

---

日本顕微鏡学会第77回学術講演会  
cryo-EMのための自動粒子Picking法と  
高精度TEMシミュレータによる評価  
Automated particle picking method for cryo-EM and the  
evaluation using high precision TEM simulator

大橋 正隆<sup>1</sup> 新川 隆郎<sup>1</sup> 岩崎 憲治<sup>2</sup>

<sup>1</sup> 株式会社バイオネット研究所 <sup>2</sup>筑波大学

**目的** Cryo EMのための自動粒子抽出ソフトウェアの開発  
TEMシミュレータを用いたソフトウェアの評価方法の提案

## 粒子抽出に関して

### 従来手法

- Template Matching法  
(Frankら, 1984; Huangら, 2004; Tangら, 2007; Langloisら, 2014; Punjaniら, 2017; Scheres, 2015)
- 深層学習  
(Wangら, 2016 ; Zhuら, 2017)

### 問題点

- 視認性の確保が必要
- 恣意性が含まれる可能性がある
- 学習データの量や品質に性能が依存

## 手法の評価に関して

### 従来手法

- ・ 公的なデータベースに登録されている実データや各自で撮影した実データを使用して手法を評価  
EMPIAR database (Iudinら, 2016)

### 問題点

- ・ 公的なデータベースから様々な実データを取得することは可能だが、取得したデータが同じ条件で取得されているとは限らないため、評価によって観測された結果の差の要因を特定することが困難

## 研究の概要

### 粒子抽出に関して

- ・ 提案手法では、抽出、2次元分類、位置補正の処理を繰り返し行うとともに、既に抽出された粒子のある領域を除外するマスク処理を導入

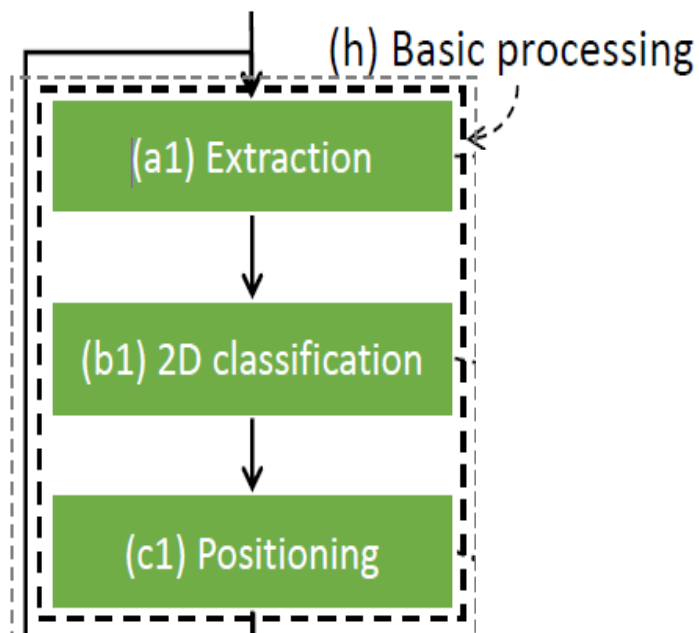
### 手法の評価に関して

- ・ 提案手法の有効性を評価するために、粒子が氷に包埋された状態のTEM画像をTEMシミュレータで生成して、生成されたデータを評価データとして用いて評価を行った

## Concept

デフォーカス値が小さい場合、粒子の視認性が低下する

粒子画像が2次元(2D)分類されるのであれば、その粒子画像には解析可能な特徴が含まれている  
2次元クラス平均画像は、個々の粒子画像と比較して、粒子が視認しやすい



基本処理(Basic Processing)

- ・ 抽出された粒子の座標が適切でないと、2次元クラス平均の粒子は欠けたりする可能性がある
- ・ 2次元クラス平均の画像中心からのずれと、各粒子のずれに基づいて、各粒子の座標を補正した
- ・ 粒子位置の事前情報はないため、最初は格子状に配置した

