

タンパク質分子の構造ダイナミクス： ウェーブレット変換による解析

鎌田 真由美¹・戸田 幹人²

(受付 2014 年 1 月 22 日；改訂 8 月 11 日；採択 8 月 21 日)

要 旨

タンパク質分子の立体構造ダイナミクスは、機能発現に深く関与していると考えられている。故に、構造ダイナミクスの特徴を捉え詳細な解析をする事は、その機能の本質的理解につながる。しかしタンパク質の構造ダイナミクスは、複雑な時空間における階層性を持っていることから、複数の時空間スケールに渡る解析は困難な課題である。本稿では、まずタンパク質の構造ダイナミクスに対するこれまでのアプローチを紹介し、さらに、連続ウェーブレット変換を用いた時系列解析手法について説明する。経時的な構造の変化に対して、時間-周波数解析手法の一手法であるウェーブレット変換と、行列分解手法である特異値分解(SVD)を共に用いる事で、運動の特徴を低次元で一様に抽出することが出来る。この解析手法の応用例として、*Thermomyces lanuginosa* lipase(TLL)に対して我々の行った研究成果について解説する。

キーワード：ウェーブレット解析，タンパク質立体構造，分子動力学，時系列解析。

1. はじめに

生体内においてタンパク質は静的に存在するのではなく、常に動きを伴った状態で存在している。この構造ダイナミクスは、タンパク質の機能発現において重要な役割を果たしていると考えられている(Frauenfelder et al., 1991)。それ故、タンパク質のダイナミクスとその機能との関係性を明らかにすることは、機能発現を引き起こす要因を知る上で重要であり、疾患のメカニズム等の様々な生理現象の本質的な理解への鍵となる。

タンパク質の構造ダイナミクスの特徴の一つに、時空間階層性がある(Henzler-Wildman and Kern, 2007)。これは、多自由度系であるタンパク質分子が、そのエネルギー地形に唯一の深いエネルギー極小点を持つのではなく、エネルギー極小点が多数、階層的に存在する複雑なエネルギー地形を持っていることに起因している。十分に低温な状況下において、タンパク質は1つのエネルギー極小付近に拘束されたfs~psの振動をしている。このような調和的な微細振動は、基準振動解析(Normal Mode Analysis, NMA)によりうまく再現される事がわかっており、これまでにNMAを用いた解析手法が報告されている(Hayward and Go, 1995; Brooks et al., 1995)。しかし、温度の上昇に伴い非調和な振動が次第に大きくなり、局所的なエネルギー極小間での遷移が偶発的に引き起こされ、ナノ秒スケールを越える構造変化のような大きな動きを現す。このように、タンパク質のダイナミクスは、複数の時間スケールにまたがる調和・非調和性を

¹ 慶應義塾大学 理工学部：〒 223-8522 神奈川県横浜市港北区日吉 3-14-1

² 奈良女子大学 研究院自然科学系：〒 630-8506 奈良県奈良市北魚屋西町

持つ複雑なものである (Zaccai, 2000).

では、どのようにこのダイナミクスを捉え、理解すれば良いのか。分子動力学 (Molecular Dynamics, MD) シミュレーションを用いると、複雑な階層性を持つタンパク質の挙動に対し、分子の運動に関する種々の物理量を各原子レベルから得ることが出来る。これまで、MD に代表されるような分子シミュレーションで観測できる時間領域はフェムト秒からナノ秒前後であり、タンパク質の特徴的な動きを示すフェムト秒から秒にまたがる領域をカバーできていなかった。しかし近年の計算機技術の発展から、ミリ秒スケールのシミュレーションが可能になりつつある (Lindorff-Larsen et al., 2011)。上述のダイナミクスの非調和性も、1990年代に入りシミュレーション時間が長くなった事により観測されるようになった一例であることから、シミュレーション時間が更に長くなることにより、これまでの既存モデルや理論では説明出来ないイベントが観測されるようになると考えられる。特に、ミリ秒スケールで起こる機能に直結するような構造変化、例えば酵素基質結合等は重要なトピックである。これまで酵素基質反応は、鍵と鍵穴モデルにより説明されてきた。即ち、酵素と基質が鍵と鍵穴の関係となっており、形状として相補的な場合のみ互いを認識し反応が起こるというモデルである。実際、これまで既に立体構造が決定されているタンパク質の多くは、この様に基質結合の際、立体構造変化をほとんど伴わない。しかし、実験的にも理論的にも様々なタンパク質に対する研究が進むにつれ、このモデルでは説明できない分子の挙動が観測されるようになった。それらを説明する為に提唱された考えとして、“induced-fit (誘導適合)” と “population shift (pre-exsiting/selected-fit model)” がある。“induced-fit” は、1958年に Koshland (Koshland, 1958) によって提唱された考え方で、分子が基質と出会う事でその構造を結合構造へと変化させる、というアイデアである。一方、“population shift” は、他分子との結合の有無に関わらず、常に複数の構造まわりで立体構造を変化させている、というものである。前者は従来最もよく受け入れられている考えであり、これまでに多くのタンパク質間相互作用やタンパク質-DNA/RNA 間相互作用に対して適用され、そのメカニズムが明らかにされている (Bui and McCammon, 2006; Levy et al., 2007; Williamson, 2000)。一方、主に免疫抗体反応や小さな分子との相互作用において、近年後者の考えが注目を集めており、これに基づく解析によっても様々なタンパク質の機能動態メカニズムが明らかにされてきている (Kumar et al., 2000; Henzler-Wildman et al., 2007; James et al., 2003; Arora and Brooks, 2007; Bahar et al., 2007; Okazaki and Takada, 2008)。よって、長時間シミュレーションにより得られるトラジェクトリデータに対して、“induced-fit” モデルだけでなく、“population shift” のような経時的な構造遷移に着目した解析も今後必要となってくる。これらのことから、複数の時空間階層を持つ複雑なダイナミクスに対して、複数の時空間スケールに渡ってその相関などを捉え、正確に特徴付ける手法開発が望まれている。本稿では、タンパク質立体構造のダイナミクスに関する解析手法について解説するとともに、ウェーブレット変換を用いたタンパク質立体構造ダイナミクスに対する新しい時系列解析手法 (Sakurai, 2010) について紹介し、我々がこれまでに行った研究成果 (Kamada et al., 2011) について報告する。

2. 構造ダイナミクスに対する MD シミュレーションを用いた解析手法

前節で述べたように、タンパク質のダイナミクスは、MD 法などの分子シミュレーションによって捉える事が出来る。シミュレーションにより得られたトラジェクトリデータを用いて解析をする事で、ダイナミクスと機能との関係を調べる事ができる。最も一般的な解析方法として、Root Mean Square Distance (RMSD) や Root Mean Square Fluctuation (RMSF) 等を用いるもの、そして主成分分析 (Principal Component Analysis, PCA) を用いるものがある。前者はトラジェクトリデータの各原子の空間座標を用いて構造のズレを計算し、その遷移を直接的に

評価するものである。しかし、タンパク質の動きは異方性が高いことから、このような各原子のデカルト座標から、揺らぎの方向を原子座標の線形結合で表す集団座標を構成し、それを用いた評価を行う方がより適切に運動を特徴付けられる。この時、タンパク質分子の各原子における主要モードを決定し要約することが出来る PCA が有効であるとされ、これまでに多くの解析で用いられている (Balsera et al., 1996; Kitao and Go, 1999)。また、グラフ理論を用いた解析事例もある。グラフとは、ノードの集合と各ノードをつなげるエッジにより構成されるものである。タンパク質への応用としては、各グラフノードをタンパク質の各残基とし、各エッジを空間上の距離や相互関係等、残基間の関係として表現するのが一般的な方法である。グラフ化することにより計算機的にも扱いやすくなることから、タンパク質相互作用ネットワーク等の解析には多数の応用事例がある (Brinda and Vishveshwara, 2005)。MD トラジェクトリ解析としては、時系列におけるタンパク質の構造的差異やイベントの検出 (Swint-Kruse, 2004; Wriggers et al., 2009; Benson and Daggett, 2012b) など適用例があるものの、未だその数は少ない。

