

# Our learning experience on Deep Learning

in learning the 3D protein structures from cryo-EM images

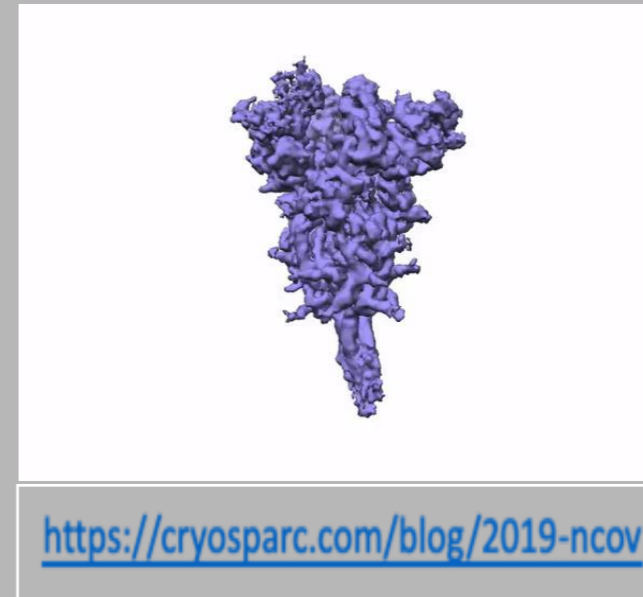
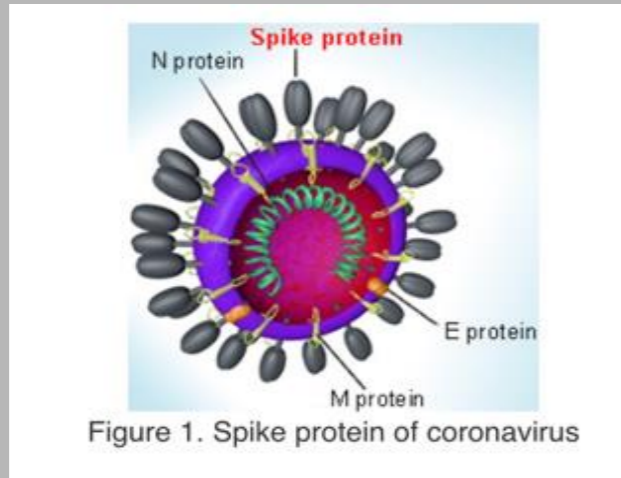
I-Ping Tu

Institute of Statistical Science, AS

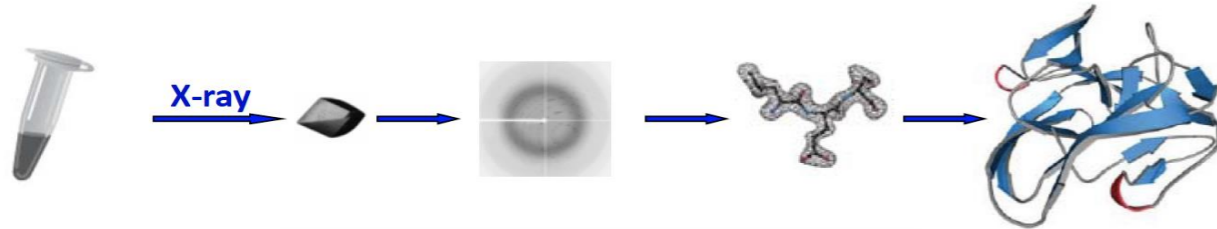
2022.04.22

# Why Cryo-Electron Microscopy

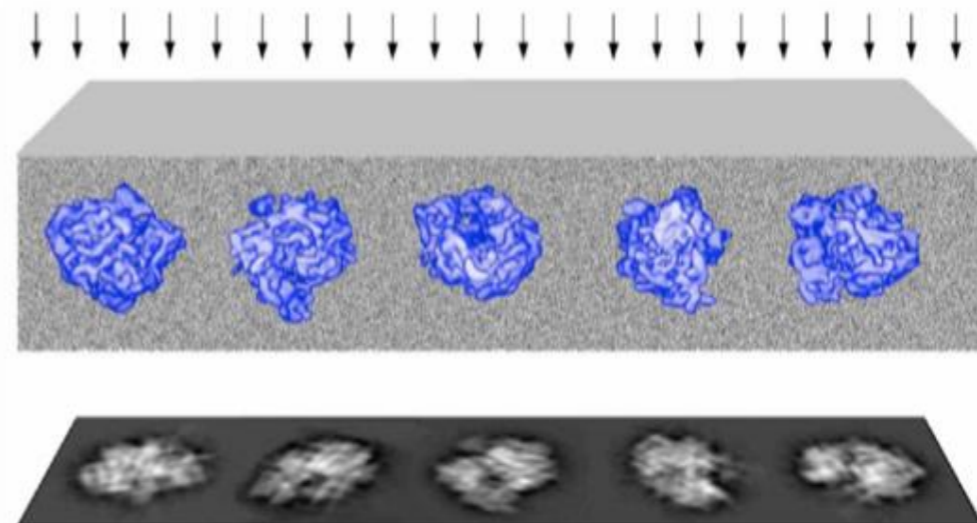
A solution for emergent problems



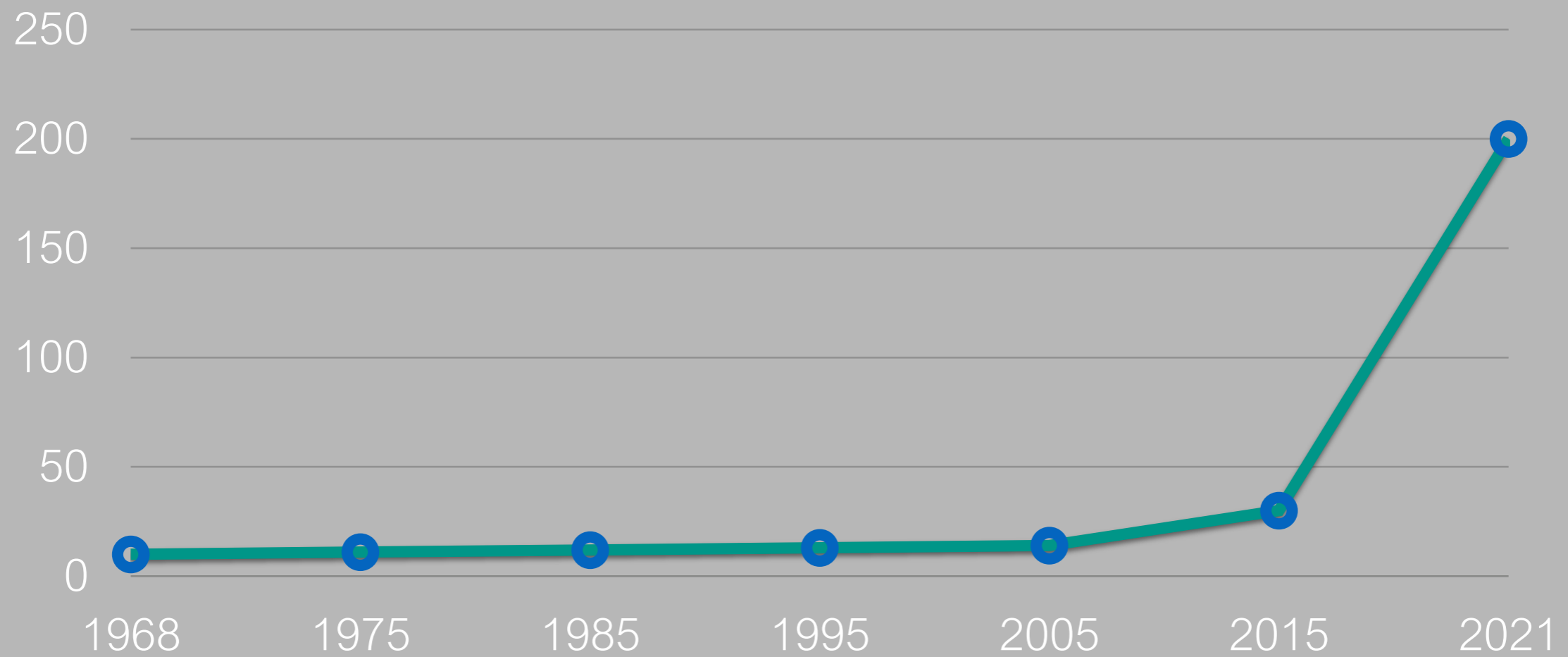
## X-ray Crystallography



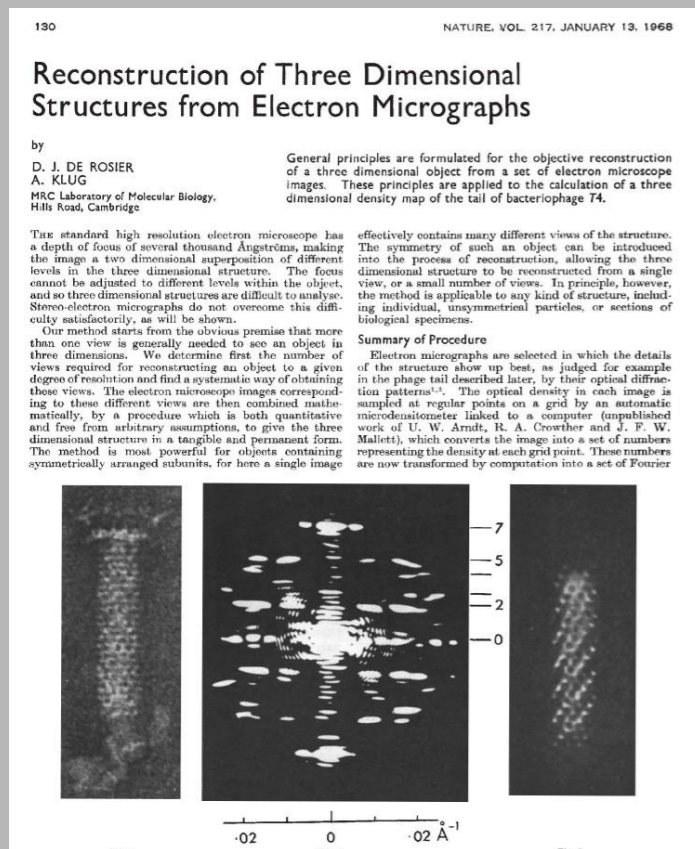
## Single Particle cryo-EM



# Brief History of cryo-EM



# Brief History of cryo-EM



50  
0

1968

1975

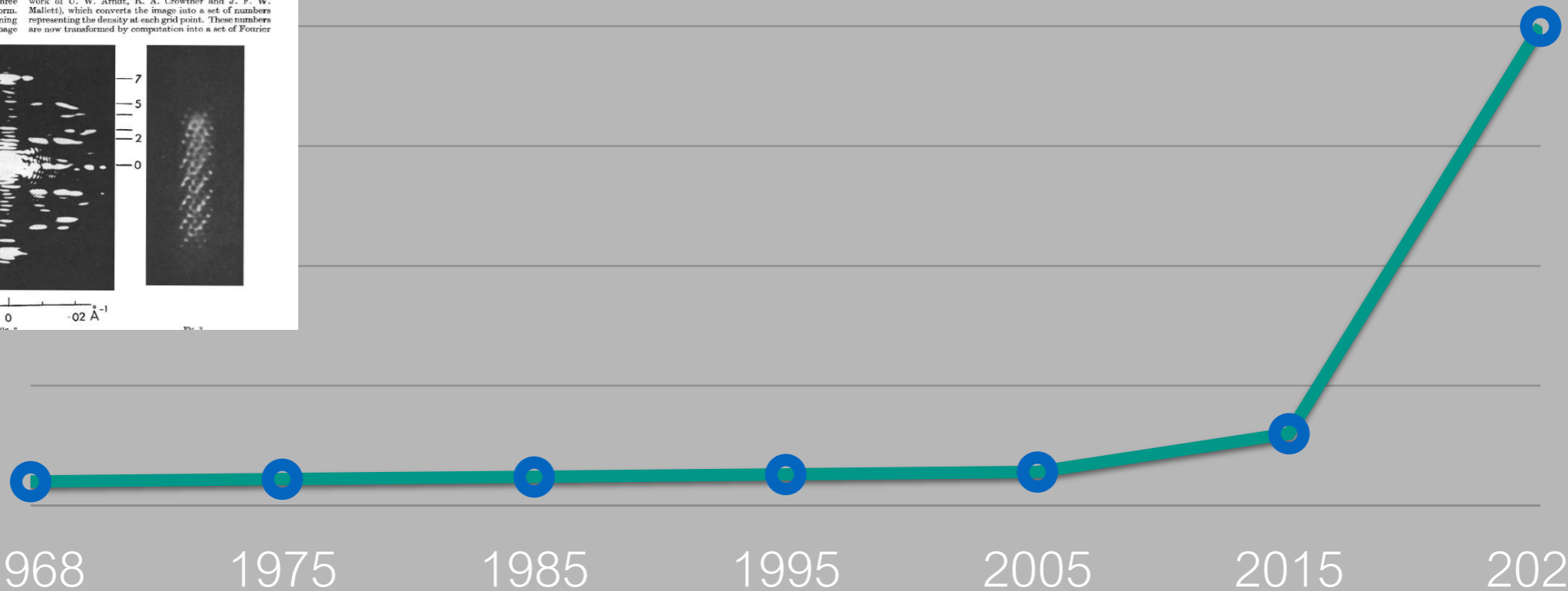
1985

1995

2005

2015

2021



# Brief History of cryo-EM

130 NATURE, VOL. 217, JANUARY 13, 1968

## Reconstruction of Three Dimensional Structures from Electron Micrographs

by  
D. J. DE ROSIER  
A. KLUG  
MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge

General principles are formulated for the objective reconstruction of a three dimensional object from a set of electron microscope images. These principles are applied to the calculation of a three dimensional density map of the tail of bacteriophage T4.

The standard high resolution electron microscope has a depth of focus of several thousand Angstroms making the image a two dimensional superposition of different levels in the three dimensional structure. The focus cannot be adjusted to different levels within the object, and so three dimensional structures are difficult to analyse. Stereo-electron micrographs do not overcome this difficulty satisfactorily, as will be shown.

Our method starts from the obvious premise that more than one view is generally needed to see an object in three dimensions. We determine first the number of views required for reconstructing an object to a given degree of resolution and find a systematic way of obtaining these views. The electron microscope images corresponding to these different views are then combined mathematically, by a procedure which is both quantitative and free from arbitrary assumptions, to give the three dimensional structure in a tangible and permanent form. The method is most powerful for objects containing symmetrically arranged subunits, for here a single image effectively contains many different views of the structure. The symmetry of such an object can be introduced into the process of reconstruction, allowing the three dimensional structure to be reconstructed from a single view, or a small number of views. In principle, however, the method is applicable to any kind of structure, including individual, unsymmetrical particles, or sections of biological specimens.

**Summary of Procedure**  
Electron micrographs are selected in which the details of the structure show up best, as judged for example in the phage tail described later, by their optical diffraction patterns<sup>1,2</sup>. The optical density in each image is sampled at regular points on a grid by an automatic microdensitometer linked to a computer (unpublished work of U. W. Arndt, K. A. Crowther and J. F. W. Mallett), which converts the image into a set of numbers representing the density at each grid point. These numbers are now transformed by computation into a set of Fourier

## Method of the Year 2015 Single Particle cryo-Electron Microscopy

Nature Methods | Editorial  
**Method of the Year 2015**  
The end of 'blob-bology': single-particle cryo-electron microscopy (cryo-EM) is now being used to solve macromolecular structures at high resolution.

Nature Methods | News Feature  
**The field that came in from the cold**  
Recent advances in cryo-electron microscopy are enabling researchers to solve protein structures at near-atomic resolutions, expanding the biological applicability of this technique. Michael Eisenstein reports.  
Michael Eisenstein

Nature Methods | Primer  
**Collect data**  
2D projections  
**Single-particle cryo-electron microscopy**  
A brief overview of how to solve a macromolecular structure using single-particle cryo-electron microscopy (cryo-EM).  
Allison Doerr

Nature Methods | Historical Commentary  
EMDB maps released

Nature Methods | Commentary  
a b

Nature Methods | Method to Watch  
a b

50

0

1968

1975

1985

1995

2005

2015

2021

# Brief History of cryo-EM

130 NATURE, VOL. 217, JANUARY 13, 1968

## Reconstruction of Three Dimensional Structures from Electron Micrographs

by  
D. J. DE ROSIER  
A. KLUG  
MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge

General principles are formulated for the objective reconstruction of a three dimensional object from a set of electron microscope images. These principles are applied to the calculation of a three dimensional density map of the tail of bacteriophage T4.

The standard high resolution electron microscope has a depth of focus of several thousand Angstroms making the image a two dimensional superposition of different levels in the three dimensional structure. The focus cannot be adjusted to different levels within the object, and so three dimensional structures are difficult to analyse. Stereo-electron micrographs do not overcome this difficulty satisfactorily, as will be shown.

Our method starts from the obvious premise that more than one view is generally needed to see an object in three dimensions. We determine first the number of views required for reconstructing an object to a given degree of resolution and find a systematic way of obtaining these views. The electron microscope images corresponding to these different views are then combined mathematically, by a procedure which is both quantitative and free from arbitrary assumptions, to give the three dimensional structure in a tangible and permanent form. The method is most powerful for objects containing symmetrically arranged subunits, for here a single image effectively contains many different views of the structure. The symmetry of such an object can be introduced into the process of reconstruction, allowing the three dimensional structure to be reconstructed from a single view, or a small number of views. In principle, however, the method is applicable to any kind of structure, including individual, unsymmetrical particles, or sections of biological specimens.

### Summary of Procedure

Electron micrographs are selected in which the details of the structure show up best, as judged for example in the phage tail described later, by their optical diffraction patterns<sup>1,2</sup>. The optical density in each image is sampled at regular points on a grid by an automatic microdensitometer linked to a computer (unpublished work of U. W. Arndt, K. A. Crowther and J. F. W. Mallett), which converts the image into a set of numbers representing the density at each grid point. These numbers are now transformed by computation into a set of Fourier

## 2017 Nobel Laureates in Chemistry

### Dubochet, Frank, Henderson

*"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules **in solution**"*

Method of Single Particle

CellPress

Congratulations 2017 Nobel Laureates in Chemistry  
Jacques Dubochet, Joachim Frank and Richard Henderson!

