

# タンパク質 X 線結晶構造解析におけるビッグデータ化と構造生物学における新たな展開

松浦滉明<sup>1</sup>, 坂井直樹<sup>1,2</sup>, 河野能顕<sup>1</sup>, 山本雅貴<sup>1</sup>, 平田邦生<sup>1</sup>

<sup>1</sup>理化学研究所 放射光科学研究センター 〒679-5148 兵庫県佐用郡佐用町光都 1-1-1

<sup>2</sup>高輝度光科学研究センター 〒679-5198 兵庫県佐用郡佐用町光都 1-1-1

■要約

タンパク質 X 線結晶構造解析では近年、自動測定の普及によりビッグデータ化が進んでいる。大量の結晶からのデータ取得の利点の1つとして、多様な結晶構造を捉えられる可能性があることが挙げられる。階層的クラスタリング (HC) は従来、単波長異常分散 (SAD) 法などによる位相決定において同型性の高いデータを抽出するために使われていたが、構造多様性解析においても有効であることが分かってきた。本稿では、SPring-8 BL32XU を中心に開発された ZOO システムなどの自動化技術の概略とともに、現在開発研究を進めている HC を用いた構造多様性解析について解説する。

## 1. はじめに

X 線結晶構造解析は半世紀以上にわたって、タンパク質の原子レベルの構造可視化における主要な実験手法として利用されてきた。最近では、高輝度光源、高速 X 線検出器に加え、自動試料交換ロボットの開発<sup>1)</sup>や、回折実験の自動化<sup>2,3)</sup>により、通常の測定 (各結晶につき 360° の振動データを取得) は、1 データセットあたりおよそ 3-5 分で完了する。回折イメージから強度データへ変換するデータ処理についても様々な自動処理パイプライン<sup>4-6)</sup>が開発されており、回折データ収集と並行してクラスター計算機によってデータ処理が進められる。SPring-8 や Photon Factory (PF) の構造生物学ビームラインにおいて自動測定を利用するユーザーは、結晶試料を専用カセット (Uni-puck) に詰めて低温輸送容器に入れて送付するだけで、来所することなくデータセットを得られる。

近年のタンパク質 X 線結晶構造解析 (MX) では、測定やデータ処理の自動化により、ビームタイムあたりに得られる回折データの量が増加 (ビッグデータ化) している。SPring-8 構造生物学ビームラインでは、理研ビームライン BL32XU を中心に開発した ZOO システム<sup>3)</sup>を用いた自動測定が利用可能であり、1 日 (24時間) でおよそ 250 データセットの回折データ収集が可能である。このような高データレートを活用したタンパク質結晶構造解析は「HDRMX (High Data-Rate MX)」<sup>7)</sup>と呼ばれ、得られるビッグデータを活用してどのような構造生物学研究を展開できるかを課題の1つとして、世界中の放射光施設において開発研究が進められている。

タンパク質結晶は同じ結晶化条件で得られた場合であっても、異なる結晶では空間群や格子定数が異なったり、結晶格子がほぼ同じであっても結晶内分子の構造 (タンパク

質分子の一部やアミノ酸残基の配向など) が異なる場合がある (これを本稿では結晶構造多型と呼ぶ)。タンパク質の構造はその機能と密接に関連しているため、複数の結晶からデータを取得し多様な結晶構造を捉えることが出来れば、そのタンパク質の機能過程において重要な構造の決定、ひいては構造ダイナミクスの理解に迫ることが出来る可能性がある。そこで、SPring-8 構造生物学ビームラインでは、大量の結晶から得られた回折データを分類し、構造多様性解析へ展開するための手法開発に取り組んでいる。回折データの分類においては、これまで「階層的クラスタリング (HC)」や「遺伝的アルゴリズム (GA)」を用いた研究が複数報告されている<sup>8-10)</sup>。最近では、複数の結晶から得られた回折データを HC を用いて分類することで、結晶構造多型をうまく分類して解析できることが分かってきた<sup>11-13)</sup>。

本稿では、まずタンパク質 X 線結晶構造解析におけるデータ収集や解析法の変遷と、SPring-8 BL32XU を中心に進められてきた回折実験やデータ解析の自動化に関する開発研究の概略を述べる。次に、現在 SPring-8 で進めている HC を用いた構造多様性解析について、実際の解析例を含めて紹介する。

## 2. タンパク質結晶からの自動回折データ収集・データ処理

### 2.1 タンパク質 X 線結晶構造解析におけるデータ収集、解析法の変遷の概略

タンパク質結晶は、回折データ測定中の放射線損傷により刻一刻と回折能が低下するため、高精度な構造解析を行うための照射線量上限がいくつか提案されている<sup>14-17)</sup>。回折データ収集を常温で行う場合には、放射線損傷の影響

が大きく1つの結晶から構造決定に必要な十分な完全データの取得が困難であるため、同型性が十分に高い結晶を複数準備し、それらから一定の振動範囲ずつ回折データを収集し、足し合わせる（マージする）ことで完全データを得ることが多い。2000年代頃からは低温（100 K）での回折データ測定（クライオ測定）が徐々に主流となり、常温における測定と比較して放射線損傷を大幅に低減できるようになったことで、単一の結晶から完全データセットを取得し構造決定するケースが一般的になってきた<sup>18)</sup>。このように現在のタンパク質 X 線結晶構造解析では単一結晶からの完全データセット収集が一般的になっているが、クライオ測定が主流になる以前は複数の結晶を用いる場合が多かった。

2000年代後半からは、複数の結晶を利用した構造解析が再び注目されてきた。これは、創薬ターゲットとして注目されている GPCR（G-Protein-Coupled Receptor；Gタンパク質共役受容体）などの膜タンパク質をはじめとする微小結晶構造解析が増えてきたことが一つの要因である<sup>19,20)</sup>。一般に膜タンパク質の結晶化は困難であるが、LCP（Lipidic Cubic Phase；脂質キュービック相）法と呼ばれる結晶化法により、膜タンパク質でも良質な結晶が得られるようになった。しかし、LCP法で得られる結晶は5-20  $\mu\text{m}$  角程度であることが多く、当時の解析可能な結晶サイズの目安（30  $\mu\text{m}$  角以上）よりもさらに小さい。このような微小結晶は照射体積に含まれる結晶格子の繰り返し数が少ないため、得られる回折強度は弱くなる。したがって、放射線損傷の影響が重篤にならないよう一定の照射線量以内で完全データセットを得ようとする、十分に大きな結晶を用いた場合と比較して高分解能データを得ることが難しい。そこで、結晶あたりに測定する振動範囲を狭めて1フレームあたりの照射線量を増やし、複数の結晶からデータを収集し、マージすることで、構造決定に必要な完全データセットを得るというアプローチ（SWSX（small-wedge synchrotron crystallography）法とよばれる）による測定、構造解析が進められるようになった。SWSX法による微小結晶からの構造決定においては、マイクロビームの利用可能なビームラインが整備されたことも大きく貢献している。SPring-8 BL32XUでは1  $\mu\text{m}$  集光の高輝度ビームが利用可能であり<sup>21)</sup>、このような高輝度微小ビームによって数多くの微小結晶からの構造解析に成功してきた<sup>22-24)</sup>。

