

科学研究費助成事業 研究成果報告書

平成 26 年 11 月 14 日現在

機関番号：22701

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23770180

研究課題名(和文)長時間全原子シミュレーションによるタンパク質揺らぎの解明

研究課題名(英文)Fluctuation dynamics of proteins revealed by long-time all-atom molecular dynamics simulations

研究代表者

淵上 壮太郎 (Fuchigami, Sotaro)

横浜市立大学・生命医科学研究科・助教

研究者番号：00381468

交付決定額(研究期間全体)：(直接経費) 3,400,000円、(間接経費) 1,020,000円

研究成果の概要(和文)：近年の計算機の目覚ましい発展にともない、タンパク質の運動を分子動力学シミュレーションで再現することができるようになった。しかし、タンパク質の複雑多様な運動の実態を把握・理解することは未だ容易ではない。

そこで我々は、シミュレーション結果からタンパク質の主要な運動を特定・抽出するために、タンパク質の「遅い運動」に着目し、シミュレーション結果から効率的に同定するための手法として「時間構造に基づいた独立成分分析(tICA)」を提案し、実際、この解析手法が有用であることを示した。

研究成果の概要(英文)：Molecular dynamics (MD) simulation is a powerful tool that is widely used to elucidate dynamic behavior of proteins and to reveal molecular mechanisms of their functions at an atomic resolution. Protein motions occur over a wide range of time scales, but not all are important for protein functions.

Because time scales of the functions are generally longer, it would be reasonable to consider that slower motions of proteins are more relevant to their functions. To identify and examine such slow dynamics of proteins from MD simulation results, we recently proposed a method of time-structure based independent component analysis (tICA). The results of its application to MD trajectories of proteins demonstrated the validity and usefulness of the tICA.

研究分野：生物学

科研費の分科・細目：生物科学・生物物理学

キーワード：生物物理 蛋白質 シミュレーション 揺らぎ

1. 研究開始当初の背景

タンパク質は柔軟性に富んだ動的な分子であり、生体内環境において大きく揺らいでいる。タンパク質はこの揺らぎを巧みに制御・活用することによって、リガンド輸送・酵素反応・シグナル伝達などの多様な機能を高精度で高効率に実現している。したがって、タンパク質の機能を解明するためには、立体構造やその分布などの静的な特徴だけでなく、タンパク質の動的側面、「タンパク質の揺らぎ」を理解することも重要である。

タンパク質の揺らぎにはメチル基の回転・ループ運動・ドメイン運動など様々な階層のものが含まれており、ピコ秒からミリ秒まで幅広い時間スケールに渡っている。一方、タンパク質の機能はマイクロ秒より遅い時間スケールで発現するのが一般的である。したがって、タンパク質の揺らぎの中でも、より遅い運動が機能に関わっている可能性が高いと考えられる。

近年の計算機の劇的な進歩に伴い、タンパク質の全原子分子動力学 (MD) シミュレーションをマイクロ秒以上の長時間に渡って実行することが可能となってきた。得られた結果にはタンパク質の動的な情報が原子レベルの精度で含まれており、それを解析することでタンパク質の機能発現メカニズムを明らかにできると期待される。しかし、タンパク質の揺らぎは大自由度空間内における複雑多彩なものであり、その運動を端的に特徴づけるのは容易なことではない。

MD シミュレーションの結果からタンパク質の主要な運動を抽出するための強力な手法として、主成分分析 (PCA) が広く利用されている。しかし、PCA による解析だけで、タンパク質の動的な振る舞いを十分に理解することは難しい。特に問題となるのが、運動の時間スケールを議論する場合であり、タンパク質の遅い運動の解明には PCA は最適な手段であるとは言えない。

