

# XFEL による生体高分子単粒子の 高分解能構造解析法にむけて

徳久淳師

独立行政法人理化学研究所 播磨研究所 放射光科学総合研究センター

XFEL 研究開発部門 ビームライン研究開発グループ データ処理系開発チーム

〒679-5148 兵庫県佐用郡佐用町光都1丁目1番1号

## 要旨

X線自由電子レーザー (XFEL) による生体高分子単粒子の高分解能構造解析法にむけた、筆者らの最近の研究を関連する研究と共に紹介する。この方法では、XFEL により発振される高輝度フェムト秒 X 線レーザーを分子方位が未知の単粒子試料に照射することで多数の 2 次元コヒーレント X 線回折強度像を観測する。開発したアルゴリズムを用いると、原子分解能の 3 次元構造を構築するために必要な信号を、高い量子雑音に埋もれた回折強度像から取り出すことが可能となる。高分解能解析に重要な役割を果たす、多数の回折強度像を分子方位の類似性に応じて分類するための方法など、開発したアルゴリズムとそれを京コンピュータへ実装する努力を合わせて報告する。

## 1. はじめに

筆者らはいま、生命の担い手である生体高分子の形を「見る」ための新たな手法の開発を進めている。2012年3月に日本の X 線自由電子レーザー (XFEL) 施設である SACLA が供用を開始した。この新しい光源からは、大強度、超短パルス X 線レーザーが得られる。高い空間コヒーレンスを持つ強い光を短時間露光することで、試料が壊れる前の構造を観測すること<sup>1)</sup>も夢ではなくなった。結晶化の必要がない生体高分子単粒子試料を対象とする構造解析法の可能性が大きく開かれた。

この新しい解析法を実現する上では、まさに試料が単粒子であることに起因する様々な問題が存在する。その筆頭は、試料からの回折光が弱いことである。X 線結晶解析に用いられる試料には、微結晶でも  $10^{12}$  個程度の分子が含まれている。それが今の場合 1 分子、あるいは単粒子なのである。試料の散乱能が小さいことを補うために強い入射光強度のコヒーレント X 線をもちいて観測を行う必要があるが、高強度 X 線は試料の放射線損傷を引き起こす。放射線損傷をできるだけ抑えるには、観測にもちいる入射光強度を可能な限り抑えることが求められる。この二律背反の関係を満たすためには、極めて微弱な光を検出することになり、検出光の統計的変動由来の雑音である量子雑音を伴う。このような回折光から、構造解析に必要な信号を検出するデータ処理技術の開発が求められる。

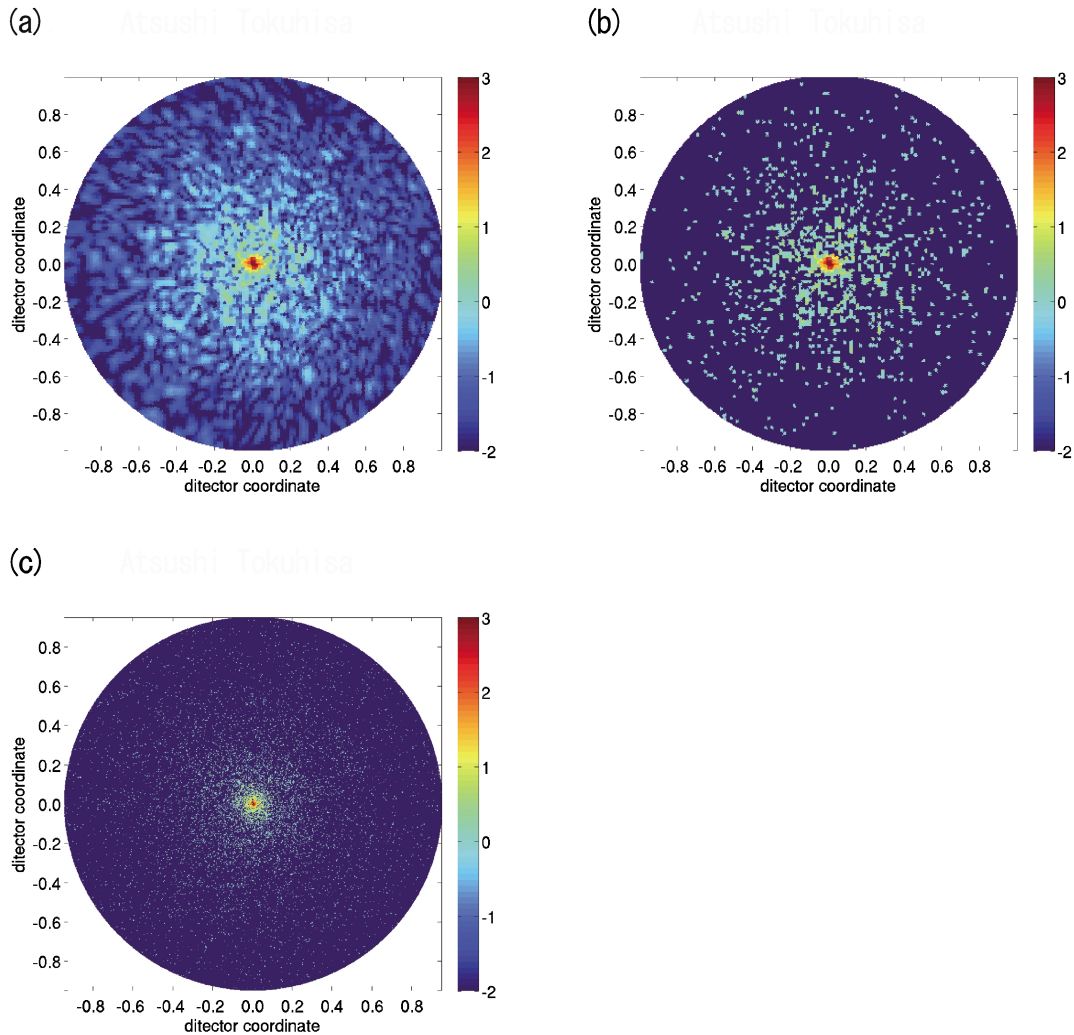
コヒーレント X 線回折法で観測されるのは 2 次元の回折強度である。これは、入射 X 線に対する試料の分子方位に応じて定まる特定の Ewald 球面上の回折強度に対応

する。回折強度は分子構造の表現である 3 次元電子密度関数のフーリエ振幅であるが、その位相情報は失われている。分子の 3 次元構造を構築するには、入射 X 線に対する分子方位を変えた測定を何度も行い、対応する Ewald 球面が波数空間を密に埋めるようにする必要がある。

Fig. 1 には、単粒子コヒーレント X 線回折強度像の、放射線損傷を無視したシミュレーション像を示している。

Fig. 1(a-b) は加水分解酵素であるリゾチーム (分子径約 60 Å) より求め、Fig. 1(c) はタンパク質合成に関わるリボソーム (分子径約 270 Å) より求めた。

Fig. 1(a) は立体構造原子座標から計算により求めた理論回折強度像であり、各画素での観測光子数期待値 (対数表示) を示している。試料の散乱能が十分強く、各画素での観測光子数が 1 よりも十分大きい場合に観測される図柄ともみなすことができる。X 線結晶解析の場合には、結晶中の分子の周期的配置のためにブラッグ条件を満たす離散点にのみ光子は回折される。一方、試料が単粒子である場合には、Fig. 1(a) に示したようなスペckルと呼ばれる連続的なパターンが得られる。図柄の細かさは、試料粒子の径の長さの逆数に比例し、大きな試料ほど細くなる。試料粒子の電子密度関数のフーリエ変換関数に関しては、分子径の逆数毎の等間隔の離散点での値がわかれば連続関数が完全に分かることを保証する、シャノンのサンプリング定理が成り立つ。この間隔よりもさらに細かな点で観測すれば、それは過剰なサンプリングということになる。過剰な点で強度値を観測すること (オーバーサンプリング) により、失われている位相情報を計算的に回復することがで



**Fig. 1** (Color online) Example of coherent X-ray diffraction patterns simulated by assuming the wavelength of the incident X-ray  $\lambda = 1 \text{ \AA}$ . These plots show the intensity in a log10 scale. (a) A simulated theoretical diffraction intensity pattern for Lysozyme by assuming incident intensity density  $I_i = 5 \times 10^{21}$  photons/pulse/mm<sup>2</sup>. (b) A simulated experimental diffraction pattern for Lysozyme by assuming  $I_i = 5 \times 10^{21}$  photons/pulse/mm<sup>2</sup>. (c) A simulated experimental diffraction pattern for 70s-Ribosome by assuming  $I_i = 2.55 \times 10^{20}$  photons/pulse/mm<sup>2</sup>.

