# タンパク質立体構造の動きに着目した時系列解析

 鎌田
 真由美<sup>†1</sup>
 戸
 田
 幹
  $\Lambda^{†2}$  

 高
 田
 雅
 美<sup>†1</sup>
 城
 和
 貴<sup>†1</sup>

本稿では,ウェーブレット変換と特異値分解により熱揺らぎをするタンパク質の時 系列データから動きの特徴を抽出できると考え,得られる動きの特徴を用いて骨格原 子の集団運動性に関する解析を行う.まず骨格原子の集団運動性と構造間での相関関 係について確認し,機能において重要となる2次構造に着目し解析を行う.その結果, 2次構造における相関関係や動き方の違いを確認する.

Motion based Time Series Analysis for 3D Protein Structures

Мауимі Камада,  $^{\dagger 1}$  Мікіто Тода,  $^{\dagger 2}$  Ма<br/>sami Таката  $^{\dagger 1}$  and Каzикі Јо<br/>E $^{\dagger 1}$ 

To extract motion feature from the temporal sequence data of proteins with heat fluctuations, we present a feature detection method using a wavelet transform and singular value decomposition in this paper. The extracted features are analysed for collective motion of framework atoms. We investigate collectivity of atom motion and correlation between structures, and analyze secondary structures that play a critical role in the function. As the result, we find correlation and difference motion between secondary structures.

# 1. はじめに

# 生命活動において重要な役割を果たすタンパク質は,生体内において熱運動により常に揺

†1 奈良女子大学大学院 人間文化研究科

Graduate School of Humanities and Sciences, Nara Women's University †2 奈良女子大学 理学部 物理科学科

Department of Physics, Faculty of Science, Nara Women's University

らいだ状態で存在している.このタンパク質のダイナミクスが機能発現などに深く関与して いると考えられている.ゆえにダイナミクスを解明し機能との相関を明らかにすることは, 構造変化に起因する疾患の解明などにもつながり,機能制御の本質を理解するためにとても 重要である.

多自由度・多成分系であるタンパク質は,エネルギー的にほぼ等価な多数のエネルギー極 小を持つ.十分に低温な状況では,ある1つのエネルギー極小近傍での細かい調和的に揺ら いでいるが,ある程度温度が上がるとエネルギー極小間の遷移といった非調和的な振動が起 こり得る.つまりタンパク質の揺らぎは調和的な動きと非調和的な動きを含んでいる<sup>2)</sup>.

局所的なエネルギー極小における振動は基準振動解析により表現が可能であるが,これは 振幅の小さな調和的振動を仮定している.つまり非調和的な極小間のジャンプを表現するに は別に手法を組み合わせる必要があり,調和的な振動と非調和的な振動が混在するタンパク 質の時々刻々の変化を捉えた解析は困難である.

そこで本研究では,ウェーブレット変換を用いて非調和的な動きを含むタンパク質の熱揺 らぎについての時系列解析を試み,機能との相関における新たな知見獲得を目指す.タンパ ク質の熱揺らぎは不規則な運動であり,時々刻々とその周波数を変化させていると考えられ る.このような経時的に不規則変動を示すタンパク質の揺らぎ解析には,ウェーブレット解 析を用いてその動きの特徴を抽出することが可能であると考える.ウェーブレット変換を用 いた解析の結果,タンパク質の熱揺らぎにおいて周波数が経時的に変化していることが確認 されている<sup>3)</sup>.本稿では,ウェーブレット変換によって得られる動きの特徴ベクトルを用い て原子の集団運動性と2次構造での動きに着目した解析を試みる.

以下,2章において特徴抽出に用いるウェーブレット変換について簡単に説明する.そして3章において,揺らぐタンパク質時系列データからウェーブレット変換と特異値分解を用いた特徴抽出の手法について説明する.4章にて,骨格原子である*C*α原子の集団運動性と2次構造との関係性について解析を行い,5章において構造ベクトルを導入し2次構造間の相関について解析する.

2. ウェーブレット変換

周波数解析の手法の1つにウェーブレット解析がある.これは時間と周波数にかかわる信 号情報を同時に抽出することができる解析手法である.本節では2.1節でウェーブレット変 換の特徴について,2.2節において実データに適応する際の注意点を述べる.

