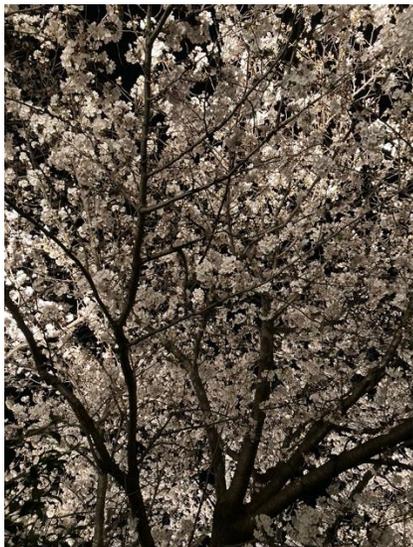


実験データ融合生体分子構造多形解析ワークフロー開発 ～ポスト「富岳」に向けた取り組みと展望～



理化学研究所 計算科学研究センター(R-CCS)
HPC/AI駆動型医薬プラットフォーム部門
バイオメディカル計算知能ユニット
上級研究員

徳久 淳師



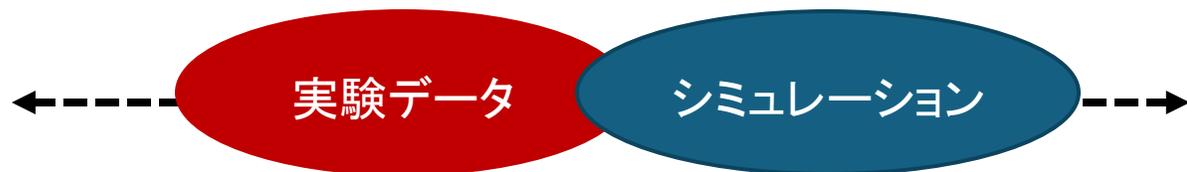
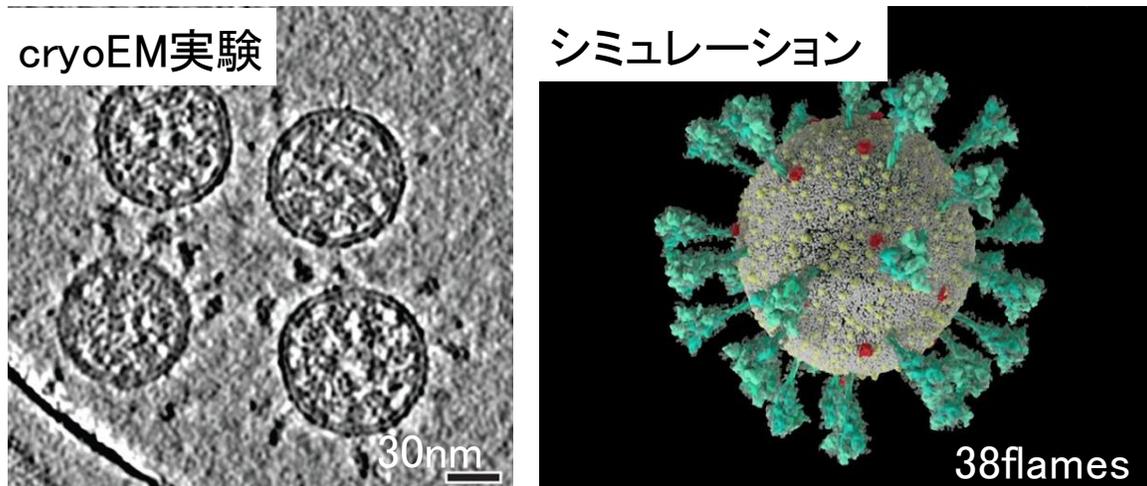
内容

- 研究背景
- これまでの取り組み(-FY2024)
 - CryoMDM構造多形解析ワークフロー開発
 - スパイクタンパク質への適用
- FY2030に目指すところ
- まとめ

背景

実験とシミュレーション(SARS-CoV-2を例に)

<https://www.nature.com/articles/d41586-021-02039-y>



時空間スケールにギャップが存在

実験 × AI × シミュレーションの融合

目的: 生体分子の多様な構造状態に原子分解能でせまる

生体分子の特徴:

- ✓ 細胞内・外の大規模な相互作用により生命活動は維持
- ✓ 「やわらか」く様々な機能を発揮する
- ✓ 分子メカニズムの理解には多様な構造状態を解明することが重要

現状の問題点:

- ✓ 残念ながら、原子分解能で時系列で実験的に生体分子を観測する万能手法はない。
- ✓ また、実験とシミュレーションには観測できる時空間スケールにギャップがある。
- ✓ そこで、HPCやAIを駆使したシミュレーションと、実験データを組み合わせ、多様な構造状態の解明を目指している。

産業における活用例:

