

修士論文

**熱運動に基づく単量体タンパク質内部の
集団的構造変化の予測**

平成22年7月

東京大学大学院

新領域創成科学研究科 情報生命科学専攻

86915 根本 崇正

指導教員：北尾 彰朗 准教授

目次

1. 序論	3
2. 方法	6
2.1 理論	6
2.1.1. 基準振動解析	6
2.1.2. Anisotropic Network Model	6
2.1.3. ヘシアン行列	9
2.1.4. タンパク質ベンチマークの構築	11
2.2 手順	13
2.2.1. 基準振動ベクトルの算出	13
2.2.2. 大規模かつ集団的な構造変化に寄与する基準振動ベクトルの選択	13
2.2.3. ドメイン運動との対応関係の検証	16
2.2.3. 構造変化カバー率の算出	16
2.2.4. 構造変化の予測	17
3. 結果と考察	19
3.1. ドメイン運動との比較	19
3.2. 寄与率と Collectivity の分布	20
3.3. 構造変化カバー率の推移	22
3.4. 予測構造と実際の構造の一致率	24
3.5. 今後の課題	26
4. 結論	27
5. 謝辞	28

I. 序論

タンパク質は常に一定の構造を保っているのではなく、多くの場合は立体構造の変化を起こすことが知られている。このような立体構造の変化は、タンパク質が他の分子と相互作用をし、適切な機能を果たす上で重要な役割を担っている。タンパク質の構造変化の規模や形式は様々であり、他分子と相互作用する領域のみが運動する構造変化も有れば、タンパク質全体が大規模に運動するような構造変化も有る。特に、ドメイン運動 [1-4] として知られている構造変化では、タンパク質の構成部位が集団的かつ剛体的に運動する。タンパク質に起こる構造変化が予測できれば、剛体ドッキングを用いた、結合構造の詳細な予測に用いることができる。

また、タンパク質が他分子と相互作用する際に起こる構造変化の形式として頻繁に議論されるシステムに、**Conformational Selection** と **Induced Fit** がある。前者のシステムは、タンパク質の構造は他の分子と相互作用する前から熱運動によって揺らいでおり、結合に適した立体構造をとった際に他の分子と会合することで結合構造に至るというシステムである。後者のシステムは、タンパク質と他分子の結合によってタンパク質の構造変化が引き起こされることによって結合構造に至るというシステムである。近年 **Dobbins** らの研究によって、タンパク質が他分子と結合する際、**Induced Fit** の寄与する度合いによらず、タンパク質自信に内在する熱運動が寄与していることが示唆された [5]。そのため、タンパク質自信の熱運動による構造のゆらぎに基づいた構造変化の予測は、多くのタンパク質に対して、剛体ドッキングに用いるための一時的な結合構造を予測するために有効であることが期待される。

しかし、ドメイン運動のような大規模な立体構造の変化は、マイクロ秒単位からミリ秒単位の比較的長いタイムスケールで起こることが知られており、シミュレーションが難しいとされている。

タンパク質構造変化の研究に最も頻繁に用いられる分子動力学法は、目的とする分子を構成する原子の座標を、時間経過に沿って順々に予測していく手法である。この手法は、タンパク質の詳細な構造変化を予測できるが、計算量が多いため現時点ではマイクロ秒単位の構造変化までしかシミュレーションできない。そればかりか、予測される構造が、もとの構造付近の安定構造にトラップされてしまう可能性があり、たとえ長時間のシミュレーションを行ったとしてもドメイン運動などの大規模な構造変化を予測できない場合がある。このような理由から、分子動力学法によるタンパク質の大規模な構造変化の予測は難しい。

タンパク質の構造変化を基準振動ベクトルの和として近似する基準振動解析[6]もまた、タンパク質の構造変化を予測するために頻繁に用いられる手法である。先行研究では、基準振動解析によってタンパク質と他分子とのドッキングによる構造変化をリファインする研究がすでに為されている[7]。この手法では、タンパク質に起こりうる構造変化を、タンパク質自身に内在する熱運動をもとに予測するため、時間経過に沿った計算を行う必要が無く、分子動力学法よりも計算時間が短いのが特徴である。そのため、タンパク質の大規模な構造変化に対しても、比較的短い時間で予測が可能となる。しかし、一般的に用いられる、タンパク質を構成する全原子を含めたモデルによる基準振動解析は、アミノ酸側鎖どうしの相互作用による小規模な構造変化に捕らわれてしまい、分子動力学法と同様、大規模な構造変化を予測できない場合がある。一方、タンパク質を構成する α 炭素のみで構成される、Anisotropic Network Model (ANM) [8,9]と呼ばれるタンパク質モデルを用いた基準振動解析の研究も進められている。このモデルではアミノ酸残基の側鎖を無視することにより、タンパク質の大規模かつ集団的な構造変化

の解析が可能となる。Tama らの研究により、基準振動解析で得られる基準振動ベクトル中の、振幅の大きい基準振動ベクトルが実際の構造変化と高い相関を持つことが明らかにされている[10]。

一方、これらの手法以外にもタンパク質に起こりうる構造変化を予測する研究が為されている。タンパク質の各原子に可能な座標変化をランダムに発生させ、予想される複数のタンパク質構造を出力するソフトウェア、CONCOORD[11]や tCONCOORD[12]は、現在でも広く構造変化の解析に利用されている手法である。これらの手法はタンパク質の構造データをもとにした安定構造周辺の構造のアンサンブルを得るために用いられる。CONCOORD を用いた構造変化予測は、残基レベルでの小さな揺らぎを含めた複数の構造を再現するため、Induced Fit のような複雑な構造変化に対しても構造変化を予測することができる。

