

平成 30 年 6 月 7 日現在

機関番号：15101

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00404

研究課題名(和文) タンパク質ダイナミクス解明のためのロバストなトラジェクトリ解析法の開発

研究課題名(英文) Robust methods for analyzing trajectories for protein dynamics

研究代表者

網崎 孝志 (Amisaki, Takashi)

鳥取大学・医学部・教授

研究者番号：20231996

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：タンパク質の静的構造の比較や動的変化(ダイナミクス)の研究には信頼のおける構造重ね合わせ法が必要である。本研究では、結晶構造の乱れに関する占有率と共分散行列の非対角成分の両方を統一して扱える理論的枠組みを開発し、プログラムmmfitcとして実装した。分子動力学シミュレーション結果の解析など、対象構造が多数の場合に、特に有用であると思われる。また、構造が大きく異なる分子が紛れ込んでいる場合に、その影響を受けにくい構造重ね合わせ法も開発した。結晶構造の比較など、対象とする構造があまり多くない場合に有用であると思われる。

研究成果の概要(英文)：Accurate structural superposition is a necessity for studying static and/or dynamic structures of proteins. In this study, we developed a theoretical framework for structural superposition that concerns occupancy factors in crystal structures and off-diagonal components of covariance matrices. The method was implemented as a computer program, mmfitc. It is useful for analyzing, for example, molecular dynamics data, where a great number of structures are subject to. For superposition, we also developed another method which is insensitive to outlier protein molecules to some extent. The method is useful for superposition of a moderate number of structures such as in the case for comparing the crystal structures.

研究分野：計算構造生物学

キーワード：構造重ね合わせ 共分散 占有率

## 1. 研究開始当初の背景

(1) タンパク質の構造と機能の関係の解明は基礎科学としても、また、創薬等の実用面でも、チャレンジングで重要な課題である。最近では、ダイナミクスと機能の間の関係が注目されている。その解明において、分子動力学シミュレーションは期待される手法のひとつである。そこで得られるのはデカルト座標系でのトラジェクトリであるが、そこから、構造の動き、ゆらぎ、多様性、遷移を調べるため、幾何学的クラスター分析による分類や主成分・独立成分分析による動的成分の同定などが行われている。

(2) このとき、前処理としての構造の重ね合わせ（並進と回転の自由度を除去）を行いデカルト座標で処理を行うか、重ね合わせを必要としない二面角などの内部座標で解析を行うのかの選択肢がある。天然変性タンパク質のように代表的静的構造が定義できないものなどを別にすると、内部座標に比べてデカルト座標での結果は直観的解釈が容易であるケースが多いため、優れた構造重ね合わせ法の開発は静的構造・動的構造両面の研究においてきわめて重要である。

## 2. 研究の目的

(1) 本研究は、結果の解釈の容易なデカルト座標系での解析を前提とし、精密なダイナミクス解析の前処理として必要な構造重ね合わせの改良を主目的としていた。まず構造重ね合わせの欠点として考えられるのが統計学的なロバスト性の問題である。たとえば、異分散性と外れ値の問題があげられる。前者については、いくつかの方法が報告されていたが、実質的に共分散（原子間相関）までもを扱うことはできなかった。また、後者の外れ値の問題については、これまでにほとんど報告がなされていなかった。

(2) そこで本研究は、タンパク質立体構造の重ね合わせにおける、これらのロバスト性についての改良を理論面・実践面で行うことを目的とした。なお、そこで取り入れようとしていた方法の一部は、ダイナミクス解析に用いられる主成分分析にも応用可能と考えられるため、当初、その改良を試みることも計画したが、後述の占有率の問題への対応を優先させた。

## 3. 研究の方法

(1) 構成尤度に基づく理論的枠組みの構築：当初は分子動力学シミュレーションが生成するトラジェクトリーデータのみを想定していたが、本手法を公開した場合のインパクトならびに今後の展開の可能性も考慮し、結晶構造へも一般化した理論的枠組みに発展させる計画とした。具体的には占有率への対応、すなわち、存在確率が1未満で、単数または複数のサイトを占める原子への対応である。このため、当初予定していた三次元行列正規分布のもとでの最尤法は、そのまま

では利用できなかった。そのため、独自の手法を考案したが、のちに、それは構成尤度に基づく手法とみなすことができることがわかった。同時に、この手法は異分散だけでなく、原子間相関を考慮した立体構造重ね合わせとできるように設計した。

(2) 詳細な数値シミュレーションによる検証：上述の理論的枠組みによる立体構造重ね合わせ法 (mmfitc) を実装し、その性能の検証数値実験を行った。共分散行列の非対角成分（原子間相関）を考慮する方法 [共分散法] と、対角成分のみを考慮する方法 [異分散法]、THESEUS [従来法] を比較した。

なお、実装にあたっては、C 基本ライブラリの他は外部ライブラリは利用せず、特に、今後の展開にあたって重要な基礎ツールとなると予想される特異値分解・固有値分解の部分も自作した。

