

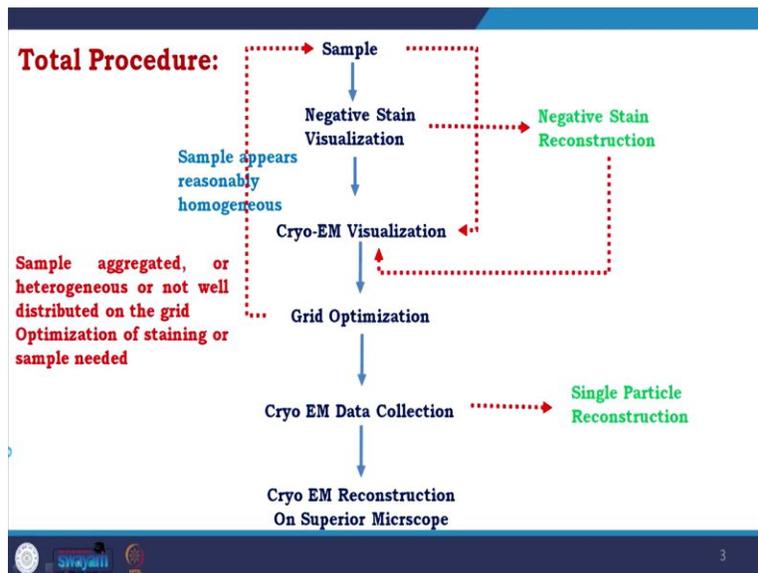
**Structural Biology**  
**Prof. Saugata Hazara**  
**Department of Biotechnology**  
**Indian Institute of Technology - Roorkee**

**Lecture – 38**  
**Cryo-electron Microscopy\_ Data Collection and Analysis**

Hi Everyone, welcome again to the course on Structural Biology. We are going for Structural Biology techniques. We have discussed high-resolution techniques like x-ray crystallography and NMR spectroscopy. Currently, we are on the module of cryo-electron microscopy. In the first class of the module, we have discussed the general principle of electron microscopy.

The second class talked about the general procedures and details of Cryo-electron microscopy. Today, we will discuss data collection and analysis, which are the core part of this technique. So, cryo-electron microscopy data collection and analysis:

**(Refer Slide Time: 01:20)**

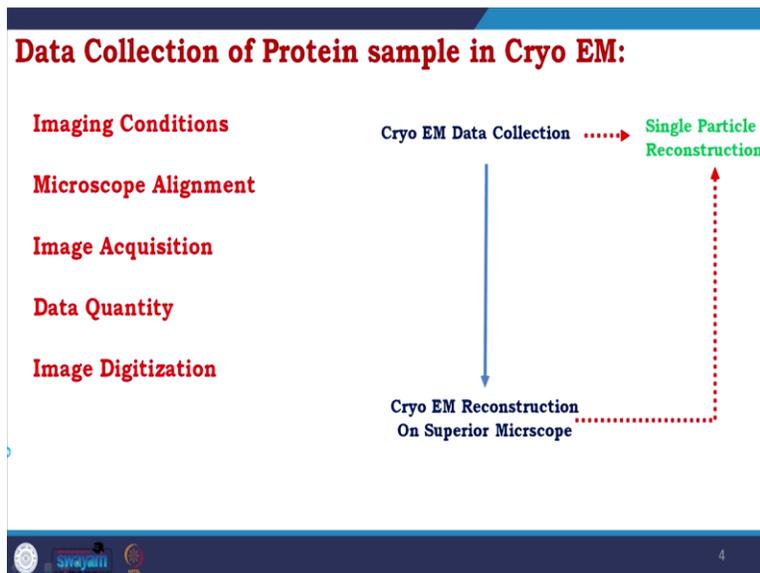


If you remember, in the last class, I have discussed the total procedure. We talked about the sample, how to prepare the sample, how to take care of the sample, then going for negative stain, visualization and then coming to Cryo-EM visualization, Grid Optimisation. Moreover, after Grid Optimization, you have to come to Cryo-EM data collection and then Cryo-EM reconstruction on Superior microscopes and the proxy dual-core.

When the sample appears reasonably homogeneous, you could also see the sample directly under Cryo-EM visualization escaping the negative visualization step. For some negative stain visualization, you could go to negative stain reconstruction and then come to Cryo-EM visualization, if the sample is aggregated or heterogeneous or not. Well distributed on the grid, or you need optimization of sample staining, you have to stop here and again go to sample preparation.

So Grid Optimisation up to that this is your checkpoints. You could do single-particle reconstruction from Cryo-EM Data Collection, you could do single-particle reconstruction, and then, from Cryo-EM reconstruction, you could go to single-particle reconstruction.

**(Refer Slide Time: 03:06)**



Today, we will talk about Cryo-EM data collection and many factors that influence the Cryo-EM data-collection, imaging condition, microscope alignment, image acquisition, data quantity, image digitization.

**(Refer Slide Time: 03:30)**

## Introduction to data collection:

Once protein density and ice thickness/quality have been optimized, an initial data collection should be performed on a mid-range electron microscope

The goal is to obtain 2D reference-free class averages which have high-resolution features at the level of secondary structure (alpha-helices or better)

Particles should be hand-picked, at least at first, to avoid the problems of template matching and "Einstein from noise" ([Henderson, 2013](#))

With modern field emission microscopes and direct detectors, only a few hundred particle images are needed to generate a 2D class of sufficient resolution to see alpha helices

Thus a dataset of a few thousand particles should be sufficient to yield a few high-resolution classes



So let us know about them. What are the procedures, and how do they influence the entire process? So, once the protein density and ice thickness of quality have been optimized. The protein makes it homogeneous, and the Ice should be transparent ice, which is called the vitrification process. An initial Data Collection should be performed on a mid-range electron microscope.

The goal is to obtain a 2D reference-free class average with high-resolution features at the level of secondary structures like alpha-helices or better. Particles should be handpicked at least at first to avoid the problem of template matching and "Einstein from noise", which is a criterion for Einstein overfitting. With modern field emission microscopes and direct detectors, only a few hundred particle images are needed to generate a 2D class of sufficient resolution to see the alpha-helices—only a few hundred with the modern field emission microscopes and definitely with the direct detectors. Thus a data set of a few thousand particles should be sufficient to yield a few high-resolution classes.

**(Refer Slide Time: 05:06)**

### Introduction to data collection:

If high-resolution 2D classes are not obtained, the specimen should be re-evaluated and optimized, including modification of protein preparations, buffers, detergents and surface modifications, and the addition of support films like amorphous carbon or graphene

Particles that don't align well may be partially or fully denatured and could lead to incorrect structures if the subsequent computation and validation is not done with great care

Many high-resolution structures to date have relied on discarding a large portion of the particles from the initial data collected

The proportion of discarded 'junk' particles can vary by orders of magnitude from one specimen to the next



If high-resolution 2D classes are not obtained, the specimens should be re-evaluated and optimized, including modification of protein preparation, buffer, detergent and surface modifications, and support films like amorphous carbon or graphene. So, if you come to that stage and do not get the high-resolution 2D classes, you have to re-evaluate the entire thing.

You have to optimize again and come back. Particles that do not align well may be partially or fully denatured and could lead to incorrect structures if the subsequent computation and validation are not done with great care. So, we have to see that while processing the sample, the protein would not be getting partially or fully denatured. To date, many high-resolution structures have relied on discovering a large portion of the particle from the initial data collected.

If you think that the data collected from the part of the protein which is denatured or not suitable, you should go and discard it. The proportion of discarded junk particles can vary by order of magnitude from one specimen to the next depending on the specimen's nature, stability, preparation, and many factors.

**(Refer Slide Time: 06:40)**

### **Introduction to data collection:**

We suggest that many, if not most of these particles are damaged during the sample preparation process

By evaluating specimen preparation using the above criteria, one can optimize the specimen and improve particle yield

This is generally preferable to collecting large datasets and discarding most particles to reach a desired resolution

Improving particle yield will make the entire process more efficient and more likely to succeed

Once a preliminary dataset is collected that generates suitable 2D classes, a larger dataset with a few hundred micrographs and several tens of thousands of particles is collected on a mid-range or high-end microscope, from which an initial 3D map can be calculated



7

So experts suggest that if not most of these particles are damaged during the sample preparation process, you could go for rejection by evaluating the specimen preparation. One can optimize the specimen and improve the particle yield; This is generally preferable to collecting large data sets and discarding most particles to reach the desired resolution. So you collect more and then your opportunity to discard. Improving particle yield will make the entire process more efficient and more likely to succeed towards the detailed analysis. Once a preliminary data set is collected that generates 2D classes, an extensive data set with a few hundred micrographs and several tens of thousands of particles are collected on a mid-range or high-end microscope from which an initial 3D map can be calculated.

