

研究終了報告書

「高分解能データの統計的推定による超高精細結晶構造解析の開拓」

研究期間： 2017 年 10 月～2021 年 3 月

(新型コロナウイルス感染症の影響を受け 2021 年 9 月まで延長)

研究者： 星野 学

1. 研究のねらい

本研究課題のねらいは、回折データを発生可能な“結晶学と統計数理・情報科学の融合技術”を開発し、計測困難な回折データを補完した結晶構造解析を実行可能にすることである。

結晶構造解析は、計測した回折データセットが与える構造因子をフーリエ変換することで得られる電子密度分布を元にして、結晶中の原子配置(結晶構造)を得るための技術である。十分な強度のX線と結晶性の良い試料を用いて、時間をかけて計測した回折データセットを使う場合には、個々の原子の位置や熱揺らぎの大きさを評価可能な結晶構造を得ることができる。しかしながら、原理的に強度が弱い高分解能回折データ(回折角 θ とX線波長 λ で与えられる分解能 $\sin \theta / \lambda$ の値が大きい回折データ)が不足すると、回折データセットから構造因子を得ることが不可能になる。構造因子が取得困難な場合には既知構造を参考にしてフーリエ変換を行うことも可能であるが、変換によって得られる電子密度分布は試料本来の分布よりもピークが鈍化し、ピーク周辺のノイズも増大する(級数打ち切り効果による影響が顕著になる)ため、解析によって得られる原子位置の精度が著しく低下する。

本研究課題では、計測困難な高分解能回折データを、統計数理および情報科学の技術を利用して取得し、上記の解析不可能あるいは解析精度低下の問題を回避可能にする。具体的な目標としては、タンパクの結晶構造解析における原子位置精度の向上と、不可逆過程の結晶構造解析の実現の2つを設定した。前者は放射線損傷の影響により、後者は放射光リングから発生されたX線のシングルパルス切り出しの計測実験条件に起因して、計測可能なデータが低分解能回折データに限定される。本研究課題では、この低分解能回折データから回折データに内在する物理量をベイズ推定し、回折データ発生モデル(任意の分解能の強度の回折データの検出を確率で表現した分布)を構築する。次に、モデルから計測不可能な高分解能領域における回折データの強度をサンプリングして、低分解能回折データを用いて事前に計算した同領域における回折データの指数との数理最適化を行う。最適化によって得られた強度と指数の組み合わせを、高分解能回折データとして計測した低分解能回折データと合わせて結晶構造解析に使用することで、計測データだけでは不可能な構造因子の取得あるいは原子位置の精度向上を達成する。

2. 研究成果

(1) 概要

本研究課題で開発した“結晶学と統計数理・情報科学の融合技術”では、第1段階として、任意の分解能 $\sin \theta / \lambda$ で回折強度 I が計測される確率分布関数 $P(I)$ のパラメータである Σ を取得する。 Σ は分布の期待値であり、原子散乱因子の2乗の総和と Debye-Waller 因子と

の積によって与えられる。原子散乱因子は任意の分解能においては定数であり、Debye-Waller 因子は試料中の原子の等方性温度因子の平均値 $\langle B \rangle$ と $\sin \theta / \lambda$ で記述されることから、 $\langle B \rangle$ を得ることで Σ を計算できる。 $\langle B \rangle$ を得る手段として本研究課題では、計測した回折データの強度平均値 $\langle I \rangle$ と $\langle B \rangle$ の関係式である Wilson の式を規格化した確率分布関数を尤度関数として用いたベイズ推定 [$P(\langle B \rangle | \langle I \rangle) \propto P(\langle I \rangle | \langle B \rangle) P(\langle B \rangle)$ で得られる事後分布から $\langle B \rangle$ を推定]を行った。ベイズ推定によって推定した被験試料の $\langle B \rangle$ を含む Debye-Waller 因子と、被験試料を構成する原子の原子散乱因子の値から、任意の分解能における Σ の値を計算した。

