Endpoint: XSEDE Comet

Request Time: 2017-10-22 01:21:38+00:00 UTC

Completion Time:

Status: ACTIVE

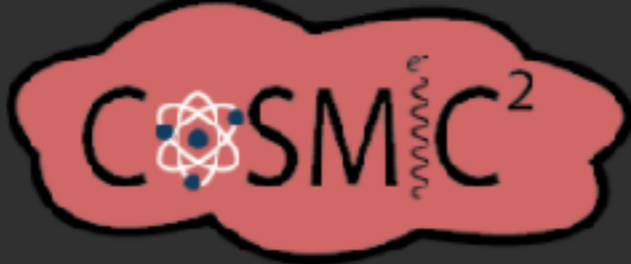
Files transferred: 0

Files skipped: 0

Bytes transferred: 0 B

Faults: 0

[Refresh](#)



Folders

Total Storage: 0 bytes

First test

Data (2)

Tasks (1)

All Data

Upload Relion directories, particle stacks, and 3D volumes using the Globus data transfer service.
Upload 3D volume and other small files (< 200 MB) using your browser.

- Globus upload
- Browser upload
- Collapse example

There is currently 1 data item in this folder.

20 records on each page

Data

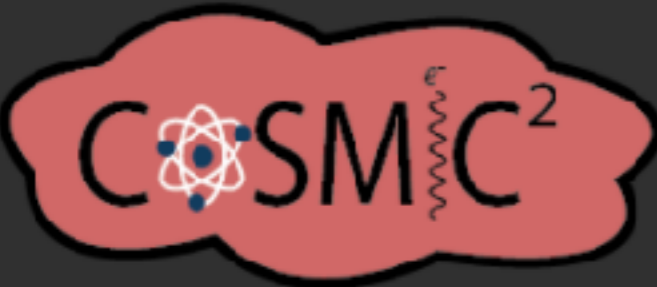
<input type="checkbox"/> Select all	Name	Bytes	Format	Date Created
<input type="checkbox"/>	betagaltest/particles.star	345460	STAR	10/9/17, 06:11
<input type="checkbox"/>	Particles/shiny_2sets.star	28378	STAR	10/21/17, 18:22



Move selected to First test GO

Delete Selected





- Folders**
Total Storage: 0 bytes
- Gateways 2017
 - Data (2)
 - Tasks (0)

Tool "RELION_2D_CLASS_COMET" successfully set to current task. ✕

Create new task

- Task Summary
- Select Data
- Select Tool
- Set Parameters

You may edit your task using the tabs above.
Current CPU Hr Usage: 0 [Explain this?](#)

Description	<input type="text" value="2D classification - 150 classes"/>
Input	<input type="button" value="1 Inputs Set"/>
Tool	<input type="button" value="Relion 2D classification"/> Click for more info
Input Parameters	<input type="button" value="14 Parameters Set"/> ← Click here to provide job parameters
<input type="button" value="Save Task"/> <input type="button" value="Save and Run Task"/> <input type="button" value="Discard Task"/>	

Saved tasks can be run later from the task list
XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.

Folders

Total Storage: 0 bytes

📁 Gateways 2017

📄 Data (2)

📄 Tasks (0)

Create new task

[Task Summary](#)[Select Data](#)[Select Tool](#)[Set Parameters](#)

Relion 2D classification: Calculate 2D class averages using Relion ([S. H. W. Scheres](#))

Simple Parameters

Number of classes *

Particle diameter (Angstroms) *

Pixel size of data (Angstroms/pixel) *

Number of iterations *

Output directory name *

In-plane angular sampling *

Pixel search range (pixels) *

Pixel search range step size, in pixels (offset_step) *

Advanced Parameters

Folders

Total Storage: 0 bytes

Gateways 2017

📁 Data (2)

📁 Tasks (0)

14 parameters successfully set to current task. ✕

Create new task

Task Summary

Select Data

Select Tool

Set Parameters

You may edit your task using the tabs above.
Current CPU Hr Usage: 0 [Explain this?](#)

Description

2D classification - 150 classes

Input

1 Inputs Set

Tool

Relion 2D classification

[Click for more info](#)

Input Parameters

15 Parameters Set

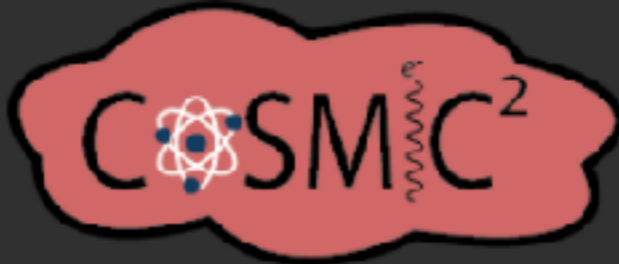
Run job!

