

## ■第23回日本放射光学会奨励賞受賞研究報告

# 高難度タンパク質微小結晶試料から迅速に構造決定を行うシステムの開発

山下恵太郎 (東京大学大学院理学系研究科)

### 1. はじめに

タンパク質は20種類のL- $\alpha$ -アミノ酸がアミド結合によって直鎖状につながった高分子である。DNAにコードされた遺伝情報に従って合成された後、自発的に折り畳まれて固有の立体構造を形成することで機能を発揮する。その大きさは数アミノ酸から成る小型のペプチドから数千アミノ酸から成る巨大なものまで幅広く、さらにはタンパク質同士や核酸と複合体を形成して機能を持つなど多種多様である。細胞内の構造体として働くものや触媒(酵素)・物質輸送など、それぞれ固有の機能を持ち、生命活動を支えている。構造生物学では、タンパク質の立体構造を詳細に調べることによって、機能発現メカニズムの解明を行っている。構造決定の手法は種々存在するが、なかでもX線結晶学は最もよく利用され(Fig. 1)、ペプチドから巨大複合体まで幅広いサイズのタンパク質構造を高い分解能で決定可能な強力な方法である。X線結晶学の利用がこれほど広がった背景として放射光の発展は欠かせなかった。タンパク質結晶構造解析には1995年頃から放射光が利用されるようになったが、それ以降、放射光の普及に伴って決定された構造の数は飛躍的に増大した<sup>1)</sup>。さらに近年では第三世代放射光施設による光源の高輝度化、検出器や試料調製技術の進歩により、これまで構造決定困難であったような試料も含めより多くの構造がより迅速に決定可能になってきている。

今日でも高難度試料として挙げられるものの一つに膜タンパク質がある。膜タンパク質は脂質二重膜で構成された

細胞膜上で機能し、細胞内外への物質の輸送や情報伝達などを担う非常に重要なタンパク質である。中でもGタンパク質共役型受容体(GPCR)は細胞外からの物理的・化学的シグナルを細胞内へ伝達する膜タンパク質であり、市販薬の40%がGPCRを標的としているなど、応用面でも重要度は高い。膜タンパク質の解析難度が高い理由は、試料調製の難度が高いことはもとより、結晶化にも困難が大きい。膜タンパク質の場合、疎水性の表面を界面活性剤で覆うことで可溶化させ結晶化を行うが、この際に通常可溶性タンパク質と同様に蒸気拡散法を適用すると、パッキングに寄与する部位が少ないことから強固な相互作用を形成できず、良質な結晶が得られにくい。そこで、膜タンパク質に特化した結晶化方法として脂質キュービックフェーズ(LCP)法が考案された<sup>2)</sup>。LCP法では、脂質二重膜に膜タンパク質を再構成してから結晶化を行うため、相互作用面積が大きく、良質な結晶ができやすい。このため、この10年ほどでLCP法の膜タンパク質への適用は大きく拡大してきている。

LCP法の利用が拡大した背景として、放射光におけるマイクロビームの貢献は大きい。LCP法による結晶は、良質ながらも数ミクロンから数十ミクロン程度の微小結晶に留まることが多いため、高品質な回折強度データを得るにはマイクロビームの利用が欠かせない。タンパク質結晶を用いたデータ収集を行う際は、クライオループを用いて結晶をその周囲の溶媒ごと拾い上げてX線を照射するため、結晶サイズとマッチしたX線ビームを用いることは、溶媒による散乱を抑え高S/N比の回折データを得る上で非常に重要だからである。SPring-8で2010年から運用されているBL32XUは真空封止ハイブリッドアンジュレタを光源とし、KBミラーによって試料位置で最小 $1 \times 1 \mu\text{m}^2$ (半値全幅)のビームサイズに最大 $2 \times 10^{12}$ 光子/秒という高フラックス微小ビームが利用できるマイクロビームラインである。BL32XUは初期の頃から膜タンパク質をはじめとした微小結晶を用いた構造決定に大きく貢献してきた。当初は横幅のある結晶を用い、ラインフォーカスビームによるヘリカルデータ収集によって1個の結晶から解析に必要なデータ(180-360°)を取り切る戦略が主流だった。ヘリカルデータ収集による放射線損傷の程度を予測し、適切なビーム強度を提示するKUMA(Kessho-wo Ugokashitari Mawashitari-suru Application)の開発によって、ユーザはデータ収集を簡便に行うことが