**(Refer Slide Time: 07:53)**

### **Introduction to data collection:**

#### **Several important factors can then be evaluated:**

First, do the 2D classes show several distinct views of the particle, with self-consistent dimensions, each with secondary structural features?

Second, is the orientation distribution sufficient to allow calculation of a 3D structure with isotropic resolution?

If the orientation distribution is not suitable, one can alter buffers, add detergents, change plasma conditions or use an alternative surface

Generally people use graphene or amorphous carbon to improve the distribution and promote additional views of the complex

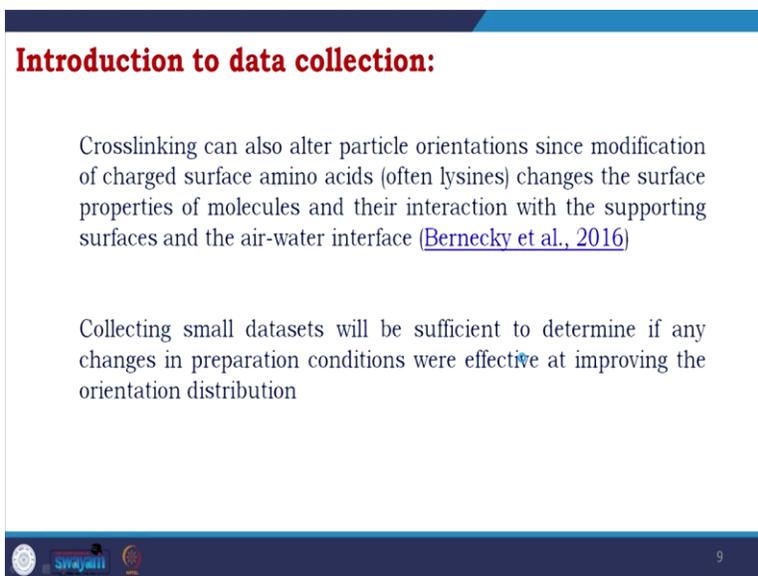


8

There are several important factors to be evaluated. First of all, do the 2D classes show several distinct views of a particle with self-consistent dimensions, each with secondary structural features? Second is the orientation distribution sufficient to calculate a 3D structure with isotropic resolution.

If the orientation distribution is unsuitable, one can alter buffers, add detergents change the plasma conditions or use alternative surfaces. Those are the options available. Generally, people use graphene or amorphous carbon to improve the distribution and promote additional complex views.

**(Refer Slide Time: 08:50)**



**Introduction to data collection:**

Crosslinking can also alter particle orientations since modification of charged surface amino acids (often lysines) changes the surface properties of molecules and their interaction with the supporting surfaces and the air-water interface ([Bernecky et al., 2016](#))

Collecting small datasets will be sufficient to determine if any changes in preparation conditions were effective at improving the orientation distribution

9

Cross-linking, which we discussed in the previous class, can also alter particle orientation by modifying charged surface amino acids. It is often applied to lysine, and it changes the surface properties of the molecule and their interaction with the supporting surface and the air-water interface. Collecting small data sets will be sufficient to determine if any changes in preparation conditions were effective at improving the orientation distribution.

So when we see that we have to check and go back, we should collect a small data set, not spending more time on the initial data collection.

**(Refer Slide Time: 09:32)**

## High-resolution data collection:

If the resolution of the initial 3D reconstruction reaches a reasonable resolution for the number of particles and image acquisition settings, a larger data collection on a high-end electron microscope should be performed with the aim of obtaining a high-resolution structure

What constitutes a “reasonable resolution” will be arbitrary and should ultimately depend on the resolution of the biological question and the state of the technology, but as a guideline, experts recommend, sub-nanometer resolution should be routinely possible from 10–50 thousand asymmetric particle images

With this in mind, a large dataset (500–1000 micrographs, ~24 hours of microscope time) should be collected with the best available microscope: Here example contains a 300 keV instrument with commercially available direct electron detectors (DEDs)

*What resolution is a good resolution!*

Suppose the initial 3D reconstruction resolution reaches a reasonable resolution for the number of particles and image acquisition setting. An extensive data collection on a high-end electron microscope should be performed to obtain a high-resolution structure. So again, we are repeatedly talking about it; when you are unsure about it, do it with a low-resolution microscope.

Take the minimum amount, then evaluate and when you reach the idea that you have a good sample, go for a high-end microscope, then go for a collection of large data set and all. What constitutes a reasonable resolution will be arbitrary. It should ultimately depend on the resolution of the biological question and the state-of-the-technology, but as a guideline, experts recommend sub-nanometer resolution from 10 to 50 thousand asymmetric particle images should be routinely possible.

When you are considering the possibility of optimizing, you will say, what is the resolution? What is a good resolution? There is no definite answer, which means the perfect answer is unavailable. It is dependent on the question you are asking. It is dependent on the setup you are working on and many other things.

A compromise is always required. With these in mind, a large data set of 500 to 1000 micrograms and 24 hours of microscope time should be collected with the best available microscope in these. We are talking about a 300 kV instrument with commercially available direct electron detector dealers.

(Refer Slide Time: 12:01)

**High-resolution data collection:**

The data collection strategy will depend on the size of the particle, the resolution desired and the particular microscopes available to the microscopist

All bio-molecular complexes exist in multiple conformational states, and for each to be resolved, even more high quality data are required

Many computational algorithms are available to help extract this information from a dataset

In some cases, ligands (e.g. small molecules, binding partners, nucleotides) can be added to stabilize particular conformations, or to interrogate a particular biological question related to how a ligand affects the conformational state of the complex

*Handwritten notes:* size ← B, Biological assembly, stability, Resolution required targeted, microscope setup, A, B, C, ligand, population A, Protein → (St), Protein + ligand/protein

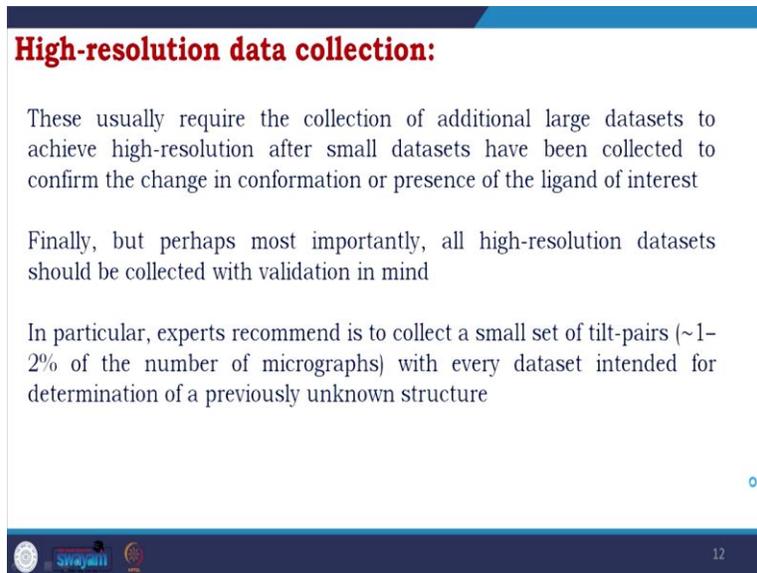
Today, data collections will depend on the size of the particle, the resolution desired, and the particular microscope available to the microscopist, so, basically, one the biological assembly. What is the size, and what is the stability? Talking about stability, that is the first thing. Second, the resolution you want. The resolution required or targeted. So this is A, and this is B. Then, the C is the microscope setup, which means the microscope available.

So, three factors influence biological assembly stability, the resolution required or targeted and microscope setup. All biomolecular complexes exist in multiple conformational states, and for each to be resolved, even more high-quality data is required. So as we are getting the opportunity to work with the native sample in its aqueous state, you could say that you will get the biological sample in different conformations and more diversity; more high-quality data is required.

Many computational algorithms help extract this information from a data set. So you have the data set, your algorithm will work and extract the information regarding the multiple conformations. In some cases, ligands, small molecule binding partners, nucleotides, cofactors can be added to stabilize particular confirmation or interrogate a particular biological question related to how a ligand affects the conformational state of the complex.

So there are different things. You have three conformations, and then you want to popularise, and you know that that would happen by binding this. So you use that ligand, then you popularize A and lesser the conformation, more accessible to get high-resolution data that is one. Second, one is the different problem: you have the protein, you already get the structure. You want to know the alteration when adding a ligand or a binding partner.

**(Refer Slide Time: 15:40)**



**High-resolution data collection:**

These usually require the collection of additional large datasets to achieve high-resolution after small datasets have been collected to confirm the change in conformation or presence of the ligand of interest

Finally, but perhaps most importantly, all high-resolution datasets should be collected with validation in mind

In particular, experts recommend is to collect a small set of tilt-pairs (~1-2% of the number of micrographs) with every dataset intended for determination of a previously unknown structure

12

So these things are possible through these techniques. These usually require collecting additional large data sets to achieve high resolution after collecting small data sets to confirm the change in conformation or presence of the ligand of interest. Finally, most importantly, all high-resolution data sets should be collected with validation in mind. In particular, experts recommend collecting a small set of tilt pairs 1 to 2% of the number of micrographs with every data set intended to determine a previously unknown structure.