50

0

1968

1975

1985

1995

2005

2015

2021

# Brief History of cryo-EM

130 NATURE, VOL. 217, JANUARY 13, 1968

## Reconstruction of Three Dimensional Structures from Electron Micrographs

by  
D. J. DE ROSIER  
A. KLUG  
MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge

General principles are formulated for the objective reconstruction of a three dimensional object from a set of electron microscope images. These principles are applied to the calculation of a three dimensional density map of the tail of bacteriophage T4.

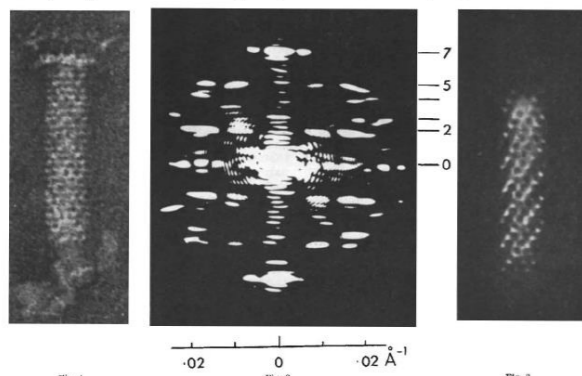
The standard high resolution electron microscope has a depth of focus of several thousand Angstroms making the image a two dimensional superposition of different levels in the three dimensional structure. The focus cannot be adjusted to different levels within the object, and so three dimensional structures are difficult to analyse. Stereo-electron micrographs do not overcome this difficulty satisfactorily, as will be shown.

Our method starts from the obvious premise that more than one view is generally needed to see an object in three dimensions. We determine first the number of views required for reconstructing an object to a given degree of resolution and find a systematic way of obtaining these views. The electron microscope images corresponding to these different views are then combined mathematically, by a procedure which is both quantitative and free from arbitrary assumptions, to give the three dimensional structure in a tangible and permanent form. The method is most powerful for objects containing symmetrically arranged subunits, for here a single image

effectively contains many different views of the structure. The symmetry of such an object can be introduced into the process of reconstruction, allowing the three dimensional structure to be reconstructed from a single view, or a small number of views. In principle, however, the method is applicable to any kind of structure, including individual, unsymmetrical particles, or sections of biological specimens.

### Summary of Procedure

Electron micrographs are selected in which the details of the structure show up best, as judged for example in the phage tail described later, by their optical diffraction patterns<sup>1,2</sup>. The optical density in each image is sampled at regular points on a grid by an automatic microdensitometer linked to a computer (unpublished work of U. W. Arndt, K. A. Crowther and J. F. W. Mallett), which converts the image into a set of numbers representing the density at each grid point. These numbers are now transformed by computation into a set of Fourier



50

0

1968

1975

1985

1995

2005

2015

2021

## RESEARCH

### CORONAVIRUS

## Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation

Daniel Wrapp<sup>1\*</sup>, Nianshuang Wang<sup>1\*</sup>, Kizmekia S. Corbett<sup>2</sup>, Jory A. Goldsmith<sup>1</sup>, Ching-Lin Hsieh<sup>1</sup>, Olubukola Abiona<sup>2</sup>, Barney S. Graham<sup>2</sup>, Jason S. McLellan<sup>1†</sup>

The outbreak of a novel coronavirus (2019-nCoV) represents a pandemic threat that has been declared a public health emergency of international concern. The CoV spike (S) glycoprotein is a key target for vaccines, therapeutic antibodies, and diagnostics. To facilitate medical countermeasure development, we determined a 3.5-angstrom-resolution cryo-electron microscopy structure of the 2019-nCoV S trimer in the prefusion conformation. The predominant state of the trimer has one of the three receptor-binding domains (RBDs) rotated up in a receptor-accessible conformation. We also provide biophysical and structural evidence that the 2019-nCoV S protein binds angiotensin-converting enzyme 2 (ACE2) with higher affinity than does severe acute respiratory syndrome (SARS)-CoV S. Additionally, we tested several published SARS-CoV RBD-specific monoclonal antibodies and found that they do not have appreciable binding to 2019-nCoV S, suggesting that antibody cross-reactivity may be limited between the two RBDs. The structure of 2019-nCoV S should enable the rapid development and evaluation of medical countermeasures to address the ongoing public health crisis.

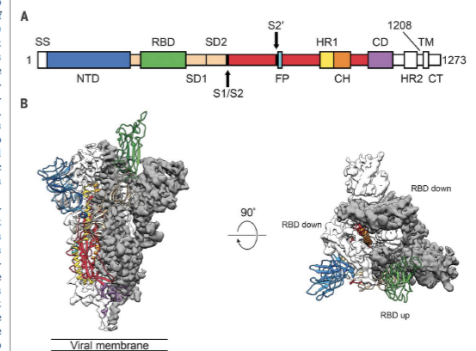
The novel coronavirus 2019-nCoV has recently emerged as a human pathogen in the city of Wuhan in China's Hubei province, causing fever, severe respiratory illness, and pneumonia—a disease recently named COVID-19 (1, 2). According to the World Health Organization (WHO), as of 16 February 2020, there had been >51,000 confirmed cases globally, leading to at least 1900 deaths. The emerging pathogen was rapidly characterized as a new member of the betacoronavirus genus, closely related to several bat coronaviruses and to severe acute respiratory syndrome coronavirus (SARS-CoV) (3, 4). Compared with SARS-CoV, 2019-nCoV appears to be more readily transmitted from human to human, spreading to multiple continents and leading to the WHO's declaration of a Public Health Emergency of International Concern (PHEIC) on 30 January 2020 (1, 5, 6).

2019-nCoV makes use of a densely glycosylated spike (S) protein to gain entry into host cells. The S protein is a trimeric class I fusion protein that exists in a metastable prefusion conformation that undergoes a substantial structural rearrangement to fuse the viral membrane with the host cell membrane (7, 8). This process is triggered when the S1 subunit binds to a host cell receptor. Receptor binding destabilizes the prefusion trimer, resulting in shedding of the S1 subunit and transition of the S2 subunit to a stable postfusion conformation (9). To engage a host cell receptor, the receptor-binding domain (RBD) of S1 undergoes hinge-like conformational movements that transiently hide or

state and up corresponds to the receptor-accessible state, which is thought to be less stable (10–13). Because of the indispensable function of the S protein, it represents a target for antibody-mediated neutralization, and characterization of the prefusion S structure would provide atomic-level information to guide vaccine design and development.

Based on the first reported genome sequence of 2019-nCoV (4), we expressed ectodomain residues 1 to 1208 of 2019-nCoV S, adding two stabilizing proline mutations in the C-terminal S2 fusion machinery using a previous stabilization strategy that proved effective for other betacoronavirus S proteins (11, 14). Figure 1A shows the domain organization of the expression construct, and figure S1 shows the purification process. We obtained ~0.5 mg/liter of the recombinant prefusion-stabilized S ectodomain from Freestyle 293 cells and purified the protein to homogeneity by affinity chromatography and size-exclusion chromatography (fig. S1). Cryo-electron microscopy (cryo-EM) grids were prepared using this purified, fully glycosylated S protein, and preliminary screening revealed a high particle density with little aggregation near the edges of the holes.

expose the determinants of receptor binding. These two states are referred to as the “down” conformation and the “up” conformation, where down corresponds to the receptor-inaccessible



**Fig. 1. Structure of 2019-nCoV S in the prefusion conformation.** (A) Schematic of 2019-nCoV S primary structure colored by domain. Domains that were excluded from the ectodomain expression construct or could not be visualized in the final map are colored white. SS, signal sequence; S2', S2' protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Arrows denote protease cleavage sites. (B) Side and top views of the prefusion structure of the 2019-nCoV S protein with a single RBD in the up conformation. The two RBD down protomers are shown as cryo-EM density in either white or gray and the RBD up protomer is shown in ribbons colored corresponding to the schematic in (A).

Wrapp et al., *Science* 367, 1260–1263 (2020) | 13 March 2020

1 of 4

## Method of Single Particle Cryo-EM

20  
"for develop"

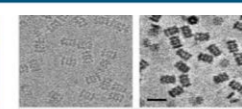


CellPress

Congratulations 2017 Nobel Laureates in Chemistry  
Jacques Dubochet, Joachim Frank and Richard Henderson!



© The Nobel Foundation



# Brief History of cryo-EM

130 NATURE, VOL. 217, JANUARY 13, 1968

## Reconstruction of Three Dimensional Structures from Electron Micrographs

by  
D. J. DE ROSIER  
A. KLUG  
MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge

General principles are formulated for the objective reconstruction of a three dimensional object from a set of electron microscope images. These principles are applied to the calculation of a three dimensional density map of the tail of bacteriophage T4.

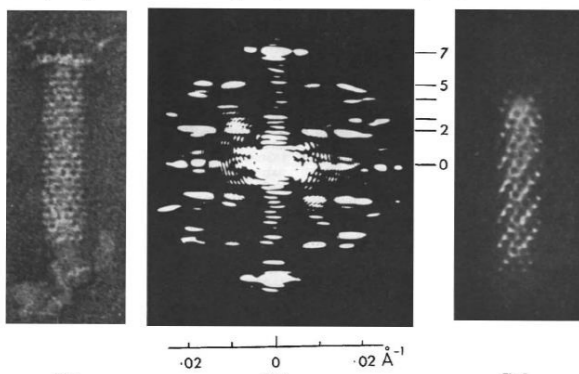
The standard high resolution electron microscope has a depth of focus of several thousand Angstroms making the image a two dimensional superposition of different levels in the three dimensional structure. The focus cannot be adjusted to different levels within the object, and so three dimensional structures are difficult to analyse. Stereo-electron micrographs do not overcome this difficulty satisfactorily, as will be shown.

effectively contains many different views of the structure. The symmetry of such an object can be introduced into the process of reconstruction, allowing the three dimensional structure to be reconstructed from a single view, or a small number of views. In principle, however, the method is applicable to any kind of structure, including individual, unsymmetrical particles, or sections of biological specimens.

### Summary of Procedure

Our method starts from the obvious premise that more than one view is generally needed to see an object in three dimensions. We determine first the number of views required for reconstructing an object to a given degree of resolution and find a systematic way of obtaining these views. The electron microscope images corresponding to these different views are then combined mathematically, by a procedure which is both quantitative and free from arbitrary assumptions, to give the three dimensional structure in a tangible and permanent form. The method is most powerful for objects containing symmetrically arranged subunits, for here a single image

Electron micrographs are selected in which the details of the structure show up best, as judged for example in the phage tail described later, by their optical diffraction patterns<sup>1,2</sup>. The optical density in each image is sampled at regular points on a grid by an automatic microdensitometer linked to a computer (unpublished work of U. W. Arndt, K. A. Crowther and J. F. W. Mallett), which converts the image into a set of numbers representing the density at each grid point. These numbers are now transformed by computation into a set of Fourier



50

0

1968

1975

1985

1995

2005

2015

2021

### RESEARCH

## CORONAVIRUS Cryo-EM structure of the prefusion conformation

Daniel Wrapp<sup>1</sup>, Nians Okubakola Abiona<sup>2</sup>, B.

The outbreak of a novel public health emergency for vaccines, therapeutic we determined a 3.5-Å S trimer in the prefusion receptor-binding domain biophysical and structural (ACE2) with higher resolution. We tested several public health appreciable binding between the two RBD: evaluation of medical

The novel coronavirus emerged in the city of Wuhan, causing illness, and presently named COVID-19. The World Health Organization on 11 February 2020, confirmed cases globally 1800 deaths. The rapidly characterized betacoronavirus general coronavirus syndrome compared with SARS to be more readily transmissible, spreading to the WHO Health Emergency (PHEIC) on 30 January 2020. COVID-19 makes a spike (S) protein that exists in a conformation that undergoes rearrangement with the host cell membrane. The receptor-binding domain (RBD) of S1 is triggered when the cell receptor, receptor-binding domain (RBD) of S1 undergoes hinge-like conformational movements that transiently hide or

Multiple factors determine the attainable resolution of reconstructions from single-particle cryo-EM. However, for biological macromolecules, the radiation damage caused by electron interactions with the sample is a fundamental limitation. To preserve the molecular structure, damage is restricted by carefully limiting the number of electrons used for imaging. The resulting counting statistics lead to high levels of noise. The signal-to-noise ratio (SNR) of cryo-EM images drops rapidly with spatial frequency, and at higher spatial frequencies the noise is typically orders of magnitude higher than the signal.

High-resolution reconstructions can still be calculated by averaging over many images of individual particles, provided that their relative orientations can be determined. However, because noise reduction scales with the square root of the number of particles, and because higher SNRs lead to more accurate orientations, acquiring more particles is often less efficient than increasing the SNR in the images. Consequently, although microscope automation and faster image processing programs have allowed reconstructions from larger datasets in recent years, increasing the SNR of the raw data may lead to bigger improvements, as is illustrated by the sudden increase in

cryo-EM resolutions with the introduction of more sensitive direct electron cameras in 2015<sup>3,4</sup>.

Here we describe the effect of three technological developments that further increase the SNR of cryo-EM images: a new cold field emission electron gun (CFEG), a new energy filter and the latest generation Falcon direct electron camera (Fig. 1a). Combined, these developments lead to a marked increase in the achievable resolution that ultimately enables the visualization of individual protein atoms in optimized samples.

Electron source optimized for energy spread

The source inside the microscope emits electrons with a range of different wavelengths, or energies. Because not all of these electrons can be focused in the same plane owing to chromatic aberration in the objective lens, the energy spread of the electrons leads to a blur in the images. The corresponding loss in SNR increases with spatial frequency and is described by an envelope on the contrast transfer function (CTF). Many state-of-the-art electron microscopes are equipped with a field emission gun (FEG) that is operated at a temperature of 1,700–1,800 K

Multiple factors determine the attainable resolution of reconstructions from single-particle cryo-EM. However, for biological macromolecules, the radiation damage caused by electron interactions with the sample is a fundamental limitation. To preserve the molecular structure, damage is restricted by carefully limiting the number of electrons used for imaging. The resulting counting statistics lead to high levels of noise. The signal-to-noise ratio (SNR) of cryo-EM images drops rapidly with spatial frequency, and at higher spatial frequencies the noise is typically orders of magnitude higher than the signal.

High-resolution reconstructions can still be calculated by averaging over many images of individual particles, provided that their relative orientations can be determined. However, because noise reduction scales with the square root of the number of particles, and because higher SNRs lead to more accurate orientations, acquiring more particles is often less efficient than increasing the SNR in the images. Consequently, although microscope automation and faster image processing programs have allowed reconstructions from larger datasets in recent years, increasing the SNR of the raw data may lead to bigger improvements, as is illustrated by the sudden increase in

cryo-EM resolutions with the introduction of more sensitive direct electron cameras in 2015<sup>3,4</sup>.

Here we describe the effect of three technological developments that further increase the SNR of cryo-EM images: a new cold field emission electron gun (CFEG), a new energy filter and the latest generation Falcon direct electron camera (Fig. 1a). Combined, these developments lead to a marked increase in the achievable resolution that ultimately enables the visualization of individual protein atoms in optimized samples.

Electron source optimized for energy spread

The source inside the microscope emits electrons with a range of different wavelengths, or energies. Because not all of these electrons can be focused in the same plane owing to chromatic aberration in the objective lens, the energy spread of the electrons leads to a blur in the images. The corresponding loss in SNR increases with spatial frequency and is described by an envelope on the contrast transfer function (CTF). Many state-of-the-art electron microscopes are equipped with a field emission gun (FEG) that is operated at a temperature of 1,700–1,800 K

### Article

## Single-particle cryo-EM at atomic resolution

https://doi.org/10.1038/s41586-020-2829-0  
Received: 22 May 2020  
Accepted: 27 August 2020  
Published online: 21 October 2020  
Check for updates

Takanori Nakane<sup>1</sup>, Abhay Kotecha<sup>1</sup>, Andrija Sente<sup>1,2</sup>, Greg McMullan<sup>1</sup>, Simonas Maslulis<sup>1,2</sup>, Patricia M. G. E. Brown<sup>1</sup>, Joana T. Grigorova<sup>1</sup>, Lina Malinauskaitė<sup>1</sup>, Tomas Malinauskas<sup>1</sup>, Jonas Mieling<sup>1</sup>, Tomasz Uchalski<sup>1,3</sup>, Lingbo Yu<sup>1</sup>, Dimple Karia<sup>1</sup>, Evgeniya V. Pechnikova<sup>1</sup>, Erwin de Jong<sup>1</sup>, Jeroen Koster<sup>1</sup>, Maarten Bisschoff<sup>1</sup>, Jamie McCosker<sup>1</sup>, Peter Tiemrijer<sup>1</sup>, Steven W. Hardwick<sup>1</sup>, Dimitri Y. Chirgadze<sup>1</sup>, Garib Murshudov<sup>1</sup>, A. Radu Aricescu<sup>1,2</sup> & Sjors H. W. Scheres<sup>1,2</sup>

The three-dimensional positions of atoms in protein molecules define their structure and their roles in biological processes. The more precisely atomic coordinates are determined, the more chemical information can be derived and the more mechanistic insights into protein function may be inferred. Electron cryo-microscopy (cryo-EM) single-particle analysis has yielded protein structures with increasing levels of detail in recent years<sup>1,2</sup>. However, it has proved difficult to obtain cryo-EM reconstructions with sufficient resolution to visualize individual atoms in proteins. Here we use a new electron source, energy filter and camera to obtain a 1.7 Å resolution cryo-EM reconstruction for a human membrane protein, the β3 GABA<sub>A</sub> receptor homopentamer<sup>3</sup>. Such maps allow a detailed understanding of small-molecule coordination, visualization of solvent molecules and alternative conformations for multiple amino acids, and unambiguous building of ordered acidic side chains and glycans. Applied to mouse apoferritin, our strategy led to a 1.22 Å resolution reconstruction that offers a genuine atomic-resolution view of a protein molecule using single-particle cryo-EM. Moreover, the scattering potential from many hydrogen atoms can be visualized in difference maps, allowing a direct analysis of hydrogen-bonding networks. Our technological advances, combined with further approaches to accelerate data acquisition and improve sample quality, provide a route towards routine application of cryo-EM in high-throughput screening of small molecule modulators and structure-based drug discovery.