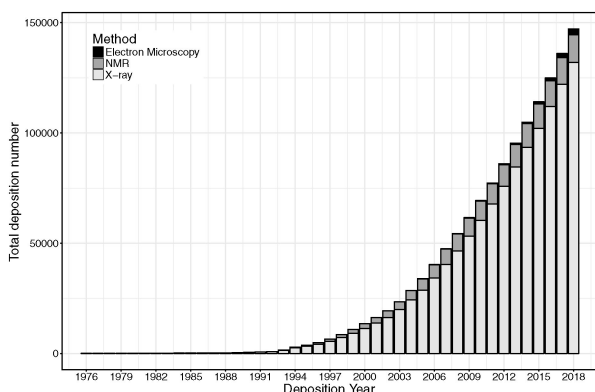
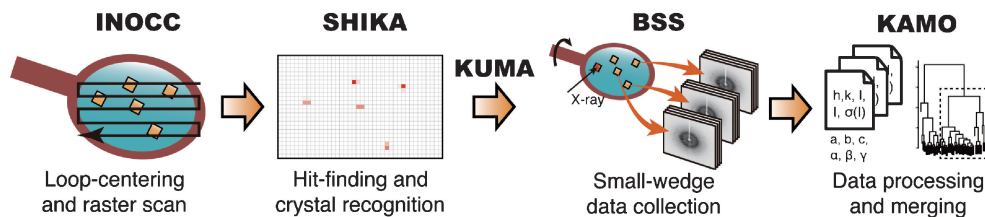


Fig. 1 The growth of the total number of PDB depositions for each major structure determination method. Plotted using data from RCSB PDB Statistics (<http://www.rcsb.org/stats/>).



**Fig. 2** (Color online) Small-wedge data collection scheme from multiple microcrystals. Crystal localization by X-ray diffraction scan, followed by data collection and processing. Slightly modified from the review<sup>5)</sup> under CC-BY 2.0 (<https://creativecommons.org/licenses/by/2.0/>).

できた<sup>3)</sup>。

マイクロビームによるヘリカルデータ収集は強力な手法だが、回折能が低く、さらに小さな各辺 5–10  $\mu\text{m}$  程度の結晶では十分なデータが取り切れない。結晶体積が小さいため、高分解能の回折強度を得ようと高強度(高光子密度)の X 線を用いると簡単に吸収線量の限界 ( $\sim 10 \text{ MGy}$ ) を超えてしまうためである。

この場合、1 結晶あたり 5–10° 程度の狭い振動範囲 (small-wedge) だけ測定し、複数の結晶に由来するデータをマージ (結合) する方法が有用であり、通常数十から数百個程度の結晶だけで十分なデータが得られる。クライオループ等の結晶ホルダに複数の結晶が収められている場合、この程度の角度領域であれば正面から見た結晶の位置を決めるだけで照射野から測定中大きくずれずにデータを収集できるので自動化においても有利である。本稿ではこの測定スキームについて述べる。

複数の微小結晶を用いたデータ収集において鍵となるのが、結晶の位置探索、放射線損傷予測に基づいた最適な露光条件の決定、得られたデータの処理である。SPring-8 BL32XU では、これらはそれぞれ SHIKA, KUMA, KAMO としてプログラム化され、これによってルーチ的にデータ収集・処理が可能になった (Fig. 2)。加えて自動ループセンタリングプログラムとして INOCC が開発され、サンプル交換ロボット SPACE の制御を含む BSS (Beamline Scheduling Software) も組み合わせて自動化 (ZOO) が達成された<sup>4)</sup>。本稿では、筆者が開発した SHIKA および KAMO について詳細に記述し、いくつかの実例を紹介する。