**(Refer Slide Time: 16:21)**

## High-resolution data collection:

They can be subsequently used to validate the reconstructed density map ([Rosenthal and Henderson, 2003](#); [Henderson et al., 2011](#); [Baker et al., 2012](#); [Wasilewski and Rosenthal, 2014](#)), measure the angular accuracy of the projection assignments ([Russo and Passmore, 2014b](#)), and determine the absolute hand of the structure ([Rosenthal and Henderson, 2003](#))

By using the systematic approach described here, the microscopist will have the best chance of efficiently going from a protein in solution to a high-resolution structure

They can be subsequently used to validate the reconstructed density map. We have given many examples so that the interested reader should go through their literature. So you have this density map, and you have to validate the reconstructed density map to measure the angular accuracy of the projects assignment because I will talk about it in data analysis and determine the absolute hand of the structure.

Using the systematic approach described here, the microscopist will have the best chance of efficiently going from a protein in solution to a high-resolution structure.

(Refer Slide Time: 17:09)

Steps	Representative Data	Evaluation Criteria	Cryo EM Type
1 Protein Preparation		<b>Composition</b> <b>Purity</b> <b>Stability (complex type, buffer composition)</b> <b>Biochemical Activity</b>	
2 Negative stain		<b>Discrete Particles</b> <b>Stability</b> <b>Particle Size and Shape</b>	Entry Level
3 Diagnostic cryo-EM		<b>Stability</b> <b>Particle Size and Shape</b> <b>Particle Distribution Vs. Concentration</b>	Entry Level Mid Range
4 Initial cryo-EM data collection		<b>High Resolution 2D class</b> <b>Initial 3D model</b> <b>Orientation Distribution</b> <b>Particle Yield</b> <b>Tilt Pair/Validation</b>	Mid Range High End
5 High res cryo-EM data collection		<b>Motion Statistics</b> <b>Angular Accuracy</b> <b>Local Overall Resolution</b> <b>Conformational States</b>	High End

So I talked about the whole thing. However, here I am explaining the case by discussing the steps taken, the representative data, how the data would be evaluated, and the type of electron microscopy used. So, first is the sample preparation. You see the SDS gel, Sodium dodecyl sulphate gel, this polyacrylamide gel. They help get the molecular weight of the complex, and then we do different spectrophotometric studies to determine the presence of protein and protein concentration.

Here, the evaluation criteria would be the composition in the mixer, the purity of a single protein or anything used, the stability of the complex type, buffer composition, and then the biochemical activity. After the protein sample is prepared with a homogeneous and considerable purity, we have a negative stain with the biochemical activities in the second step. So you see how the particles look in the negative stain.

Here, we need to see the discrete particles. You will see the presence of discrete particles. Again you will evaluate the stability. You will measure the particle size and shape. The microscope required to measure those things is a basic entry-level transmission cryo-transmission electron microscope. Then going for Cryo-EM visualization is also called Diagnostic Cryo-EM.

Here, you will look for stability particle size and shape particle distribution versus concentration. It would help if you had an entry-level midrange Cryo-electron microscope for that. Then, initial Cryo-EM Data Collection, you see the particle distribution. You need a high-resolution 2D class initial 3D model that has to be developed from that. You are to get the orientation distribution, and you are to measure the particle yield.

To do that, you need a mid-range, high-end microscope and then high-resolution Cryo-EM data collection, which gives you the molecular level representation here. You see the ribosome structure, which is a favourite thing for the Cryo-electron microscope to work on. Here, you identify the different conformations, motion statistics, angular accuracy, local overall resolution, and conformational state, and you need a high-end microscope.

**(Refer Slide Time: 20:40)**

As with most technological devices, there is a tradeoff between ultimate performance of the current microscopes and other requirements

Like type of detectors, the amount of money and effort required to achieve this performance

But since the goal is ultimately to determine structures at resolutions which are sufficient to unambiguously answer biological questions, the compromises should be reasonable

Between the performance of the instrument and the amount and quality of data required to achieve a particular resolution that is appropriate for the stage of structure determination

As with most technological devices, there is a trade-off always between the ultimate performance of the current microscope and the other requirements, like the type of detector the amount of money and effort required to achieve this performance. However, since the goal is ultimately to determine a sufficient structured resolution to answer the biological questions unambiguously, the compromises should be reasonable.

There should always be a balance to get the biological information we are looking for at a high confidence level, between the instrument's performance and the amount and quality of data required to achieve a particular resolution appropriate for the stage of structure determination.

**(Refer Slide Time: 21:37)**

In general, we think it is better to spend more effort on preparing optimal specimens, as now, and even more so in the future, this is where the biggest differences in the resolution and interpretability of the resulting density maps are likely to come from

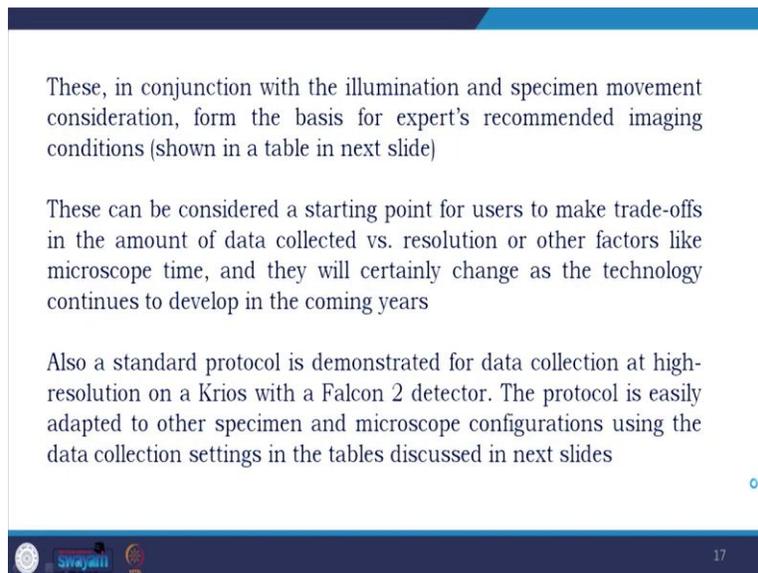
Since biological imaging is ultimately limited by damage to the specimen, currently the most important single factor governing the quality of the micrographs is the efficiency of the electron detector

Using the current understanding of detective quantum efficiency (DQE) and published measurements of DQE for the various commercially available detectors (McMullan et al., 2014; Chiu et al., 2015; Kujiper et al., 2015),

So trade-offs are still there. In general, we think it is better to spend more effort on preparing optimal specimens; that is, we must put more time on sample preparation as now and even more so in the future; this is where the most significant difference in the resolution and interpretability of the resulting density maps are likely to confirm. If your protein sample is homogeneous is stable, you always have an advantage.

Since biological imaging is ultimately limited by damage to the specimen, the most critical single factor governing the micrograph's quality is the electron detector's efficiency using the current understanding of detective Quantum efficiency DQE and published measurement of DQE for the various commercially available detectors.

**(Refer Slide Time: 22:38)**



These, in conjunction with the illumination and specimen movement consideration, form the basis for expert's recommended imaging conditions (shown in a table in next slide)

These can be considered a starting point for users to make trade-offs in the amount of data collected vs. resolution or other factors like microscope time, and they will certainly change as the technology continues to develop in the coming years

Also a standard protocol is demonstrated for data collection at high-resolution on a Krios with a Falcon 2 detector. The protocol is easily adapted to other specimen and microscope configurations using the data collection settings in the tables discussed in next slides

17

These, in conjunction with the illumination and specimen movement consideration, form the basis of experts who recommend imaging conditions, which we are talking about in the next slide, developed a table; This can be considered a starting point for the users to make trade-offs in the amount of data collected versus resolution or other factors like the microscope time, and they will undoubtedly change as the technology continues to develop the coming years.

As you see, we have understood the fundamental physics principles to improve this. Indeed, in the coming years, many more improvements will be possible. Also, here I am trying to show a standard protocol for data collection and high resolution on the Krios with the Falcon 2 detector.

The protocol is easily adapted to other placements we are looking for; it targets specific protocols, but it could be easily adapted to other specimens and microscope configurations using the data collection setting.