Multiple factors determine the attainable resolution of reconstructions from single-particle cryo-EM. However, for biological macromolecules, the radiation damage caused by electron interactions with the sample is a fundamental limitation. To preserve the molecular structure, damage is restricted by carefully limiting the number of electrons used for imaging. The resulting counting statistics lead to high levels of noise. The signal-to-noise ratio (SNR) of cryo-EM images drops rapidly with spatial frequency, and at higher spatial frequencies the noise is typically orders of magnitude higher than the signal.

High-resolution reconstructions can still be calculated by averaging over many images of individual particles, provided that their relative orientations can be determined. However, because noise reduction scales with the square root of the number of particles, and because higher SNRs lead to more accurate orientations, acquiring more particles is often less efficient than increasing the SNR in the images. Consequently, although microscope automation and faster image processing programs have allowed reconstructions from larger datasets in recent years, increasing the SNR of the raw data may lead to bigger improvements, as is illustrated by the sudden increase in

cryo-EM resolutions with the introduction of more sensitive direct electron cameras in 2015<sup>3,4</sup>.

Here we describe the effect of three technological developments that further increase the SNR of cryo-EM images: a new cold field emission electron gun (CFEG), a new energy filter and the latest generation Falcon direct electron camera (Fig. 1a). Combined, these developments lead to a marked increase in the achievable resolution that ultimately enables the visualization of individual protein atoms in optimized samples.

Electron source optimized for energy spread

The source inside the microscope emits electrons with a range of different wavelengths, or energies. Because not all of these electrons can be focused in the same plane owing to chromatic aberration in the objective lens, the energy spread of the electrons leads to a blur in the images. The corresponding loss in SNR increases with spatial frequency and is described by an envelope on the contrast transfer function (CTF). Many state-of-the-art electron microscopes are equipped with a field emission gun (FEG) that is operated at a temperature of 1,700–1,800 K

Multiple factors determine the attainable resolution of reconstructions from single-particle cryo-EM. However, for biological macromolecules, the radiation damage caused by electron interactions with the sample is a fundamental limitation. To preserve the molecular structure, damage is restricted by carefully limiting the number of electrons used for imaging. The resulting counting statistics lead to high levels of noise. The signal-to-noise ratio (SNR) of cryo-EM images drops rapidly with spatial frequency, and at higher spatial frequencies the noise is typically orders of magnitude higher than the signal.

High-resolution reconstructions can still be calculated by averaging over many images of individual particles, provided that their relative orientations can be determined. However, because noise reduction scales with the square root of the number of particles, and because higher SNRs lead to more accurate orientations, acquiring more particles is often less efficient than increasing the SNR in the images. Consequently, although microscope automation and faster image processing programs have allowed reconstructions from larger datasets in recent years, increasing the SNR of the raw data may lead to bigger improvements, as is illustrated by the sudden increase in

cryo-EM resolutions with the introduction of more sensitive direct electron cameras in 2015<sup>3,4</sup>.

Here we describe the effect of three technological developments that further increase the SNR of cryo-EM images: a new cold field emission electron gun (CFEG), a new energy filter and the latest generation Falcon direct electron camera (Fig. 1a). Combined, these developments lead to a marked increase in the achievable resolution that ultimately enables the visualization of individual protein atoms in optimized samples.

Electron source optimized for energy spread

The source inside the microscope emits electrons with a range of different wavelengths, or energies. Because not all of these electrons can be focused in the same plane owing to chromatic aberration in the objective lens, the energy spread of the electrons leads to a blur in the images. The corresponding loss in SNR increases with spatial frequency and is described by an envelope on the contrast transfer function (CTF). Many state-of-the-art electron microscopes are equipped with a field emission gun (FEG) that is operated at a temperature of 1,700–1,800 K

Method of Single Particle Cryo-EM

"for development"



Nature  
Met  
The  
cryo  
now  
stru

Wrapp et al., Science 367, 1260–1263 (2020) | 13 March 2020

CellPress

Congratulations 2017 Nobel Laureates in Chemistry  
Jacques Dubochet, Joachim Frank and Richard Henderson!

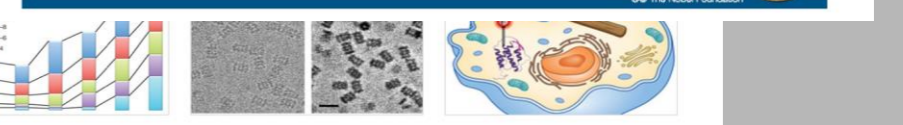


Fig. 1. Structure of 2019-nCoV S in the prefusion conformation. (A) Schematic of 2019-nCoV S primary structure colored by domain. Domains that were excluded from the ectodomain expression construct or could not be visualized in the final map are colored white. SS, signal sequence; S2', S2' protease cleavage site; FP, fusion peptide; HRI, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Arrows denote protease cleavage sites. (B) Side and top views of the prefusion structure of the 2019-nCoV S protein with a single RBD in the up conformation. The two RBD down protomers are shown as cryo-EM density in either white or gray and the RBD up protomer is shown in ribbons colored corresponding to the schematic in (A).

1 of 4



# Single-particle cryo-EM at atomic resolution

<https://doi.org/10.1038/s41586-020-2829-0>

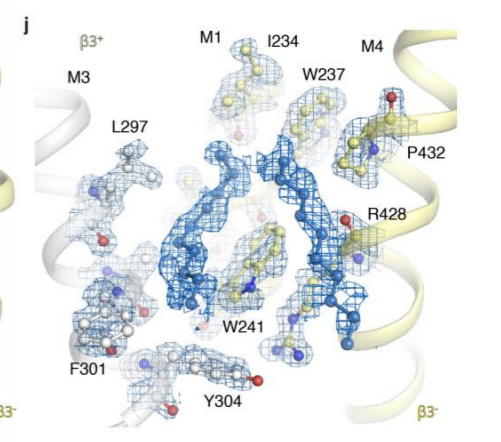
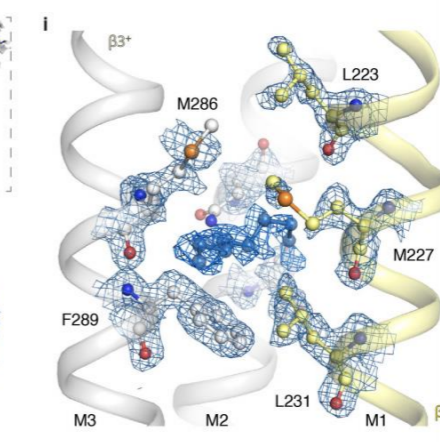
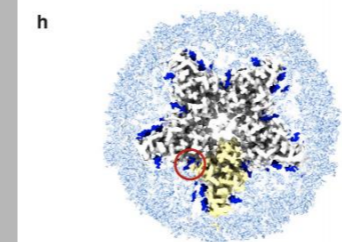
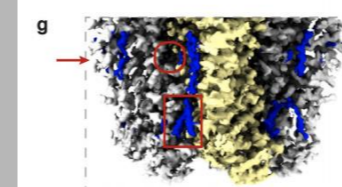
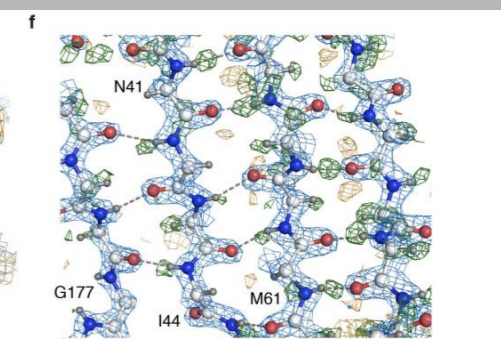
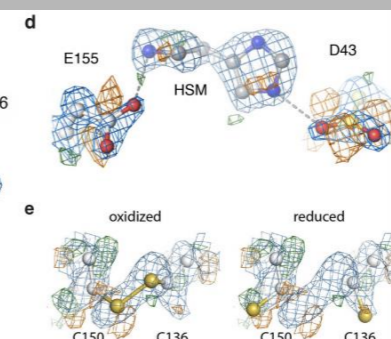
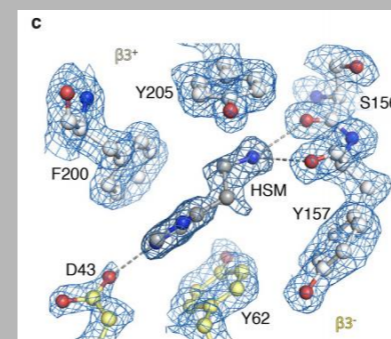
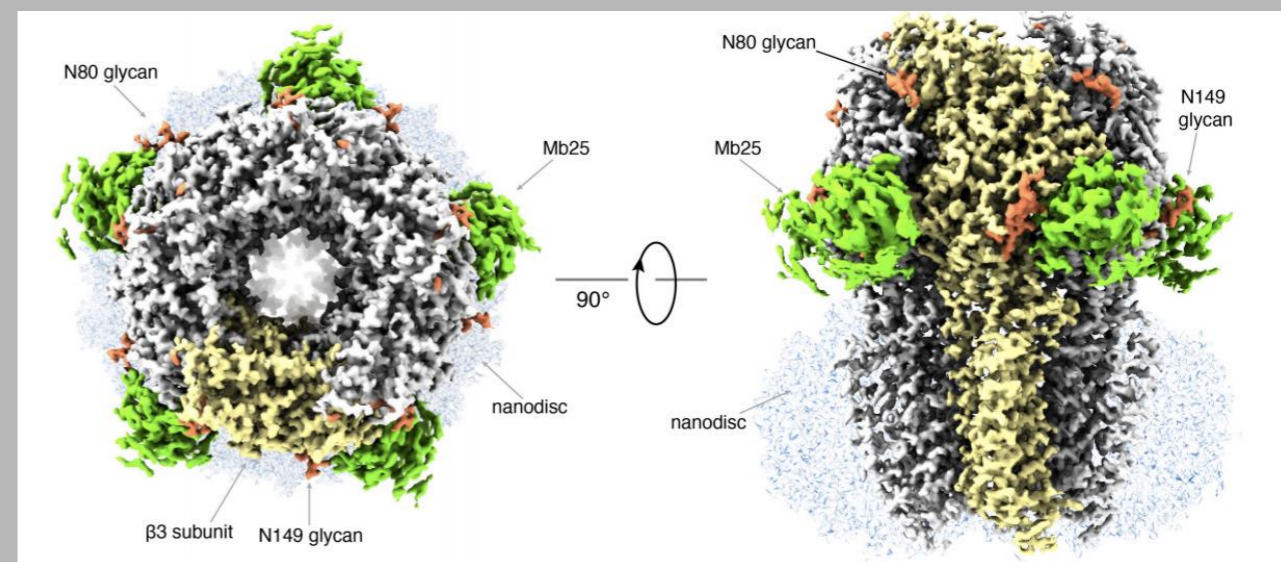
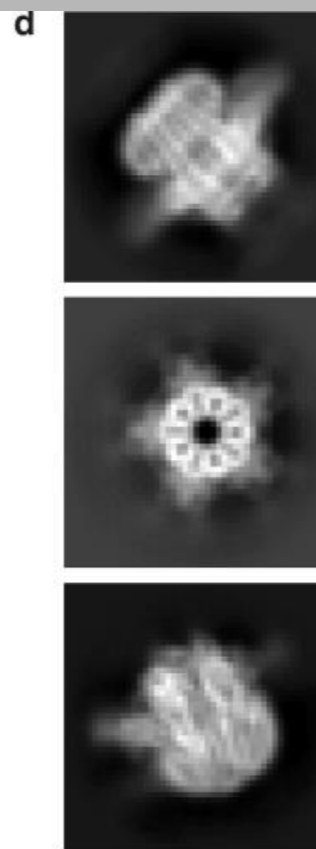
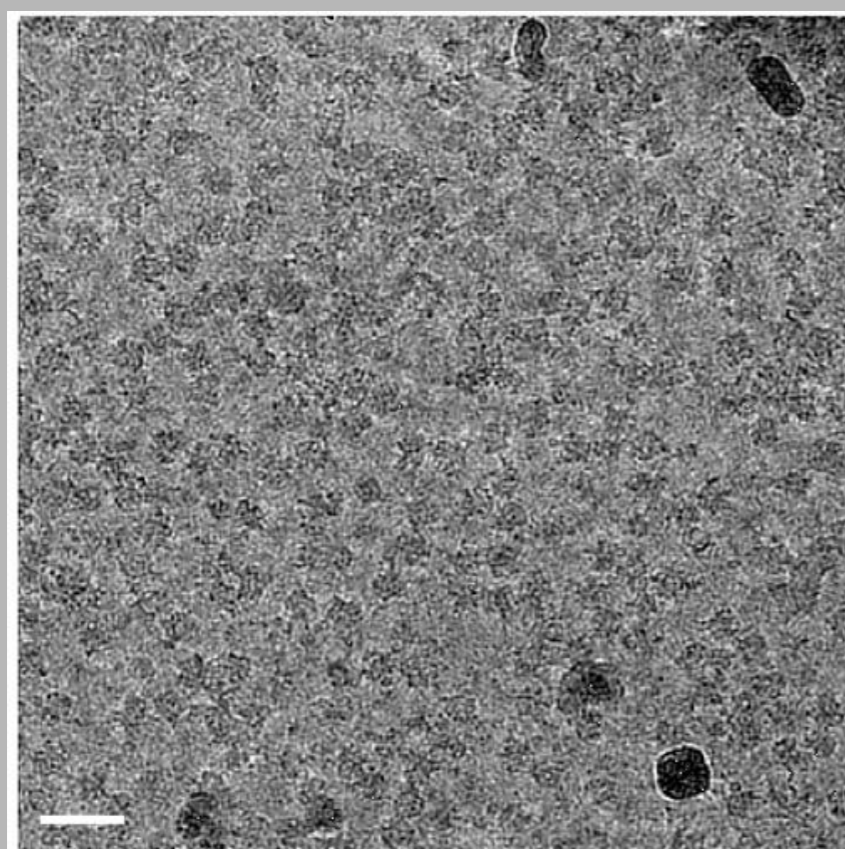
Received: 22 May 2020

Accepted: 27 August 2020

Published online: 21 October 2020

 Check for updates

Takanori Nakane<sup>1,9</sup>, Abhay Kotecha<sup>2,9</sup>, Andrija Sente<sup>1,9</sup>, Greg McMullan<sup>1</sup>, Simonas Masiulis<sup>1,7</sup>, Patricia M. G. E. Brown<sup>1</sup>, Ioana T. Grigoras<sup>1,8</sup>, Lina Malinauskaite<sup>1</sup>, Tomas Malinauskas<sup>3</sup>, Jonas Miehling<sup>1</sup>, Tomasz Uchański<sup>4,5</sup>, Lingbo Yu<sup>2</sup>, Dimple Karia<sup>2</sup>, Evgeniya V. Pechnikova<sup>2</sup>, Erwin de Jong<sup>2</sup>, Jeroen Keizer<sup>2</sup>, Maarten Bischoff<sup>2</sup>, Jamie McCormack<sup>2</sup>, Peter Tiemeijer<sup>2</sup>, Steven W. Hardwick<sup>6</sup>, Dimitri Y. Chirgadze<sup>6</sup>, Garib Murshudov<sup>1</sup>, A. Radu Aricescu<sup>1</sup> & Sjors H. W. Scheres<sup>1</sup>✉