3. ウェーブレット変換を用いた構造ダイナミクスの特徴抽出

非調和性を含んだタンパク質のダイナミクスは、時間変化に伴いその動きの振幅と各周波数成分の大きさを変化させる。つまり、タンパク質のダイナミクスは周波数と振幅の変化として捉える事が出来る。これに対して、時間-周波数解析手法である“ウェーブレット解析”を用いることで、時間軸を残したまま経時的な動きの特徴をうまく捉える事が出来る。ウェーブレット変換は、連続ウェーブレットと離散ウェーブレットに大別され、共にこれまで画像圧縮や計測、音響信号処理等の多分野に渡り応用されている。生体分子に対しても応用されており、MD 解析への適用事例もいくつかある (Askar et al., 1996; Vela-Arevalo and Wiggins, 2001; Liò, 2003; Benson and Daggett, 2012a)。これらの解析では、構造間遷移の様なレアイベントや重要な動きを含む時間領域の検出、ノイズを含む高周波層の不要データを取り除いて時間周波数領域で局在した特徴を得ること等を行っており、ウェーブレット解析が MD のトラジェクトリ解析に有用であることを示している。また、Matsunaga 等の研究では、離散ウェーブレット解析と PCA を組み合わせた方法により、タンパク質の粗視化モデルを対象に、フォールディングにおいて振動エネルギーが異なる周波数領域間を移動する過程について解析を行っている (Matsunaga et al., 2013)。ただ、個々のタンパク質の機能発現とダイナミクスとの関係を明らかにしたいと考えたとき、具体的にどの残基が重要な動きを示しているのか、また、その残基と他の残基との関係がどうなっているのかを経時的に解析する事が求められる。そこで、すでに示されているウェーブレット変換の時間-周波数における局在的な特徴抽出能力を生かし、個々の残基の動きを特徴付けるような指標を導入して、ダイナミクスの時系列解析へ適用する。以降、連続ウェーブレット変換をベースとした解析手法 (Sakurai, 2010) について紹介する。手法の大きな流れを図 1 に示す。

この方法は、連続ウェーブレット変換と特異値分解 (Singular Value Decomposition, SVD) の大きく分けて 2 つのステップから成る。まず、ウェーブレット変換は、ウェーブレット関数 $\psi_{a,b}(t)$ と対象となる信号 $f(t)$ の内積により下記のように定義される。

$$(3.1) \quad W(a,b) = \langle \psi_{a,b}(t), f(t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^* \left(\frac{t-b}{a} \right) f(t) dt$$

ここで、 $W(a,b)$ はウェーブレット係数、 $a > 0$, $b \in \mathbb{R}$ は、各々スケールパラメータ、シフトパラメータと呼ばれる。ウェーブレット関数 $\psi_{a,b}(t)$ は、マザーウェーブレットと呼ばれる基底関数 $\psi(t)$ とスケールパラメータ a 、シフトパラメータ b を用いて次のように定義される。

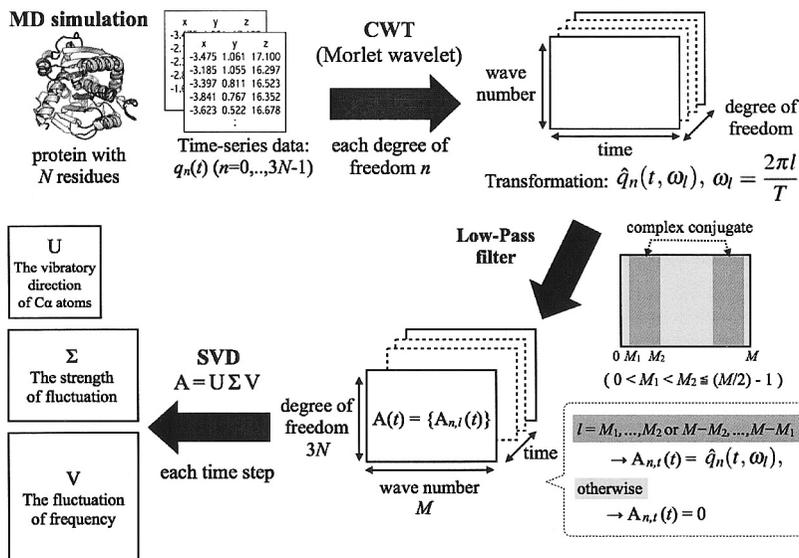


図 1. 解析手法全体の流れ。MD シミュレーションにより得られる C α 原子の時系列データに対し、各自由度において連続ウェーブレット変換 (Continuous Wavelet Transform, CWT) を行う。次に、得られた変換データに対しローパスフィルタを適用する。ローパスフィルタでは、対象とする低周波層のみに変換データの値を保持し、その他には 0 を当てはめる。フィルタリング後のデータに対し、次元圧縮の為に特異値分解 (Singular Value Decomposition, SVD) を適用する。SVD により得られる 3 つの行列 U , Σ , V は、各々、C α 原子の振動方向、振動の大きさ、周波数成分の変動に関する特徴を表す。

$$(3.2) \quad \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right)$$

ウェーブレット変換では、このウェーブレット関数を 2 つのパラメーター a , b で拡大縮小させることにより、フーリエ変換等では失われてしまう時間領域の情報を残したまま、時間-周波数の解析を行う。連続ウェーブレット変換で用いられる主なマザーウェーブレットとして、Mexican Hat, Meyer wavelet, Morlet wavelet 等があげられる。ウェーブレット変換では使用するマザーウェーブレットにより、計算速度や時間-周波数の各領域で得られる情報の精度等が異なる。とはいえ、式の形から各マザーウェーブレットの大まかな性格等はわかるものの、一概にどのマザーウェーブレットがどのようなデータに適しているのか断言する事は難しく、その選択には対象毎に考察が必要であると考えられる。本解析手法では、次式で表される Morlet wavelet を用いる。

$$(3.3) \quad \psi(t) = C e^{-\frac{t^2}{\sigma^2}} \left(e^{-i\pi t} - e^{-\frac{\pi^2 \sigma^2}{4}} \right)$$

ここで σ は窓関数の窓の中に入っている波の数を表す。Morlet wavelet はガウシアンカーネルによって重み付けされた正弦曲線の形を持っていることから、時系列における局所的な振動成分を捉える事が出来る。その為、脳波や脳磁図の解析にも良く用いられている。また σ を無限大にする極限で通常のフーリエ変換に漸近することから、フーリエ変換との比較がしやすい。ゆえに、本解析でも C α 原子の時系列データに対して、Morlet wavelet を用いた連続ウェーブレッ