Σ を得た後には第 2 段階として、回折データの強度発生を行う。この回折データとは強度と指数の組であり、指数は計測可能な低分解能回折データから決定可能な結晶の周期単位長と対称性を元に計算することで得られる。注目する分解能における Σ を含む $P(I)$ から、マルコフ連鎖モンテカルロサンプリングのアルゴリズムを用いて、事前に計算した上記の分解能近傍に含まれる回折データの指数の数だけ強度を発生させた。

発生させた強度は、第 3 段階として、指数との組み合わせを適切に行い、結晶構造解析に利用可能な回折データセットを得る。強度と指数を適切に組み合わせる手段として、本研究課題では数理最適化のアルゴリズムを応用した。最適化計算によって得られた回折データセットを結晶構造解析に利用したところ、電子密度分布において原子位置のピークが先鋭化したことが確認できた。以上により、本研究課題の目的である高分解能回折データの欠損（級数打ち切り効果の影響増大）によって低下した原子位置精度の向上を達成した。

新型コロナウイルス感染症の感染拡大の影響による研究遅延を補うために、6 ヶ月間研究期間を延長した。延長期間では、計測実験条件の制約による高分解能データが欠損するシングルショット X 線回折計測実験を実施した。当該計測実験を放射光施設にて実施するまでの準備期間には、タンパク結晶試料から計測した回折データに対して本研究課題で開発した融合技術の適用も並行して行った。

(2) 詳細

研究テーマ A 「少ない回折データを用いた試料固有パラメータ $\langle B \rangle$ のベイズ推定」

試料固有のパラメータである $\langle B \rangle$ を得るために、結晶学において $\langle I \rangle$ と $\langle B \rangle$ の関係式として知られる Wilson の式 [$\langle I \rangle / \Sigma f^2 = (1/C)^2 \exp[-2\langle B \rangle (\sin \theta / \lambda)^2]$ 、C: スケール因子、f: 原子散乱因子]に注目した。本研究課題では、Wilson の式を「 $\sin \theta / \lambda$ において $\langle I \rangle / \Sigma f^2$ 回発生する事象の度数分布」として捉え、規格化して確率分布関数を得ることでベイズの理論における尤度関数 $P(\langle I \rangle | \langle B \rangle)$ として利用可能にした。事前分布は一様分布と正規分布の 2 種類を検討し、推定値に対する事前分布の影響は無視できるほど小さかったことから、最尤推定で解析的に推定値を計算する目的で一様分布を用いることにした。結晶構造解析の標準試料である 2-ジメチルスフラニリデン-1,3-インダンジオン(YLID)を被験試料として、予備計測データ(202 個の回折データ、通常は試料の状態確認のために 1 分程度で計測するデータ)を用いて推定した $\langle B \rangle$ の値は、 1.64 \AA^2 であった。この値は、同じ結晶で 4015 個の回折データ(4 時間程度で計測)を計測して結晶構造解析を行った結果として得られた $\langle B \rangle = 1.82(46) \text{ \AA}^2$ と、解析誤差範囲内で一致した。この推定方法における $\langle B \rangle$ の推定精度の高さを実証する

目的で、2つの異なる溶媒条件から析出させた[12]-シクロパラフェニレンの単結晶について、予備計測データから $\langle B \rangle$ を推定したところ、同一分子の結晶にもかかわらず $\langle B \rangle$ の値はそれぞれ 3.50 \AA^2 と 3.13 \AA^2 であった。この違いは、結晶に内包された溶媒分子が異なることを強く示唆している。そこで同じ結晶を用いて長時間かけて多数の回折データを計測し結晶構造解析を行ったところ、 $\langle B \rangle$ の推定値の違いから示唆された通り、2つの結晶は内包する溶媒分子が異なる[12]-シクロパラフェニレンの結晶であった(図1)。この結果から、本研究課題で開発したベイズ推定技術が、少ない回折データからであっても結晶構造解析結果に対応した $\langle B \rangle$ の推定値を与えることが確認できた。本成果は論文(代表的な論文発表1)として報告している。

