Lecture IX: Sub-tomogram averaging, theory, approach and perspective

Misha Kudryashev

June 6th 2015

IWAIP-CryoEM Workshop: June 3-7, 2015, Beijing
Motivation for CET + STA

• Study structures of molecular complexes
• Preferably inside native context
• Preferably at high(er) resolution
• Observe different conformational / functional states
• *Quick and easy enough*
• *Reliably*
Outline

• Workflow of sub-tomogram averaging
• Applications of sub-tomogram averaging
• Challenges in STA
  – CTF correction
  – Impact of the reconstruction method
  – STA at subnanometer resolution
  – Classification
• Limits and perspective
Tomography

Resolution is/may be limited by:
1. Non-perfect alignment of tilt series
2. Contrast transfer function (CTF)
3. Noise
4. Angular sampling (Crowther criterion)
5. Missing wedge (resolution anisotropy)

Sub-volume averaging helps with 3-5

From: Lucic et al, Annu Rev Biochem, 2005
0. Generate tomograms
1. Pick particles from tomograms
2. Align each particle to a reference taking missing wedge into account

Cross-Correlation

\[ f(x) \times g(x) = FT^{-1} \left[ F(u) \cdot G^*(u) \right] \]

Cross-correlation of \( f \) with \( g \) in real space is slow. It can be done much faster by calculating their FFTs \( F \) and \( G \), taking the complex conjugate of \( G^* \), multiplying \( F \) with \( G^* \), and calculating the inverse FFT of the result.
Cross correlation in 2D

**STEPS:**
1. Fourier transforming 2 images
2. Point – multiplication of the FT of the first image with the complex conjugated FT of a second image
3. Finding the highest value in the CC result, calculating the shifts

Source: Mathworks website
Constrained cross-correlation

- During alignment of a **fully sampled reference** and a **partially sampled sub-tomogram** (with a missing wedge) cross-correlation is calculated in the Fourier space only for the components sampled in the sub-tomogram.

Subtomogram averaging workflow

1. Pick particles from tomograms
2. Align each particle to a reference taking missing wedge into account
3. Sum up the rotated particles in Fourier Space
4. Sum up Fourier weights and divide the (3) by this sum
5. Take the result of (4) as a new reference and repeat (2)-(4) till the process converges
6. Analyze the results

Three-dimensional Reconstruction and Averaging of 30 S Ribosomal Subunits of *Escherichia coli* from Electron Micrographs

V. KNAUER, R. HEGERL AND W. HOPPE

Max-Planck-Institut fuer Biochemie
Abteilung fuer Strukturforschung I
D-8033 Martinsried, Federal Republic of Germany

(Received 28 December 1981, and in revised form 30 August 1982)

From the micrographs of a tilt series, several particles of negatively stained 30 S ribosomal subunits of *Escherichia coli* were three-dimensionally reconstructed. Three of them showing similar orientation with respect to the supporting foil were averaged after alignment by newly developed three-dimensional correlation methods. As a main result we found a stained channel-like structure inside the particle. We tentatively propose that this corresponds, at least partially, to positively stained segments of the 16 S RNA.
Electron Tomography of Single Ice-Embedded Macromolecules: Three-Dimensional Alignment and Classification

Jochen Walz, Dieter Tiptke, Michael Nitsch, Abraham J. Koster, Reiner Hegerl, and Wolfgang Baumeister

Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, 82152 Martinsried, Germany

Received October 6, 1997

From 3-D reconstructions of automatically recorded tilt series of ice-embedded macromolecules, several hundred 3-D images of single particles can be extracted. Here we describe correlation-based techniques to align the particles with respect to translation and orientation in 3-D and the calculation of an averaged reconstruction after application of the correct weighting function to the particle projections. Multivariate statistical analysis and classification are applied to the set of three-dimensionally reconstructed particles to investigate interimage variations on the 3-D level.

FIG. 5. MSA was applied to the data set consisting of 307 reconstructed particles. (a) Three eigenvectors with the highest eigenvalues. For analysis, the x-y slices of the volumes were pasted to 2-D images with a size of 520 x 200 pixels. (b) To visualize characteristic features, the eigenvectors were scaled and added to and subtracted from the average. The scaling factors were chosen such that 80% of the particles were between the points in coordinate space represented by the importance images. (c) The data set was split into 6 classes using the coordinates with respect to the three eigenvectors shown in (a) for cluster analysis. The numbers of particles contributing to each class are shown below the isosurface representations of the class averages. For all isosurfaces, the threshold was set to include the mass of the molecule (630 kDa) at an estimated protein density of 1.3 g/cm³.
EMDB deposited maps on May 26\textsuperscript{th} 2015

http://www.ebi.ac.uk/pdbe/emdb/statistics_main.html/
Applications of sub-tomogram averaging
Structure and variations of something large in the cellular context