### 情報処理学会研究報告 IPSJ SIG Technical Report



### 2.1 概 念

時刻 t によって変化する信号 f(t) があるとする.その信号に対し周波数解析としてフー リエ変換を適応するとスペクトルに関する情報が得られるが,得られるスペクトルは周波数 についての関数であり,時間の情報は失われてしまう.そこで考えだされたのが窓フーリエ 変換である.経時的に対象を解析するために信号の一部を"窓関数"を用いて切り出し,こ の窓をずらしながら信号のスペクトルを次々に解析する.しかし一般に周波数の分解能は窓 関数のサイズに反比例する.つまり決まった大きさの窓関数を用いる窓フーリエ変換では, 様々な周波数に対する解像度は一定にならない.

ー方,ウェーブレット変換は,マザーウェーブレット $\psi(t)$ という基底関数を拡大縮小・ 平行移動させるウェーブレット関数を用いることで,各々の周波数に合わせた時間幅を適応 して解析を行うことが可能である.これにより様々な周波数に対する解像度が一定となる. 図1に窓フーリエ変換との違いを示す.ウェーブレット関数は伸張と位置を決める2つの 実数パラメータa,bを用いて

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

のように定義される<sup>1)</sup>. 任意の信号 f(t) に対する連続ウェーブレット変換 W(a,b) は,信号とウェーブレット関数の内積により以下のように定義される.

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

W(a,b)は,t = bにおいて信号 f(t)の中に $\psi_{a,b}(t)$ の成分がどれだけ含まれるかを表す.

ウェーブレット変換には,連続ウェーブレット変換と離散ウェーブレット変換の2種類が存在し,マザーウェーブレットの選び方で無数に分けることができる.本研究ではマザーウェーブレットに Morlet ウェーブレットを用いる. Morlet ウェーブレットの窓関数は以下

のように定義される.

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

σ は窓の中に入っている波の数を表している. 基準振動解析 では小さな振幅で調和的な動き を仮定しているため,準安定構造間における状態遷移のような非調和的な動きを解析するこ とができない.ウェーブレット変換は上記に述べたように,周波数に合わせた解析が可能で ある.よって,振幅にあわせて周波数が変化するような非調和的な動きの解析も,ウェーブ レット変換では対象とすることができる.

2.2 解析に用いる際の注意

前項で既述したウェーブレットの定義式(2.1)では, t の範囲が -∞ から ∞ であること を前提としている.しかし実際に本稿での解析に用いる実データは,有限の幅しか持つこと はできない.ゆえに境界付近のデータは信頼性が低くいため,解析を行う際には時系列の 端部分を除くことが必要である.また,一般に有限の幅のデータ解析において低周波数は 境界の影響を特に受け,高周波数においては1つの波の中にデータが少なく離散的になり, データの信頼性は低くなる.そのため,低周波数と高周波数を除いた部分を用いて解析を行 うことが必要である.

### 動きの特徴を抽出

本稿では,残基数 269 の "1tib"を対象タンパク質として用いる.この 1tib に対して,分 子動力学シミュレーションを 2n 秒実行し,200 タイムステップの時系列を得る.1 ステッ プの間隔は 10p 秒である.この時系列データには,重心を合わせて回転運動を取り除いた  $C\alpha$ の空間座標が記述されている.

MD シミュレーションの結果得る各自由度の時系列データに対し,2節で詳述した連続 ウェーブレット変換を施す.式(2)における $\sigma$ は5とする.ウェーブレット変換により, 各自由度が持つ周波数成分の経時的な変化を得る.本稿の対象タンパク質1tibは,269個の  $C\alpha$ からなる.よって系は,807の自由度を持つため,ウェーブレット変換により807個の 出力ファイルが得られる.各ファイルには,タイムステップと波数,その波数における周波 数成分が記述される.また,2.2項で述べた注意事項を考慮し本稿では,時系列の端部分を 除いた時刻 $t:20 \sim 180$ を用いる.本稿で用いるMolretウェーブレットは複素ウェーブレッ トであるため,199から101までの周波数成分は各々1から99までの周波数成分の複素共役 になっている.ゆえに周波数は,低周波数と高周波数を除いた部分, $\omega:15 \sim 50,150 \sim 185$ のデータを用いて解析を行う.