ト変換を行う。ここで、 N 個の $C\alpha$ 原子の時系列データ $q_n(t)$ ($n = 0, \dots, 3N - 1$) を、 N 残基のタンパク質に対する MD シミュレーションによって取得したとする。まず、ウェーブレット変換を各自由度に対して適用し、変換データ $\hat{q}_n(t, \omega_l)$, $\omega_l = \frac{2\pi l}{T}$ を得る。ここで T は時系列の長さであり、実際の計算では、連続ウェーブレット変換を M 個の離散的な時系列データで近似している。変換された時系列データから、ゆっくりとした動きを取得する為に、低周波成分のみを残すローパスフィルタを施す。周波数 ω_l の範囲を、 $l = M_1, \dots, M_2$, ($0 < M_1 < M_2 \leq \frac{M}{2} - 1$) および $l = M - M_2, \dots, M - M_1$ とする。ここで、 M_1 は、時系列データ端の不連続性に起因する効果を避けるカットオフであり、 M_2 がローパスフィルタで残す低周波領域を指定する。 l の範囲が $\frac{M}{2}$ に関して対称なのは、有限個の実時系列データの変換で出てくる複素共役な振動成分を残すためである。以上のウェーブレット成分を用いて、行列 $A(t) = \{A_{n,l}(t)\}$ を作成する。行列の各成分は、上記のローパスフィルタの範囲 l に含まれる場合に、 $A_{n,l}(t) = \hat{q}_n(t, \omega_l)$ の値を持ち、それ以外は $A_{n,l}(t) = 0$ を持つ。次に各時刻 t に対して、SVD を行列 $A(t)$ に適用する。これにより、左特異ベクトルからなる行列 $U(t)$ 、右特異ベクトルから成る行列 $V(t)$ 、そして非負の特異値を対角成分に持つ行列 $\Sigma(t)$ を取得する。この際に、行列 $U(t)$ は実行列に取ることができることに注意する。また、行列 $V(t)$ は互いに複素共役な振動成分から構成される。 $U(t), V(t), \Sigma(t)$ は各々、 $C\alpha$ 原子の振動に関する空間での動きの特徴、それらの周波数における振動の特徴、そして振幅を表している。

4. タンパク質への応用

前節で紹介したウェーブレット変換による特徴抽出手法を用いた応用事例として、我々が行った解析について報告する。

4.1 対象タンパク質

すでに記述したように、タンパク質立体構造は時々刻々と変化し、複数の構造を渡り歩いていると考えられる。そこで、同一分子に対して複数の構造が実験的に明らかになっている、*Thermomyces lanuginosa* lipase (TLL) を解析対象として扱うこととする。TLL は、残基数 269 の比較的小さなタンパク質であり、脂肪の加水分解を触媒する酵素群 (EC: 3.1.1.3) の一つである。リパーゼはその構造内に α/β -hydrolase 構造を持っており、活性部位の安定性に関わっている。この活性部位は、“lid” と呼ばれる蓋のようなもので覆われており、“lid” に相当する残基等も明らかになっている (PDB: 1TIB, 残基番号 84~93)。TLL は、この “lid” が開くと活性部位に基質が結合出来るようになり、酵素活性を示すようになる。よって、これ以降、“lid の開いた構造” を “open (活性) 構造” と呼び、“lid が閉じている構造”、つまり活性を示さない状態の構造を “closed (不活性) 構造” と呼ぶこととする。TLL は、タンパク質立体構造データベースである PDB に X 線結晶構造解析などで明らかになった 7 つの立体構造が登録されており、3 つの open 構造と 4 つの closed 構造を含んでいる (Shu et al., 2009)。本解析では、closed 構造の 1 つである 1TIB を分子動力学シミュレーションの初期構造として用いる。図 2 に、1TIB の 2 次構造と 3 次構造を示す。更に、7 つの TLL 構造すべての B-factor を図 3 に示す。B-factor は原子の熱振動の大きさを表す温度因子であり、その大きさは原子の運動性の高さの判断基準となることから、各々の B-factor の違いは各構造の揺らぎの違いとして捉えることが出来る。

4.2 データ

ここでは解析に用いた時系列データについて説明する。時系列データは分子動力学 (MD) シミュレーションによって取得する。今回は AMBER9 の SANDER プログラムを用いて TLL の

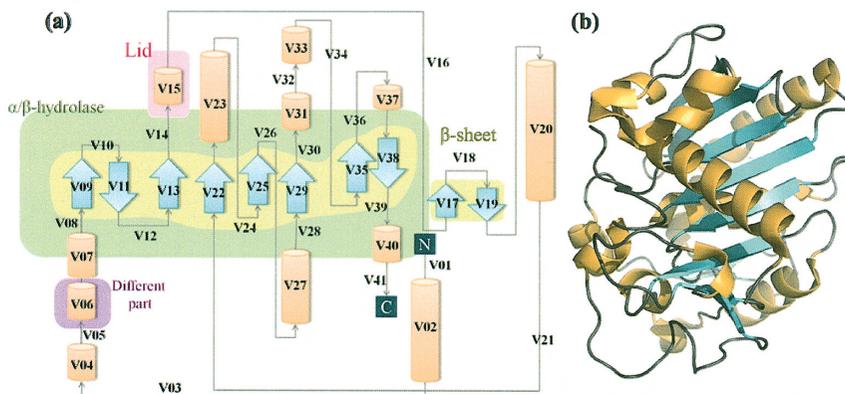


図 2. TLL の closed 構造の 1 つである 1TIB の (a) 2 次構造と (b) 3 次構造。円柱と矢印は各々 α -helix と β -strand を示している。1TIB は 12 の α -helix と 10 の β -strand, 19 の loop から構成されている。また、黄緑色で囲まれた、8 つの β -strand と 2 つの β -sheet から成る、2 つの β -sheet を持っている。本稿では、各 2 次構造に対し N 末端から順番に番号 k ($k = 1, \dots, 41$) を割当て、例えば、v01 や v12 の様に識別する。“lid” (v15) と v06 は、開閉構造で変化する部分である。TLL 構造の特徴の 1 つである、 α/β -hydrolase は、v09, v11, v13, v22, v25, v29, v35, v38 の β -strand から成る。2 次構造に関するこれらの情報はタンパク質立体構造データベース「PDBsum」から取得した。(a) is reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

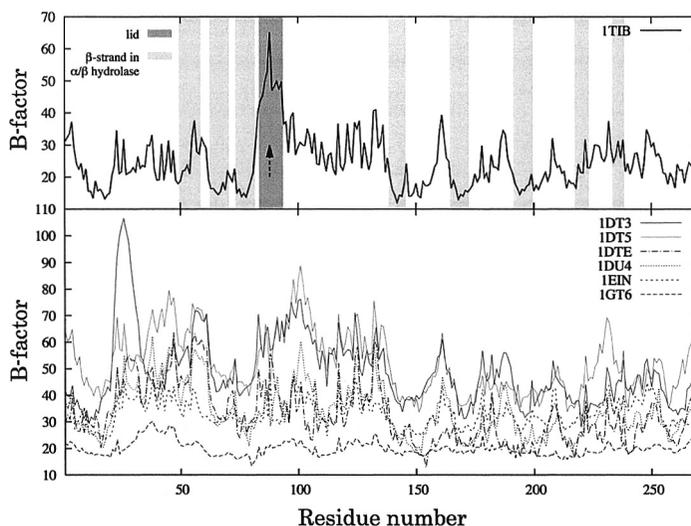


図 3. TLL 構造の持つ B-factor プロット。横軸は残基番号、縦軸は B-factor の値である。計 7 つの TLL 構造、3 つの open 構造 (1DT5, 1EIN, 1GT6) と 4 つの closed 構造 (1DTE, 1DU4, 1DT3, 1TIB) について、(a) 1TIB, (b) その他の構造を示している。(a) 背景の色は、薄い灰色が α/β -hydrolase を構成する 8 つの β -strand, 濃い灰色が “lid” に相当する部分を表している。矢印は “lid” を指し、他の構造とは異なり、1TIB の B-factor は大きな値 (ピーク) を示すことが分かる。

MD シミュレーションを行った。シミュレーションの手順は以下の通りである。まず初期構造（今回は 1TIB）を PDB から取得し、共役勾配法を用いてポテンシャルエネルギーの最小化を 2000 ステップ行う。次に系を 24Å の cut-off を用いて 0 K から 300 K まで 10 ps 毎に 5 K ずつ構造緩和を行う。温度設定には、AMBER ff99 force field と温度圧力一定の NTP アンサンブルを用い、溶媒分子は Hawkins 等による Generalized Born (GB) モデルを用いて陰に扱う。これらの処理の後、300 K でシミュレーションを行った。計算は、24Å cut-off の GB モデルを用いて 300 K で 1 fs のタイムステップで行い、10 ps ごとに間引いてトラジェクトリデータを得る。解析に用いた時系列データは、C α 原子の空間座標のみから成り、回転・並進は取り除いている。