そこで、我々はタンパク質の遅い運動の実態を明らかにすべく、新たな解析手法を提案し、その手法を「時間構造に基づいた独立成分分析 (tICA)」と名付けた。この tICA では、タンパク質の静的な情報だけでなく、動的な情報をも利用することから、タンパク質の遅い揺らぎを効率的に記述・理解できると期待される。この手法をベースとして、タンパク質揺らぎの動的機構・重要アミノ酸残基を明らかにすべく研究を開始した。

2. 研究の目的

タンパク質の MD シミュレーションの結果に埋もれている有用情報を効果的に抽出する手法を開発し、タンパク質の複雑多様な揺らぎの実態、その動的機構、機能との相関を解明する一連の枠組みを創設することを目指し、以下の3つの課題に取り組む。

(1) 解析手法としての tICA の確立

我々が提案する解析手法である tICA が、タンパク質揺らぎの動的な性質を抽出・理解するのに有用であることを確認すべく、様々なタンパク質のシミュレーション結果に tICA を適用することで、tICA の効果的な使用法および適用限界を見定め、タンパク質揺らぎの解析手法としての確立を目指す。

(2) タンパク質の遅い揺らぎの実態解明

タンパク質の遅い時間スケールで起こる運動は、タンパク質の機能実現に決定的な役割を果たしている可能性が高く、その運動の特定はタンパク質の機能発現機構の解明にとって重要である。そこで、マイクロ秒スケールの長時間シミュレーションを実行し、タンパク質の遅い揺らぎがどのような運動であるのかを明らかにすることを目指す。解析対象として、基質結合にともない大規模な立体構造変化を示すタンパク質を取り上げる。

(3) 動的情報に基づいた重要残基の特定

タンパク質揺らぎは機能を実現する上で重要な要素であり、突然変異体を用いた種々の実験から揺らぎの鍵となるアミノ酸残基の存在が示唆されている。このようなアミノ酸残基を特定すべく、シミュレーションで得られた原子レベルの情報をもとに、タンパク質揺らぎを支配するアミノ酸残基を推定する手法を開発する。また、推定したアミノ酸残基の妥当性を変異体のシミュレーションによって検証する。

3. 研究の方法

(1) タンパク質の長時間シミュレーション

タンパク質の複雑多様な揺らぎの実態を解明すべく、リガンドの結合に伴い大規模な立体構造変化を起こすことが知られている2つのタンパク質、「リジン・アルギニン・オルニチン結合タンパク質 (LAO)」と「マルトース結合タンパク質 (MBP)」を解析対象として選択した。

各タンパク質のリガンド非結合時の構造を、大量の水分子の中に埋め込むことで水溶液中の状態を再現した系を構築し、1マイクロ秒の MD シミュレーションを実行した。シミュレーションの実行には横浜市大の池口により開発された MD シミュレーションソフトウェア MARBLE を使用し、力場には CHARMM22/CMAP を用いた。

(2) tICA による遅い運動の特定

n 次元時系列データ $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$ に対して、その共分散行列

$$\mathbf{C} = \langle (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle) (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle)^T \rangle$$

および、時間遅れ共分散行列

$$\bar{\mathbf{C}} = \langle (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle) (\mathbf{x}(t + \tau_0) - \langle \mathbf{x}(t) \rangle)^T \rangle$$

を用意する (ここで、 τ_0 は遅延時間パラメータ)。この2つの行列を用いて、一般化固有値問題 $\bar{\mathbf{C}}\mathbf{F} = \mathbf{C}\mathbf{F}\mathbf{K}$ を解き、固有値行列 \mathbf{K}

と固有ベクトル行列 \mathbf{F} を求める。ここで、固有値や固有ベクトルの要素が複素数となるのを避けるため、対称化した $\bar{\mathbf{C}}$ を用いた。この対称化は、時系列が時間反転に関して対称であれば正当化される。

tICA で得られる固有ベクトル \mathbf{f}_i と、それと対になるベクトル $\mathbf{g}_i = \mathbf{C}\mathbf{f}_i$ を用いると、時系列データは

$$\mathbf{x}(t) = \sum_i \mathbf{g}_i^i \mathbf{f}_i^i \mathbf{x}(t) = \sum_i a_i(t) \mathbf{g}_i$$

