

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 24 年 5 月 31 日現在

機関番号：24506

研究種目：挑戦的萌芽研究

研究期間：2011

課題番号：23654129

研究課題名（和文） 生体超高分子の立体構造および電子構造の統合的決定技術の開発

研究課題名（英文） Development of an integrated structural determination system employing crystallographic diffraction data and *ab initio* electronic structure calculation of biological macromolecules

研究代表者

館野 賢 (TATENO MASARU)

兵庫県立大学大学院・生命理学研究科・教授

研究者番号：40291926

研究成果の概要（和文）：近年，生体高分子の X 線結晶構造解析の精度が飛躍的に向上し，それらの中には，生体高分子の活性中心における(外殻)電子の軌道自体を実験的に求める研究も見られる。本研究は，非経験的 (*ab initio*) 電子構造計算を用いて，回折実験データとの理論的な比較を行うことにより，生体高分子の活性中心の立体構造と電子構造の双方を最良の精度で同時に獲得するための解析技術の開発を試みた。立体構造と電子構造を同時に比較・評価するための情報統計学的な理論スキームを開発し，それを実装した計算システムの主要部分を構築することによって，本解析法の妥当性・実現性を評価した。その結果，上述の目的を実現するための基盤となる知見を得ると共に，解析システムのプロトタイプを構築し，本研究の目的を達成した。

研究成果の概要（英文）：In this study, we developed a theoretical scheme and a computational system to simultaneously determine the three-dimensional and electronic structures of biological macromolecular systems by integrating crystallographic diffraction data and electronic states obtained by *ab initio* quantum mechanics calculation. In our algorithm, informatical approaches play a crucial role to obtain the important electronic/geometric structural information involved in the crystallographic data and the electronic states obtained by using *ab initio* calculations. We evaluated the scheme and the computational system that were developed in this study. As a result of the analysis, we found the critical issues to achieve our goal; then, we created the improved theoretical schemes by resolving the issues, and thus, produced a prototype system to actualize the strategy for the above-mentioned structural determination. This is a solid basis to further sophisticate the obtained informatical system, thereby leading to create a new strategy to determine the electronic and geometric structures of biological macromolecular systems with high accuracy.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	3,100,000	930,000	4,030,000

研究分野：

科研費の分科・細目：

キーワード：電子構造計算，X線結晶構造解析，電子密度，*ab initio* 計算，構造精密化，生体高分子，バイオインフォマティクス，情報科学，計算科学

1. 研究開始当初の背景

近年、生体高分子の X 線結晶構造解析の精度が飛躍的に向上し、それらの中には、生体高分子の活性中心における(外殻)電子の軌道自体を実験的に求める研究さえ見られる。したがって、低分子量化合物に対する構造解析のみならず、生体高分子に対してもその電子構造の決定までを同時に行うことの可能な技術開発が望まれている。他方で、遷移金属が結合したタンパク質の活性サイトなどにおいては、結晶構造解析による原子座標の配置(構造精密化)が十分正確にできない場合も多く、そうした際の立体構造決定技術の改善もまた望まれている。

例えば、ある酸化還元酵素における電子移動部位では、2つの Cu イオンが共有結合を形成し、6個のアミノ酸残基がこれらに配位している。この酵素の酸化型及び還元型の立体構造の回折データは、それぞれ 1.8, 1.9 Å であり、比較的良質(中解像度)であるといえる。しかし得られた上記の活性部位の立体構造には原子間の重なりが見られ、エネルギーの高い状態が見出された [1] (文献番号は、後掲の文献リストの番号に対応する)。これは、Cu 原子などの電子の軌道が広がりをもつために (delocalization), 実験データのみにより正確な原子位置を決定することが極めて困難であるためと考えられる。

上記の研究の中では、高エネルギー状態構造を解消するための理論的な試みも行った上で、生物機能のメカニズムを解析した(その方法と成果は、以下の本研究においても実際に用いられ生かされている) [1]。しかしこれらの課題は本来、結晶構造解析のストラテジの中で、回折データを高精度に処理する過程において果たすべきものである。こうした現実の解析事例が動機となり、それらの課題を根本的に解決するための技術開発を模索し始めたことが、本研究の出発点にある。

2. 研究の目的

そこで本研究では、非経験的 (*ab initio*) 電

子構造計算を用いて、回折実験データとの理論的な比較を行うことにより、生体高分子の活性中心の立体構造と電子構造の双方を最良の精度で同時に獲得するための理論スキームを開発する。さらに、それを実装した計算システムの主要部分を構築することにより、当該解析法の妥当性・実現性を評価する。