きる。位相情報が回復されれば、フーリエ逆変換により実像を得ることができる。

**Fig. 1(a)**は各画素での観測光子数期待値を表示しているが、実験的に実際に観測される強度は、その画素に到達した光子数であり整数値である。**Fig. 1(b, c)**には入射光強度を仮定したときに、観測光子数期待値から確率的に生成した観測光子数の例を示している。試料の散乱能に対して入射光強度が十分ではない測定では、回折光はまばらにしか飛んでこないため、**Fig. 1(b, c)**に示した様な回折強度像が実験的に得られることになる。

様々な分子方位に対応する、2次元の回折強度像を多数観測した上で、それらをデータ処理し、試料の3次元構造を導く道筋は、大きく2つに分けることができる。

第1の道筋では、観測された2次元の回折強度像に対して、オーバーサンプリング法によって位相を回復し、それに基づいて観測した回折強度像毎に2次元実空間像を

求める。この2次元実空間像は、近似的に試料単粒子の入射光軸方向の射影像であることが知られている。多数の回折強度像に対してこの手続きを実行することにより、試料単粒子の様々な方向への近似的射影像が得られる。これらの射影像に対しては、電子顕微鏡による構造解析法で開発された3次元再構成法が適用可能である。これが第1の道筋の概要で、適用対象によっては有力な方法であるため現在この道筋に沿った手法開発が、装置面および計算手法面の両面でなされている。ただし、問題は**Fig. 1(a)**の様なデータに対しては上手く働くオーバーサンプリングによる位相回復計算が、**Fig. 1(b, c)**の様なデータに対しては、画素当たりの観測光子数が1よりもある程度大きい範囲でしか機能しない点にある。従って、この道筋による構造決定法においては、観測光子数期待値が1となる波数が得られる構造分解能を決める。

この事情を克服し、画期的に高い構造分解能達成が期待

できるのが、現在筆者らが開発に携わっている第2の道筋による解析法<sup>2)</sup>である。いまのところ、観測光子数期待値が0.1程度となる波数に対応する分解能達成が視野に入っている。この方法においては、はじめに、観測された Fig. 1(b, c) の様な多数の回折強度像をその図柄により、単粒子に対する入射 X 線軸方向が近い類似の像ごとに分類する。2つの回折強度像が類似しているかを判定するために、2つの像の相関図を計算的に求める。類似している場合には、相関図に相関線が出現するので、それを検出することで類似性を判定することができる。類似性の高い像を集めて、その平均を求めることによって、量子雑音の影響を低減する。図柄の似た Fig. 1(b, c) の様な実験回折強度像を多数集めることで、その平均から Fig. 1(a) の様な図を求めるのである。この操作により得られた、雑音の低減された1対の像からは、波数空間内における相対配置に関する情報を読み取ることができる。それをもとに、総ての回折強度像を波数空間内に配置することで、3次元の回折強度関数を構築する。この3次元回折強度関数に対してオーバーサンプリング法を適用し位相情報を回復し、さらには実空間3次元構造を導く。これが第2の道筋の概要である。第1の道筋と第2の道筋とでは、位相情報を回復する手順と2次元情報を3次元空間に配置する手順とが逆になっている。

第2の道筋においては、画像全体の情報を総合的に取り込むことで、類似性を判定するため、画素当たりの観測光子数期待値が0.1程度の画素における観測値からも、有効な情報を読み取ることができる。このため、第2の道筋を採用することで、より高分解能の構造解析を達成することができる。本記事では、関連する研究を紹介すると共に、第2の道筋によるデータ解析アルゴリズムと、それによって達成される分解能について紹介する。

結晶化の必要のない生体高分子単粒子試料を対象とする XFEL を用いた高分解能立体構造解析法を実現するためには、第2の道筋によるデータ解析を採用すべきと考える。それに照準を合わせたハード面の早急な整備が必要であり、実現のためには効率良く観測が行える実験系の構築が重要な要素となる。ソフト面では、開発したアルゴリズムを理化学研究所・計算科学研究機構に設置された「京」コンピュータに実装することで高速処理を可能とするシステム作りを進めており、この努力も紹介する。

## 2. 単粒子構造解析法概要

生体高分子の立体構造を構築する既存の手法には、X 線結晶構造解析法、電子顕微鏡法、核磁気共鳴法などがある。現状で最も高分解能を達成している X 線結晶構造解析法では、少なくとも  $10^{12}$  個の分子が規則正しく配列した、結晶状態の試料をもちいることで、X 線回折強度から重原子同型置換法などにより位相情報を決定し、分子構

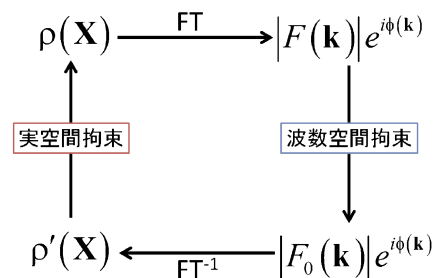


Fig. 2 (Color online) Block diagram of the system for hybrid input-output algorithm.

造を表す電子密度関数を原子分解能で構築することができる。X 線結晶構造解析により高分解能構造を得るには良質の結晶が必要となる。良質の結晶試料を準備できるタンパク質は限られており、このことが X 線結晶構造解析法に対して大きな制限を与えている。

空間的に X 線の位相がそろったコヒーレント X 線をプローブとして用いることで、結晶以外の試料に対して回折による構造解析を行うことができる。コヒーレント X 線回折では、連続的な図柄をもつスペckルが観測されるため、回折強度をオーバーサンプリングすることにより、観測値では失われている位相情報を取り戻すことができる。

オーバーサンプリングによる位相回復の可能性は、D. Sayer により1952年に提唱された<sup>3)</sup>。オーバーサンプリング法における、位相回復計算法には幾つかの手法が存在している。反復的位相回復方法は1972年に Gerchberg と Saxton により提案され<sup>4)</sup>、その後 J. Fienup により改良された。現在は、J. Fienup により提案された HIO (Hybrid Input Output) アルゴリズム<sup>5)</sup>による反復的位相回復計算が広く用いられている。

HIO アルゴリズムによる反復的位相回復計算の模式図を Fig. 2 に示す。HIO アルゴリズムでは実空間拘束と波数空間拘束のもと、フーリエ変換を繰り返し行う。おのおのの格子点上の回折強度より散乱振幅  $|F_0(\mathbf{k})|$  をもとめ、それに対して乱数により初期位相をランダムに与えることで、位相回復計算を始める。実空間拘束では、サポート領域以外の画素に対して値がゼロに近づくように拘束を加える。ここでいうサポート領域とは、求めたい像が存在する領域である。一方、波数空間拘束では、フーリエ振幅  $|F(\mathbf{k})|$  を回折強度から求めた散乱振幅  $|F_0(\mathbf{k})|$  におきかえ、位相のみ次の計算に引き継ぐ。この計算を繰り返し行うことでやがて計算は収束し、波数空間では位相が回復され実空間では電子密度関数が復元される。

コヒーレント X 線回折法により構造解析が行われている主な試料は、現在のところタンパク質などの生体高分子ではなく、より試料の散乱能が高い、金属試料<sup>6)</sup>、バクテリア<sup>7)</sup>、細胞<sup>8)</sup>、染色体<sup>9)</sup>、ウイルス<sup>10,11)</sup>である。