しかし、これまでの先行研究は、タンパク質の構造変化が集団的であるかどうかを考慮に入れていない。また、集団的構造変化はドメイン運動のように比較的単純な構造変化が多いと期待される。そのため、集団的構造変化に予測対象を絞ることによって、大規模な構造変化をより簡単に予測できることが期待される。本研究では、タンパク質の立体構造に基づいて、集団的構造変化を予測する手法を開発することを目的とした。そのために、まず集団的構造変化をするか否かの判定基準を作り、次に変化後の構造を予測して実際の構造との一致率を検証した。

具体的な手順の概要としては、構造変化前後の構造が得られている 415 種類のタンパク質 X 線結晶構造データを用いて基準振動解析を行い、集団的構造変化に寄与している基準振動ベクトルを検出する方法を検討した。次に、予測された変化と実際に観測された変化後の構造を比較した。これによって、基準振動解析の結果からタンパク質の大規模な構造変化をどの程度の確率で予測できるかを網羅的に評価した。

2. 方法

2.1. 理論

2.1.1 基準振動解析

基準振動解析とは、タンパク質分子の熱運動によって起こる構造変化を、基準振動ベクトルの線形結合として近似する手法である。そのため、タンパク質の構造変化による各原子座標の変化と各基準振動ベクトルの関係は以下のように表せる。

$$\Delta \mathbf{r} = \sum_{j=1}^{3N-6} (\mathbf{u}_j \cdot \Delta \mathbf{r}) \mathbf{u}_j \quad (2-1)$$

N はタンパク質モデルを構成する原子数である。本研究における基準振動解析では、モデル構成原子を α 炭素のみとしたので、 N はタンパク質を構成する α 炭素の数である。 $\Delta \mathbf{r}$ はタンパク質における各原子の座標変化を表す $3N$ 次元ベクトル、 \mathbf{u}_j は j 番目の $3N$ 次元の基準振動ベクトルであり、いずれも各原子座標の変位を意味する。基準振動解析の特徴はタンパク質の構造変化を熱運動のみによって表現する点であり、他の分子との相互作用にともなう Induced Fit によってタンパク質構造が複雑に歪む場合、単量体タンパク質に対する基準振動解析では上手く構造変化が予測できない場合がある。

2.1.2 Anisotropic Network Model

本研究での基準振動解析には、 α 炭素のみを用いたタンパク質モデルの一つ、Anisotropic Network Model (ANM) を用いた。ANM はタンパク質の内部に働く電磁気力やファンデルワールス相互作用、水素結合などといった相互作用の種類の違いを無視し、カットオフ距離 r_c 以内の α 炭素原子間を連結するバネの力として近似する。ANM における各原子の運動方程式は、タンパク質全体のポテンシャルエネルギーの 2 次微分行列（ヘシアン行列）を用いて以下のように表される。

$$\frac{d^2}{dt^2} \mathbf{R} = -\mathbf{H} \cdot (\mathbf{R} - \langle \mathbf{R} \rangle) \quad (2-2)$$

\mathbf{R} は各原子の質量加重座標、 \mathbf{H} は $3N \times 3N$ 次元のヘシアン行列を示す。

また、ANM における基準振動ベクトルは、ヘシアン行列を対角化ことで得られる。

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (2-3)$$

\mathbf{U} は $3N$ 次元の基準振動ベクトル $3N$ 個からなる $3N \times 3N$ の基準振動行列、 $\mathbf{\Lambda}$ は各基準振動ベクトルの固有値を持つ $3N \times 3N$ の対角行列である。各基準振動ベクトルの固有値は、対応する基準振動ベクトルの振動数の 2 乗と等しい。また、エネルギーの等分配則により、低振動数の基準振動ベクトルは大きな振幅を持つ。

