

STRUCTURAL BIOLOGISTS

ASSIGNEE

EDWARD T ENG

SIMONS ELECTRON MICROSCOPY CENTER



SIMONS ELECTRON MICROSCOPY CENTER







SIMONS ELECTRON MICROSCOPY CENTER

NEW YORK STRUCTURAL BIOLOGY CENTER Center Microscopy Electron Simons









Carl Negro



Kotaro Kelly Post Doc.



Yong Zi Tan Grad. Student



Post Doc.



Venkat Dandey Alex Noble Post Doc.



Bill Rice EM Manager Ed Eng

Ashleigh Raczkowski Staff Scientist Senior Technician





National Resource for Automated Molecular Microscopy http://nramm.nysbc.org





Alex Wei Technician

Zhening Zhang

Res. Scientist





Clint Potter Director



Bridget Carragher Director



Micah Rapp Grad Student









Kelsey Jordan Technician



Laura Kim Research Associate



Daija Bobe lechnician



Crystal Premo Administrator



What is possible? "LIFE IS REALLY SIMPLE, BUT WE **INSIST ON** MAKING IT COMPLICATED."

-CONFUCIUS (551-479 BCE)





What is possible today?

12h

*****2.5Å within a day





Is this routinely do Glutamate Aldolase dehydrogenase D2 D3 ~150kDa 334kDa rabbit muscle cow liver h

one?	Workflow validation/testing	
Apoferritin	20S proteasome	60S/80S ribosome
0	D7	C1
443kDa	750kDa	~2-4MDa
orse spleen	Thermoplasma or Mycoplasma	human





What type of computing challenge do you have?

Infrastructure to do cryo-EM processing for a research project/lab

Infrastructure to support a multi-user/ instrument EM facility









SAS/SATA/SSD bays

 8x hot-swap 3.5" drives bays + 3x fixed 5.2.5" drives bays + 1x fixed 3.5" drive bays

GPUs

 32GB DDR4 Memory installed (Supports up to 128GB DDR4; 256GB with optional Xeon Processor) Up to 4x NVIDIA Tesla Pascal

- GPUs
- Tower/4U Rackmount Convertible
- Up to 1.5TB DDR4 Memory Up to 8x NVIDIA Tesla Pascal 24x 2.5" hot-swappable
 - Up to 1024GB Memory Up to 4x NVIDIA Tesla Pascal GPUs 8 x SATA3 6Gb/s ports Optional:
 - SAS 6G RAID Card

working nplexes.

ded for

exxactcorp.com

sing

* > SYSTEMS > WORKSTATIONS > STRUCTURAL BIOLOGY GRID > SBGRID CYROEM WORKSTATION



thinkmate.com

5-2600 v4 AM) A Drive Trays	FORMANC STER NOD	Computation	
figured for nnig. NOW 25	of GB ECC DDR4 (up to 1TB)	Storage Computers Deals	cs • n History • Laptops
2.4	4x GPUs in 1U or 8x GPUs in 4U	Software	stern"
Zx ^z Ir	SSD nfiniband 56Gb	SBGrid Supported Software Computing Resources Get Help About SBGrid Join us Home Supported Applications	
ć	rackmount applications.	All Crystallography NMR Electron Microscopy Computational Chemistry Structure Visualization & Analysis Other Click to open explanation	anded vi
lepa	rticle.com	Search by title or developer: Filter by keywords: Filter by member type: available keywords - All - Active filters: Electron Microscopy x Clear all Software count 67 Filtering is on with any keywords.	



The challenge of cryo-EM computation



Infrastructure to do cryo-EM processing for a research project/lab

Infrastructure to support a multi-user/ instrument EM facility



The challenge of cryo-EM computation

How many instruments?

How many users?

What do they want to do?

What support would you like to provide?

Baldwin, et al. Current Opinion in Microbiology, Vol 43, 2017 (in press)

Infrastructure to support a multi-user/ instrument EM facility





400 registered users, 150 active Krios users









What software do they request to use?