(3) L21 ノルム重みによるロバスト重ね合わせ法の開発：従来の RMSD（立体構造の違いの測度の一種）は原子単位と分子単位の違いの階層構造になっている。これを利用して「外れ分子」の影響を受けにくい立体構造重ね合わせ法を考案し、実装した。

## 4. 研究成果

(1) 占有率1未満の原子を考慮する構造重ね合わせ法の開発

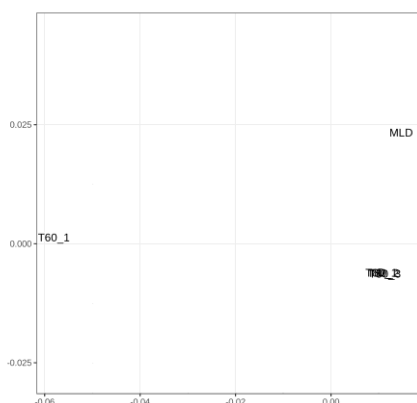
① 結晶構造のディスオーダー（構造乱れ）には、天然変性タンパク質にみられるように規則的な静的構造をもたないものと、一部の原子あるいは原子群が複数のサイトを占める場合がある。いずれもタンパク質構造のダイナミクスを反映しており、その現象の解析自体が有意義である。前者のようなダイナミクスは本研究の範疇ではないため、後者の場合のみを考えた。構造変化により複数の配置をとる場合などである。

② 本研究でのアイデアは、占有率が1または0だけの原子で構成された構造 (conformer) を考え、それに占有率を分配して計算した重みを乗じて合成することにより元の構造 (archetype) を構成できるというものである。実際に行う操作は逆向きで、archetype を分割して conformers を得る。そのために占有率の接頭部和に基づくアルゴリズムを考案した。

③ 原子位置には、三次元行列正規分布を仮定し、占有率が0の原子（位置が観測されていない原子）の座標については、全空間にわたって積分して消去し、周辺確率密度を求めた。これに占有率の分割から得られた重みを用いて構成尤度 (composite likelihood) とする方法とした。

④ 構成尤度の最大化は EM アルゴリズムに準拠した反復法とし、各構造 (conformer) を重ね合わせるための回転行列と並進要素を推定する方法とした。

⑤ 実際の適用例を示す。対象は hMTH1 (酸化スクレオチド分解酵素) の結晶構造 PDB エントリー 28 個で、総計 44 個の archetype が 303 個の conformers に分配された。この比 303/44 が比較的高く、占有率 0 超 1 未満の原子が多く含まれる例といえる。異分散の対角共分散行列を用いる場合 [異分散法] と共分散行列の一部の非対角要素も考慮する場合 [共分散法]、並びに、異分散最尤法に基づく従来法 (THESEUS-3. 3. 0, 2015) のそれぞれでこれらの構造の同時重ね合わせを行い、それぞれに



ついて平均構造を共分散逆行列での重み付き平均構造を求めた。これら推定法ごとの平均構造相互間の RMSD を求め、それを多次元尺度構成法により二次元図とした。

この図の MLD が従来法の結果である。右下の重なりには、共分散行列の対角成分のみを用いた場合と、ごく限定した非対角成分も用いた場合の結果が重なっている。この限定のためこれらの結果に大きな違いを生じず、重なって表示されている。限定の程度を緩めて、非対角成分の数がやや多い場合の結果が T60\_1 で示されている。従来法では占有率 1 未満の原子のすべてのサイトの情報は利用できないが、それを合理的に考慮した本提案法と比較すると、このように推定平均構造にある程度の違いがみられた。このことは構造重ね合わせ一般において、本手法の有用性・必要性を示している。

(2) 原子間相関 (共分散行列非対角成分) を考慮する構造重ね合わせ法の開発

① 前(1)で述べた構成尤度に基づく手法では、すべての構造に同一の共分散行列を仮定することにより、その非対角成分をも考慮した表式を比較的簡単に導出できる。この仮定はむしろ合理的と考えられ、また、その結果、分布を規定する共分散行列は全体で 1 個だが、欠測のある原子のパターンにより作業用共分散行列は異なる。これは一般化推定方程式 (GEE) においても、同様の定式化の例がみられる。

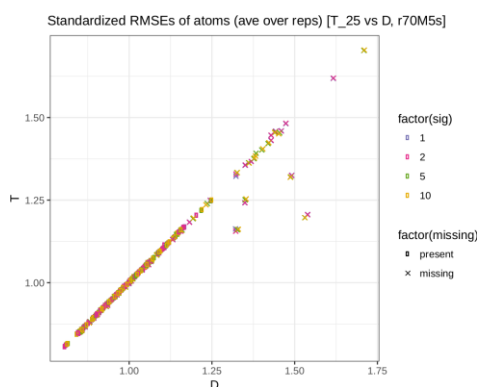
② もともと、構造データ数の規模から、推定方程式に利用できるような精度で共分散成分を求めることは不可能であるから、共分