ここに示した例では、2次元構造解析と3次元構造解析が混在している。散乱角が狭い回折強度像では Ewald 球

面を平面とみなすことができ、位相回復計算により試料の2次元平面への射影実像を求めることができる。入射X線に対する分子方位を様々に変えた測定を行うことで、試料の3次元構造も構築することができる。ヒト染色体を用いた例では、サンプルホルダーに固定した試料の姿勢を制御することで、単一の試料に対して幾つかのX線入射角度で測定を行い、実験的に得られる相対位置情報をもとに波数空間で3次元回折強度関数を構築し、3次元関数に対して位相回復計算を行うことで、120 nm 分解能の3次元構造が構築されている。

結晶構造解析に比べると現状到達している分解能は～10 nm と高くはない。この理由の一つは、オーバーサンプリング法による反復的位相回復計算を行うには、一定以上の回折強度が要請されることにある。回折強度が弱く S/N 比が悪い回折強度像に対しては、反復的位相回復計算の収束性が悪くなり、位相回復をうまく行うことができない。高い分解能を得るには、広い散乱角からの信号が必要であるが、散乱角が広がるほど回折強度は一般に下がる。つまり、コヒーレント X 線回折法においては、到達可能な分解能は位相回復計算の適用限界が決めることになる。

このため、コヒーレント X 線回折実験では、位相回復計算の精度と、それに耐えうる強度の回折強度像をいかに広い波数範囲で、獲得するかということが高分解能構造を得るための鍵となる。

XFEL により発振されるフェムト秒 X 線レーザーを用いると、より強い光を試料に対して一瞬照射することができる。強い X 線を試料に照射すれば、照射とともに非弾性散乱により分子の放射線損傷が始まるが、測定は損傷の影響が大きくなる前に終わる。測定が終わったあとには、試料分子はやがてクーロン爆発をおこして崩壊することになる。このため、XFEL を用いた単粒子構造解析では、測定毎に新たな試料粒子を導入する。

測定時に新規試料粒子を導入する方法は幾つか提案されている。高分解能構造を構築するという観点からは、試料分子の周りに散乱体が無く、オーバーサンプリング条件を純粋に満たすことができる、分子飛翔法<sup>12)</sup>が推奨される。分子飛翔法ではインjekター機構を用いて試料粒子を真空チャンバー内に飛ばし、そこに X 線パルスレーザーを照射することで、単粒子試料からのコヒーレント回折強度像を2次元検出器により測定する (Fig. 3)。3次元構造を構築するには、このような観測を何度か行う。この方法では、測定時の分子方位は制御できず、分子方位に関する直接の情報は測定時には得ることができない。試料が単粒子であるがゆえ、回折強度は弱く S/N 比の悪い回折強度像が得られる。回折強度を稼ぐには、試料位置の入射光の強度密度をさらに上げることも一つの方法であるが、分子の放射線損傷を考えると、必要以上に入射光強度密度を上げることはできない。加えて、パルス通過後は分子が

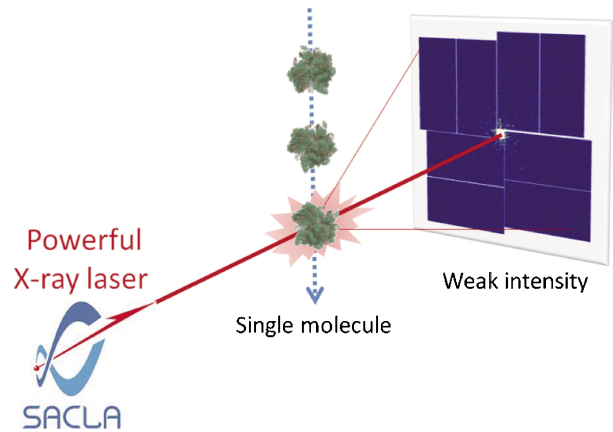


Fig. 3 (Color online) Schematic description of single bio-macular X-ray coherent imaging using XFEL and a molecular beam injection system.

壊れてしまうために、同じ試料に対して何度も X 線を照射し回折強度の積算を行うことはできない。このような事情から XFEL 単粒子構造解析では、量子雑音に埋もれた微弱な信号を相手にすることになる。

このような背景のもと、高分解能構造を得る方法として、回折強度像を分類・平均することでその精度を向上するという案が示された<sup>13)</sup>。筆者らはこの流れのもと研究を進め、細部を詳細に検討することで、回折強度像の分類法および3次元回折強度関数の構築法を新たに開発した。開発した方法は量子雑音に対して強く、雑音に埋もれた弱いシグナルから高い分解能の構造を構築することが可能である。

いま一度、我々が取り組むべき問題を整理する。生体分子の立体構造を求めるとは、実空間に広がる3次元の電子密度関数の分布を求めることである。一方、X線回折実験の観測量である回折強度  $s(\mathbf{k})$  は、構造因子  $F(\mathbf{k})$  の絶対値の二乗に比例し、偏光因子を除いた式は次式で与えられる。

$$s(\mathbf{k}) = I_i r_{CE}^2 \omega |F(\mathbf{k})|^2 \quad (1)$$

ここで、 $I_i$  は入射 X 線強度密度であり本稿では photons/pulse/mm<sup>2</sup> で表すことにする。 $r_{CE}$  は古典電子半径を表し定数である。 $\omega$  は画素あたりの立体角であり、入射 X 線波長、分子径  $L$  および次元あたりのオーバーサンプリング比  $\sigma$  により  $\omega = (\lambda/\sigma L)^2$  で表される。構造因子と電子密度関数は次のフーリエ変換の関係にあり、波数空間において構造因子と位相情報が得られれば、実空間における電子密度関数  $\rho(\mathbf{x})$  を求めることができる。

$$\rho(\mathbf{x}) = \int d\mathbf{k} F(\mathbf{k}) \exp(2\pi i \mathbf{k} \cdot \mathbf{x}) \quad (2)$$

この式が示すとおり，実空間で3次元電子密度関数を求めるには，波数空間においても3次元の情報が必要である。観測量である2次元検出器像は，波数空間に広がる回折強度関数を，原点に接し半径が入射X線の波長分の1であるEwald球面上の回折強度の一部に対応する。入射X線に対する分子方位が異なると，切り取るEwald球面の方向が変わる。相対的な配置が分かっている，位相情報が欠落したS/N比の悪い2次元コヒーレント回折強度像が多数得られたときに，3次元電子密度関数を求めることが今の問題設定である。

筆者らが具体化した単粒子構造解析法の概要は次のとおりである。手順1. 一対の回折強度像の類似度を求めることにより，回折強度像を分類する。分類後，グループ化を行いグループ内で回折強度像を平均することでS/N比を向上する。手順2. S/N比向上後の回折強度像を用いて，回折強度像の共通部分から3次元回折強度関数を構築する。手順3. 精度が向上された3次元回折強度関数をもとに，オーバーサンプリング法により位相情報を回復し3次元電子密度関数を構築する。

この解析法では，分類・平均操作により，精度が向上された回折強度像を作成することで，単体では位相を回復することが難しい，微弱な強度の回折強度像をも活用することが可能となる。以降で，手順1および手順2に対して開発したアルゴリズムの概要を説明する。

### 3. 回折強度像分類法

高分解能達成の鍵となる回折強度像の分類・平均操作では，回折強度像をその図柄をもとに分類する。これにより，似ている回折強度像を選択し，それらを平均することで回折強度像の精度を向上することができる。このため，1対の回折強度像から図柄の相関値を求め類似度を求める。回折強度像の図柄の類似度は，入射X線に対する分子方位の類似度を意味する。同じことを波数空間に焼きなおすと，各々の測定で観測されたEwald球が波数空間で互いにどの程度近傍に位置しているのかを計ることとなる。1対のEwald球中心座標が原点となす角度 $\beta_{ij}$ を近傍値と呼ぶことにする。ここで，添え字*i, j*は回折強度像につけられた通し番号のうち任意のものを表す。分類計算ではこの近傍値に密接に関係した量をいかに精度よく求めることができるかが重要となる。

