

シリアルフェムト秒結晶学における異常分散を用いた位相決定

山下恵太郎¹, 岩田 想^{1,2}, 吾郷日出夫¹, 山本雅貴¹, 中津 亨^{1,3}

¹理化学研究所放射光科学総合研究センター 〒679-5148 兵庫県佐用郡佐用町光都 1-1-1

²京都大学大学院医学研究科 〒606-8501 京都府左京区吉田近衛町

³京都大学大学院薬学研究科 〒606-8501 京都府左京区吉田下阿達町46-29

■ 研究

X線自由電子レーザー (XFEL; X-ray free electron laser) は、その超短パルス・超高輝度特性によって、試料損壊前にX線回折を得ることを可能にした。シリアルフェムト秒結晶学(SFX; serial femtosecond crystallography) は、大量の微小結晶試料から次々にXFELによる回折パターンを得て構造解析を行う手法であり、無損傷かつ高分解能のタンパク質構造を常温でも得ることを可能にし、時分割測定への応用も拡がりつつある。しかしながら、SFXによる構造解析の大部分は既知の類似構造が必要な分子置換法によるものであり、新規構造の決定に応用された事例は存在しない。私たちはSACLAにおいて12.6 keVのXFELを用いて、ネイティブおよび水銀誘導体のタンパク質結晶から、異常分散を用いた同型置換法(SIRAS; single isomorphous replacement with anomalous scattering)によって効率的に位相決定が可能であることを示した。さらに、データ処理方法が改良されたことで、誘導体結晶だけを用いた単波長異常分散(SAD; single-wavelength anomalous dispersion)法による位相決定も可能になった。

1. はじめに

生体中における化学反応など様々な生命現象を司るタンパク質の機能を理解するためには、その正確な立体構造情報が欠かせない。現在主流の構造解析手法であるX線結晶回折では、質の良い結晶さえ得られれば高分解能構造解析が可能であり、結晶作製技術の向上やシンクロトロン放射光によるX線の高輝度化などにより、構造解析の対象は次第に拡大してきている。しかしながら、質の良い結晶が得られにくい試料も多く、数ミクロン程度の微小結晶や、回折能の低い結晶しか得られないことは多い。近年利用可能になったX線自由電子レーザー(XFEL; X-ray free electron laser)は、試料位置で1ミクロン角あたり 10^{10-13} 光子という大強度の超短パルス(幅数十フェムト秒)光をもたらし、試料損傷前の回折X線測定を可能にした。ただし試料はXFEL照射後に即座に損壊するため、連続的に試料を供給する必要がある。そこで、液体ストリーム等により微小結晶試料を連続的に供給するシリアルフェムト秒結晶学(SFX; serial femtosecond crystallography)が考案された¹⁾。これにより、微小結晶を用いた無損傷・高分解能構造解析の可能性が拡がり、時間分割測定への応用など、世界で多くの取り組みがなされている。

既にSFXによる様々な構造決定事例が報告されている²⁻⁵⁾が、大部分の構造解析は分子置換法によるもの、すなわち既知構造を利用した方法で位相決定がなされており、新規構造解析に応用された事例は無い。2014年に

Barendsら⁶⁾が、ガドリニウム(Gd)原子を結合させたlysozymeの構造を、SAD法によって決定できたという報告をおこなった。Lysozyme 1分子あたり2原子のGdが結合しており、非常に大きな異常分散寄与があったが、それでも6万枚もの回折パターンを必要とした。

本研究では、水溶性タンパク質であるルシフェリン再生酵素(LRE; Luciferin-regenerating enzyme)の水銀誘導体結晶を用いてSFX実験を行い、位相決定を試みた。本ケースは、先のlysozymeケースと比べて空間群の対称性は半分、異常分散シグナルの寄与率は3分の1程度であり、より現実的な系と言えるだろう。当初、収集した約86,000枚のイメージを用いてもSADによる位相決定は成功しなかったが、ネイティブデータを併用して異常分散を用いた単一同型置換法(SIRAS; single isomorphous replacement with anomalous scattering)を適用したところ、合計20,000枚のイメージで位相決定に成功し、SFXにおいて同型置換法が新規構造決定の効果的な方法であることを示した⁷⁾。その後、データ処理プログラムの性能向上により、データの質が劇的に向上し、13,000枚のイメージからSAD法による位相決定が可能となった⁸⁾。

2. SFXによる新規構造決定

A. ターゲット試料

本実験のターゲット試料として、水銀を結合させたルシフェリン再生酵素(LRE)の微小結晶を使用した。LRE

は308アミノ酸から成る水溶性タンパク質であり、システイン残基を1つだけ持つため、水銀結合部位は1つである。

LREは大腸菌発現系によって大量発現・精製され、バッチ法によって $2\text{--}5\ \mu\text{m} \times 10\text{--}30\ \mu\text{m}$ の棒状の結晶が得られた。水銀誘導体結晶は、ネイティブ結晶が安定に存在できるストック溶液に $1\ \text{mM HgO}$ を加えた溶液に6日間浸漬し、結合せずに結晶溶媒中に残った水銀を除くためにストック溶液へ1時間バックソークすることで得られた。この結晶は空間群 $P2_12_12_1$ であり、非対称単位中に1分子のLREが含まれていた。本実験の条件($12.6\ \text{keV}$)では水銀の持つ異常分散寄与は $f'' = 9.71e$ であり、Bijvoet ratio $\langle |AF| \rangle / \langle |F| \rangle$ は4.2%である。