**(Refer Slide Time: 23:50)**

**General classes of Cryo Transmission Electron Microscopes:**

Microscope class	Typical examples	~ marginal cost/day (in 2016 €, including detectors)
Entry level	FEI T12, JEOL 1400	250
Mid-range	FEI F20 Talos, JEOL 2100F	600
Upper-mid-range	FEI F30 Polara, JEOL 3200FS	1000
High-end	FEI Titan Krios	3000

**Specimen support geometries for particular applications:**

Mode	Grid	Foil	Film
Negative stain	400 mesh	none	50-100 Å am-C
Diagnostic cryo	300 mesh	1.2/1.3 µm	none / 20 Å am-C
Medium-resolution (≥3.5 Å) cryo	300 mesh	1.2/1.3 µm	none / 20 Å am-C
High-resolution cryo (<3.5 Å) >400 kDa	300 mesh	1.2/1.3 µm	none / 20 Å am-C
High-resolution cryo (<3.5 Å) <400 kDa	300 mesh	1.2/1.3 µm	none / graphene
Very high-resolution cryo (<2.8 Å)	300 mesh	0.6/1.0 µm	none / graphene
Cryo-tomography: cellular (>30 Å)	200 mesh	2.0/2.0 µm *	none
Cryo-tomography: high-resolution sub-tomogram avg. (<15 Å)	300 mesh	1.2/1.3 µm	none / 20 Å am-C

We are going to discuss in table no. 1. The General classes of Cryo-Transmission electron microscopes: We are talking about the entry-level, mid-range, Upper-end range, and high-end levels. So Entry-level a typical example is FEI 12 JEOL 1400, and the marginal cost by day in 2016 is 250 Euro so for that the entry-level because you have to consider the cost also. In the mid-range FEI F20 Talos JEOL 2100F. The cost per day is 600 Euro.

In the Upper mid-range, the cost is 1000 per day, 1000 Euro, and for high-end FEI Titan Krios, we are talking about 3000 Euro per day. Specimens support geometries for particular applications; you need a 400 mesh grid for the negative stain. You do not need the foil and the film for 50 to 100 angstrom, and there are different modes as we have discussed in the figure before the negative stain, Diagnostic cryo, medium resolution, which is around 3.5 angstrom.

Very high-resolution cryotomography and cryo tomography with high-resolution tomograph averaging you have to have the grid, foil and film according to that.

**(Refer Slide Time: 25:43)**

**Currently used electron detectors for cryo-microscopy:**

Detector type	Current examples	Pixel pitch ( $\mu\text{m}$ )	Num. pixels	Frame rate (Hz)	Recommended flux (optimum [range] $\text{e}^-/\text{px}^2/\text{s}$ )
Phosphor - CCD	Gatan Orius 830	7.4	2048 $\times$ 2048	1	-
	Gatan US1000	14	2048 $\times$ 2048	1.5/15	-
Phosphor - CMOS	Tietz TemCam	15.6	4096 $\times$ 4096	1	-
	Gatan OneView	15	4096 $\times$ 4096	25	-
	FEI Ceta	14	4096 $\times$ 4096	32	-
Direct integrating	Direct EL DE16*	6.5	4096 $\times$ 4096	60	$\sim$ 240
	Direct EL DE20*	6.4	5120 $\times$ 3840	32	$\sim$ 100
Direct counting	FEI Falcon 2*	14	4096 $\times$ 4096	18	50 [10 – 60] (300 keV) 40 [8 – 48] (200 keV) <sup>†</sup>
	FEI Falcon 3*	14	4096 $\times$ 4096	32	100 [20 – 120] (300 keV)
	Gatan K2	5	3838 $\times$ 3710	400	5 [2 – 8] (300 keV)

\*These detectors can also be operated in electron counting mode but current frame rates make this impractical for normal data collection

†The flux,  $f_0$  at one energy can be scaled with reasonable accuracy to other energies relevant to transmission electron microscopy using the equation  $f_1 = f_0(\beta^2_1/\beta^2_0)$  where  $\beta$  is the ratio of the electron velocity to the speed of light

The detectors with different detector types like Phosphor CCD, Phosphor CMOS, direct integrating, and direct counting with a specific cryo-electron microscope have taken a huge and vital role in improving resolution in the electron microscope.

(Refer Slide Time: 26:09)

**Protocol for High-resolution data collection using a Krios / Falcon 2 and all-gold supports:**

1. Load specimens in cartridges and mount in autoloader, including a calibration grid (graphitized carbon or similar)
2. Discard any that are bent or broken during handling
3. Load a specimen in the column and use low mag to check grid is intact and contains several squares with an appropriate thickness of ice

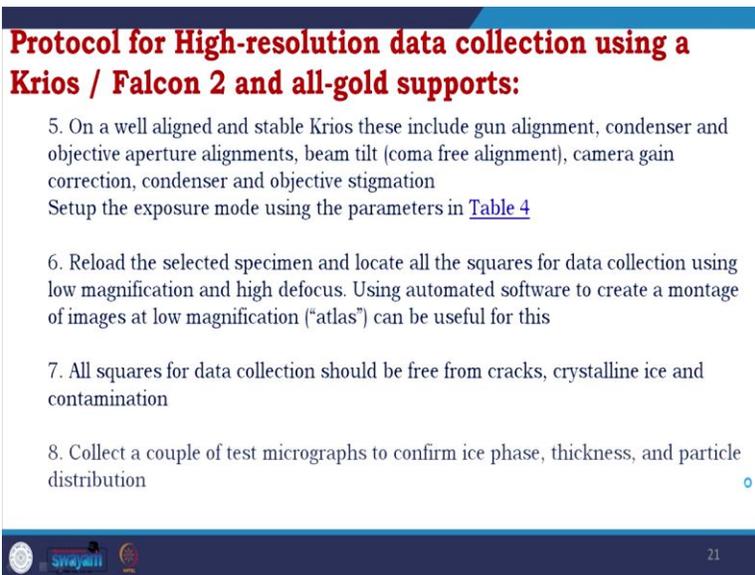
Ice thickness is best judged at low magnification by increasing the defocus to hundreds of micrometers to improve the contrast

4. Once a grid is selected for imaging, load the calibration grid and perform standard microscope alignments in low dose exposure mode

So coming to the protocol for high-resolution data collection, you have to load the specimen in cartridges and mount it in an autoloader, including a calibration grade, which could be graphitized carbon or similar material. Discard any that are bent or broken during handling. Load a specimen in the column is low magnification to check grid is intact and contains several squares with an appropriate thickness of ice so, again, the vitrification.

Ice thickness is the best judge at low magnification by increasing the defocus to hundreds of micrometres to improve the contrast. Once a grid is selected for imaging, load the calibration grid and perform standard microscope alignment in low dose exposure mode.

**(Refer Slide Time: 27:05)**



**Protocol for High-resolution data collection using a Krios / Falcon 2 and all-gold supports:**

5. On a well aligned and stable Krios these include gun alignment, condenser and objective aperture alignments, beam tilt (coma free alignment), camera gain correction, condenser and objective stigmation  
Setup the exposure mode using the parameters in [Table 4](#)
6. Reload the selected specimen and locate all the squares for data collection using low magnification and high defocus. Using automated software to create a montage of images at low magnification (“atlas”) can be useful for this
7. All squares for data collection should be free from cracks, crystalline ice and contamination
8. Collect a couple of test micrographs to confirm ice phase, thickness, and particle distribution

21

A well-aligned and stable Krios includes gun alignment, condenser and objective aperture alignments, beam tilt coma-free alignment, camera gain correction, condenser and objective stigmation. We have discussed setup the exposure mode using the parameters we have in the table. Reloading the selected specimen to locate all the squares for data collection using low magnification and hiding the focus using automated software to create a montage of images at low magnification can be helpful for that purpose.

All squares for data collection should be free from cracks, crystalline ice and contamination. Collect a couple of test micrographs to confirm ice phase, thickness and particle distribution.

**(Refer Slide Time: 28:03)**

## Protocol for High-resolution data collection using a Krios / Falcon 2 and all-gold supports:

9. Set focus in exposure mode using beam tilt or by looking at the fringes / bright spots at the edge of the hole ([Russo and Passmore, 2016b](#))

10. Check the illumination geometry of the beam in exposure mode: it should be round, centered on the imaging axis, and centered on the hole while encompassing an annulus of the support around the hole which is explained in the next figure

11. Begin collecting data, either manually or using automated data collection software [ch BC]

12. Check the focus and illumination every 20–30 holes. If mounted correctly, focus should vary by  $<1-2\ \mu\text{m}$  across an  $80\ \mu\text{m}$  square

13. Collect a few (1–2% of the total micrographs) tilt-pair micrographs for validation. Tilt angles and exposures of  $[0, 15^\circ]$  and  $[1, 3\ \text{s}]$  respectively, are reasonable for validation

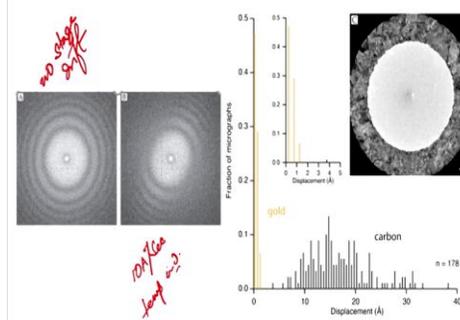


Set focus in exposure mode using beam tilt or by looking at the fringes or bright spot at the edge of the hole. These researchers who have seen the changes take the illumination geometry of the beam in exposure mode; it should be round, centred on the imaging axis and centred on the hole while encompassing an annulus of the support around the hole, which is explained in our following figure.

