

令和元年6月25日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00407

研究課題名(和文) タンパク質の動的特性を特徴づけるための分子内動的構造ネットワーク解析法の開発

研究課題名(英文) Development of intra-molecule dynamic structure network analysis method to characterize protein dynamics

研究代表者

輪湖 博 (Wako, Hiroshi)

早稲田大学・社会科学総合学院・教授

研究者番号：60158607

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：タンパク質の動的構造を理解するために基準振動解析計算を行い、低周波数の基準振動モードそれぞれについて、隣接し、大きな正の相関をもって動くアミノ酸残基対を結合して分子内動的構造ネットワークを定義した。そしてそれらネットワークの特性からタンパク質の動的構造を特徴づけることを考えた。酵素、アロステリックタンパク質、多量体などに応用し、ネットワーク解析で使われる中心性指標の一つ betweenness がタンパク質の構造特性や機能とよい相関をもつことを見出した。低振動モードの動きは機能と関係することが知られており、タンパク質の動的構造と機能の相関をよりの確に表現することが可能となった。

研究成果の学術的意義や社会的意義

タンパク質の機能と立体構造の関係を理解するためには、その静的構造だけでなく、動的構造との関係を明らかにすることが必要である。コンピュータはそうした動的構造に関する情報を、実験では得られない原子レベルの分解能で提供してくれる反面、その膨大なデータからわれわれの理解につながるような特性量をいかに抽出するかが問題となっている。本研究は、こうした問題に新たな方法論を提供する。われわれは膨大な数のタンパク質について基準振動解析計算を行い、それをPDBjを通して公開しているが、それらをより積極的に利用して、タンパク質の動的構造と機能の相関を理解していくための有用なツールを提供できるものと考えられる。

研究成果の概要(英文)：We have performed normal mode analysis for many proteins for understanding their dynamic structures. Using these results, we defined an intra-molecular dynamic protein structure network (dPSN) for characterizing them. We applied this method to various types of proteins such as enzymes, allosteric proteins, and oligomeric proteins and found that a centrality measure, betweenness, is the most effective property because it is well correlated with structural and functional characteristics of amino acid residues. Since the motions of the lowest-frequency normal modes are related to functional motions of a protein, we defined dPSN individually for such normal modes and revealed the significant features of their motions. For example, the residues with high betweenness were found in active sites of enzyme proteins, in an allosteric pathway between an active and effective sites, and on the inter-subunit interface of oligomeric proteins, implying effectiveness of this method.

研究分野：生物物理学、バイオインフォ マティクス

キーワード：生物物理学 バイオインフォ マティクス 計算化学 タンパク質の立体構造 基準振動解析 動的構造 ネットワーク解析

1. 研究開始当初の背景

タンパク質の機能はその立体構造と密接に関係している。しかし、その構造-機能相関を理解するためには、静的な構造情報だけでなく、動的な振舞い（ここでは動的構造と呼ぶことにする）との対応関係を調べるのが重要である。この問題へアプローチするための有力な手法の一つがコンピュータを利用した分子動力学（MD）シミュレーションおよび基準振動解析（NMA）である。これらは、実験では得ることの難しい原子レベルの分解能でタンパク質立体構造のゆらぎや構造変化の情報を提供してくれる。しかし反面、出力される情報量が極めて多く、そこからどのような情報を抽出すれば動的構造に関するわれわれの理解が進むのかが必ずしも明確に示されているわけではない。

こうした中で、ここ10年くらいの間に、ネットワーク解析法のタンパク質立体構造（静的・動的を含め）への応用が提案され、活発に議論されるようになった。しかし、ネットワーク解析法でも多様な特性量が提案されており、どのような解析がタンパク質の動的構造解析に最適であるかなど、解決されていない問題も多く、本研究を始める動機となった。

一方、われわれは、PDBj（大阪大学蛋白質研究所）に二次データベース ProMode-Elastic を構築し、タンパク質の基準振動解析データを公開してきた。現在も新規計算結果の追加登録を行い、充実を図っているが、こうして蓄積された動的構造に関する膨大な計算データからいかに有益な情報を抽出するかは喫緊の課題でもあった。