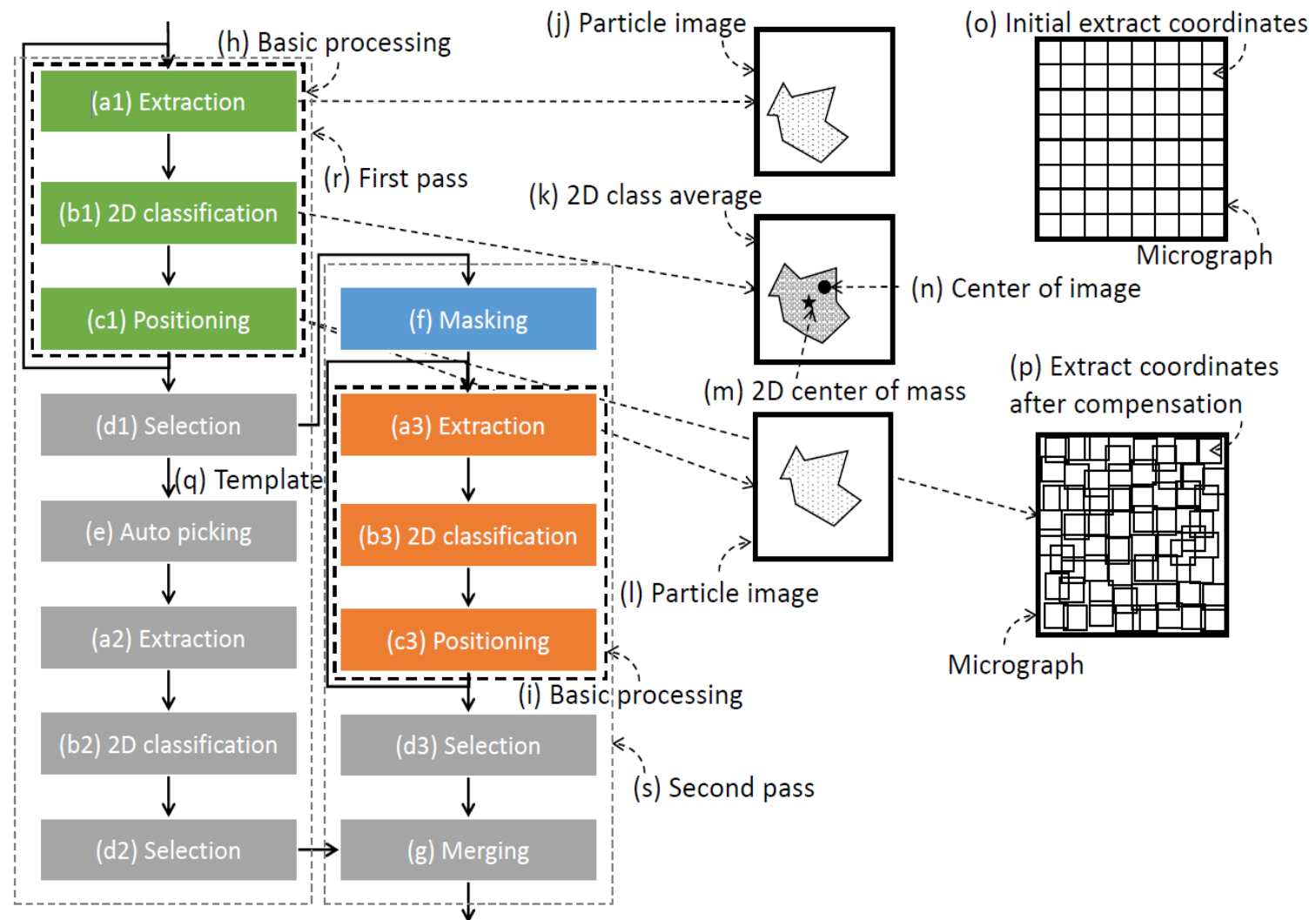
## Two-Pass Picking

基本処理(Basic Processing)は処理時間がかかるという問題がある

Template Matchingと組み合わせることで高速化を図る

粒子の向きに偏りがある場合、Templateに含まれない可能性がある

First Passで抽出した粒子位置をマスクし、それ以外の領域に対して基本処理を実施する