Structure of membrane/viral proteins in lipids/viruses

- Retrovirus envelope complex
  From: Forster et al, PNAS, 2005

- HIV gp120 with b12-FAB

- Bacterial multidrug efflux pump
  From: Trepout et al, BBA, 2010
Classification of conformational states *in situ*

Present and absent “C-ring” of bacterial flagellar motor

Structure of proteins with preferential orientation on the grid

Poorly ordered Myosin V crystals

Bovine Respirasome in detergent over carbon layer in preferential orientation
From: Dudkina et al, PNAS, 2011
“Template matching”

- From Lucic et al, 2005, Annu. Rev. Biochem
Combination of template matching and STA

Human Polysomes in situ
From: Brandt et al, Mol cell, 2010
STA followed by placing averages to the tomograms

Structure of something very small

Retroviral RNA Packaging Element, 43 kDa, 10 nm long
From: Miyazaki et al, JMB, 2010
STA average as an initial template for single particle refinement

GAG protein (HIV-1)
Challenges in SVA
STA history / before direct detectors

Don’t report resolution higher then first zero of your CTF unless you did CTF correction

Independent half set refinement ("gold standard refinement")

Dataset

Halfset 1
- Alignment averaging
- Average 1

Halfset 2
- Alignment averaging
- Average 2

Average
- Detect resolution

Iterative refinement Till convergence

Merge Estimate resolution Sharpen
Frequencies to use

• Safely use the frequencies up to $K_c$;

From: Lucic et al, Annu Rev Biochem, 2005
STA history / before direct detectors


<table>
<thead>
<tr>
<th></th>
<th>Correlation with publication year</th>
<th>Correlation with Resolution Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of particles</td>
<td>0.01</td>
<td>-0.28</td>
</tr>
<tr>
<td>Symmetry order</td>
<td>-0.15</td>
<td>0.21</td>
</tr>
<tr>
<td>Number of asymmetric units</td>
<td>0.02</td>
<td>-0.27</td>
</tr>
<tr>
<td>Number of classes</td>
<td>0.10</td>
<td>-0.14</td>
</tr>
<tr>
<td>Publication year</td>
<td>1.00</td>
<td>-0.13</td>
</tr>
<tr>
<td>Pixel size</td>
<td>-0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Camera size</td>
<td>0.25</td>
<td>0.05</td>
</tr>
<tr>
<td>Acceleration voltage</td>
<td>0.03</td>
<td>-0.27</td>
</tr>
<tr>
<td>Minimal underfocus</td>
<td>0.05</td>
<td>0.60</td>
</tr>
<tr>
<td>Presence of an energy imaging filter</td>
<td>0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>Angular coverage</td>
<td>0.19</td>
<td>-0.22</td>
</tr>
<tr>
<td>Electron dose</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>Ice thickness</td>
<td>0.14</td>
<td>0.35</td>
</tr>
<tr>
<td>Largest linear sample size</td>
<td>0.12</td>
<td>0.44</td>
</tr>
</tbody>
</table>
Focal pair tomography

-15 µm

-3 µm

Contrast transfer function in tomography

Difficulties:

- Tilted samples have defocus gradient across the image, increasing with tilting angle
- Nominal defocus varies during tilting
- Electron dose is spread over tilt series
- Sample thickness adds defocus gradient across the sample
- Simple model of CTF is not applicable, more complex mathematics is needed (TCTF, Philippsen, et al, Ultramicroscopy, 2007)

From: Fernandez, Li and Crowther, JSB, 2006
Fitting CTF on a projection with DD

1 e/A², untilted image, field of view 600 x 600 nm², ~4 microns underfocus, no carbon, recorded on Titan Krios and K2 Summit

CTPlotter from Imod
Xiong et al, 2009, J Struct Biol
Impact of tomographic reconstruction method on STA
Weighted back projection vs iterative methods

- from Chen and Forster, 2014, J Struct Biol

(a) WBP

(b) Iterative Nonuniform fast Fourier transform based Reconstruction method (INFR)
SART vs WBP

Kunz and Frangakis, 2014, J Struct Biol
Subnanometer STA
Refinement of tomographic geometry for resolution improvement

- Constrained single particle tomography
- $\sim 100 \text{ e/A}^2 @ 80 \text{ kV}$ on CCD, 1.74 $\text{ A/pix}$, 2-3 microns defocus, $+/-45 \text{ deg}$ with 2 deg increment

Bartesaghi et al, Structure, 2012
Refinement of tomographic geometry for resolution improvement