B. SFX 実験

SACLA BL3においてSFX実験を行った。LREのネイティブ結晶および水銀誘導体結晶の両方からデータ収集を行った。サンプル消費量を抑えるため、サンプル媒体にはgrease matrix法⁹⁾を用いた。XFELの繰り返し周波数は30 Hzで、パルス幅は10 fs以下である。LRE微小結晶を混合したグリースは、湿潤ヘリウム雰囲気中のDAPHNISチャンパー¹⁰⁾中に取り付けられたシリンジから供給され、回折パターンはMPCCD検出器¹¹⁾を用いて記録された。検出器はセンサー8枚からなるshort working distance (SWD)の配置で用いた (Fig. 1)。試料とカメラ間の距離は約81 mmで、 $12.6\ \text{keV}$ (波長 $0.984\ \text{\AA}$)ではエッジで約 $1.7\ \text{\AA}$ 解能まで回折斑点が記録可能である。

C. CrystFEL 0.5.3a によるデータ処理および位相決定

まずはじめに、X線回折強度データを収集した時点

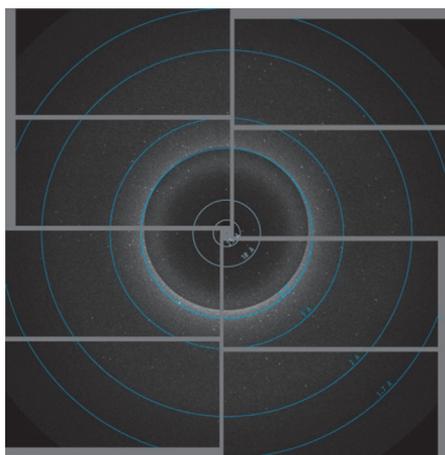


Fig. 1 A diffraction image of Hg-bound LRE microcrystal with MPCCD detector in SWD octal configuration. The rings are drawn at 25, 10, 4, 3, 2, and $1.7\ \text{\AA}$ resolutions. The low-angle X-ray absorber was used for Hg-derivative crystals to mitigate detector saturation by low resolution strong spots. The figure was prepared using hdfsee in CrystFEL package.

(2014年5月)で利用可能であったCrystFEL 0.5.3a¹²⁾によってデータ処理を行った。次小節で改良された方法による結果を紹介するが、ここでは文献⁷⁾で発表した解析を紹介する。SFXにおけるデータ処理の概要およびCrystFELに関しては、Appendixを参照されたい。

ネイティブ結晶からは133,958枚のイメージが収集され、このうち20%にあたる26,238枚が検出器の中心領域強度に基いて選択された (Low-level filter)。さらにその41%の10,792枚がCrystFELによる指数付けに成功し、Monte Carlo積分によって平均多重度222の $1.5\ \text{\AA}$ 分解能のデータが得られた。水銀誘導体結晶からは583,291枚が収集され、51%の298,061枚が選択され、さらに29%の85,747枚が指数付けに成功した。Monte Carlo積分によって平均多重度908の $1.6\ \text{\AA}$ 分解能のデータが得られた。

異常分散シグナルの評価には、ランダムに2つに分けたデータセット間で $I(hkl) - I(\bar{h}\bar{k}\bar{l})$ の相関係数を評価した CC_{ano} がしばしば使われるが、本ケースではわずか0.0453しか無く、絶望的な数値に思えた。Gd結合 lysozyme のケースでは、位相決定に成功した60,000枚の回折パターンで $CC_{\text{ano}} = 0.48$ であり、対称性によって決まる等価反射の多重度やBijvoet ratioの違いを考えても、非常に小さい数値であった。SHELXC/D/E¹³⁾を用いてSAD法による位相決定を試みたが、成功しなかった。しかしながら、異常分散差 Patterson 図を確認してみると、マージする枚数を増やすごとに明瞭になるピークが存在しており、実際これは水銀のサイトであった (Fig. 2)。2万枚以上マージした場合には、SHELXDでも正しくこの位置が決められていたが、SHELXEで解釈可能な電子密度が得られるほどデータの質は良くはなかった。

残念ながら、85,747枚の回折パターンを用いても、SAD法の適用ができるほどの質のデータには到達できなかったと結論されたが、幸いネイティブ結晶の回折データも得ていたため、重原子同型置換法の適用を試みる事ができた。SIRAS法を適用したところ、ネイティブ結晶10,792枚と水銀誘導体結晶10,000枚の回折パターンから高品質の位相決定が可能なが判明した (Table 1)。SHELXEで平均FOM (figure of merit; 位相誤差の余弦の期待値) 0.615の位相が得られ、主鎖トレースでは197残基が構築され、構築したモデルとネイティブデータの相関を表すCC値 (規格化構造因子から計算) は27%であった。この時点で得られた電子密度図は、容易に解釈可能なものであった (Fig. 3)。この位相を用いてARP/wARP¹⁷⁾で自動モデル構築を行ったところ304残基がほぼ正確に構築された。

マージする枚数を様々に変えて位相決定を試みたところ、位相決定の成功には常に合計2万枚程度は必要であった (Fig. 4)。重原子同型置換法の適用には、ネイティブ・重原子誘導体結晶間での同型性が非常に重要となるが、今回のケースでは格子定数の差異は0.2%未満であり、高