#### 情報処理学会研究報告

IPSJ SIG Technical Report

ウェーブレット変換により得られた時系列データから, *C*<sup>α</sup>の大きくゆったりとした動きを取り出すために,非正方行列から主要な特徴を取り出す手法としてよく知られている特異値分解を用いる.特異値分解は以下の式で定義される.

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^{\dagger} \tag{3}$$

与える行列 A は,あるタイムステップにおける  $C\alpha$  原子の自由度を行成分に,周波数を列 成分として作成する.波数のデータは,既述のように  $15 \sim 50 \ge 150 \sim 185$ を用い,それ 以外の要素には 0 を代入する.列成分は 199 個となる.また, $C\alpha$  全体からなる系の自由度 は 807 である.ゆえに行列 A は 807 行 199 列の行列となる.この分解によって得られる左 特異ベクトルであるユニタリ行列 U は, $C\alpha$  原子の動きを表す.また,右特異ベクトルで あるユニタリ行列 V は周波数の動きを表す.

上記手法により得る特徴に対して行った妥当性の検証によって,多自由度で表現される動きが少数の特徴ベクトルを用いることで再現できることが確認されている<sup>5)</sup>.また,タンパク質の揺らぎにおける大まかな動きの解析に対して第1主成分を用いることが有用であることが確認されている.よって本稿では,以降第1成分に対応するユニタリ行列 U の第1 列ベクトルを用いて解析を行う.

### 4. 特徴を用いた集団運動性についての解析

タンパク質において機能発現と2次構造は密接に関与していると考えられる.そこで本 稿では,骨格原子の集団運動性を2次構造と関連付けて解析を行う.まず4.1節にて集団 運動性を評価する指標の定義を行い,4.2節にて2次構造との対応関係を確認する.本節以 降,単に「原子」と表現する場合,それは*C*α原子を指すこととする.

### 4.1 集団運動性を示す指標の定義

タンパク質を構成する原子 p を中心とした原子集団の動きの強さを  $x_p$  とする.まず, 複数の原子から成る集団の定義の仕方を考える.図 2 がその概念図であり,赤色の丸が原子 p を示す.1 つは,図 2-(a) に示すように原子 p から両方向に原子 p を含めて各々n 原子を 原子集団の範囲 (2n-1) としてとる.これは骨格上における隣接原子間でのまとまりを考 える1 次元での定義である. $\alpha$  ヘリックスやループは骨格上隣接した原子がつながって構 成されているので,この1 次元での指標を用いることでその動きが捉えられる.しかし, $\beta$  シートは3 次元空間において隣接した  $\beta$  ストランドにより構成されるため,ストランド間 に多数の原子を挟むと1 次元の指標ではうまく  $\beta$  シートとしての集団運動性が捉えられな いと考える.そこで,2 つ目として図 2-(b) に示すような3 次元空間における集団の定義



考える.これは3次元空間で原子 p を中心とした半径 r の球を考え,この球の中に含まれる原子を原子 p を中心とした原子集団の構成員とする定義である.これを用いることで図に示すようなストランド間に大きなループを含んだとしても, β シートとしての集団運動性を捉えることができると考えられる.

動きの類似性の定義については,原子 p と原子 q の内積を計算する場合と余弦を計算する場合の 2 つを考える.余弦では原子間の向きが揃っているかどうかだけが値に反映されており,ベクトルの長さ,つまりその動きの強さは加味されない.一方,内積を用いる場合は原子が揃った方向に振動しているかどうかの評価に動きの強さも加わる.タンパク質の 2 次構造において,ループは構造の外側にあることから比較的激しい動きを持つとされ, $\alpha$  へリックスや  $\beta$  シートは比較的構造の内側に組み込まれる傾向が強いことから,動きの大きさとしては強く現れないと考えられる.そのため,動きの大きさが加味される内積を用いた場合ループの動きを強く捉え,一方余弦では,まとまって動く  $\alpha$  へリックスや  $\beta$  シートの動きを捉えることができると考える.