4.3 特徴抽出手法の適用

3 節で説明した手法を同様の手順で TLL の時系列データに適用する。まず、トラジェクトリデータの各自由度に対してウェーブレット変換を行う。本解析では、Morlet wavelet の窓関数パラメータ σ の値として $\sigma = 5$ を用いる。TLL は $N = 269$ 残基から成り、保存された構造は 10 ps 刻み、時系列全体の長さは 2 ns である。故に、時系列の数 M は合計 $M = 200$ である。時系列の有限性からの影響をさける為に、これ以降の解析ではウェーブレット変換で得られた時系列のうち 0.2~1.8 ns のデータを用いる。また、構造のゆったりとした大きな動きに着目する為、 $M_1 = 15$, $M_2 = 50$ のローパスフィルタを適用する。これらは周波数 133~40 ps に対応している。よって、ウェーブレット変換とローパスフィルタの結果、 $[3N = 807 \text{ 行}] \times [2(M_2 - M_1 + 1) = 72 \text{ 列}]$ の行列 $A(t)$ を得る。次に、得られる各時刻 t での行列 $A(t)$ に対して SVD を適用する。その結果、高々 $K = \min(3N, 2(M_2 - M_1 + 1))$ の非零特異値と、各々に対応する左特異・右特異ベクトルが得られる。

SVD を適用する目的は、自由度の次元削減である。そこで、どのくらいの削減が可能なのか、各特異値の重みを以下のように計算し、確認を行う。

$$(4.1) \quad W_j(t) = \frac{(\lambda_j(t))^2}{\sum_{i=1}^K (\lambda_i(t))^2},$$

ここで、 $\lambda_i(t)$ は時刻 t における第 i 特異値である。図 4 に、第 1, 2 特異値の重み $W_1(t)$, $W_2(t)$, 更にその合計 $W(t) \equiv W_1(t) + W_2(t)$ の時間変化について示す。グラフから第 1 特異値が全体を通しておよそ 30% を、第 2 特異値までを考慮するとおよそ 50% を占める事が解る。

以降の解析では第 1 特異値のみに着目し、ここでは第 1 左特異ベクトル $e_{i=1}(t)$ を用いた解析について述べる。SVD の結果得られる左特異ベクトルは、ダイナミクスの空間における動きの特徴を現している。まず 4.4 節で、C α 原子の集団運動を特徴付ける指標を導入し解析を行う。次に 4.5 節で、2 次構造の動きを定量化し、2 次構造間の動きの相関関係について解析を行う。

4.4 特異ベクトルを用いた解析：タンパク質の集団運動

4.4.1 集団運動を特徴付ける

本稿での“タンパク質の集団運動”とは、隣接する C α 原子のまとまった動きの事を指す。隣接する C α 原子の範囲を広くとれば、剛体的な運動が抽出され、狭く選べばフレキシブルな運動も抽出される。この集団運動を評価する為、上記手法により得られた特徴ベクトルを用い、指標を導入する。ここで、 $\mathbf{u}_p(t)$ を時刻 t における p 番目の C α 原子の振動ベクトルと呼ぶこととし、第 1 左特異ベクトル $e_{i=1}(t) = (\mathbf{u}_1(t), \dots, \mathbf{u}_p(t), \dots, \mathbf{u}_N(t))$ を用いて定義する。振動ベクトル間の類似度、つまり C α 原子間の振動の類似性を、内積もしくは余弦を用いて定義する。また、集団(=C α 原子のまとまり)を“1 次構造上の近さ(1 dimension)”もしくは“3 次元構造上での近さ(3 dimension)”の 2 種類で定義する。前者は 1 次構造であるアミノ酸配列上での近さの

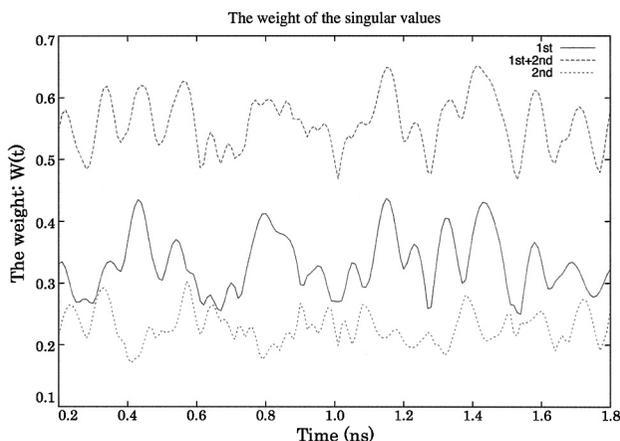


図4. 特異値分解により得られる特異値について, 第1特異値 $W_1(t)$ と第2特異値 $W_2(t)$, そしてその合計 $W(t) \equiv W_1(t) + W_2(t)$ が全特異値に占める割合を重みとして示している. 重みは(4.1)により計算される. 横軸は時刻 t ns, 縦軸は重みである. Reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

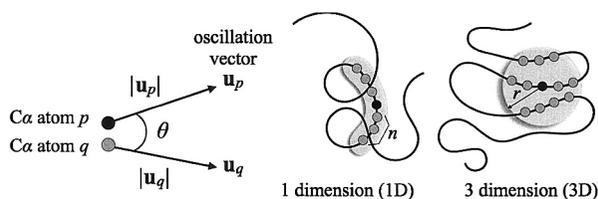


図5. 左: p 番目と q 番目の $C\alpha$ 原子の振動ベクトル $\mathbf{u}_p, \mathbf{u}_q$. この2つのベクトル間の類似度として, 内積と余弦を用いる. 右: 原子集団の定義. n は p 番目の $C\alpha$ 原子からのアミノ酸配列上の隣接原子の数, r は p 番目の $C\alpha$ 原子からの空間上の距離である. この n, r 内に含まれる原子が, 各 $C\alpha$ 原子の隣接原子(=集団)となる.

みを考慮するものである. 後者は立体構造の3次元空間上での近さを考慮しており, β -strandの様な, アミノ酸配列上では離れていても空間的に近くまとまった動きを示すと考えられるものに関して, その集団性を捉える事を意図している. また, 振動ベクトルの内積は振幅の大きさを考慮した集団性であるのに対して, 余弦を用いると振幅は小さくとも方向が揃った集団運動を抽出できることが期待される. これらの定義を図5に示す.

類似度の定義と集団の定義から, 表1にまとめたように, 集団運動に対する4種類の指標を導入する. n は p 番目の $C\alpha$ 原子から隣接する $C\alpha$ 原子の数を表している. また, r は p 番目の $C\alpha$ 原子からの空間上での距離である. つまり, 3次元構造での定義では, p 番目の $C\alpha$ 原子を中心とした半径 r の球を考え, その球内に含まれる $C\alpha$ 原子が集団の構成原子となる.

4.4.2 集団運動の時間変化

図6は, 先述の4つの集団運動指標 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) の時間変化を各々プロットしたものである. ここでは, $n = 10, r = 10\text{\AA}$ とした. 明るい色ほど各指標 $x_p^{(i)}(t)$ で定義される集団運

表 1. 集団運動性の評価指標. 2つの集団の定義(1 dimension, 1D; 3 dimension, 3D)と2つの類似性の定義(内積, 余弦)から, 計4つの指標を導入する. $\mathbf{u}_p(t)$ と $\mathbf{r}_p(t)$ は, 各々, 時刻 t における p 番目の C α 原子の振動ベクトルと座標ベクトルである. $b_{p,r}(t)$ は, p 番目の C α 原子から半径 r 内の距離にある C α の数である.