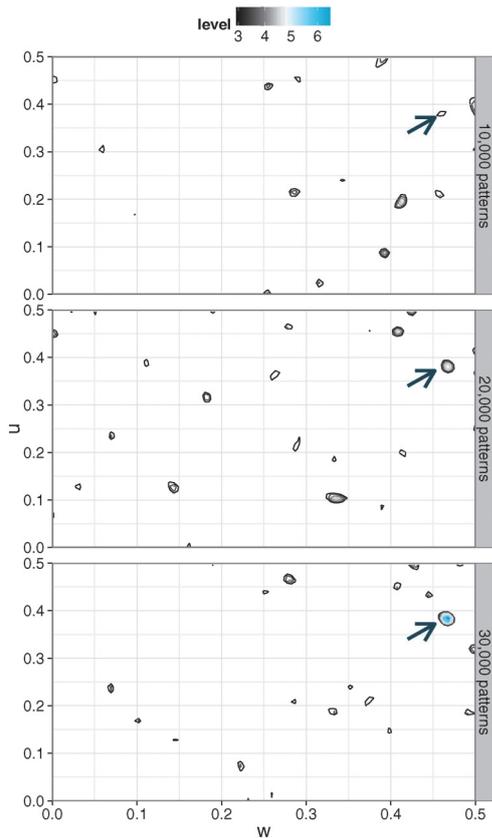


Fig. 2 Anomalous difference Patterson maps. The $\nu=1/2$ Harker section is calculated using different numbers of indexed Hg-derivative patterns. When 10,000 patterns are used, a peak of 3.2σ is observed at the Hg position; however, there are other higher false peaks. The correct Hg peaks (indicated by arrows) with heights of 5.0σ and 6.9σ using 20,000 and 30,000 patterns, respectively, are observed as the highest peaks. The Patterson maps were calculated using the CCTBX functionality¹⁴⁾ and the figure was prepared using R¹⁵⁾ with ggplot2 package¹⁶⁾.

い同型性を有していた。また、両データ間での不一致を表す $R_{\text{iso}} = \sum_{hkl} |F_{\text{deriv}}| - |F_{\text{native}}| / \sum_{hkl} |F_{\text{native}}|$ (Scaleit²¹⁾ による計算) は23.0%であった。

D. CrystFEL 0.6.1を用いた再解析

2015年8月に CrystFEL 0.6.1が発表され、多くの新機能が利用可能になった²³⁾。先の解析で用いていた0.5.3a (2014年4月に発表) と比べると、以下のような機能が追加されている。i) prediction refinement (回折パターンごとに結晶方位行列と検出器の水平面内シフトを精密化), ii) geoptimiser²⁴⁾ (各センサーの幾何学配置: 回転および位置を精密化), iii) profile_radius (逆格子点半径)

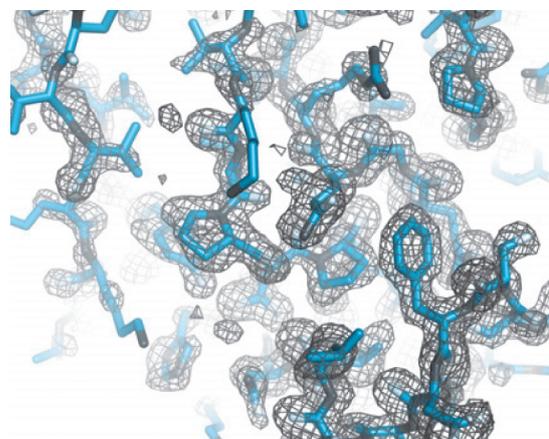


Fig. 3 The SIRAS electron density map and a refined model of LRE. For the calculation, 10,000 Hg-derivative and 10,792 native diffraction patterns were used. The SIRAS electron density map produced by SHELXE is contoured at 1.0σ . The model was manually and automatically refined using Coot¹⁸⁾ and phenix.refine¹⁹⁾ after automated model building by ARP/wARP¹⁷⁾. The figure was prepared using PyMOL²⁰⁾.

Table 1 Data collection statistics.

	Native	Hg-derivative	
	10,792 patterns (CrystFEL 0.5.3a)	10,000 patterns (CrystFEL 0.5.3a)	13,000 patterns (CrystFEL 0.6.1)
Wavelength (Å)	0.981	0.984	
Beam energy	30.0 μJ/pulse	86.4 μJ/pulse	
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a, b, c ; Å)	48.2, 77.6, 84.8	48.1, 77.5, 84.8	48.1, 77.5, 84.8
Resolution range† (Å)	10–1.50 (1.56–1.50)	25–1.60 (1.66–1.60)	25–1.60 (1.66–1.60)
Completeness† (%)	100 (100)	100 (100)	100.0 (99.98)
SFX multiplicity†	222.2 (196)	106.2 (44)	62.3 (13)
R_{split}^\dagger	0.2727 (1.671)	0.3727 (5.772)	0.2238 (1.457)
$\langle I/\sigma(I) \rangle^\dagger$	2.6 (0.63)	1.8 (0.19)	3.3 (0.83)
$CC_{1/2}^\dagger$	0.893 (0.1988)	0.835 (0.0279)	0.935 (0.2356)
CC_{ano}	n/a	-0.066	0.020

† Values in parenthesis are for the highest resolution shell.