散行列に対する何らかのモデル化あるいは正則化が必要である。本研究では、GEE に倣い、推定に利用する共分散行列を作業共分散行列と呼び、推定する標本共分散行列と区別する。そして作業共分散には、Stein タイプの縮小 (shrinkage) および相関行列上での閾値法と条件数を指標とする縮小を組み合わせた正則化についての検討を行った。

③ 詳細な数値シミュレーションの結果、以下の知見が得られた。

作業共分散行列の非対角成分は縮小により対角優位性を増大させるよりも、ノンゼロ成分の個数を限定するほうが効果的であった。相関を持たせたデータにおいては、閾値・条件数正則化による作業共分散を使用する方法が非対角成分をゼロとする異分散法よりも、平均構造、分散成分ともに正確な推定が可能であった。ただし、その改善の程度は大きくなく、構造数が数百以上の場合に有効であった。

また、欠測がある原子が含まれる場合でも、従来法に比べてロバストであることが示された。たとえば、以下の図は構造数 N=200、原子数 n=154 の場合の例で、平均構造の各原子の標準化平方誤差の平均値の平方根を異分散法 (横軸) に対して共分散法の結果を縦軸にプロットしたものである。欠測がある原子 (×) は大きな誤差を示しているが、相関に関わる原子については、いずれも直線の下側にシフトし、誤差が縮小しているのがわかる。

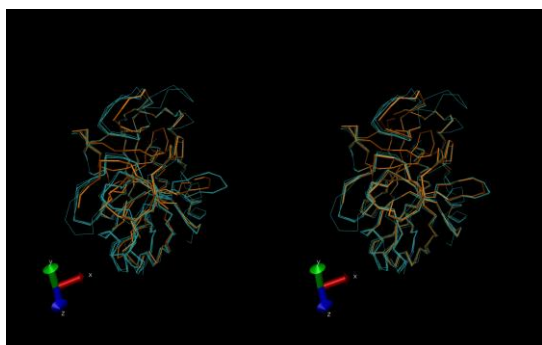


相関係数 (共分散、非対角成分) については、閾値を設けて相関の有無の一致度 (F1 スコア) で判定を行ったが、構造数が 50 以上の場合には異分散法よりも高い値を示した。また、比較に用いた THESEUS (の拡張機能で共分散法に相当) は、欠測がある原子の位置 (残基番号) の影響を受けたが、本研究の手法は非対角成分を考慮する方法でも、そうでない場合も、欠測原子の位置、また、欠測の程度にも比較的影響を受けにくかった。

(3) L21 ノルム重みを用いた「外れ構造」の影響を受けにくい構造重ね合わせ法の開発

① 一对の構造の違いの指標としては RMSD (root-mean-squared deviation) が広く用いられるが、それを 3 個以上の構造の同時重ね合わせのコスト関数として用いたものが等分散性の通常の最小二乗法である。その改良版が異分散法であるが、本研究での異分散法は、(暫定推定) 平均構造との重み付き MSD の和を最小化する方法とみることにもできる。このとき、RMSD を一種の L21 ノルムとみて、それをさらに重みとして加える方法を考案した。

② 人血液凝固第 X 因子の重鎖の 6 個の構造 (PDBID: 1EZQ, 1FOR, 1FOS, 1FJS, 1FXV, 1G2L) を本研究の異分散法とそれに RMSD の L21 ノルムの重みを加えた方法でそれぞれ重ね合わせを行った (下図の左と右)。太線 (オレンジ色) は推定平均構造である。左の図では、外れ分子構造 (1G2L) の影響を受け、そちらに平均構造が寄っているが (たとえば、最上部手前のループ)、右の図ではその影響が少なく、平均構造は外れ分子構造を除く他の分子構造の近くに位置しているのがわかる。平均構造に対してそれぞれの構造は重ね合わされるので、平均構造の違いは、当然、重ね合わせの違いにつながり、また、それは平均のズレが目立つ部分だけではなく、構造全体に及んでいる。



③ 原子単位の外れ値は、実用上、異分散性の導入で対応できているため、さらに、分子単位で外れ値に対応するものである。この手法は、占有率への対応とあわせて、結晶構造の比較など、対象とする構造があまり多くない場合に有用であると思われる。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 1 件)

① 網崎孝志、藤原伸一. Regression method for comparison of multiple protein conformations. 日本生物物理学会第 55 回年会(熊本大学、2017 年)

## 6. 研究組織

### (1) 研究代表者

網崎 孝志 (AMISAKI, Takashi)

鳥取大学・医学部・教授

研究者番号：20231996

### (2) 連携研究者

藤原 伸一 (FUJIWARA, Shin-ichi)

鳥取大学・医学部・准教授

研究者番号：00362880