Save Task

Save and Run Task

Discard Task

Saved tasks can be run later from the task list
XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.



Folders

Total Storage: 0 bytes

📁 Gateways 2017

📁 Data (2)

📁 Tasks (1)

Task Details

Task	2D classification - 150 classes
Owner	michaelc
Group	michaelc
Date Created	10/22/17, 16:45
Tool	Felion 2D classification
Input	View (1)
Parameters	View (14)
Output	View (2)
Intermediate Results	None
Status	COMPLETED

Task Messages

Sun Oct 22 16:49:08 PDT 2017 > QUEUE : SUCCESS : NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310 : Added to run queue.

Sun Oct 22 16:49:17 PDT 2017 > COMMANDRENDERING : SUCCESS : NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310 : Command rendered successfully: relion_refine_mpi --angpix 1 --offset_step 2 --psi_step 10 --itor 20 --particle_diameter 200 --K 150 --o Class2D --offset_range 5 --ctf --strict_highres_exp -1 --tau2_fudge 2 --i Particles/shiny_2sets.star

Sun Oct 22 16:49:17 PDT 2017 > INPUTSTAGING : SUCCESS : NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310 : Staging input files to COMET

Sun Oct 22 16:49:17 PDT 2017 > INPUTSTAGING : SUCCESS : NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310 : Input files staged successfully to /projects/cosmic2/gateway/workspace/NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310/

Sun Oct 22 16:49:28 PDT 2017 > SUBMITTED : SUCCESS : NGBW-JOB-RELION_2D_CLASS_COMET-9637B07DE52F4F7C9DD9055063697310 : Submitted to COMET as job '12082615'.

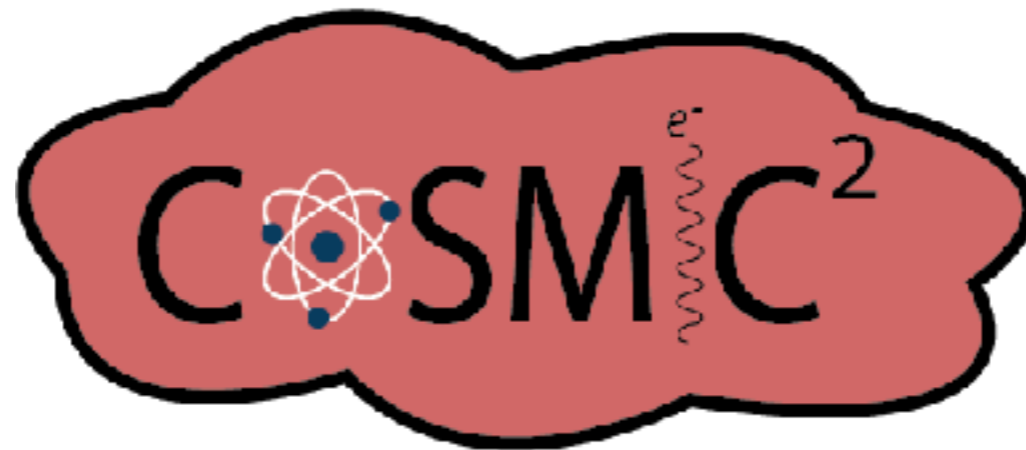
Lessons learned

- Globus integration was straightforward
- User account management & authentication
- Very easy data movement

Next steps:

Linking up to cloud storage

Moving data between NSF HPC resources using Globus



For more information

Website:

cosmic-cryoem.org

Github repo:

<https://github.com/cianfrocco-lab/COSMIC-CryoEM-Gateway>

PEARC17 paper:

<https://goo.gl/gnQ79a>

XSEDE

Extreme Science and Engineering
Discovery Environment

ECSS

SGCI

Science Gateways
Community Institute