以上のことから,本稿では表1に示す4つの指標 $x_p^1(t) \sim x_p^4(t)$ を定義する.

4.2 2次構造との対応関係

まず表1の指標を用いて集団運動性と2次構造との関係性について調査する. 立体構造 に関する情報は、PDBsumから取得する<sup>7)</sup>.図3に PDBsumから取得した1tibの2次構 造を示す.本稿では以降、各構造や構造間に対する解析を行うため、各々の2次構造に対し て図中に示すような独自の名称を付ける.N末端から順番に $\alpha$ へリックス・ $\beta$ シート・ルー プの3つの2次構造をひとまとめにして v01~v41とする.

まず,2次構造と時刻 60 における集団運動の強さをプロットしたものを図 4 に例示する. それぞれのグラフは,横軸が残基番号 resをとり縦軸にその残基を中心とした集団の動きの大きさ $x_{res}(60)$ をとる.また背後のピンクの帯は $\alpha$ ヘリックスを,青の帯は $\beta$ シートを

3

#### Vol.2010-BIO-20 No.5 2010/3/4

#### 情報処理学会研究報告

IPSJ SIG Technical Report

	Inner Product	Cos	
1 次元	$x_i^1(t) \equiv \left  \sum_{ i-j  < n} v_i(t) \cdot v_j(t) \right $	$x_i^3(t) \equiv \left  \sum_{ i-j  < n} \frac{v_i(t) \cdot v_j(t)}{ v_i(t)   v_j(t) } \right $	
3 次元	$x_i^2(t) \equiv \left  \sum_{ r_i - r_j  < r} v_i(t) \cdot v_j(t) \right $	$x_{i}^{4}(t) \equiv \left  \sum_{ r_{i} - r_{j}  < r} \frac{v_{i}(t) \cdot v_{j}(t)}{ v_{i}(t)   v_{j}(t) } \right $	

#### 表1 集団運動性の指標

それぞれ表している.時刻 60 において,内積のグラフより矢印で示した残基 86~91 と残 基 111~135 のヘリックスが大きく動いていることが確認できる.内積のグラフではその計 算式からも明らかであるが,大きく動く部分とそうでない部分の違いが顕著に表される,一方,余弦のグラフでは内積のグラフに比べ動きのある部分は全体的に同じような大きさで現 れている.1 次元と3 次元とでは,3 次元のほうが各 2 次構造の集団での動きを良く捉えら れていると確認できる.しかし全体的には,図3 に示される  $\beta$  シートの動きがあまり捉え られていない.

次に集団運動の経時的な変化を確認するために,各々の指標を全ての時刻 ( $t: 20 \sim 180$ ) で計算し,3次元プロットしたものを図5に示す.横軸はタイムステップ,縦軸は残基番号 を表し,動きの強いものほど明るくプロットされる.内積のグラフから時刻45付近で大き な特徴的な動きがあることが確認できる.余弦のグラフでは全体的に動きがあることを示 しており,動きの経時的な変化を目視により確認することができる.これらの考察から,上 記で定義した4つの指標はそれぞれ解析目的に応じて使い分けられることがわかる.2次構 造の経時的な動きを得る場合は余弦3次元の指標を,構造の動きの強さと特徴的な動きの 有無を得る場合は内積の指標を用いることが適切であると考える.1次元と3次元ではあま り大きな違いはない.今回ターゲットとしている1tibの持つ2つの $\beta$ シートは,1つは2 つのストランドから成る小さいものであるが,一方は8つのストランドから成る大きな $\beta$ シートであり,空間的にも広がりを持っている.これを捉えるには,集団を定義する球の範



囲をとても大きくしなければならず,これでは精度の低下などの理由から構造解析には適さ ない.よって β シートとしての集団運動性の解析には,また別に新たな指標を定義する必 要性が考えられる.

5. 構造ベクトルによる定量化と相関関係の解析

本節では前節での2次構造と集団運動との関係性を定量的に評価するため,5.1項で構造 ベクトルを定義する.5.2項において定義した構造ベクトルを用いて2次構造間の相関関 係を確認し,5.3項にて対象タンパク質において重要だと考えられる部位に着目した解析を 行う.