のように展開される。これより、tICA では固有ベクトル \mathbf{f}_i ではなく、対となる \mathbf{g}_i が運動の方向を表わす独立成分 (IC) となる。固有値は運動の時間スケールを特徴づける量になっており、固有値が大きいほど対応する IC が表わす運動の時間スケールが遅いと考えられることができる。したがって、最大の固有値をもつ第一独立成分 (IC1) \mathbf{g}_1 が最も遅い運動の方向を表わしており、射影トラジェクトリー $\mathbf{a}_1(t)$ によってその動的振る舞いを確認することができる。

4. 研究成果

(1) tICA の有用性の検証

我々が提案する解析手法である tICA がタンパク質の遅い揺らぎを効率的に特定・理解できるかどうかを検証するために、タンパク質の長時間 MD シミュレーションを実行し、得られた時系列データに tICA を適用した。

解析対象のタンパク質として、2 つのドメインから成る「リジン・アルギニン・オルニチン結合タンパク質 (LAO)」を選択した。LAO は、ドメイン間に存在する裂け目部分にリガンドを特異的結合し、その際、ドメイン運動による大きな構造変化を示す。

この LAO の MD シミュレーションを 600 ナノ秒実行したところ、LAO が大きく揺らぐ様子が観察され、初期構造からの RMSD (C_α 原子のみ) の時間発展からも 100 ナノ秒オーダーの遅い揺らぎが含まれていることが明らかとなった。一方で、LAO の 2 つのドメインはいずれも安定であり、その立体構造はほとんど変化していなかった。よって、LAO の揺らぎは、ドメイン運動が支配的であり、剛体的なドメインの運動として近似できる。

そこで、剛体ドメイン解析によってこの近似を実行し、自由度を削減したトラジェクトリーについて tICA による解析を行った。tICA によって得られた IC について、各 IC が表わす運動の時間スケールを確認するため、それぞれの射影トラジェクトリーから相関時間を求めてみたところ、IC は期待通り時間スケールの遅い順に並んでいることがわかった。一方、各 IC による運動が全体の揺らぎに対して占める割合を求めたところ、遅い運動だからといって必ずしも揺らぎが大きいわけではないことも明らかとなった。

得られた上位 3 つの IC (IC1, IC2, IC3) が表わす運動 (図 1) を見てみると、PCA で得られる主成分 (PC) が表すものとは異なっ

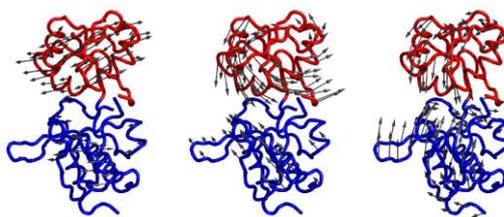


図 1: tICA によって特定された LAO の遅い時間スケールのドメイン運動。左から順に IC1, IC2, IC3 によって表わされる運動。

ていることが確認できた。特に、RMSD に見られた LAO の遅い運動に着目すると、PCA の場合、上位 2 つの PC にその運動が含まれていたのに対し、tICA では最も遅い IC である IC1 のみでうまく特徴づけられ、遅い運動を的確に抽出できていることがわかった。

以上により、tICA によってタンパク質の遅い運動を効率的に抽出し、その動的振る舞いを明らかにできることが検証できた。

(2) tICA によるタンパク質主鎖が示す遅い運動の解析

タンパク質の機能実現に必要と考えられる遅い運動は、ドメイン運動のような全体的な集団運動だけでなく、ループの揺らぎなどの局所的な運動として生じている可能性も高い。そこで、タンパク質の主鎖を対象として、どのような遅い運動が起こっているのかを tICA を用いて調べた。ただし、主鎖の運動は C_α 原子だけで十分に表現され得ると考えられるため、 C_α 原子のみを解析対象とした。