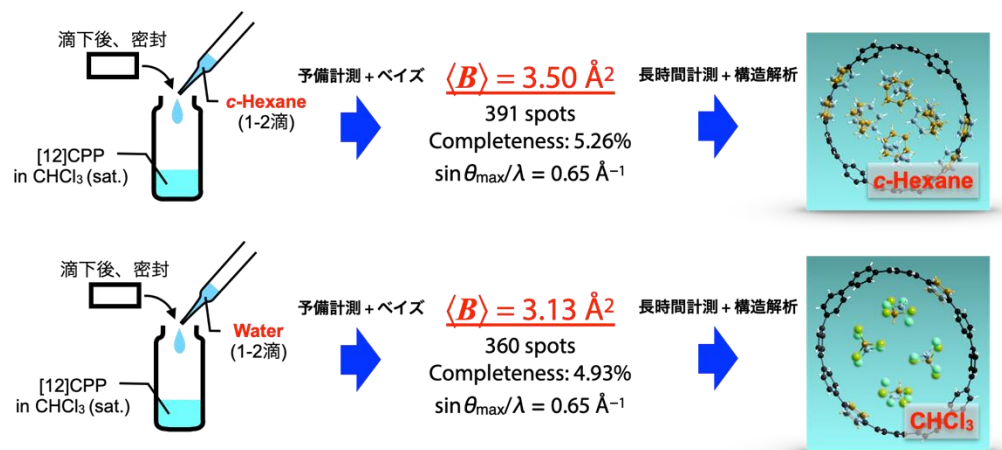


図 1. 異なる溶媒条件で析出させた[12]-シクロパラフェニレンの回折データを用いたベイズ推定と結晶構造解析結果

研究テーマ B 「低分解能データから高分解能データを発生させた結晶構造解析の実証」

注目する分解能近傍で計測が期待される回折強度を、その分解能における $P(I)$ からのマルコフ連鎖モンテカルロサンプリングによって発生させた。発生させた強度に指数を付与して回折データを得るために、強度と指数の組み合わせ最適化を行った。YLID の単結晶から計測した $\sin \theta / \lambda < 0.35 \text{ \AA}^{-1}$ の回折データセットから $\langle B \rangle$ を推定し、 $\sin \Delta \theta / \lambda = 0.005 \text{ \AA}^{-1}$ ずつ高分解能領域の回折強度を発生させて最適化を行うことを繰り返し、 $\sin \theta / \lambda = 0.65 \text{ \AA}^{-1}$ までの回折データセットを得た。最適化した回折データセットを、計測した回折データセットに加える前と後の結晶構造解析結果の比較を図2に示す。計測した回折データセットだけでは、電子密度分布が原子分解能を持たず、既知の分子構造を剛体として(分子中の原子の相対配置を固定し、分子全体を電子密度に当てはめて)解析する必要があった。一方で、計測データセットに最適化した回折データセットを加えると、電子密度分布が先鋭化して原子分解能の観察が可能になった。残渣電子密度が大きいために、結晶構造解析における構造精密化では原子間距離が理想的な値から逸脱しないように束縛をかける必要があるが、個々の原子ごとに位置を精密化可能な程に原子位置精度を高めることができた。