- ✓ 創薬 (低分子から中分子、高分子へのモダリティシフト)
- ✓ 機能性食品の開発
- ✓ 触媒などの材料系

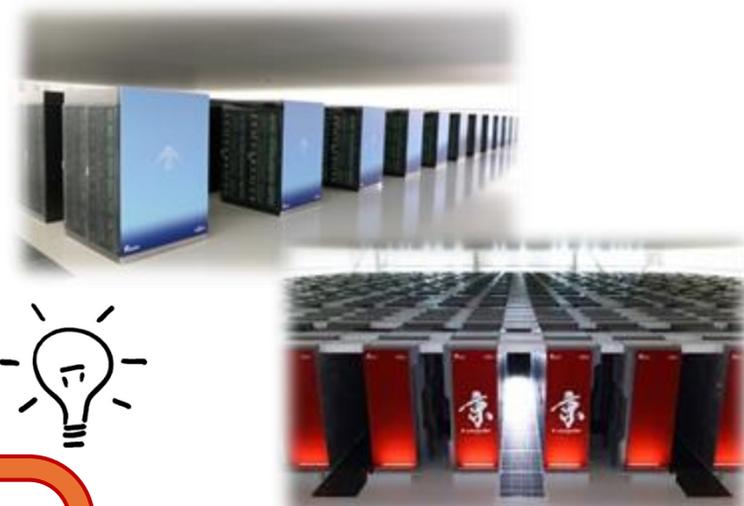
概要

SACLA @SPring-8

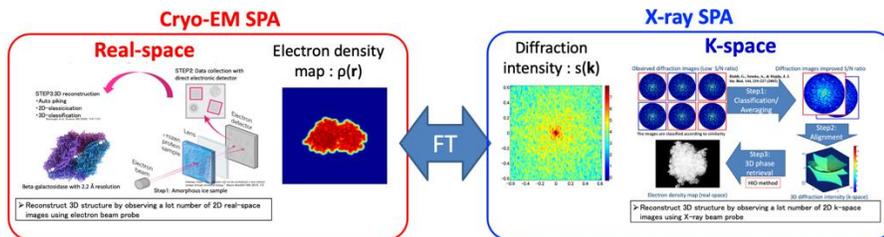


コラボレーション

Fugaku @R-CCS



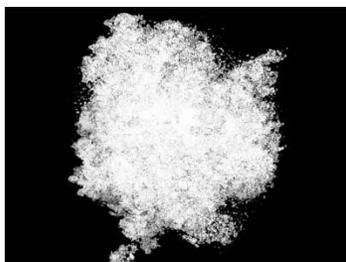
生体分子の構造多形解析の実現を目指し
単粒子構造解析(SPA)に注目



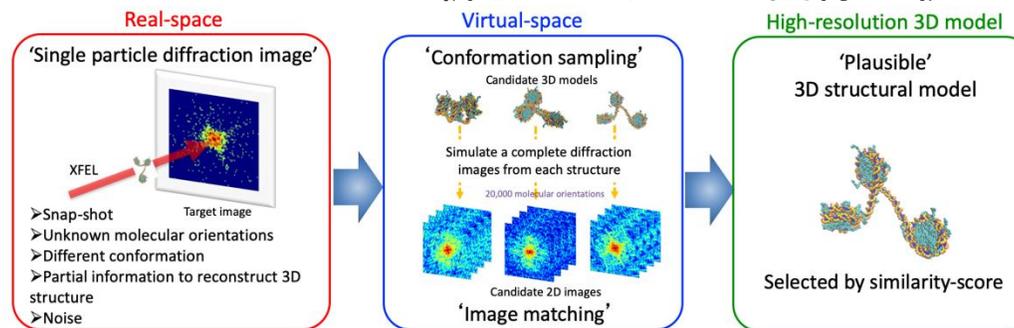
通常、数万枚程度の画像を測定する

K-computer @AICS

SPAの従来型スキーム:
多数の像から主要な構造(密度マップ)を構築



SPAの新しいスキーム: 画像一枚の可能性を追求



モデリング精度

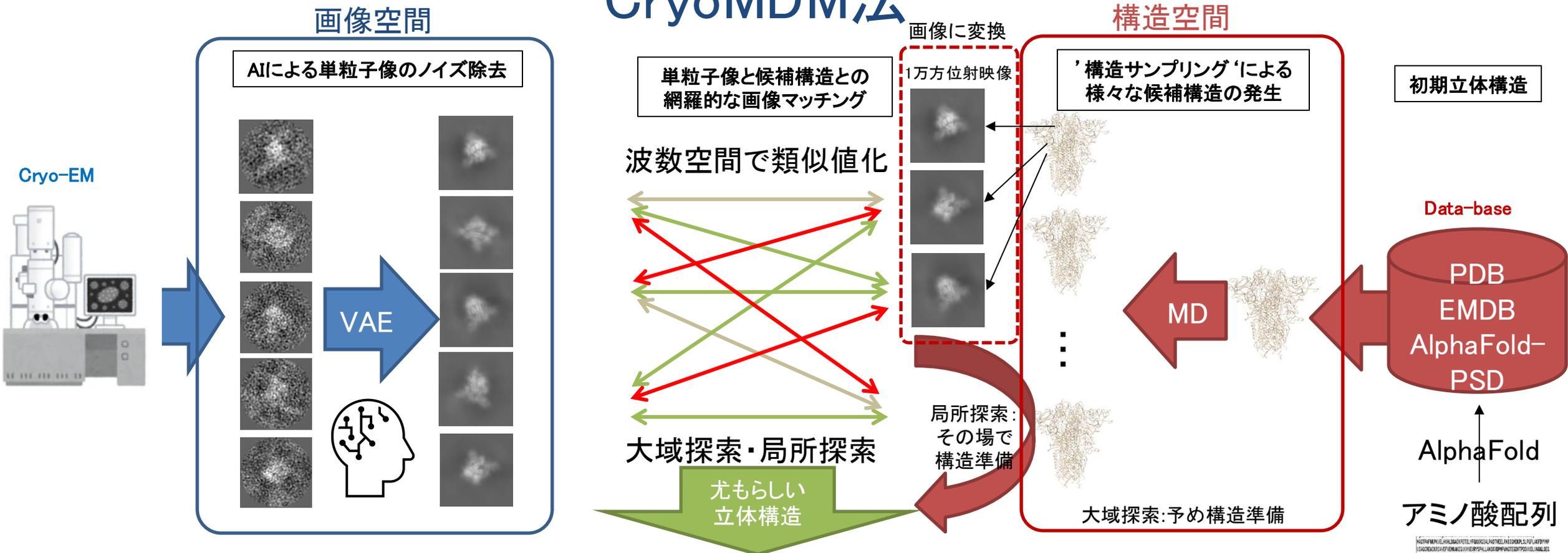
計算コスト

AI技術により調和

到達点

➤ データ同化より、単一の実験像から直接原子分解能の立体構造を推定することで構造多形を評価可能

CryoMDM法



利点: 個々の単粒子像に対する多様な立体構造を網羅的に高解像度で推定可能

アイデア: 仮想空間において実験データの不完備性を補う

- ① '構造サンプリング'により、様々な候補構造を発生（予め & その場で）。
- ② 1候補構造あたり1万分子方位で'画像マッチング'を実施。
- ③ 標的像と候補像の間の類似値により'尤もらしい'立体構造を選択。