筆者らは，1対の回折強度像の相関値 $c_{ij}$ を次式で定義した。光子数期待値の波数依存性により回折強度像を規格化した上で相関値を求めることで，散乱角がより広い領域の相関を拾うことができ，先行研究に比べより微弱な信号をも活用することが可能となった。

$$\begin{aligned} c_{ij}(\xi, \alpha) &= \frac{\Psi_{ij}(\xi, \alpha)}{\bar{s}_Q(i; \xi) \bar{s}_Q(j; \xi)} - 1, \\ \Psi_{ij}(\xi, \alpha) &= \frac{1}{N_\xi} \sum_{l=0}^{N_\xi-1} s_Q\left(i; \xi, \frac{2\pi l}{N_\xi}\right) s_Q\left(j; \xi, \frac{2\pi l}{N_\xi} + \alpha\right), \\ \bar{s}_Q(i; \xi) &= \frac{1}{N_\xi} \sum_{l=0}^{N_\xi-1} s_Q\left(i; \xi, \frac{2\pi l}{N_\xi}\right). \end{aligned} \quad (3)$$

入射X線軸に対する分子の回転は，回折強度像ではビーム中心を支点とした回転として現れる。上記相関係数は，この回転自由度 $\alpha$ を考慮することで，図柄を2次元平面に対して360度回転した相関値を一度に求めることができる。相関値を $\alpha$ と $k = (2/\lambda) \sin(\xi/2)$ の関数として表示したものを，相関図とよぶ。ここで， $\xi = 2\theta$ であり散乱角をあらわす。また， $s$ は観測光子数， $\bar{s}$ は光子数期待値， $\lambda$ は入射X線波長， $N_\xi$ は同心円上の画素の離散数をそれぞれあらわす。

1対の回折強度像の類似度が高く，測定時の分子方位が互いに似ている場合には相関図には高い相関値を示す相関線が出現する。Fig. 4(a-c)には，シミュレーションより求めた，雑音のない1対の理論回折強度像を用いて計算した相関図を示している。相対的な分子方位を $\beta_{ij} = 0.58^\circ, 1.48^\circ, 3.00^\circ$ と変えた場合の相関図を示している。

相関線が出現する方向は，入射X線軸に対する分子の相対回転角である $\alpha$ に対応する。Fig. 4では入射X線軸に対する相対回転角はゼロとした。この場合相関線は3時方向に出現する。測定時の分子方位がたがいにずれ， $\beta_{ij}$ 角が増すにつれ，相関線の長さが短くなるのが見て取れる。つまり，高い波数領域まで相関線が伸びているほど， $\beta_{ij}$ 角度が小さく1対の回折強度像が波数空間では近傍に位置する。

相関図を用いて，ある代表回折強度像に対して波数 $\hat{k}$ まで相関線が伸びている回折強度像によりグループを形成したとする。この時，グループ内で平均像を求めると， $\hat{k}$ よりも小さな波数範囲では図柄の精度を向上することができる。このグループの角を $\delta_G$ とよぶことにする。一方， $\hat{k}$ よりも高角側の波数領域では，平均像を求めることで図柄の情報が壊れる。高分解能を達成するためには，相関線ができるだけ長く表れた回折強度像セットを用いて平均像を求め，その精度を向上するのが良いということになる。

実際の実験データでは，この基本的な相関線の振る舞いに，量子雑音の影響が加わる。Fig. 5(a-c)は，Fig. 4(a)と同じ $\beta_{ij} = 0.58^\circ$ の場合であるが，量子雑音を加えた1対の実験回折強度像を用いて相関図を求めた。入射X線強度密度を $I_i = 2.55 \times 10^{20}, 8 \times 10^{19}, 5 \times 10^{19}$  photons/pulse/mm<sup>2</sup>と変えた図を示している。

Fig. 5を見ると，入射X線強度密度が下がるにつれ，高い波数領域では相関線を認識することが難しくなってくる。これは，波数の大きな領域ほど回折強度は弱く，相関図に量子雑音の影響が強くなるためである。このため，

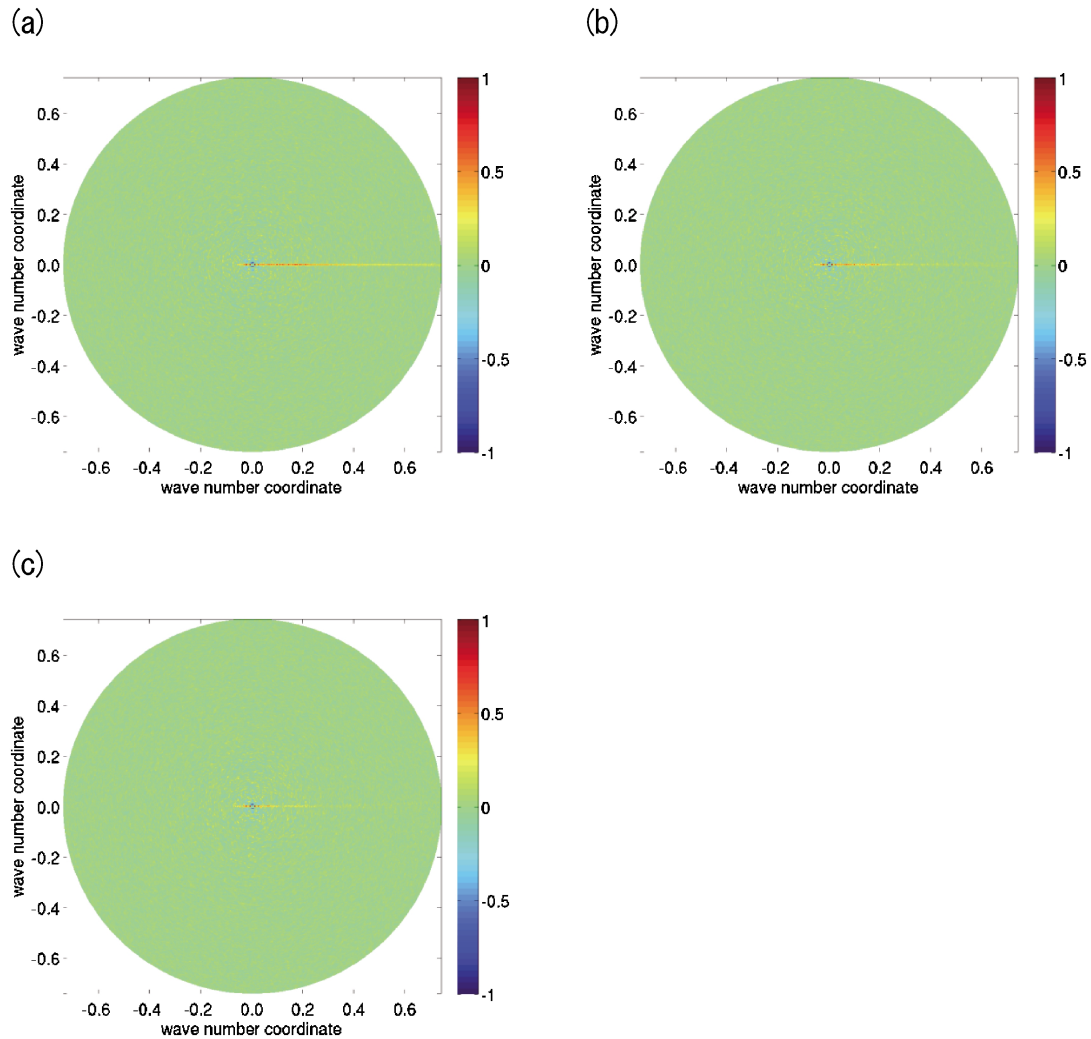


Fig. 4 (Color online) Correlation pattern for the case of 70s-Ribosome calculated with pair of theoretical diffraction patterns. (a) In case of the angle  $\beta_{ij}=0.58^\circ$ . (b) In case of the angle  $\beta_{ij}=1.48^\circ$ . (c) In case of the angle  $\beta_{ij}=3.00^\circ$ .