Inner Product - 1D	Inner Product - 3D
$x_p^{(1)}(t) \equiv \frac{1}{2n-1} \left \sum_{ p-q < n} \mathbf{u}_p(t) \cdot \mathbf{u}_q(t) \right $	$x_p^{(2)}(t) \equiv \frac{1}{b_{p,r}(t)} \left \sum_{ \mathbf{r}_p(t) - \mathbf{r}_q(t) < r} \mathbf{u}_p(t) \cdot \mathbf{u}_q(t) \right $
Cosine - 1D	Cosine - 3D
$x_p^{(3)}(t) \equiv \frac{1}{2n-1} \left \sum_{ p-q < n} \frac{\mathbf{u}_p(t) \cdot \mathbf{u}_q(t)}{ \mathbf{u}_p(t) \mathbf{u}_q(t) } \right $	$x_p^{(4)}(t) \equiv \frac{1}{b_{p,r}(t)} \left \sum_{ \mathbf{r}_p(t) - \mathbf{r}_q(t) < r} \frac{\mathbf{u}_p(t) \cdot \mathbf{u}_q(t)}{ \mathbf{u}_p(t) \mathbf{u}_q(t) } \right $

動が強い事を示している. 類似度に内積を用いた場合の結果から, 残基 90 の C α 原子が時刻 $t = 0.45$ ns あたりで特徴的な動きを示している事がわかる. 一方で, 余弦を用いた結果からは, 広範囲にわたる集団運動の隆起が確認された. また, 集団の定義による違いは見て取れなかった. これは今回の対象タンパク質が比較的小さい分子であることから, 集団定義に関するパラメータ n と r により定められる集団自体に大きな差が無い為だと考えられる. つまり, 対象タンパク質が更に大きな分子であったり, β -strand を多数含む様な場合に, この“1D”と“3D”の集団の定義の違いが有意な差として現れると考えている.

4.4.3 異なる時刻における集団運動性の違いについて

4.4.2 節において, 時間変化に伴い各原子の集団運動が異なる動きの特徴を示していることを述べた. そこで, クラスタリング手法を用いてこれらの特徴を時刻ごとに分類することで, C α 原子の集団運動性と各 TLL 構造との相関関係について調べる.

●クラスタリング

異なる時刻における集団運動を, Affinity Propagation (AP) と呼ばれるクラスタリング手法 (Frey and Dueck, 2007) を用いて分類する. この手法では, クラスタ数を “preference” と呼ばれるパラメータで調整する事が出来, 各クラスタは “example” と呼ばれるクラスタ中心によって表される. ここで, 異なる時刻 t_i, t_j 間での集団運動の類似度を下記のように定義する.

$$(4.2) \quad s(t_i, t_j) = -\sqrt{\sum_{p=1}^N |x_p^{(2)}(t_i) - x_p^{(2)}(t_j)|^2},$$

今回は4つの指標のうち $x_p^{(2)}(t)$ を用い, クラスタ数が4つになるようパラメータを決定した. 4つのクラスタにおけるクラスタ中心は, 時刻 $t_{c_1} = 0.42$ ns, $t_{c_2} = 0.97$ ns, $t_{c_3} = 1.34$ ns, $t_{c_4} = 1.54$ ns である. 図6中にこれら4つのクラスタ中心を赤い矢印として示している.

●集団運動と TLL の構造間での相関関係

各クラスタ中心 t_{c_j} ($j = 1, \dots, 4$) に対して計算された $x_p^{(2)}(t_{c_j})$ と7つの TLL 構造が持つ B-factor 間での相関値を表2(a)に示す. クラスタ中心 $t_{c_1} = 0.42$ ns は, 初期構造である 1TIB

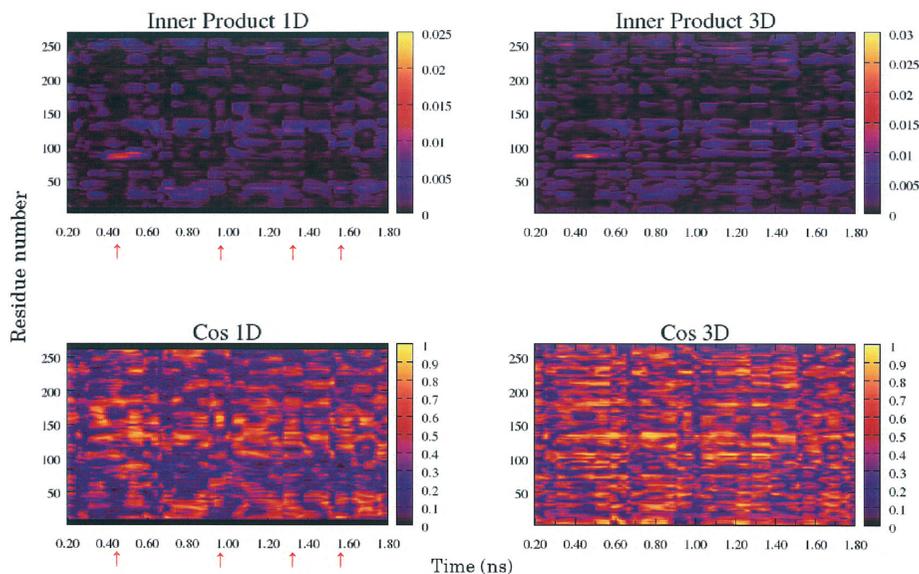


図 6. 集団運動性を評価する 4 つの指標 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) の時間発展. 横軸は時刻 (ns), 縦軸は残基番号 p である. 本稿では, $n = 10$, $r = 10 \text{ \AA}$ を使用している. 明るい色ほど指標 $x_p^{(i)}(t)$ により評価される集団運動が強い事を示している. 1D のプロットの下に添えた赤矢印は, 4.4.3 節で説明するクラスタの中心となった 4 つの時刻である. Reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

の B-factor と大きな相関を示している. これは “lid” が時刻 0.42 ns 辺りに特徴的な動きを示している事や, 図 3 において “lid” が 1TIB の B-factor で特徴的なピークを示していたことから妥当な結果であるといえる. 他のクラスタ中心に関しては, 複数の構造との間で大きな相関が確認できる. 立体構造での類似度を確認する為に, 構造を重ね合わせた際のズレの大きさを表す値, 最小二乗平均距離 (Root Mean Square Distance, RMSD) を, 各々の TLL 構造と各クラスタに属する全てのメンバー間で GASH アルゴリズム (Standley et al., 2007) を用いて計算する. 各クラスタごとの平均値を表 2 (b) に示す. $t_{c_1} = 0.42 \text{ ns}$ をクラスタ中心とする cluster1 は, 1TIB の構造と最も小さい RMSD 値を持つが, その他の cluster は, 他の構造とより近い構造を取っている事が確認できる. このような遷移の特徴は, この時間スケールのダイナミクスが, 異なる構造に対応する複数のポテンシャル井戸近傍の運動特性を反映していることを示している. より長い時間スケールのダイナミクスの特性に着目すれば, 個々のポテンシャル井戸近傍の局所的な振動とは質的に異なる, 新たな運動特性を示す可能性もあると予想されるが, この方向の研究は今後の課題である.

4.5 特異ベクトルを用いた解析: 2 次構造の集団運動

次に, 2 次構造の集団運動に関して解析を行う. 図 7 に時刻 $t = 0.6 \text{ ns}$ における各指標 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) と 2 次構造との関係性をプロットした. 横軸は残基番号 p , 縦軸は時刻 $t = 0.6 \text{ ns}$ における各指標の値 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) である. 更に, 代表的な 3 つの 2 次構造, α -helix, β -strand, loop を, 各々ピンク, ブルー, 白色で表している. 内積を用いた指標の結果におい

表 2. (a) 4 つのクラスタ中心 t_{c_j} ($j = 1, \dots, 4$) の指標 $x_p^{(2)}(t_{c_j})$ と各 TLL 構造の持つ B-factor との相関係数. *が付く相関係数は, 有意水準 5% で有意でないと判定されたものである. (b) 各 TLL 構造と各クラスタ内メンバー間での RMSD の平均. 平均は各クラスタに含まれるメンバーの数で算出しており, RMSD の計算は GASH プログラムを用いた.