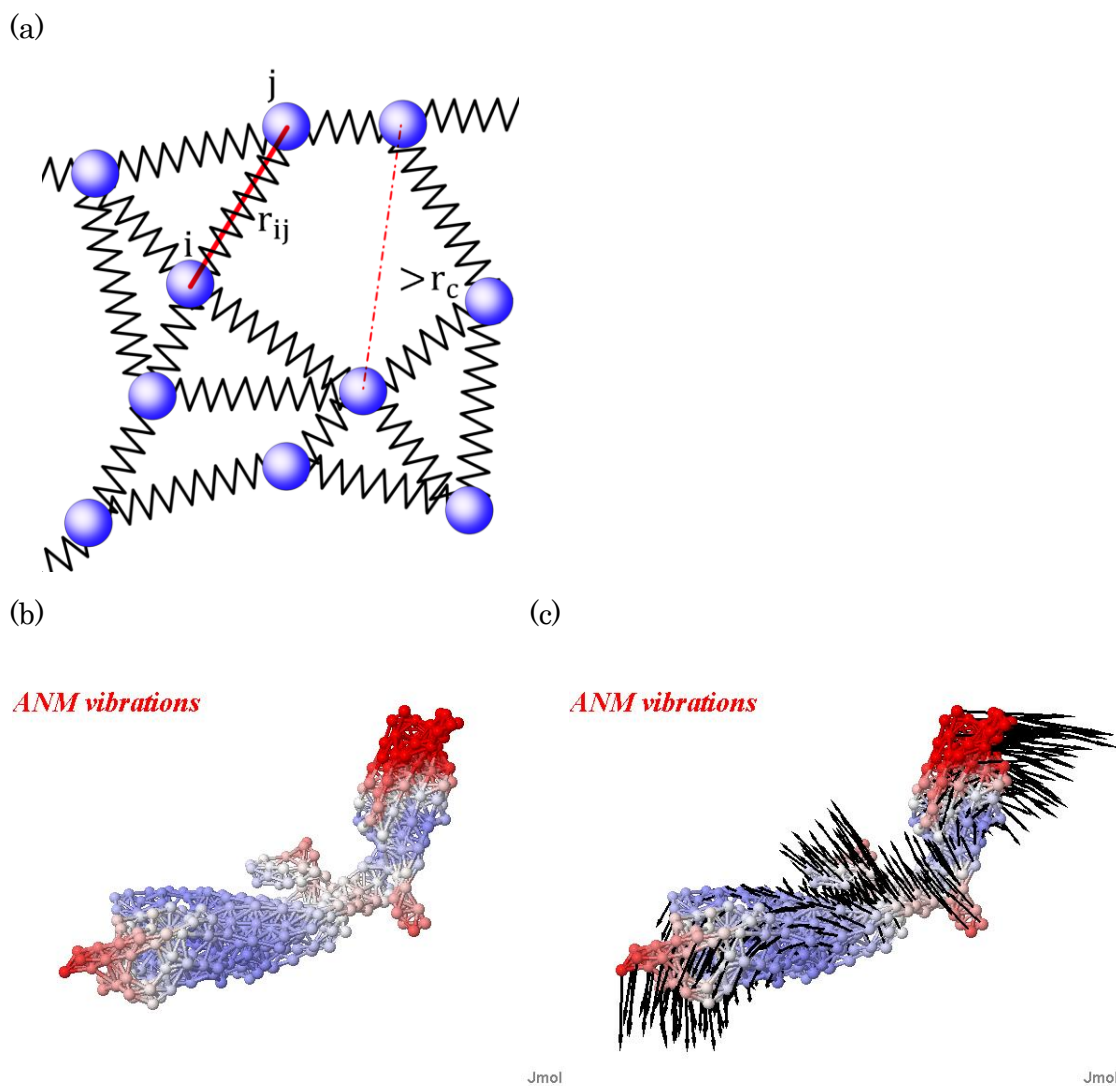


Fig 2-1 ANM を用いた基準振動解析の図

(a)は ANM の模式図を示す。青い球は α 炭素、波線は ANM によって想定されるバネを意味する。 α 炭素 ij の間の距離 r_{ij} がカットオフ距離 r_c よりも短い場合、その 2 原子間にバネを想定する。(b)はタンパク質 1S3R を ANM で表示した物である。球体の部分は α 炭素、それらをつなぐ線は ANM で想定されるバネを意味する。(c)は ANM の基準振動ベクトルのひとつを表した物であり、ベクトルひとつが各原子の移動する方向を意味する。出展 ANM Web Server

(<http://ignmtest.cccb.pitt.edu/cgi-bin/anm/anm1.cgi>)

2.1.3 ヘシアン行列

ANM におけるヘシアン行列の計算方法を示す。

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N1} & \cdots & \mathbf{H}_{NN} \end{pmatrix} \quad (2-4)$$

ヘシアン行列の部分行列 \mathbf{H}_{ij} は原子 ij 間に働く力の定数であり、異方性の情報を含む。

ヘシアン行列の部分行列のうち、非対角成分は以下のように表わされる。

$$H_{ij} = \begin{pmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{pmatrix} \quad (i \neq j) \quad (2-5)$$

X, Y, Z は原子のデカルト座標を示す。

また、対角成分は以下のようになる。

$$H_{ii} = - \sum_{i \neq j}^N H_{ij} \quad (i = j) \quad (2-6)$$

原子が互いにバネで連結されているものと近似するため、ポテンシャルエネルギー V は以下のように定義される。

$$V = \sum_{i \neq j} \frac{1}{2} c (r_{ij} - r_{ij}^0)^2 \\ = \frac{c}{2} \sum_{i \neq j} \left\{ \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2} - r_{ij}^0 \right\}^2 \quad (2-7)$$

c はバネ定数、 r_{ij} は原子 i と原子 j の間の距離、 r_{ij}^0 は原子 i と原子 j の間の平均距離である。

この式を X_i で微分することで、以下の式を得る。

$$\frac{\partial V}{\partial X_i} = c(X_i - X_j) \left(1 - \frac{r_{ij}^0}{r_{ij}} \right) \quad (2-8)$$

この式を Y_j または X_j で微分することで、以下の2階微分の式を得る。

$$\frac{\partial^2 V}{\partial X_i \partial X_j} = c \left\{ 1 - \frac{r_{ij}^0}{r_{ij}} + \frac{(X_i - X_j)^2 r_{ij}^0}{r_{ij}^3} \right\} \quad (2-9a)$$

$$\frac{\partial^2 V}{\partial X_i \partial Y_j} = c \frac{r_{ij}^0}{r_{ij}^3} (X_i - X_j)(Y_i - Y_j) \quad (2-9b)$$

平衡状態では $r_{ij}^0 = r_{ij}$ となるので以下のように書ける。

$$\frac{\partial^2 V}{\partial X_i \partial X_j} = \frac{c}{r_{ij}^2} (X_i - X_j)(X_i - X_j) \quad (2-10a)$$

$$\frac{\partial^2 V}{\partial X_i \partial Y_j} = \frac{c}{r_{ij}^2} (X_i - X_j)(Y_i - Y_j) \quad (2-10b)$$