## 1. TEMシミュレータにより生成したシミュレーションデータを使用した評価

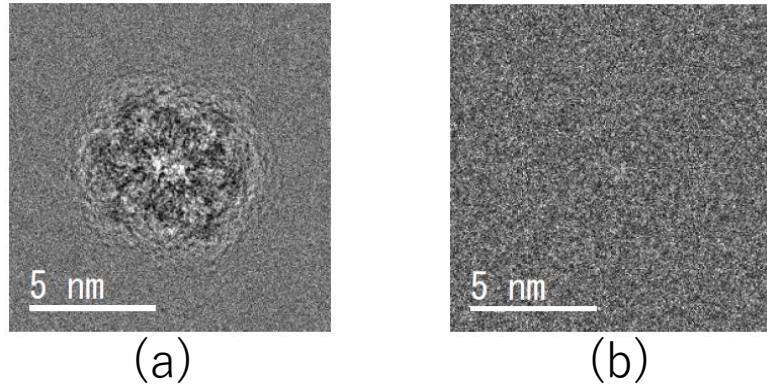
(a) 様々なDefocusによる評価

(b) 粒子の向きの分布に偏りがある場合

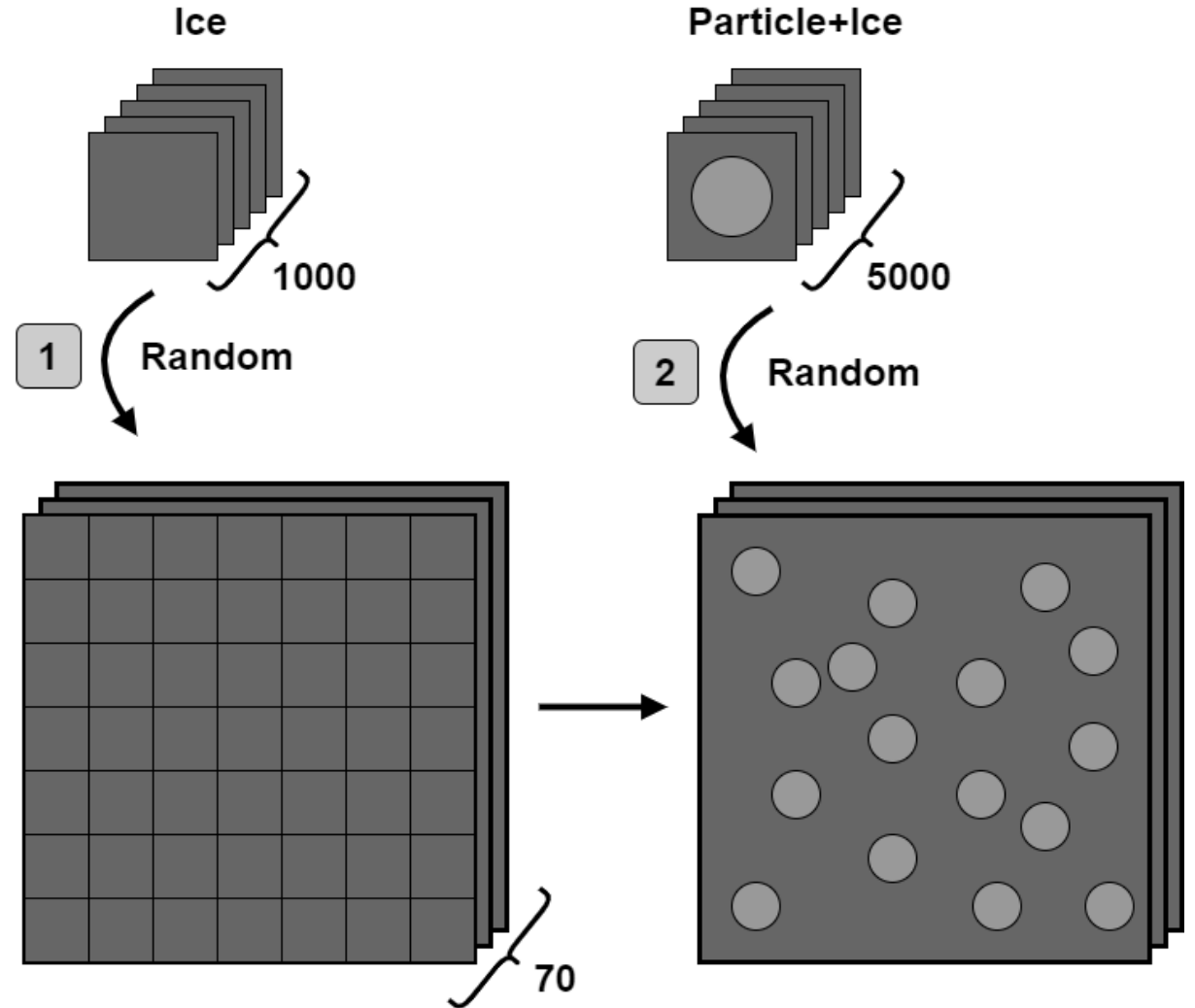
## 2. 実データを使用した評価

## Simulation Data 作成

TEM Simulator: elbis (BioNet Lab. Inc.)