(a)	Open			Closed			
	1DT5	1EIN	1GT6	1DTE	1DU4	1DT3	1TIB
	Correlation						
exemplar1 (0.42 ns)	0.15899	0.02312*	-0.05581*	0.23075	0.06115*	0.21374	0.51528
exemplar2 (0.97 ns)	0.17992	0.27449	0.24979	0.22586	0.22090	0.06142*	0.18510
exemplar3 (1.34 ns)	0.40421	0.36016	0.1806	0.26506	0.33449	0.30310	0.42954
exemplar4 (1.54 ns)	0.27327	0.24812	0.36855	0.18698	0.27374	0.00776*	0.27273
(b)	RMSD						
cluster1 (0.42 ns)	1.4284	1.4063	1.3946	1.4684	1.4350	1.5169	1.3916
cluster2 (0.97 ns)	1.4996	1.4419	1.4576	1.4896	1.5037	1.5477	1.4847
cluster3 (1.34 ns)	1.4856	1.4191	1.4266	1.4576	1.4967	1.5420	1.4666
cluster4 (1.54 ns)	1.5080	1.4271	1.4512	1.4932	1.5103	1.5594	1.4910

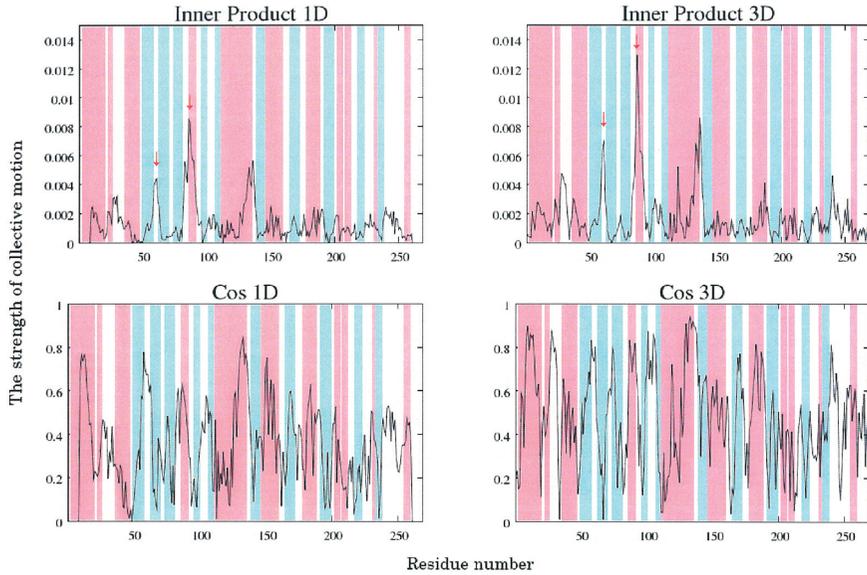


図 7. 時刻 $t = 0.6$ ns における 4 つの指標 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) と 2 次構造との関係. 横軸は残基番号 p , 縦軸は指標 $x_p^{(i)}(t)$ ($i = 1, \dots, 4$) の値である. 3 つの 2 次構造, α -helix, β -strand, loop を, 各々, ピンク, 青, 白で表している. 内積による指標では, 特徴的なピーク (赤矢印でしめした部分) が, 残基 55 付近の loop と残基 80 辺りの α -helix の一部と一致性を示している. 一方, 余弦による指標では, 主に α -helix と β -strand と一致するピークが多数見られる. 他方で, 残基 110 辺りから 130 辺りにある比較的大きな α -helix は, 一体となった集団運動を示さず, 後半部分が C 末端側にある loop に引ずられた運動を示している. Reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

て、特徴的なピークを示す部分に赤い矢印で印をつけている。残基 60 辺りの loop と残基 90 辺りの α -helix の一部分が集団運動のピークと一致していることが見てわかる。一方、余弦を用いた結果においては、 α -helix と β -strand に対応する様々なピークが見られる。他方で残基 110 辺りから 130 辺りにある比較的大きな α -helix は、一体となった集団運動を示さず、C 末端側の loop に引っ張られるように中間部分でねじれたような動きをしていることがこれらの図からわかる。

上述の結果は、一般的な傾向として、 $C\alpha$ 原子の集団運動と 2 次構造間との間に関連性があること、しかし 2 次構造によっては、必ずしもまとまった集団運動を示さない場合も有り得ることを示している。このように、提案指標を用いる事で、2 次構造単位の原子のまとまりとそれらが方向性をもって動いているか否かを、定量的に評価出来ると考えられる。この関連性の定量的な評価のため、2 次構造 k の集団運動を表す新しい指標 $R_k^{(i)}(t)$ を下記のように定義する。

$$(4.3) \quad R_k^{(i)}(t) = \sum_{p=1}^N v_{k,p} x_p^{(i)}(t)$$

2 次構造 k の構造ベクトル $v_k = \{v_{k,p}\}$ は、 $C\alpha$ 原子 p が 2 次構造 k に属する場合は $v_{k,p} = 1/N_k$ 、それ以外は $v_{k,p} = 0$ となるよう定義する。ここで、 N_k は 2 次構造 k に含まれる $C\alpha$ 原子の数 (=残基数)である。図 7 から、余弦を用いた指標が loop 構造のみならず α -helix や β -strand の集団運動を捉えていることから、以降の解析では、 $R_k^{(i=4)}(t)$ を用いることとする。

4.5.1 2 次構造間の集団運動の相関

異なる 2 次構造 k, k' 間での $R_k^{(i=4)}(t)$ の相関と $C\alpha$ 原子の時間平均距離を比較する。比較結果を図 8 に示す。図 8 の左上は、 $R_k^{(i=4)}(t)$ と $R_{k'}^{(i=4)}(t)$ の相関値を示しており、赤、青、白色は、各々、正の相関、負の相関、無相関を表している。図 8 の右下は、 $C\alpha$ 原子 p と q 間の距離 $\langle |r_p(t) - r_q(t)| \rangle$ の時間平均であり、明るい色ほど 2 原子間距離 (p, q) が近いことを示している。横軸は $C\alpha$ 原子の残基番号と、3 つの 2 次構造 α -helix, β -strand, loop をピンク、青、白色で表している。図 8 から、空間上近い位置に存在する 2 次構造間で集団性に正の相関が見られ、一方、構造の反対位置にある 2 次構造間では集団性に負の相関が確認できる。

ここで集団性における正の相関とは、片方が集団的に運動している時に、他方も集団的に運動しているという相関であって、運動方向の相関を必ずしも意味しないことを注意しておく。その点で集団性の相関は、PCA 等の方法で得られる、分布の空間的な広がり方向の相関と等価ではなく、集団性を示す指標の導入と、その指標の値の時間相関を見ることで初めて特徴付けられるものであることを強調しておく。

4.5.2 重要部位における集団運動

ここで、TLL の機能において重要な部分に着目してみることにする。既に記述のように、TLL の機能発現において“lid”が重要な役割を果たしている。この“lid”と活性部位を図 9(A)に示す。“lid”は図 2 に示す 2 次構造 v15 に相当し、v15 はその形を open-closed の構造変化に伴い、図 9(B-I)に示すように変化させる。“lid”に加え、2 次構造 v06 も open-closed に伴いその形を、図 9(B-II)に示すよう、loop から α -helix と変化させる。図 9(B-I, II)共に、open 構造と closed 構造を緑とピンクで示している。図 9(C-I, II)は、各々 v15 と v06 に関する集団運動の相関係数の値を示している。横軸は 2 次構造の番号 k 、縦軸は相関値である。相関値は有意水準 5% で判断し、有意と出たものに関して、正の相関・負の相関・無相関をそれぞれ赤・青・灰色で印をつけている。ここで特筆すべきは、構造 v15 (“lid”)と v06 が正の相関を互いに持っている事である。これは、これら 2 つの構造変化に関する集団運動が存在していることを示していると考える。また、v15 と v06 共に α/β -hydrolase 構造に含まれる 8 つの β -strand と無相関であるこ