また、巨大タンパク質の基準振動解析はコンピュータに大きな負荷がかかり、今まで計算を敬遠してきた。しかし、巨大タンパク質の動的構造は興味ある研究対象であり、精度低下を最小限に抑えた近似計算を開発し、巨大タンパク質の基準振動解析ができるような方法論を開発することも、重要な課題であると考えていた。

こうした状況を踏まえ、タンパク質の動的構造を特徴づけるための基準振動解析計算を基礎にした分子内動的構造ネットワークを新たに定義し、さまざまなタンパク質に適用しながら、最も適した特性量を探索し、また、それによってタンパク質の動的構造がいかに理解できるかを考えながら、研究をスタートさせた。

2. 研究の目的

われわれはこれまでに自ら開発した基準振動解析プログラムを用いてタンパク質立体構造の動的振舞いを研究してきた。基準振動解析では、低周波数の基準振動モード（低振動モード）の運動がタンパク質全体の協奏的な運動を特徴づけており、その組み合わせで天然構造の周りの揺らぎが構成されていると考えられている。また、基質が結合している場合と結合していない場合の立体構造がともに分かっているタンパク質において、

両者間で観察される構造変化と低振動モードの動きが良く対応する事例が数多く報告されており、機能との関連が深い運動であると考えられている。

基準振動解析では、タンパク質の運動は個々の原子の変位ベクトルで示される。それを基にした運動のアニメーション表示も可能である。しかし、こうした視覚に訴える表現は、原子レベルの分解能をもった詳細な情報をもたらし、われわれの理解を深めてくれる一方、その解釈やその動きの表現に主観的要素が入ることも否めない。したがって、なんらかの客観的指標によって運動を表現し、議論することが望まれている。

そこで本研究では、各低振動モードに対して分子内動的構造ネットワークを定義し、これにネットワーク解析の手法を適用して、何らかの指標によって各モードの動きを特徴づけることを考えた。そうした指標を用いることで、タンパク質の構造-機能相関をよりの確に表現できる方法の開発をめざした。

3. 研究の方法

(1) 分子内動的構造ネットワーク

タンパク質の立体構造を、グラフ理論におけるネットワークとして表現することを考えた。ネットワークは頂点とそれらを結ぶ辺からなるが、頂点として各アミノ酸残基のC β 原子（GlyについてはC α 原子）を選び、それらを辺で結ぶ条件を以下のように定めた。このとき、その条件によって、静的構造ネットワーク（sPSN）と動的構造ネットワーク（dPSN）を定義した（PSNはProtein Structure Networkの略）。

静的構造ネットワーク static PSN

PDBの構造において、あるアミノ酸残基の少なくとも1つの原子と、他のアミノ酸残基の少なくとも1つの原子との距離がある閾値（ここでは5Å）以下であるとき、その2つの残基（C β 原子で代表する）は接触しているとし、辺で結ぶ。

動的構造ネットワーク dynamic PSN

静的構造ネットワークの定義における2つの残基の接触条件に加え、それぞれの残基のC β 原子の変位ベクトルのcosineがある閾値（ここでは0.8）以上のとき、両者を辺で結ぶ。

の条件は、2つのC β 原子の動きが同方向であり、両者の動きが相関していると判断したことを意味している。ちなみに、dPSNはsPSNの部分グラフであり、基準振動モードごとに異なったものとなる。

他の多くの研究は、sPSNのみを考えている。dPSNに類するものを定義している場合でも、すべての基準振動モードの平均（時間平均にあたる）で2つの残基の運動の相関を求め、それを辺の定義に使っていることが多い。しかし、時間平均を使うと、相関の高い残基対のほとんどは空間的に近い残基対となり、sPSNとdPSNの差が小さくなるのが一般的である。本研究の特徴の一つは、個々の振動モードについてdPSNを定義し、各振動