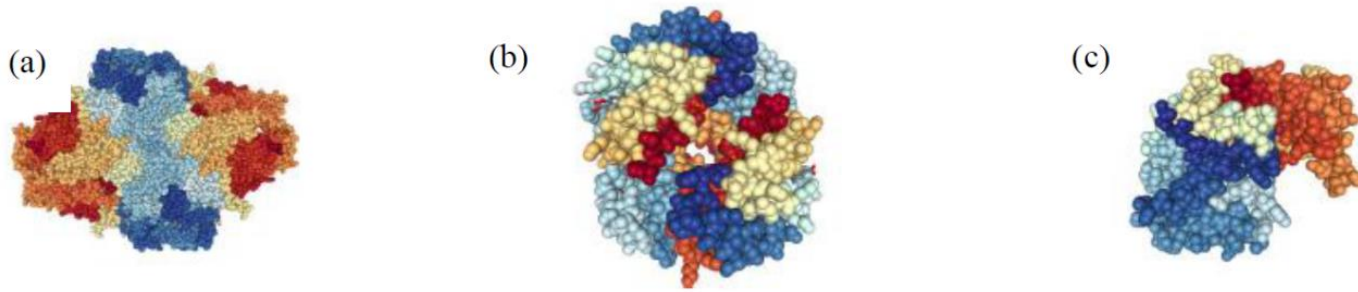
Simulated projection images of deoxyhemoglobin  
(a) TEM image without amorphous ice.  
(b) TEM image at a total dose of  $40 \text{ e}/\text{\AA}^2$ . The defocus value of these images was  $2500 \text{ \AA}$ . TEM images were obtained under the following conditions: accelerating voltage, 200 kV; spherical aberration, 1.0 mm; chromatic aberration, 1.1 mm; and electron energy spread, 0.3 eV.



**Table 1** Simulated data conditions

Accelerated voltage	200 kV	Spherical aberration	1.0mm
Chromatic aberration	1.1mm	Electron energy spread	0.3 eV
Total Dose (Exposure)	40 e/ Å <sup>2</sup>	Defocus	1000/2500/5000/15000 Å

## Sample



(a)  $\beta$ -galactosidase (PDB ID: 1px3). (b) Deoxyhemoglobin (PDB ID: 2hhb). (c) Pepsin (PDB ID: 5pep).  
The 3D view images were obtained from wwPDB (<https://www.wwpdb.org/>) (Berman *et al.*, 2000).