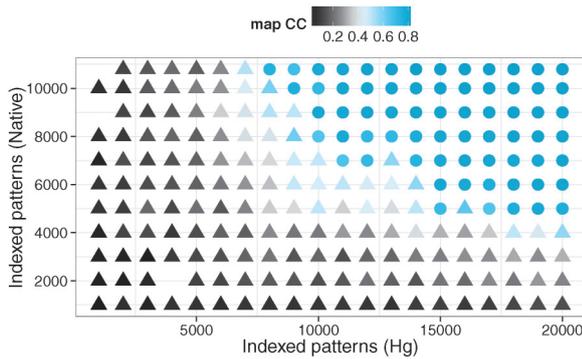
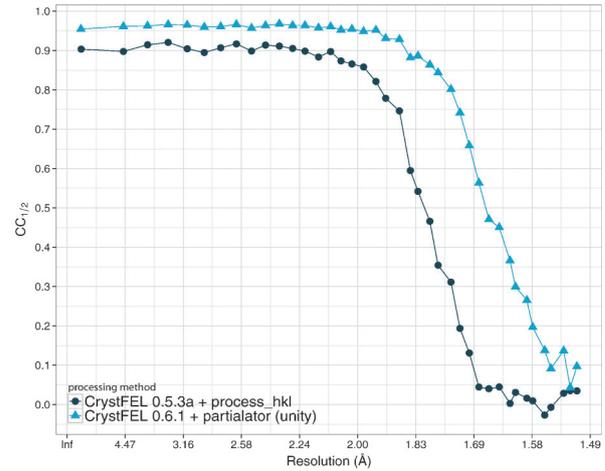
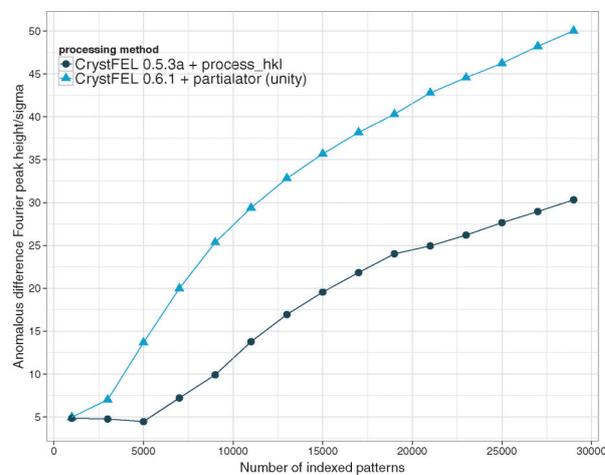


Fig. 4 Phase quality (map CC) for the different numbers of the native and the derivative indexed patterns. The success and failure of phasing are represented as circular and triangular symbols, respectively. The map CC is defined as the correlation coefficient between the SIRAS electron density map produced by SHELXE and the F_c map calculated using the final refined structure. The result was judged as a success if the map CC was greater than 0.65. The map CC was calculated using phenix.get_cc_mtz_pdb²²⁾ and the figure was prepared using R¹⁵⁾ with ggplot2 package¹⁶⁾.

の自動決定, iv) 安定して動作するようになった partialator²⁵⁾ (分解能に伴う強度減衰を表す B 値を含めた スケーリング, あるいは partiality 補正込みのスケールリングおよび post-refinement)。LREの水銀誘導体データを CrystFEL 0.6.1を用いて再解析を行ったところ, 大きな改善が見られた。指数付け枚数は同程度だが, i) および ii) は, prediction (回折斑点の出現予測位置)の精度を向上させ, 劇的なデータの質的向上をもたらした。更に iv) は, partiality 補正の性能は (少なくとも本ケースに関しては) まだ実用的とは言えなかったが, partiality 補正無し のスケールリングでは, これまでの方法と比べて改善があった。これまで単純な Monte Carlo 積分と, フレームスケールリングを行ってから Monte Carlo 積分を行う 2 通りがあったが, 前者だと低角, 後者だと高角のデータが悪化することが多かった。Partialator による B 値を含めたスケールリングを, 回折パターンごとの分解能カットオフとともに用いることで, 高-低分解能までの範囲でデータの質 ($CC_{1/2}$; Appendix 参照) を保たせることができた (Fig. 5a)。同時に, 異常分散差データの質も大きく改善した。例えば, 決定した構造の位相を用いて計算した異常分散差 Fourier ピーク高さは, 同じ枚数の回折パターン数でも再解析によって大きく向上している (Fig. 5b)。再解析後のデータを用いて SAD 法による位相決定を試みたところ, 13,000枚の回折パターンから位相決定に成功した (Table 1)。位相決定には, 前小節同様に SHELX C/D/Eを用いた。SHELXEは前小節で用いたバージョン (2014/2) の期限が切れたため, 最新版 (2014/4) を用いている。また, 本再解析によって, SIRASに必要な枚数も合計7,000枚にまで減少することが確認された。



(a)



(b)

Fig. 5 Comparison of data quality with two CrystFEL versions. In CrystFEL 0.5.3a, process_hkl (the simple Monte Carlo integration) was used for merging. In CrystFEL-0.6.1, partialator with unity option (Monte Carlo integration after scaling patterns with linear scale and relative B -factors and with by-pattern resolution cutoffs) was used. (a) Comparison of $CC_{1/2}$ by resolution. In CrystFEL 0.5.3a 34,393 patterns were merged while 35,231 patterns were merged in CrystFEL 0.6.1. (b) Comparison of anomalous difference Fourier peak heights (calculated with ANODE²⁶⁾) at the mercury site by the number of indexed patterns. These figures were prepared using the ggplot2¹⁶⁾ in R¹⁵⁾.

ただし, 本再解析の結果でも CC_{ano} は低く, 13,000枚で 0.02であった。 CC_{ano} はランダムに 2 分したデータ間の相関を見ており, 再解析でもデータ精度 (precision) は低いままであることを示している。一方, 異常分散差 Fourier のピーク高さが大きく改善していることから異常分散データの確度 (accuracy) が改善している, すなわち真値に接近していることは明らかである。本結果はデータ処理方法の改善によってデータ確度が向上したことを示してい

る。元々 CC_{ano} がゼロ付近であったときから異常分散 Patterson のピークは観察され、異常分散 Fourier ピークも有意にあったことから、 CC_{ano} は少なくとも SFX における異常分散シグナルの指標としては適切でない可能性があると言えるだろう。