To begin the data collection either manually or using automated Data Collection software. Check the focus and illumination every 20 to 30 holes. If mounted correctly, the focus should vary to less than 1 to 2 micrometres across an 80-micrometre square. That would be the checking criteria. Collecting a few 1 to 2% of the total micrograph tilt pair micrograph of validation, tilt angle and exposure of 0 to 15 degrees and 13 seconds, respectively, are generally reasonable for validation.

**(Refer Slide Time: 29:16)**

## Testing for specimen movement during or after data collection:



Large amounts of specimen movement or stage drift can be detected during data acquisition using real-time fast Fourier transforms (FFTs) of the collected micrographs. FFTs of specimens are shown with no stage drift (A) and 10 Å/sec temperature induced stage drift (B). Micrograph C shows the recommended, symmetric illumination of a frozen specimen suspended across a hole in an all-gold support foil. Histogram is the in-plane movement statistics for 1 second micrographs (16 e<sup>-</sup>/Å<sup>2</sup>) on all-gold supports vs. amorphous carbon on gold (Quantifoil) under the same symmetric illumination conditions shown in C. Inset is enlargement of histogram near origin.

So we talked about the focusing, so it is testing for specimen movement during or after data collection. A large amount of specimen movement or stage drift can be detected during the acquisition using real-time First Fourier Transforms FFTs of the collected micrographs. FFTs are shown with no stage drift here, ten angstroms per the second temperature-induced stage drift.

So here, no stage drift is observed, and 10 angstroms per second temperature-induced stage drift. This micrograph C shows the recommended symmetric illumination of a frozen specimen suspended across a hole in an all gold support foil. The histogram is the in-plane movement statistics for a 1-second micrograph on all gold support process amorphous carbon on gold under the same symmetric illumination condition, shown here; This is the enlargement of the histogram near the origin to get a better look.

**(Refer Slide Time: 30:49)**

Every dataset collected using a direct electron detector should be checked to verify that the specimen preparation and illumination conditions are optimized to minimize particle motion

Currently, the simplest way to do this is using a per-micrograph motion correction program like motioncorr (Li et al., 2013)

Each micrograph is collected as a movie where the total dose is subdivided into individual frames

The algorithm is used to determine the overall movement of the specimen in the micrograph with time, and then offset and re-sum the frames thus removing some of the information loss

The program is fast, as it is implemented for a GPU architecture, and therefore can be used to quickly check the overall movement of the specimen support during imaging

Every data set collected using a direct electron detector should be checked to verify that the specimen preparation in illumination conditions is optimized to minimize the particle motion. The simplest way is to use a per-micrograph motion correction program like motioncorr. Each micrograph is collected as the movie, where the total dose is subdivided into individual frames.

The algorithm is used to determine the overall movement of the specimen in the micrograph with time and then offset the re-sum the frames that remove some of the information loss. The program is fast as it is implemented for a GPU architecture and, therefore, can quickly check the overall movement of the specimen support during imaging.

**(Refer Slide Time: 31:45)**

Distributions of movement for micrographs in a one second exposure at  $16 \text{ e}^-/\text{\AA}^2$  (Titan Krios on a Falcon 2 direct electron detector with 1.7 Å pixels)

Both sets were collected with the recommended symmetric conditions

On all-gold supports, the movement is less than 1.5 Å, which is less than one pixel in the image

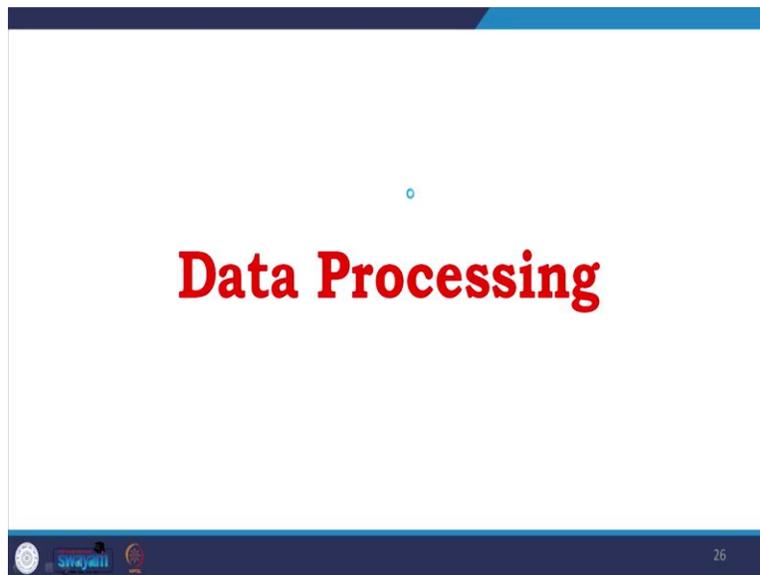
On standard Quantifoil supports under the same conditions, the per-micrograph movement is an order of magnitude larger

Thus, with ultra-stable supports and modern low-drift microscope stages, if you use micrograph motion tracking algorithms to check for incorrect data collection settings, stage drift, or damage to the foil or grid since under normal conditions with ultra-stable supports, the overall movement of the specimen support should be essentially undetectable

Distribution of movement for micrographs in one-second exposure at 16electron by A square Krios on Falcon 2 direct electron detector with 1.7-angstrom pixels. Both sets should be collected with the recommended symmetric conditions. On all gold support, the movement is less than 1.5 angstrom which is less than 1 pixel in the image.

So it could be a here, on standard Quantifoil support under the same conditions for per-micrograph movement is an order of magnitude larger than this. Thus, with Ultra stable support and modern load rip microscope stages, if you use microgram motion tracking algorithms to check for incorrect Data Collection settings, stage drift or damage to the foil or grid since under normal conditions with ultra-stable supports, the overall movement of the specimen support should be essentially undetectable.

**(Refer Slide Time: 32:56)**



Coming to the next stage is data processing.

**(Refer Slide Time: 32:59)**

The image processing pipeline in single particle cryo-EM is required to solve the 3D electron density of a target molecule, in potentially many conformational states, from noisy 2D images collected using cryoTEM

Each collected image is a movie of dose-fractionated frames that require motion estimation and correction

The corrected images (micrographs) are then used to estimate the microscope CTF during the exposure, as well as to find and pick out single particles

The single particles are extracted from the micrographs, and then are sorted and filtered using 2D classification methods

The resulting filtered particle stacks are used to perform ab initio 3D structure determination of potentially multiple discrete states or targets



The image processing pipeline in single-particle cryo-EM is required to solve the 3D electron density of a target molecule in potentially many conformational states from noisy 2D images, collected using the priority is collected images the movie of fractionated films that require motion, Estimation and correction we talked in the previous section. Also, the corrected images are then used to estimate the microscope CTF during the exposure and to find and pick out the single particles.

The single particles are extracted from the micrographs and then are sorted and filtered using the 2D classification method. So you have to look at the different particles in a different orientation and then separate them or classify them. The resulting filtered particles perform ab initio 3D structure determination of potentially multiple discrete states or targets.

**(Refer Slide Time: 34:09)**

These coarse structures are then further classified and refined in 3D to yield interpretable molecular density maps and achieve state of the art resolutions

Each of the processing stages above can have parameters and inputs that requires decisions made by the scientist, using structural or other insight as prior knowledge to guide processing

Typical workflows on difficult molecules often involve multiple iterations of segments of the processing pipeline, each time with changes in the selection of input data or parameters, to yield optimal results

In particular, curating data and separating a sample in silico into multiple homogenous subsets representing different conformations can require knowledge both of the target and of the underlying algorithms

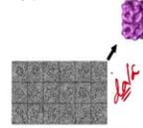


These coarse structures are classified and defined in 3D to yield interpretable molecular density maps and achieve state-of-the-art resolutions. Each of the processing stages above can have parameters and inputs that require the decision made by the scientist, using structural in other insight as prior knowledge to guide the processing

Typical workflows on complex molecules often involve multiple iterations of segments of the processing pipeline. So you do the same thing multiple times towards going from an initial model to a refined, corrected model, each time with changes in the selection of input data parameters to kill the optimal results. In particular, curating data and separating a sample in silico into multiple homogeneous subsets representing different conformations requires knowledge of both the target and the underlying algorithms.