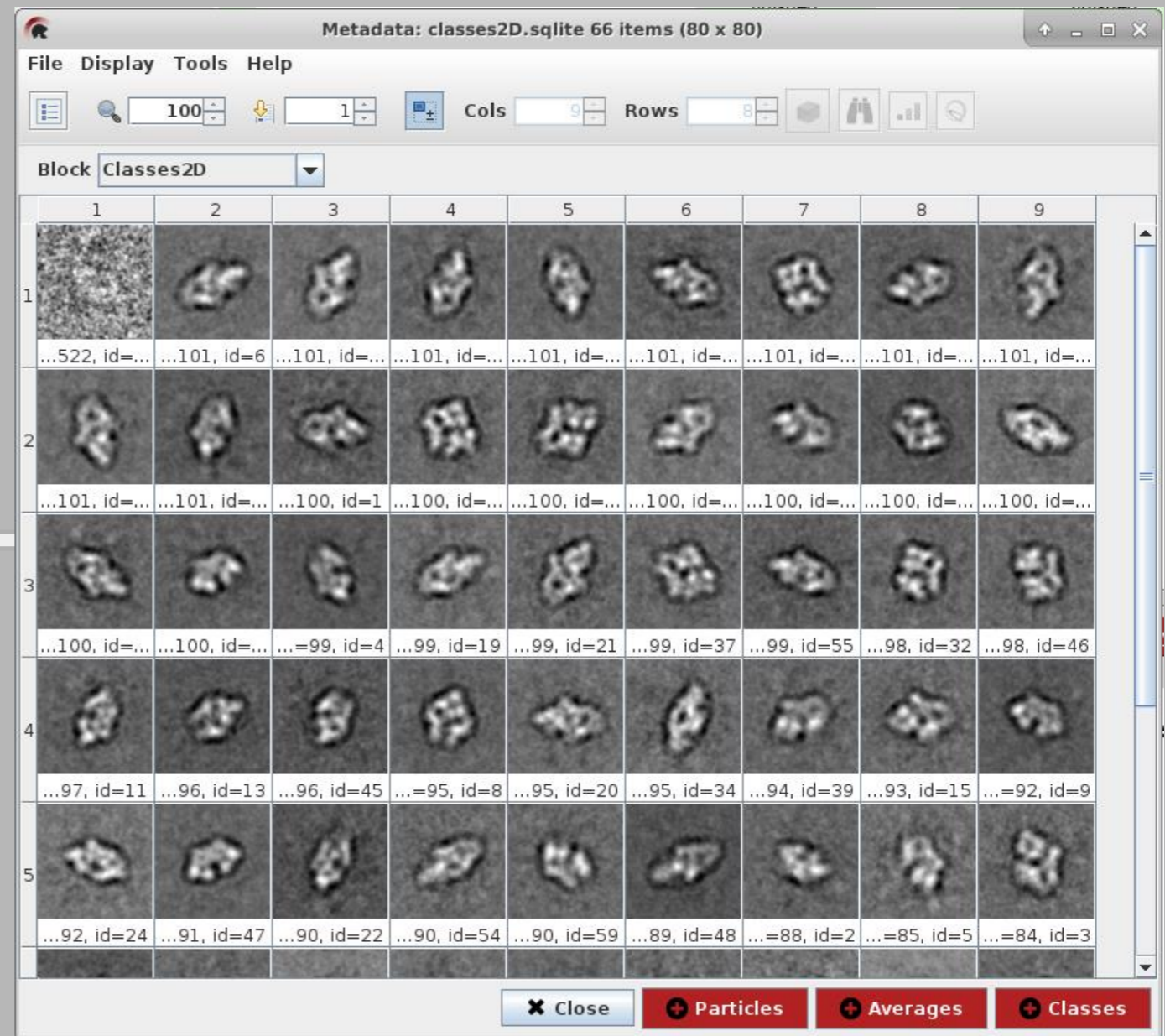
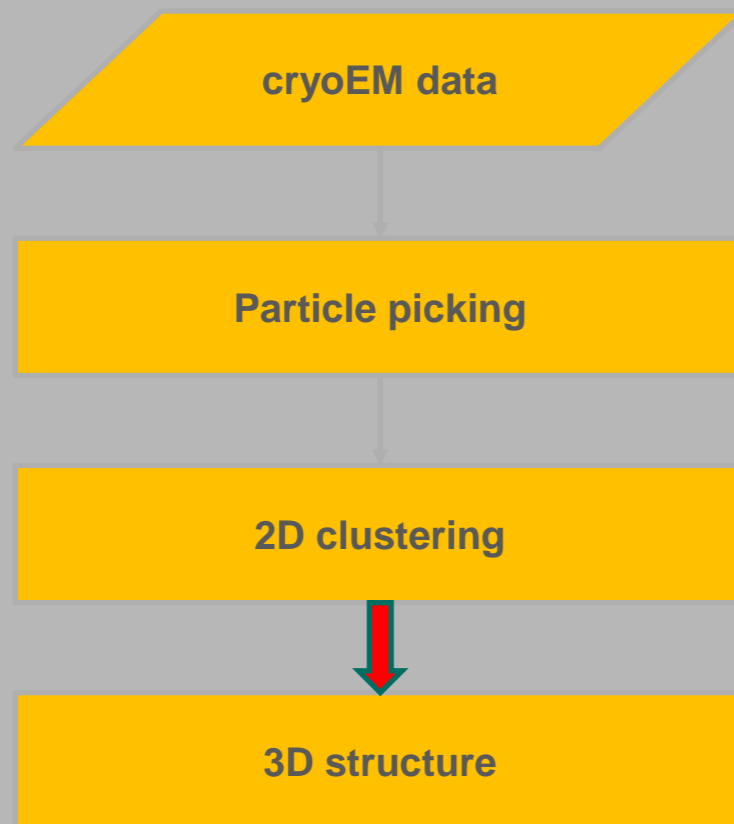


	Apoferritin	GABA <sub>A</sub> -β3	GABA <sub>A</sub> -β3	GABA <sub>A</sub> -β3	GABA <sub>A</sub> -β3
Number of collected movies	3370	8733	1244	1268	1556
Initial particle images (no.)	428590	1105069	192053	199867	242059
Final particle images (no.)	363126	233576	33166	42780	45009
Map resolution at FSC=0.143 (Å)	1.22	1.73	1.97	1.89	1.87

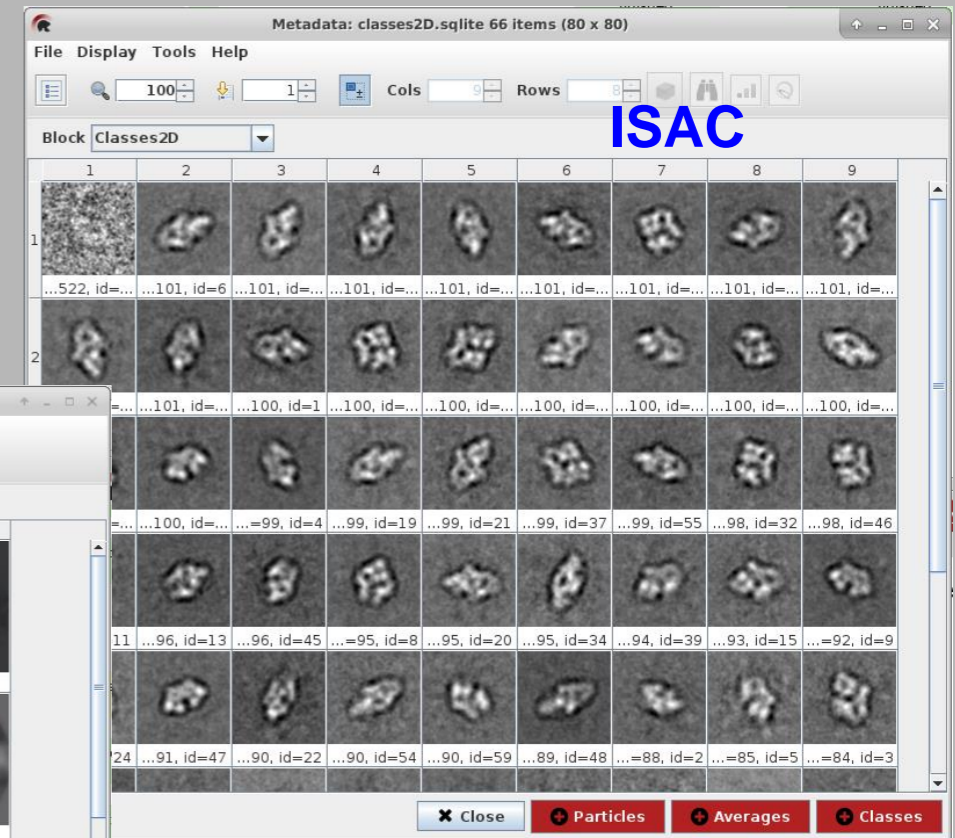
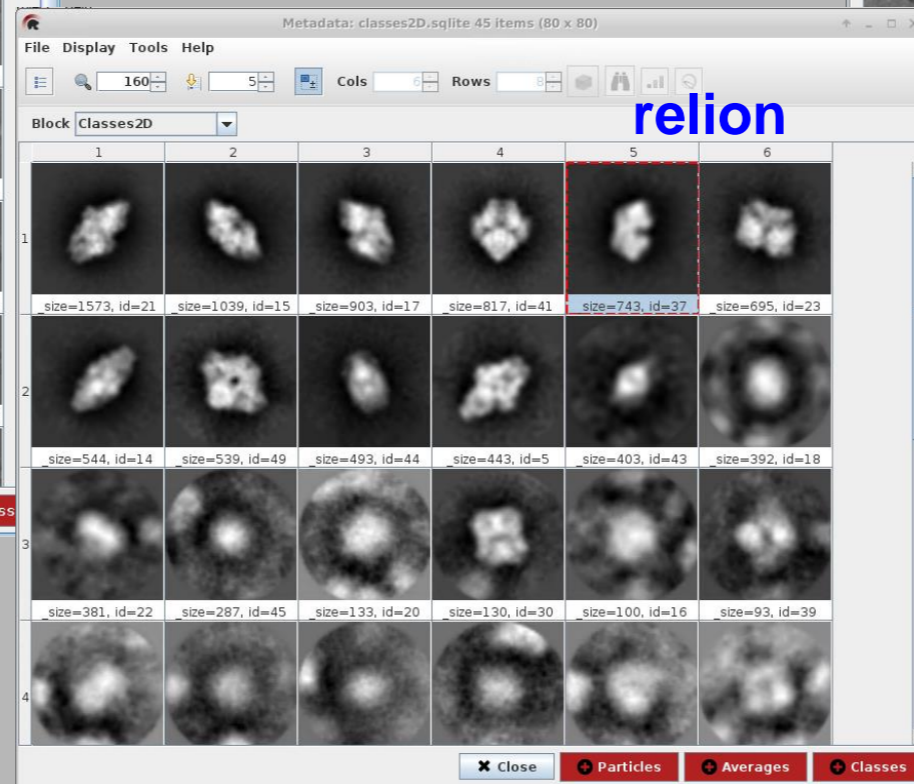
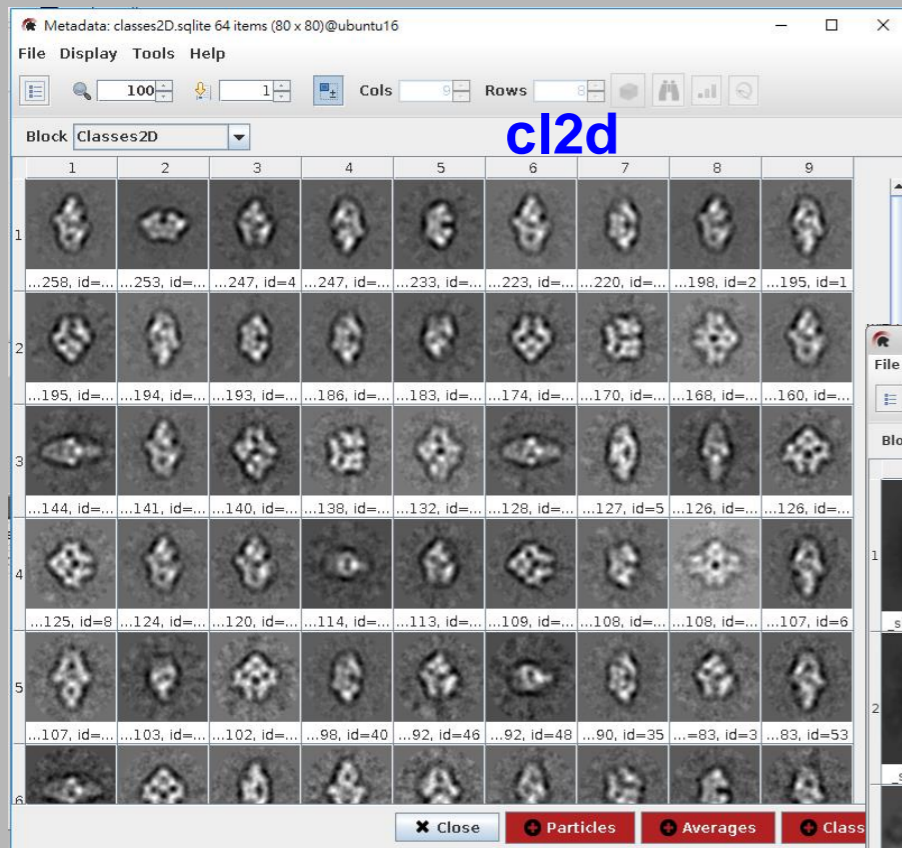
# Challenges

- Large Data Set (Thousands of movies)
- High Dimension (Pixel number)
- Heavy Noise
- Heterogeneous Conformations
- Missing Value Problem

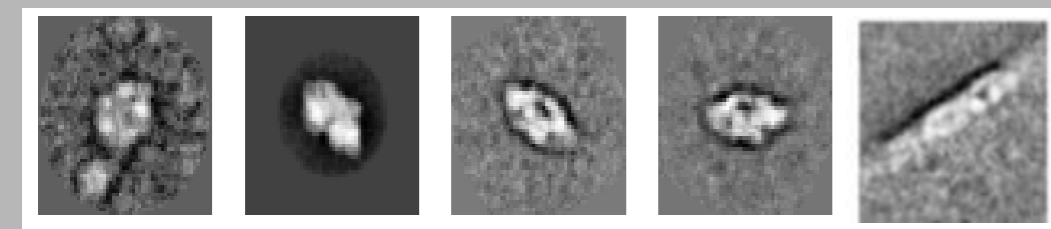
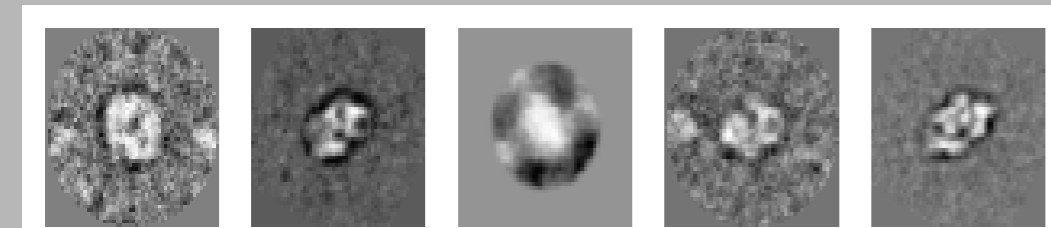
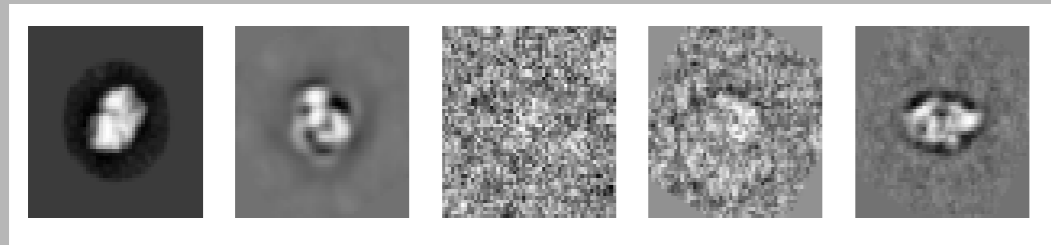
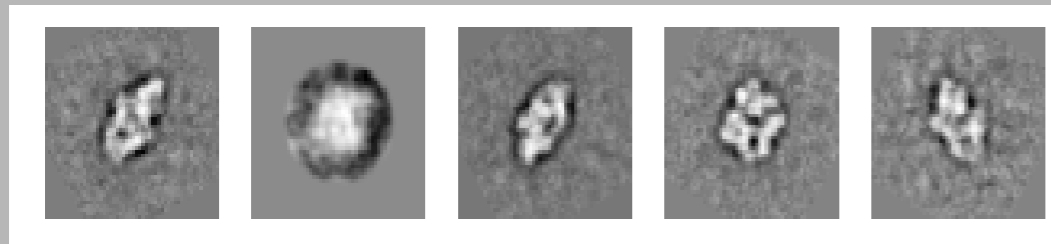
# We employ CNN to classify 2D clustering averages.



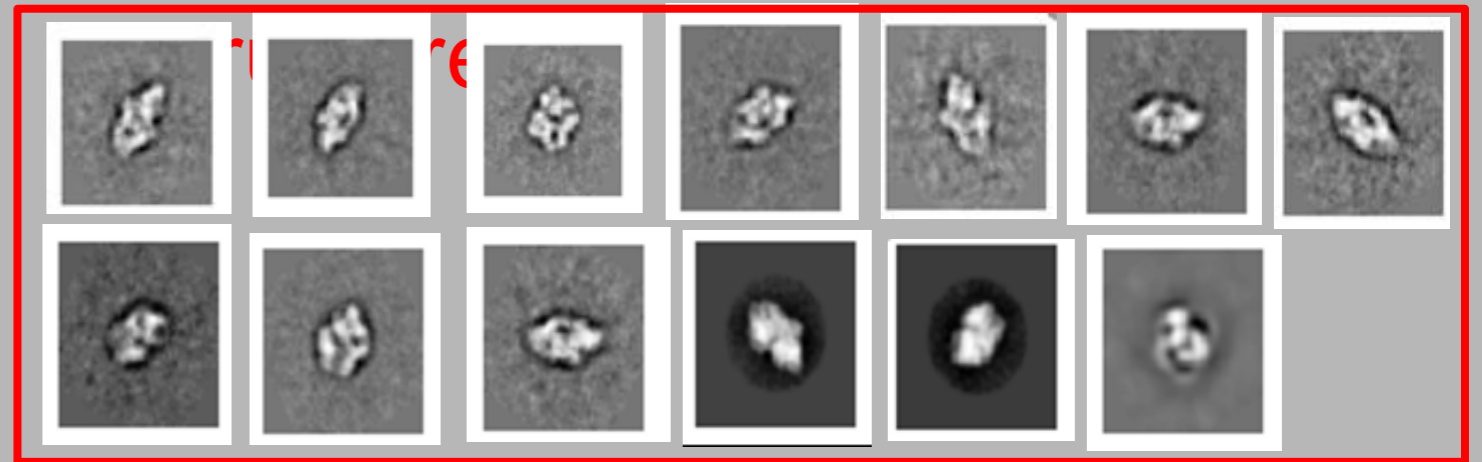
# One important function of 2D clustering is to sift valid particle images. It is manually executed!