モードの運動を特徴づけることを考えた点にある。

(2) 基準振動解析法の発展

巨大タンパク質の基準振動解析を行うために、精度低下を最小限に抑えた近似計算について検討を行った。

(3) データベースの充実

PDBj (大阪大学蛋白質研究所) の二次データベース ProMode-Elastic の登録データの追加・更新を行った。

4. 研究成果

(1) 中心性指標の評価 酵素を例に

周波数の最も低い基準振動モードから 10 番目のモードまで (以後、単純に第 1 モード、第 2 モード、...、第 10 モードと参照する) のそれぞれについて定義された dPSN と、対照として定義された sPSN について、それらを特徴づける量として何が適当かを検討した。そのために、ネットワーク解析で使われる 3 つの中心性指標 density、closeness、betweenness を各アミノ酸残基について計算し、解析した。

まず対象としたのは酵素で、基質が結合している場合と結合していない場合の両方の立体構造が PDB に登録されており、両者間での構造変化が比較的大きいものを選択した。

その結果、3 つの指標の中では betweenness が次のような特徴をもち、dPSN の特性量としてもっとも適していると評価した。

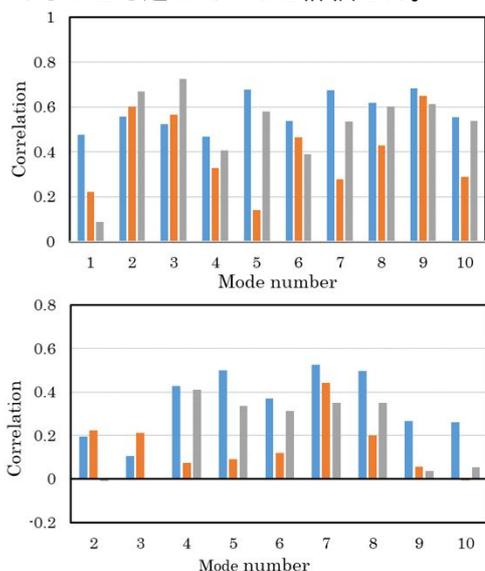


図 1. 中心性指標のモード間の相関

3 つの中心性指標、density (青)、betweenness (橙)、closeness (灰色) に対して、(上段) sPSN と各モードとの相関、(下段) 第 1 モードとそれ以外のモードとの相関を示した。いずれも、betweenness の相関が他の指標の相関よりも小さく、それだけ各モードの違いを反映した指標であることを示唆している。

a) その値が大きい残基と小さい残基がはっきりと分かれる。

b) あるモードに対して各残基について求められた値を sPSN および他のモードで求め

られた値と比較すると (両者の相関を求めると) density や closeness に比べ、比較的相関が小さい。すなわち sPSN や他のモードとの違いを比較的よく反映する指標と思われる (図 1)。

c) 活性部位やドメインとドメインの連結部位など、構造特性や機能と関連する残基が高い値を示す。

d) 進化で保存される残基が高い値を示す。

特に、c) においては、アポ状態での基質の結合部位に betweenness の高い残基があることがわかった。また d) は、betweenness が高い残基は重要であることを示唆している。

こうした結果から、density や closeness よりも betweenness が dPSN の特性量として適していると評価し、以後の解析では betweenness を基盤にした議論を展開することとした。

(2) アロステリック・タンパク質

アロステリック効果を示すタンパク質では、リガンドの結合部位とエフェクター分子が結合するアロステリック部位が空間的に離れており、両者がどのような方法で情報交換を行っているかを明らかにすることが課題の一つとなっている。そこで、互いに接触し、大きな正の相関をもって動く残基が作るネットワーク、dPSN がその情報伝達の経路を構築しているという予想のもとで、以下のような解析を行った。

例として異化活性化タンパク質 (Catabolite Activator Protein; CAP) を取り上げた。CAP は転写に関わるホモ 2 量体のタンパク質で、サブユニットは 2 つのドメイン、N 端側の cAMP 結合ドメイン (CBD) と C 端側の DNA 結合ドメイン (DBD) からなる。A~F の 6 本の α ヘリックスのうち、CBD にあり cAMP 結合部位に近い C ヘリックスは 2 つのサブユニットの接触面に存在し、DBD にある F ヘリックスが DNA 結合部位を構成する。