**(Refer Slide Time: 32:24)**

## Single Particle Analysis:



Perhaps the most commonly used variant of cryo-electron microscopy is single-particle analysis

In this technique, data from a large number of 2D projection images, featuring identical copies of a protein complex in different orientations, are combined to generate a 3D reconstruction of the structure

When atomic models are available for some or all of the sub-components of the complex, they can be placed or fitted into the density map to provide pseudo-atomic models, considerably extending the information obtained by electron microscopy



29

Coming to the single-particle analysis from the data to construction of the single-particle, this is perhaps the most commonly used variant of Cryo-electron microscopy, the single-particle analysis in this technique. Data from many projection images featuring identical copies of a protein complex in Different orientations are combined to generate a 3D structure reconstruction.

When atomic models are available for some or all of the complex's subcomponents, they can be fitted into the density map to provide pseudo atomic models considerably, extending the information obtained by electron microscopy.

(Refer Slide Time: 36:11)

## Single Particles:

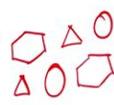
### Isolated macromolecular complexes

*Big advantage*

Randomly oriented in solution

*multiple conformation*

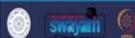
Can be trapped in different reaction states by Vitrification



No crystallization or ordered assembly needed

The position and orientation of each particle must be determined for 3D reconstruction

The more particles used, the higher the resolution (<3 Å)



30

So, isolated macromolecular complexes, randomly oriented in solution, can be trapped in different reactions States by vitrification. No crystallization or order assembly is needed. The position and orientation of each particle must be determined for 3D reconstruction. So the macromolecular complex would be randomly oriented means they would stay in multiple confirmations. However, they could be trapped in different reactions states.

No crystallization or order assembly is needed, which is a significant advantage as we typically talk in crystallography. The position and orientation of each particle must be determined for 3D reconstruction. The more particles used, the higher the resolution.

(Refer Slide Time: 37:39)

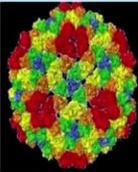
**Single Particles:**

**Isolated macromolecular complexes**

Mixed states can be separated ("purification in the computer")

Ultimate limit to resolution from radiation damage

Interpretation by atomic structure docking or direct determination of backbone

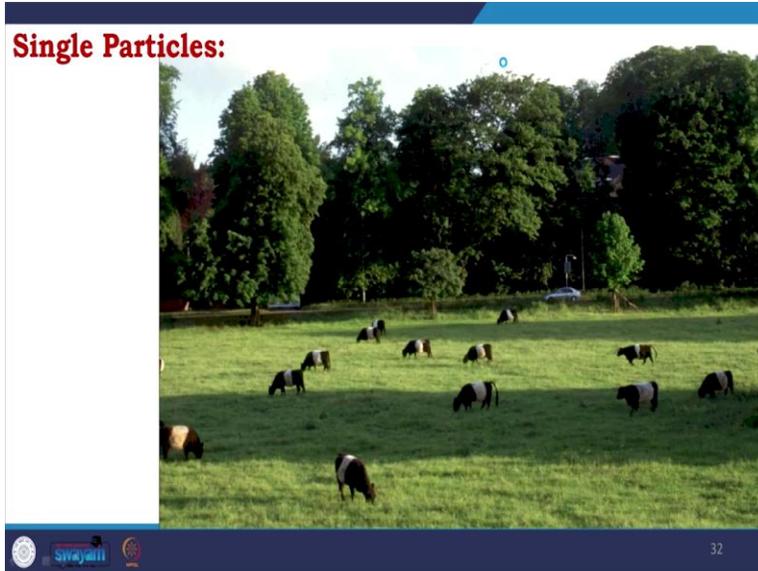


*Handwritten notes in red:* "Separate" with an arrow pointing to "Purification (computer)"

31

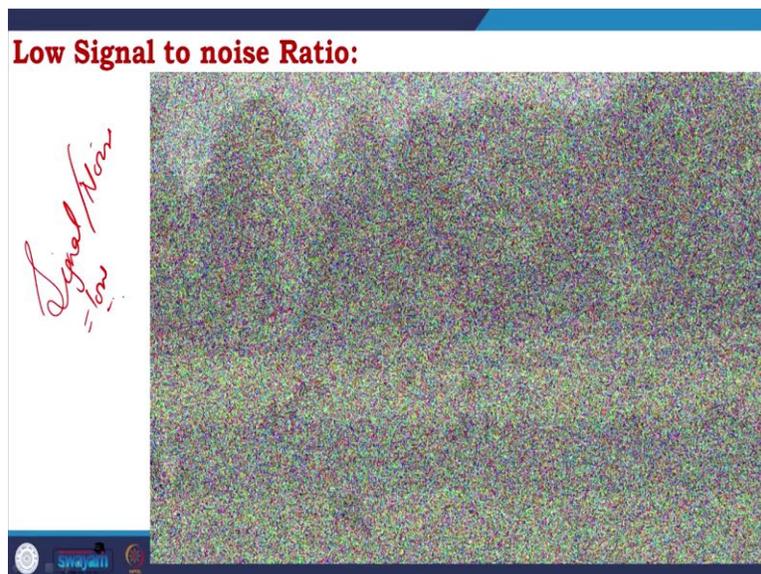
So for those isolated macromolecular complexes, mixed states can be separated. So you get the different states and then separate them; Using a computer, you get different alignments, then you separate them and classify them. The ultimate limit to resolution from radiation damage, Interpretation by atomic structure docking or direct determination of the backbone is possible.

(Refer Slide Time: 38:18)



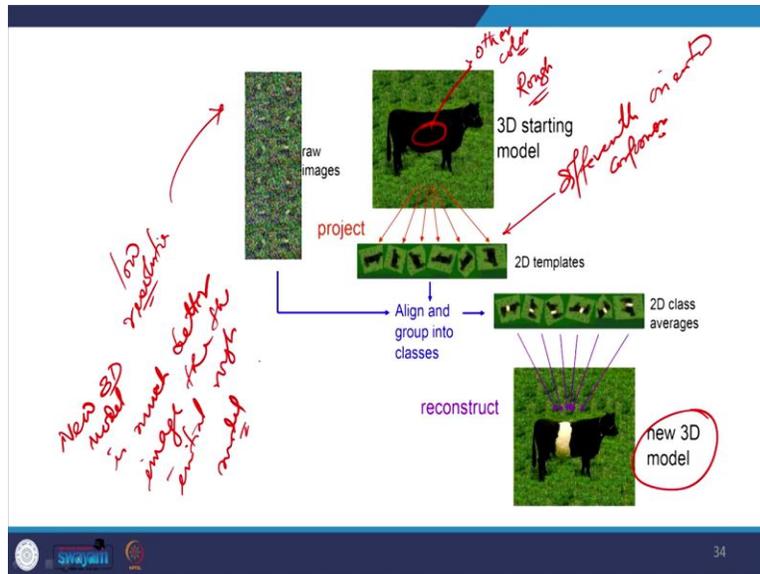
So let us start thinking about a single particle. All the cows there, roaming here considering the house as a single particle. You consider it as a single particle then you collect data.

**(Refer Slide Time: 38:45)**



The data is noise signal is low as you see you could not see anything.

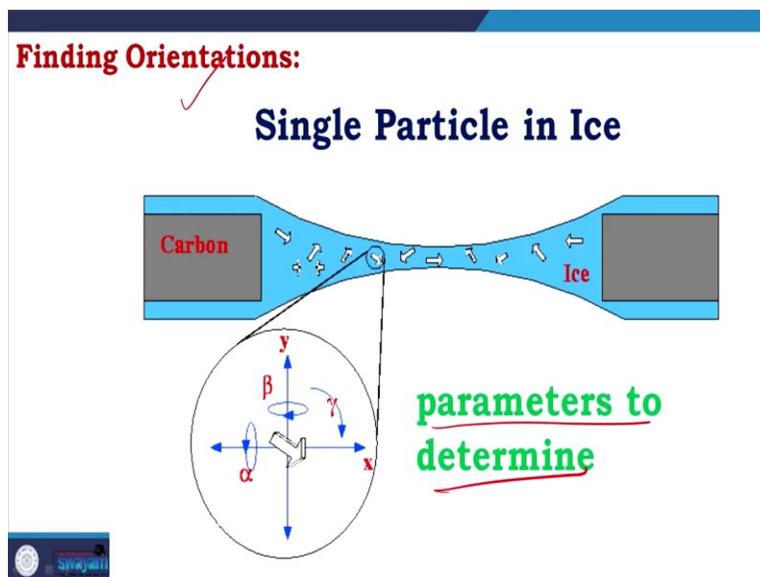
**(Refer Slide Time: 38:56)**



However, they knew the process to get these images, the low-resolution images. And then, you have the initial rough 3D model rough. If you compare, you will see the cow had some other colour that you do not get here because of the low-resolution model. However, with this rough model, you identify differently oriented conformers. Now you get the raw images, align and group them into classes, and get 2D class averages.