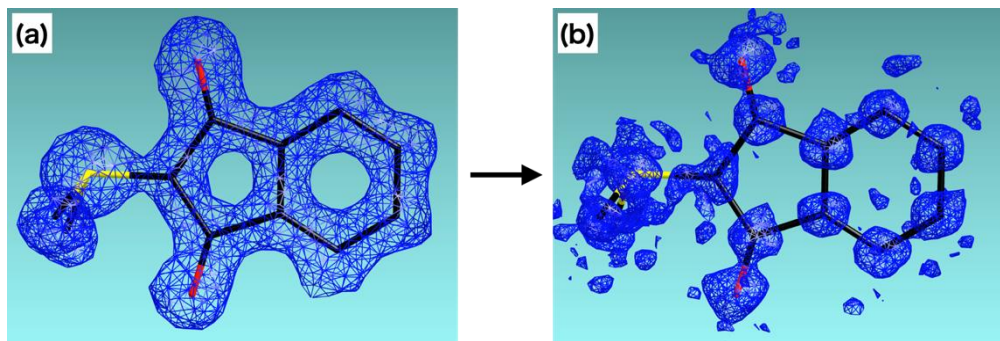


図 2. 低分解能の計測データだけ(a)と、計測データに高分解能の発生データを加えたデータ(b)による結晶構造解析によって得られた電子密度分布

研究テーマ C 「回折強度発生を利用した計測実験条件最適化の提案」

本研究課題で開発した技術のうち、回折データの強度発生の特徴を生かすことができる研究実例を示す目的で、精密な電子密度分布を与える回折データセットを得るための X 線露光時間の最適化を提案した。X 線露光時間は回折強度とその誤差(σ)の比である I/σ の目標値から計算する。本研究テーマでは、予備計測データから研究テーマ 1 の技術で $P(I)$ を得て、研究テーマ 2 で行った強度発生によって得られた I が誤差範囲内で区別可能である I/σ を、精密な電子密度分布を与える回折データセットが満たすべき I/σ の基準値として提案した。実際に提案した基準値を満たす場合と下回る場合の回折データセットを計測し、結晶構造解析結果を比較したところ、提案した I/σ の基準値を満たすことで理論的に妥当な電子密度分布が可視化されたことを示した(図 3)。本研究テーマの成果は、研究テーマ A を発表した論文(代表的な論文発表 1)に含まれる。

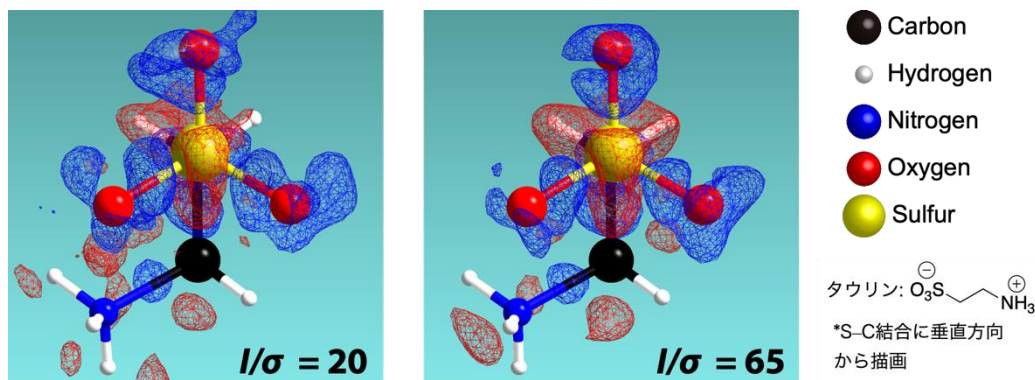


図 3. X 線露光時間最適化による電子密度分布の精度向上。タウリンについて、球対称電子密度分布からの変形(正:青、負:赤)を表示。 $I/\sigma = 65$ では分布が理想的な 3 回対称性に近い。

研究テーマ D 「回折強度を発生可能な分解能条件の検討」

本研究課題における技術開発を通じて、適用可能な回折データの分解能下限値と、発生可能な回折データの分解能上限値が明確になった。

適用可能な回折データの分解能下限値は、本技術が Wilson の式に立脚しているため、

Wilson の式の適用範囲と関係する。例えば、回折強度に対する消衰効果(回折 X 線が別の結晶面で散乱を起こす、あるいは入射 X 線が結晶表面から順次散乱されて強度が減衰する、などの効果)の影響が大きい低分解能回折データでは、 $\langle B \rangle$ の値を正しく推定できなかった(代表的な論文発表1にて報告)。また、卵白リゾチーム結晶の回折データでは、タンパク結晶に特有の約 4 Å の相互作用(水素結合)が多く存在することに起因して、 $\sin \theta / \lambda = 0.125 \text{ \AA}^{-1}$ 近傍で回折データが Wilson の式から逸脱しているため、この分解能近傍のデータはベイズ推定を適用できないことがわかった。これらのデータを考慮すると、本技術を適用するデータの分解能下限値は $\sin \theta / \lambda = 0.165 \text{ \AA}^{-1}$ が妥当であり、この分解能よりも低分解能な回折データは現状では本技術が適用できないと言える。