内容

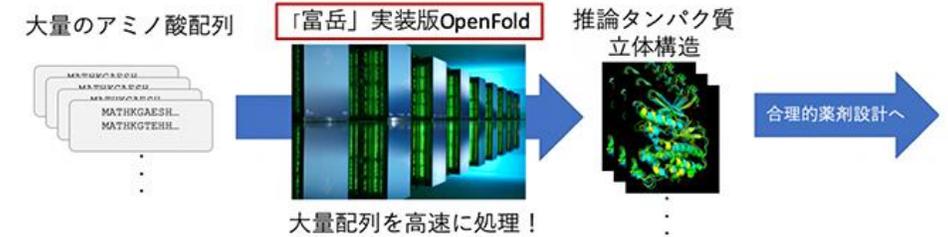
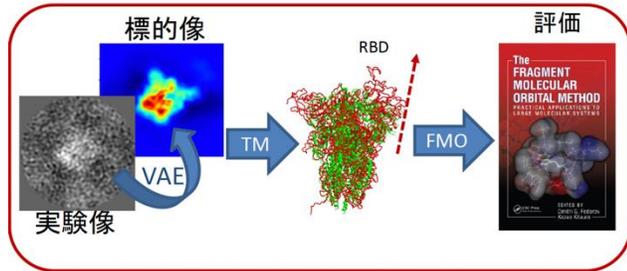
- 研究背景
- これまでの取り組み(-FY2024)
 - CryoMDM構造多形解析ワークフロー開発
 - スパイクタンパク質への適用
- FY2030に目指すところ
- まとめ

運用・開発中のワークフロー(WF)

複数手順を組み合わせ、かつ複数の可能性を広範に探索する為、FY2018からWFを構築し始めた

FY2019: 京大スパコンにてTM-PaCS-MD-WFを構築・運用

FY2020: 「富岳」試行的利用課題にて「富岳」にWF実装開始



Project1: 実験データと原子モデルの融合

CryoMDM構造探索-WF → FY2024-: WF統合化

- ・AIによる実験画像改善(プリポスト)
- ・初期立体構造生成
 - OpenFold-for-Fugaku(FY2024: 計算ノード)
- ・構造アンサンブル準備
 - OFLOOD-WF(FY2021: 計算ノード)
 - BS-PaCS-WF(FY2023: 計算ノード)
- ・大域探索
 - CryoEM-Matching-WF(FY2021: 計算ノード)
- ・局所探索
 - TM-PaCS-MD-WF(FY2021: 計算ノード)
- ・構造評価
 - FMO-WF(FY2022: 計算ノード、Prefectを利用)
- ・可視化

Project2: ゲノム情報と原子モデルの融合

Genome2MD-WF → FY2024-: 開発中

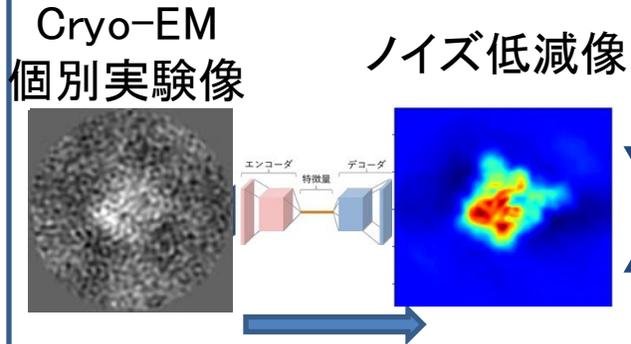
- ・配列解析
- ・OpenFold-for-Fugaku(計算ノード)
- ・MD-analysis(計算ノード)
- ・静的構造解析
- ・ダイナミクス解析
- ・可視化

以降、PJ1を中心に紹介します

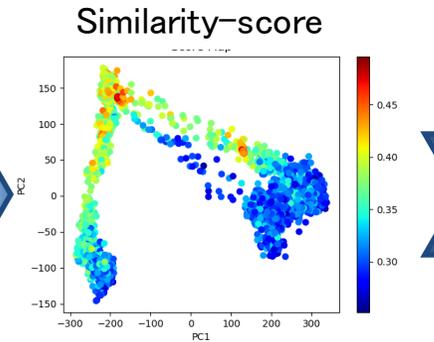
「富岳」における生体分子構造多型解析ワークフロー

2段階の構造探索で効率化

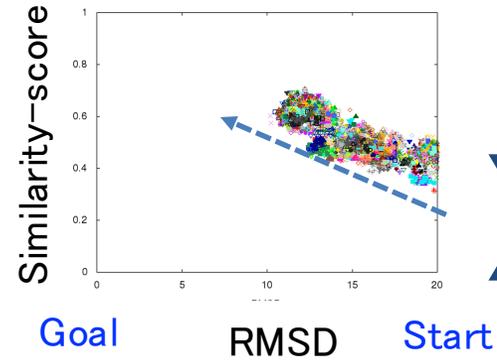
個別実験像改善



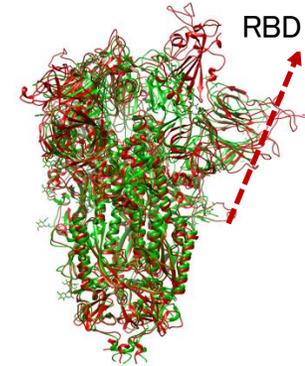
大域構造探索



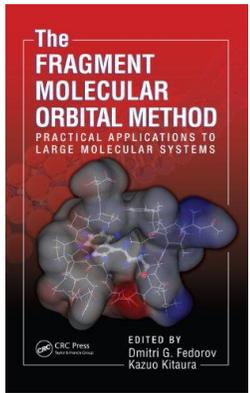
局所構造探索



尤もらしい3D構造



構造評価



- 原子分解能で構造を推定するための一連の構造多型解析ワークフローを富岳上に実装
 - 手順0: 実験データ事前解析 (Pre-postノードへ実装)
 - 手順1: VAE(Variational Autoencoder)により標的画像の画質を改善 (Pre-postノードへ実装).
 - 手順2: CryoEM-Matching-WFにより大域構造探索を実施 (計算ノードへ実装).
 - 手順3: TM-PaCS-MDにより局所構造探索を実施 (計算ノードへ実装).
 - 手順4: 各々の個別像から尤もらしい立体構造を推定
 - 手順5: FMO-WFにより推定構造を評価 (FMODD連携)