3. 研究の方法

本研究では、1) 通常の構造精密化によって得られた原子核位置から、活性化サイトの立体構造を多数生成し、2) それらに対して非経験的 (*ab initio*) 電子構造計算を行った後、3) 得られた電子密度を、回折実験によって得た電子密度と比較することにより、高精度の立体構造及び電子構造を同時に得るための理論スキームと解析システムの開発を行った。

1) では、活性化サイトのコンフォメーションを多数生成するために、モンテカルロ・シミュレーションおよび分子動力学計算を用いた。この際、正確なエネルギー関数(古典場)を用いることが、立体構造の定量的な評価を正確に行うために不可欠である。例えば、遷移金属が関わる場合やカチオンと π 電子との相互作用 ($\text{Na}^+ - \pi$ 相互作用) など、汎用のエネルギー関数では正しく評価することが著しく困難な場合も多くある。こうした問題を回避するために我々は、計算時間を増大させることなく(通常のエネルギー関数と同程度の計算量で)、高精度のエネルギーの値を得ることができる計算法を開発し、その有効性を実証してきたところである (Y. Hagiwara *et al.*, *J. Am. Chem. Soc.* 2009, **131**, 16697; Y. Hagiwara *et al.*, *J. Chem. Theory Comput.*, 2011, **7**, 2593-99)。

2) では、我々の研究室で構築した、高精度・高並列化されたハイブリッド量子力学/分子力学 (QM/MM) 計算システムを適応し、生体高分子の活性部位の電子密度を求めた。このハイブリッド QM/MM 計算法では、活性化サイト以外の部分とその周囲の溶媒の効果

を *ab initio* 電子構造計算に導入することができ、活性化サイトの周囲の環境の効果を含めた電子状態を高効率に求めることができ、高精度の電子構造計算が可能である。

3) では、以下で概説する新たな理論スキームと解析システムを用いて、前項から得られた2つの電子密度を比較・評価することによって、構造決定を行うための情報統計学的なスキームを構築する。

4. 研究成果

まず、ふたつの電子密度分布を定量的かつ組織的・自動的に比較するための理論を以下のように構築した。

前項 1) に記したように、活性部位のコンフォメーション・セットを生成し、これを

$$\mathbf{C}_0 = \{c_1, c_2, c_3, \dots, c_N\},$$

とする。それぞれのコンフォメーションに対して、ハイブリッド QM/MM 計算を実行し、電子密度分布のセットを得て、

$$\mathbf{R} = \{\rho_1(\mathbf{r}), \rho_2(\mathbf{r}), \rho_3(\mathbf{r}), \dots, \rho_N(\mathbf{r})\}.$$

とする。次にこれらの計算結果を用いて、実験データから得られた電子密度関数 $\rho_{\text{exp}}(\mathbf{r})$ との相関関数 $\text{Cor}(i)$ を、それぞれのコンフォメーションに対して以下のように計算する。

$$\text{Cor}(i) = \int \rho_i(\mathbf{r}) \cdot \rho_{\text{exp}}(\mathbf{r}) d\mathbf{r}$$

しかしこの方法では、精密な比較が困難であり、厳密な立体構造の決定には適さないことがわかった (unpublished data)。その理由は、実験データの解像度が十分ではない点にある。すなわち、1.5 Å 程度の中解像度データでは、データに含まれる誤差が大きいため、*ab initio* 計算によって得られた電子密度分布との正確な比較は困難である。

他方で、生体高分子に関する現在の回折データにおいて、ターゲットとなり得るデータの多くは 1.5Å 前後の解像度を有する。そのため、こうした実験誤差を考慮した上で、精密な比較法を導入することが必要である。ここで実験誤差とは、例えるならば、画像のエッジがぼんやりとぼけた状況に相当する。実

際、上記の解像度における電子密度マップは、*ab initio* 電子構造計算によって得られた電子密度分布に比較して、そのエッジがぼけていることがわかる。そこで、こうしたボケ画像を情報科学的に高画質へと再構成し直す画像処理技術を、以下のように導入する (インフォマティクスの応用)。

解像度の差を取り扱うために、*ab initio* 計算から得られた全電子密度分布関数と、ここで新たに導入するガウシアン関数に対して、以下の相関関数を新たに $\text{Cor}(i)$ として定義する。

$$\text{Cor}(i)$$

$$= \int \left(\int g(\boldsymbol{\tau}) \rho_i(\mathbf{r} - \boldsymbol{\tau}) d\boldsymbol{\tau} \right) \cdot \rho_{\text{exp}}(\mathbf{r}) d\mathbf{r}$$