最近では、マイクロビームを微小結晶解析のみならず、十分に大きな結晶が得られる場合にも活用する例が増えてきた。単点露光で測定する場合には結晶と同程度のサイズの X 線ビームを利用するのが一般的であるが、マイクロビーム利用すると、結晶の端からもう一方の端に向かって並進移動しながら振動データを取得する「ヘリカルデータ収集（ヘリカル測定）」<sup>25)</sup>というデータ収集スキームが利用できる。露光点を移動させることで放射線損傷が分散さ

れるため、単点露光で完全データを取得する場合と比較して、1フレームあたりに照射可能な X 線量を増やすことが出来る。また、単点露光の場合には測定の後半のフレームほど放射線損傷の影響が蓄積されていくが、ヘリカル測定では各フレームにおいて照射された X 線量は均一であると考えられることができる。この特徴を利用すると、不均質な結晶しか得られない場合であっても、良質な回折データをもたらす部分のデータのみを利用して構造解析を行うといった応用的な解析が可能になる<sup>26,27)</sup>。

## 2.2 タンパク質 X 線結晶構造解析における回折データ収集～構造解析の概略および ZOO システム/NABE システムによる自動化

タンパク質結晶の回折測定<sup>28)</sup>は、一般に、結晶化プレートで作製したタンパク質結晶をループに拵ってゴニオメータにマウントして行われる。手動測定の場合には、ゴニオメータにマウントしたループを同軸カメラで観察し、X線照射位置に結晶を移動し、回転中心を合わせる（センタリング）。50-100  $\mu\text{m}$  程度以上の十分な大きさの結晶を用いる場合は、1つのループに1つの結晶を拵って測定を進める。この場合、同軸カメラによる観察で結晶位置を視認できることがほとんどである。一方、微小結晶の場合には、1ループに1結晶だけマウントすると測定の効率が悪く、また大量の微小結晶が析出した結晶化ドロップから1つの結晶だけを拵うことも難しい。さらに同軸カメラの画像から結晶を探し、センタリングすることも難しい。

微小結晶のように結晶位置の視認が困難な場合には、ループ上を弱い X 線で走査し、タンパク質結晶様の回折点の有無に基づいて結晶の有無を判断する、ラスタースキャンが有効である。ラスタースキャンを手動で行う場合、SPring-8 構造生物学ビームラインでは BSS（Beamline Scheduling Software）<sup>29)</sup>とよばれる測定用プログラムから実行することが出来るが、各点で得られた回折像はユーザー自身が確認しながら、タンパク質結晶様の回折の有無を記録する必要がある。さらに SWSX 測定の場合には、大量の微小結晶について各結晶の測定条件を設定し、測定を進めるのは非常に手間がかかり、ヒューマンエラーも生じかねない。

このようなルーティン測定を自動化するため、BL32XUを中心に自動データ収集システム ZOO の開発を進めた。ZOOには SPring-8 周辺の動物の名前を冠した様々なプログラムがあり、INOCC がループセンタリングを行い、SHIKA がラスタースキャンの結果から結晶位置を特定し、KUMA が測定条件を設定するなど、個々の作業を行う各プログラムを ZooNavigator が統合することによって自動測定が実現されている。BL32XUにおける自動測定の様子、個別のプログラムや開発背景などの詳細は参考文献を参照して頂きたい<sup>30-33)</sup>。

ZOOでは前述した SWSX 測定、ヘリカル測定などの主

要な測定スキームに加えて、より微小な結晶を対象としたSSROX (Serial Synchrotron Rotation Crystallography) 法<sup>34,35)</sup>による測定も可能である。単点露光やヘリカル測定の場合には、1結晶あたりの測定時間は試料の交換も含めて5-6分程度で、1日連続測定を行えば250データセットほどが得られる。得られるデータが増えると、次に問題になるのはデータ解析である。回折データ取得後に必要な処理としては、データ処理(空間群、格子定数の決定や各指数の反射強度のリスト化)、位相決定、モデル構築、構造精密化がある。ZOOの中にはKAMOとよばれる自動データ処理パイプラインがあり、測定されたデータを自動検出し、測定と並行してデータ処理を行っている。しかし、データ処理によって得られる空間群、格子定数、統計値テーブルなどの結果のみでは得られた複数のデータの間にもどのような違いがあるかなどを判断することは困難であり、結晶構造の違いなどを調べるには、位相決定などデータ解析を進め、電子密度マップを確認しなければならない。

位相決定では、類似のタンパク質の構造を利用したMR (Molecular Replacement; 分子置換)法がよく用いられる。MR法で位相決定できない新規のタンパク質構造の場合には、SAD (Single Anomalous Dispersion; 単波長異常分散)法などの実験的位相決定が必要となる。同じタンパク質試料について大量の結晶から構造の違いなどを検討する際には、一度構造が決まればその構造モデルを利用してMR法を適用することができる。MR法を用いた位相決定、モデル構築、構造精密化については様々なプログラムが開発されており、ルーティン的な解析を自動化するためのパイプラインなども進められている<sup>36)</sup>。SPRING-8構造生物学ビームラインでは、KAMOによるデータ処理に続くデータ解析の自動化のため、BL32XUを中心に自動構造解析パイプラインNABEの開発を進めている<sup>37)</sup>。

創薬におけるフラグメントスクリーニングでは、複数の薬剤候補化合物と標的タンパク質との複合体の結晶構造を決定することで相互作用様式の決定や薬剤候補化合物のさらなる合理的デザインを行う。薬剤候補化合物の結合の有無や相互作用様式はデータ処理後の強度データのみでは分からないので、作用部位(化合物結合部位)周辺の電子密度マップの確認が必要である。NABEの中のプログラムの1つであるMIZUTAKIは、KAMOによる自動データ処理で得られた強度データから、位相決定、構造精密化を自動で行い、指定残基周辺の電子密度マップを一覧できるレポートHTMLを出力する。これにより、自動測定で大量に得られたデータの中から結合する化合物などを、簡単かつ効率的に電子密度マップで確認することができる。