5.1 構造ベクトルの定義

4.2 項での指標と 2 次構造の動きの対応関係に対する評価は目視によるものである.そこ で本節以降,定量的に解析・評価を行うために各々の 2 次構造 k に対する構造ベクトル  $v_k$ を定義する.構造ベクトルは構造を構成する骨格原子数 n で規格化する.n 個の原子で構成される 2 次構造 k に対する構造ベクトル  $v_k$  は,残基番号 res の原子がその構成要素であ

#### 情報処理学会研究報告

IPSJ SIG Technical Report



る場合,  $v_k(res)$ 値は 1/nをとり, そうでない場合は 0をとる. 各々の構造に対して定義した構造ベクトル  $v_k$ を用いて,時刻 tにおける 2次構造 kの動きの強さ  $R_k(t)$ を以下のように算出する.

$$R_k(t) = \sum_{r=r=1}^{269} v_k(res) x_{res}(t)$$
(4)

本研究では,まず時系列における2次構造の動きに対する解析として,3次元での集団に対し原子間の余弦を計算する「 $x_p^4(t)$ 」に着目する.

### 5.2 2次構造間の相関関係

経時的な 2 次構造の動きにおける構造間での相関の有無を検証するために,  $R_k(t)$  を用い て構造間での相関係数を計算する.相関係数の算出には, Octave<sup>4)</sup> を用いる.全ての構造 間で総当り的に相関係数を計算した結果を図 6 に示す.横軸と縦軸は各構造を示す.全体と して強い相関はあまり確認できず, ほぼ無相関に近いものが多数を占めている.正の相関は 骨格上 1 次元的に隣接している構造間で確認でき,その他で正の相関を示すものは,ほとん どが空間的に隣接している構造間である.また,負の相関は立体構造上両極に位置する構造





間で確認できる.

5.3 重要部位に関する解析

前項では, すべての2次構造についての相関関係を大まかに確認している.本稿での対象 タンパク質 1tib と同一アミノ酸配列を持つタンパク質が存在する.そこで,以下5.3.1目 においてそれらタンパク質との関係から重要であると考えられる部位について述べ,5.3.2 目と5.3.3目において重要部位に着目した解析および考察を行う.

5.3.1 対象タンパク質における重要部位

本研究で対象とするタンパク質 1tib と同一アミノ酸配列を持つタンパク質に 1GT6A, 1GT6B がある.これは水素結合によりホモダイマーを形成し,結合する部位の違いによっ てA構造とB構造とに分けられる.図7に1tibと1GT6の構造とその関係を示す.立体構 造情報からこれらA構造とB構造間に2次構造上の違いは確認できない.ゆえに,1GT6A と1GT6B 間に構造ではなく動きとしての違いを抽出することができれば,ホモダイマー

### 情報処理学会研究報告 IPSJ SIG Technical Report



形成に関する解析において非常に有用であると考える.よって本稿ではこれ以降,1tibにおける 1GT6A と 1GT6B の結合部位に対応する部分,またこれらの構造と1tib との相違部分に対応する 2 次構造を重要部位として着目し更に詳細な相関関係の考察を行う.図8に,1tib と 1GT6A,B の相違部分と 1GT6A・1GT6B における結合部位を示す.図中の構造名は図3 で示したものである.1tib と 1GT6A,B との相違部位は v06 と v15,1tib における 1GT6A の結合部位(結合部位A)は v18 と v23,1GT6B の結合部位(結合部位B)は v36 と v02 に各々対応している.

### 5.3.2 重要部位に着目した構造間の相関関係

まず 6 つ各々の構造部位と他の 2 次構造との相関を確認する.図 9 に各重要部位とその他の構造との相関係数のグラフを示す.横軸は v01 から v41 の構造,縦軸は相関係数であ





る.図中の赤色の点線よりも上部にあるものは5%の有意水準で正の相関を,青色の点線よりも下部にあるものは負の相関を持つ.

得られる相関係数から,図10に重要部位と相関を持つ構造を示す.各重要部位と正の相 関を持つ構造は赤色,負の相関を持つ構造は青色で表す.結合部位Bは構造上対極に位置 する部位といくつか負の相関を持つが,その相関と構造に秩序だったものはない.また,結 合部位Aも構造的に近い部位と正の相関が確認でき,その多くはシートやループ構造を持 つ.相違部位に関しては,近くのヘリックス部位と正の相関が得られ,v15に関しては構造 的に離れた部分に位置するいくつかの構造と正の相関を持っている.