## 2. SHIKA による迅速な微小結晶探索

回折データを収集する際、ビームに対して結晶位置をあわせるセンタリング作業が必要になる。微小結晶、特に LCP 法によって結晶化された試料は凍結後に顕微鏡下で目視しづらいことが多く、高い位置精度も要求されるため、X 線を用いた結晶位置の決定が必須である。このとき極力損傷を起ささない程度の低線量 (光子密度  $\sim 2 \times 10^7 / \mu\text{m}^2/\text{frame}$ ) でスキャンを行う。結晶位置の確認のためには強度の大きい低分解能の回折斑点が観測できれば十分なので、低線量でも問題ない。スキャン領域はサンプルホル

ダの大きさに依存するが、多いときは 1 万箇所以上をスキャンして結晶位置を探索する。BL32XU では EIGER X 9M (DECTRIS Ltd.) を導入して通常 50 Hz の連続的読み出しでスキャンを行っている。高速に生み出される大量のイメージを即座に解析して結晶位置を提示するため、プログラム SHIKA (Spot-wo Hirote Ichiwo Kimeru Application) を開発した。EIGER の streaming interface は ZeroMQ (ØMQ) ネットワークライブラリを利用してイメージデータをネットワーク上に順次送信 (PUSH socket) しており、SHIKA はそれを複数プロセスから受け取る (PULL socket) ことによって並列処理を実現している。SHIKA の仕組みを Fig. 3 に示した。SHIKA は、SFX 実験のために開発されたプログラム Cheetah<sup>6)</sup> に含まれるピークサーチ関数を利用して、分解能 5 Å 以下の低角領域に存在する回折斑点を数えてスコアとする。SHIKA GUI は保存されたファイルを読むことで、結果を表示する。レポート HTML ファイルも生成され、ユーザは持ち帰ったデータからも Web ブラウザを利用して結果を確認できる。GUI から評価基準を選択し、しきい値・最小距離を入力することで自動的にスコアの高い座標を選択でき、KUMA へのゴニオメータ座標情報転送機能によって即座にデータ収集を構築できる。

## 3. KAMO による自動データ処理

データ処理の自動化は非常に重要である。多数の結晶を用いる場合、より多くの結晶を用いることによってより高分解能や多型の解析が可能になることが期待できる。データ処理が自動化されていないと、多量のデータを収集することに対して心理的抵抗が大きい。あるいは実験中にこれ以上データを取る必要があるかどうかを判断する上でも自動処理は重要である。そこでプログラム KAMO (Katapashikara Atsumeta data wo Manual yorimoiikanjide Okaeshisuru) を開発した<sup>7)</sup>。KAMO は、i) 個別 wedge の適切な処理 (指数付け・積分)、ii) 空間群の決定、iii) indexing ambiguity の解決 (あれば)、iv) 同型性に基づいたデータの分類、v) 異常値の除去とスケール・マージ、vi) 結果の視覚的な出力の機能を持つ。KAMO を用いた処理の流れを Fig. 4 に示した。個別データの処理とマージは XDS パッケージ<sup>8)</sup>を利用する。個別データの処理には