同様の方法で、最終的にヘシアン行列の非対角成分は以下のようにして求めることができる。

$$H_{ij} = c \begin{pmatrix} \frac{(X_i - X_j)(X_i - X_j)}{r_{ij}^2} & \frac{(Y_i - Y_j)(X_i - X_j)}{r_{ij}^2} & \frac{(Z_i - Z_j)(X_i - X_j)}{r_{ij}^2} \\ \frac{(X_i - X_j)(Y_i - Y_j)}{r_{ij}^2} & \frac{(Y_i - Y_j)(Y_i - Y_j)}{r_{ij}^2} & \frac{(Z_i - Z_j)(Y_i - Y_j)}{r_{ij}^2} \\ \frac{(X_i - X_j)(Z_i - Z_j)}{r_{ij}^2} & \frac{(Y_i - Y_j)(Z_i - Z_j)}{r_{ij}^2} & \frac{(Z_i - Z_j)(Z_i - Z_j)}{r_{ij}^2} \end{pmatrix} \quad (2-11)$$

バネ定数は調節可能な値であり、そのためヘシアン行列の値の大きさは相対的な値である。本研究ではバネ定数の値を 1 とした。また、ANM を用いた基準振動解析では PDB データ中の各原子の座標を平衡状態の座標として用いる。

これによって得られたヘシアン行列の対角化により、 $3N$ 個の基準振動ベクトルとそれらに対応する $3N \times 3N$ の対角行列が得られる。また、対角化によって得られた固有値は基準振動ベクトルの振動数の二乗に対応している。本研究では α 炭素間のカットオフ距離を 15 \AA とした。この値は Eran Eyal らの研究によって、ANM におけるカットオフ距離に適しているとされる 15 \AA から 21 \AA の範囲内の値である [10]。

2.1.4. タンパク質ベンチマークの構成

本研究では、Guoying Qi らの提供するタンパク質データセット[13,14]をもとに、構造変化予測を行うためのベンチマークを作成した。このデータセットは、複数構造の得られているタンパク質 3120 種類の PDB データからなる。これらのタンパク質はすべて、構造変化前後で 0.5Å以上の RMSD を持ち、異なるタンパク質ファミリーに属している。よって、このデータセットは様々な種類の構造変化を含む網羅的なデータセットである。

本研究では、構造変化前後のタンパク質を 1 対 1 で比較するため、構造変化前後で同一のアミノ酸配列を持つ 1 本鎖のタンパク質 415 種類を選択し、本研究で用いるベンチマークとした。また、先行研究により、タンパク質の約半数がドメイン運動をすることが知られているため、ベンチマーク中の約半数のタンパク質に対して集団的構造変化の予測が有効であると期待される。

タンパク質の構造変化前後を区別する基準として、本研究では OPEN 構造と CLOSE 構造を定義し、OPEN 構造を構造変化前、CLOSE 構造を構造変化後として扱うことにした。OPEN 構造と CLOSE 構造の判定は、タンパク質を構成する全 α 炭素原子間での平均距離 $\langle\Delta r\rangle$ を比較することで行った。

$$\langle\Delta r\rangle = \sum_{i \neq j} \frac{\sqrt{(\mathbf{r}_i - \mathbf{r}_j)^2}}{C_1^N} \quad (2 - 12)$$

$$\langle\Delta r\rangle_{\text{OPEN}} > \langle\Delta r\rangle_{\text{CLOSE}}$$

C_1^N は原子間の組み合わせの総数、 \mathbf{r}_i と \mathbf{r}_j はそれぞれ i 番目と j 番目の原子の座標を意味する。本研究では、 $\langle\Delta r\rangle$ の大きい方の構造は OPEN 構造、 $\langle\Delta r\rangle$ が小さい方の構造は CLOSE 構造と定義した。本研究では OPEN 構造に対して基準振動解析を行い、構造

変化を予測するが、それは OPEN 構造の方が CLOSE 構造に比べて大きな構造変化がよりあらわになると期待したためである。

2.2. 手順

本研究ではベンチマークの 415 種類のタンパク質の OPEN 構造に対して ANM を用いた基準振動解析を行った。その後、得られた基準振動ベクトルと実際の構造変化のベクトルを比較し、集団的構造変化の基準を検討し、基準振動ベクトルによってどの程度の一致率で実際の集団的構造変化を予測できるかを検討した。研究の手順は以下のようになる。

2.2.1. 基準振動ベクトルの算出

基準振動解析によって、各タンパク質につきそれぞれ $3N$ 個の基準振動ベクトルと固有値を得た。各タンパク質は固有値がほぼ 0 に等しい基準振動ベクトルをそれぞれ 6 個ずつ持っていた。これらはタンパク質の構造変化に寄与しない、並進と回転に対応する基準振動ベクトルである。本研究ではこれらの 6 個の基準振動ベクトルを除外した、 $3N-6$ 個の基準振動ベクトルを用いて構造変化の予測を行った。

2.2.2. 大規模かつ集団的な構造変化に寄与する基準振動ベクトルの選択

本研究では、ヘシアン行列の対角化によって得られた基準振動ベクトルと固有値を用いて、大規模かつ集団的な構造変化に対応する基準振動ベクトルを以下のようにして検出した。

本研究では j 番目の基準振動ベクトルに対応する寄与率 A_j と Collectivity K_j を計算した。寄与率は、基準振動ベクトルが構造変化による原子の平均二乗変位に寄与する割合を示す値であり、以下の式で求められる。

$$A_j = \frac{\frac{1}{\lambda_j}}{\sum_{i=1}^{3N-6} \frac{1}{\lambda_i}} \quad (2-13)$$