まず、一つ目の対象として、tICA の検証に用いたのと同じ LAO を選び、MD シミュレーションを 1 マイクロ秒の長時間に渡って実行した。得られた C_α 原子のトラジェクトリーに対して tICA を適用し、LAO 主鎖の遅い運動を抽出した。得られた上位の IC について、その運動の時間スケールを確認したところ、数十ナノ秒から数ナノ秒の時間スケールをもつ遅い運動であることが確認できた。

続いて、遅い方から 5 つの IC (IC1~IC5) がどのような主鎖の運動を表しているかを調べたところ、IC2 は全体的な運動であるドメイン運動であったのに対し、残りの 4 つ (IC1, IC3, IC4, IC5) ではごく少数の C_α 原子が著しい動きを示していたことから、それらの近傍で局所的な運動が生じていると考えられる。これらの局所運動の詳細を明らかにすべく、同定された部位の周辺について、主鎖二面角の変化や周辺残基との相互作用の変化を調べたところ、確かに遅い運動が生じていることが確認できた。

一例として、最も遅い運動を表す IC1 の結果を以下に示す。IC1 では 3 つの残基 (R218, Q219, D220) において顕著な変位が見られ、その周辺で局所的な運動が生じたことがわかった。これら 3 つの残基は、結晶構造において α -ヘリックスの末端に位置し、複数の水素結合によって安定な構造を形成している

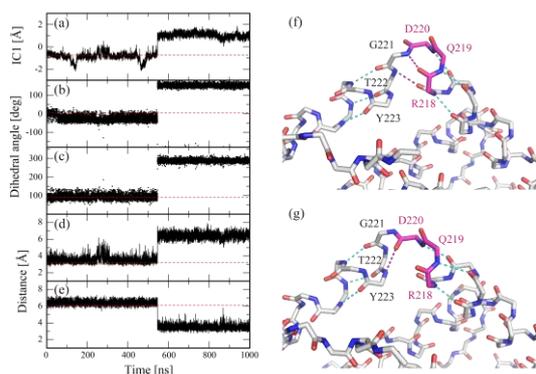


図 2: IC1 によって特定された LAO 主鎖の局所運動. (a)–(e) それぞれ, IC1, 主鎖二面角 $D220\psi$, 主鎖二面角 $G221\phi$, 原子間距離 $R218O-G221N$, 原子間距離 $D220O-Y223N$ の時間発展. 赤の破線は初期構造に用いた結晶構造の値. (f, g) 著しい動きを示す $C\alpha$ 原子周辺の結晶構造, および, 1 マイクロ秒後の構造における LAO 主鎖. 顕著な変位を示す $C\alpha$ 原子を含む残基の主鎖炭素原子を紫色で示した. また, 水色と紫色の破線は, 安定, および, 不安定な主鎖間の水素結合を表している.

(図 2(f)). 図 2(a) に示した IC1 の時間発展からは, この部分の立体構造が 545.2 ナノ秒で遷移的に変化し, その後シミュレーションが終わるまで遷移後の構造のまま安定でいたことが見て取れる. 実際, 結晶構造(図 2(f))と 1 マイクロ秒後の立体構造(図 2(g))とを比較すると, 残基 $D220$ と $G221$ の間のペプチド結合部分がクランクシャフト運動を起こすことで局所的に構造が変化していることがわかる. また, この運動に伴って 2 つの水素結合 $R218O-G221N$ と $D217O-T222N$ が切断され, 新たに $D220O$ と $Y223N$ の間に水素結合が形成されたこともわかる. クランクシャフト運動の遷移的な挙動は, 主鎖二面角 $D220\psi$ と $G221\phi$ の遷移や, 関連した水素結合の形成・切断が, IC1 と同一時刻でただ 1 回のみ起こっていることから確認することができる(図 3(b)–(e)). 以上より, IC1 で示唆された局所運動は確かに起こっており, その詳細を原子レベルで明らかにすることができた.