### 3. 階層的クラスタリング (HC) を用いた回折データの分類

#### 3.1 HCによる回折データ分類の概略

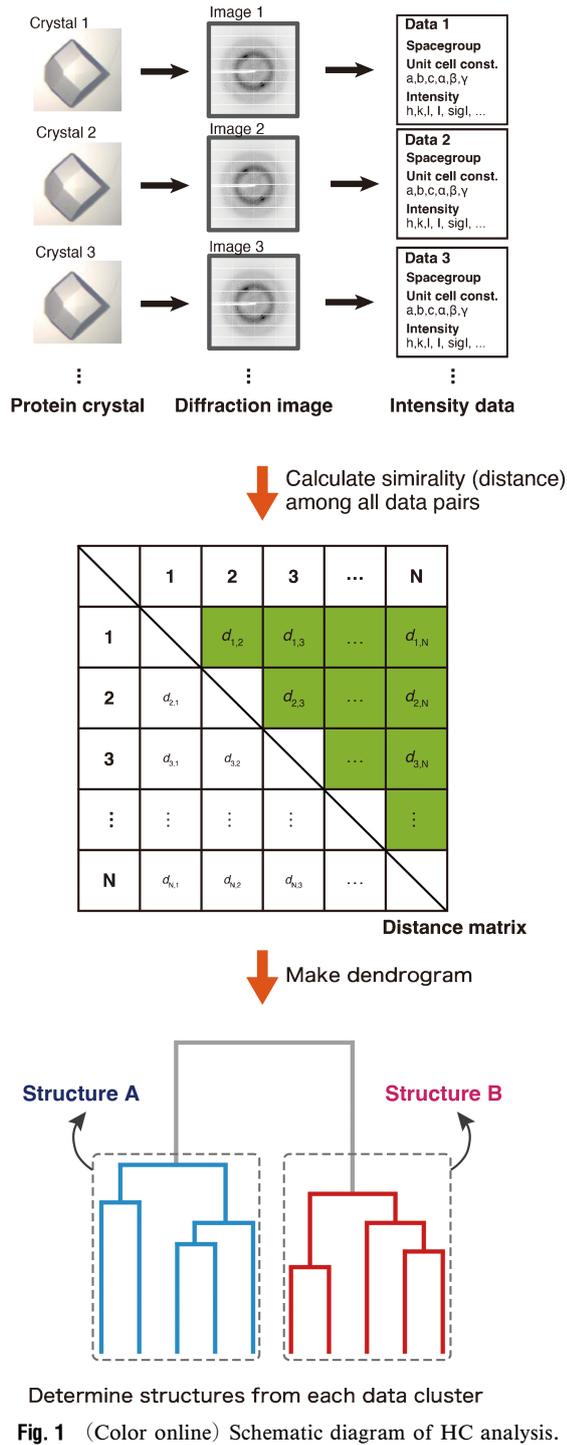
同じ条件下で結晶化して得られたタンパク質結晶であっても、結晶構造が異なる場合がある。このことは、同じタンパク質試料であっても大量の結晶からデータを取得することで、多様な構造を観測できる可能性を示唆している。「階層的クラスタリング (HC)」は同型性の高いデータの選別に用いられており、データ間の類似性のみに基づいて分類を行うため、得られたデータの中にいくつ構造多型が含まれるかなどの事前情報を必要としない。HCによる解析の概略をFig. 1に示した。

しかし、実際のケースではどのデータが結晶構造が異なるかという正解が分からないので、HCを適用して分類しただけでは、どのデータ群(クラスター)に注目して解析を進めれば良いかは分からない。HCによる分類結果に大きく寄与するのは、「類似性」の定義と「リンケージ法」とよばれるクラスター同士の距離の定義の2つである。一般的にHCによるデータ分類では、類似度やリンケージ法を変えて得られた結果を見て、どの組み合わせ(条件)が良いかを総合的に判断する。回折データの分類においては、類似度として格子定数または共通反射の強度の相関、リンケージ法としてWard法や完全連結法がよく利用されている<sup>4,8,9)</sup>。HCを適用するとデータ同士の距離関係を可視化するデンドログラム(樹形図)が得られるので、大きく距離の離れているクラスターをいくつか選んで、電子密度マップを比較しどのような構造の違いを確認する。

KAMOは複数データのマージも自動化しており、マージ前にHCを適用してデータ分類を行う。NABEによる自動解析と組み合わせ、様々な条件での分類結果を検討することもできるが、構造の違いがどこに見られるかなどはユーザー自身が得られた電子密度マップを見て詳細に調べなければならない。とはいえ、ZOOは一定の照射X線量で測定を行っており、結晶試料のばらつき以外の条件はある程度一律に揃っているため、適切な分類条件のもとでHCを適用すれば、同型性の高いデータはある程度決まった距離(閾値)以内にクラスター化されるのではないかと考えた。そこで、構造既知の標準試料データを用いて構造多型の分類に適した条件や、同型データがまとまる際の距離の指標などについて検討を行った。