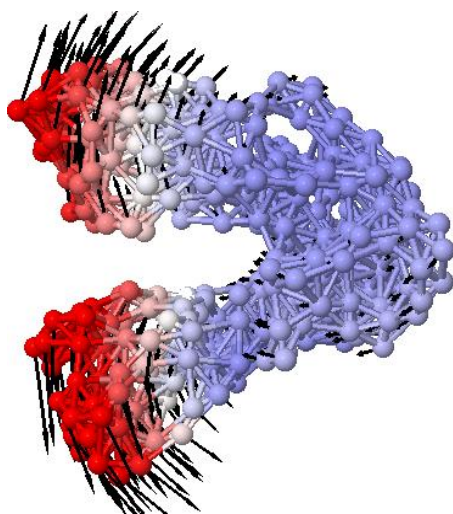
寄与率は 0 から 1 までの値をとり、寄与率が 1 に近い基準振動ベクトルならばタンパク質構成原子の座標変位を 1 ベクトルだけで表すことができる。一方、寄与率が 0 に近い基準振動ベクトルは構造変化による原子の座標変位に寄与しない。よって、タンパク質の大規模な構造変化は寄与率の高い基準振動ベクトルによって支配される。

一方、Collectivity は、対応する基準振動ベクトルの構造変化がタンパク質全体にどの程度分散しているかを示す値であり、以下のようにして計算できる。

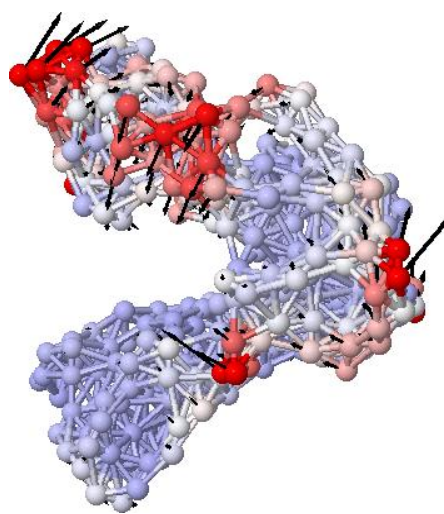
$$K_j = \frac{1}{N} \exp\left(-\sum_{i=1}^N u_{ij}^2 \log u_{ij}^2\right) \quad (2-14)$$

u_{ij} は、j 番目の基準振動ベクトルによる i 番目の原子の座標変位の大きさを指す。タンパク質全体に原子の座標変位が分散しているほど K_j の値は 1 に近づき、座標変位が局在しているほど $1/N$ に近づく。よって、集団的構造変化は Collectivity の高い基準振動ベクトルによって支配される。

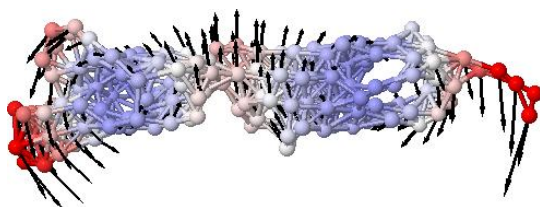
よって、高寄与率、高 Collectivity の基準振動ベクトルを検出することで、タンパク質の大規模かつ集団的構造変化を予測することができる。



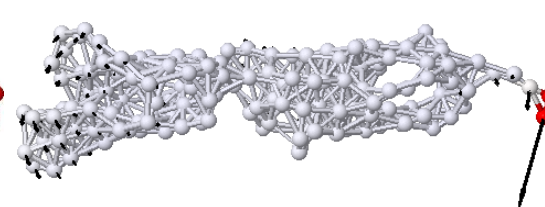
(a) 寄与率 0.18



(b) 寄与率 0.01



(c) Collectivity 0.59



(d) Collectivity 0.03

Fig2-2. 寄与率や Collectivity の大きさと、基準振動ベクトルによる構造変化の関係

黒い矢印は任意のタンパク質に対する基準振動ベクトルの相対的な振幅の大きさを示す。寄与率の高い基準振動ベクトル(a)は、寄与率の低い基準振動ベクトル(b)よりも大きな振幅を持ち、タンパク質の大規模な構造変化に大きく寄与する。

また、Collectivity の高い基準振動ベクトルによる構造変化は(c)のようにタンパク質全体に構造変化が分散している。一方、Collectivity の低い基準振動ベクトルでの構造変化は(d)のように局所的である場合が多い。

2.2.3. ドメイン運動との対応関係の検証

本研究で集団的構造変化の指標とした Collectivity と、先行研究の為されているドメイン運動の比較を行った。まず、構造変化前後の立体構造データをもとにタンパク質のドメイン運動を判定するソフト DynDom を用いて、415 種類のタンパク質に対してドメイン運動の有無を判定した。

次に、各タンパク質の実際の構造変化と最も高い内積を持つ基準振動ベクトルを選択し、その Collectivity の分布がドメイン運動の有無によってどのように異なるかを検証した。

2.2.4. 構造変化カバー率の算出

j 番目の基準振動ベクトルと実際に観測される構造変化のベクトルの類似性を評価する基準として、内積 $\Delta\mathbf{r} \cdot \mathbf{u}_j$ を計算した。内積は以下の式を満たす。

$$\sum_{j=1}^{3N-6} (\Delta\mathbf{r} \cdot \mathbf{u}_j)^2 = |\Delta\mathbf{r}|^2 \quad (2-15)$$

内積を用いて、目的のタンパク質の構造変化を M 個の基準振動ベクトルで表現できる割合を示す値、カバー率 Q を定義した。

$$Q = \sum_{j=1}^M \left(\frac{\Delta\mathbf{r} \cdot \mathbf{u}_j}{|\Delta\mathbf{r}|} \right)^2 \quad (2-16)$$