3. 国内外における SFX による新規構造決定の取り組み

前述のとおり、Barends らは Gd 結合 lysozyme を用いた SAD 法による位相決定を成功させ、SFX においても新規構造決定が可能であることを示した。彼らは米 Stanford にある LCLS の CXI ステーションで実験を行い、8.5 keV の X 線を用いた。当初 60,000 枚の回折パターンが位相決定には必要という結果だったが、データ処理の改善により、現在では 7,000 枚でも位相決定が可能になっている²⁷⁾。

SACLA における SFX では、本稿で解説した LRE の他、非標識 lysozyme の硫黄および塩化物イオンを用いた SAD 法による位相決定の成功が中根らによって報告された²⁸⁾。さらに福田らによって、天然状態で Cu が結合した *A. faecalis* 由来亜硝酸還元酵素 Af NiR (nitrite reductase) も SAD 法による位相決定に成功したとの報告もあった²⁹⁾。また LCLS では、Nass らから thaumatin の S-SAD に成功したとの報告があった²⁷⁾。

Table 2 に、現在までに報告されたケースをまとめた。Bijvoet ratio と空間群の対称性によって必要になる枚数は変化するが、Bijvoet ratio の低いケースでも許容範囲と思われる程度の枚数で SAD 法による位相決定に成功している。この表にある全てのケースでは、CrystFEL によってデータ処理が行われ、Monte Carlo 積分によってマージされたデータが使用されている。

4. まとめ

12.6 keV という SACLA の得意とする高エネルギー XFEL を利用して、微小結晶を用いたタンパク質の新規立体構造決定が可能であることが示された。データ処理方法の改善によって SAD 法も成功に至ったが、SIRAS はより効率的に構造決定できる可能性があるため、現実のケースでは可能であればネイティブデータも収集しておくことが望ましいだろう。本稿では紹介できなかったが、SACLA では他にも SeMet 置換体を利用した SAD 法による位相決定や、その他の原子種を用いた新規構造決定の取り組みがなされている。今回のケースでは 1.7 Å 程度の高分解能データが用いられており、今後はより低分解能のケースでも検証していく必要がある。今後もデータ処理プログラムの性能向上によって、より少ないデータ量から位相決定可能な高品質なデータが得られるようになるだろう。LRE の回折パターンは、XFEL の実験データデータバンク CXIDB ID 31 (<http://cxidb.org/id-31.html>) から誰でも自由にアクセスができるようになっている。CXIDB では他の SFX 実験の生データも多く利用可能になっているので、興味のある方はデータ処理に挑戦してみたい。

5. APPENDICES

A. タンパク質結晶における位相決定

X 線回折では、電子密度分布の Fourier 変換である構造因子の位相情報を直接観測することができない。この問題は位相問題と呼ばれ、結晶学の中心的な問題の一つである。タンパク質結晶学では、構成原子数が非常に多く、また分解能も典型的には 2-3 Å 程度であるため、直接法の適用は通常困難である。現在主流の位相決定法は、第一に分子置換法であり、既知の類似構造を用いる方法である。類似構造が存在せず分子置換法の適用が困難な場合、つまり新規構造決定の場合は、重原子を用いる方法が一般的で

Table 2 Published experimental phasing cases with SFX.

Sample	Method	Facility	Energy (keV)	Element	Space group	Required patterns	Bijvoet ratio	Year
Lysozyme	SAD	LCLS	8.5	Gd	$P4_32_12$	60,000	11.9%	2014 ⁶⁾
						7,000 [†]		2016 ²⁷⁾
LRE	SIRAS	SACLA	12.6	Hg	$P2_12_12_1$	20,000 [*]	4.2%	2015 ⁷⁾
						7,000 ^{*,†}		2016
Lysozyme	SAD	SACLA	7	S, Cl	$P4_32_12$	150,000	1.6%	2015 ²⁸⁾
AfNiR	SAD	SACLA	10.8	Cu	$P2_12_12_1$	156,000 [§]	1.7%	2016 ²⁹⁾
Thaumatin	SAD	LCLS	6	S	$P4_12_12$	125,000	2.1%	2016 ²⁷⁾
LRE	SAD	SACLA	12.6	Hg	$P2_12_12_1$	13,000	4.2%	2016 ⁸⁾

[†] Updated number by reprocessing the same data since the first publication. ^{*} The total number (native and derivative) of indexed patterns.

[§] Not necessarily the minimum required number, which was not studied in the reference.

ある。重原子を用いる方法には、ネイティブタンパク質結晶および特定位置に重原子を導入した誘導体結晶を用いる単一同型置換法 (SIR; single isomorphous replacement), ネイティブ結晶および2種以上の誘導体結晶を用いる多重同型置換法 (MIR; multiple isomorphous replacement), これらに異常散乱シグナルの利用を加えた SIRAS および MIRAS (SIR/MIR with anomalous scattering), 1種の誘導体結晶のみを用いた単波長異常分散法 (SAD) 法, 複数の波長のデータを用いる多波長異常分散法 (MAD; multiple-wavelength anomalous dispersion) 法がある (Table 3)。

異常分散とは、原子核に強く束縛された内殻電子が入射 X 線のエネルギー依存的に起こす共鳴散乱のことである。原子散乱因子 f_{atom} のうち、構成電子を自由電子とみなした場合の散乱因子 f^0 に対する補正項として現れ、実部を f' 、虚部を f'' と呼ぶ。これらは内殻電子に起因するため一般に回折角 (分解能) には非依存である。

$$f_{\text{atom}} = f^0 (\sin \theta / \lambda) + f'(\lambda) + if''(\lambda) \quad (1)$$