発生可能な回折データの分解能上限値は、推定した $\langle B \rangle$ から計算して得られる $\langle I \rangle$ の値とその不確かさによって決まる。回折強度は正の値であることから、 $\langle I \rangle$ が誤差範囲内で負の値を取り得ることは不適である。実際に卵白リゾチーム結晶の回折データに本技術を適用して解析したところ、 $\sin \theta / \lambda = 0.43 \text{ \AA}^{-1}$ を超えると $\langle I \rangle$ が誤差範囲内で負の値になる結果が得られた。この上限値を実際の計測データと比較したところ、計測実験開始段階では $\sin \theta / \lambda = 0.43 \text{ \AA}^{-1}$ の回折データが検出されていたが、実験の進行と共に放射線損傷が起り、最終的な回折データセットとしては、 $\sin \theta / \lambda = 0.40 \text{ \AA}^{-1}$ まで低分解能化した。以上の結果から、本技術で発生可能な回折データの分解能上限値は、放射線損傷がゼロであれば計測できた分解能の上限値に相当すると言える。

3. 今後の展開

本研究課題で開発した計測不可能な回折データを取得可能にする技術によって、従来は結晶構造解析が不可能であった物質や現象について、結晶構造解析による原子分解能の構造評価が適用可能になる。例えば数マイクロメートルスケールの微小結晶は、回折強度が結晶の体積に比例する性質を持つため、放射光を利用して高分解能回折データを計測することは困難である。高分解能データを発生させて微小結晶の結晶構造解析が実現できれば、医薬品代謝物や貴重天然抽出物などの痕跡量物質の構造評価に結晶構造解析が利用可能になる。また、不均一系の触媒の構造評価は、粉末 X 線回折計測によって行われるが、粉末回折データは原理的に分解能が上がるほど回折データの重なりが多くなり、加えて不均一系故の多成分の回折データが混合したデータになるため、結晶構造解析を行うことは不可能である。しかし本研究の技術が利用可能であれば、高分解能回折データの計測よりも低分解能回折データの分離を優先させて計測(長波長の X 線の利用や試料と検出器の距離を離れた計測実験で実現可能)すれば、計測後に各成分の回折データを分けて高分解能化し、結晶構造解析を実行することができる。今後は上記のように、本研究課題によって開発した技術を、結晶構造解析の適用範囲外の物質へと適用し、物質開発を原子分解能の構造評価で底上げする基盤技術として確立させる。

4. 自己評価

本研究の目的である「計測した低分解能回折データを元にして、高分解能回折データを発生させる」技術開発について、結晶学とベイズの理論および数理最適化との融合によって概ね達成す