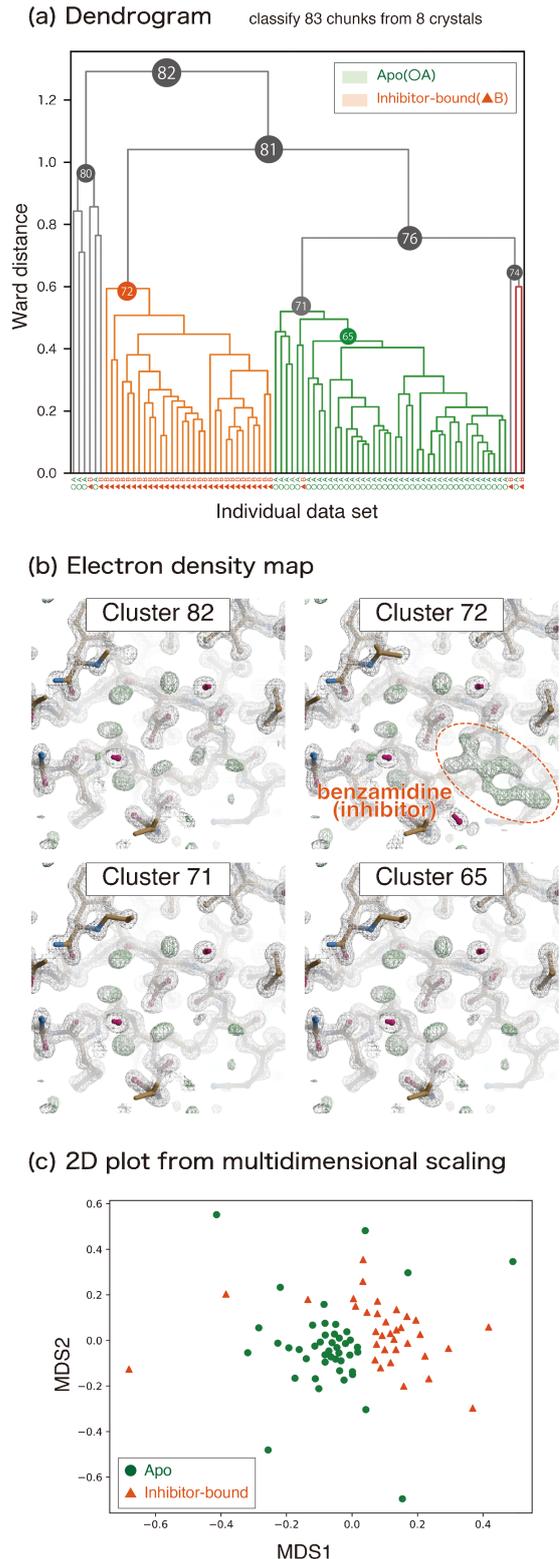
#### 3.2 構造既知の標準試料データを用いた条件検討

HCによる構造多型の分類に適した条件を探るため、標準結晶としても利用されるトリプシン結晶(長辺が200 $\mu\text{m}$ 程度)のデータを利用した検討を行った。トリプシンの阻害剤(ベンズアミジン)結合型と非結合(アポ)型の結晶をそれぞれ4個ずつ調製し、ZOOを利用してヘリカル測定スキームで1結晶あたり360°のデータセットを取



得した。得られたデータはいずれも30°ずつの部分データ (chunk と呼ぶ) に分割してデータ処理を行った。これは、ヘリカル測定データの特徴を活かして結晶内の構造多型についても考慮して解析を行うためである。

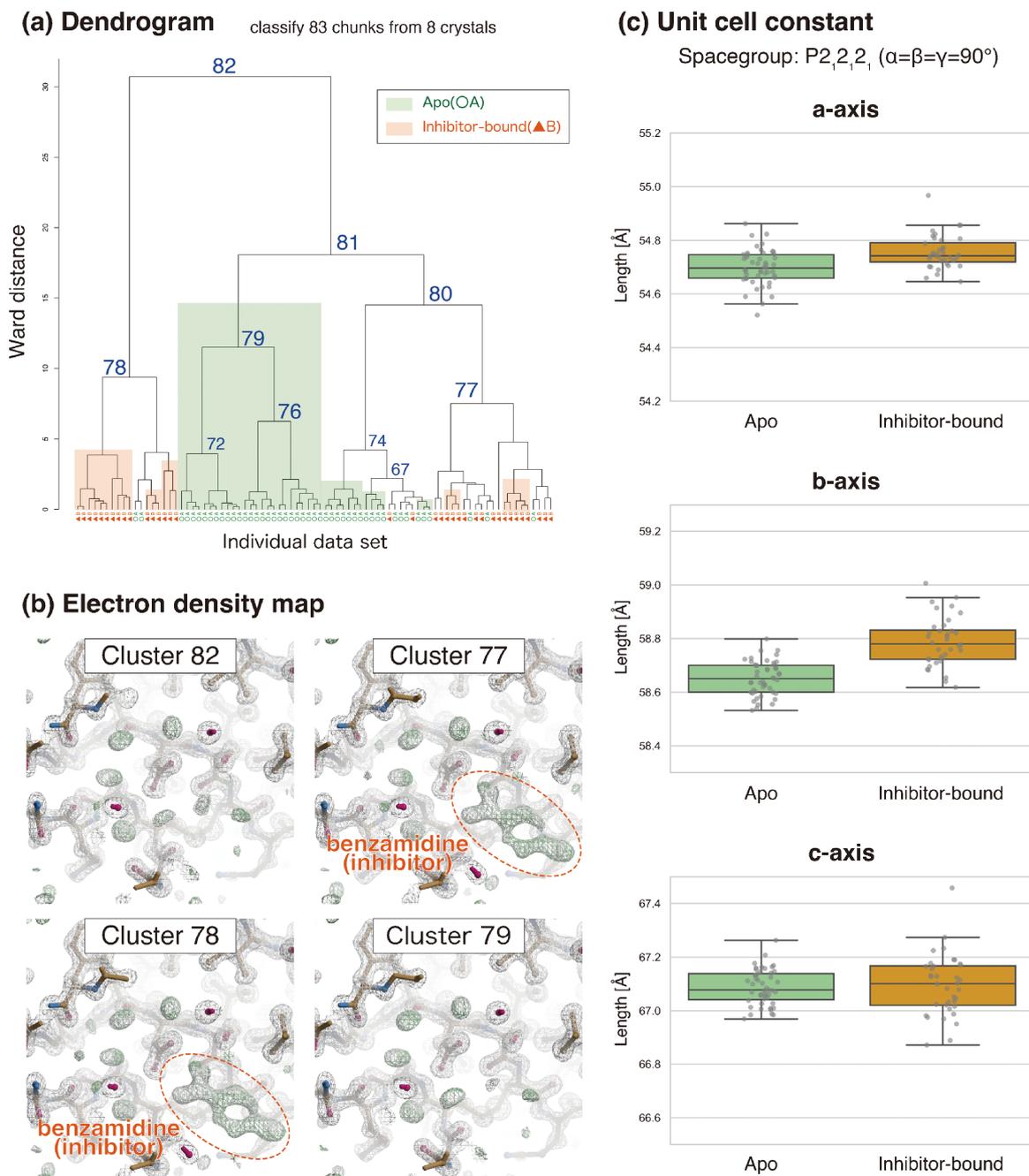
HC による構造多型データ分類の条件検討では、阻害剤結合型およびアポ型トリプシンの chunk をそれぞれ同数混ぜたデータ群を作成 (これを *in silico mixing* と呼んでいる) し、このデータ群に対して種々の条件で HC を適用



**Fig. 2** (Color online) Result of intensity-based HC analysis on *in silico* mixed data set including apo- and inhibitor-bound trypsin. (a) dendrogram and (b) electron density map from the merged data at some clusters (c) data similarity visualized by multidimensional scaling.