CryoEM-Matching-WF

WFの役割: 実験標的画像と候補構造(異なる分子方位の画像を準備し方位探索)の類似性を判定する

タスク: 一対の画像の類似性をスコア化



波数空間においてFFTにより相関値を計算
 →現状CPUで実施
 →GPU対応で高速化が期待できる

1標的画像あたり1億タスク: 1万候補構造x1万分子方位



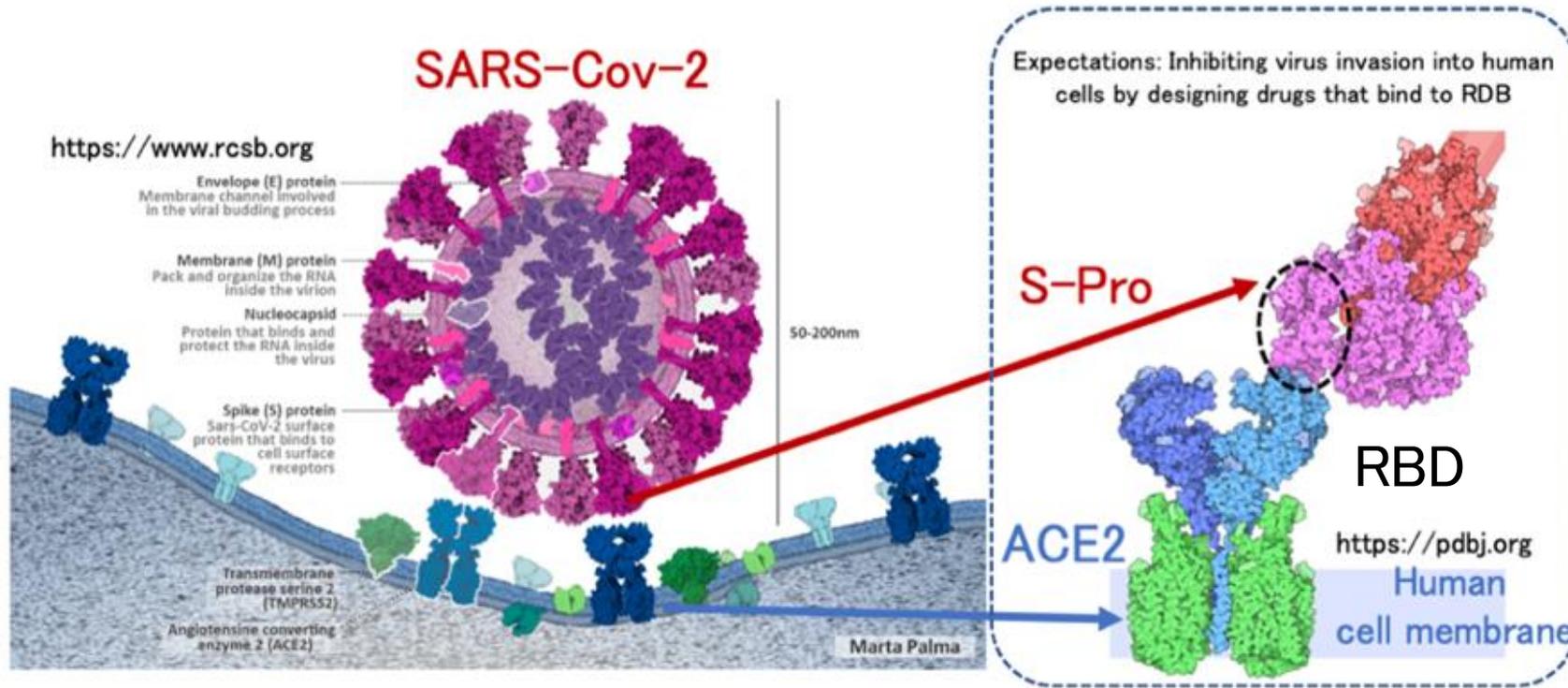
- ✓ ファイル繋がり版 (各ランクにおいて数GB程度の入力データ読み込み。中間ファイル群はLLIO第1階層に書き出し。)
- ✓ メモリ繋がり版 (方位探索においてBO導入の際にも利用)
- ✓ 結果は, ストレージの圧迫を避けるためLLIO第1階層上で.tgz形式にアーカイブしてからFEFSにコピー。
- ✓ 共有テンポラリからノード内テンポラリに変更することで, 大規模並列時の性能が大幅に向上(FY2024)

ノード数	ELAPSE(I/Oあり)			ELAPSE(I/Oなし)		
	Singularity適用	ノート内テンポラリ	共有テンポラリ	Singularity適用	ノート内テンポラリ	共有テンポラリ
451	00:13:39	00:16:08	00:20:36	00:13:22	00:13:42	00:16:59
4501	00:23:36	00:54:31	02:20:35	00:23:32	00:25:41	01:51:07

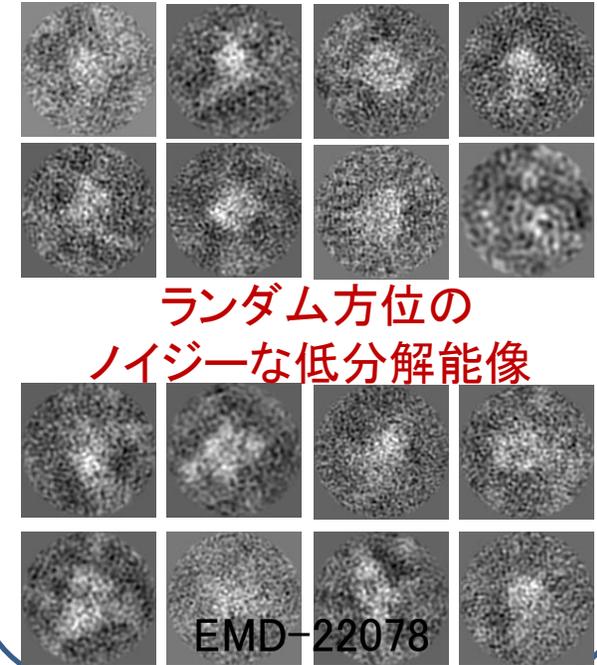
内容

- 研究背景
- これまでの取り組み(-FY2024)
 - CryoMDM構造多形解析ワークフロー開発
 - スパイクタンパク質への適用
- FY2030に目指すところ
- まとめ