タンパク質結晶学において通常使用する波長範囲 (~1.0–2.0 Å) の場合、タンパク質を構成する軽元素では硫黄を除いて異常分散効果は無視できるほど小さく、位相決定に用いることはできない。一方、たとえばセレンは K 吸収端を波長 0.9793 Å 付近に持ち、有意な異常分散効果 ($f'' \approx 3.85e$) を有するため、セレンメチオニン置換体タンパク質を大腸菌発現系などで大量発現させ結晶化する方法が近年広く用いられている。タンパク質結晶学において長い歴史を持つ水銀は、波長 1.0 Å においても高い異常分散効果 ($f'' \approx 10.0e$) を有し、しばしばシステイン残基に結合することが知られている。シンクロトロン放射光におけるタンパク質結晶学では波長 1.0 Å 程度が最も一般的に使われているため、異常分散効果を用いた位相決定では、この波長付近で高い異常分散効果を持つ元素の利用が一般的である。

歴史的には同型置換法 (SIR, MIR, SIRAS, MIRAS) が多く使われてきたが、シンクロトロン放射光の利用により波長の変更が可能になったため MAD 法の利用が広がった。さらに、データの質の向上や、位相決定ソフトウェア

の性能改善によって、単一の重原子誘導体結晶から単一波長のデータのみで位相決定が可能な SAD 法も十分に実用性が高まり、現在では分子置換法に次いで最も一般的な方法となっている。ただし、それでも SAD 法で求められるデータの質のハードルは高い。SAD 法では、異常分散差 $\Delta_{\text{ano}}(hkl) = |F(hkl)| - |F(\bar{h}\bar{k}\bar{l})|$ を正確に測定する必要があるが、この値は $|F(hkl)|$ に対して一般的に小さい。Bijvoet ratio $\langle |\Delta_{\text{ano}}| \rangle / \langle |F| \rangle$ は近似的に以下のように計算できる³⁰⁾。

$$\frac{\langle |\Delta_{\text{ano}}| \rangle}{\langle |F| \rangle} = \sqrt{2} \sqrt{\frac{N_A f_A''}{N_P Z_{\text{eff}}}} \quad (2)$$

この数値は一般には 5% 程度以下であり、このようなわずかな差を正確に決定できるような高精度な測定が必要となる (ここで N_A は異常散乱原子の数, N_P はタンパク質原子の数, Z_{eff} はタンパク質の平均原子番号である)。一方、同型置換法ではネイティブ結晶と重原子誘導体結晶間の差を用い、この差 $\Delta_{\text{iso}}(hkl) = |F_{\text{deriv}}(hkl)| - |F_{\text{native}}(hkl)|$ は近似的に以下のように計算できる³¹⁾。

$$\frac{\langle |\Delta_{\text{iso}}| \rangle}{\langle |F| \rangle} = \begin{cases} \frac{1}{\sqrt{2}} \sqrt{\frac{N_A Z_A}{N_P Z_{\text{eff}}}} & (\text{acentrics}) \\ \sqrt{\frac{N_A Z_A}{N_P Z_{\text{eff}}}} & (\text{centrics}) \end{cases} \quad (3)$$

この差は 10% 程度以上あることが多く、検出は比較的容易である。ただし、両結晶間で同型性が維持されていることが大前提である。すなわち、格子定数の差が無く、構造の差異も重原子の有無のみであることが理想的な条件である。SIRAS/MIRAS 法では、この Δ_{iso} に加えて Δ_{ano} も同時に利用する。したがって、同型性さえ維持されていれば、SIRAS 法は SAD 法に対して大きなシグナルが使えるという優位性を持つ。

同型置換法および SAD/MAD 法では、以下のような共通の手続きを取る。i) まず重原子のみの部分構造の構造振幅を推定し、ii) Patterson 法や直接法などにより重原子座標を決定、iii) 部分構造の構造因子から全体構造の位相を計算、iv) 密度改良法により位相を改良して解釈可能な電子密度図を得る。代表的な位相決定プログラムである SHELXC/D/E は、SHELXC が i) を、SHELXD が ii) を、SHELXE が iii) および iv) を行う。位相決定の成功には 2 つのハードルがあり、まず ii) において重原子が正確に見つかること、そして iv) でモデル構築が可能な質の電子密度図を得ることである。これらの点においては現在でもプログラムの性能向上が続けられており、SHELXE は従来の密度改良に加えてタンパク質の主鎖構造のトレースを行うサイクルを繰り返すことで位相改良の性能を高めている¹³⁾。

Table 3 Experimental phasing methods using heavy atom derivative(s).

Method	Crystals	Wavelength
SIR	native & derivative	any
MIR	native & multiple derivatives	any
SIRAS	native & derivative	that gives large f''
MIRAS	native & multiple derivatives	that gives large f''
SAD	a derivative	that gives large f''
MAD	a derivative	>1 varying f' and f''

B. SFX におけるデータ処理

現在、シンクロトロン放射光または実験室系 X 線発生装置におけるタンパク質結晶学では、単色光を用いた振動写真法によるデータ収集が最も一般的である。すなわちゴニオメータを用いて結晶を任意の軸まわりに回転させながら X 線を露光し、2 次元 X 線検出器を用いて連続的に回折パターンを収集する。結晶を回転させながらイメージを収集することで、逆格子点体積を積分した強度を実験的に得ることができる。放射線損傷の問題が存在するが、数十 μm 程度以上の大きさがあれば、構造解析可能な完全なデータを 1 つの結晶から得ることが可能である。