M 個の基準振動ベクトルとの内積が大きければカバー率は 1 に近づき、内積が小さければカバー率は 0 に近づく。

本研究では、より高いカバー率を得られる基準振動ベクトル M 個の選択方法を検討した。また、カバー率と寄与率、Collectivity との対応関係を検証した。

2.2.5. 構造変化の予測

ANMの基準振動解析によって得られた基準振動ベクトルには振幅の情報が無いため、どのようにして振幅を割り振るかが問題となる。

本研究では、基準振動ベクトルに対してランダムの振幅を加えることによってタンパク質の構造を変化させ、それらの立体構造データを出力した。そしてそれらの構造変化後の構造と、実測の CLOSE 構造との間の RMSD を計測した。ランダムの重みを発生させる手法としては、ボックス=ミュラー法を用いてガウス乱数を発生させる方法を用いた。ボックス=ミュラー法によってガウス乱数 R を発生させる式は以下のよう

$$\begin{cases} R = \sigma\sqrt{-2\ln R_1} \sin(2\pi R_2) \\ R = \sigma\sqrt{-2\ln R_1} \cos(2\pi R_2) \end{cases} \quad (2-17)$$

R_1, R_2 はそれぞれ 0 から 1 までの一様乱数である。また、 σ の値はタンパク質ごとに適した値が異なり、基準振動解析の結果との相関が見られなかったため、0 から 10 までの一様乱数とした。

本来のタンパク質の構造変化による各原子座標の変位 $\Delta\mathbf{r}$ は以下のようにして表すことができる。

$$\Delta\mathbf{r} = \sum_{j=1}^{3N-6} (\Delta\mathbf{r} \cdot \mathbf{u}_j) \mathbf{u}_j = \sum_{j=1}^{3N-6} c_j |\Delta\mathbf{r}| \mathbf{u}_j \quad (2-18)$$

ここで、基準振動ベクトル M 個によって構造変化が記述できるものとして以下のように近似する。

$$\Delta\mathbf{r} = \sum_{j=1}^M c_j |\Delta\mathbf{r}| \mathbf{u}_j = \sum_{j=1}^M \frac{c'_j}{\omega_j} \mathbf{u}_j \quad (2-19)$$

ω_k は基準振動ベクトルの振動数であり、固有値の $1/2$ 乗で得られる。 c_j, c'_j は構造変化後の構造が無くては求めることができない。本研究では、 c'_j をガウス乱数で発生させた場合と、実測の構造変化をもとに得た振幅を割り当てた場合との比較を行い、どの程度の確率で予測構造が CLOSE 構造と一致するかを調べた。

3. 結果と考察

3.1. ドメイン運動との比較

タンパク質のドメイン運動を数値的に判定するソフトウェア DynDom を用いて、ドメイン運動と Collectivity の対応関係を検証した。その結果、DynDom によってドメイン運動有り と判定されたタンパク質は、実際の構造変化と最大の内積を持つ基準振動ベクトルにおいて、高い Collectivity を持つ傾向が有ることが分かった(Fig3-1)。この結果は、Collectivity が集団的構造変化と相関していることを示す。