回折強度に依存して、相関線を認識できる波数が決まることになる。雑音の影響が顕著になる波数を  $k_N$  とする。Fig. 4(a) が示す通り、 $\beta_{ij}$  角度が小さい場合には  $k_N$  を超えて相関線が存在するはずであるが、Fig. 5 が示す通り雑音はその認識の邪魔をする。相関線が認識できる最大の波数は雑音の影響が顕著になる波数  $k_N$  が与えることになり、これが到達可能な分解能を決める。

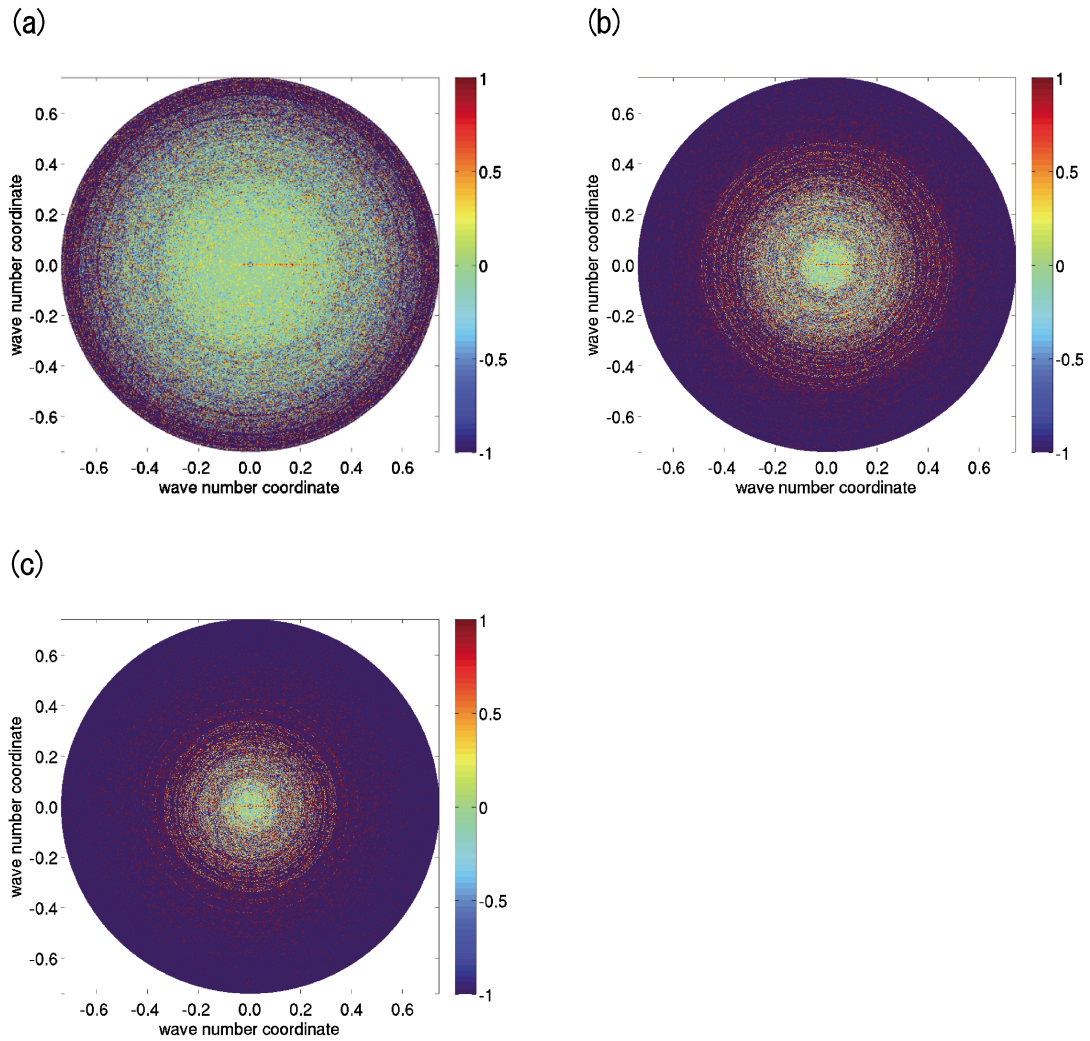
#### 4. 三次元回折強度関数構築法

精度が向上された回折強度像を用いることで、その共通部分を求めることで 3 次元回折強度関数を構築することができる。分子方位が異なる測定で得られる回折強度像は、異なる Ewald 球面上の関数に対応するが、どの Ewald 球も原点に接するため、1 対の Ewald 球は交わりを持つことになる。1 対の回折強度像の差を計算することで、この共通部分を求めることができ、波数空間における

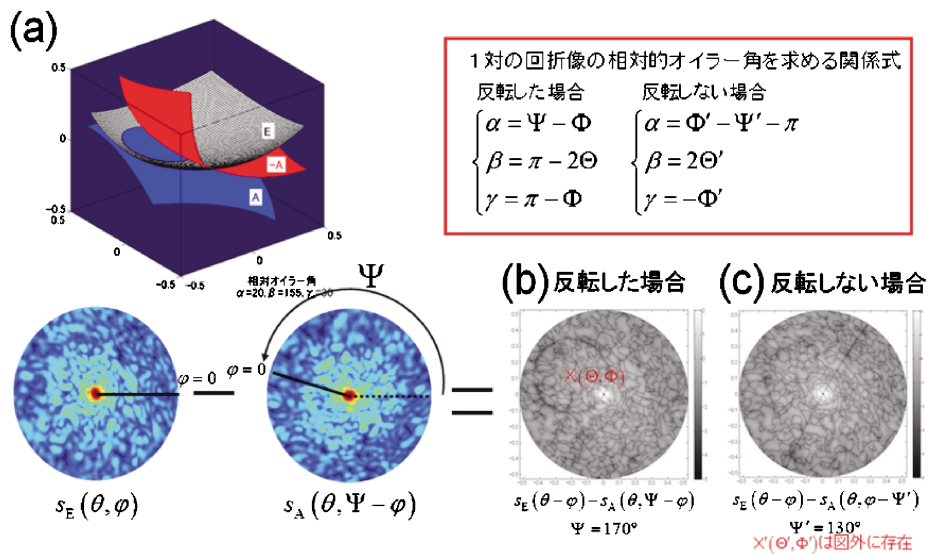
相対位置が分かる。

Fig. 6(a) には 3 つの Ewald 球平面の波数空間における相対位置を示している。円あるいは円弧として交わっているのが分かる。1 対の Ewald 球面の波数空間での相対位置は、オイラー角  $(\alpha, \beta, \gamma)$  を用いて表現できる。図中の平面 E を基準 Ewald 球面とすると、図中の例では Ewald 球面 A の相対的なオイラー角は  $(\alpha, \beta, \gamma) = (20^\circ, 155^\circ, 30^\circ)$  で与えられる。回折強度像の相対配置解析では、この相対的オイラー角を各対に対して求めることが目的となる。Ewald 球面-A は Ewald 球面 A に対して、波数空間で原点に対して 180 度反対側に位置する。実像のフーリエ振幅である回折強度は、波数空間で点対称性を持つことから、Ewald 球面-A の回折強度像は Ewald 球面 A の回折強度像を裏返した図柄を持つ。このことから、E と A の回折強度像からは、E と A の共通部分と E と-A の共通部分の両方を求めることができる。

1 対の回折強度像の一方を固定し、他方を反転した上で



**Fig. 5** (Color online) Correlation pattern for the case of 70s-Ribosome calculated with pair of experimental diffraction intensity patterns in a case of the angle  $\beta_{ij} = 0.58^\circ$ . (a) In case of incident X-ray intensity density of  $I_i = 2.25 \times 10^{20}$  photons/pulse/mm<sup>2</sup>. (b) In case of  $I_i = 8 \times 10^{19}$  photons/pulse/mm<sup>2</sup>. (c) In case of  $I_i = 5 \times 10^{19}$  photons/pulse/mm<sup>2</sup>.



**Fig. 6** (Color online) Schematic description of three different orientation Ewald spheres in the wave number space and plot for detecting an intersecting circle between a pair of enhanced diffraction patterns.