一方、XFEL におけるデータ収集はこれと大きく異なる。数~数十 fs の超短パルス光は、その照射直後に試料とその周囲に物理的損傷を引き起こし、また、フェムト秒オーダーの時間で結晶を回転させることも不可能であるから、振動写真法の適用はできない。結晶は XFEL 露光中にはほぼ静止状態であるので、得られる回折パターンは静止 (still) 写真と呼ばれるものになる。自己増幅自発放射 (SASE; self-amplified spontaneous emission) による XFEL は $\Delta E/E \approx 0.5\%$ のエネルギー幅を持つため、Ewald 球に厚みを生じ、静止写真上に得られる強度は逆格子点のうち一定の有限体積が積分されたもの (つまり部分強度のみの観測) となる。SFX ではすべての回折写真は別々の結晶に由来しており、露光体積や結晶方位、さらには XFEL の強度やエネルギー分布などが回折写真ごとに変化していることになる。これらの独立な測定結果から、構造解析に用いる構造振幅を得るためにはどうすれば良いか。理想的には個々の部分反射強度から完全反射強度を補完し、適切な結晶間スケール・強度補正を経て、マージするべきであろう。しかしこのような方法開発はまだ発展途上であり、現在最もロバストな方法は、Monte Carlo 積分によるもの、つまり各 hkl に関して、全ての部分反射強度を平均化したものを構造振幅の二乗として扱うことである^{32,33}。

CrystFEL^{12,23}は SFX のデータ処理のために新規開発されたソフトウェア・パッケージであり、現在 ver. 0.6.2 が最新である。大量の入力画像に対して、ピークサーチを行い、DirAx³⁴や MOSFLM³⁵などの既存のプログラムあるいは DirAx アルゴリズムの再実装である asdf を使用して指数付けを行い、Bragg 反射強度の積分およびバックグラウンド減算を行う。これを担うのが indexamajig と呼ばれるプログラムであり、その結果は stream ファイルと呼ばれる形式で保存される。Monte Carlo 積分によるマージは process_hkl で行われる。このとき、一度マージした強度に対して各フレームの線形スケール因子を決め、それを適用して再度マージするオプション (--scale) も存在する。Ver. 0.6.1 から実用可能となった partialator は、process_hkl の上位互換で、線形スケールに加えて B 値によるスケール (--model=unity) や、さらに結晶方位

や格子定数の精密化 (post-refinement) を通じて観測強度の部分度を計算し完全反射強度へのスケールリング (--model=scsphere) も可能である。いずれのプログラムでも、偏光因子に対する補正は行われる。

データの質の判断には $CC_{1/2}$ や R_{split} がよく用いられる。これらは、得られた回折パターンのセットをランダムまたは交互に 2 つのセットに分け、それぞれの中でマージし、2 セット間でマージされた強度を比較するものである。 $CC_{1/2}$ は両者の相関係数であり、 R_{split} は以下の式で計算される¹²。 R_{split} はマージ後のデータ精度を表す R 値であり、 R_{pim} と類似した統計的性質を持つ。

$$R_{\text{split}} = \frac{1}{\sqrt{2}} \frac{\sum_{hkl} |\langle I \rangle_{\text{set1}} - \langle I \rangle_{\text{set2}}|}{\sum_{hkl} (\langle I \rangle_{\text{set1}} + \langle I \rangle_{\text{set2}}) / 2} \quad (4)$$

振動写真法では $CC_{1/2}$ の他にも $\langle \langle I \rangle / \sigma(\langle I \rangle) \rangle$ がよく用いられるが、CrystFEL の算出する $\sigma(\langle I \rangle)$ は測定多重度が低い場合に過小又は過大評価されることがあり¹²、振動写真法の場合と単純に比較はできない。

謝辞

本研究は理研 RSC, JASRI, 京都大学, 大阪大学, キッコーマン株式会社のグループによる共同研究で行われました。特に結晶試料調製については潘東青博士・村井智洋さんをはじめとする方々にご尽力頂きました。また、本研究は JST/X 線自由電子レーザー施設重点戦略課題推進事業「創薬ターゲット蛋白質の迅速構造解析法の開発」および JSPS 科研費 15K14941 の支援を受け実施したものです。本実験にご協力頂いた JASRI, RSC エンジニアリングチームの皆様を始めとする関係各位に深く感謝申し上げます。本稿に関する実験は SACLA 利用実験課題 (課題番号: 2014A8032 および 2013B8045) でデータ収集を行った成果です。本稿における計算解析は SACLA に併設された HPC システムおよびミニ京スーパーコンピュータシステムを利用して実行しました。また本稿執筆にあたり、原稿を丁寧にチェックして頂いた東京大学の中根研究員に感謝いたします。

参考文献

- 1) H. N. Chapman *et al.*: Nature **470**, 73 (2011).
- 2) L. Redecke *et al.*: Science **339**, 227 (2013).
- 3) C. Kupitz *et al.*: Nature **513**, 261 (2014).
- 4) J. Tenboer *et al.*: Science **346**, 1242 (2014).
- 5) Y. Kang *et al.*: Nature **523**, 561 (2015).
- 6) T. R. M. Barends *et al.*: Nature **505**, 244 (2014).
- 7) K. Yamashita *et al.*: Sci. Rep. **5**, 14017 (2015).
- 8) K. Yamashita *et al.*: submitted (2016).
- 9) M. Sugahara *et al.*: Nat. Meth. **12**, 61 (2015).
- 10) K. Tono *et al.*: J. Synchrotron Rad. **22**, 532 (2015).
- 11) T. Kameshima *et al.*: Rev. Sci. Instrum. **85**, 033110 (2014).
- 12) T. A. White *et al.*: J. Appl. Cryst. **45**, 335 (2012).