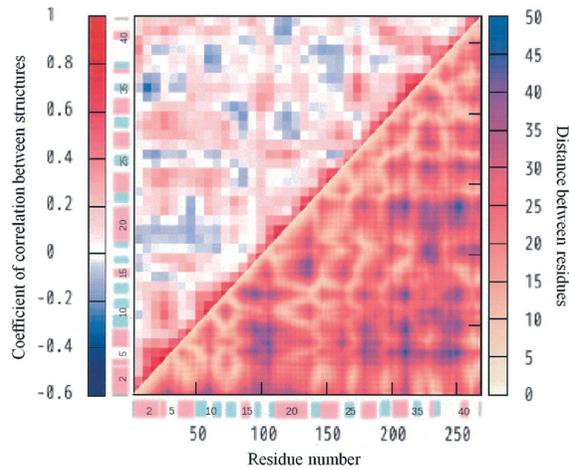


図 8. 2 次構造間の相関 $R_k^{(i=4)}(t)$ と、 $C\alpha$ 原子の時間平均距離の比較. 左上の図は、2 次構造 k と k' 間での相関 $R_k^{(i=4)}(t)$ を表しており、正の相関、無相関、負の相関を各々、赤、白、青色で示している. 右下の図は、 p 番目と q 番目の $C\alpha$ 原子間の距離の時間平均 $\langle |\mathbf{r}_p(t) - \mathbf{r}_q(t)| \rangle$ を示しており、 (p, q) の色が明るいほど、2 つの原子が近くに位置していることを表している. 縦軸・横軸ともに、3 つの 2 次構造を表すピンク (α -helix)、青 (β -strand)、白 (loop) と $C\alpha$ 原子の番号を表している. 各 2 次構造の番号は、図 2 に示されているものと同一である. Reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

とも、ここで記しておく. 更に興味深い事に、構造 v15 は活性部位を含む構造、v23, v30, v40 と、とても弱い正の相関を持っており、v06 はそれらとは無相関を示している. これらの結果から、v15 と v06 に関する集団運動は、活性部位のような機能的に重要な部位にあまり影響しないようになっていいると考えられる. これは、活性部位の様な重要部位が、構造の動きによって変化してしまうと酵素としての機能が失われてしまうと考えことから、妥当であろう.

本研究で見出された、大きく構造変化する部位と機能部位の間の「無相関」に注目すると、今回の結果の検証として次のような研究が考えられる. 即ち、“lid” の部位のアミノ残基を置換、あるいは “lid” そのものを削除した変異体を作成し、実験的に機能変化を測定するとともに、それらの変異体に対する MD によって、大きく構造変化する部位と機能部位の間の相関の程度を計算する. このようにして、タンパク質の全体としての構造変化が活性部位の配置を変える「相関」が、酵素としての機能低下や喪失の原因と成りうることを示すといったシナリオである.

今回の解析では、機能に関して重要な部位とその他に位置する原子や 2 次構造との時系列における相関関係を見た. 今後はこの様な解析で、アロステリーのように機能発現を誘因する“トリガー”となる原子や部位を特定したいと考えている. アロステリーでは今回の結果とは逆に、3 次元的に離れた部位の間の集団性の相関、特にタンパク質の全体としての構造変化が機能発現に結びつく機構が興味の焦点となるであろう. また、今回はウェーブレット解析の強みである時間情報を引き出すような指標導入にまで解析が及ばなかったが、時間相関を含めた解析への発展や、周波数の時間変化の特徴を表す右特異ベクトルも用いた指標の提案を行いたい. それによってシグナル伝達など、生体分子内の情報伝達のダイナミックな機構の解明を目指していきたい. このような研究によって、PCA や NMA 等の方法論を越えて、生体分子の動的な描像を獲得することが重要であろう.

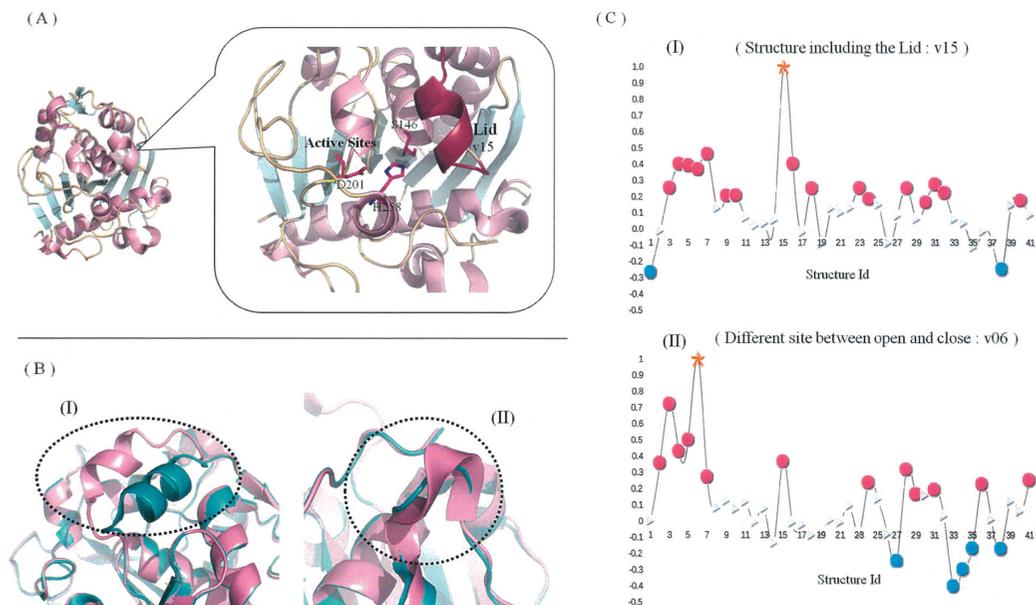


図9. TLLの重要部位。(A) 1TIBにおける“lid”と活性部位。“lid”は濃赤色で示されており、図2の構造v15と一致している。活性部位を構成する残基Ser146, Asp201, His258は、各々、2次構造v23, v30, v40に含まれている。(B) open-closedにより変化する部位、(I)と(II)について、open構造とclosed構造を各々、緑とピンクで表している。変化する部分の1つは“lid”である(I)。もう1つは、残基36から40に相当する部分であり、構造v06に含まれている。後者は、open構造においてはloopであるがclosed構造では α -helixを構成している(II)。各図はPymolによって描画している。(C)重要部位に関する相関係数。横軸は構造番号 k 、縦軸は相関係数である。赤い丸と青い丸は各々正の相関、負の相関を示しており、グレーの点は無相関を表している。(I) - “lid”を含む構造v15と他の2次構造との相関、(II) - 構造v06と他の2次構造との相関。Reprinted with permission from Kamada, M. et al. (2011). Chem. Phys. Lett., 502, p.241, Copyright 2011, Elsevier.