回折強度像中心を支点に回転しながら差を計算することで、ある回転角で共通部分がゼロとなり浮かび上がる (Fig. 6(b))。回折強度の点対称性から、一方の回折強度像を反転しない差像からも同様に共通部分を求めることができる (Fig. 6(c))。

一方の像を反転した差像から得られた、共通交円の中心座標  $(\Theta, \Phi)$  と回転角  $\Psi$  から 1 対の Ewald 球の相対的なオイラー角  $(\alpha, \beta, \gamma)$  を次式により求めることができる。

$$\alpha = \Psi - \Phi, \beta = \pi - 2\Theta, \gamma = \pi - \Phi. \quad (4)$$

どちらの回折強度像も裏返さずに差像を求めた場合は、その共通交円の中心座標  $(\Theta', \Phi')$  と回転角  $\Psi'$  から、相対的なオイラー角は次式で与えられる。

$$\alpha = \Phi' - \Psi' - \pi, \beta = 2\Theta', \gamma = -\Phi' \quad (5)$$

共通交円の中心座標の推定には、画像処理法の一つであるハフ変換を用いることができる。相対的なオイラー角を、平均により精度が向上された回折像対ごとに求めることで、波数空間で回折強度像を立体的に配置することができる。

上記手順により構築される 3 次元回折強度関数は、多数の Ewald 球の一部により構成される。この関数をもとに内挿等により 3 次元格子状に変換することで、S/N 比が向上された 3 次元格子状の回折強度関数を構築できる。

構築した、3 次元回折強度関数に反復的位相回復法を適用し位相を回復する。位相回復において、量子雑音の影響は収束性に現れる。雑音の影響がひどい場合には、計算は収束することはなく、振動するような挙動を示すこともある。また量子雑音下では、初期位相を変えて計算を行うたびに、回復した位相とそこから得られる構造が異なり、どれが正しい構造か判断がつかなくなる場合がある。このような状況では、多数の異なる収束構造の平均像が解析の結果得られた構造として用いられることがあるが、得られる分解能はその分下がる。筆者らの経験では、HIO アルゴリズムによる位相回復法が収束するには、回折強度期待値が 1 光子程度の信号が必要であるという印象である。筆者らが採用する道筋 2 では手順 1 で、回折強度像を平均し雑音の影響を抑えることで、位相回復における雑音の影響を著しく軽減できる。これにより、HIO アルゴリズムによる位相回復においても、収束性の問題が分解能に与える影響を低減できる。

この方法では、手順 1 で回折強度像の精度をどの波数領域まで向上することができるかにより、位相回復計算を適用できる波数領域が決まり、これにより構築できる構造の分解能が決まる。

## 5. 到達可能分解能の予測

平均操作で精度が向上する波数、つまり分解能波数  $k_R$  は、相関線が認識できる最大の波数によって与えられる。相関線が認識できる最大の波数は、量子雑音が顕著になる波数  $k_N$  と密接に関係しており、これは有効画素に到達する光子数によって決まる。このため、波数空間において一辺が相関長  $k_c$  に対応する大きさをもつ有効画素に到達する光子数の期待値から分解能波数を見積もることができる。ここでいう相関長とは、波数空間における回折強度がもつ相関距離のことである。また、回折強度の標準偏差  $\sigma_c$  を量子雑音の指標として用いることで、分解能波数  $k_R$  を回折強度から見積もることができる。標準偏差が相関線上の相関値と等しくなるまで相関線を認識できることから、相関線上の相関値を控えめに与えることで、 $\sigma_c = 0.6$  となる波数を  $k_N$  の定義とした。

回折強度像がもつ量子雑音由来のポアソン分布と生体高分子の複雑な構造に由来する構造非周期性分布の 2 つの分布を考慮することで、回折強度の分散値は回折強度期待値である  $\bar{s}(k)$  の関数として次式で与えられることを筆者らは示した。この式を用いることにより、回折強度期待値から到達可能な分解能を見積もることができる。

$$\sigma_c^2 = \frac{g(\bar{s}(k))}{2\pi k L \left[ 1 - \left( \frac{k\lambda}{2} \right)^2 \right]^{1/2}} \quad (6)$$

$$g(x) = \frac{5x^2 + 6x + 1}{x^2} \quad (7)$$

$L$  は分子径、 $\lambda$  は入射 X 線波長、 $k$  は波数をそれぞれ表す。回折強度の期待値  $\bar{s}(k)$  は、3 次元波数空間の回折強度を波数一定の球面上で平均したものである。回折強度像のビーム中心を支点とした、同心円上で強度を平均したものは、この  $\bar{s}(k)$  と精度よく一致するため、一枚の回折強度像データから分解能波数を与えることも可能である。また筆者らは  $\bar{s}(k)$  関数の標準タンパク質モデルを作成した。標準タンパク質モデルでは分子径と分子形状の球形からのずれを表すパラメータの二つにより  $\bar{s}(k)$  を精度よく求めることができる。これにより分子径の関数として、到達可能な分解能を見積もることも可能である。

(6)式より、分解能波数が与えられると、相関長と分解能波数の関係からグループの角  $\delta_G$  を見積もる事ができ、3 次元構造を構築するのに必要な有効回折強度像の数を見積もることが可能になる。

リボソーム分子を例として、入射 X 線強度密度  $2.55 \times 10^{20}$  photons/pulse/mm<sup>2</sup> を仮定した場合に、到達可能分解能、立体構造を構築するのに必要な有効回折強度像数などを見積もった。結果を Table 1 にまとめる。ここで仮定した入射 X 線強度密度は、SACLA 仕様値<sup>14)</sup> のパルス強



**Table 1** Attainable resolution and necessary number of diffraction patterns and amount of data for 70s-Ribosome

分子	70s-Ribosome
分子径 or 回折強度の相関長: $1/k_C=L$	270 Å
入射強度密度 [photons/pulse/mm <sup>2</sup> ]	$2.55 \times 10^{20}$
雑音が顕著になる波数 $k_N$ [Å <sup>-1</sup> ]	0.33
$k_N$ における光子数期待値 $s_N$	0.09
分解能: $1/k_N$ [Å]	3.0
グループの角: $\delta_G = k_C/k_N$ [degree]	0.64
グループ数: $N_G = 2/\delta_G^2$	15,878
各グループでの平均枚数: $N_A = 8/s_N$	88
必要な有効回折像の数: $N_{\text{measure}} = N_A \cdot N_G$	$1.4 \times 10^6$
分類計算の数: $N_{\text{compare}} = N_A \cdot N_G^2$	$2.2 \times 10^{10}$
回折像総データ量: 16.8 MB ( $2,048 \times 2,048 \times 4$ byte) $\cdot N_{\text{measure}}$	23.5 TB

度を50 nm (FWHM) 集光した理想的な値に対応する。

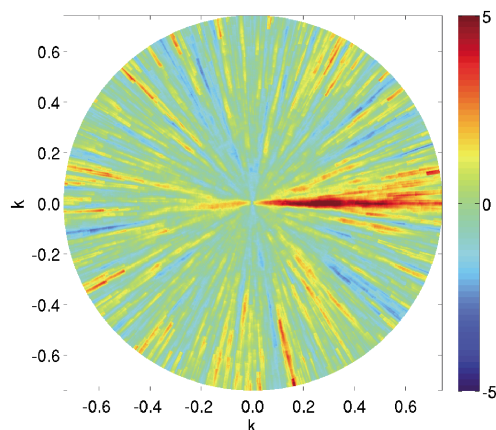
この時、量子雑音が顕著になる波数  $k_N$  は  $0.33 \text{ \AA}^{-1}$  であり、3 Å 分解能を達成できる見込みである。リボソーム分子において3 Å 分解能の達成は結晶構造解析に匹敵する数値であり、単粒子構造解析では結晶化の必要がないことからその有用性は非常に高い。一度、この方法が確立されてしまえば同サイズの他の多くの分子に対して同様の高分解能構造解析が期待できるのである。

$k_N$  における回折強度期待値  $s_N$  は0.09光子であり、ここで紹介した解析法では約1/10光子までも有効信号として扱っていることになる。SACLA に設置されたMP-CCDs (multi-port charge-coupled devices) 検出器を用いると、上記解析に必要な波数領域を測定するには試料と検出の間の距離を10.8 cm とすれば良い。MP-CCDs 検出器は50 μm の物理ピクセルが2048×2048並んだ構造をしている。このジオメトリを採用することで、観測可能な最大の波数は  $0.44 \text{ [Å}^{-1}]$  となり、一次元あたりのオーバーサンプリング比は8となる。

この条件では、平均グループの角  $\delta_G$  は0.64度となる。立体構造を得るには、波数空間で半球分の立体角をこのグループ角で埋め尽くす必要があることから、グループ数は約16,000となる。グループ内で必要な平均枚数は、共通交点を認識可能な枚数から88枚と見積もることができる。これより、3 Å の立体構造を構築するのに必要な、有効回折強度像の数は、 $1.4 \times 10^6$  となる。この数の回折強度像を分類するのに必要な、分類計算の数は、 $2.2 \times 10^{10}$  という一見すると膨大な数となる。このような膨大な分類計算を行うには、相関線の自動同定法と、大規模計算を実現するための計算機構が必要となる。

## 6. 相関線自動同定法

$1.4 \times 10^6$  の回折強度像に対して、 $2.2 \times 10^{10}$  回という多数回の相関計算を実行しさえすれば、原理的にはこの方法



**Fig. 7** (Color online) Significance map for the case of 70s-Ribosome calculated with pair of experimental diffraction intensity patterns in a case of the angle  $\beta_{ij} = 0.58^\circ$  and  $I_i = 5 \times 10^{19}$  photons/pulse/mm<sup>2</sup>.