した。*KAMO*はHCを行う前に、まず全てのデータ群の中から等価な結晶格子をもつデータセットを抽出し、さらにTukeyが提唱した四分位範囲を基準とする外れ値検出法を用いて、格子定数が大きく異なるデータセットを除去する。前処理の後残ったデータ群に対して格子定数または共通反射の強度相関に基づくHCを適用し、2種類のトリプシンのデータを分類することを試みた。リンケージ法についてはWard法や完全連結法以外にも利用可能なすべての方法について検討した。

その結果、トリプシンの阻害剤結合型/非結合型のデータ分類においては、強度相関に基づくHCで、リンケージ法にWard法を用いた場合に、2種類のデータがうまく分類された( Fig. 2a)。電子密度マップからも、強度相関に基づくHCによってデータが分類されていることが確認できた( Fig. 2b)。デンドログラムを見ると、アポ型および阻害剤結合型のデータはいずれもWard距離が概ね0.6以下でクラスタ化していた。HCで得られた距離行列に基づいて、多次元尺度構成(MDS)法によって2次的に各



**Fig. 3** (Color online) Result of unit cell-based HC analysis on *in silico* mixed data set including apo- and inhibitor (benzamidine)-bound trypsin. (a) dendrogram and (b) electron density map from the merged data at some clusters (c) Variation of unit cell constants.

chunk の距離関係を可視化すると (Fig. 2c), 「Ward 距離  $\leq 0.6$ 」でクラスタ化したアポ型と阻害剤結合型の chunk はある程度まとまって分布しており, うまくクラスタを形成しなかったクラスタ80やクラスタ74に属するデータは, 分布の中心から大きく離れたところに位置していた。本稿では説明を割愛するが, 数値シミュレーションの結果からも「Ward 距離  $\leq 0.6-0.7$ 」が同型性の高いデータ群の良い指標となることが示唆されており<sup>13)</sup>, この閾値をもとに得られたクラスタを解析することで効率よく構造多型を見つけることが出来る。

一方, 格子定数に基づく HC では, 2種類のデータをうまく分類できなかった。デンドログラム (Fig. 3a) や電子密度マップ (Fig. 3b) を見ると, 一部のクラスタでは片方のデータのみが含まれるクラスタも見られるが, 2種類のデータはそれぞれ複数の小さなクラスタに分かれ, かつ混ざっていた。格子定数に基づく HC による分類が困難であった理由としては, 2種類のデータではいずれの軸においても化合物結合に伴う格子長の変化が微小であったことが考えられる (Fig. 3c)。

#### 4. 実際の試料における構造多様性解析例

アポ型および阻害剤結合型トリプシンのモデルデータを利用した検討において, 共通反射の強度相関に基づいた階層的クラスタリング (HC) が構造多型の分類に有効であり, 同型性の高いデータが概ね Ward 距離  $\leq 0.6-0.7$  でクラスタ化することが分かった。この結果を利用して実際に構造多型解析を行った例を, ここでは2例紹介する。

##### 4.1 *Aquifex aeolicus* 由来 HypD の結晶構造解析

1つ目の例として, 水素変換酵素である [NiFe] ヒドロゲナーゼの成熟化因子の1つである *Aquifex aeolicus* 由来 HypD (*AaHypD*)<sup>38)</sup> における解析結果を紹介する。同じ精製ロットかつ結晶化条件で得られた *AaHypD* 結晶について, 異なる結晶化ドロップから得た計6個の結晶を対象にヘリカル測定を行ったデータを用いた。各結晶からは  $360^\circ$  ずつデータを取得した後,  $30^\circ$  の chunk に分割してデータ処理を行い, 得られた72個の強度ファイルについて, *KAMO* を利用して強度相関に基づく HC およびデータのマージを行った。

HC によるデータ分類の結果, いくつかのクラスタに分かれたが, 大きく2つのクラスタに分かれることがわかった (Fig. 4a)。一方 (クラスタ50や51) では N 末端が2次構造をとっておらず, もう一方 (クラスタ52や42) では N 末端が2次構造をとり, タンパク質側へ折りたたまれていた (Fig. 4b)。さらに, この N 末端の構造変化と対応するように, N 末端がほどけて不安定化されると, HypD 中心付近にある [4Fe-4S] 周辺が不安定化している様子が温度因子の比較から明らかになった。N 末端の

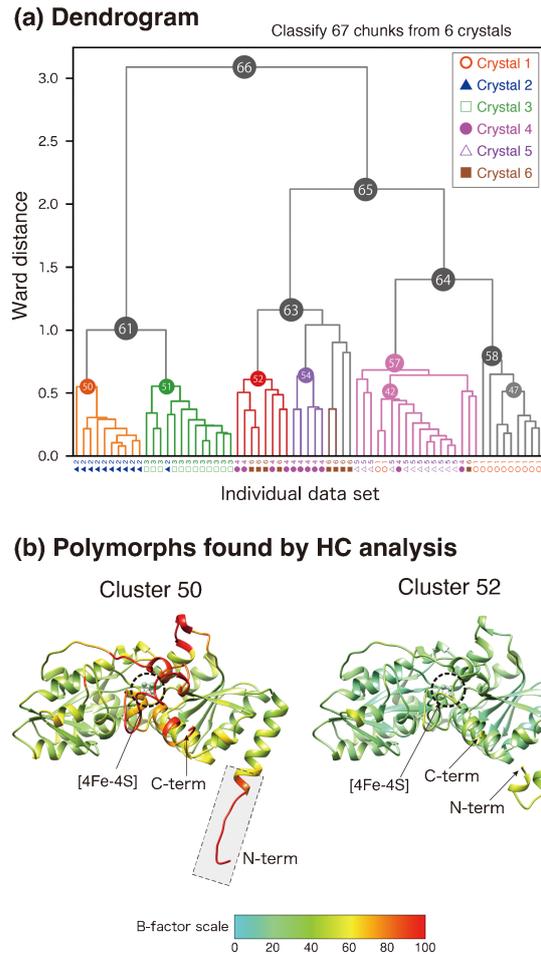


Fig. 4 (Color online) Result of intensity-based HC analysis on *AaHypD*. (a) dendrogram, (b) structural polymorphs found at clusters 50 and 52.

状態の異なる2つのデータ群は Ward 距離  $\leq 0.6-0.7$  を基準に見てみるとさらに細かく分かれているが, それらの中には [4Fe4S] の占有率の違いなど, より細かな構造の違いが見出された。観測された構造変化が HypD の機能とどのように関連しているかについては, 現在調査中である。Ward 距離  $\leq 0.6-0.7$  で分けられたクラスタについて構成するデータを調べてみると, *AaHypD* 結晶の場合には概ね結晶ごとに分かれていることが分かった。したがって, 同じ結晶化条件で得られた結晶であっても, 結晶化ドロップの違いやわずかな沈殿剤条件の差異によって結晶構造が異なる場合があることが示された。

##### 4.2 Transportin-1-ペプチド複合体の結晶構造解析

2つ目の例として, Transportin-1 (Trn1) とペプチドとの複合体の解析例を紹介する。Trn1は標的となるタンパク質を核内へ輸送するための核内輸送受容体であり, 核局在化配列という核への移行シグナルとして働くアミノ酸配列をもつペプチドと共結晶化することで Trn1-ペプチド複合体の結晶を得た。同じ結晶化条件で結晶化し, さら

に同じ結晶化ドロップに析出した結晶を用いて回折データ測定を行った。4つの結晶から、いずれも720°ずつヘリカル測定でデータを取得した後、30°のchunkに分割してデータ処理し、得られた96個の強度ファイルについて、KAMOを用いて強度相関に基づくHCおよびデータのマージを行った。