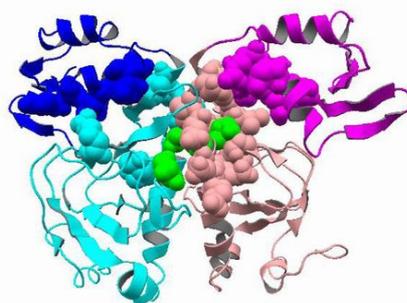


図 2. apo CAP (PDB ID: 2wc2) において高い betweenness をもつ残基の原子を球で示した。A 鎖の CBD、DBD、B 鎖の CBD、DBD をそれぞれ薄茶色、赤紫色、水色、青色で、また、cAMP の結合部位近くの残基を緑色で示した。

10 個の低振動モードについて調べ、いずれかのモードで高い betweenness をもつアミノ酸残基を図 2 に示した。興味深いことに、cAMP 結合部位から C ヘリックスの C 末端を経由して、DNA 結合部位である F ヘリックスまで、betweenness の大きな残基が連なっていることがわかった。すなわち、互いに接触

し、大きな正の相関をもって動く残基で構成された経路が、cAMP 結合部位と DNA 結合部位の間に形成されていることが示された。この経路によってどのように情報が伝達されるかはこの結果からは説明できないが、重要な示唆を提供しているものと考えている。

(3) 多量体タンパク質

多量体タンパク質の動的特性を調べるため、ホモの 3 量体、4 量体、6 量体の基準振動解析を行った。2 量体については過去に行った経験があり、5 量体に関しては分解能の高いデータが PDB にはなかったため、本研究では解析の対象としなかった。

基準振動解析では、各基準振動モードに対して各原子の変位ベクトルを計算する。それらは豊富な情報を提供してくれる反面、われわれがそこから何を学ぶかが見えにくいという側面もある。そこで、その変位ベクトルのデータから各サブユニットの剛体としての運動を計算し、複数のサブユニットの相対運動を見ることで多量体タンパク質の運動を記述することを試みた。

具体的には、変位ベクトルからサブユニットの剛体としての並進運動と回転運動のベクトルを計算し、さらに得られたベクトルを多量体に対して定義した円柱座標系の成分で表現することを考えた。ここで、円柱の軸方向を z 軸、半径方向の軸を r 軸、円柱の接戦方向を t 軸と呼ぶことにすると、10 個の低振動モードについて、図 3 のような結果を得た (4 量体の例を示す)。

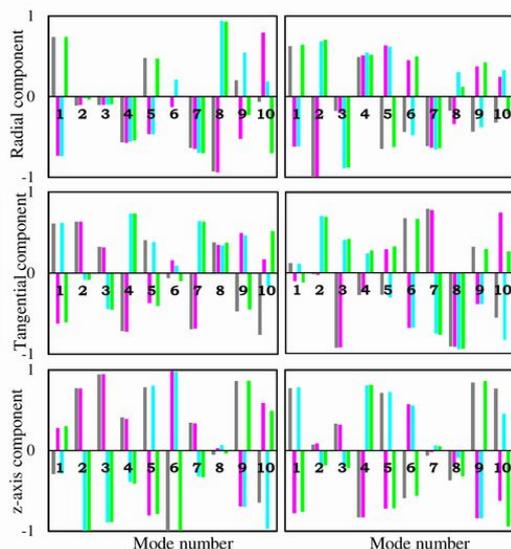


図 3 . 円柱座標系で示した各サブユニットの各基準振動モードにおける運動。4 量体タンパク質 L-アスパラギナーゼ (PDB ID: 1o7j) の例。左側が並進運動、右側が回転運動。上から r 軸、 t 軸、 z 軸成分。それぞれの図の中で色は 4 つのサブユニットを表し、1~10 の数字は基準振動モード番号を表す。単位ベクトルの成分なので絶対値の最大値は 1 である。

こうした表現は、各サブユニットの運動をイメージしやすくしてくれる。特に、 r 、 t 、 z 軸のうちの一つの成分だけが顕著に大きい場合はより明確であり、実際、いくつかそう