さらに, 2 つ目のターゲットとして, LAO と同じファミリーに属し, リガンドの結合にともない大きな構造変化を示すタンパク質であるマルトース結合タンパク質 (MBP) を取り上げ, 長時間のシミュレーションを実行した. LAO と同じく, MBP も 100 ナノ秒オーダーの遅いドメイン揺らぎを示すことが確認できた. また tICA による解析を行ったところ, LAO と同じく, 遅い運動としてドメイン運動と局所運動が特定され, tICA の有用性を確認することができた.

以上より, tICA が, タンパク質の遅い運動を, その運動が全体的であるか局所的であるかにかかわらず, うまく抽出することがで

きる有用な解析手法であることがわかった.

(3) tICA によるタンパク質主鎖が示す遅い運動の多様性の解析

tICA による解析によって遅い運動として特定されたものは, ゆっくりと変化するものだけではなく, 急激な変化が稀に起こるといった特徴を持っているもの (レアイベント) も特定された. このような運動は, その発生の有無・頻度がシミュレーションごとに異なることが予想される. そこで, 複数のシミュレーション結果を比較することによって, タンパク質揺らぎの複雑多様な実態を解明することを目指した.

LAO の 1 マイクロ秒のシミュレーションを 3 回実行したところ, いずれにおいても 100 ナノ秒オーダーの遅いドメイン運動が観察された. tICA によって LAO 主鎖の遅い運動を抽出したところ, その結果は 3 回のシミュレーションで大きく異なっており, 揺らぎの多様性が明らかとなった. 一方, 複数のシミュレーションで共通する遅い局所運動もみつけた. このような再現性のある運動をする部位は, 機能発現にとって重要である可能性が高いと考えられる.

(4) 立体構造変化の鍵となるアミノ酸残基の変異体シミュレーションによる検証

tICA の解析結果をもとに, LAO の遅い揺らぎで鍵となっていそうなアミノ酸残基として $K186$ と $F233$ を推定した. この推定を確認するため, それぞれをアラニンに置換した変異体 ($K186A$ と $F233A$) のシミュレーションを実行し, 変異の影響を調べた.

まず, 変異体 $K186A$ について 5 回のシミュレーションを行ったところ, いずれにおいても野生型とは異なったドメイン運動が見られ, 構造がリガンド結合型の方に大きく変化してリガンドの結合には不利な状況となる様子が観察された. 置換したリジン残基 $K186$ は野生型においてドメイン間に跨った塩橋を形成しており, この塩橋の消失が原因であったと考えられる.

比較のため, 野生型のシミュレーションも 5 回実行してみたところ, 5 回のうち 4 回はドメイン運動をしつつもリガンドが結合できるような構造を維持していた. ただし, 残りの 1 回は, $K186$ が $D238$ とついていた塩橋が壊れ, クローズ型へと変化しており, 野生型であっても鍵となる相互作用がなくなれば, 構造変化が起きることがわかった.

一方, 変異体 $F233A$ では $K186A$ で見られたような構造変化は起きておらず, 野生型と変わらない挙動であった. したがって, $F233$ は LAO の遅い揺らぎの鍵となっているわけではないようである.