強度相関に基づくHCによってデータを分類した結果、ペプチドの結合様式には2種類あり、データによってその比率が大きく異なることが分かった (Fig. 5)。一方の結合様式 (Form2) ではペプチドは2次構造をとっており、もう一方の結合様式 (Form1) ではペプチドは2次構造をとっていなかった。また、4.1節で紹介した *AaHypD* 結晶とは異なり、Trn1-ペプチド複合体の場合には、同じ結晶の中でも Form1 優勢のクラスター (クラスター76) に属する chunk と Form2 優勢のクラスター (クラスター72) に属する chunk が見られた (Fig. 5b)。この結果から、マイクロビームを用いたヘリカル測定と30°程度の部分データへの分割処理、およびHCによる回折データ分類を組み合わせることで、結晶内の構造多型の分類も可能になることが示

された。さらに、Trn1-ペプチド複合体結晶の共結晶化を pH の異なる結晶化条件で行い、得られた結晶を用いて同様の解析を行ったところ、pH の変化に従って Form1/Form2 のデータ数の比率が変化した。これらの結果を応用すると、結晶化条件の変化やリガンド (基質) のソーキング (浸漬) 等により構造変化を誘発させ、フリーズトラップによって中間状態の混在した状況でデータ測定を行い、HCを用いて複数の構造が含まれるデータを分類することで、タンパク質の機能過程における構造変化を X 線結晶構造解析法で可視化できるのではないかと期待される。

## 5. まとめと今後の展望

本稿では、SPring-8 BL32XU における微小結晶構造解析を中心に発展してきたタンパク質 X 線結晶構造解析の自動化と、それらの自動化技術によって実現された高データレートを活用したタンパク質 X 線結晶構造解析の一例として、複数の結晶から得られた回折データを階層的クラスタリング (HC) を用いて分類し、多様な構造を捉える

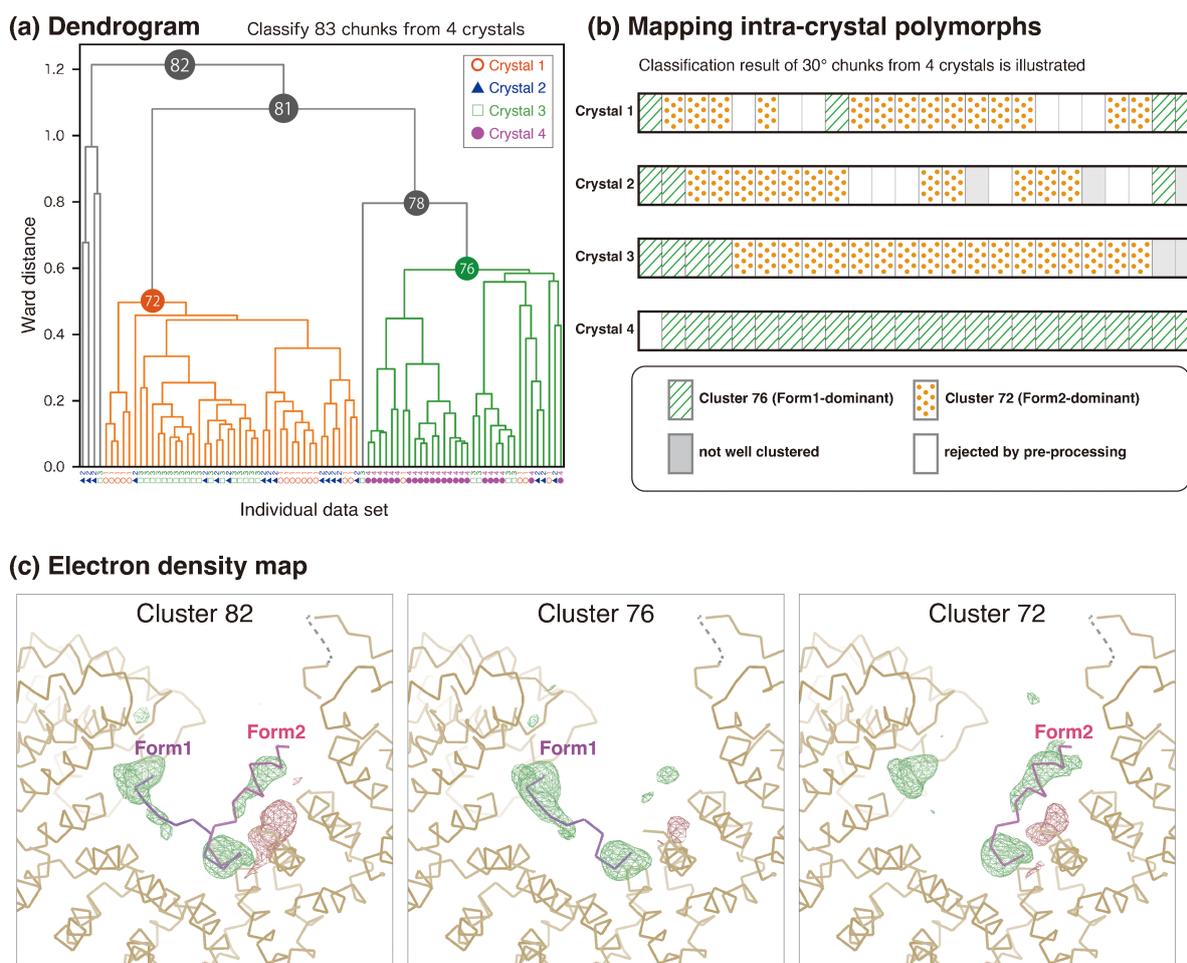


Fig. 5 (Color online) Result of intensity-based HC analysis on Trn1-peptide complex. (a) dendrogram, (b) mapping of intra-crystal polymorph, and (c) electron density map from merged data at some clusters.