Bartesaghi et al, Structure, 2012

Figure 4. Progressive Improvement in Resolution Achieved by Each Component of Constrained Single-Particle Tomography and Comparison to Highest Resolution Map of GroEL Reported Using Cryo-EM
Maps are represented as iso-surfaces with the fitted X-ray coordinates.
(A) Reconstruction by conventional subvolume averaging.
(B–D) Fourier-based CTF-corrected reconstructions using only first 11 exposures in the tilt series and alignments from subvolume averaging (B), after traditional projection-matching refinement of image shifts (C), and after constrained projection-matching refinement (D).
(E–F) FSC plots of maps in (A)–(D) obtained from the correlation of reconstructions between random halves of the image data set, indicating resolutions measured by the 0.5 cutoff criteria of 24.5, 15.3, 10.8, and 8.4 Å, respectively.
(G–H) Maps of the entire complex corresponding to the subunits shown in (A)–(D).
(I–J) Maps (4.2 Å) of GroEL (EMGS E. coli) obtained by traditional single-particle cryo-EM (Juddie et al., 2008), using 20,401 particles, 25–38 e⁻/Å² and 300 kV imaging (iso-surfaces shown at suggested contour levels of 0.597).
• 102 tomograms acquired, 80 selected, 133 individual tubes
• CCD data @ 200 kV
• Particles divided into 2 individual datasets and processed separately
• 121,346 sub-tomograms with eventually applied C2 symmetry

Flexible fitting with molecular dynamics

Cryo-EM map of the *E. coli* ribosome at 6.7-Å resolution

Trabuco L et al, 2008, Structure
Classification

• PCA + K-Means
  – Calculate eigenvolumes to reduce the dimensionality
  – Separate particles according to eigenvalues

• MRA / ML
  – Generate M seeds and align each particle to each seed
  – Iterate till convergence

• Challenges
  – Risk to classify direction of missing wedge, defocus, etc
  – Computationally expensive
Limits and future technologies
Further improvements in resolution: phase plates

Volta Phase Plate: Danev et al, 2014, PNAS

Asano et al, 2015, Science
Limiting factors in atomic resolution cryo electron microscopy: No simple tricks

Xing Zhang, Z. Hong Zhou

Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, 237 BSRR, 615 Charles E. Young Dr. S., Los Angeles, CA 90095-7364, USA
California NanoSystems Institute, University of California, Los Angeles, Box 957227, 570 Westwood Plaza, Bldg 114, Los Angeles, CA 90095-7227, USA

ARTICLE INFO

Article history:
Received 22 November 2010
Received in revised form 3 May 2011
Accepted 5 May 2011
Available online 24 May 2011

Keywords:
cryoEM
Atomic resolution
Beam tilt
Dynamic scattering
Defocus gradient

ABSTRACT

To bring cryo electron microscopy (cryoEM) of large biological complexes to atomic resolution, several factors – in both cryoEM image acquisition and 3D reconstruction – that may be neglected at low resolution become significantly limiting. Here we present thorough analyses of four limiting factors; (a) electron-beam tilt, (b) inaccurate determination of defocus values, (c) focus gradient through particles, and (d) particularly for large particles, dynamic (multiple) scattering of electrons. We also propose strategies to cope with these factors; (a) the divergence and direction tilt components of electron-beam tilt could be reduced by maintaining parallel illumination and by using a coma-free alignment procedure, respectively. Moreover, the effect of all beam tilt components, including spiral tilt, could be eliminated by use of a spherical aberration corrector. (b) More accurate measurement of defocus value could be obtained by imaging areas adjacent to the target area at high electron dose and by measuring the image shift induced by tilting the electron beam. (c) Each known Fourier coefficient in the Fourier transform of a cryoEM image is the sum of two Fourier coefficients of the 3D structure, one on each of two curved ‘characteristic surfaces’ in 3D Fourier space. We describe a simple model-based iterative method that could recover these two Fourier coefficients on the two characteristic surfaces. (d) The effect of dynamic scattering could be corrected by deconvolution of a transfer function. These analyses and our proposed strategies offer useful guidance for future experimental designs targeting atomic resolution cryoEM reconstruction.

© 2011 Elsevier Inc. All rights reserved.
1. Accuracy of defocus determination

- Use higher dose per micrograph
- Store focusing images

\[ d_{\text{max}} = \sqrt{2\varepsilon\lambda} \]

- \( e \) – defocus variation
- Lambda – wavelength

<table>
<thead>
<tr>
<th>Res. (Å)</th>
<th>100 kV</th>
<th>200 kV</th>
<th>300 kV</th>
<th>400 kV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>54 Å</td>
<td>80 Å</td>
<td>102 Å</td>
<td>122 Å</td>
</tr>
<tr>
<td>3.0</td>
<td>122 Å</td>
<td>179 Å</td>
<td>228 Å</td>
<td>274 Å</td>
</tr>
<tr>
<td>4.0</td>
<td>216 Å</td>
<td>319 Å</td>
<td>406 Å</td>
<td>488 Å</td>
</tr>
<tr>
<td>7.0</td>
<td>662 Å</td>
<td>976 Å</td>
<td>1244 Å</td>
<td>1494 Å</td>
</tr>
</tbody>
</table>
2. Gradient defocus across the sample