したモードがあることがわかる。また、各サブユニットの成分が互いにほぼ等しい場合には、全体として対称的に運動していることを示している。一方、サブユニットごとに異なる場合は、全体の運動は定常波となり、それぞれのサブユニットは位相をずらしながら同じ動きをしていることになる。この場合、定常波を構成する波の数によってその動きをさらに分類することができる。実際、図 3 は、多くの基準振動モードが、こうした運動のどれかに分類できることを示している。

各サブユニットの表面にある残基に注目した解析も行った。各サブユニットの運動を剛体運動と変形運動に分け両者の変位ベクトルの cosine を調べると、サブユニット内部の原子は一様分布するのに対して、表面の原子では極端に -1 に近い値を取り、剛体運動と変形運動が反対方向を向いていることを示していた。これは、サブユニットの数が多くなるにつれ全体として連続体としての性質を帯びるようになるとの予想に反し、各サブユニット本来の動きがあくまで基本となり、他のサブユニットとの接触面でのみ、多量体としての特徴が現れることを示唆していた。

ただし、上で述べたような、多量体全体が示す対称的な、あるいは定常波的な運動は多量体特有の運動であり、こうした運動との整合性をいかに付けるかは、今後さらに解析が必要である。

また、サブユニットの接触面にあって高い betweenness の値をもつ残基のうち、相手のサブユニットの接触面にあってやはり betweenness が高い残基と接触しているものを特定した (図 4)。このような残基対によってサブユニットの接触面に形成されるネットワークは多量体のダイナミクスを考える上で重要な役割を果たしていると考えられる。さらに、サブユニット内部に形成される betweenness の高い残基のネットワークとの連結も確認されたが、これは今後の検討課題として残された。

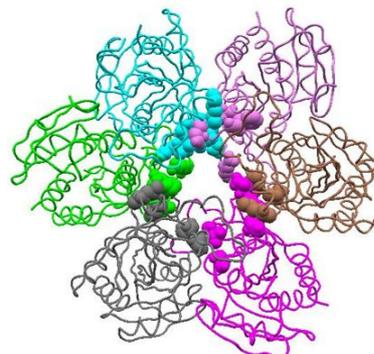


図 4 . サブユニットの接触面にある betweenness の高い残基 (球で表現されている) によって形成されたネットワーク。6 量体のウリジンホスホリラーゼ (PDB ID: 4r2x) の例。鎖の違いに応じて色を変えて表現している。

(4) TCR-pMHC 複合体

ヘテロな多量体 (複合体) の動的特性も調べた。T 細胞受容体-ペプチド-MHC (TCR-pMHC) 複合体については多くの立

構造データが PDB に登録されている。TCR は認識するペプチドの違い、特異性の強さ、交差性の有無など、多様性に富んでいるのが特徴であるが、複数の複合体について基準振動解析を行うことによって、その共通点や、それぞれの複合体に特有の性質が見られるのではないかと考えた。

27 個の複合体データについて動的構造ネットワーク解析を行った。図 5 に結果の一部を示した。27 個の複合体のマルチプル・アライメントを行い、betweenness の高い残基を色で示している。共通して betweenness が高い残基がある領域と限られた複合体にのみそうした残基がみられる領域があり、普遍性と多様性を垣間見ることができる。

TCR には、ペプチド-MHC 複合体の認識に関わる 3 つの相補性決定領域 CDR1、CDR2、CDR3 がある。図からは、CDR1 と CDR3、とりわけ後者に betweenness の高い残基が集中しているのがみてとれる。こうした動的性質が、認識機能にどのように関わるかはまだ十分に解析できていないが、興味深い結果であると考えている。

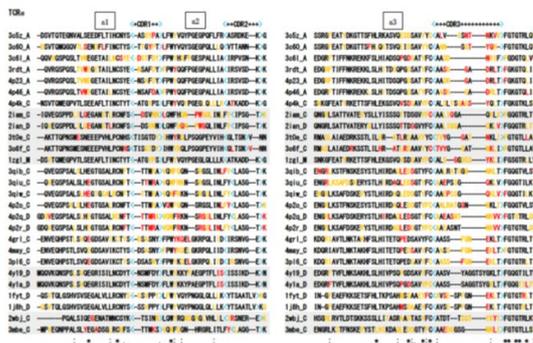


図 5. TCR-pMHC 複合体のマルチプル・アライメント (TCR の一部を表示している)。betweenness が高いアミノ酸残基を黄色と赤色で示した (赤色がより高い値であることを示す)。上端の四角で示したのはヘリックスの領域であり、それと並ぶ 3 本の横線は左から CDR1、CDR2、CDR3 の領域である。

