

実験データから「見えない内部状態の分布」を再構成する新手法を開発 —情報幾何の応用により、生体分子の動きの可視化に成功—

慶應義塾大学理工学部物理学科の荳口友隆専任講師、京都大学複合原子力科学研究所の井上倫太郎教授、杉山正明教授らの研究グループは、実験データから直接観測できない「内部状態の分布」を高精度で再構成する新しい計算手法を開発しました。

実験データから対象の内部状態の分布を推定する問題は「逆問題」と呼ばれ、生命科学、物理学、材料科学といった幅広い分野で重要な研究課題となっています。本研究では、情報幾何に基づき、分布空間の多様体構造を厳密に考慮した新しい最適化手法を開発しました。これにより、従来法では困難であった高精度な分布再構成が可能となり、実験データから溶液中におけるタンパク質の動きを可視化することにも成功しました。本成果は、実験データから「見えない内部状態の分布」を可視化するための新しい計算科学的基盤となるものであり、今後、さまざまな分野における逆問題解析への応用が期待されます。

本成果は、2026年6月15日付で『Physical Review Research』に掲載されました。

1. 本研究のポイント

- ・分布空間がもつ多様体構造を考慮した新しい最適化計算手法、**manifold-based gradient descent (MBGD)** 法を確立しました。
- ・通常は非対称で距離とはみなされない **Kullback-Leibler (KL)** ダイバージェンスが、特定の幾何学的条件のもとでは実質的に距離として機能することを示しました。
- ・新手法を小角 X 線散乱 (SAXS) データの解析に応用し、実験データや分子動力学 (MD) シミュレーションだけでは捉えにくかった溶液中タンパク質の動きを可視化できることを実証しました。

2. 研究背景

多くの物理・化学・生命科学の系では、対象は単一の状態ではなく、複数の内部状態がある確率で共存しています。このような「内部状態の分布」は、系の性質や機能を決める重要な要素です。たとえば、溶液中のタンパク質はさまざまな立体構造をとり、それらの構造状態の分布（構造アンサンブル）が機能と密接に関係しています。一方、実験で得られるデータは、プローブに対する各内部状態の応答を、その分布にわたって平均した情報です。そのため、実験データから内部状態の分布を直接知ることはできず、観測データをもとに元の分布を推定する「逆問題」を解く必要があります。

この逆問題では、内部状態の分布は「総和が1に規格化されていること（正規化条件）」、「各確率が負にならないこと（非負性）」という制約を満たす必要があります。そのため、分布空間を通常のエウクリッド空間と仮定する従来の最適化手法では、これらの制約によって決まる分布空間の幾何学的構造を十分に反映できず、正しい分布への収束が困難でした。このことが、幅広い逆問題に適用できる汎用的な解析手法の開発を難しくしていました [1,2]。

本研究では、分布空間を幾何学的な「多様体」として扱う情報幾何学を逆問題に導入しました。分布空間の多様体構造に沿って内部状態の分布を更新することで、実験データと整合する分布を高精度に再構成し、これまで直接見ることはできなかった内部状態の分布を可視化する新しい計算手法を実現しました。

3. 研究内容・成果

本研究ではまず、分布空間における KL ダイバージェンスの幾何学的性質に注目しました。KL ダイバージェンスは2つの確率分布の違いを測る量として広く用いられていますが、一般には対称性をもたないため、数学的に厳密な意味での距離ではありません。本研究では、分布空間上の微小な変位が指数接続の測地線に沿う場合には、KL ダイバージェンスが実質的に距離として振る舞うことを明らかにしました。この結果に基づき、分布空間の多様体構造に沿って確率分布を更新する MBGD 法を定式化しました。MBGD 法では、この KL ダイバージェンスの幾何学的性質を利用して確率分布を更新するため、正規化条件と非負性を常に満たしながら、実験データとの一致度を表す目的関数を安定に最小化できます。

手法の有効性を検証するため、溶液中の粒子サイズ分布を SAXS データから再構成するモデル計算を行いました。その結果、MBGD 法は真の粒子サイズ分布を高精度に再構成できた一方、従来法では計算が不安定となり、真の分布を再構成できませんでした。この結果は、確率分布を対象とする逆問題では、分布空間の幾何学的構造を正しく扱うことが本質的に重要であることが示しています。

さらに、本手法をタンパク質の構造アンサンブルの再構成に適用しました。タンパク質は溶液中で単一の構造をとるのではなく、複数の構造状態の間を揺らぎながら機能しています。しかし、結晶構造解析から得られるのは静的な構造であり、また MD シミュレーションも計算に用いる力場関数の近似やそのパラメータに依存するため、動きに関する実験データを必ずしも再現できません。