What is asked	Support required
ELION / FREALIGN / oSPARC/EMAN2/etc	Single Particle Analysis
IMOD / Protomo / namo/PEET/PyTom/etc	. Tomography
Amira / IMOD / Dragonfly/etc	Segmentation/ Annotation
6	14% 3%
r/Novice	Intermediate Expert



How many instruments are used?

data generation at SEMC circa 2017





FEI Titan Krios#1 / #2 / #3 Falcon3 x3 K2 x3



FEI Tecnai F20 DE20 TVIPS 4K CMOS



FEI Tecnai Biotwin TVIPS 4K CMOS



JEOL 1230 Gatan US4000 CCD

> FEI Helios 650 ETD, TLD, ICE





How much data is generated?



Other scopes & CMOS/CCD

TEM Exposure images in 2015 & 2016:

1,069,315

#TEM Exposure images in 2017:

2,689,276

TEM Exposure images: 3,758,591*

*Total number of saved images since 2015: 766,329,392





The challenge of cryo-EM computation scalable

How many instruments?

How many users?

What do they want to do?

What support would you like to provide?

Not enough and getting more

Growing exponentially

Everything

As much as possible



SEMC solutions





The overall mission of NRAMM is to develop, test and apply technology for automating and streamlining cryo-electron microscopy (cryoEM) for structural biology.







NRAM



Image acquisition



Data processing









SEMC solutions



- HPC Server and storage (DDN):
- 2 x 42U rack enclosures
- DDN GRIDScaler GS7K appliance with 1.1PB GPFS paralegal file system
- 420TB DDN WOS object storage for archival
- 1056 x CPU cores. 44 x SuperMicro nodes each with 24 x CPU cores and 256GB RAM
- 4 x GPU nodes each with one GPU and 128GB RAM. One GPU server with 8 x GPUs and 512GB RAM and 2 x GPU servers each with 4 x GPUs and 512GB RAM.
- 4 buffer servers each with 51TB local storage, 2 x GPUs, 128GB RAM and 10G Fiber Network cards.
- 5 x 36 QSFP port 56Gb/s FDR InfiniBand switches.
- Bright Cluster Manager
- Basic Onsite Support; 7x24 remote support



SEMC solutions



Central MySQL Database and web server Size of images: 158.06 TB 766,329,392 **# DB records:**

Size of database: 7.44 GB

3,758,591 images

3,064 tilt series



since 2015

Appion

3/4 users who use Leginon also have Appion sessions

LEGINON





Example: Single-particle workflow **During EM session** After EM session SEMC Æ Appion Leginon computing session session Micrograph/ Setting up Frame alignment Particle sorting workflow Initial 2D Data acquisition **CTF** estimation classification Workflow Initial model Particle picking Optimization generation if needed Micrograph/ **3D** refinement Particle curating

Home institution/ cloud/SEMC 2D classification **3D** classification 3D refinement Model building



Example: Single-particle workflow

During EM session



After EM session SEMC 45 Home institution/ computing cloud/SEMC 2D classification Appion RELION/cryoSPARC/ **3D** classification EMAN2/Xmipp/SPIDER/ IMAGIC/sparx/etc... VIPER/SIMPLE/SPARX/ cryoSPARC/RELION/ 3D refinement Optimod/EMAN2/etc... **RELION/FREALIGN/** cryoSPARC/EMAN2/Xmipp/ Model building IMAGIC/spider/etc...





What is the timeline?



cryoSPARC

RELION









What is the timeline?







RFI ION workstation A GeForce GTX TITAN X Pascal 2 x Ten-Core 2.20GHz 25MB Cache 8 x 32GB 2400MHz DDR4 1x180GB STA SSD, 1x750GB SATA SSD



The challenge of cryo-EM computation



Infrastructure to do cryo-EM processing for a research project/lab





"It does not matter how slowly you go as long as you do not stop."