- CTF Estimation にはCTFFIND4を使用した
- 2次元分類にはRELIONを使用した



# 評価結果

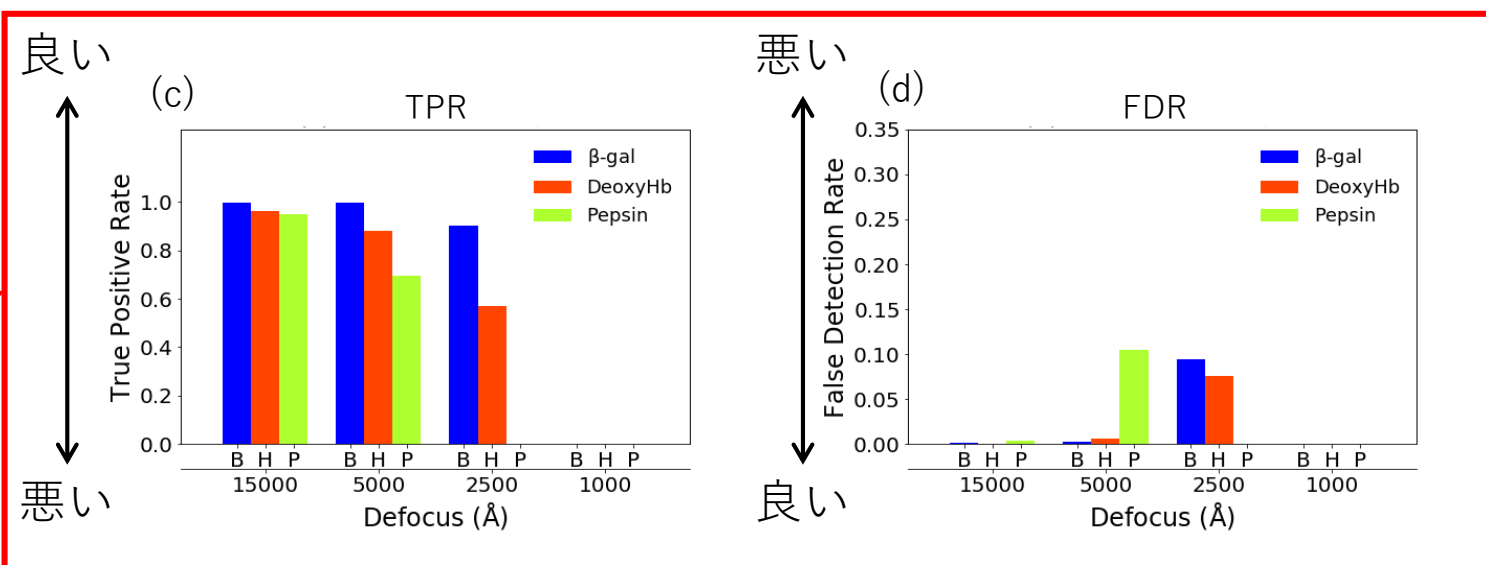
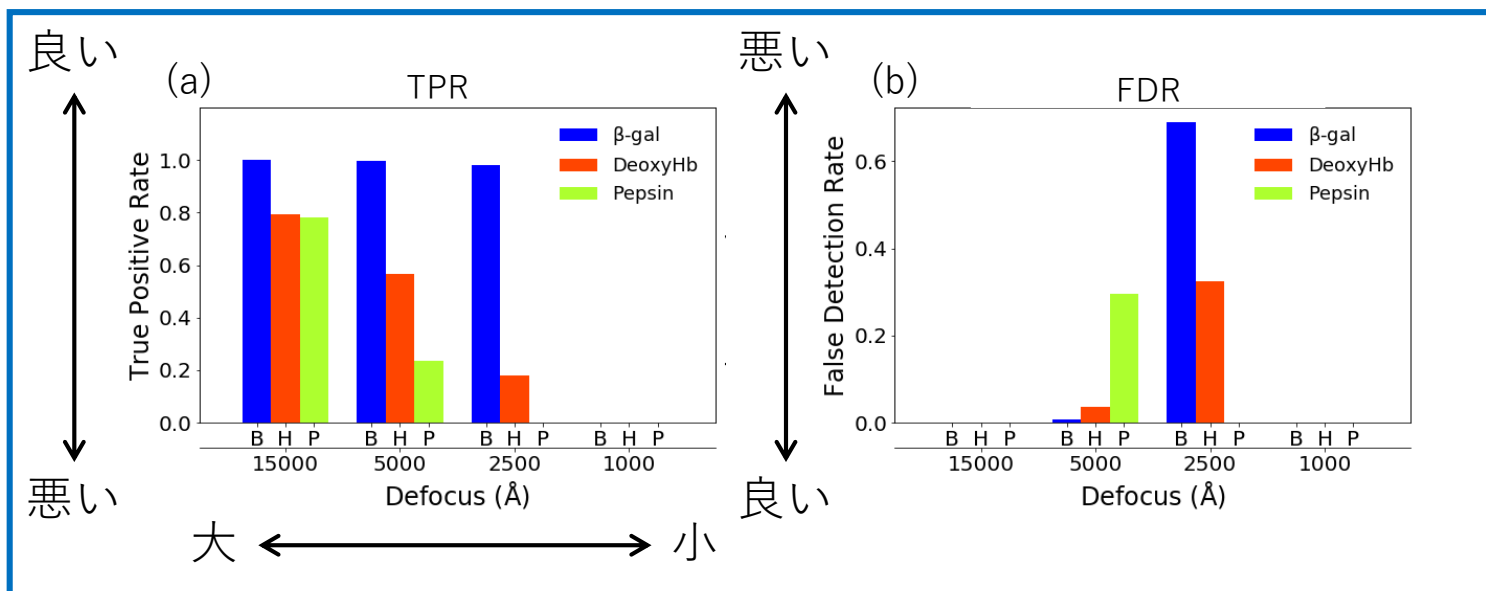
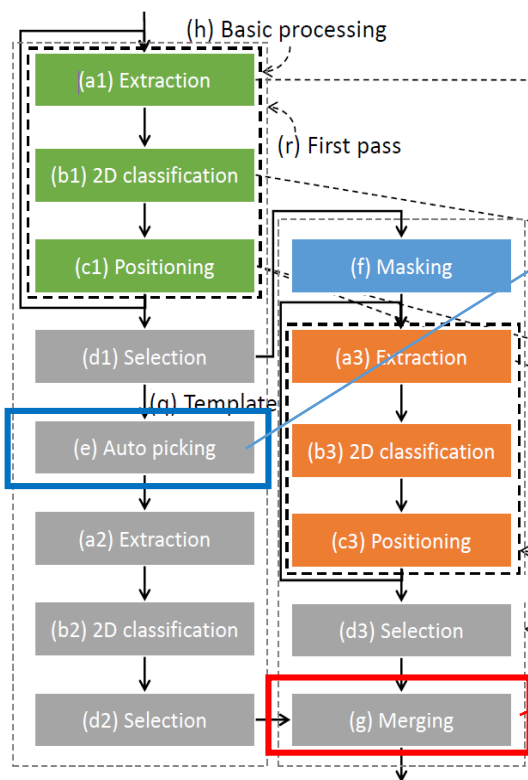
## 評価指標