5. まとめ

本稿では、ウェーブレット変換を用いた時系列解析の新しい手法を紹介し、その一例として、リパーゼ(TLL)を用いた解析について紹介した。この解析では、タンパク質分子における原子の集団運動を特徴付けるような指標を導入し、以下に示す2つの結果を得た。まず、タンパク質の振動運動が、1つの構造に対応するある特定のポテンシャル井戸に限られたものではなく、複数の構造をとっていることを、集団運動の時間発展から示した。これは、タンパク質がその時間発展において複数の構造を渡り歩いている事を示唆するものである。次に、集団運動の2次構造間での相関解析により、複数の2次構造に関与した集団運動が存在することを確認した。更に、構造変化に関与するような集団運動が、機能的に重要な部位には影響を与えていないと考えられることを示した。これらの結果から、今回用いたウェーブレットによる解析手法が、集団運動に関する解析に十分に適用可能であることが確認出来た。

今回の解析では、ウェーブレットと特異値分解によって、時間空間方向の次元縮約を行う新た

な手法を提案し、さらにこの手法によって可能となる集団運動の指標の導入を行った。ここでは集団運動の指標に第1特異ベクトルのみを用いたが、本文中の解析で示した様に、今後はさらに複数の特異値を考慮する必要があると考えられ、更なる発展が望まれる。また、“intrinsically disordered proteins” (IDP)が今後の重要な対象タンパクであると考えている。IDPは激しく大きな構造揺らぎを示し、その構造変化は各々の機能を発現する上で、重要な役割を担っていると考えられている。これらのことから、非調和運動を含んだ構造変化を示す分子に対して適用可能な新しい手法開発が望まれている。我々が今回用いた手法は、非調和運動に適用可能であることから、IDPのダイナミカルな挙動を理解するのに重要なツールとなる事を期待している。

謝 辞

本研究で解析した時系列データを作成していただいた関嶋政和氏(東工大)、共同研究者の城和貴氏(奈良女)・高田雅美氏(奈良女)・木村沙知氏(当時奈良女院生)、ウェーブレットとSVDを組み合わせた時系列解析の研究に関して櫻井乃梨子氏(当時奈良女院生)・木寺詔紀氏(横浜市大)・淵上壮太郎氏(横浜市大)に感謝します。本研究は、科研費特定領域研究「実在系の分子理論」、科研費基盤研究C、科研費挑戦的萌芽研究、奈良女子大学プロジェクト経費から援助を受けました。

参 考 文 献

- Arora, K. and Brooks, C. L. III (2007). Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism, *Proceedings of the National Academy of Sciences*, **104**, 18496–18501.
- Askar, A., Cetin, A. E. and Rabitz, H. (1996). Wavelet transform for the analysis of molecular dynamics, *The Journal of Physical Chemistry*, **100**, 19165–19173.
- Bahar, I., Chennubhotla, C. and Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation, *Current Opinion in Structural Biology*, **17**, 633–640.
- Balsera, M. A., Wriggers, W., Oono, Y. and Schulten, K. (1996). Principal component analysis and long time protein dynamics, *The Journal of Physical Chemistry*, **100**, 2567–2572.
- Benson, N. C. and Daggett, V. (2012a). Wavelet analysis of protein motion, *International Journal of Wavelets, Multiresolution and Information Processing*, **10**, 1250040.
- Benson, N. C. and Daggett, V. (2012b). A chemical group graph representation for efficient high-throughput analysis of atomistic protein simulations, *Journal of Bioinformatics and Computational Biology*, **10**, 1250008–1250024.
- Brinda, K. V. and Vishveshwara, S. (2005). A network representation of protein structures: Implications for protein stability, *Biophysical Journal*, **89**, 4159–4170.
- Brooks, B. R., Janezic, D. and Karplus, M. (1995). Harmonic analysis of large systems. I. Methodology, *Journal of Computational Chemistry*, **16**, 1522–1542.
- Bui, J. M. and McCammon, J. A. (2006). Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism, *Proceedings of the National Academy of Sciences*, **103**, 15451–15456.
- Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991). The energy landscapes and motions of proteins, *Science*, **254**, 1598–1603.
- Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points, *Science*, **315**, 972–976.

- Hayward, S. and Go, N. (1995). Collective variable description of native protein dynamics, *Annual Review of Physical Chemistry*, **46**, 223–250.
- Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hubner, C. G. and Kern, D. (2007). Intrinsic motions along an enzymatic reaction trajectory, *Nature*, **450**, 838–844.
- Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins, *Nature*, **450**, 964–972.
- James, L. C., Roversi, P. and Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity, *Science*, **299**, 1362–1367.
- Kamada, M., Toda, M., Sekijima, M., Takata, M. and Joe, K. (2011). Analysis of motion features for molecular dynamics simulation of proteins, *Chemical Physics Letters*, **502**, 241–247.
- Kitao, A. and Go, N. (1999). Investigating protein dynamics in collective coordinate space, *Current Opinion in Structural Biology*, **9**, 164–169.
- Koshland, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis, *Proceedings of the National Academy of Sciences*, **44**, 98–104.
- Kumar, S., Ma, B., Tsai, C. J., Sinha, N. and Nussinov, R. (2000). Folding and binding cascades: dynamic landscapes and population shifts, *Protein Science*, **9**, 10–19.
- Levy, Y., Onuchic, J. N. and Wolynes, P. G. (2007). Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition, *Journal of the American Chemical Society*, **129**, 738–739.
- Lindorff-Larsen, K., Piana, S., Dror, R. O. and Shaw, D. E. (2011). How fast-folding proteins fold, *Science*, **334**, 517–520.
- Liò, P. (2003). Wavelets in bioinformatics and computational biology: State of art and perspectives, *Bioinformatics*, **19**, 2–9.
- Matsunaga, Y., Baba, A., Li, C. B., Straub, J. E., Toda, M., Komatsuzaki, T. and Berry, R. S. (2013). Spatio-temporal hierarchy in the dynamics of a minimalist protein model, *The Journal of Chemical Physics*, **139**, 215101.
- Okazaki, K. and Takada, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms, *Proceedings of the National Academy of Sciences*, **105**, 11182–11187.
- Sakurai, N. (2010). Time series analysis using wavelet toward molecular dynamics simulation of proteins, Masters Thesis, Graduate School of Humanities and Sciences, Nara Womens University, Japan.
- Shu, Z., Duan, M., Yang, J., Xu, L. and Yan, Y. (2009). Aspergillus niger lipase: Heterologous expression in Pichia pastoris, molecular modeling prediction and the importance of the hinge domains at both sides of the lid domain to interfacial activation, *Biotechnology Progress*, **25**, 409–416.
- Standley, D. M., Toh, H. and Nakamura, H. (2007). ASH structure alignment package: Sensitivity and selectivity in domain classification, *BMC Bioinformatics*, **8**, 116.
- Swint-Kruse, L. (2004). Using networks to identify fine structural differences between functionally distinct protein states, *Biochemistry*, **43**, 10886–10895.
- Vela-Arevalo, L. V. and Wiggins, S. (2001). Time-frequency analysis of classical trajectories of polyatomic molecules, *International Journal of Bifurcation and Chaos*, **11**, 1359–1380.
- Williamson, J. R. (2000). Induced fit in RNA-protein recognition, *Nature Structural & Molecular Biology*, **10**, 834–837.
- Wriggers, W., Stafford, K. A., Shan, Y., Piana, S., Maragakis, P., Lindorff-Larsen, K., Miller, P. J., Gullingsrud, J., Rendleman, C. A., Eastwood, M. P., Dror, R. O. and Shaw, D. E. (2009). Automated event detection and activity monitoring in long molecular dynamics simulations, *Journal of Chemical Theory and Computation*, **5**, 2595–2605.

Zaccai, G. (2000). How soft is a protein? A protein dynamics force constant measured by neutron scattering, *Science*, **288**, 1604–1607.

Dynamics of Protein Structure: Analysis with Wavelet Transformation

Mayumi Kamada¹ and Mikito Toda²

¹Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University

²Department of Physics, Nara Women's University

Structural dynamics of proteins are closely related to activity of their functions. Therefore, characterizing and analyzing structural dynamics provide us with important clues to understanding the essence of these functions. However, dynamics of proteins exhibit complex temporal-spatial hierarchy, which makes it difficult to capture motion features involving multiple temporal-spatial scales. After reviewing previous approaches toward protein dynamics, this article presents a new method for extracting features of protein motion from time-series data. This novel method uses the wavelet transformation together with the singular value decomposition (SVD). The wavelet analysis enables us to characterize time varying features of the dynamics and SVD reduces the degrees of freedom of the data. We apply the method to extract structural dynamics of *Thermomyces lanuginosa* lipase (TLL) from time-series data of molecular dynamics.