Automated systematic evaluation of cryo-EM specimens with SmartScope

Jonathan Bouvette
Ph.D, visiting fellow
National Institute of Environmental Health Sciences

April 6th 2022, NYSBC
CryoEM workflow

- Protein purification
- Grid preparation
- Screening
- Dataset collection
- Data processing
- Structure
CryoEM workflow

Sample optimization

- Multiple cycles are required to obtain a good sample
- Most projects require preparing and screening >100 grids
- Each grids take >30 min to screen
## Grid Screening
### Goal of a screening session

| Learn as much as possible about the specimen | Thorough sampling | Good grid? |
|--------------------------------------------|-------------------|------------|
| Freezing conditions                         | Different ice thickness | Where are the best areas? |
| Sample quality                              | Find what is good   | Enough for a dataset? |
|                                            | And what is bad    | Improvements?   |

Diagnose and plan the next optimization cycle
Ease the optimization process
Maximize dataset quality
CryoEM workflow
Weekly on the NIEHS Arctica

• 80-100 grids screened:
  – 30 hours of active screening
  – 10 hours of grid preparation

• ~4-7 grid collected:
  – 20 hours of active setup
  – 80 hours of collection
Grid Screening is repetitive

- Record atlas
- Save Image
- Choose Area
- Move stage
- Save Image
- Choose square
- Move stage
- Eucentric
- Save image
- Center on hole
- Autofocus
- Save image
- Star over
- 36000 x
Manual grid screening – Cutting corners to speed up

- Incomplete metadata
- Suboptimal images
- Hard to navigate the results
- Subjective sampling

Save Image
Choose Area
Move stage

Save Image
Choose square
Move stage
Eucentric

Save image
Star over

Record atlas

62 x

210 x

2300 x

36000 x

Center on hole
Autofocus
Goals

• Automate screening
• Provide good sampling
• Complete data
• Intuitive interface
SmartScope – Automated workflow overview
## SmartScope – Layered modular approach to area selection

| Finders                  | Classifiers                                      | Selectors                                     |
|--------------------------|--------------------------------------------------|-----------------------------------------------|
| Object detection         | Named labels                                    | Clustering                                    |
|                          | Finite number of categories                      | Tunable number of categories                  |

- **Finders**
  - Object detection

- **Classifiers**
  - Named labels
  - Finite number of categories

- **Selectors**
  - Clustering
  - Tunable number of categories
SmartScope – Layered approach to area selection

1. **ROIs Detection and classification**
   - Good □ □ □ □
   - Bad □ □ □ □
   - Cracked □ □ □ □
   - Partial □ □ □ □

2. **Cluster by area size**
   - Smallest □ □ □ □
   - Largest □ □ □ □

3. **Filtered clusters**
   - Smallest □ □ □ □
   - Largest □ □ □ □

4. **Selection from different clusters**
   - Good □ □ □ □
   - Bad □ □ □ □
   - Cracked □ □ □ □
   - Partial □ □ □ □
   - Queued □
SmartScope – Layered approach to area selection

TOIs Detection

Cluster by intensity

Group for BIS

Selection from different clusters

Target

Darkest Brightest

Darkest Brightest

Queued
SmartScope – Layered modular approach to area selection

Finders
- Object detection
  Can also act as a classifier
- RCNN square finder/classifier
- YOLO hole finder
- Binary square finder
- FFT hole finder
- Regular pattern

Classifiers
- Named labels
  Finite number of categories
- Flow-based square classifier
- RCNN square classifier

Selectors
- Clustering
  Tunable number of categories
- Area size clustering
- Signal intensity clustering

Create custom workflows
Add new methods as plugins
Web Interface

- Real-time tracking
- Microscope interaction
Web Interface

- Real-time tracking
- Microscope interaction
Web Interface

- Real-time tracking
- Microscope interaction
- Preprocessing
Web Interface

- Real-time tracking
- Microscope interaction
- Preprocessing
- Annotation
Supervised Automatic screening

Giving the users some freedom

- Change Label
- Modify selection
- Annotation
- Changing parameters

- Micrograph curation (still under work)
Automatic screening
Leveraging early metadata

Faster R-CNN architecture

Identify and classify

Training set:

~ 1500 labeled squares
Hole Finder

- YOLO-based architecture
- AI hole finder is being trained to find holes on multiple grid types.
- Currently 10,000 holes in the training set.
- Precision of 98%, 89% recall
  - Mean-average precision 87%
Screening statistics

The graph shows the relationship between the number of holes sampled and the time spent on the specimen. The data is divided into two groups: BIS and No BIS. The BIS group consists of 38 samples, while the No BIS group consists of 942 samples.

- **BIS** samples are represented by black circles, with a median line indicated in blue.
- **No BIS** samples are represented by gray circles.

Key points:
- The median number of holes sampled is indicated by a blue dashed line.
- The graph includes a legend showing the number of squares sampled, with 1 and 7 highlighted.

The Talos Arctica K2 detector is mentioned in the context of the data collection process.
Automatic data collection
Quick setup and high-resolution capabilities
Conclusions

• Automated screening procedure
  – Square finder and classifier
  – Hole finder
  – Clustering methods

• Interactive interface
  – Ability to choose and modify area selection
  – Easy result access and complete bookkeeping

• Data persistence and organization

• Fast data collection setup

• Overnight screening sessions
CryoEM workflow
Weekly at the NIEHS Arctica

• >120 80-100 grids screened:
  – 30 hours of active screening
  – Lightly supervised automatic screening
  – 10 hours of grid preparation

• ~4-7 grid collected:
  – <10 20-hours of active setup
  – >90 80 hours of collection
**Short term goals – More Flexibility with modular protocols**

**Protocol recipe**

| Magnification level | Acquisition method | Finder (1) | Classifier (0 or more) | Selectors (1 or more) |
|---------------------|---------------------|------------|------------------------|-----------------------|
|                     |                     |            |                        |                       |

- Allow easy addition of Finders, Classifiers, Selectors as external plugins.
- Add acquisition methods to the microscope interface also as plugins.
- Create protocols by mixing existing methods.

**Ease the integration of new workflows**

Sample variety: virions, filaments, cells

**Tomography**
Sample-specific navigation roadmap

1. Sample specific state selection
2. User annotation to drive the selection on-the-fly
3. Using preprocessing information as feedback to drive the selection
4. Train AI models to drive the selection and “learn” about the samples
Acknowledgements

Bartesaghi Lab:
Dr. Alberto Bartesaghi
Wendy Qinwen Huang

Copeland Lab:
Dr. William Copeland
Dr. Amanda Riccio

UNC CryoEM:
Dr. Joshua Strauss
Jared Peck

Funding:
NIEHS/NIH

Elizabeth Viverette
Dr. Venkata Dandey
Dr. Kedar Sharma
Dr. Mario J. Borgnia
Kevin John U. Butay
Dr. Kanda Borgognoni
Dr. Amanda Riccio