そこで、MD シミュレーションから得られた構造アンサンブルを SAXS 実験データと整合するように MBGD 法で最適化したところ、真の構造アンサンブルを高精度に再構成できることを確認しました。さらに、本手法をグルコサミン・リン酸化酵素 (SjGlcNK) の実験 SAXS データに適用した結果、溶液中における酵素の動きの可視化に成功し、基質認識には構造アンサンブルが重要な役割を果たすことを明らかにしました (図 1)。これは、実験データと MD シミュレーションを MBGD 法によって統合することで、それぞれ単独では捉えることが難しかった溶液中タンパク質の動きを可視化できることを示した成果です。

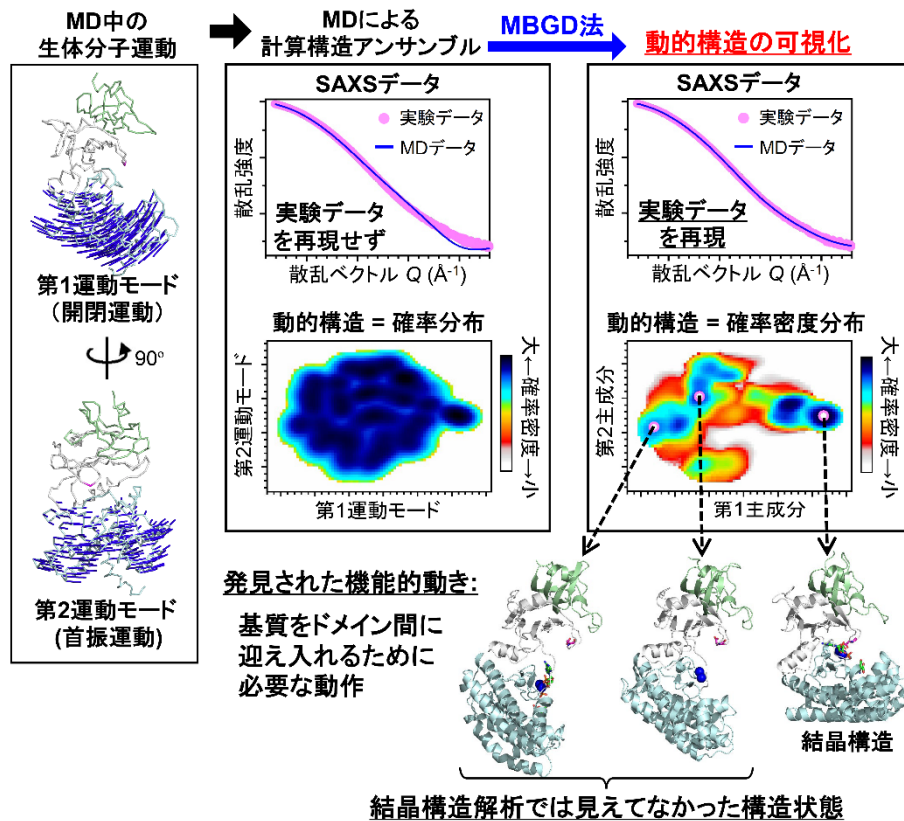


図 1. MBGD 法による実験データと MD シミュレーションの統合解析：可視化された SjGlcNK の動的構造

4. 今後の展開

本成果は、実験データから直接観測できない「内部状態の分布」を読み解くための新しい数学的・計算科学的基盤を提供するものです。

今後は、SAXS に加えて中性子散乱、核磁気共鳴、電子共鳴といった複数の実験データを MBGD 法によって MD シミュレーションと統合していきます。これにより、核酸や超分子複合体、天然変性タンパク質など、より複雑な動きを行う生体分子の可視化が期待されます。このような可視化によって得られる動的構造データを蓄積すれば、将来的には機械学習 (AI) を用いて、生体分子の機能的動きの予測に繋げることも可能になるでしょう。さらには、生体分子に限らないソフトマター系や材料科学への適用も期待されます。

<参考文献>

1. J. Köfinger, L. S. Stelzl, K. Reuter, C. Allande, K. Reichel and G. Hummer, Efficient ensemble refinement by reweighting, *J. Chem. Theory Comput.* **15**, 3390 (2019).
2. S. Bottaro, T. Bengsten and K. Lindorff-Larsen, Integrating molecular simulation and experimental data: A Bayesian/maximum entropy reweighting approach, *Methods Mol. Biol.* **2112**, 219 (2020).
3. S. -I. Amari and H. Nagaoka, *Methods of information geometry*, Vol. 191 (Oxford University, New York, 2000).

<原論文情報>

Tomotaka Oroguchi, Rintaro Inoue, Masaaki Sugiyama, Manifold-based transformation of probability distributions: Application to the inverse problem of reconstructing distributions from experimental data, *Physical Review Research* **8**, 023291 (2026).

• DOI: 10.1103/vkck-1pcm