$$TPR = \frac{TP}{TP + FN}$$

正しく粒子が拾えたか

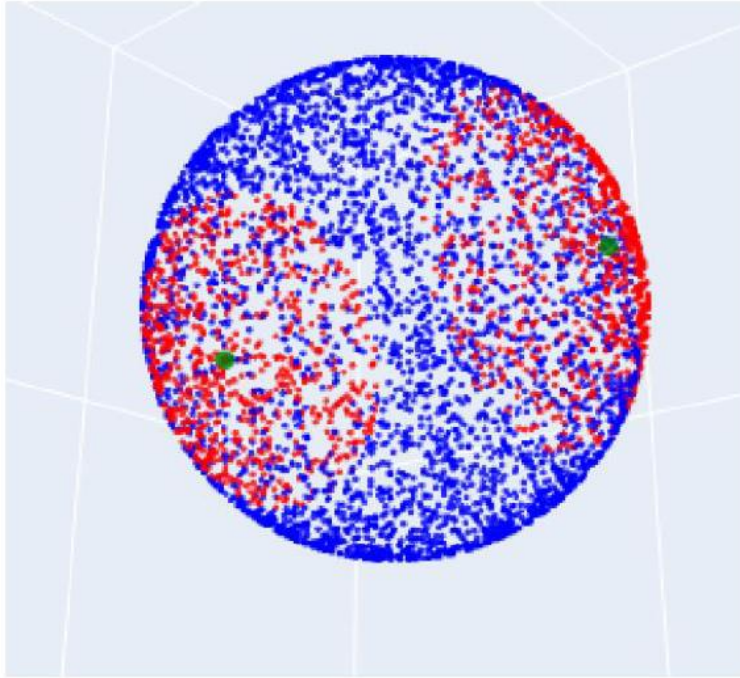
$$FDR = \frac{FP}{FP + TP}$$

誤って粒子以外を拾ったか

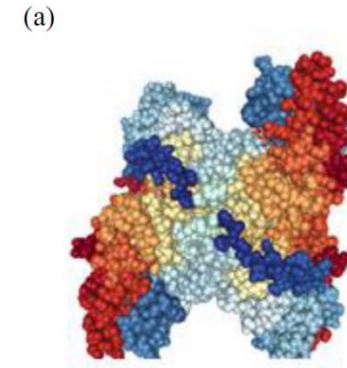


## データセット作成方法

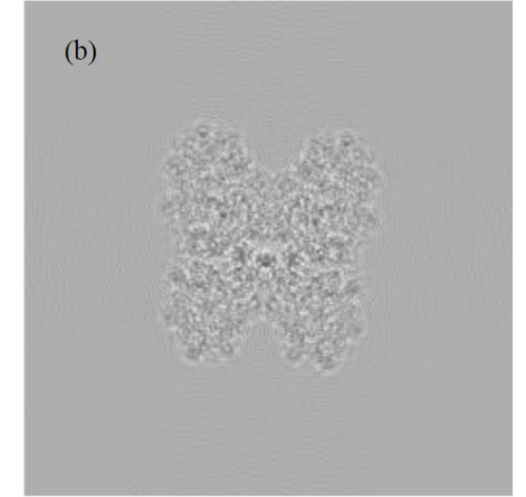
## Sample



投影方向



aldolase (PDB ID: 6ald)