試みについて紹介した。

本稿で紹介した解析例では、同じ結晶化条件で得られた結晶において異なる構造が見られたケースのみを紹介したが、実際には構造多型が含まれる場合もあれば、含まれない場合もあり、HCを適用して調べてみないと分からない。一般には、HCを適用して分類された結果の中にいくつ異なるデータ群が含まれるかは、分けられたデータを丁寧に見ながら検討しなければならない。しかし、自動測定データのように測定条件などが一律的に決められていれば、同型性の高いデータ群は概ね一定の距離(閾値)以内でクラスタ化するのではないかと考え、筆者らのグループでは、構造既知の標準結晶トリプシンのデータを用いた検討を行った。この結果得られた「Ward 距離 $\leq 0.6-0.7$ 」は、実際の解析例において構造多型を見出すための指標として有効であることを示した。本稿では紹介できなかったが、数値シミュレーションの結果からも、この指標が構造多型分類において有効であることを示唆する結果が得られている。ただし、この指標は限られた試料、データで検討を行った結果得られたものであり、今後さらに多くの試料において解析、検討を進めていく中で、より適切な指標に置き換わる可能性がある。特に、HC適用前のデータの前処理は高精度な分類を行う上で重要であり、分解能が大きく異なるデータが混ざっているケースなどについて、今後検討していきたい。

HCによるデータ分類を活用したタンパク質X線結晶構造解析のさらなる応用として、現在、構造ダイナミクス解析への展開に向けた研究開発を進めている<sup>39)</sup>。これまでの結晶構造解析ベースの動的構造解析では、結晶内での反応(構造変化)を高精度に同期することが必須であり、結晶試料において中間状態を制御する方法の確立が動的構造解析におけるボトルネックとなっていた。しかし、HCを用いたデータの分類によって、完全な中間状態の制御が出来なくても、混在する中間状態を分類して捉えられる可能性が見えてきた。例えば、タンパク質結晶をリガンドや阻害剤を含む溶液にソーキング(浸漬)し、浸漬時間の異なるいくつかの条件で結晶凍結を行う(フリーズトラップ)することで、反応時間が秒～分単位であれば、その反応過程におけるタンパク質の構造変化を捉えられる可能性がある。常温測定の場合にはインクジェットやレーザーなどを利用したオンラインでの反応誘起が出来れば、反応開始直前から反応中にかけて連続測定して得られたデータを分類することで、連続的に変化するタンパク質構造を解析できるかもしれない。しかし、HCは連続的に変化するデータの解析には向かないので、今後は、連続的に構造変化するデータを分類、解釈するための新たな解析手法の開発についても、測定系の開発と並行して必要になると考えられる。

## 謝辞

本稿で紹介した解析手法の研究開発は文科省科研費・新学術領域「高速分子動画」や、AMED・創薬等先端技術支援基盤プラットフォーム事業(BINDS)の支援を受けて進められているものです。本稿の執筆にあたりご協力頂いた、理化学研究所・放射光科学研究センター・生物系ビームライン基盤グループおよび高輝度光科学研究センター・構造生物学推進室の関係者の皆様に感謝いたします。また、4節で紹介した実際の解析例について貴重なデータをご提供頂いた、分子科学研究所の村木則文博士(現 慶應義塾大学)、奈良先端科学技術大学院大学の藤間祥子博士に厚く御礼申し上げます。

## 参考文献

- 1) H. Murakami, K. Hasegawa, G. Ueno, N. Yagi, M. Yamamoto and T. Kumasaka: *Acta Crystallogr. D* **76**, 155 (2020).
- 2) U. Zander, G. Bourenkov, A. N. Popov, D. de Sanctis, O. Svensson, A. A. McCarthy, E. Round, V. Gordeliy, C. Mueller-Ciekmann and G. A. Leonard: *Acta Crystallogr. D* **71**, 2328 (2015).
- 3) K. Hirata, K. Yamashita, G. Ueno, Y. Kawano, K. Hasegawa, T. Kumasaka and M. Yamamoto: *Acta Crystallogr. D* **75**, 138 (2019).
- 4) K. Yamashita, K. Hirata and M. Yamamoto: *Acta Crystallogr. D* **74**, 441 (2018).
- 5) S. Monaco, E. Gordon, M. W. Bowler, S. Delagenière, M. Guijarro, D. Spruce, O. Svensson, S. M. McSweeney, A. A. McCarthy, G. Leonard and M. H. Nanao: *J. Appl. Crystallogr.* **46**, 804 (2013).
- 6) G. Winter and K. E. McAuley: *Methods* **55**, 81 (2011).
- 7) H. J. Bernstein, L. C. Andrews, J. A. Diaz Jr, J. Jakoncic, T. Nguyen, N. K. Sauter, A. S. Soares, J. Y. Wei, M. R. Wlodek and M. A. Xerri: *Struct. Dyn.* **7**, 014302 (2020).
- 8) J. Foadi, P. Aller, Y. Alguel, A. Cameron, D. Axford, R. L. Owen, W. Armour, D. G. Waterman, S. Iwata and G. Evans: *Acta Crystallogr. D* **69**, 1617 (2013).
- 9) G. Santoni, U. Zander, C. Mueller-Dieckmann, G. Leonard and A. Popov: *J. Appl. Crystallogr.* **50**, 1844 (2017).
- 10) U. Zander, M. Cianci, N. Foos, C. S. Silva, L. Mazzei, C. Zubieta, A. de Maria and M. H. Nanao: *Acta Crystallogr. D* **72**, 1026 (2016).
- 11) T. Nguyen, K. L. Phan, D. Kozakov, S. B. Gabelli, D. F. Kreitler, L. C. Andrews, J. Jakoncic, R. M. Sweet, A. S. Soares and H. J. Bernstein: *Acta Crystallogr. D* **78**, 268 (2022).
- 12) A. S. Soares, Y. Yamada, J. Jakoncic, S. McSweeney, R. M. Sweet, J. Skinner, J. Foadi, M. R. Fuchs, D. K. Schneider, W. Shi, B. Andi, L. C. Andrews and H. J. Bernstein: *Acta Crystallogr. F* **78**, 281 (2022).
- 13) H. Matsuura, N. Sakai, S. Toma-Fukai, N. Muraki, K. Hayama, H. Kamikubo, S. Aono, Y. Kawano, M. Yamamoto and K. Hirata: *bioRxiv*, 2022.09.13.507775 (2022).
- 14) R. Henderson: *Proc. R. Soc. London Ser. B* **241**, 6 (1990).
- 15) E. F. Garman and R. L. Owen: *Acta Crystallogr. D* **62**, 32 (2006).
- 16) O. B. Zeldin, S. Brockhauser, J. Bremridge, J. M. Holton and E. F. Garman: *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20551 (2013).