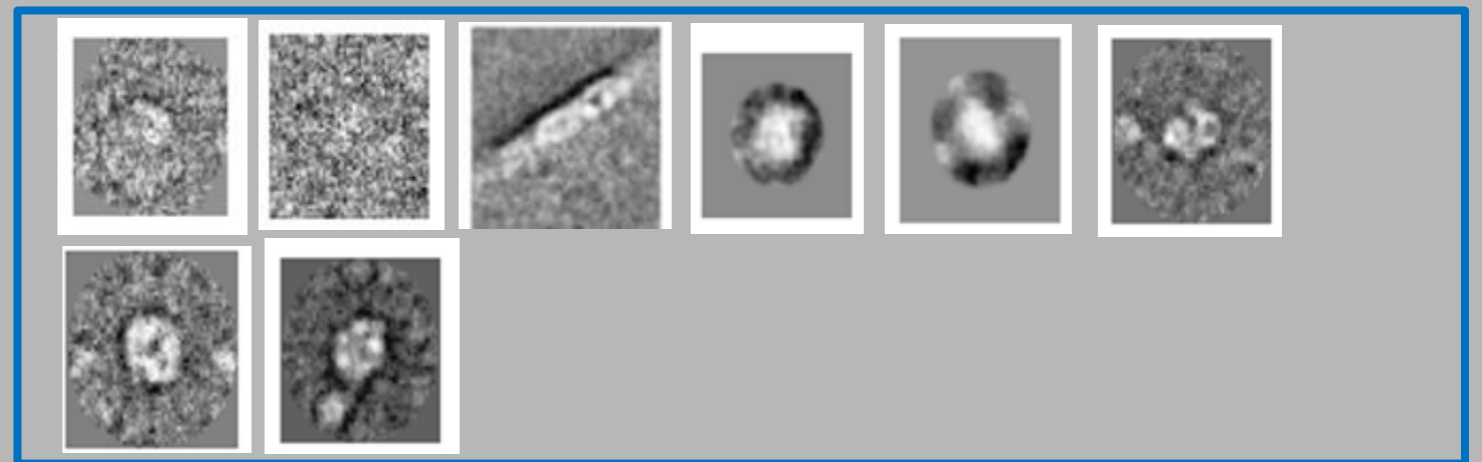
# Classification for 2D clustering averages





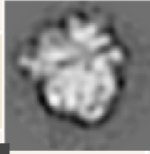
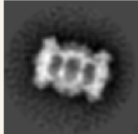
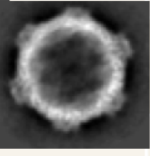
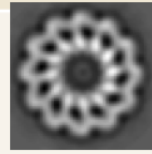
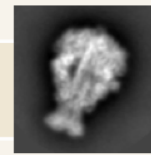

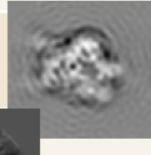
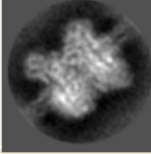
Accept for 3D



Discard



# Training Set

Figure	Name	id	Image size	Number(1/0)	Total Number
	Volta phase plate?		100×100	158(92/66)	3792
	betagal		100×100	200(47/153)	4800
	ribosome		130×130	500(218/282)	12000
	T20S proteasome	10057	200×200 250×250	79(7/72)	1869
	cowpea mosaic virus	10205	400×400	59(18/41)	1416
	Activated NAIP2/NLRC4 Inflammasome	10063	530×530	40(13/27)	960
	TcdA1	10089	350×350	60(27/33)	1440
	thermostabilized avian CFTR	10219	400×400	100(25/75)	2400
	Synaptic RAG1-RAG2 Complex	10049	192×192	150(64/86)	3600
	bovine liver glutamate dehydrogenase	10217	300×300	95(27/68)	2280
			Total	1441(538/903)	34584

# Model

---



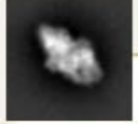
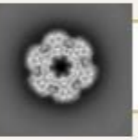
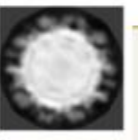
```
def BuildModel(training, training_label, size, num_filter,name):
    model=Sequential()
    model.add(Conv2D(filters=20,kernel_size=(30,30),input_shape=(100,100,1),activation='sigmoid',padding='same'))
    model.add(Conv2D(filters=20,kernel_size=(20,20),padding='same'))
    model.add(MaxPooling2D(pool_size=(10,10)))
    model.add(BatchNormalization())
    model.add(Activation( 'relu' ))
    model.add(Flatten())
    model.add(Dense(100,activation='sigmoid'))
    model.add(Dense(1,activation='sigmoid'))
    model.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])
    checkpoint = ModelCheckpoint(name+'model-{epoch:03d}.h5', verbose=1, monitor='val_loss',save_best_only=True, mode='auto')
    model.fit(training,training_label,epochs=20,batch_size=500,verbose=1, validation_data=val_set, callbacks=[checkpoint])
```

# Training & validation

	Training set		Validation set		
	loss	accuracy	loss	accuracy	
1	0.1545	0.9226	-0.158	0.9392	0.945
2	0.0318	0.9288	0.1343	0.9435	0.955
3	0.0324	0.9149	0.1495	0.9467	0.95
4	0.0259	0.9471	0.1953	0.9325	0.945
5	0.0506	0.9148	0.1497	0.9373	0.94
6	0.0032	0.9471	0.1698	0.9373	0.95
7	0.1042	0.9528	-0.2351	0.9498	0.955
8	0.1117	0.9013	0.1499	0.9429	0.945
9	0.0937	0.959	-0.1671	0.9275	0.925
10	-0.0071	0.9406	0.216	0.9329	0.935



# Testing

Particle		Set	loss	accuracy	Number(1/0)
<u>Betagal</u>		1	0.1874	0.97	100 (23/77)
		2	0.1104	0.9755	89 (21/68)
T20S proteasome		1	0.0941	0.9667	30 (4/26)
		2	0.2554	0.8667	30 (7/23)
<u>Betagal(Example)</u>		1	0.015	0.9792	48 (10/38)
		2	0.0057	1.00	30 (8/22)
GSP		1	0.5762	0.725	40 (9/31)
		2	1.3221	0.5789	95 (27/68)
		3	0.5056	0.8471	85 (17/68)
<u>Macrobrachium rosenbergii Nodavirus</u>		1	2.3056	0.7	20 (6/14)
		2	3.1729	0.6	10 (4/6)

# The performance of DCGAN model on single frame image super-resolution (SR).

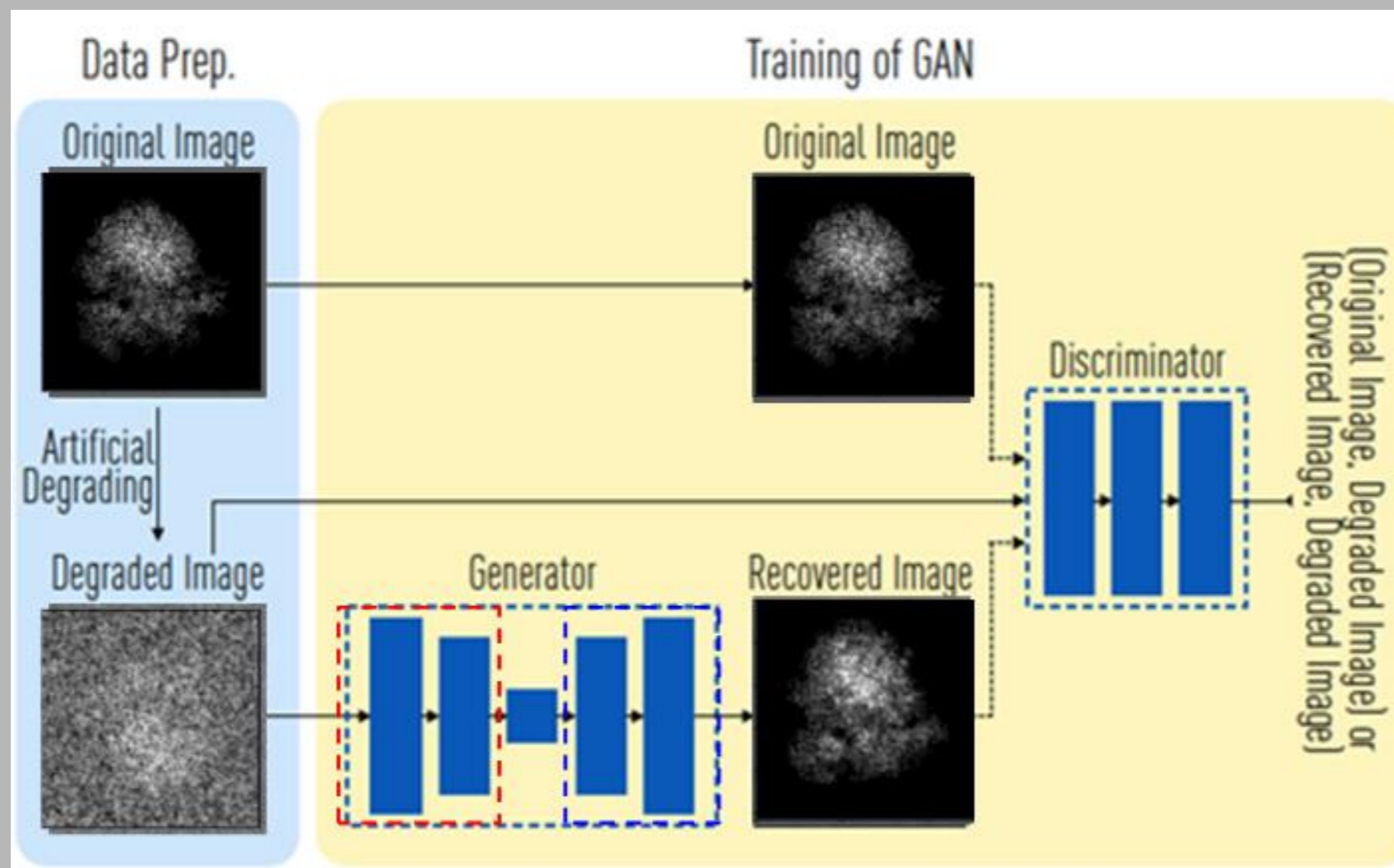
- We can see that DCGAN achieves slightly lower mean PSNR on the test set than the conventional bicubic method. However, if examining closely into the results of the DCGAN based superresolution images (2c), we can see that, even with some distortion, DCGAN provides finer details on the resulting images, which actually agrees more with human recognizing conventions.

## **DCGANs for image super-resolution, denoising and deblurring**

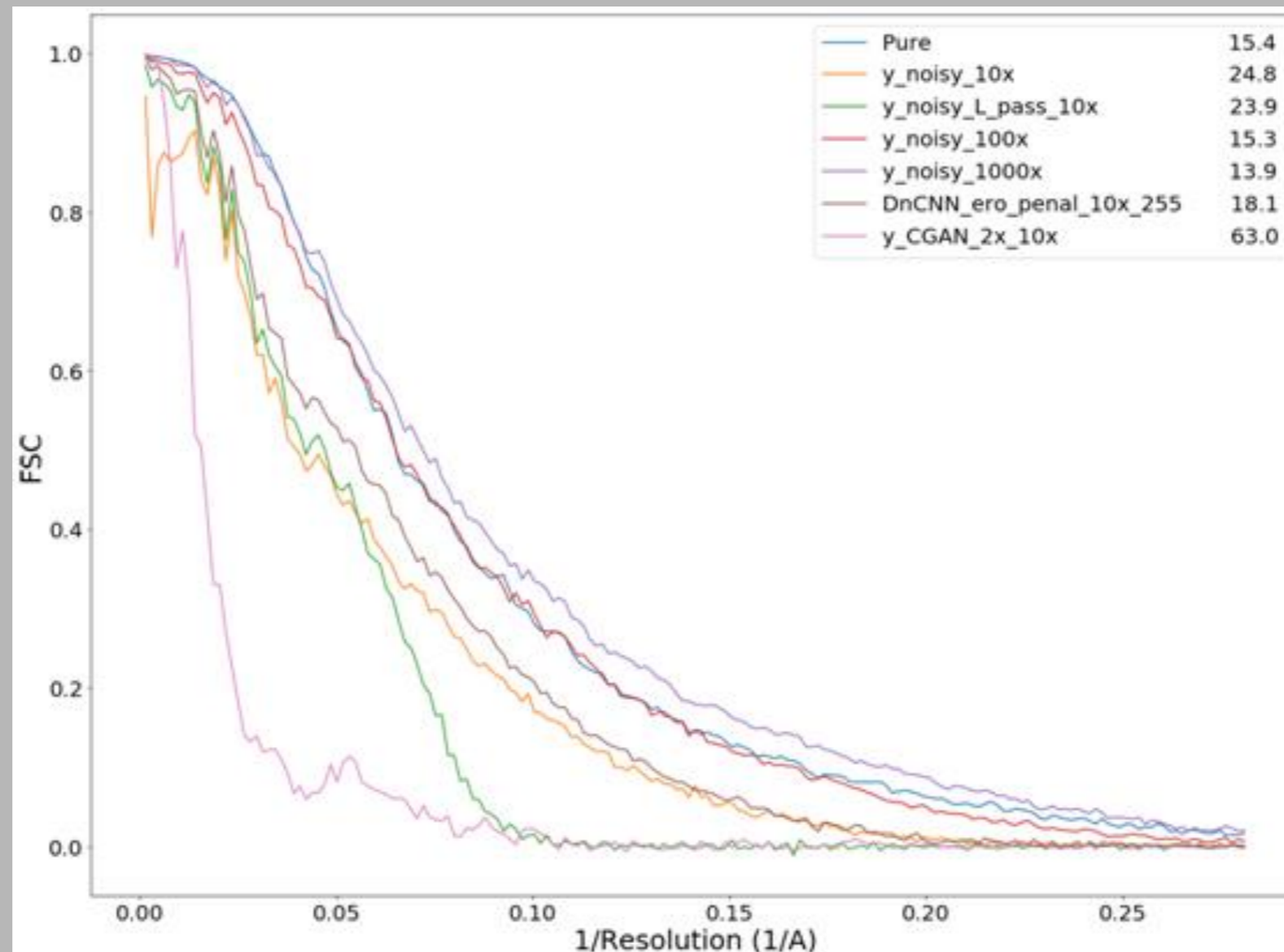
Qiaojing Yan  
Stanford University  
Electrical Engineering  
qiaojing@stanford.edu

Wei Wang  
Stanford University  
Electrical Engineering  
wwang23@stanford.edu

An ambitious project: Pushing the resolution by GAN, collaborated with SW Chen.

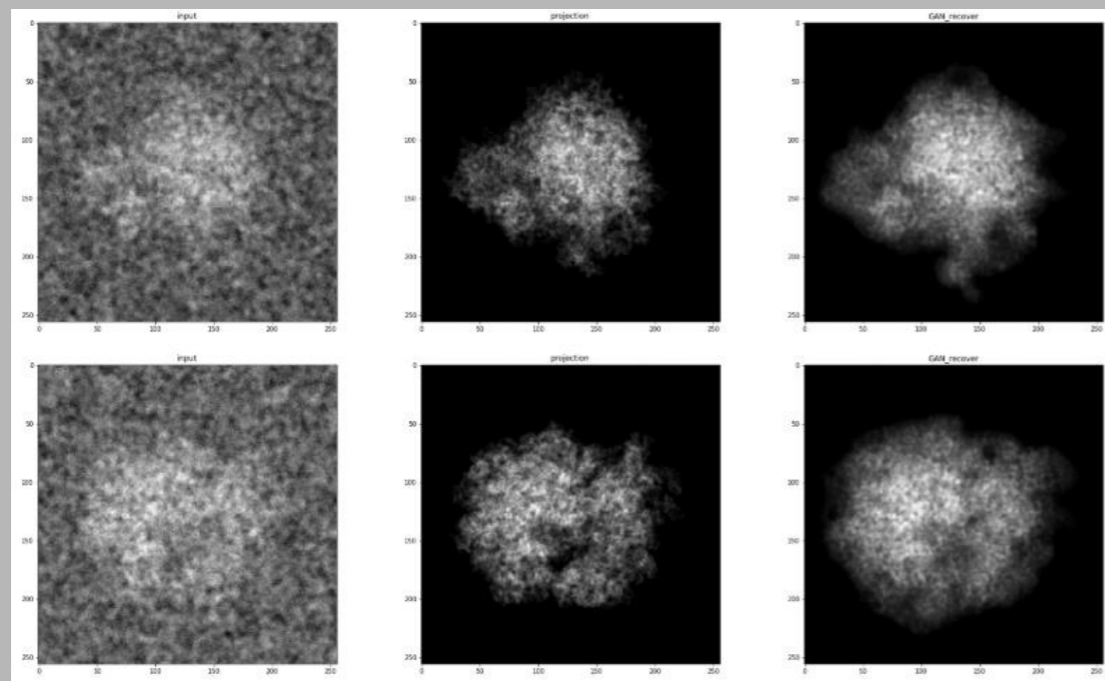


# 3D reconstruction using GAN de-noised images is not as expected,



# cGAN Approach: 2D de-noised images look great!

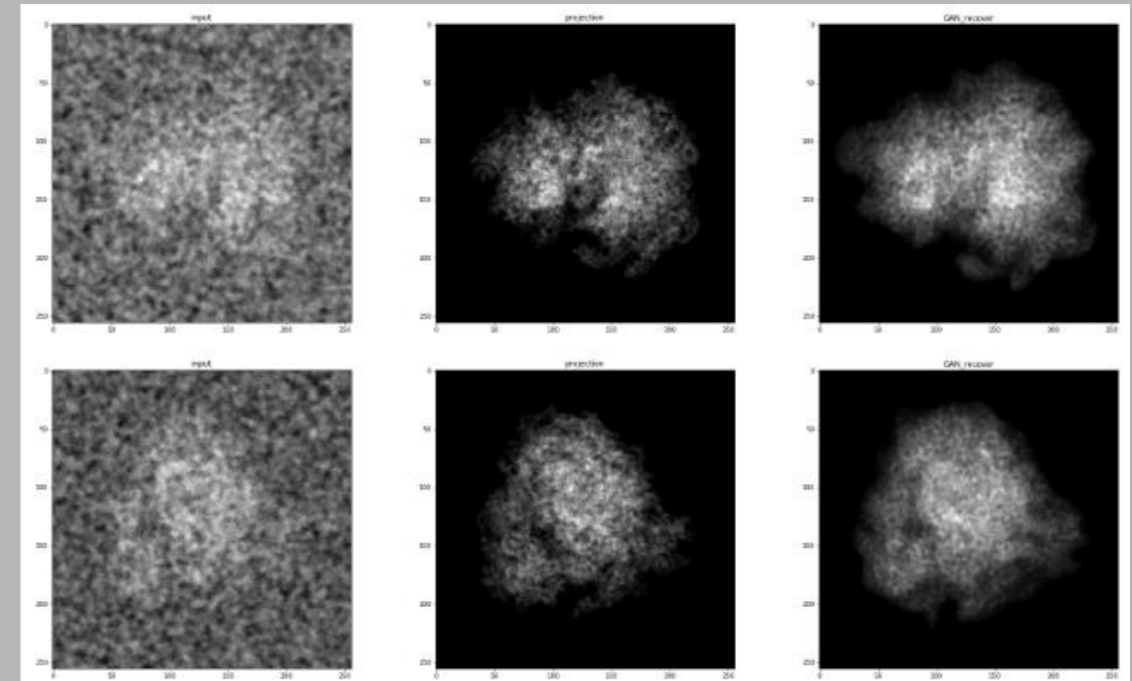
- Training set
  - 231 projections of ribosome plus Poisson noise
  - 8251 paired data
- Testing set
  - 720 images



Input

Ground Truth

Output



Input

Ground Truth

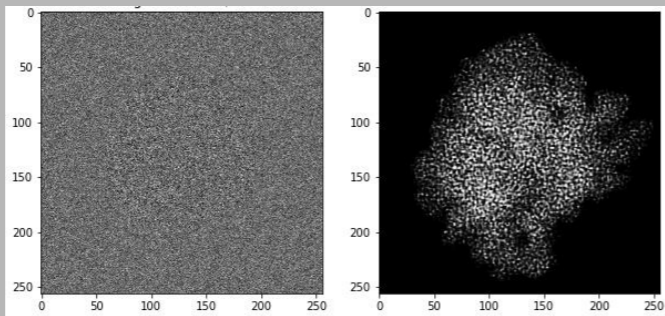
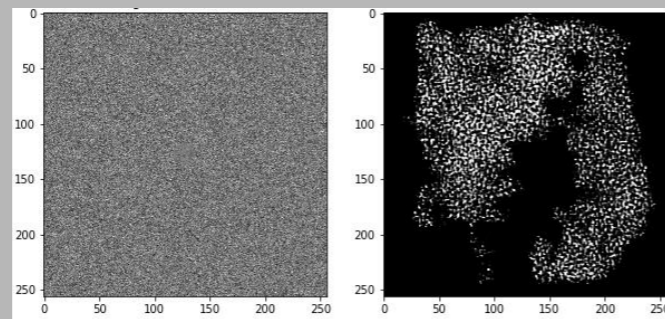
Output

# Is this real science? cGAN is so powerful that it can transform purely noise images.

- The output images by cGAN could not yield reasonable 3D structure reconstruction.
- Pure or nearly noise images may also output likely particle images.
  - It is dangerous to use in field like Cryo-EM where the image SNR is low.