ることができた。現時点では、残渣電子密度の水準が高い、具体的な物質に対する適用例が限定的、といった課題が残るが、「高分解能回折データの欠損」という計測限界の影響により電子密度分布の原子分解能が失われる問題を解決することができたことは、評価に値する。さきがけ研究費の執行については、高額装置である単結晶X線回折装置の導入ができたことにより、本研究開発を計画通りかつ臨機応変に推進することができた。研究の推進においては、領域内研究者と積極的に連携し、中西義典研究者(2期生)との共同研究は論文発表へと至る成果を得ることができた。この研究成果はプレスリリースを行い、その後、新聞報道やメディアインタビューへと繋がるなど、社会的に大きなインパクトを与える結果となった。また、領域会議やクラスター会議を通じて積極的に情報技術に対する知見を深め、その結果、数理最適化の技術を回折データ取得のための重要なプロセスに導入する発想を得ることができた。この最適化プロセスの導入においては、小野峻佑研究者(1期生、最適化問題の専門家)や加藤健一研究者(3期生、情報エントロピーを用いた回折データ解析の専門家)との議論を行うなど、さきがけのネットワークを活用して効果的に推進できた。その一方で、自身の専門である計測研究だけでなく、融合技術開発の方針設計やプログラムコード作成といった情報研究部分についても自らが行うことで、個人型研究であるさきがけの特性に沿って、自らを情報計測の専門家として成長させるように努めた。その他にも、さきがけコンバージェンスキャンプや SciFoS 活動にも参加して、自らの研究開発を社会実装する上で必要な知見を獲得するだけでなく、産学官の垣根を超えた人脈を得ることもできた。これらの知見や人脈は、ポストさきがけとして本研究課題の成果を、我が国の研究開発力の向上と経済効果を高める物質構造評価基盤技術として社会実装する上で、機能することが期待できる。加えて、領域内クラスター会議にて領域外の情報技術応用を推進する研究者との交流機会の設計や、研究実施場所である理化学研究所で情報計測関連のオンライン会議(情報計測領域の研究者3名が招待講演を実施)を企画し、情報計測分野の開拓と発展に対して貢献した。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数:1件

1. Manabu Hoshino, Yoshinori Nakanishi-Ohno, Daisuke Hashizume, “Inference-assisted intelligent crystallography based on preliminary data”, *Sci. Rep.*, 2019, **9**, 11886, <https://doi.org/10.1038/s41598-019-48362-3>

通常は被験試料のチェックのために予備的に計測する少数の回折データをベイズ推定に用いて、未計測な領域も含めたデータ分布のモデルを得る方法論を構築した論文。この方法論を用いた、「結晶構造解析前に結晶内の分子を区別」と「高精度の結晶構造解析結果を得るための計測条件提案」の2つの実用例も報告した。

(2) 特許出願

研究期間累積件数:0件(特許公開前のものも含む)



(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

招待講演(国際会議)

1. Manabu Hoshino, “Evaluation of ‘*to-be-collected*’ diffraction data using information technology”, CEMS Topical Meeting Online – Advanced Visualization of Electrons, Atoms, and Molecules, Online, 2020 年 10 月.
2. Manabu Hoshino, “Bayesian Inference using X-ray Diffraction Data: Estimation of Crystal Structure Parameter and Application to Host–Guest Chemistry”, CEMSupra 2019, Tokyo, Japan, 2019 年 12 月.

招待講演(国内会議)

1. 星野学, “動く構造を可視化する単結晶構造解析のための計測・解析法開発”, 東京大学物性研究所 LASOR セミナー, 柏, 2019 年 10 月.
2. 星野学, “少ない X 線回折データから結晶構造の“乱れ”や“ゆらぎ”の情報を抽出”, 日本分析化学会第 67 年会, 仙台, 2018 年 9 月.

解説記事

1. 星野学, “結晶学とベイズ理論: 強度補正や構造パラメーター推定から回折データの統計モデリングまで”, 日本結晶学会誌, **62**, 217–218 (2020).

プレスリリース

1. 科学技術振興機構、理化学研究所、東京大学 “熟練の研究者の「勘と経験」を誰でも簡単に再現 ～たった数分で単結晶構造解析の結果の事前評価が可能に～” 2019 年 8 月 22 日

新聞報道

1. 日刊工業新聞, X線結晶構造解析、事前評価を簡易化 理研が技術, 2019 年 9 月 19 日.
2. PC WATCH, 理研・東大などが、熟練研究者の「勘と経験」をコンピュータで再現。単結晶構造解析を高速化, 2019 年 8 月 23 日 (*同日の Yahoo ニュースでも掲載).
3. 化学日報工業, 単結晶構造解析を高効率・高精度化 理研-東大, 2019 年 8 月 22 日 (*1面掲載).

メディアインタビュー

1. “高分解能データの推定技術で、超高精細な結晶構造解析を実現する”, 株式会社 経営共創基盤(IGPI) Top Researchers (<https://top-researchers.com/?p=3828>), 2020 年 5 月 8 日.