でリボソーム分子の3 Å 分解能を達成することができると。このためには、相関図から相関線を敏速に自動同定し、近傍情報を抽出するための方法を構築する必要がある。相関図から近傍情報を抽出するために求めるべき項目は次の4つである。①相関線の有無②相関線の方向  $\hat{\alpha}$ ③相関線の長さ  $\hat{k}$ ④相関線の強さ  $I_c(\hat{k}, \hat{\alpha})$ 。論文投稿前であることから細部は割愛するが、筆者らは相関値を積分し、さらに相関線の有意度を計る量を定義することで、相関線を自動同定する方法を開発した。微弱な相関値を積分により集積することにより、強い量子雑音下において上記4項目を精度よく求めることができる。

**Fig. 5** に示した一連の相関図をみると、入射 X 線強度密度が下がるにつれ、目視では相関線が認識しづらくなる。**Fig. 7** には、**Fig. 5(c)** より求めた相関値の有意度を示す量を表示している。**Fig. 5(c)** では認識しづらい相関線を、**Fig. 7** では3時方向に相関線が現れていることを容易に認識することができる。このように、開発した相関線の自動同定法を導入することにより、より弱い相関線も拾いあげることが可能となる。

**Fig. 8** には、リボソーム、入射 X 線強度密度  $2.55 \times 10^{20}$  photons/pulse/mm<sup>2</sup> の相関線自動同定の結果を示す。**Fig. 8** においては、各々の点は一回の分類計算の結果を表しており、約800回の相関計算の結果を示している。横軸はEwald 球の近傍値である  $\beta_{ij}$  を示しており、求めるべき答えに対応する。縦軸は相関線の強さを表す  $I_c(\hat{k}, \hat{\alpha})$  を示しており、自動同定法により抽出された近傍情報である。

相関線上の相関値の積分強度  $I_c^{(1)}$  は、 $\beta_{ij}$  の関数として理論的には次式で与えられる。

$$I_c^{(1)} = \frac{\pi \lambda^2}{\sqrt{2} L^2 \beta_{ij}} \quad (8)$$

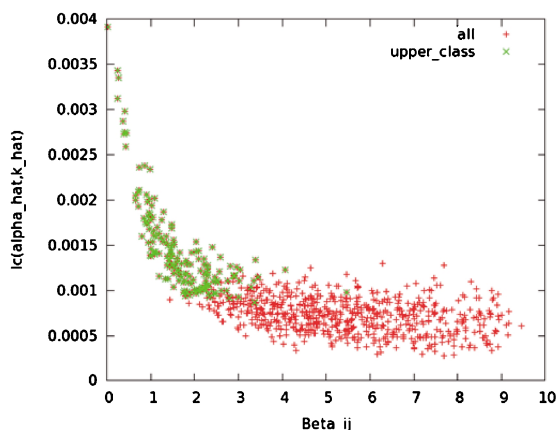


Fig. 8 (Color online) Plot the  $I_c(\hat{\alpha}, \hat{k})$  which is the neighborhood value between two images estimated by the automatic identification method, versus the angle of  $\beta_{ij}$  which is answer.

Fig. 8 に示した  $I_c(\hat{k}, \hat{\alpha})$  と  $\beta_{ij}$  の分布図は、この理論式の周りに点が分布しており、入射 X 線強度密度が一定水準以上である場合には、 $I_c(\hat{k}, \hat{\alpha})$  より  $\beta_{ij}$  を推定することも可能であることを示している。一方で、入射 X 線強度密度が下がると自動同定法に与える量子雑音の影響が強くなり、 $I_c(\hat{k}, \hat{\alpha})$  と  $\beta_{ij}$  の関係が次第に崩れる。

いまの解析法ではグループを形成することが目的となるため、得られた近傍情報をもとに一定の基準を満たす上位クラスを判定することになる。図中緑で示した点は、上位クラスの判定基準を通過したことを示している。ここに示した  $2.55 \times 10^{20}$  photons/pulse/mm<sup>2</sup> の場合には、 $I_c(\hat{k}, \hat{\alpha})$  と  $\beta_{ij}$  の関係が理論値に近く、 $I_c(\hat{k}, \hat{\alpha})$  の値によりグループ内で  $\beta_{ij}$  に対する序列をつけることも可能である。これは判定基準をより厳しくすることで、より狭い  $\beta_{ij}$  角度の回折強度像セットのみを抽出することができることを示しており、より近傍の回折強度像のみでグループ化を行うことで、高い波数領域まで精度を向上することが可能となり、より高分解能な構造も構築することができることを示唆している。このように、相関線の自動判定法の導入により、前述で見積もった分解能よりもさらに高い分解能をも狙える可能性も出てきた。

## 7. 「京」と連携した大規模計算機構の構築

前述した相関線自動同定法を用いて大規模な分類計算を実現するため、計算システムのインフラ整備にも取り組んでいる。理化学研究所・計算科学研究機構と連携することで、京速コンピュータ「京」を用いた大規模回折強度像データ転送・分類解析機構の整備を進めている。「京」と SACLA は共に、兵庫県に位置しており、SACLA で測定した回折強度像データをオンラインで「京」に転送することで、大量の回折強度像に対する、大規模な分類解析を敏速に行うことができると期待している。

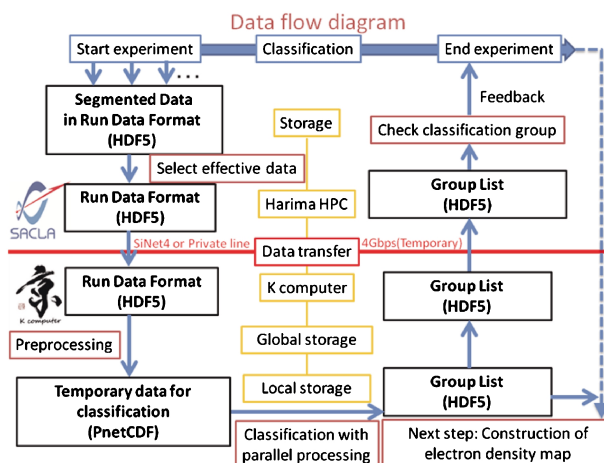


Fig. 9 (Color online) Data flow diagram for single bio-molecular structure analysis system with the K computer.