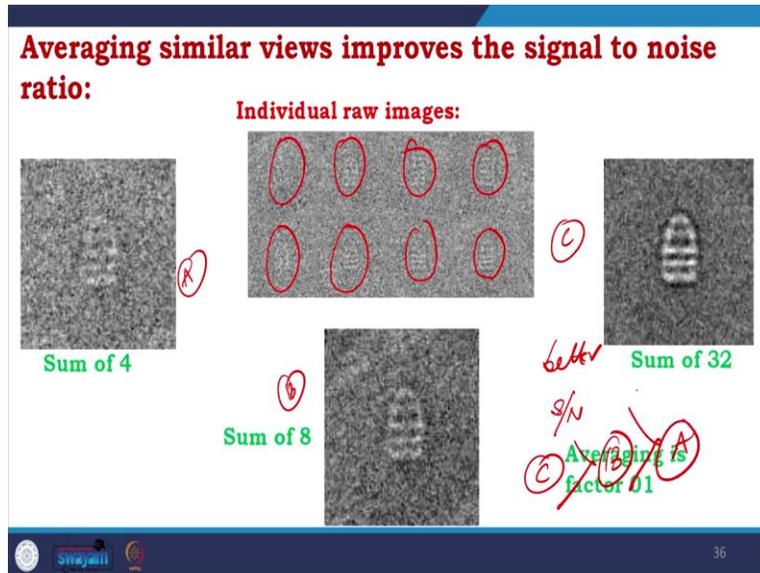
From the 2D class average, you reconstruct a new get new 3D model. The new 3D model is better than the initial rough model. That is where image processing helps.

(Refer Slide Time: 40:28)



Also, you have to find the orientation. If you have the carbon base, you have the vitrified ice, and the particles are in a different orientation. It would help if you had the parameters to determine those orientations; this is the key here.

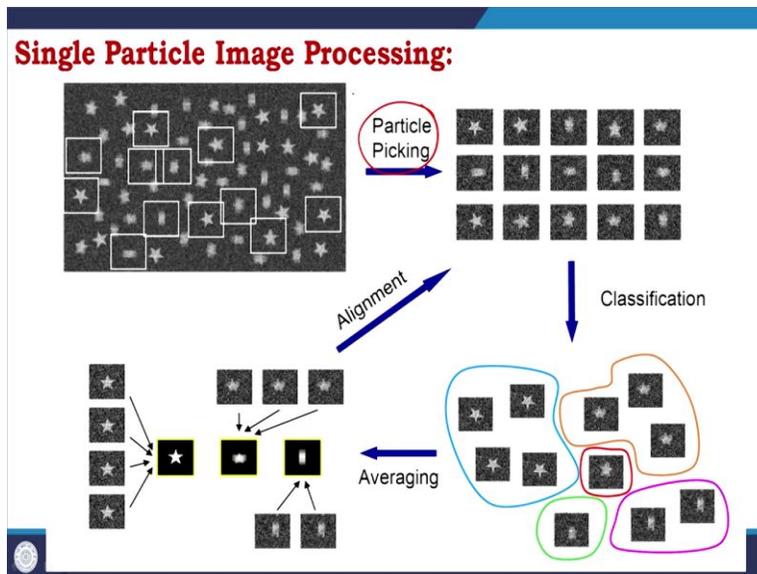
(Refer Slide Time: 40:53)



Also, if you see your average similar views, you improve the signal. So, this is the individual raw images. You could see something here, but they are under very low resolution. Now, if you average, let us say, you make a sum of 4 to average them. You get a picture like that, a better picture with the sum of 8. So, you get more samples; this is better. And then, it is much better because of the sum of 32.

So the more you take, the more you do average better, signal-to-noise you get. So, from this scenario to this scenario, C is better than B better than A.

(Refer Slide Time: 42:57)



So, if you summarise what we perform for single particle image processing, we get the data. Then we identify single conformations or orientations and pick them up; this is called particle picking. And then, when you separate them, we make the classification, and once we do the classification, we do the averaging. Averaging gives a better picture than then we align with them. That is the single-particle image processing summary.

**(Refer Slide Time: 42:40)**

### 3D reconstruction:

- The goal of 3D reconstruction in Cryo electron microscopy is to take a 2D projection and turn it into its 3D volume

This allows for the visualization of small structures within the sample

It is essential to reconstruct the 2D projection because the information gained from the projection itself is limited

Coming to the 3D reconstruction, the goal of 3D reconstruction in a Cryo-electron microscope is to take 2D projects and turn them into their 3D volume; This allows for the visualization of small structures within the sample. It is essential to reconstruct the 2D projection because the

information gained from the projection itself is limited. So, the projection you get towards 3D is limited. So you need to reconstruct it. So this is what we perform.

**(Refer Slide Time: 43:26)**

A projection occurs when a 3D image is taken and those points are then put onto a 2D plane

Each point in the 2D plane corresponds to a slice in the 3D volume and is based on Radon's theorem. It is possible to obtain a 3D volume, get the 2D projection, subsequently take the Fourier transform of the projection and use the Fourier transform of the projection to obtain a "slice" of the 3D volume

Thereafter, it is possible to take the inverse Fourier transform of the "slice" yielding a 3D volume.

39

A projection occurs when 3D images are taken. Moreover, those points are then put on a 2D plane. So we take a 3D object, and then the points are put into the 2D plane. Now, what happened? Each point in the 2D plane corresponds to the slice in the 3D volume and is based on Radon's theorem. It is possible to obtain its 3D volume, get the 2D projection, take the Fourier Transform projection, and use the Fourier Transform projection to a slice in the 3D volume.

You only get the points in 2D, so, 3D object. You start with condition A. You come here, and taking that 2D projection, you do 2D Fourier Transform you get a 2D FT. Then you get a 3D FT of the object, giving you a slice. Now that you do the 3D inverse, it is possible to take the inverse Fourier Transform of the slice, yielding the same volume. So this is C, this is D and goes to A.

**(Refer Slide Time: 45:02)**

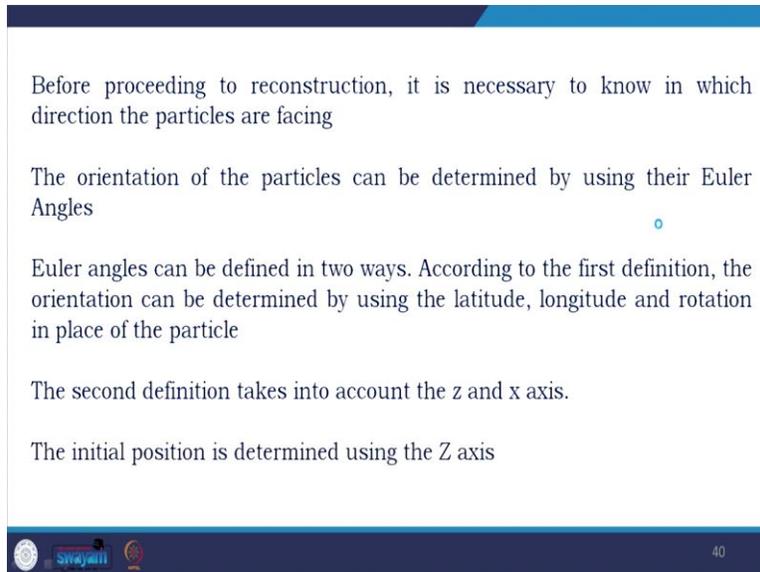
Before proceeding to reconstruction, it is necessary to know in which direction the particles are facing

The orientation of the particles can be determined by using their Euler Angles

Euler angles can be defined in two ways. According to the first definition, the orientation can be determined by using the latitude, longitude and rotation in plane of the particle

The second definition takes into account the z and x axis.

The initial position is determined using the Z axis

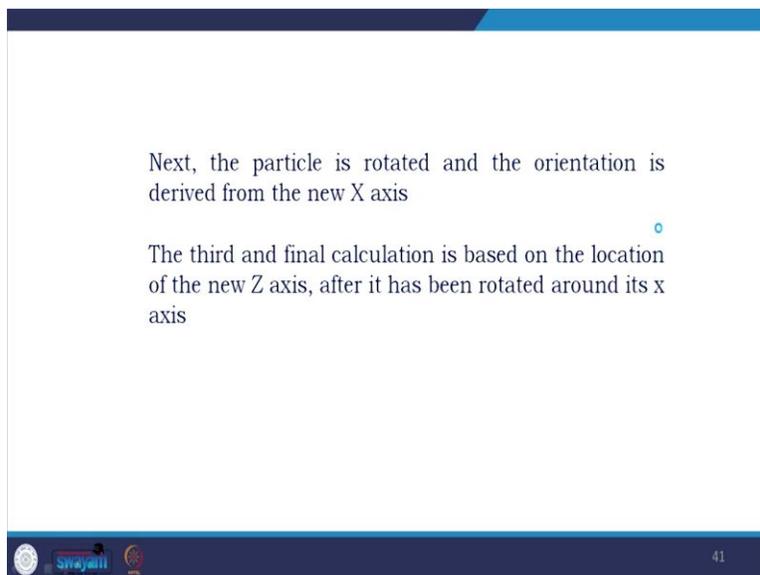
A presentation slide with a white background and a blue header and footer. The text is centered and discusses the determination of particle orientation using Euler Angles. The footer contains logos for 'Swayam' and 'eGangotri' and the number '40'.