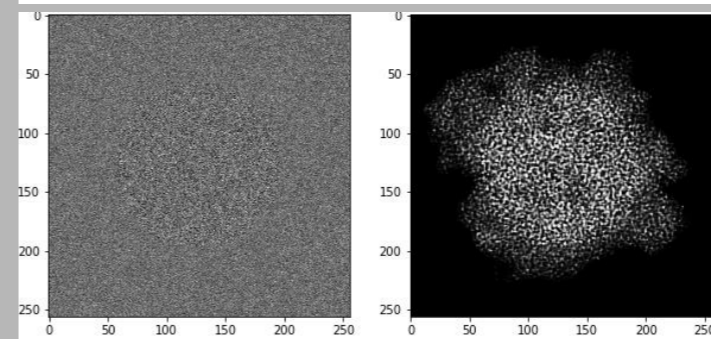
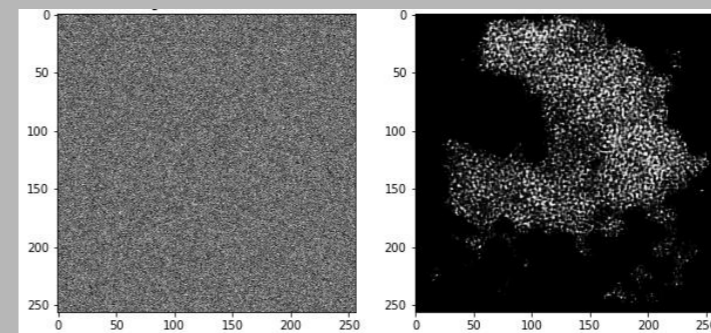
Input  
Random Noise

Output

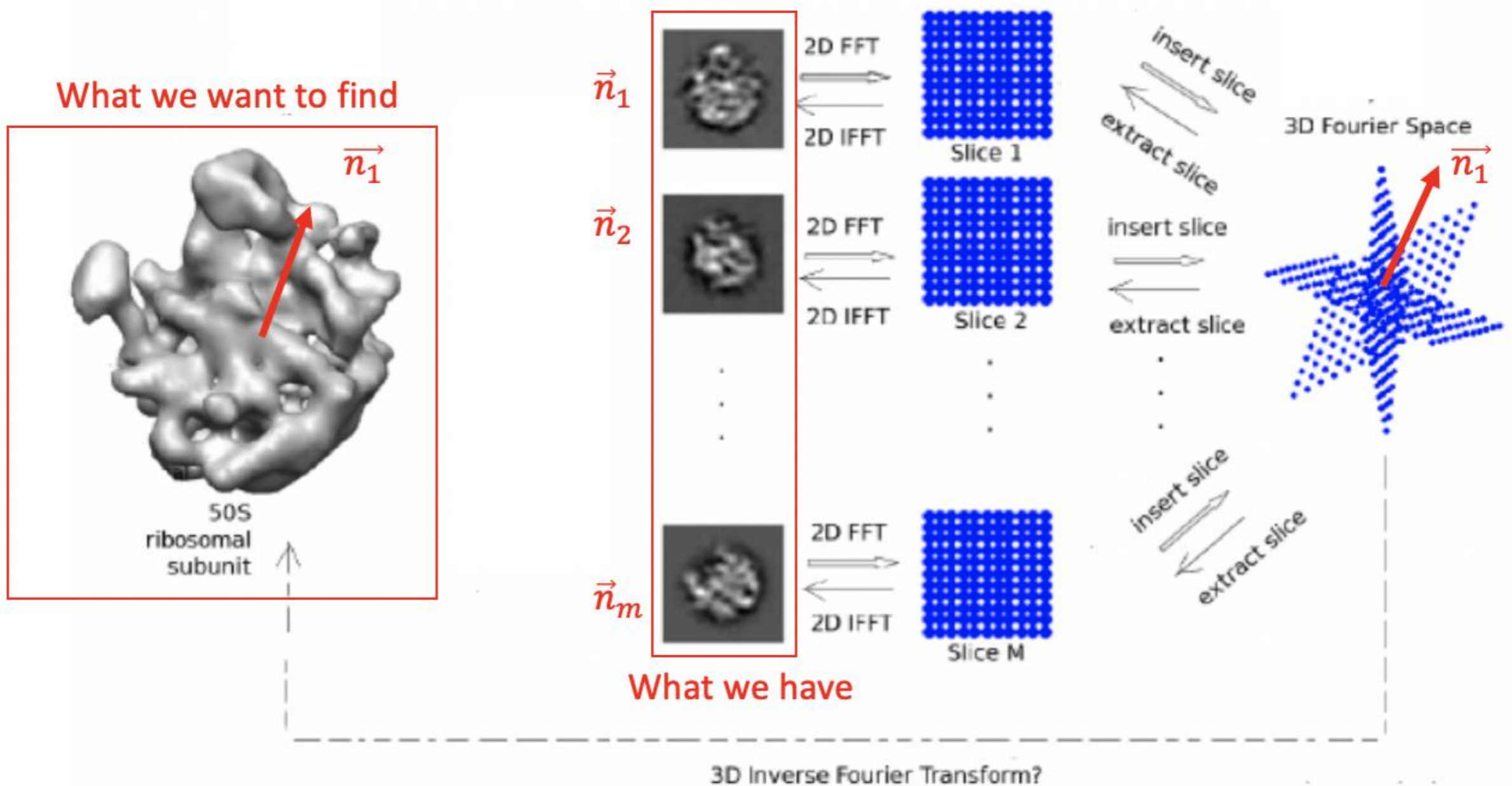


Input  
Random Noise

Output



Domain Knowledge:  
Fourier Slice Theorem and Common Line.  
Common Line implicitly holds the particle images as  
a structure, but neglected in Deep Learning.



# Future work: Integration of AlphaFold and ASCEP

## Article

# Improved protein structure prediction using potentials from deep learning

## AlphaFold

<https://doi.org/10.1038/s41586-019-1923-7>

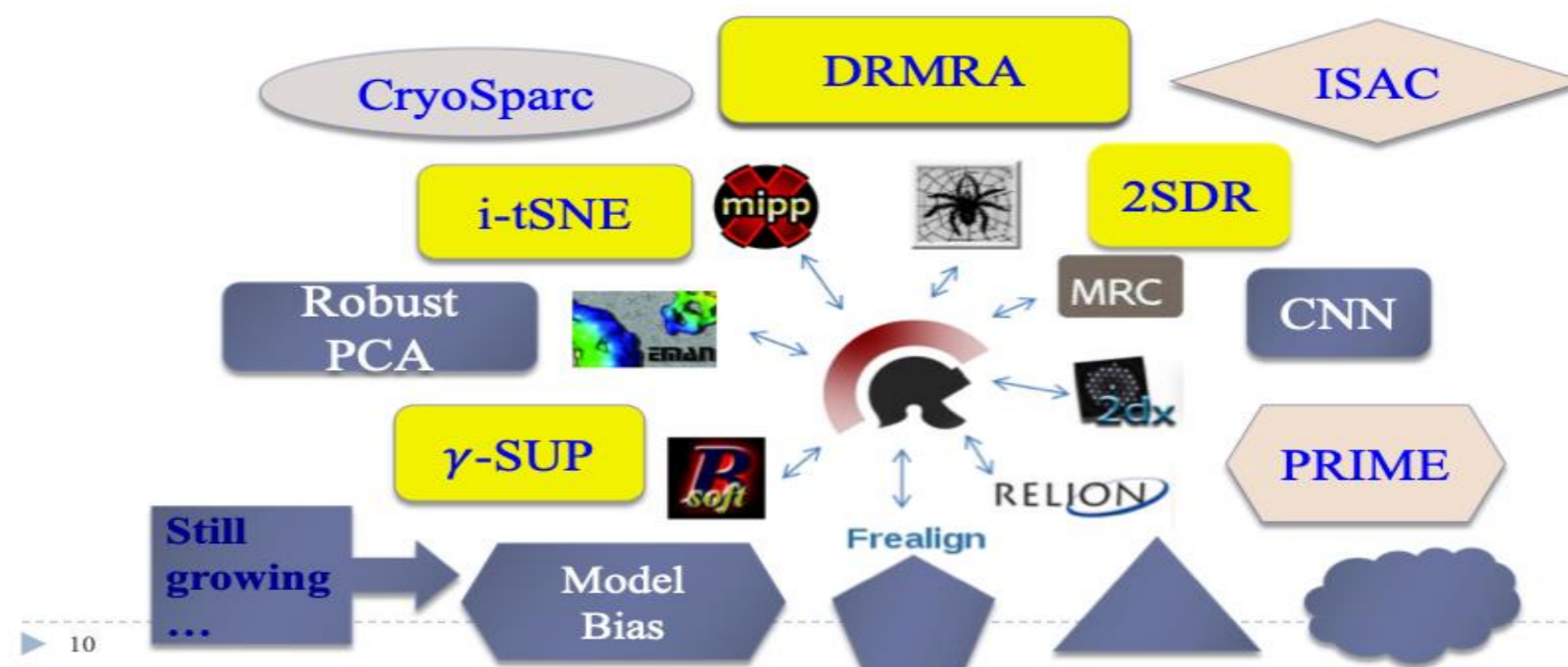
Received: 2 April 2019

Accepted: 10 December 2019

Published online: 15 January 2020

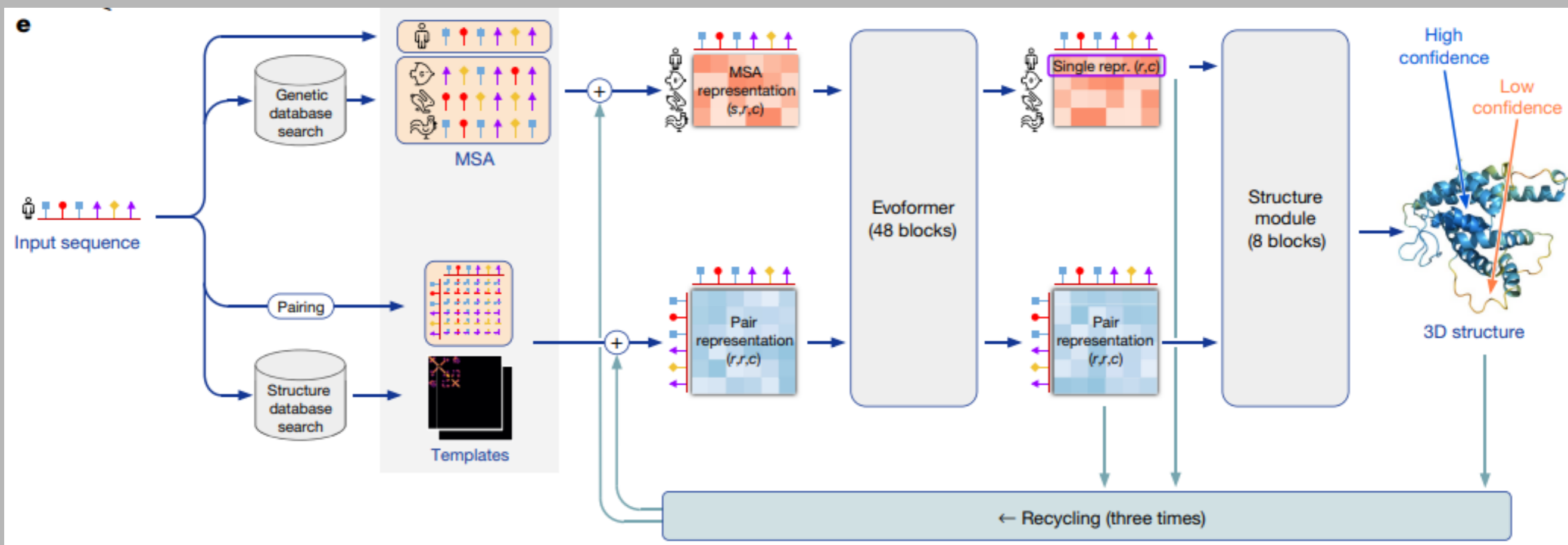
Andrew W. Senior<sup>1,4\*</sup>, Richard Evans<sup>1,4</sup>, John Jumper<sup>1,4</sup>, James Kirkpatrick<sup>1,4</sup>, Laurent Sifre<sup>1,4</sup>, Tim Green<sup>1</sup>, Chongli Qin<sup>1</sup>, Augustin Židek<sup>1</sup>, Alexander W. R. Nelson<sup>1</sup>, Alex Bridgland<sup>1</sup>, Hugo Penedones<sup>1</sup>, Stig Petersen<sup>1</sup>, Karen Simonyan<sup>1</sup>, Steve Crossan<sup>1</sup>, Pushmeet Kohli<sup>1</sup>, David T. Jones<sup>2,3</sup>, David Silver<sup>1</sup>, Koray Kavukcuoglu<sup>1</sup> & Demis Hassabis<sup>1</sup>

**ASCEP**: We employ Scipion to integrate competitive algorithms to process experimental cryo-EM data.





# AlphaFold 2

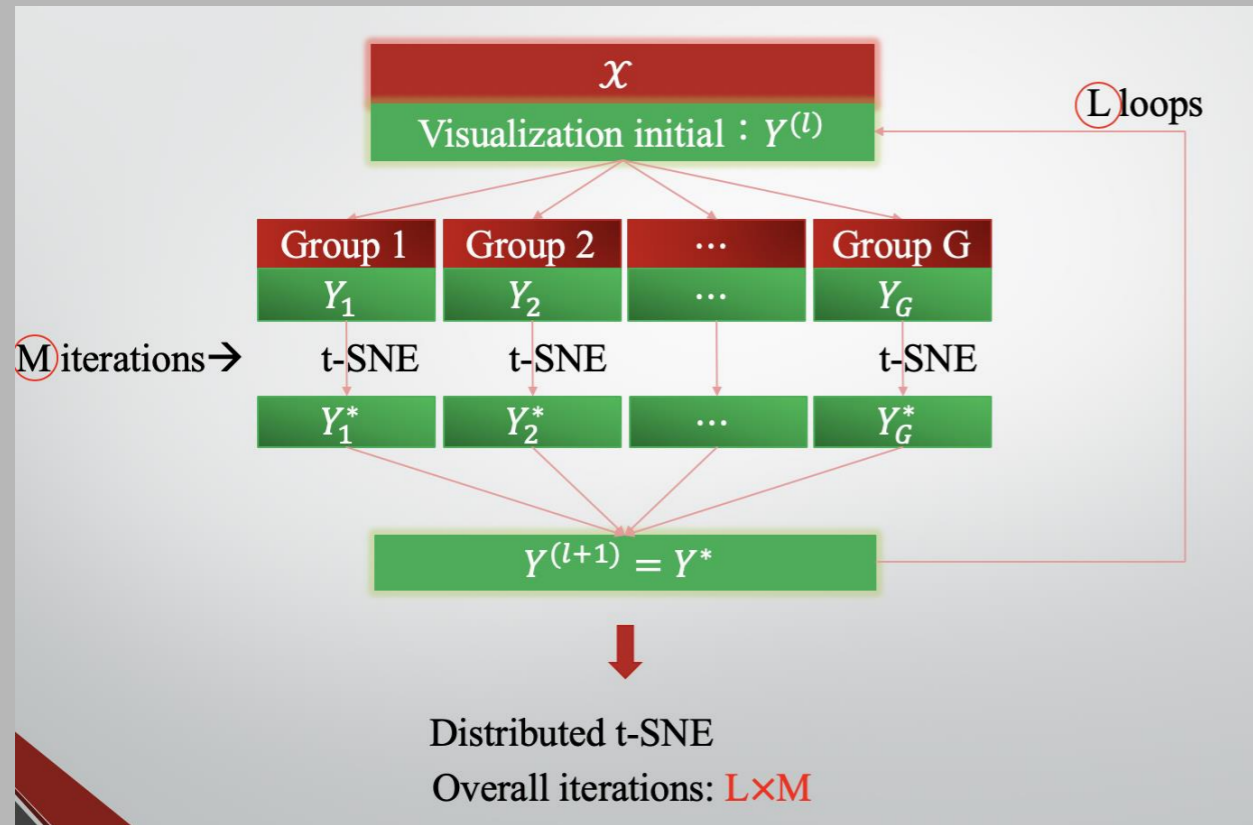


# ASCEP' approach is to develop

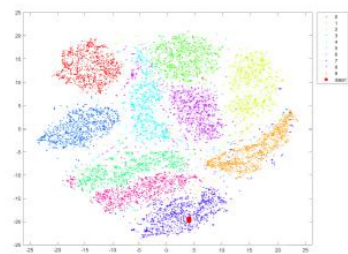
- Efficient Dimension Reduction Method to subdue the noise impact and save computation complexity.
- Robust Clustering Algorithm to reduce the iteration times yet output comparable results (if not better).
- Distributed Computation Algorithms.
- Interactive Visualization Tool.
- Mathematical Modeling for better interpretation.