次にこの評価関数を用いて、最適なコンフォメーションを求める方法を以下のように構築した。前述のようにコンフォメーション・セット (初期コンフォメーション・セット) を作成し、その中で最も $\text{Cor}(i)$ の値が大きいコンフォメーション i を考える。その立体構造の各原子核の座標セットを

$$\mathbf{X}(i) = \{x_1, x_2, x_3, \dots, x_{N_{\text{atom}}}\}$$

とし、各原子核の周囲の電子密度に関する相関関数を以下のように定義する：

$$\text{Cor}(i, X_a)$$

$$= \int_{\mathbf{r} \in V_{X_a}} \left(\int g(\boldsymbol{\tau}) \rho_i(\mathbf{r} - \boldsymbol{\tau}) d\boldsymbol{\tau} \right) \cdot \rho_{\text{exp}}(\mathbf{r}) d\mathbf{r}$$

この時、原子座標の微小変化に対する $\text{Cor}(i, X_a)$ の変化量

$$\frac{\partial \text{Cor}(i, X_a)}{\partial X_a},$$

をハミルトニアンに導入し、これに従って、同時に Stimulated Annealing (SA) 法を適用したハイブリッド QM/MM 分子動力学計算を実行する (我々が構築した上述の計算システムを用いた)。また適宜、構造最適化 (エネルギー最小化) 計算をこれらに併用することによ

って、局所的な安定構造を得て、新たなコンフォメーション・セットを生成し、

$$\mathbf{C}_1 = \{c_1, c_2, c_3, \dots, c_N\}$$

とする。

次に、この \mathbf{C}_1 から、相関関数 $\text{Cor}(i)$ の最大値を求め、前回のイタレーションにおけるコンフォメーション・セットから計算された相関関数 $\text{Cor}(i)$ の最大値との差、すなわち、

$$\Delta\text{Cor} = \text{Cor}(i_{\max, \mathbf{C}_{n+1}}) - \text{Cor}(i_{\max, \mathbf{C}_n})$$

を定義する。ここで \mathbf{C}_n は n 番目のコンフォメーション・セットを表し、その中の各コンフォメーションにおいて、以下の条件 (Cor の値が最大) を満たすものを i_{\max, \mathbf{C}_n} とする。すなわち、

$$\text{Cor}(i_{\max, \mathbf{C}_n}) = \text{Max}_{i \in \mathbf{C}_n}(\text{Cor}(i))$$

を満たすコンフォメーションとして、 i_{\max, \mathbf{C}_n} を定義する。

次に、 i_{\max, \mathbf{C}_n} と $i_{\max, \mathbf{C}_{n+1}}$ の立体構造の違いを

$$\Delta\text{RMSD} = \sqrt{\frac{1}{N_{\text{atom}}} \sum_{X_a}^{N_{\text{atom}}} |X_{a, i_{\max, \mathbf{C}_n}} - X_{a, i_{\max, \mathbf{C}_{n+1}}}|^2}$$

として定義する。ここで ΔCor および ΔRMSD が、それぞれに設定した基準値よりも小さな値になったときに、そのコンフォメーションを「最適化された立体構造」として採用する。

以上の改良された理論スキームを、先に開発した解析システムに実装した。さらにこれを用いて、現実の系へ応用し、その評価を行ったところ、当初の目的に対して良好な結果が得られた (unpublished data)。

このように、情報統計学的な解析技術 (インフォマティクス) の応用は、大規模・大量データを用いた解析において不可欠である。言うまでもなく、生命科学における新技術の開発においても極めて重要である。

最近我々は、ゲノム DNA を網羅的に解析したデータにおける大量の塩基配列から、機能的に重要な部分配列を正確かつ自動的に抽出するための、独自の新たな解析アルゴリズム MODIC を開発した。これを実装し、現

在利用可能な他のプログラムによる解析結果と比較したところ、我々のアルゴリズムが最も優れた結果を出力することが明らかになった。このアルゴリズムは、ゲノム DNA の塩基配列と遺伝子位置の情報のみを用いて (他の生物学的な知見を一切用いない)、機能的に重要な配列を情報科学的に抽出する *ab initio* (非経験的) モチーフ同定法である [4]。

こうしたバイオインフォマティクス分野における解析システムは、現在の先端的生命科学の研究において必要不可欠であり、これら無しに実験の解析・設計さえできず、研究を遂行し得ない状況にある。我々自身も MODIC を用いた共同研究を、実験グループと共に現在推進しているところであるが、今後ますますその必要性は高まる一方である。

このアルゴリズムの開発で用いた、確率密度関数によるノイズとシグナルとの分離法などの考え方は、こうしたゲノム科学分野のみならず、本課題における構造生物学分野においても今後さらに応用することによって、立体構造・電子構造の決定技術の洗練化に大きな寄与をもたらし得るものと期待される。