CryoMDM法によるスパイクタンパク質構造多形解析



Cryo-EM個別実験データ



- 本研究は、抗体の主要な標的であるRBDやNTDに注目し、スパイクタンパク質の多様な構造状態（遷移状態、準安定状態）の解明を目指す。
- SARS-CoV-2スパイクタンパク質のCryo-EM実験データに対してcryoMDM法を適用した。
 - FY2021-FY2022: 野生型
 - FY2022-FY2023: 変異型(D614G)
 - FY2023-FY2024: 抗体複合体(感染増強抗体、感染抑制抗体)
- FY2025: 感染抑制抗体の親和性成熟を目指しバーチャル空間解析の実施を予定(hp250078)。

構造探索(大域・局所)の例

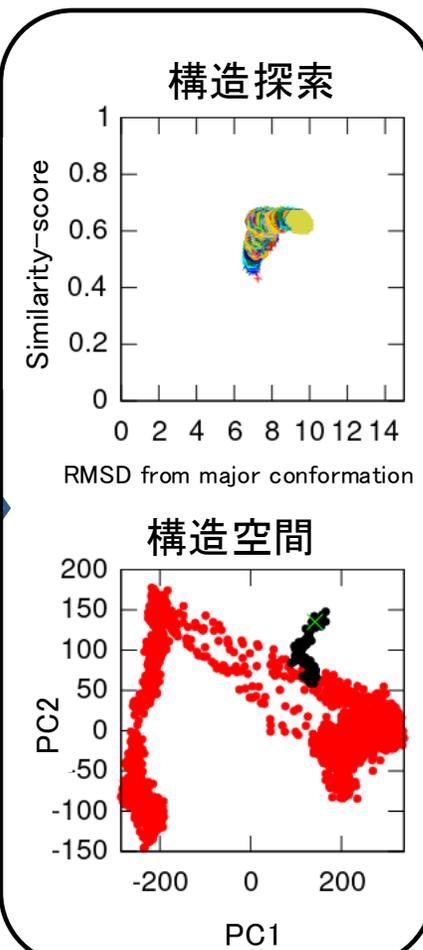
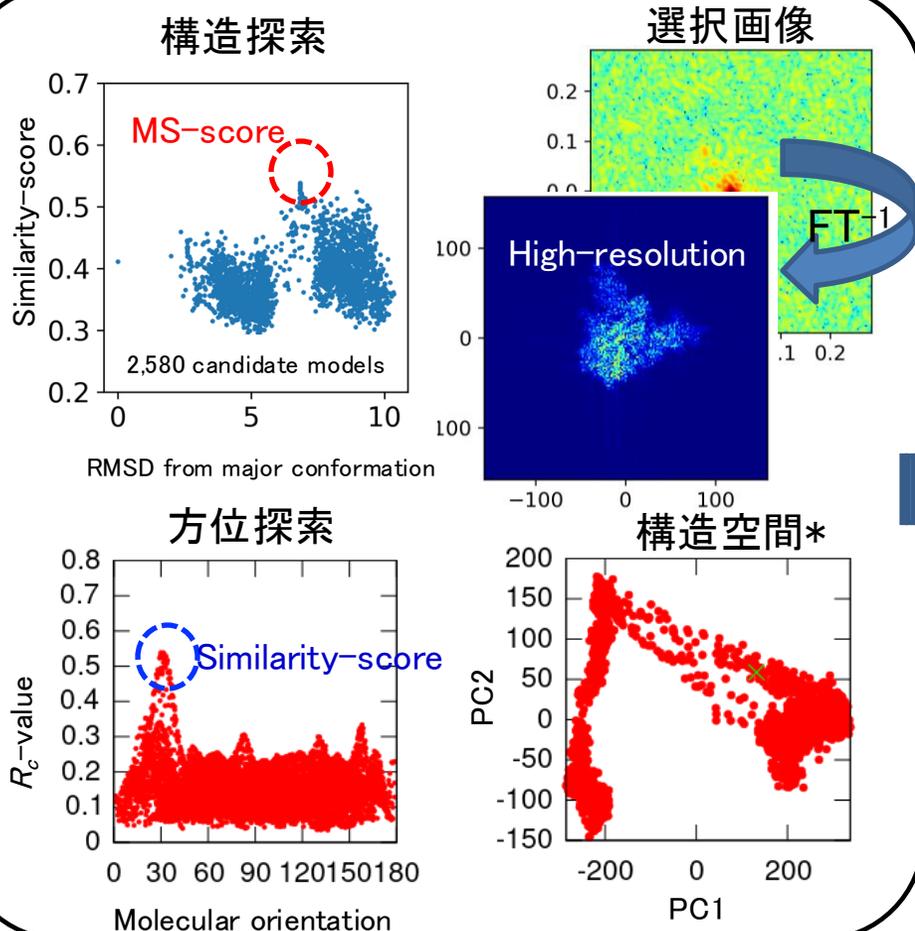
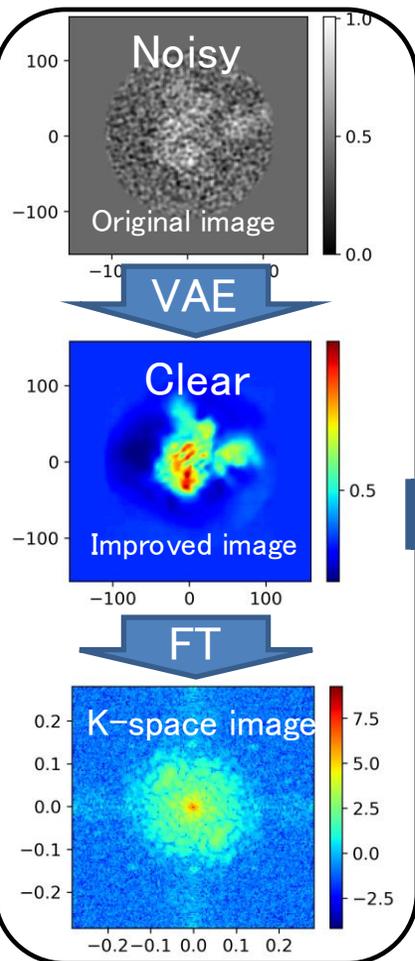
大域構造探索:Cryo-EM Matching