次に重要部位6つの構造間における相関を確認すると,結合部位Aは相違部位v15とは

### 情報処理学会研究報告 IPSJ SIG Technical Report



図 10 重要部位と相関を持つ構造

正の相関を持ち,一方,結合部位 B は相違部位 v06 と正の相関を持つ.各結合部位が各々 異なる相違部位と相関を持っていることは注目すべき結果であり,このことから,相違部位 v06 は B への構造変異に関係しており,相違部位 v15 は A への構造変化に関係しているの ではないかと考えられる.また,結合部位間の相関関係に有意性は確認できないが,更に データの個数を増やして上記の結果を検証することは有用であると思われる.

5.3.3 クラスタ解析

上記の解析において結合部位間においては有意であると判定されなかったものの,重要部 位間に興味深い相関関係が得られている.そこで,重要部位の動きが経時的にどのように 変化し関わり合っているのかを確認するため,重要部位に対応する6つの2次構造のみを 用いた時刻間でのクラスタ分析を行う.時刻 t<sub>i</sub>と時刻 t<sub>j</sub>間における類似度は,以下の式に よって計算する.

$$s(t_i, t_j) = -\sqrt{\sum_{vec=1}^{6} \left| I_{t_i}(vec) - I_{t_j}(vec) \right|^2}$$
(5)



図 11 重要部位のみの類似度と動きの変化との対応関係

 $I_t(vec)[vec: 1 ~ 6]$ は、時刻 t における  $R_k(t)[k = 2, 6, 15, 18, 23, 36]$ に対応している、類 似度を図 11 の右側に示す。図から時系列においていくつかのクラスタ性を持つことが確認 できる.ここで時刻 110~140 に着目する。図 11 の左側に、時刻 110~140 における 6 つの 重要部位に相当する 2 次構造の動きと類似度の対応関係を示す.赤の線が結合部位 A, 緑 の線が結合部位 B, 青の線が相違部位の動きの大きさの変化を表している.この 3 つの部 位において,主にまとまって動いている部位が入れ替わる箇所で区切り,黒の点線で囲んで いる.この区切った部分と類似度の図で確認できるまとまりが対応していることがわかる.つまり,他の構造よりも優位にまとまって大きく動く重要部位が,経時的に遷移していると いうことが確認できる.

次に,この類似度を用いてクラスタリングを行う.本稿ではクラスタリング手法として Affinity Propagation (AP)<sup>6)</sup>を用いる.APは2007年にトロント大学のBrendan J. Frey と Delbert Dueck らによって提案された,メッセージ伝播をベースとするクラスタリング手 法である.APは入力データとして,類似度と自身がどれだけクラスタ中心としてふさわし いかの評価に用いるパラメータ Preference を用いる.Preference を徐々に変化させて AP クラスタリングを行い,本稿ではクラスタ数5が得られる場合に着目する.この時クラスタ 中心として,時刻58・72・101・109・124が得られる.ここでクラスタ中心となる時刻で, 結合部位 A と結合部位 B,相違部位の各々の動きの大きさを確認する.その値と,結合部 位 A と結合部位 B とで動きが大きい方に色を付け,表2に示す.表からも確認できるよう に,時刻58では結合部位 A の動きが大きく,B の動きは小さい.一方,時刻101では結

#### 情報処理学会研究報告

IPSJ SIG Technical Report

時刻	結合部位A	結合部位B	相違部位
58	1.233288	0.909616	1.383803
101	0.257595	0.345029	0.331418
109	1.211622	0.447874	0.690929
124	0.724414	1.433488	1.277808
72	0.565064	1.421617	0.64962