- Limits the resolution
- Need to correct for wave change from top to bottom of the sample

\[ d \geq \sqrt{0.714D\lambda} \quad \text{and} \quad D \leq 1.4 \frac{d^2}{\lambda} \]

<table>
<thead>
<tr>
<th>Dia. (Å)</th>
<th>100 kV</th>
<th>200 kV</th>
<th>300 kV</th>
<th>400 kV</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>3.33 Å</td>
<td>2.74 Å</td>
<td>2.43 Å</td>
<td>2.22 Å</td>
</tr>
<tr>
<td>600</td>
<td>4.71 Å</td>
<td>3.88 Å</td>
<td>3.44 Å</td>
<td>3.14 Å</td>
</tr>
<tr>
<td>900</td>
<td>5.77 Å</td>
<td>4.75 Å</td>
<td>4.21 Å</td>
<td>3.84 Å</td>
</tr>
<tr>
<td>1300</td>
<td>6.94 Å</td>
<td>5.71 Å</td>
<td>5.06 Å</td>
<td>4.61 Å</td>
</tr>
</tbody>
</table>

*Limitation based on formula by Spence (1988).*
3. Multiple (dynamic) scattering

- Adds another oscillating transfer function

\[
\text{DTF} = \frac{\sin(0.5\pi\lambda s^2 D)}{0.5\pi^2 s^2} = \lambda D \sin c(0.5\pi\lambda s^2 D)
\]

<table>
<thead>
<tr>
<th>Table 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution limit imposed by dynamic scattering.</td>
</tr>
<tr>
<td>Dia. (Å)</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>300</td>
</tr>
<tr>
<td>600</td>
</tr>
<tr>
<td>900</td>
</tr>
<tr>
<td>1300</td>
</tr>
</tbody>
</table>
Limits of electron tomography

Electron tomography at 2.4-Ångström resolution

M. C. Scott, Chien-Chun Chen, Matthew Mecklenburg, Chun Zhu, Rui Xu, Peter Ercius, Ulrich Dahmen, B. C. Regan, & Jianwei Miao

Transmission electron microscopy is a powerful imaging tool that has found broad application in materials science, nanoscience and biology. With the introduction of aberration-corrected electron lenses, both the spatial resolution and the image quality in transmission electron microscopy have been significantly improved and resolution below 0.5 Ångstroms has been demonstrated. To reveal the three-dimensional (3D) structure of thin samples, electron tomography is the method of choice, with cubic-nanometre resolution currently achievable. Discrete tomography has recently been used to generate a 3D atomic reconstruction of a silver nanoparticle two to three nanometres in diameter, but this statistical method assumes prior knowledge of the particle’s lattice structure and requires that the atoms fit rigidly on that lattice. Here we report the experimental demonstration of a general electron tomography method that achieves atomic-scale resolution without initial assumptions about the sample structure. By combining a novel projection alignment and tomographic reconstruction method with scanning transmission electron microscopy, we have determined the 3D structure of an approximately ten-nanometre gold nanoparticle at 2.4 Ångström resolution. Although we cannot definitively locate all of the atoms inside the nanoparticle, individual atoms are observed in some regions of the particle and several grains are identified in three dimensions. The 3D surface morphology and internal lattice structure revealed are consistent with a distorted icosahedral multiply twinned particle. We anticipate that this general method can be applied not only to determine the 3D structure of nanomaterials at atomic-scale resolution, but also to improve the spatial resolution and image quality in other tomography fields.
Sub-Tomogram Averaging / Single particle analysis

• Applicable to protein complexes inside native context
• Every particle is 3D -> more stable convergence and classification
• Helps if isolated single particles / 2d crystals have preferred orientation
• Possible to insert average structure/classes back to the tomogram
• Allows observing structural variation in situ
• STA average can be used as a reference for single particle refinement

• Lower resolution (~2 nm)
• Challenging CTF correction
• 3D-volume processing requires missing wedge compensation and longer processing time
Thanks!
For the Dynamo tutorial tomorrow:

• Download and install a free trial version of Matlab (Linux, Mac)
• Download Dynamo from www.dynamo-em.org

• Questions please
  misha.kudryahsev@gmail.com
References:

• **Reviews:**
  - Limiting factors in single particle cryo electron tomography, M Kudryashev, D Castaño-Díez, H Stahlberg, Computational and structural biotechnology journal 1 (2), 1-6

• **Original articles:**
  - In situ structural analysis of the Yersinia enterocolitica injectisome M Kudryashev, et al, Elife 2, e00792
Some useful tricks for STA

Take 30 minutes to learn the basic tricks of Matlab

• Variables
• Loops
• Functions