(5) tICA によるタンパク質主鎖二面角が示す遅い運動の解析

LAO の主鎖で生じている遅い運動を別角

度から検証すべく、これまで用いていた C_{α} 原子のデカルト座標とは異なる主鎖二面角の時系列データに tICA を適用した。その結果、抽出された運動は C_{α} 原子のデカルト座標に対する tICA の結果とよく一致しており、tICA は座標系に依らず遅い運動を抽出できることがわかった。

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 4 件)

- ① Yusuke Naritomi and Sotaro Fuchigami, Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions, *J. Chem. Phys.* 134(6), 065101 (2011), 査読有, DOI:10.1063/1.3554380
- ② browse 上 壮 太 郎, 独立成分分析 tICA によるタンパク質ダイナミクスの解析, 分子シミュレーション研究会会誌「アンサンブル」, Vol. 13, No. 4, 61-166 (2011), 査読無, DOI:10.11436/mssj.13.161
- ③ Yusuke Naritomi and Sotaro Fuchigami, Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis, *J. Chem. Phys.* 139(21), 215102 (2013), 査読有, DOI:10.1063/1.4834695.
- ④ browse 上 壮 太 郎, 独立成分分析 tICA でタンパク質の複雑な運動を解きほぐす, 統計数理, 印刷中 (第 62 巻, 第 2 号, 2014), 査読有, <http://www.ism.ac.jp/editsec/toukei/index.html>.

[学会発表] (計 12 件)

- ① 成富佑輔、 browse 上 壮 太 郎、分子動力学シミュレーションによるタンパク質の遅い運動の解析、第 5 回分子科学討論会、2011/9/20、札幌市・札幌コンベンションセンター。
- ② Yusuke Naritomi, Sotaro Fuchigami, Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis, Biophysical Society 56th Annual Meeting, 2012/2/29, San Diego Convention Center, San Diego, California.
- ③ browse 上 壮 太 郎、独立成分分析 tICA によるタンパク質ダイナミクスの解析、日本物理学会第 67 回年次大会、2012/3/24、西宮市・関西学院大学西宮上ヶ原キャンパス。
- ④ browse 上 壮 太 郎、タンパク質の長時間シミュレーション中に含まれるレアイベントの解析、第 6 回分子科学討論会、2012/9/18、文京区・東京大学本郷キャンパス。

- ⑤ browse 上 壮 太 郎、独立成分分析 tICA によるタンパク質シミュレーションのレアイベント解析、日本物理学会 2012 年秋季大会、2012/9/20、横浜市・横浜国立大学常盤台キャンパス。
- ⑥ Sotaro Fuchigami, Rare events in protein simulation revealed by using time-structure based independent component analysis, Biophysical Society 57th Annual Meeting, 2013/2/3, Pennsylvania Convention Center, Philadelphia, Pennsylvania.
- ⑦ browse 上 壮 太 郎、タンパク質シミュレーションのレアイベント解析と複数比較、日本物理学会第 68 回年次大会、2013/3/27、東広島市・広島大学東広島キャンパス。
- ⑧ browse 上 壮 太 郎、リジン、アルギニン、オルニチン結合タンパク質の立体構造変化の鍵となる相互作用：変異体シミュレーションによる検証、第 13 回日本蛋白質科学会年会、2013/6/14、鳥取市・とりぎん文化会館。
- ⑨ browse 上 壮 太 郎、独立成分分析 tICA によるタンパク質主鎖二面角の遅い運動の解析、第 7 回分子科学討論会、2013/9/26、京都市・京都テルサ。
- ⑩ browse 上 壮 太 郎、独立成分分析 tICA を用いたタンパク質主鎖二面角のダイナミクス解析、日本物理学会 2013 年秋季大会、2013/9/27、徳島市・徳島大学常三島キャンパス。
- ⑪ Sotaro Fuchigami, Slow dynamics of protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis, 第 51 回日本生物物理学会年会、2013/10/28, 京都市・国立京都国際会館。
- ⑫ browse 上 壮 太 郎、独立成分分析 tICA を用いたタンパク質主鎖二面角のダイナミクス解析 II、日本物理学会第 69 回年次大会、2014/3/30、平塚市・東海大学湘南キャンパス。

[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]

特になし

6. 研究組織

(1) 研究代表者

browse 上 壮 太 郎 (FUCHIGAMI, Sotaro)
横浜市立大学・大学院生命医科学研究科・助教
研究者番号：00381468