また、図には示されていないが、TCR および MHC それぞれがもつドメインの境界領域に betweenness の高い残基が共通して多く存在することもわかった。TCR-pMHC 複合体は、上で述べたホモの多量体のような対称性をもたない棒状の形状であり、動的構造という視点からも特異的な様相を示している。

この複合体については分子動力学計算を行い、TCR の特異性の違いについても検討を行った。TCR とペプチドと MHC の結合様式の違いがエントロピーおよびエンタルピーの観点から異なり、特異性の強さに変化を生み出している可能性を示唆する結果を得た。

(5) 巨大タンパク質の基準振動解析

コンピュータによるタンパク質の研究では、タンパク質のサイズが大きくなるにつれ計算時間の問題が顕在化する。高速化するための方法の一つとして、モデルの簡略化がある。われわれは、基準振動解析計算において

二面角を独立変数として採用していることを生かした高速化を考え、巨大タンパク質に適用し、その有効性を検証した。

通常、モデルの簡略化といった場合、一つのアミノ酸残基を $C\alpha$ 原子で代表させることが一般的である。しかし、われわれは、基準振動解析を二面角を独立変数として行っていることを生かし、全原子モデル (ただし、水素原子は除く) を保ったまま、側鎖二面角および主鎖の ω を固定し、主鎖の二面角 ϕ と ψ だけを変数とする近似 (PP モデル) と、隣り合う $C\alpha$ を結ぶ仮想結合の回りの二面角を変数とする近似 (VB モデル) を考え、その有効性を検討した。基準振動解析では、Hessian 行列の一般化固有値問題を解くのに計算時間のほとんどが費やされるため、相互作用の数を減らすことよりも独立変数の数を減少させる近似の方が高速化にはもっとも有効と考えたからである。

すべての回転可能な二面角を変数とするオリジナルのモデルを FULL モデルと呼ぶと、原子の揺らぎは、PP モデル、VB モデルともに FULL モデルの結果をよく再現した。また、主鎖の二面角 ϕ と ψ の揺らぎに関しても、PP モデルは FULL モデルの結果をよく再現した。さらに、各原子の変位ベクトルについて、cosine 類似度 (2 つのモデルの対応する原子の変位ベクトルの cosine の平均) をモードごとに調べると、多くの場合、互いに同じモード番号どうしが高い類似度をもち、その値も低振動モードでは 1.0 に近かった。類似度が高いモードが隣り合うモードで入れ替わるケースがあったが、それはその 2 つのモードが縮退しているためであり、さらに類似度が小さくなる場合もあるが、それは運動の位相の違いによるものと考えられた。したがって、本質的にはよく一致していることがわかった。このことは、これらの近似モデルが、平均量だけでなく、低振動モードの分子の運動をよく再現できていることを示している。

こうした結果は、全原子モデルを採用したことにより、分子の動き方に強い制約がかかり、独立変数の取り方にあまり依存しなくなっているためと考えられる。また、二面角という変数が鎖状分子の運動を記述するのに適しているからでもある。固定した側鎖の二面角は局所的な運動に関与し、低振動モードが示すような分子全体の動きにはあまり影響を与えないからである。

(6) データベース ProMode-Elastic

PDB のデータを解析した二次データベースとして、ProMode-Elastic を PDBj に構築し、基準振動解析の結果を提供してきた。動的構造の計算結果を提供するサイトは少ないので、現在でも解析結果の追加登録を行っている。2019 年 6 月 7 日現在、PDB に登録された 152,500 件のデータうち、われわれのプログラムで対応可能な 141,095 件のデータについて計算を行い、その結果を登録した。また、複数の鎖を含むものについてはそれぞれの鎖についても別々に解析を行っており、それらが 249,631 件登録されている。