Before proceeding to reconstruction, it is necessary to know the direction the particles are facing. The orientation of the particle can be determined using the Euler angles. Now the Euler angle can be defined in 2 ways. According to the first definition, the orientation can be determined by using the latitude, longitude and rotation in the plane of the particle. The second definition takes into the z and x-axis. The initial position is determined using the Z-axis.

**(Refer Slide Time: 45:42)**

Next, the particle is rotated and the orientation is derived from the new X axis

The third and final calculation is based on the location of the new Z axis, after it has been rotated around its x axis

A presentation slide with a white background and a blue header and footer. The text is centered and describes the final calculation based on the new Z-axis after rotation around the x-axis. The footer contains logos for 'Swayam' and 'eGangotri' and the number '41'.

Next, the particle is rotated, and the orientation is derived from the new x-axis. The third and final calculation is based on the location of the new Z-axis after it has been rotated around the x-axis.

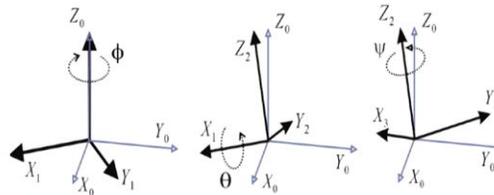
**(Refer Slide Time: 45:56)**

A backprojection is basically the inverse of a Radon transform

It is the process where a 2D projection goes into a 3D model. It is the only method that can be used for this purpose

A back projection reconstructs an image by taking each view and smearing it along the same path it was acquired

The result from this is a blurry version of the correct image (Voss 2013)



The back-projection is the inverse of the red on transformation. It is the 2D projection that goes into a 3D model. It is the only method that can be used for this purpose. A back projection and reconstruct an image by taking each view and smearing it along the same path it was acquired. So you could see how the rotation happens in the Euler angle. The result from this is a blurry version of the correct image.

(Refer Slide Time: 46:38)

### 3D Reconstruction Process Steps:

**3D reconstruction process estimates the unknown orientations and 3D structure at the same time;**

3D electron density maps are created from 2D projections

Angles of projections relative to each other are determined

Find common line projections to determine relative angles

So, what are the 3D reconstruction process steps? 3D reconstruction process estimates the unknown orientation to distract at the same time. The 3D electron density maps are created from the 2D projection; as we discussed, the angle of projection relative to each other are first determined. Then, find the common line projections to determine the relative angles.

(Refer Slide Time: 47:04)

**Observation of the Specimen:**

**The contrast of the specimen depends on:**

- Specimen itself
- Defocus value of the objective lens
- Thickness of the ice

**There are three methods of observing and recording images:**

- Fluorescent Screen
- Photographic Film
- CCD Cameras
- DED Set up

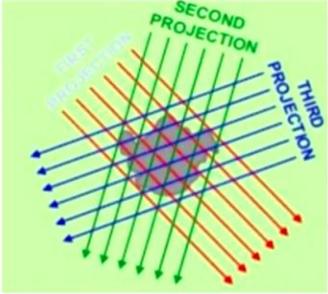


You have to do the observation of the specimen. The contrast of the specimen depends on the specimen itself, the defocus value of the objective lens and the thickness of the ice. There are three methods of observing and recording images: Fluorescent screen, we talked about photographic film, we talked about CCD cameras and, just as an extension, the DED setup.

(Refer Slide Time: 47:33)

3D images of biological sample is obtained by a process called Cryo-Electron Tomography(CET), where a 3D reconstruction of a sample is created from tilted 2D images

Raw Images from different projections are recorded and at last they are stitched together to form the required 3D structures using computer software.



So, 3D images of the biological sample are obtained by Cryo-electron tomography or CET, where 3D reconstruction of the sample is created from the tilted 2D images because of the tiltation get the angular changes. The raw images from different projects the recorded, and at last,

they are stitched together to form the required 3D structure using the computer software. So if you see this is the object and first, second, third, those projections are taken.

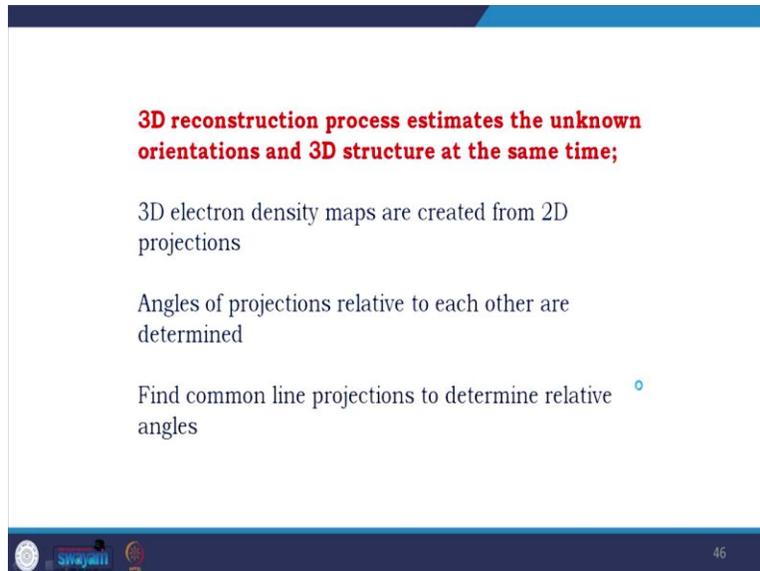
**(Refer Slide Time: 48:13)**

**3D reconstruction process estimates the unknown orientations and 3D structure at the same time;**

3D electron density maps are created from 2D projections

Angles of projections relative to each other are determined

Find common line projections to determine relative angles

A presentation slide with a white background and a blue header and footer. The text is centered and includes a red heading, three bullet points, and a footer with logos and the number 46.

The 3D reconstruction process estimates the unknown orientation and the 3D structure simultaneously. The 3D electron density maps are created from the 2D projection. The projection angle relative to each other is determined, and the, find common line production to determine the relative angles.

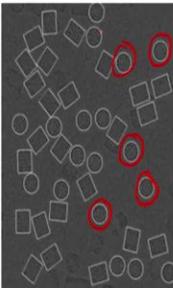
**(Refer Slide Time: 48:41)**

**Automated Particle Picking:**

Identify particles in micrograph and cut out patches containing one particle each

This can be done automatically

Manual process is tedious and difficult

A micrograph showing a collection of small, irregularly shaped particles. Some particles are highlighted with red circles, indicating they have been automatically picked. The background is dark, and the particles are light gray.

So Re-projection, the 3D density map can generate the projection that can be used to realign the raw images. The process may have to be repeated at different times. Then there would be

automated particle picking, identifying the particles in the micrograph and cutting out patches containing the one particle each. Like here, if you see, you have to find out these once in that way; This can be done automatically because the manual process is tedious and difficult.

**(Refer Slide Time: 49:23)**

**Images Enhancement:**

Image Noise is the random variation of brightness or color

Information in images produced by the sensor

Cryo EM images are very noisy and have very low contrast

Smooth the noise as well as enhance the contrast



49

Then we have image Enhancement; image noise is the random variation of brightness or colour, Information in images produced by the sensor. The Cryo-EM images are very noisy and have very low contrast. So here you do the enhancement, and you get the better picture. Smooth the noise as well as enhance the contrast.

**(Refer Slide Time: 50:02)**

**Summary:**

Cryo-EM is a form of Transmission Electron Microscopy (TEM) where the sample is studied in its native state at cryogenic temperatures

Used for 3D visualization of biological molecules

Resolution of Cryo-EM is not high enough but it is improving using different computer techniques and many other additions

With the advancement of technology, this technique will certainly improve and the effect have already started to be observed

50

In summary, cryo-EM is a form of Transmission electron microscopy where the sample is produced in the native state at cryogenic temperatures used for 3D visualization of biological macromolecules. The resolution of Cryo-EM is not high enough as we have seen here, but it is improving using different computer techniques and many other additions. With the advancement of technology, this technique will undoubtedly improve.

Moreover, the effects have already started to be observed. In the next class, we will talk about how the advancement of electron microscopy happened. From the 1970s started at the initial stage, how the other innovation has contributed towards high-resolution starting around 100s to 50s and the resolution now they are regularly attending lower than three and some.

So we will talk about the resolution Revolution. We will talk about the journey. We will talk about the future and some of the modern innovation recent innovations, help the resolution revolution. Thank you very much for listening.