Fig. 9 に計画しているデータフロー図を示す。SACLA 施設で測定された検出器生データは、播磨に設置されているストレージへ高速でもれなく保存される。ここで得られた、実験データには、大量のごみデータが含まれることが予想される。このため、分類解析を行う対象である有効回折強度像を選別する為のフィルタリング機構が必要となり、現在整備が進められている。フィルタリング解析を通過した回折強度像データは、HDF5 形式で保存され、Run-Data-Format と呼ばれるフォーマットで保存される。Run-Data-Format に収められるデータ項目は筆者らが策定を行い、後の解析に必要な項目が収められている。この Run-Data-Format を播磨から京へ転送する。このため、データ転送機構の整備も並行して進められている。SFTP プロトコルにより SINET4 を介した転送試験が行われており、今後専用回線やデータ圧縮などが検討され、転送機構と転送形式に対してのチューニングがさらに進められる予定である。

「京」に送られた有効回折強度像データは、回折強度像の分類解析に適した形式に変換し Temporal-Data-for-Classification として「京」のストレージに保存しておく。保存形式は I/O 速度向上のため、netCDF 形式を採用した。回折強度像の分類解析では、おのおのの相関計算は独立性が高いため並列処理が適している。本稿で紹介した解析法を多数の計算ノードを用いて並列処理することで、高速化が期待できる。各々の計算ノードで異なる回折強度像対にたいして相関値を求め、自動同定法により近傍判定を行い、近傍情報リストを作成する。近傍情報リストをもとに閾値を与えることで、類似回折強度像でグループを形成する。回折強度像分類解析における出力は近傍情報リストであり、この近傍リストを後の解析に引き継ぐことで、2次元平均回折強度像の作成→3次元回折強度関数の構築→位相回復という、電子密度関数を構築する一連の解析を行うことができる。

「京」リソースの約半分である、4万ノードを用いて24時間以内に約100万枚の回折強度像を分類することを目標としている。年度内には「京」において、リボソームを対象とした100万枚の模擬回折強度像データを用いたベンチマークテストを行う予定にしている。目標値が達成されると、実験と同時進行で回折強度像の分類解析を行うことが可能となる。実験中に分類解析結果が得られることで、実験系のセットアップに対するフィードバック機構を構築することができる。フィードバック機構の働きにより、試料濃度の調整や、試料と検出器間の距離などの測定条件を精査することができ、効率よい測定への貢献が期待される。

前述したように、生体高分子の高分解能構造解析を行うには道筋2が必要である。一方、道筋2では一見膨大な分類計算を行う必要があり、現実離れしているのではないかという指摘を受けることもある。しかし、多数の計算ノードを用いて並列処理を行うことにより、実験と並行して分類計算を行うことさえ可能となるのである。開発した構造構築法と現存する計算機を用いることで、リボソーム程度の大きさの分子で3 Åという高分解能構造解析を実現するための計算機構を構築することができる。

## 8. まとめ

本稿で紹介した回折強度像の分類解析は似ているものを選択し平均することで、回折強度が弱く単体では構造情報を得ることが難しい回折強度像を有効活用するための方法であるといえる。グループ化を行う方法では、分類作業で回折強度像が似ているか似ていないのみを判断すればよいが、相関線の自動同定法は、回折強度像の近傍情報を定量的に評価することもできるため、それ以上の情報を提供する。入射 X 線強度密度が一定以上あれば、分類計算により詳細な近傍情報を得ることができ、この情報を用いることで回折強度像の平均操作を行うことなく、直接3次元波数空間へ回折強度像をマッピングする方法への展開も期待できる。この発展的方法では必要測定枚数の減少が期待できる。

現在、SACLA では様々な利用実験が行われている。世界的に見ても、微結晶試料を用いた実験や、生体試料を用いたコヒーレント回折実験など成果が報告されつつある<sup>15)</sup>。筆者らがめざすところは3次元高分解能構造解析であるが、効率のよい測定が行える実験系が確立されるまでは、精度の向上した2次元の折強度像に対して位相回復計算を行うことで、単体では解析が難しい回折強度像セットから、低分解能の2次元射影像を求めるといった、利用法に対しても回折強度像の分類解析は有効である。

生体高分子を試料とした高分解能構造解析を行うためには、弱い信号を拾うために大量の回折強度像を分類する。これを実現するための計算機構は、「京」と連携することで可能となる。また入射 X 線強度密度も高分解能達

成には重要な要素である。高い入射 X 線強度密度を達成するには、X 線領域における集光技術が重要な役割を果たすが、2010年には硬 X 線領域の10 nm 集光が実現されている<sup>16)</sup>。これ以外にも、XFEL 単粒子構造解析が実現するまでには、効率よく測定を行うための実験系の構築、X 線強度とパルス幅と分子の放射線損傷の関係の解明、試料周りの水からの散乱の影響、生体分子がもつ構造多形への対応、ビームストップなどの実際の検出器形状を反映したより現実的なデータ処理など、ソフト面とハード面が密接にかかわった取り組むべき難しい課題が山積している。しかし、革新的な技術が開発されるその数年前には、不可能とさえ思われた様々な手法が世の中には数多く存在する。この事実を励みに、皆様のご協力のもと生体高分子の形を高分解能で見るための手法の開発を追求していきたい。

## 謝辞

本稿で紹介した内容は多くの方々との共同研究の成果です。特に単粒子構造解析アルゴリズムの開発は郷信広名誉教授（京都大学）、河野秀俊博士（JAEA）との共同研究であり、アルゴリズムの京コンピュータへの実装は、初井宇記博士（RIKEN）、城地保昌博士（JASRI）、石川裕教授（東京大学）、新井淳也氏（東京大学）、ならびに堀敦史博士をはじめとした理化学研究所・計算科学研究機構の方々との共同研究です。謹んで感謝いたします。

## 参考文献

- 1) R. Neutze, R. Wouts, D. Spoel, E. Weckert and J. Hajdu: Nature, VOL 406, 17, August (2000).
- 2) A. Tokuhiwa, J. Taka, H. Kono and N. Go: Acta. Cryst. A68, 366 (2012).
- 3) D. Sayre: Acta. Cryst. 5, 843 (1952).
- 4) R. W. Gerchberg and W. O. Saxton: Optik 35, 237 (1972).
- 5) J. R. Fienup: App. Opt. 21, 2758-2769 (1982).
- 6) Y. Takahashi, N. Zetsu, Y. Nishino, R. Tsutsumi, E. Matsuura, T. Ishikawa and K. Yamauchi: Nano. Lett. 10, 1922-1926 (2010).
- 7) J. Miao, K. O. Hodgson, T. Ishikawa, C. A. Larabell, M. A. Le Gros and Y. Nishino: Proc. Natl. Acad. Sci. U.S.A. 100, 110 (2003).
- 8) H. Jiang, *et al.*: PNAS. vol. 107, no 25, 11234-11239 (2010).
- 9) Y. Nishino, Y. Takahashi, N. Imamoto, T. Ishikawa and K. Maeshima: Phys. Rev. Lett. 102, 018101 (2009).
- 10) C. Song, *et al.*: Phys. Rev. Lett. 101, 158101 (2008).
- 11) M. M. Seibert, *et al.*: Nature 470, 78-81 (2011).
- 12) S. Kassemeyer, *et al.*: Opt. Exp. Vol. 20, Issue 4, pp. 4149-4158 (2012).
- 13) G. Huldt, A. Szoeké and J. Hajdu: J. Str. Biol. 144, 219-227 (2003).
- 14) 矢橋牧名, 石川哲也編: XFEL/SPRING-8 Beamline Technical Design Report Ver. 1.0, 理研・合同推進本部利用グループ (2008).
- 15) L. Schlichting and J. Miao: Cur. Opi. in Str. Biol. 22, 613-626 (2012).
- 16) H. Mimura, *et al.*: Nature Phys. 6, 122-125 (2010).

著者紹介



徳久 淳師

理化学研究所 播磨研究所 放射光科学総合研究センター XFEL 研究開発部門  
ビームライン研究開発グループ データ処理系開発チーム 基礎科学特別研究員  
E-mail: tokuhisa@spring8.or.jp

専門：生物物理

【略歴】

2007年奈良先端科学技術大学院大学物質創成科学研究科物質創成科学専攻博士課程修了，博士（理学）。同年日本原子力研究開発機構博士研究員，2010年より現職。

## Possibility of atomic-resolution single bio-molecular imaging with X-ray free electron laser

**Atsushi TOKUHISA** RIKEN Harima Institute, RIKEN SPring-8 Center, XFEL Research and Development Division, Beam Line Research and Development Group, Data Acquisition Team  
1-1-1, KOUTO, SAYO-CHO, SAYO-GUN, HYOGO, 679-5148 JAPAN

**Abstract** We report our recent study on the development of algorithm for imaging single bio-molecules by using the X-ray free electron laser. In this method, high intensity coherent femto-second X-ray laser is irradiated onto a target of a single bio-molecule with random orientation, and two-dimensional diffraction intensity patterns are recorded repeatedly, each time for a different orientation. The algorithm enables to extract signals necessary to construct an atomic-resolution three-dimensional structure from data deeply immersed in the quantum noise. We also report the present status of our effort of implementing the algorithm in the K computer, especially for classification of many diffraction patterns according to similarity of orientations.