5. 主な発表論文等

[雑誌論文](計 4 件)

Tsuchiya Yuko, Namiuchi Yoshiki, Wako Hiroshi, and Tsurui Hiromichi. A study of CDR3 loop dynamics reveals distinct mechanisms of peptide recognition by T-cell receptors exhibiting different levels of cross-reactivity. *Immunology* **153**, 466–478 (2018). (査読有)

DOI: 10.1111/imm.12849

Wako Hiroshi and Endo Shigeru. Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank. *Biophys. Rev.* **9**, 877–893 (2017). (査読有)

DOI: 10.1007/s12551-017-0330-2

Kinjo R. Akira, Bekker Gert-Jan, Wako Hiroshi, Endo Shigeru, Tsuchiya Yuko, et al. New tools and functions in data-out activities at Protein Data Bank Japan (PDBj). *Protein Sci.* **27**, 95–102 (2017). (査読有) DOI: 10.1002/pro.3273

Wako Hiroshi and Abe Haruo. Characterization of protein folding by a Φ -value calculation with a statistical mechanical model. *Biophys. Physicobiol.* **13**, 263–279 (2016). (査読有)

DOI: 10.2142/biophysico.13.0_263

[学会発表](計 13 件)

猿渡茂、輪湖博 「二面角系粗視化モデルによる巨大タンパク質の立体構造ゆらぎ - X 線構造の温度因子との比較」日本蛋白質科学会 2018 年

輪湖博、土屋裕子、猿渡茂 「基準振動のネットワーク解析による TCR-pMHC 複合体の動的構造」日本生物物理学会 2018 年

猿渡茂、輪湖博 「二面角系粗視化モデルによる巨大核酸分子の立体構造ゆらぎ - X 線構造の温度因子との比較」日本生物物理学会 2018 年

輪湖博、猿渡茂 「タンパク質の基準振動モードのネットワーク解析: 中心性指標 Betweenness と活性部位」日本蛋白質科学会 2017 年

猿渡茂、青木拓弥、輪湖博 「二面角系粗視化モデルによる超分子複合体の基準振動解析」日本蛋白質科学会 2017 年

輪湖博、猿渡茂 「タンパク質の基準振動モードのネットワーク解析: 中心性指標 Betweenness とアロステリック機構」日本生物物理学会 2017 年

土屋裕子、波内良樹、輪湖博、鶴井博理 「T 細胞受容体による特異的および交差反応的な抗原認識機構の解明」日本生物物理学会 2017 年

H. Tsurui, Y. Tsuchiya, Y. Namiuchi, and H. Wako. Characterization of TCR-pMHC interaction at residue-level based on FMO and PIEDA. *CBI 学会* 2017 年

Y. Tsuchiya, Y. Namiuchi, H. Wako, and H. Tsurui. The distinct dynamic mechanisms of peptide recognition by specific and cross-reactive T-cell receptors. *CBI 学会* 2017 年

猿渡茂、輪湖博 「二面角系基準振動解析プログラムの巨大分子への適用—側鎖自由度の固定」日本蛋白質科学会 2016 年

土屋裕子、波内良樹、輪湖博、鶴井博理 「動的構造解析に基づく特異性および交差性を有する TCR の抗原認識機構の解明」日本蛋白質科学会 2016 年

土屋裕子、波内良樹、輪湖博、鶴井博理 「T 細胞受容体による特異的および交差反応的な抗原認識機構の解明」日本生物物理学会 2016 年

輪湖博、猿渡茂 「タンパク質の基準振動モードのネットワーク解析: 中心性指標の計算」日本生物物理学会 2016 年

[図書](計 0 件)

[産業財産権]

出願状況(計 0 件)

取得状況(計 0 件)

[その他]

PDBj の二次データベース ProMode-Elastic <https://pdbj.org/promode-elastic>

6. 研究組織

(1)研究分担者

なし

(2)研究協力者

研究協力者氏名: 猿渡 茂、土屋 裕子

ローマ字氏名: ENDO Shigeru, TSUCHIYA Yuko