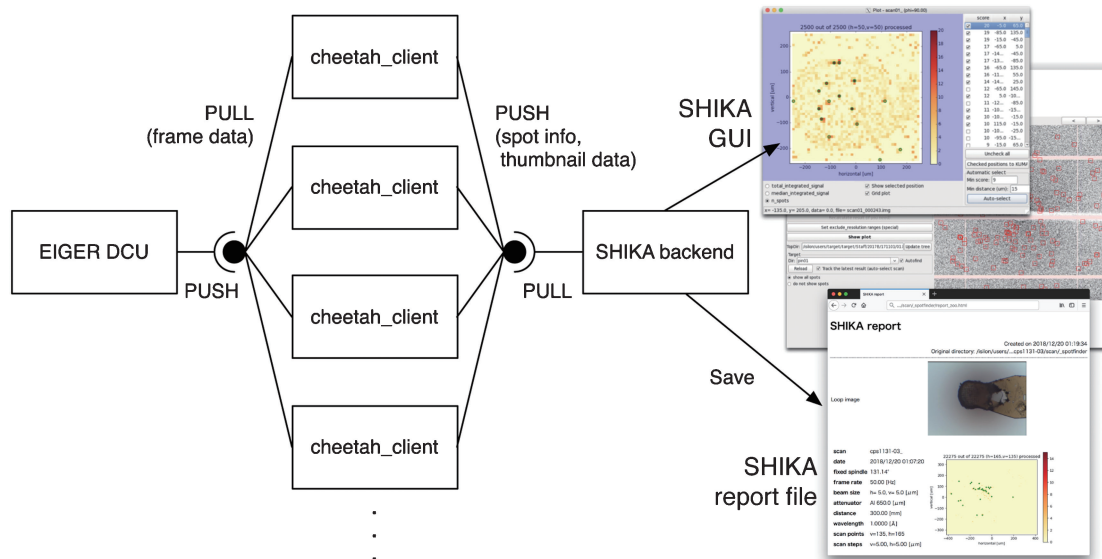


Fig. 3 (Color online) The mechanism of on-the-fly hit-finding using EIGER detector by SHIKA.

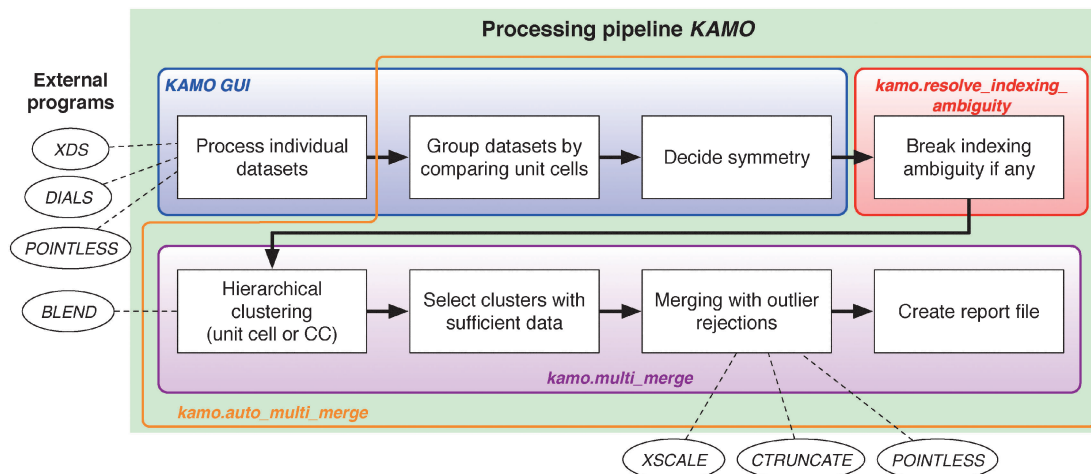


Fig. 4 (Color online) The flow of KAMO programs. Copied from the Yamashita et al. (2018)<sup>7)</sup> under CC-BY 2.0 (<https://creativecommons.org/licenses/by/2.0/>).

DIALLS<sup>9)</sup>も利用できるが、small-wedgeの処理には現時点ではXDSのほうが実績がある。CCP4パッケージ<sup>10)</sup>からはPOINTLESS<sup>11)</sup>, CTRUNCATE, BLEND<sup>12)</sup>を利用する。KAMOはもともとSPring-8ビームライン、特にZOOの一部として利用するために開発したが、現在ではその他の施設でも利用可能である。SPring-8ビームラインで利用する場合は、BSSのログを監視することでデータ収集と並行した自動処理ができるようになっている。

個別データセットの処理が完了したら、マージ処理に進む。各データセットについて、格子定数が類似しているもの同士をグループ化し、各グループに関して、可能な点群対称性がリストアップされ、この中からマージ時に仮定する対称性を選択する。ここで強度の対称性よりも格子の対称性が高い場合（たとえばP4など）、反射の指数の付け

方に幾何学的に等価な複数の可能性が生じる（indexing ambiguity）ため、実際に強度を比較して指数の付け方を合わせる作業が必要になる。KAMOでは、selective breedingアルゴリズム<sup>13)</sup>をデフォルトの解決方法として採用している。