# Distributed t-SNE

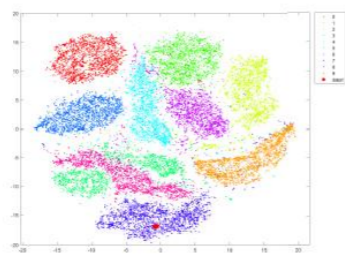
Szu-Han Lin, Ting-Li Chen\* and I-Ping Tu (in preparation)



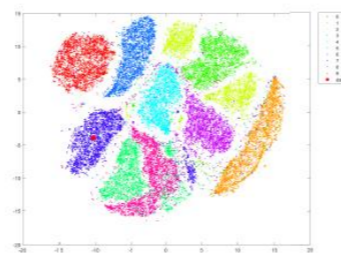
Sample Size	t-SNE (2008)		Barnes-Hut t-SNE (2014)	
	Time (s)	Peak memory (GB)	Time (s)	Peak memory (GB)
10,000	1625.607	0.781555	261.272	0.685524
30,000	14342.144	6.732208	918.285	1.955463
50,000	Out of memory	Out of memory	1656.379	3.227379
70,000			2526.174	4.493526
90,000			3335.760	5.783798
110,000			7945.461	7.090229
130,000			5144.081	8.418926
200,000			11405.644	13.084904
300,000			Out of memory	Out of memory



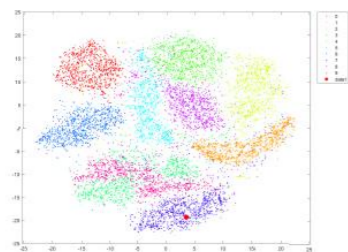
(a) t-SNE,  $n = 10,000$



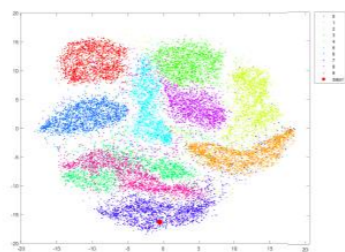
(b) t-SNE,  $n = 20,000$



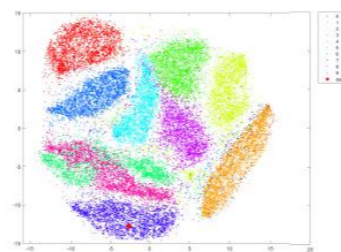
(c) t-SNE,  $n = 30,000$



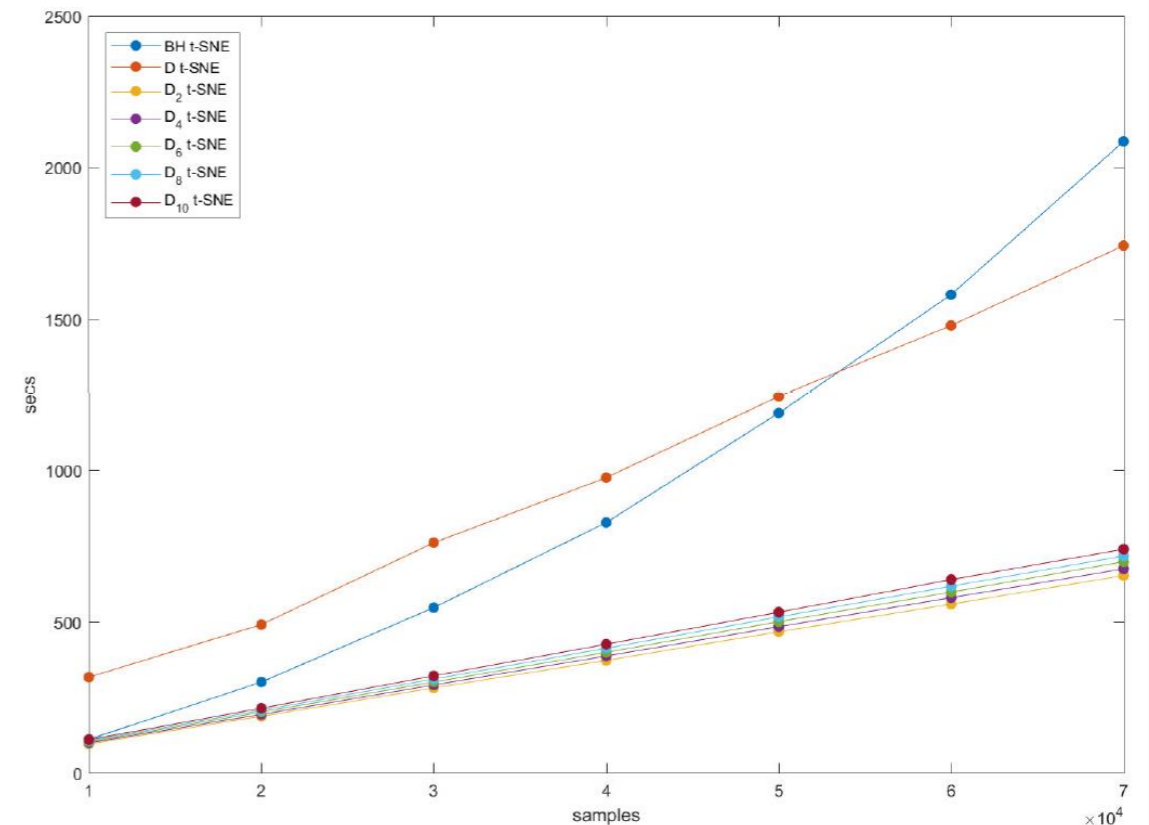
(d) Dt-SNE,  $n = 10,000$



(e) Dt-SNE,  $n = 20,000$



(f) Dt-SNE,  $n = 30,000$



Szu-Chi Chung, Shao-Hsuan Wang, Po-Yao Niu, Su-Yun Huang, Wei-Hau Chang and I-Ping Tu\* (2020). “Two-stage dimension reduction for noisy high-dimensional images and application to Cryogenic Electron Microscopy”. *Annals of Mathematical Sciences and Applications* 5, 283-316.

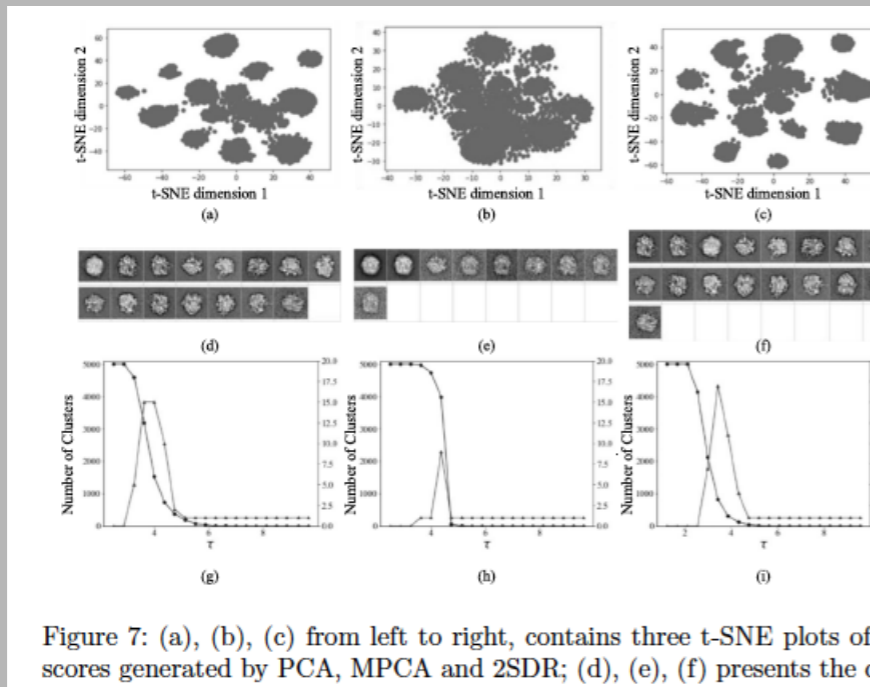
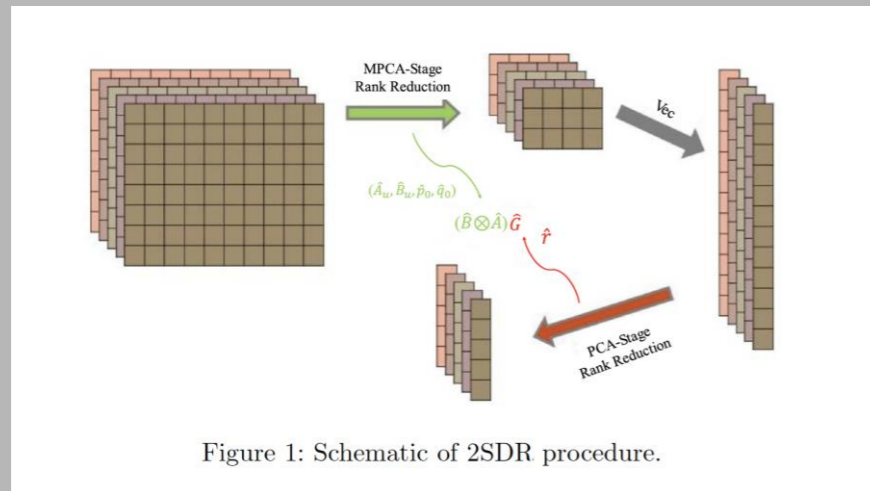


Figure 7: (a), (b), (c) from left to right, contains three t-SNE plots of scores generated by PCA, MPCA and 2SDR; (d), (e), (f) presents the cluster visualizations; (g), (h), (i) are line plots showing the number of clusters versus a parameter  $\tau$  for PCA, MPCA, and 2SDR.

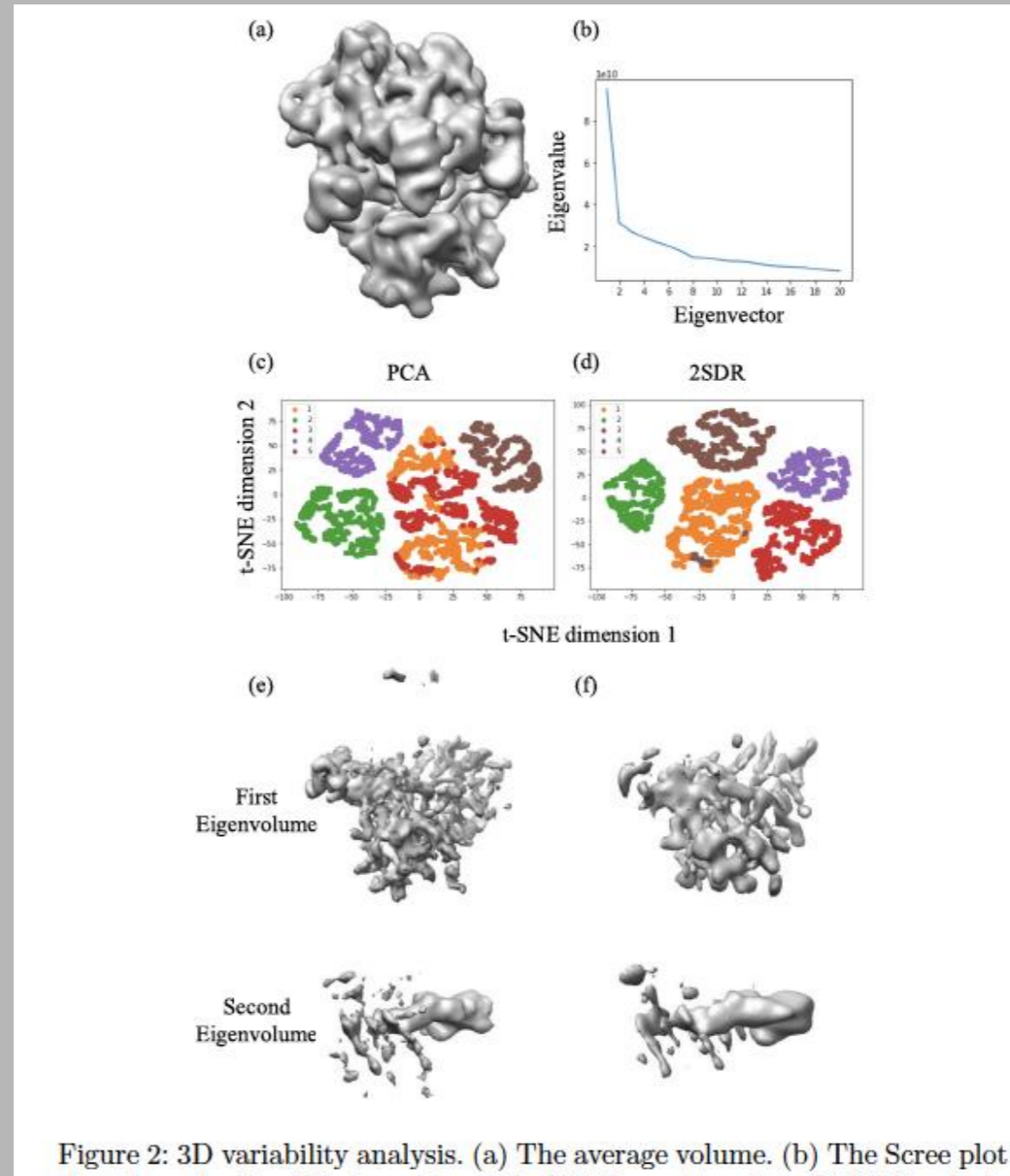
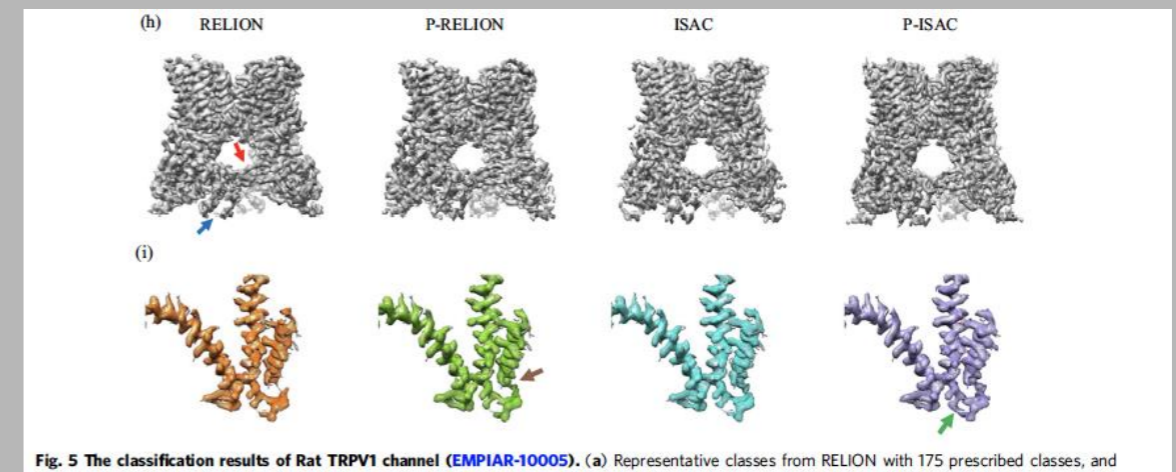
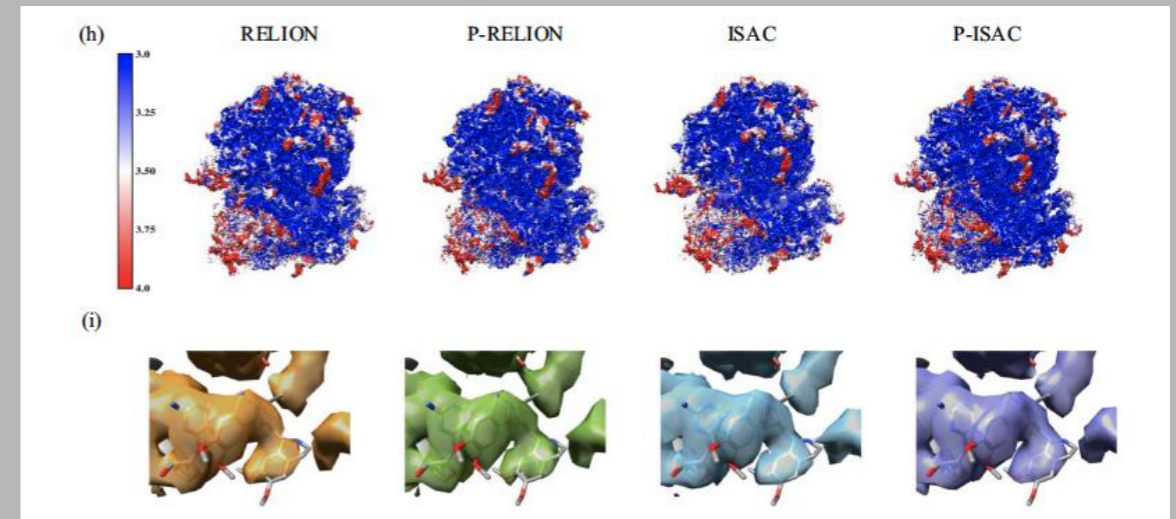
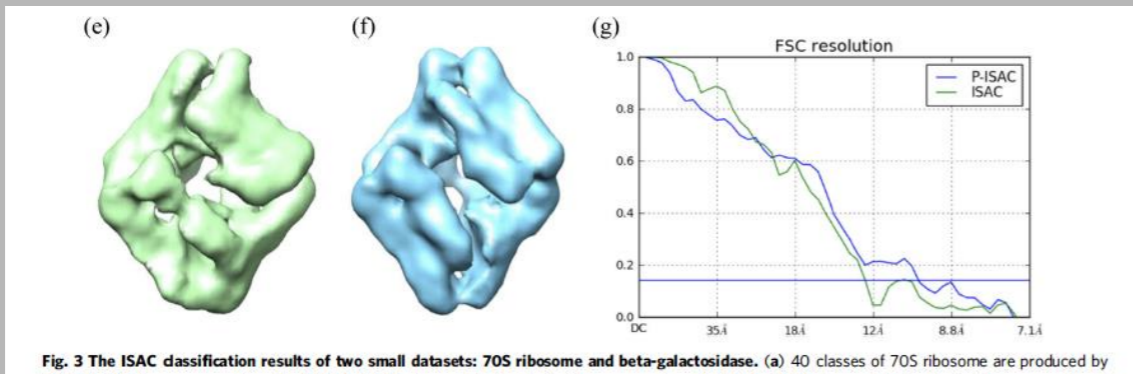
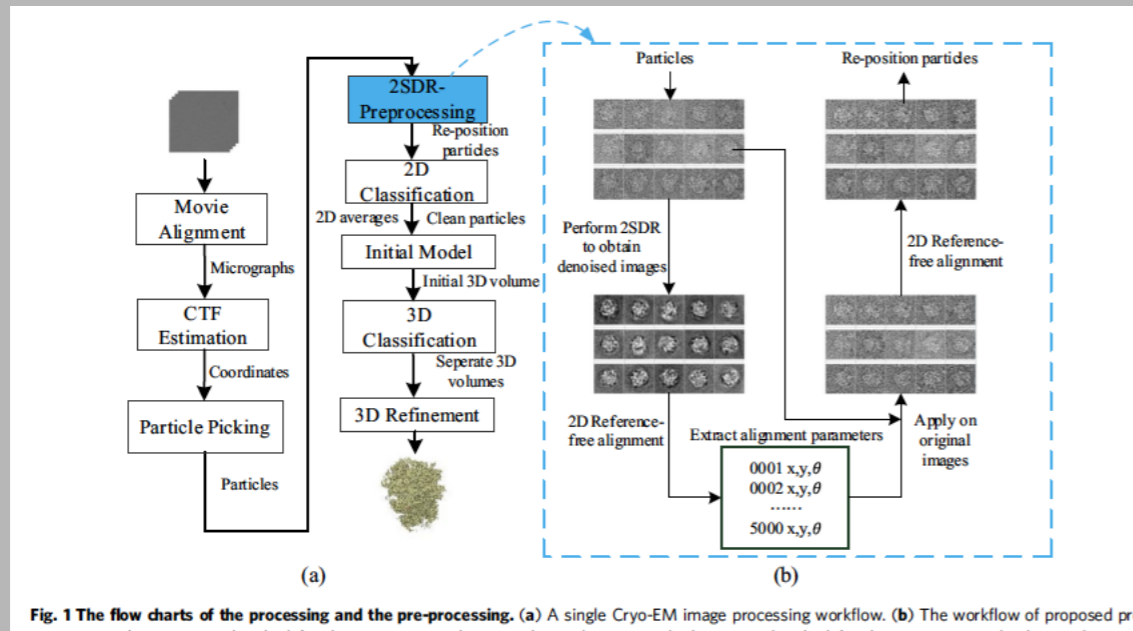