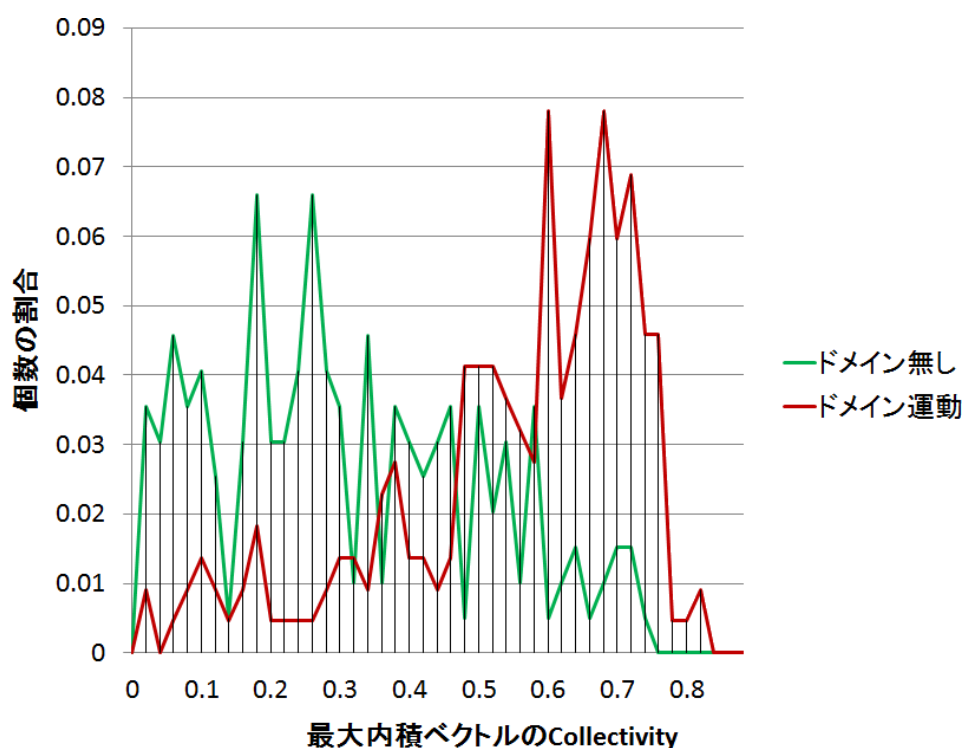


Fig3-1 ドメイン運動の有無によって生じるカバー率の違い

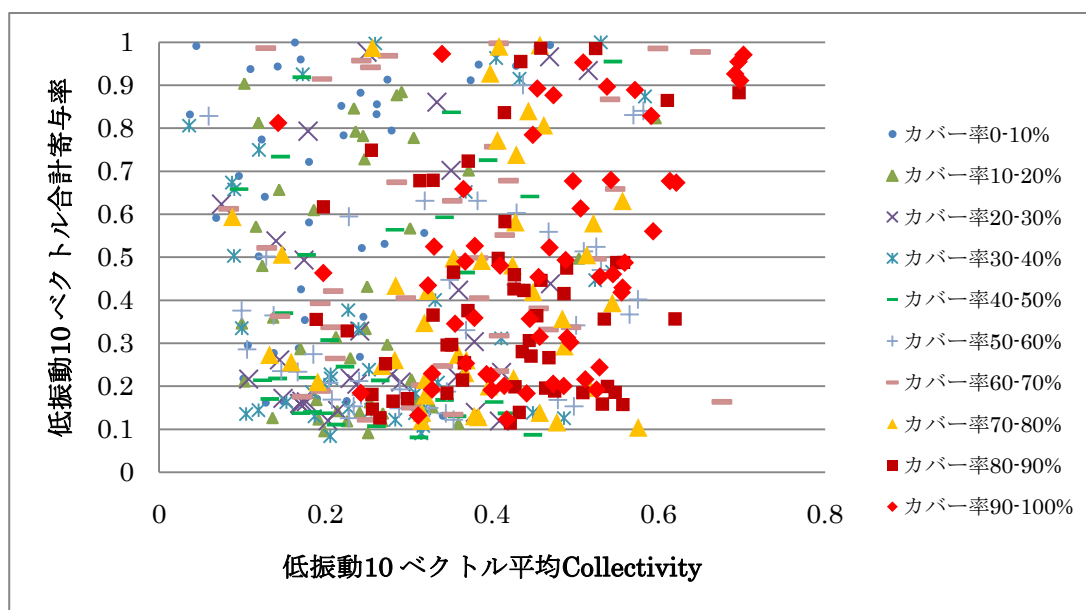
DynDom によってドメイン運動有り と判定されたタンパク質は大きい Collectivity を持つ傾向が有る。

3.2. 寄与率と Collectivity の分布

基準振動解析によって得られた寄与率と Collectivity の対応関係を求めた。低振動数 10 ベクトルに対して、合計寄与率と平均 Collectivity を求め、それらの相関を調べたがこの二つの値の間には相関関係は見受けられなかった (Fig3-2(a))。しかし、構造変化カバー率の期待値は平均 Collectivity の増加に伴って増加している。

続いて、Collectivity の高い 10 個の基準振動ベクトルに対して同様のプロットをした (Fig3-2(b))。これによってより高いカバー率が得られることを期待したが、結果的にカバー率の期待値は低下した。これはカバー率が低く Collectivity と振動数が大きい基準振動ベクトルの存在を示している。そこで、Fig3-3 のように、高 Collectivity ベクトルの選出を固有値 5 以内に限定し、寄与率の低い基準振動を除外することを試みた。これにより、カバー率の期待値はわずかに上昇したが、低振動数 10 ベクトルを用いた場合の平均カバー率には届いていない。