マージの準備が整ったら、クラスタリングとマージの作業へ進む。階層的クラスタリングの基準として格子定数と相関係数の2種類を用意している。格子定数ベースにはBLEND<sup>12)</sup>のクラスタリング機能を用いる。相関係数ベースの場合は、各データ間の共通反射から相関係数（correlation coefficient: CC）を計算し、 $d(i, j) = \sqrt{1 - CC(i, j)}$ をデータセット*i, j*間の距離としてWard法で階層的クラスタリングを実行する。またデータセット間で分解能に対する強度の落ち方が大きく異なる場合、そのままCCを計算

することは不適切なので、規格化構造因子の二乗  $|E|^2$  を使うオプションも用意している。

クラスタリング計算を行ったら、各クラスタについて completeness と multiplicity を計算し、両者が高い（デフォルトではそれぞれ  $\geq 90\%$  か  $\geq 2$ ）クラスタのみ、マージ処理を行う。スケールには XSCALE を用い、その出力に基いて異常値 (outlier) を検出し、除外する。除くべきものは、マージ後の結果を悪化させる悪いフレームあるいは悪い結晶の2種類ある。これらを順次検出・除外するため、スケール・マージ計算のサイクルは最大3回行われる。まず最初に全データセットをスケール・マージし、各フレーム上の強度とマージ後の強度から相関係数を計算し、その極端に低いものを Tukey の基準を用いて検出・除外する<sup>14)</sup>。悪いフレームを除いたデータセットをマージし、そのうち特に大きな系統誤差あるいは極端な  $B$  値を持つデータセットを Tukey の基準に照らして検出・除外し、最終結果とする。結果のレポートは HTML ファイルで出力され、クラスタリング結果と統計値を視覚的に確認できるようになっている。

#### 4. 利用成果

BL32XU において、SHIKA および KAMO は既にほとんどの実験に利用されている。BL32XU 以外からの利用も含め、KAMO を用いた構造解析成果として現在27報(2017年以降)の論文が出版されている。BL32XU において本スキームを利用して最初に出版された、細胞内結晶化によって得られた多角体タンパク質の微小結晶解析では100以上収集したうち40個程度の結晶をマージして分解能  $1.55 \text{ \AA}$  で構造決定を行った<sup>15)</sup>。GPCR であるリゾホスファチジン酸受容体 LPA<sub>6</sub><sup>16)</sup>、ヒト由来オレキシン2受容体<sup>17)</sup>、ロイコトリエン B4 受容体 BLT1<sup>18)</sup> はいずれも数百個の結晶をマージし、それぞれ分解能  $3.2, 3.7, 1.96 \text{ \AA}$  で決定された。また同じく GPCR であるムスカリン M<sub>2</sub> 受容体の解析<sup>19)</sup> では、水銀置換体の結晶から収集した  $5^\circ \times 459$  のデータセットを用い、 $2.5 \text{ \AA}$  分解能において SAD 法による位相決定にも成功した。このとき32個のクライオルーブを用い、ZOO による自動測定で7時間弱をかけて671データセットを収集した。他にもトリオースリン酸/リン酸輸送体 (TPT)<sup>20)</sup>・真核生物由来 MATE トランスポーター<sup>21)</sup> などの輸送体や、鉄還元酵素 Dcytb<sup>22)</sup> の解析にも成功している。いずれも1結晶あたり  $5\text{--}10^\circ$  程度のデータを数十から数百個の結晶から収集し、構造解析に成功している。利用成果の多くは LCP 法によって結晶化された膜タンパク質だが、本稿で述べた手法の適用範囲は膜タンパク質に限らない。最初に挙げた多角体タンパク質は細胞内結晶化によって得られた可溶性タンパク質試料であるし、また、同じく可溶性タンパク質で、砕けた状態の針状結晶を含んだ1ルーブのみから構造決定に成功した例もある (unpublished)。その他にも、微小結晶でもない

が、多数のデータをマージすることによってフレキシブルなドメインの電子密度が改善した例もある (unpublished)。微小結晶はもちろんのこと、微小結晶に限らなくても、上記のようなケースでお困りの方は、利用を検討していただければ幸いである。