- 13) G. M. Sheldrick: *Acta Crystallogr. D* **66**, 479 (2010).
- 14) R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty and P. D. Adams: *J. Appl. Cryst.* **35**, 13 126 (2002).
- 15) R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2008).
- 16) H. Wickham: *ggplot2: elegant graphics for data analysis* (Springer New York, 2009).
- 17) G. G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis: *Nat. Protoc.* **3**, 1171 (2008).
- 18) P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan: *Acta Crystallogr. D* **66**, 486 (2010).
- 19) P. V. Afonine *et al.*: *Acta Crystallogr. D* **68**, 352 (2012).
- 20) Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.3r1 (2010).
- 21) P. L. Howell and G. D. Smith: *J. Appl. Cryst.* **25**, 81 (1992).
- 22) P. D. Adams *et al.*: *Acta Crystallogr. D* **66**, 213 (2010).
- 23) T. A. White *et al.*: *J. Appl. Cryst.* **49**, 680 (2016).
- 24) O. Yefanov *et al.*: *Opt. Express* **23**, 28459 (2015).
- 25) T. A. White: *Philosophical Transactions of the Royal Society B: Biological Sciences* **369** (2014).
- 26) A. Thorn and G. M. Sheldrick: *J. Appl. Cryst.* **44**, 1285 (2011).
- 27) K. Nass *et al.*: *IUCrJ* **3** (2016).
- 28) T. Nakane *et al.*: *Acta Crystallogr. D* **71** (2015).
- 29) Y. Fukuda *et al.*: *Proc. Natl. Acad. Sci. USA* **113**, 2928–2933 (2016).
- 30) W. A. Hendrickson and M. M. Teeter: *Nature* **290**, 107 (1981).
- 31) F. H. C. Crick and B. S. Magdoff: *Acta Crystallogr.* **9**, 901 (1956).
- 32) R. A. Kirian *et al.*: *Opt. Exp.* **18**, 5713 (2010).
- 33) R. A. Kirian *et al.*: *Acta Crystallogr. A* **67**, 131 (2011).
- 34) A. J. M. Duisenberg: *J. Appl. Cryst.* **25**, 92 (1992).
- 35) A. G. W. Leslie and H. R. Powell: *Processing diffraction data with mosfm*, vol. **245**, 41 (Springer 5 Netherlands, 2007).

著者紹介



山下 恵太郎

国立研究開発法人理化学研究所 放射光科学総合研究センター 利用システム開発研究部門 ビームライン基盤研究部 生命系放射光利用システム開発ユニット 基礎科学特別研究員

E-mail: k.yamashita@spring8.or.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)。同年, 理化学研究所に入所, 上記所属にて特別研究員。2015年より現職。



吾郷 日出夫

国立研究開発法人理化学研究所 放射光科学総合研究センター 利用システム開発研究部門 ビームライン基盤研究部 生命系放射光利用システム開発ユニット 専任研究員

E-mail: ago@spring8.or.jp

専門: 放射光構造生物学

【略歴】

1991年大阪大学大学院理学研究科博士課程前期修了。日本たばこ産業株式会社, 理化学研究所研究員, 前任研究員を経て2013年より現職。この間1995年博士(理学)(東京工業大学)。



岩田 想

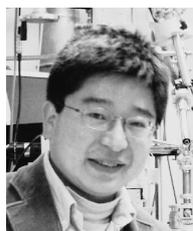
国立研究開発法人理化学研究所 放射光科学総合研究センター SACLAL利用技術開拓グループ グループディレクター

E-mail: s.iwata@mfour.med.kyoto-u.ac.jp

専門: 膜タンパク質の構造生物学, X線結晶学

【略歴】

1991年東京大学大学院農学系研究科博士課程修了, 博士(農学)。2000年より英国インペリアルカレッジ生命科学科教授(~2015年)。理化学研究所ゲノム科学総合研究センター客員主管研究員を経て2012年より現職。2007年より京都大学大学院医学研究科教授(兼任)。



山本 雅貴

国立研究開発法人理化学研究所 放射光科学総合研究センター 利用システム開発研究部門長(兼)ビームライン基盤研究部長

E-mail: yamamoto@riken.jp

専門: 放射光構造生物学

【略歴】

1991年大阪大学大学院理学研究科博士後期課程修了, 博士(理学)。同年, 理化学研究所に入所, 2011年より利用システム開発研究部門長。兵庫県立大学大学院客員教授および神戸大学連携創造本部客員教授を兼任。



中津 亨

京都大学大学院薬学研究科 薬科学専攻 准教授

E-mail: nakatsu@pharm.kyoto-u.ac.jp

専門: タンパク質結晶学, 構造生物学

【略歴】

1997年京都大学農学研究科博士後期課程修了, 博士(農学)。京都大学化学研究所教務職員, 理化学研究所播磨研究所連携研究員, 京都大学大学院薬学研究科助教授を経て2007年より現職。

Experimental phasing with serial femtosecond crystallography using anomalous dispersion

Keitaro YAMASHITA¹, So IWATA^{1,2}, Hideo AGO¹,
Masaki YAMAMOTO¹, Toru NAKATSU^{1,3}

¹RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan,

²Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan,

³Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimoadachi-cho, Sakyo-ku Kyoto 606-8501, Japan

Abstract XFEL, X-ray free electron laser, has enabled the X-ray diffraction data collection before the sample destruction due to its super-short pulse and ultra-high brilliance. Serial femtosecond crystallography (SFX), which successively collects XFEL diffraction data from many micro-crystals, has opened a new door for damage-free, room-temperature, and high-resolution protein crystallography using micro-crystals. However, most of structures determined by SFX were phased using molecular replacement, which requires a previously solved similar structure, and SFX has never been applied to the proteins whose structures were previously unknown. Using native and mercury derivative crystals and 12.6 keV XFELs at SACLA, we showed single isomorphous replacement with anomalous scattering (SIRAS) effectively worked for *de novo* phasing with SFX. Moreover, the improved data processing enabled single-wavelength anomalous dispersion (SAD) phasing using mercury derivative crystals only.