Compositional and conformational variability in CryoEM and CryoET: structural biology in context

Steve Ludtke

Baylor College of Medicine
Charles C. Bell Jr., Professor of Structural Biology
Biochemistry and Molecular Biology
Deputy Director, Advanced Technical Cores
Director, CryoEM/CryoET Core

NOTE: This talk makes extensive use of movies. The YouTube video is likely a better resource for the material.
Structural Biology

• NMR
  • local structure, specific distances, local dynamics

• X-ray Crystallography
  • high resolution structures, often non-native

• CryoEM
  • intermediate - high resolution structure, in-vitro flexibility

• CryoET + CryoFIB
  • low - intermediate resolution structure, 3-D variability, cellular context

• Super-resolution Fluorescence (dynamic localization/co-localization)
Structural Biology

- NMR
  - local structure, specific distances, local dynamics
- X-ray Crystallography
  - high resolution structures, often non-native
- CryoEM
  - intermediate - high resolution structure, in-vitro flexibility
- CryoET + CryoFIB
  - low - intermediate resolution structure, 3-D variability, cellular context
- Super-resolution Fluorescence (dynamic localization/co-localization)

Alphafold →
Structural Biology

- NMR
  - local structure, specific distances, local dynamics
- X-ray Crystallography
  - high resolution structures, often non-native
- CryoEM
  - intermediate - high resolution structure, in-vitro flexibility
- CryoET + CryoFIB
  - low - intermediate resolution structure, 3-D variability, cellular context
- Super-resolution Fluorescence (dynamic localization/co-localization)
Structural Biology

- NMR
  - local structure, specific distances, local dynamics

- X-ray Crystallography
  - high resolution structures, often non-native

- CryoEM
  - intermediate - high resolution structure, in-vitro flexibility
    - Still needed for info on complexes & flexibility/dynamics

  - CryoET + CryoFIB
    - Still needed to observe native structures, observe native assembly, etc.
    - low - intermediate resolution structure, 3-D variability, cellular context

- Super-resolution Fluorescence (dynamic localization/co-localization)

Alphafold — ok... it's pretty decent. Positive and negative impacts?
EMAN2 Deep Learning Strategies

- Cellular Annotation/Particle Picking
  - Identify localized features in images or tomograms
  - Convolutional neural network

- Deep Learning Gaussian Mixture Model
  - Particle Based Conformational and Compositional Variability
  - Conventional dense neural network (similar to autoencoder)
Comprehensive structure and functional adaptations of the yeast nuclear pore complex

Christopher W. Akey,1,18,* Digvijay Singh,2,18 Christina Ouch,1,3,18 Ignacia Echeverria,4,12,18 Ilona Nudelman,5 Joseph M. Varberg,9 Zulin Yu,9 Fei Fang,6 Yi Shi,6 Junjie Wang,7 Daniel Salzberg,4 Kangkang Song,3 Chen Xu,3 James C. Gumbart,8 Sergey Suslov,2 Jay Unruh,9 Sue L. Jaspersen,9,10 Brian T. Chait,7 Andrej Sali,4,13,14 Javier Fernandez-Martinez,5,16,17,* Steven J. Ludtke,11,* Elizabeth Villa,2,15,* and Michael P. Rout5,19,*

1Department of Physiology and Biophysics, Boston University School of Medicine, 700 Albany Street, Boston, MA 02118, USA
2Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA
3Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA
4Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA
5Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA
6Department of Cell Biology, University of Pittsburgh, Pittsburgh, PA, USA
7Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA
8School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
9Stowers Institute for Medical Research, Kansas City, MO, USA
10Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS, USA
11Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA
12Department of Cellular and Molecular Pharmacology, San Francisco, San Francisco, CA 94158, USA
13Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA 94158, USA
14Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158, USA
15Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92039, USA
16Present address: Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain
17Present address: Instituto Biofisika (UPV/EHU, CSIC), University of the Basque Country, 48940 Leioa, Spain
18These authors contributed equally
19Lead contact

Single Particle CryoEM

~26,000 NPCs
7-11 Å resolution
ThermoFisher Aquilos 2
Cryo FIB/SEM
with EasyLift

FIB = Focused Ion Beam
SEM = Scanning Electron Microscope

Used for:
• cutting 100-500 nm cellular lamella
• tissue lift-out
• automatic slice-and-view
• Cryo-SEM imaging
• Platinum sputter coater
• Platinum GIS deposition
• Gallium FIB

To be purchased this year:
• iFLM (widefield fluorescence)*
• EM-ICE high pressure freezer
S. cerevisiae Nuclear Pore Complex
~52 MDa, ~30 unique Nups
~550 total proteins in complex

+ recent developments
Electron cryo-tomography (cryoET)

Stage tilt angle: -60°

e-beam

image

tilt series

-60°

Courtesy J.M. Bell
Electron cryo-tomography (cryoET)

Stage tilt angle: -45°
Electron cryo-tomography (cryoET)

Stage tilt angle: -30°

e⁻ beam

tilt series

image

-60°
-45°
-30°

Courtesy J.M. Bell
Electron cryo-tomography (cryoET)