Figure 2: 3D variability analysis. (a) The average volume. (b) The Scree plot showing eigenvalues versus eigenvectors. (c) PCA t-SNE plot. (d) 2SDR t-SNE plot. (e) First eigenvolume. (f) Second eigenvolume.

Szu-Chi Chung, Hsin-Hung Lin, Po-Yao Niu, Shih-Hsin Huang, I-Ping Tu\* and Wei-Hau Chang\* (2020). “Pre-pro is a fast pre-processor for single-particle cryo-EM by enhancing 2D classification”. *Communications Biology* 3, 508.



# DRMRA: We have successfully found a hidden conformation

Szu-Chi Chung, Hsin-Hung Lin, Tien-You Liu, Kuen-Phon Wu, [Ting-Li Chen](#), Wei-Hau Chang\*, and [I-Ping Tu\\*](#), (2022), "Distributed Rapid Multi Reference Alignment (DRMRA) for accelerating single particle cryo-EM analysis", (in preparation).

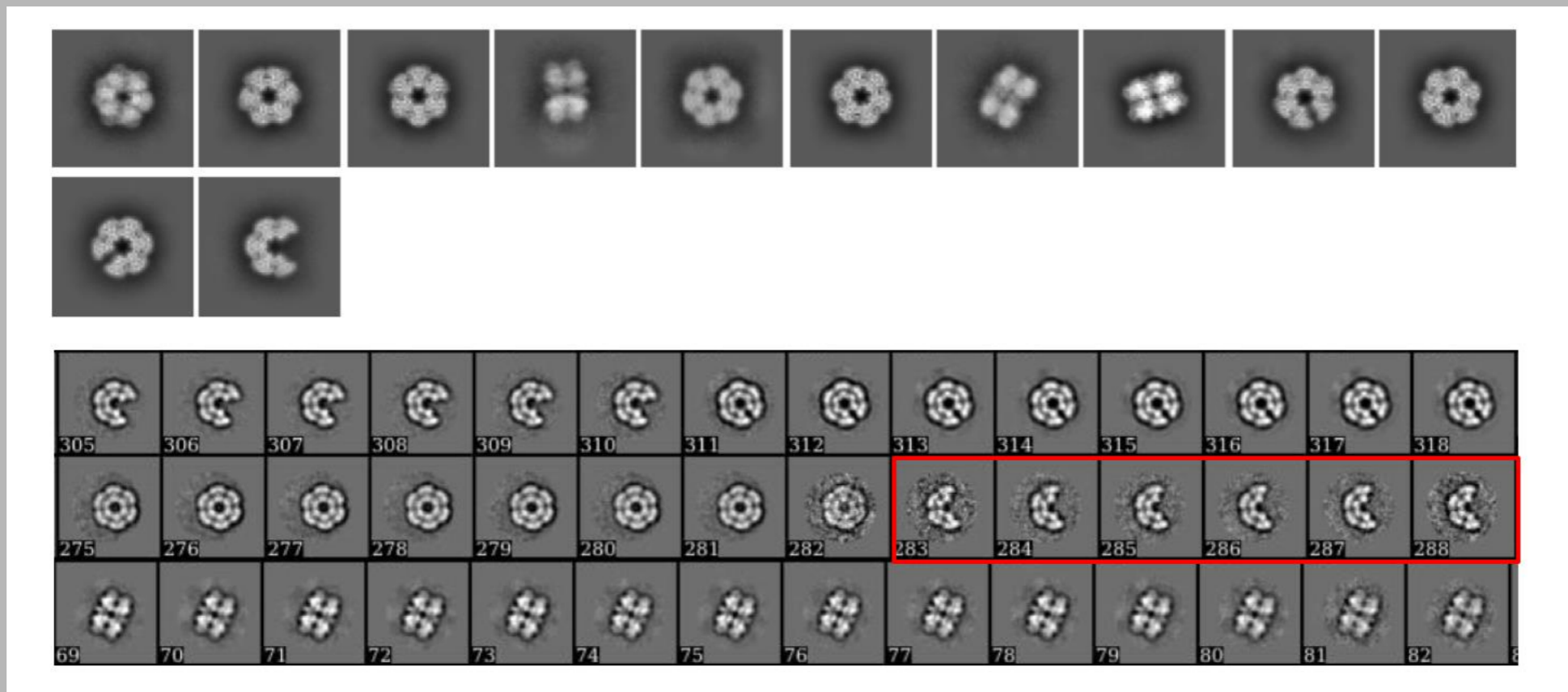


Table 1: The classification time of the CPU implementation on five datasets. Notice that we only use the CPU implementation from RELION.

Algorithm \ Dataset	RELION	ISAC	DRMRA
70S ribosome	0.43 (30 classes)	1.78 (200 members)	0.28
GS Protein	22.91 (150 classes)	50.92 (1000 members)	1.82
80S ribosome	62.32 (520 classes)	124.50 (4X binning, 200 members)	2.54 (4X binning)
NC-TRPV1	16.03 (200 classes)	55.61 (3X binning, 1000 members)	1.76 (3X binning)
NanoD-TRPV1	12.36 (100 classes)	*	3.86 (3X binning)

Table 2: The classification time of the GPU implementation on five datasets.

Algorithm \ Dataset	RELION	ISAC	DRMRA
70S ribosome	0.28 (30 classes)	0.33 (200 members)	0.18
GS Protein	8.68 (150 classes)	4.08 (1000 members)	1.58
80S ribosome	30.12(f) (520 classes)	14.55 (4X binning, 200 members)	2.02 (4X binning)
NC-TRPV1	13.56(f) (200 classes)	5.38 (3X binning, 1000 members)	1.51 (3X binning)
NanoD-TRPV1	9.49(f) (100 classes)	12.65 (3X binning, 1000 members)	3.33 (3X binning)

# 2.5 Å cryo-EM structure of particulate methane monooxygenase

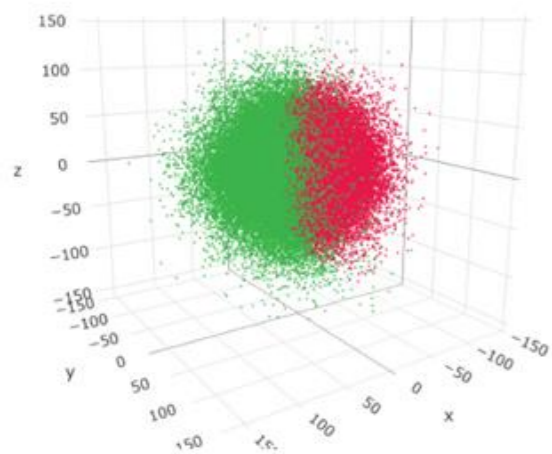
Wei-hau Chang\*, I-Kuen Tsai, Shih-Hsin Huang, Hsin-Hung Lin, Szu-Chi Chung, I-Ping Tu, Steve S.-F. Yu\* & Sunney I. Chan\* (2021) Cryo-EM structures of the functional particulate methane monooxygenase (pMMO) from *Methylococcus capsulatus* (Bath) reveals the sites of the copper centers. *Journal of American Chemical Society* (accepted).

## 3DVA - Clustering

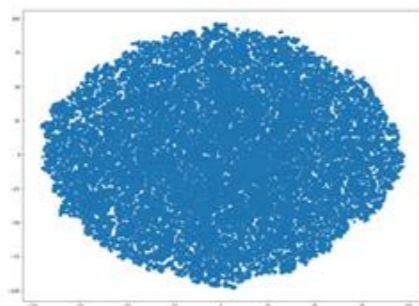
► Classify into 2 class  
(Using GMM)

- With C3 symmetry
- Non-uniform refinement

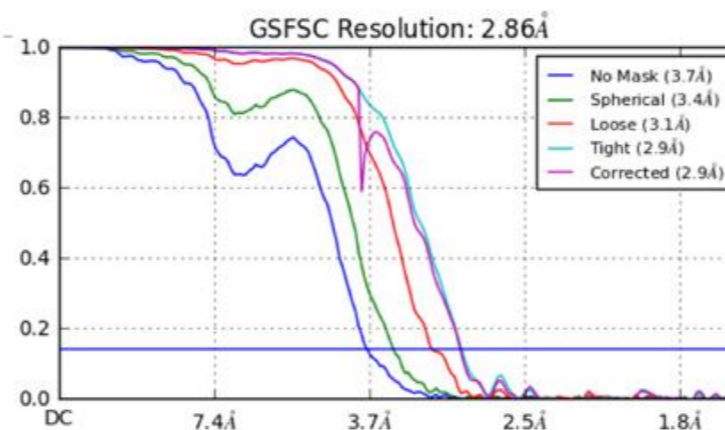
PC 1,2,3 (x,y,z)



t-SNE

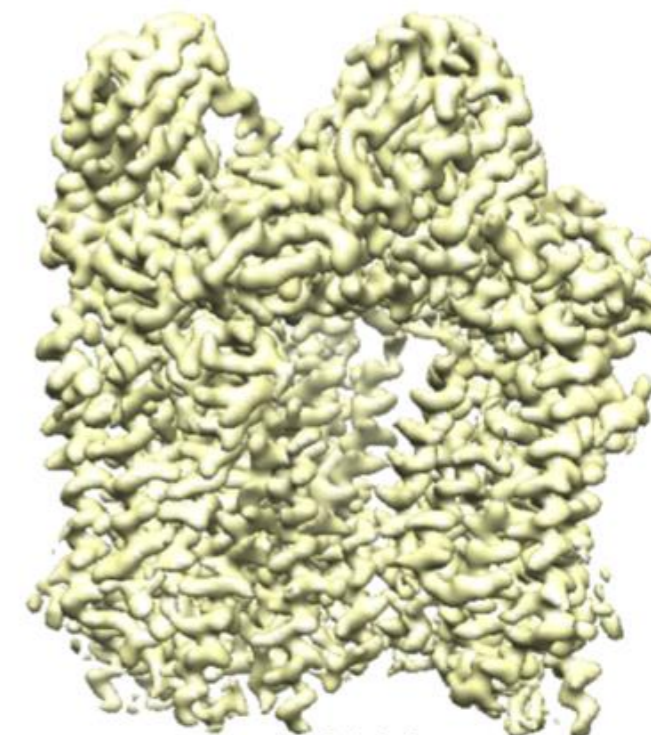
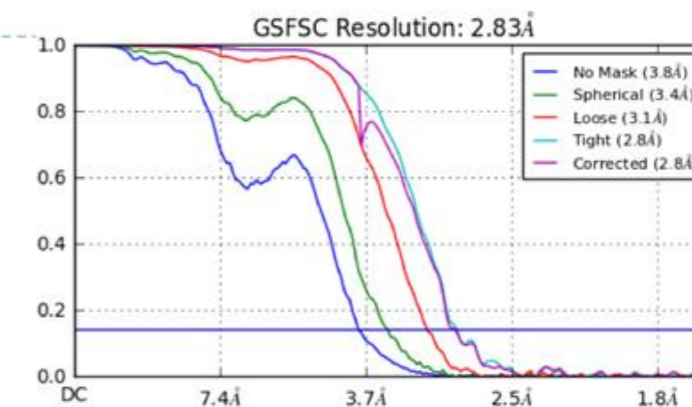


Conformation 1



243147

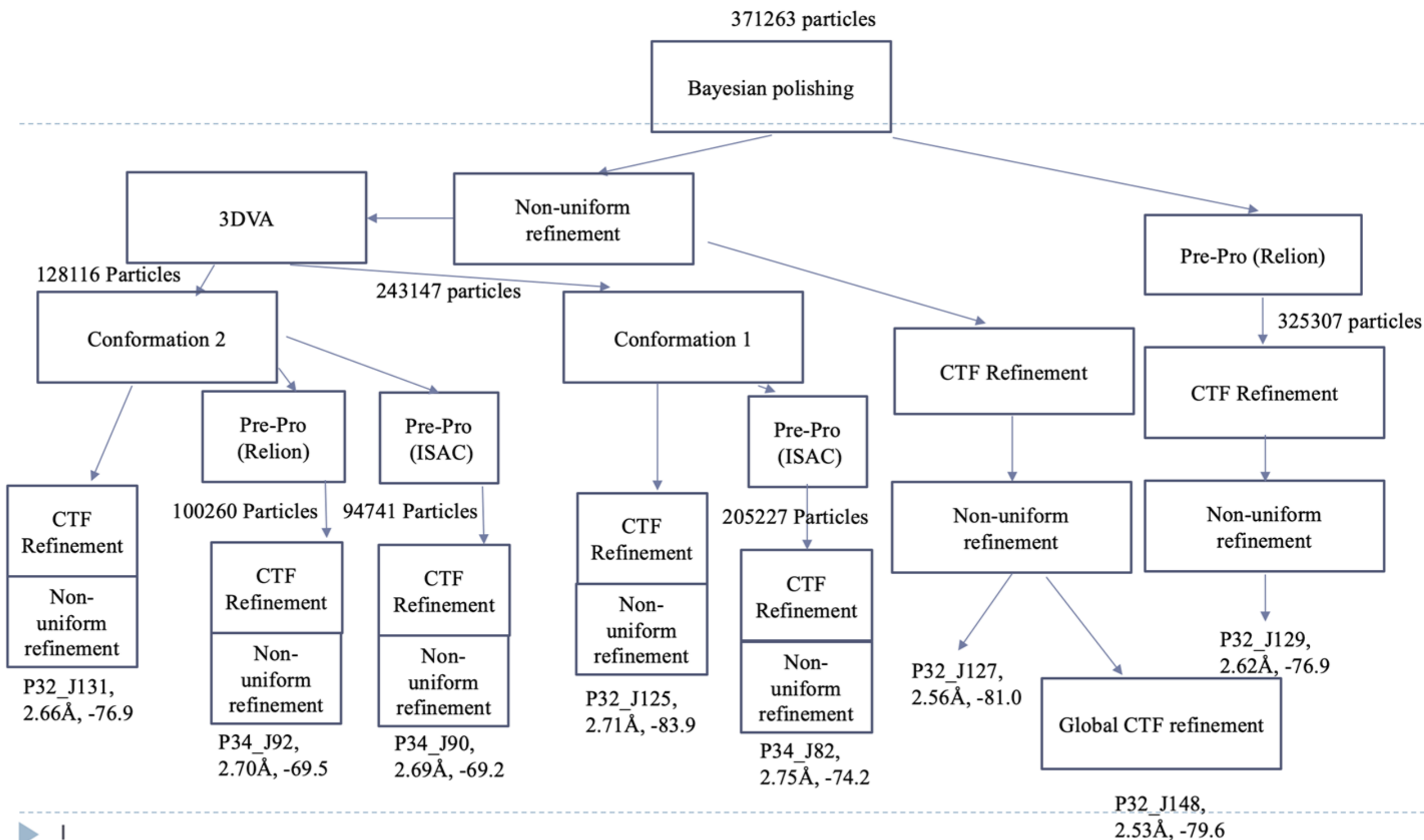
Conformation 2



128116



Finished within a week in one server  
 We have saved a lot of energy!



# Acknowledgements

## Collaborators and Grants

Institute of Statistical Science

Cheng-Yu Hung

Huei-Lun Siao

Szu-Han Lin

Yu-Hsiang Lien

Tien-Yu Liu

Po-Yao Niu

PI

Yi-Ching Yao

Ting-Li Chen

Su-Yun Huang

Institute of Chemistry

Hsin-Hung Lin

Wei-Hau Chang

Institute of Biological Chemistry

Kuen-Phon Wu

NSYSU, Applied Math

Szu-Chi Chung

NCU, Institute of Statistics

Shao-Hsuan Wang

- ▶ Investigator Award, Academia Sinica, 2021-2025
- ▶ Grand Challenge Seed Program, 2018, 2019 and 2020