- 
- 17) S. Baba, H. Matsuura, T. Kawamura, N. Saskai, Y. Nakamura, Y. Kawano, N. Mizuno, T. Kumasaka, M. Yamamoto and K. Hirata: *J. Synchrotron Rad.* **28**, 1284 (2021).
  - 18) W. P. Brumeister: *Acta Crystallogr. D* **56**, 328 (2000).
  - 19) V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Wies, B. K. Kobilka and R. C. Stevens: *Science* **318**, 1258 (2007).
  - 20) D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, X. J. Yao, W. I. Weis, R. C. Stevens and B. K. Kobilka: *Science* **318**, 1266 (2007).
  - 21) K. Hirata, Y. Kawano, G. Ueno, K. Hashimoto, H. Murakami, K. Hasegawa, T. Hikima, T. Kumasaka and M. Yamamoto: *J. Phys. Conf. Ser.* **425**, 012002 (2013).
  - 22) H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A. D. Maturana, R. Ishitani and O. Nureki: *Nature* **482**, 369 (2012).
  - 23) R. Taniguchi, A. Inoue, M. Sayama, A. Uwamizu, K. Yamashita, K. Hirata, M. Yoshida, Y. Tanaka, H. E. Kato, Y. Nakada-Nakura, Y. Otani, T. Nishizawa, T. Doi, T. Ohwada, R. Ishitani, J. Aoki and O. Nureki: *Nature* **548**, 356 (2017).
  - 24) S. Abe, J. Tanaka, M. Kojima, S. Kanamaru, K. Hirata, K. Yamashita, A. Kobayashi and T. Ueno: *Sci. Rep.* **12**, 16031 (2022).
  - 25) D. Flot, T. Mairs, T. Giraud, M. Guijarro, M. Lesourd, V. Rey, D. van Brussel, C. Morawe, C. Borel, O. Hignette, J. Chavanne, D. Nurizzo, S. McSweeney and E. Mitchell: *J. Synchrotron Rad.* **17**, 107 (2010).
  - 26) T. Katoh, T. Sengoku, K. Hirata, K. Ogata and H. Suga: *Nat. Chem.* **12**, 1081 (2020).
  - 27) 沼本修孝: *日本結晶学会誌* **63**, 229 (2021).
  - 28) <https://www.youtube.com/watch?v=bjOjXhEhKvg>
  - 29) G. Ueno, H. Kanda, T. Kumasaka and M. Yamamoto: *J. Synchrotron. Rad.* **12**, 380 (2005).
  - 30) <https://www.youtube.com/watch?v=fZQo1Oo9Aeo>
  - 31) 山下恵太郎: *日本結晶学会誌* **59**, 207 (2017).
  - 32) 山下恵太郎: *日本結晶学会誌* **60**, 104 (2018).
  - 33) 平田邦生: *日本結晶学会誌* **62**, 84 (2020).
  - 34) C. Gati, G. Bourenkov, M. Klinge, D. Rehders, F. Stellato, D. Oberthür, O. Yefanov, B. P. Sommer, S. Mogk, M. Duszhenko, C. Betzel, T. R. Schneider, H. N. Chapman and L. Redecke: *IUCrJ* **1**, 87 (2014).
  - 35) M. Yamamoto, K. Hirata, K. Yamashita, K. Hasegawa, G. Ueno, H. Ago and T. Kumasaka: *IUCrJ* **4**, 529 (2017).
  - 36) M. Wojdyr, R. Keegan, G. Winter and A. Ashton: *Acta Crystallogr. A* **69**, 299 (2013).
  - 37) 松浦況明, 坂井直樹: *日本結晶学会誌* **63**, 219 (2021).
  - 38) N. Muraki, K. Ishii, S. Uchiyama, S. G. Itoh, H. Okumura and S. Aono: *Commun. Biol.* **2**, 385 (2019).
  - 39) 山本雅貴: *日本結晶学会誌* **65**, 55 (2023).

## 著者紹介



### 松浦 晃明

理化学研究所・放射光科学研究センター  
特別研究員

E-mail: hiroaki.matsuura@riken.jp

専門：構造生物化学

#### 【略歴】

2020年3月 兵庫県立大学生命理学研究科  
博士課程修了。博士（理学）。2020年4月  
より現職。



### 坂井 直樹

高輝度光科学研究センター・構造生物学推  
進室 主幹研究員

E-mail: nsakai@spring8.or.jp

専門：構造生物学

#### 【略歴】

2003年3月 北海道大学大学院理学研究科  
博士課程修了。博士（理学）。2003年 同  
大学研究員，特任助手を経て2006年同大  
学 助教。2011年 独・Luebeck 大学  
Senior scientist。2015年 理化学研究所・  
ライフサイエンス技術基盤研究センター  
研究員。2018年 同・放射光科学研究セン  
ター 研究員。2022年 高輝度光科学研究  
センター 研究員。2023年より現職。



### 河野 能顕

理化学研究所・放射光科学研究センター  
専任技師

E-mail: yoshiaki.kawano@riken.jp

専門：放射光構造生物学

#### 【略歴】

1996年 北海道大学大学院理学研究科博士  
課程修了。博士（理学）。1996年 理化学  
研究所 放射光構造生物学研究推進グルー  
プ 協力研究員。1998年 同 生化学システ  
ム研究室 基礎科学特別研究員。1999年  
同 研究技術開発室 研究員。2004年 同  
播磨研究所 城生体金属科学研究室 研究員。  
2008年より現職。



### 山本 雅貴

理化学研究所・放射光科学研究センター  
部門長

E-mail: yamamoto@riken.jp

専門：放射光構造生物学

#### 【略歴】

1991年9月 大阪大学大学院理学研究科博  
士課程修了。博士（理学）。1991年 理化  
学研究所，研究員。2004年 同 播磨研究  
所 研究技術開発室 室長。2008年 同 放  
射光科学総合研究センター 基盤研究部 部  
長。2010年より現職。



### 平田 邦生

理化学研究所・放射光科学研究センター  
専任技師

E-mail: kunio.hirata@riken.jp

専門：放射光構造生物学

#### 【略歴】

2003年12月 大阪大学大学院理学研究科博  
士課程修了。博士（理学）。2003年 高輝  
度光科学研究センター協力研究員。2006  
年 理化学研究所・放射光科学研究セン  
ター 協力研究員。2009年 同 技師。2012  
年より現職。2014-2017年，科学技術振興  
機構さきがけ研究員兼務。

---

# New trends of structural biology utilizing high data-rate and big-data of macromolecular crystallography

Hiroaki MATSUURA<sup>1</sup>, Naoki SAKAI<sup>1,2</sup>, Yoshiaki KAWANO<sup>1</sup>,  
Masaki YAMAMOTO<sup>1</sup>, Kunio HIRATA<sup>1</sup>

<sup>1</sup>RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

<sup>2</sup>Japan Synchrotron Radiation Research Institute, SPring-8, 1-1-1 Kouto, Sayo, Hyogo 679-5198, Japan

---

**Abstract** Data amount in macromolecular crystallography has been increasing due to a high data-rate measurement realized by an automated measurement. Diverse crystal structure may be observed by using diffraction data sets from multiple crystals. “Hierarchical clustering (HC)” is an algorithm which has been used for extracting highly isomorphous data sets in experimental phasing such as the Single Anomalous Dispersion (SAD) method. Recently HC found to be effective for classifying structural polymorph from multiple diffraction data sets. In this article, we describe an outline of automated pipelines, for example *ZOO*, which have been developed at BL32XU, SPring-8. In addition, we introduce our recent developments on structural polymorph analysis using HC.