特定方向

特定方向グループとそれ以外のグループに分類

特定方向のグループの発生頻度を少なくする

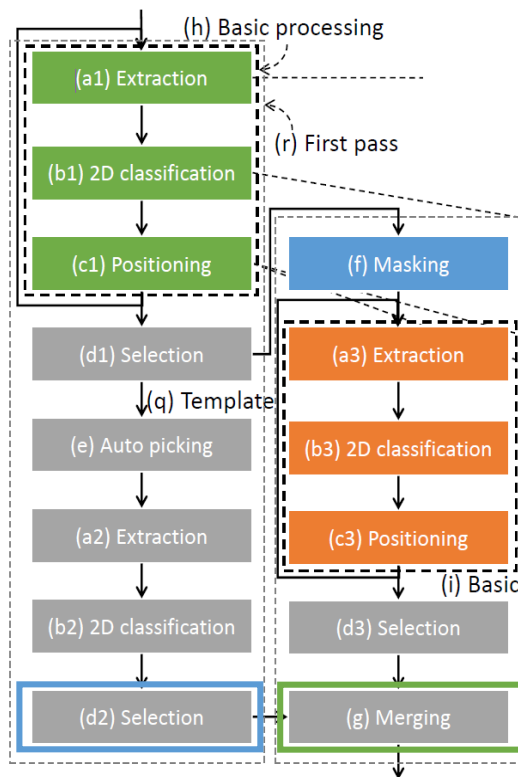
1%, 3%, 5%, 10%

Table 2 TPRs and FDRs

発生頻度	10%	5%	3%	1%
TPR(%)	98.43	98.95	99.30	98.88
FDR(%)	0.29	0.16	0.16	0.13

Table 3 特定方向グループのTPR

発生頻度	10%	5%	3%	1%
First Pass(%)	90.79	85.90	90.04	54.55
Final Result(%)	93.16	91.12	93.94	70.13



Prob.	Template	Result of Picking
10%		
5%		
3%		
1%		

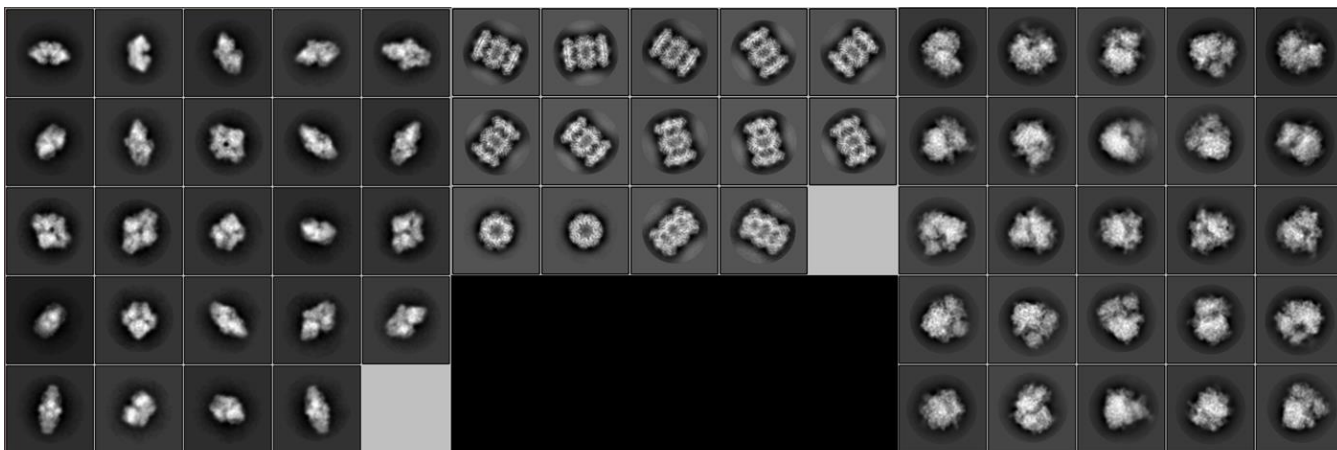
## Sample

Sample	(a) $\beta$ -galactosidase	(b) T20S proteasome	(c) 80S ribosome
EMPIAR	EMPIAR-10017	EMPIAR-10025	EMPIAR-10028
Detector	FALCON II detector	GATAN K2 Summit	FALCON II detector
Micrograph size	4096 × 4096	7420 × 7676	4096 × 4096
# of micrographs	84	196	1081
Pixel size	1.77 Å	0.66 Å	1.34 Å
Box size	180 × 180	450 × 450	360 × 360
Symmetry	D2	D7	C1

- (a)  $\beta$ -galactosidase EMPIAR-10017 (Scheres, 2015)
- (b) T20S proteasome EMPIAR-10025 (Campbell *et al.*, 2015)
- (c) 80S ribosome EMPIAR-10028 (Wong *et al.*, 2014)