局所構造探索:PaCS-MD

cluster0

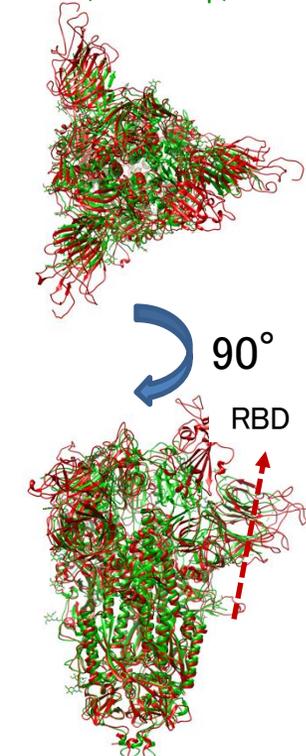
入力:粒子像

出力



推定した尤もらしい構造

cf. 主要な構造 (PDB:6x6p)



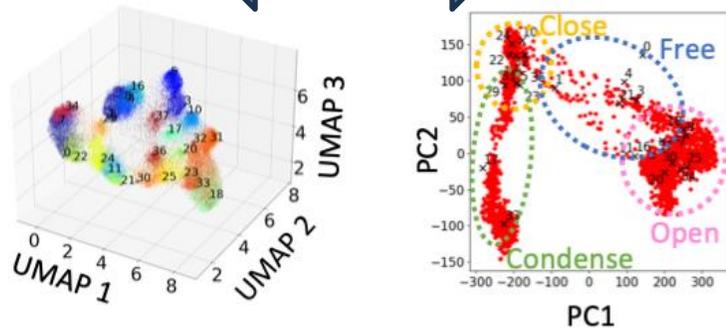
- 1枚の標的実験画像から、主要な構造とは異なる構造状態の推定に成功した。
- 粒子像に映し出されたスパイクタンパク質の多様な構造状態を原子分解能で評価することができる。
- 推定した立体構造に対してFMO計算を実施することで分子内部相互作用解析を実施できる。

*大域構造空間は公開MDデータを活用し構築した: (1) Mori T, Jung J, Kobayashi C, Dokainish HM, Re S, Sugita Y., Biophys J.16;120(6):1060-1071. (2021), (2) Dokainish HM, Re S, Mori T, Kobayashi C., Jung J., Sugita Y., eLife 11:e75720. (2022)

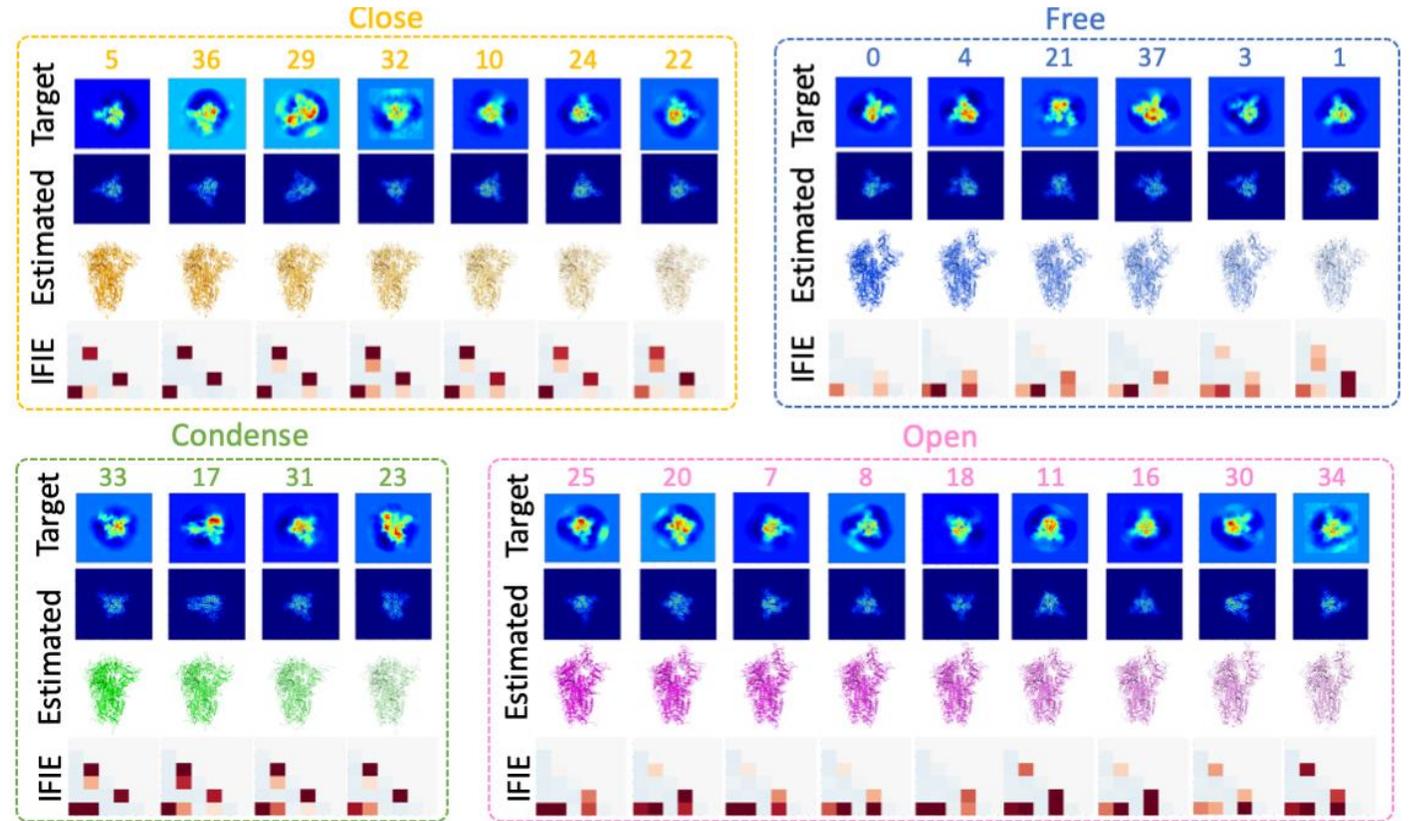
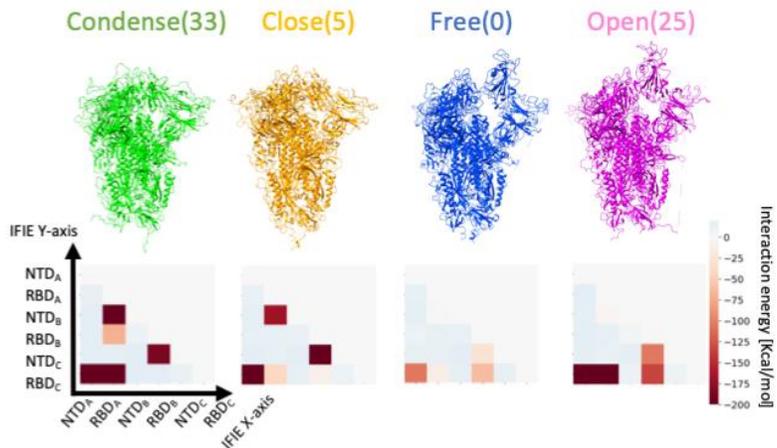
実験データ(野生型)の網羅的解析結果