Stage tilt angle: 60°

e⁻ beam

image

tilt series

-60°
-45°
-30°
-15°
0°
15°
30°
45°
60°

Courtesy J.M. Bell
3D Fourier reconstruction

“Missing wedge”

Reconstructed Tomogram

**Fourier Volume**

**Inverse FFT**

Courtesy J.M. Bell
S. cerevisiae Nuclear Pore Complex
~52 MDa, ~30 unique Nups
~550 total proteins in complex
~1000 Tomograms
~1000 NPCs
Resolution gets "stuck" at ~30 Å
3 Orthogonal Views
Traditional Classification (subtomogram)
Fatty Acid Synthase

Back to Single Particle Analysis
Single View CryoEM Average
Fatty Acid Synthase, ~30 Å motion
We thank the NIH for its support: R01GM080139.

Gaussian representation

\[
map(x) = \sum A_j e^{-\frac{(x - p_j)^2}{\sigma_j^2}}
\]

- Adjustable complexity
- Easier to model continuous motion

Amplitude
Sigma
3D Position vector
Particle-projection comparison

Particle-projection

Fourier ring correlation (FRC)

Raw particle

Model projection

Target resolution
Feedforward Neural Network

input layer

hidden layer

hidden layer

output layer

64 weights

\[ a = \text{ReLU} \]
Feedforward Neural Network

100x100 pixel image = 10,000 neurons

10,000^2 connections = 10^8 weights per layer

... we want to operate on 4k x 4k x 1k tomograms!

... and it's extremely inefficient (no translational equivalence)
Using gradient from neural structure (conformation = 0) as input
L17-Depleted 50S Ribosomal Assembly Intermediates (EMPIAR-10076)

18 submitted maps

Using GMM on the Same Data
L17-Depleted 50S Ribosomal Intermediate
(EMPIAR-10076)
Precatalytic Spliceosome
(EMPIAR) 10180

327490 particles
~7 Å resolution

Acknowledgements

Irina Serysheva’s Lab:
- Guizhen Fan
- Mariah Baker
- Alexander Seryshev
- Raj Sharma
- Venkata Mallampalli

Matthew Baker

Medical School UTHealth at Houston

Steven Ludtke’s Lab:
- Muyuan Chen
- Zhao Wang

Baylor College of Medicine (Houston)

University of Rochester (Rochester, NY)
David Yule’s Lab:
- Lara Terry
- Vikas Arige

KU Leuven (Leuven, Belgium)
Geert Bultynck’s Lab:
- Hristina Ivanova
- Ian de Ridder

UTHealth Cryo-EM Core Facility: uthealth.corefacilities.org

Supported by NIGMS, NINDS, NSF, CPRIT, AHA & Welch Foundation
Inositol 1,4,5 - Trisphosphate Receptors

IP₃-gated Ca²⁺ Release Channels

- Expressed in virtually all eukaryotic cells
- Intracellular ion channels
- Response to many extracellular stimuli (hormones, growth factors, neurotransmitters, neurotrophins, odorants, light, and etc.)
- Ligand-gated ion channels: primary ligands - IP₃ & Ca²⁺
- Associate in vivo with multiple modulatory proteins (>100)

Slide courtesy of Irina Serysheva
Ca\textsuperscript{2+} - Dependent Activation of IP\textsubscript{3}R I


Slide courtesy of Irina Serysheva
Intrinsically Dynamic IP$_3$R1 Structure

- Deep-learning based analysis of structural variability from 2D cryo-EM images!

**Particles**
- 133k apo
- 133k Ca+IP3+ATP
- 133k high Ca

Slide courtesy of Irina Serysheva

2022, under review
Intrinsic flexibility of ARM2 domain

- Deep-leaning analysis revealed an extended-retracted motion of ARM2
Problems

• 5 parameter Gaussian representation -> large RAM, limited model size
• Gradients on noisy data -> poor latent space accuracy
• Subtomogram averaging
  • 2-D or 3-D representation?
  • Per-particle tilt high noise levels
• Requires large batch size (GPU RAM)
Deep Learning

d(score)/d(Gaussian parameters) for each particle

Conformation latent space

Gaussian parameters) for each particle

M x 5

1xN

M x 5

4

Gaussian

Delta Function

Projection

Fourier image

Particle image

(S/2) x S

Fourier ring correlation

1 x (S/2)

Score

Amplitude "gradient" only!
(matches finite difference)

Rotation matrix

Resolution Mismatch Allowed
Problems

- 5 parameter Gaussian representation -> large RAM, limited model size
- Gradients on noisy data -> poor latent space accuracy
- Subtomogram averaging
  - 2-D or 3-D representation? -> 2-D
  - Per-particle tilt high noise levels -> Subtilt series uses average gradient
  - Requires large batch size (GPU RAM) -> Solved with delta functions
Acknowledgements

Beckman Foundation: 2021FIB-41
CPRIT Core Award: RP190602
NIH: R01GM080139

also:
Erik Anderson
Phil Baldwin
Zhili Yu
Adam Fluty

Steve Ludtke
Zhao Wang
Sneka Raveendran
Isaac Forrester
Sana Qureshi
Robyn Leidel

Core Staff