表 2 クラスタ中心となる時刻における各部位の動きの大きさ

合部位 B の動きが大きく, A の動きは小さい.また,時刻 109 は時刻 58 と,時刻 124 は 時刻 101 と各々同じ結合部位が動きの大きさとして優位となっているが,相違部位の動き の大きさが時刻 58・101 とは異なっている.このように動きの大きさが優位となる構造が 時刻によって変化し,それによって分類できることが確認できる.結合部位に関して前目 では有意な相関は確認できないが,上記の結果より動き自体にはその大きさに逆相関のよう なものが確認できる.つまり,ある状況下において結合部位 A が活動的に動いている分子 は 1GT6A に,逆に結合部位 B が不活動な分子は 1GT6B になりやすく,このとき結合が 起こりやすいものと考えることができる.また,結合部位が各々異なる相違部位と相関を示 すことから,その結合の起こりやすさには相違部位に対応している2次構造の動きも関係 していると考えられる.1GT6A・1GT6B によるホモダイマー形成において2次構造の動 き方が関与しているかどうかについては,実際にパスサンプリングなどによる,さらに長い 時間スケールでの MD シミュレーションを行い検証する必要性があると考える.

### 6. ま と め

本稿では,多自由度からなる非調和的な振動を含んだタンパク質のダイナミクスの解析を 目的とし,その動きの特徴を取り出す為に分子動力学シミュレーションの結果得られる時系 列データに対し,ウェーブレット変換と特異値分解を適応した.そして得られる特徴を用い て様々な解析を行った.

特異値分解の結果得られる特異ベクトルは, 左特異ベクトルが骨格原子の空間における振動方向の特徴を示し, 右特異ベクトルは揺らぐタンパク質における周波数成分の動きを示している. 左特異ベクトルを用いて, タンパク質において重要とされる原子の集団運動性を示す指標の定義を行った.指標として原子の動きの評価の算出に内積と余弦を用いた2つを定義し, 原子集団の範囲設定として, 骨格上で1次元的に集団を定義するものと3次元的に集団を定義するものを用意した. これらを各々を組み合わせた4つの指標を用いた解析と各指標の評価を行い, 原子の振動方向での集団運動性がタンパク質の2次構造と対応し

ていることを確認した.

更に2次構造としての動きを見る為に,集団運動性の指標に2次構造に対応した構造ベクトルを導入して解析を行った.まず,構造ベクトルを用いて2次構造間の相関関係についての解析を行った.結果,全体的に立体構造上に近いものに正の相関が見られ,離れているものには負の相関がみられた.そして,対象タンパク質と同一アミノ酸配列を持つタンパク質が結合によりホモダイマーを形成することから,ホモダイマーを形成する際の各々の結合部位と対象タンパク質との相違部位にあたる2次構造に着目し解析を行った.その結果,2つの結合部位が各々異なる相違部位と正の相関を持つことがわかった.

重要部位に限定してクラスタ解析を行った結果,動きの大きさに優位を見せる構造部位が 時刻よって異なることから,ホモダイマー形成には,各結合部位に相当する2次構造の動き の方の違い,また結合部位と相違部位各々の動きの大きさにおける関係性が関与しているも のと考えられる.2次構造間の相関関係と上記の仮説に対する妥当性の検証には,さらに長 い時間スケールでのホモダイマー形成を考慮した分子シミュレーションを行う必要がある.

### 謝 辞

本稿で使用した分子動力学のデータを提供して頂いた産業総合技術研究所の関嶋政和先 生(現在東京工業大学)に感謝致します.

### 参考文献

- 1) I. ドブシー.: ウェーブレット 10 講, Springer, (2003)
- 2) 北尾彰朗.:分子シミュレーションと中性子散乱による蛋白質のダイナミクス研究,日本 中性子科学会学会誌「波紋」,12(2), p80-84 (2002)
- 3) 木村紗知, 熱揺らぎを反映したタンパク質の結合予測に関する研究 (2008).
- 4) GNU Octave Home Page. http://www.gnu.org/software/octave/index.html.
- 5) Mayumi Kamada., Sachi Kimura., Mikito Toda., Masami Takata., Kazuki Joe.: Clustering the Temporal Sequences Feature of 3D Protain Structure, In Proceedings of International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'09), July 2009, (Now Printing)
- Brendan J. Frey, Delbert Dueck.: Clustering by Passing Messages Between Data Points, Science, 315(5814), 972-976, 2007.
- 7) PDBsum Home Page, http://www.ebi.ac.uk/pdbsum/