画像空間 ← cryoTM法 → 構造空間

26の標的像に対する網羅的解析結果



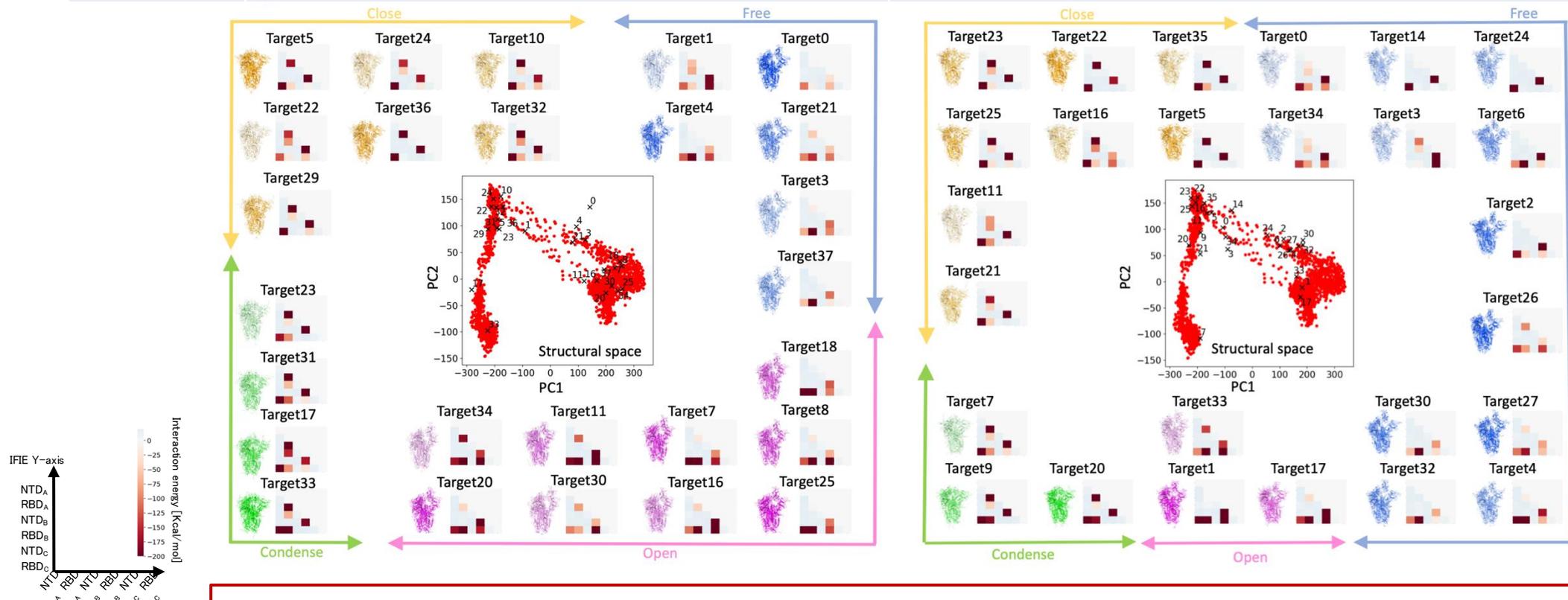
特徴的な4状態



- 26の標的像に対して、cryoMDM法により網羅的な解析を実施した。
- 推定構造に対してアンサンブルFMO計算を行うことで、ドメイン間相互作用エネルギー(IFIE)を求めた。
- 4つの特徴的な状態(Condense, Close, Free, Open)に大別できることを示し、これまで実験的に得られていない2つの新しい中間状態(Condense, Free)を解明した。

ドメインIFIEにおける野生型と変異型(D614G)の比較

	野生型	変異型(D614G): 感染力増大
Condense	17, 23, 31, 33 (4, 15.3 %)	7, 9, 20 (3, 11.5 %)
Close	5, 10, 22, 24, 29, 32, 36 (7, 26.9 %)	5, 11, 16, 21, 22, 23, 25, 35 (8, 30.7 %)
Free	0, 1*, 3, 4, 21, 37 (6, 23.1%) *はCloseに比較的近い状態	0*, 2, 3*, 4, 6, 14, 24, 26, 27, 30, 32, 34* (12, 46.1 %)
Open	7, 8, 11, 16, 18, 20, 25, 30, 34 (9, 34.6%)	1, 17, 33 (3, 11.5 %)



➤ 変異型(D614G)ではFree状態にシフトしていることを解明し、感染力増大の要因を示唆

内容

- 研究背景
- これまでの取り組み(-FY2024)
 - CryoMDM構造多形解析ワークフロー開発
 - スパイクタンパク質への適用
- FY2030に目指すところ
- まとめ