# 評価結果

## 2次元分類結果

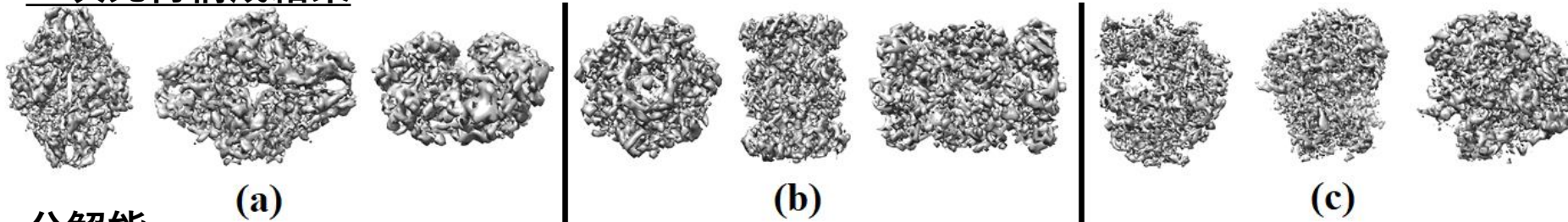


(a)

(b)

(c)

## 3次元再構成結果



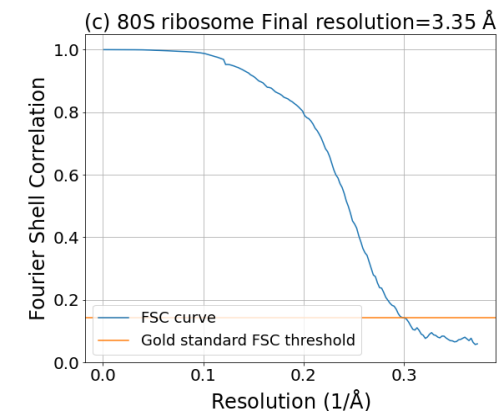
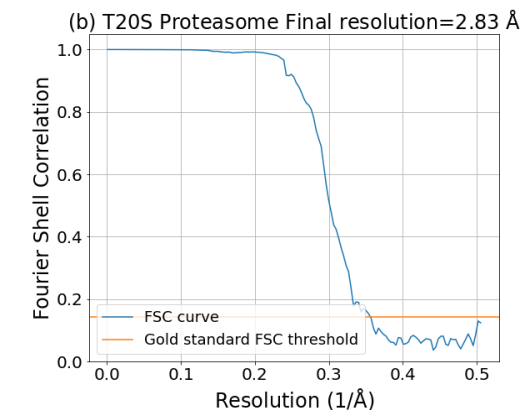
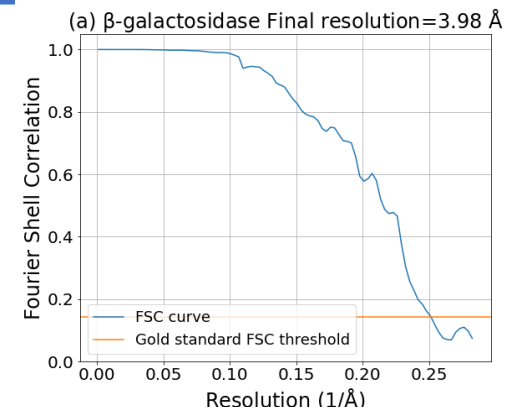
(a)

(b)

(c)

## 分解能

Sample	(a) $\beta$ -galactosidase	(b) T20S proteasome	(c) 80S ribosome
分解能	3.98 Å	2.83 Å	3.35 Å
オリジナル論文で報告された分解能	4.2 Å	2.8 Å	3.2 Å



## 粒子抽出ソフトウェア

- ・シンプルな構成
- ・提案手法で粒子抽出ができなければ、撮影された電顕画像は二次元分類には適していないと判断でき、解析時間の無駄を回避できる

## TEMシミュレータを使用した評価

### (a)様々なデフォーカスの評価

- ・提案した手法は約90%のTPRを達成した
- ・デフォーカス値が小さくなるにつれてFDRが増加した

### (b)偏った粒子の向きの分布のデータセットによる評価

- ・発生頻度の低い投影方向から投影された粒子を表す特定方向グループにおいて、最終結果のTPRがファーストパス結果のTPRに比べて向上することを示した

## 実データによる評価

- ・提案手法によって抽出された粒子から3次元再構成を行った結果によって得られた分解能から、実際のTEM画像にも適用できることを示した