#### 5. おわりに

マイクロビーム技術の登場によって微小結晶を用いた構造解析が技術的に可能になり、これまで大きく良質な結晶が得られず構造解析できなかったような試料も解析対象になった。本稿で述べた small-wedge データ収集では、なるべく多くのデータを集め、適切にマージを行うことで、より高分解能かつ高精度のデータが得られることが期待される。SHIKA・KAMO を含む微小結晶のためのワークフローおよびソフトウェア環境整備と、その自動化システム ZOO の開発によって、大量の微小結晶を用いた構造解析が非常に簡易・高速になった。KAMO は GitHub にてオープンソースで公開しており (<https://github.com/keitaroyam/yamtbx>)、SPring-8 のデータに限らず利用可能である。出版済みの構造解析に関して、生データへのリンクやデータ処理の方法等も同ドキュメントに掲載しているので、興味のある方はあわせて参照頂きたい。

#### 謝辞

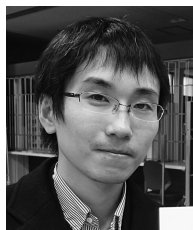
本研究は、日本医療研究開発機構 (AMED) 「創薬等支援技術基盤プラットフォーム事業」および日本学術振興会 (JSPS) 科学研究費補助金研究活動スタート支援「多数の微小結晶を用いた結晶構造解析手法および自動化システムの開発」などの支援を受けて行われました。関係各所に深く感謝します。また本研究の遂行にあたり、SPring-8 構造生物学ビームライン関係者の方々、特に理化学研究所放射光科学総合研究センターの山本雅貴部長・平田邦生専任技師には大変お世話になりました。さらに、SPring-8 BL32XU のユーザの方々には実サンプルの提供やフィードバックを通じて多大なご協力を頂きました。この場を借りてお礼申し上げます。

#### 参考文献

- 1) H. M. Berman: *Protein Science* **21**, 1587 (2012).
- 2) M. Caffrey: *Acta Cryst. F* **71**, 3 (2015).
- 3) K. Hirata *et al.*: *J. Physics: Conf. Series* **425**, 012002 (2013).
- 4) K. Hirata *et al.*: *Acta Cryst. D* **75**, 138 (2019).
- 5) M. Yamamoto *et al.*: *IUCrJ* **4**, 529 (2017).
- 6) A. Barty *et al.*: *J. Appl. Cryst.* **47**, 1118 (2014).
- 7) K. Yamashita, K. Hirata and M. Yamamoto: *Acta Cryst. D* **74**, 441 (2018).
- 8) W. Kabsch: *Acta Cryst. D* **66**, 125 (2010).
- 9) D. G. Waterman *et al.*: *CCP4 Newsletter on Protein Crystallography* **49**, 13 (2013).
- 10) M. D. Winn *et al.*: *Acta Cryst. D* **67**, 235 (2011).
- 11) P. R. Evans: *Acta Cryst. D* **67**, 282 (2011).

- 12) J. Foadi *et al.*: Acta Cryst. D **69**, 1617 (2013).
- 13) W. Kabsch: Acta Cryst. D **70**, 2204 (2014).
- 14) J. W. Tukey: Exploratory data analysis (Addison-Wesley, 1977).
- 15) S. Abe *et al.*: ACS Nano **11**, 2410 (2017).
- 16) R. Taniguchi *et al.*: Nature **548**, 356 (2017).
- 17) R. Suno *et al.*: Structure **26**, 7 (2017).
- 18) T. Hori *et al.*: Nature Chemical Biology **14**, 262 (2018).
- 19) R. Suno *et al.*: Nature Chemical Biology **14**, 1150 (2018).
- 20) Y. Lee *et al.*: Nature Plants **3**, 825 (2017).
- 21) H. Miyauchi *et al.*: Nature Communications **8**, 1633 (2017).
- 22) M. Ganasen *et al.*: Communications Biology **1**, 120 (2018).

● 著者紹介 ●



**山下恵太郎**

東京大学大学院理学系研究科 助教

E-mail: keitaro.yamashita@bs.s.u-tokyo.ac.jp

専門: X線構造生物学

**【略歴】**

2013年北海道大学大学院生命科学院博士後期課程修了, 博士(生命科学)取得。同年より理化学研究所放射光科学研究センターにて勤務。2018年より同客員・東京大学大学院理学系研究科。