FY2030に目指すところ

- ① 実験融合構造多形解析→**点から線へ**
 - 現状: 数十の構造状態を原子分解能で解明
 - ポスト富岳: パスウェイ上の連続的な中間状態を原子分解能で解明
- ② 構造多形を考慮した創薬の実現→**抗体創薬支援**
 - 現状: 数百変異を導入し親和性成熟を検討(FY2025)
 - ポスト富岳: 数万変異を導入し親和性成熟を検討
- ③ WF構築→**サイト横断的WF構築**
 - 現状: 1つのサイト(富岳&Pre-post)でWF構築
 - ポスト富岳: 複数サイトの多様な計算機(CPU,GPU,量子コン)をまたぐWF構築

①実験融合構造多形解析→点から線へ

現状：実験データから原子分解能で数十の構造状態を解明

ポスト富岳：複数の観測点を繋ぎ、パスウェイ上の連続的な構造変化を原子分解能で解明

要素技術開発(FY2025-)：“生成AI構造サンプリングエンジン”

→エネルギー解析により遷移状態、準安定状態の解明

FY2024

FY2025- (Preliminarily)

confidential

②構造多形を考慮した創薬の実現→抗体創薬支援

現状: 数百変異を導入し親和性成熟を検討(FY2025-:1,500万NH)

FMO計算に必要な資源: 600変異パターン x 5rank x 96node x 20hour = 576万NH

ポスト富岳: 数万変異を導入し親和性成熟を検討

要素技術開発: 変異体解析WF構築

→創薬DXプラットフォーム高度化(低分子創薬→高分子創薬)

感染抑制抗体複合体に対する網羅的な構造推定結果(FY2024)

相互作用エネルギー(fab-EとSpro-B)

confidential

③WF構築→サイト横断的WF構築

現状:1つのサイト(富岳&Pre-post)でWF構築

ポスト富岳:複数サイトの多様な計算機(CPU,GPU,量子コン)間でWF構築

- 実施例1:実験観測施設からポスト「富岳」へTB規模のデータを転送し、On-the-flyで解析を実施することで、実験画像から3次元原子構造を瞬時に取得
- 実施例2:WF中で(1)シミュレーションによる合成データ生成(CPU)と(2)それを用いたAI学習(GPU)を反復的に行い、“生成AI構造サンプリングエンジン”を高効率に構築
- 実施例3:構造探索を量子コンピュータを用いて加速

サイト横断的に1つのWFを構築



まとめ

- 生体分子の多様な構造状態を解明するため、実験データ融合生体分子構造多形解析ワークフローを「富岳」に構築した
 - cryoMDM法により1枚の実験像から原子分解能で直接立体構造を推定可能
- スパイクタンパク質のcryoEM実験データ解析に適用した
 - Free状態(新規)を含む複数の構造状態を解明
 - 変異体においてFree状態に状態シフトが起こっていることを解明
 - 抗体複合体解析において多様な結合様式を解明
 - 数百変異の導入により抗体の親和性成熟を目指した解析を実施中
- FY2030に目指すところ
 - 実験データに基づきパスウェイ上の連続的な構造変化を原子分解能で解明
 - 数万変異の抗体に対する親和性成熟の検討により高分子創薬の支援を実現
 - サイト横断的なWFを構築することで実験データ融合解析をon-the-flyで実現

謝辞

- ・公益財団法人計算科学振興財団 研究教育拠点(COE)形成推進事業「分子シミュレーションに基づくゲノム医療・ゲノム創薬基盤の構築」
- ・理研-富士通共同研究費「HPCとAIを活用した創薬イノベーション」
- ・AMED BINS「クライオ電子顕微鏡による生体高分子構造解析の支援と高分解能化・高速化・自動化を目指した高度化」
- ・文部科学省「富岳」高度化・利用拡大枠利用課題「バイオメディカル応用のためのシミュレーション・AI技術開発による「富岳」の高度化」
- ・一般課題「低分子化抗体デザインを目指したcryoMDM法によるスパイクタンパク質-抗体の構造多形解析に関する研究」等(課題番号: hp210126, hp220078, hp230102, hp240109)
- ・「HPCIを活用したFMO創薬プラットフォームの構築」(課題番号hp210130, hp220143, hp230131, hp240162)
- ・Society 5.0推進利用課題『「富岳」を機軸とした創薬DXプラットフォームの構築』(課題番号: hp220284)
- ・成果創出課題『「富岳」で目指すシミュレーション・AI駆動型次世代医療・創薬』(課題番号: hp230216, hp240211)
- ・cryoMDMの開発においては、次の方々に特に感謝申し上げます。
 奥野恭史教授(京都大学)、難波啓一教授(大阪大学)、加藤貴之教授(大阪大学)、水口賢司教授(大阪大学)、Gert-Jan Bekker博士(大阪大学)、神谷成敏教授(兵庫県立大)、城地保昌博士(理研・SPRING-8)、松本篤幸准教授(京都大学)、池口満徳教授(横浜市立大)、寺山慧准教授(横浜市立大)、金田亮博士(理研)、馬彪(理研)、井阪悠太博士(理研)、千葉峻太郎博士(理研)、秋永宣伸博士(理研)、笹倉陽子様(理研)、郷信広名誉教授(京大)、杉田有治先生(理研)、Florence TAMA教授(名古屋大学)、宮下治博士(理研)、高田彰二教授(京都大学)、仁田亮教授(神戸大学)、川鍋友宏博士(理研)、松田元彦博士(理研)
- ・FMO計算の実施においては、次の方々に特に感謝申し上げます。
 福沢薫教授(大阪大学)、本間光貴先生(理研)、加藤幸一郎准教授(九州大学)、渡邊千鶴博士(理研)
- ・OpenFold-for-Fugakuの開発においては、次の方々に感謝申し上げます。
 富士通株式会社 コンピューティング研究所
 大山 洋介博士、田淵晶大博士
- ・PaStELの開発においては、次の方々に特に感謝申し上げます。
 富士通株式会社 人工知能研究所
 山崎公裕様、和田裕一郎博士、和田睦世様、藤間光徳博士、吉川和様、相賀健人様、石井美幸様、河東孝博士、中川章博士

本研究は、上記の支援を受けて実施しました、改めて感謝申し上げます。