そもそも現在の X 線結晶構造解析法において、回折データ取得後の処理については、ほとんどすべてがコンピュータ・プログラムによって自動化されているため、その処理内容の詳細を理解する研究者はいまや少ない。定式化に至っては一層、マスクされてしまっている。しかしここでもまた、最早表面的に見えないだけで、大量データ解析におけるインフォマティクス技術が用いられており、そもそもその技術の重要性は強調し過ぎることではない。そして今再び、電子構造と立体構造の統合決定技術の開発においても、前述のように本研究において、こうした情報統計学的な技術の応用が自ずと不可欠のものとなったのである。

今後、本研究で開発された上述の計算アルゴリズムをさらに洗練させることにより、中解像度の回折データに対する広く一般性のある解析法へと一層発展させたいと考えて

いる。こうした解析結果を用いて、正確な立体構造と電子構造を得て、それらに基づいて生物機能のメカニズムの研究を推進したいと考えている。

本研究は、そのための重要な一歩を築いたものといえる。

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)

[雑誌論文] (計5件)

- ① Jiyoung Kang, Yohsuke Hagiwara, and Masaru Tateno, Biological applications of hybrid quantum mechanics / molecular mechanics calculation, *J. Biomed. Biotech.*, **2012**, 236157, 2012, 審査有
DOI:10.1155/2012/236157
- ② Yohsuke Hagiwara, Jiyoung Kang, and Masaru Tateno, Structural Instability of the Active Site of T1 Lipase Induced by Replacement of Na⁺ with Water Complexed with the Phenylalanine Aromatic Ring, *J. Chem. Theory Comput.*, **7**, 2593-99, 2011, 審査有
DOI:10.1021/ct100752y
- ③ Jiyoung Kang, Hiori Kino, and Masaru Tateno, A theoretical investigation of the functional role of the axial methionine ligand of the Cu_A site in cytochrome c oxidase, *Biochim. Biophys. Acta Bioenerg.*, **1807**, 1314-27, 2011, 審査有
DOI:10.1016/j.bbabi.2011.06.014
- ④ Ryo Nakaki, Jiyoung Kang, and Masaru Tateno, A novel ab initio identification system of transcriptional regulation motifs in genome DNA sequences based on direct comparison scheme of signal/noise distributions, *Nucleic Acids Research*, in press, 審査有
- ⑤ 姜 志始, 萩原陽介, 舘野 賢, コンピュータ・シミュレーションによる生体触媒反応機構の研究, 応用物理, 第80巻, 第7号, p.610-14, 2011, 審査有

[学会発表] (計5件)

- ① Jiyoung Kang and Masaru Tateno, A computational study of functional role of

axial methionine ligand of Cu_A site in cytochrome c oxidase, 第49回日本生物物理学会年会, 2011年9月16日~9月18日, 兵庫県立大学・姫路書写キャンパス.

- ② Masaru Tateno, Jiyoung Kang, and Ryo Nakaki, Theoretical and computational quantum structural biology for understanding functional mechanisms of biological macromolecular systems, 第49回日本生物物理学会年会, 2011年9月16日~9月18日, 兵庫県立大学・姫路書写キャンパス.
- ③ Jiyoung Kang, 舘野 賢, シトクロム酸化酵素の Cu_A サイトにおける Cu 配位 Met 残基の機能的役割, 第84回日本生化学会大会, 2011年9月21日~9月24日, 国立京都国際会館.
- ④ 忒岡 亨, Jiyoung Kang, 舘野 賢, シトクロム酸化酵素の Heme a₃ 及び Cu_B サイトにおける電子構造の解析, 第84回日本生化学会大会, 2011年9月21日~9月24日, 国立京都国際会館.
- ⑤ Jiyoung Kang, 舘野 賢, Computational investigation of functional role of axial methionine ligand of Cu_A site in cytochrome c oxidases, 第34回日本分子生物学会年会, 2011年12月13日~12月16日, パシフィコ横浜.

[図書] (計2件)

- ① Jiyoung Kang, Yohsuke Hagiwara, and Masaru Tateno, Recent Applications of Hybrid Ab Initio Quantum Mechanics / Molecular Mechanics Simulations to Biological Macromolecules, In Some Applications of Quantum Mechanics, 359-384, 2012.
- ② 萩原陽介, 姜 志始, 舘野 賢, 生体反応の量子ハイブリッド分子動力学シミュレーション, 赤井久純, 白井光雲編 「密度汎関数理論の発展とマテリアルデザインへの応用」 シュプリンガー・ジャパン, 第III部/第8章, 327-344, 2011.

[産業財産権]

○出願状況 (計0件)

○取得状況 (計0件)

〔その他〕
ホームページ等
<http://www.sci.u-hyogo.ac.jp/edu/kenkyuu/base42.html>

6. 研究組織

(1) 研究代表者

館野 賢 (TATENO MASARU)
兵庫県立大学大学院
生命理学研究科・教授
研究者番号：40291926

(2) 研究分担者

無し

(3) 連携研究者

無し