(a) 平均カバー率 : 54%



(b) 平均カバー率 : 33%

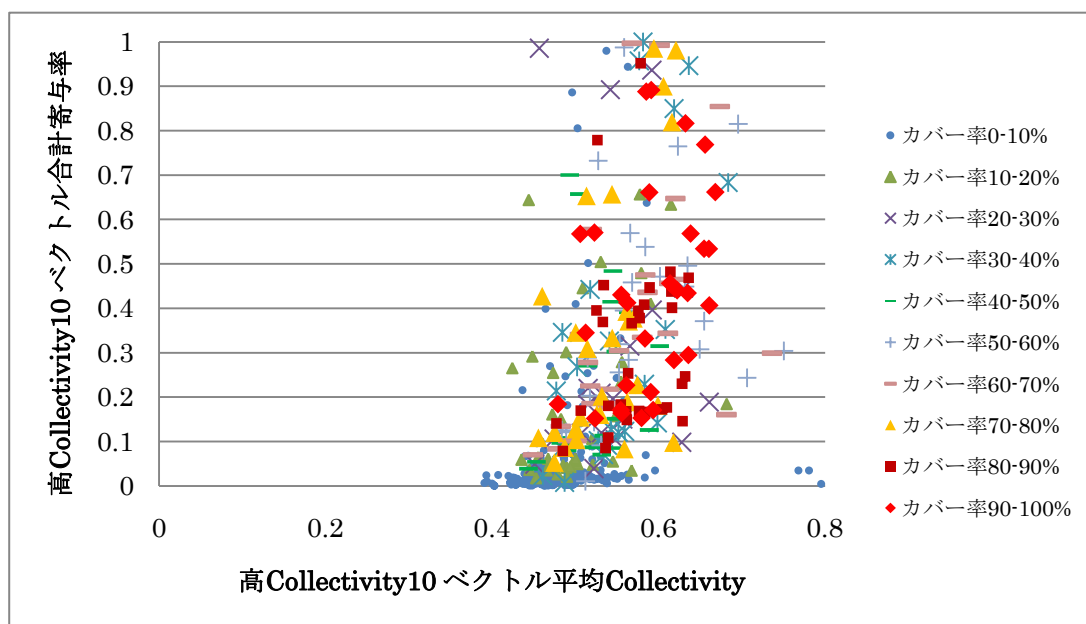


Fig 3-2 タンパク質の構造変化を記述するのに必要なベクトル数

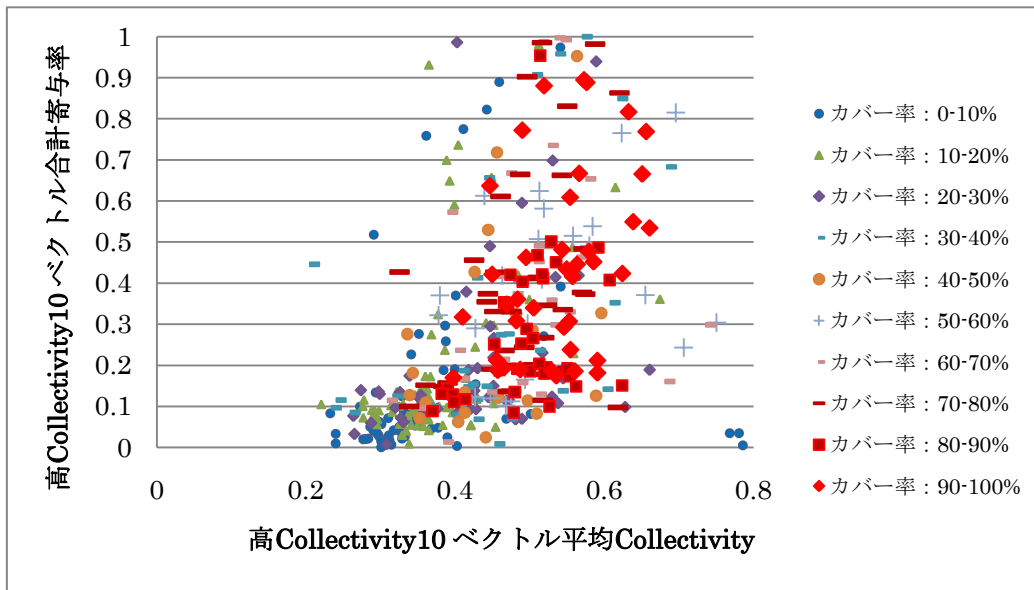


Fig3-3 高振動数側（固有値 5 以上）のベクトルを排除した上での高 Collectivity10 ベクトルと合計寄与率、カバー率の関係（平均カバー率：42%）

3.3.構造変化カバー率の推移

Fig3-2(a)の図をもとに、低振動 10 ベクトルの合計寄与率、平均 Collectivity の値に対応するカバー率の期待値をマップにした(Fig3-4)。その結果、カバー率と合計寄与率との間には強い相関は見られなかったが、カバー率と平均 Collectivity の間には相関が見られ、平均 Collectivity の増加にともなってカバー率の期待値も上昇することが分かった。

これによって、低振動 10 ベクトルの平均 Collectivity が低いタンパク質に対しては、基準振動解析による構造変化予測が高いカバー率で行えないことが分かった。そこで本研究では、低振動 10 ベクトル平均 Collectivity が 0.35 以上のタンパク質に対して、それを集団的構造変化と判定し、構造変化の予測を行うことにした。この平均 Collectivity の値は、低振動数 10 ベクトルによってカバー率約 70%、約半数のタンパク質に対して集団的構造変化と判定できる値である。

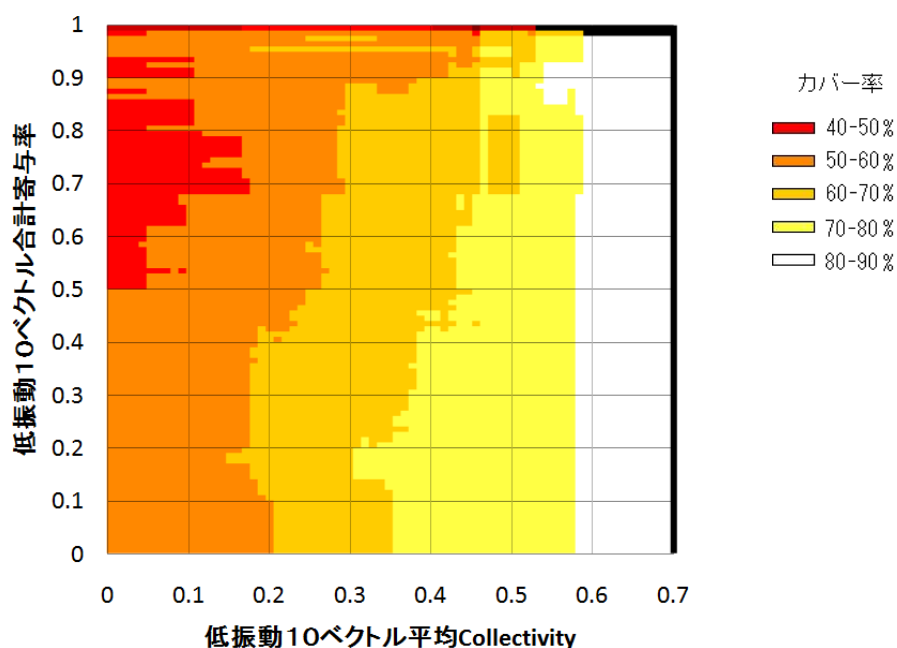


Fig3-4 タンパク質の寄与率と Collectivity から予測されるカバー率の期待値
 低振動数 10 ベクトルの平均 Collectivity と合計寄与率に対するカバー率期待値をマップにした図である。

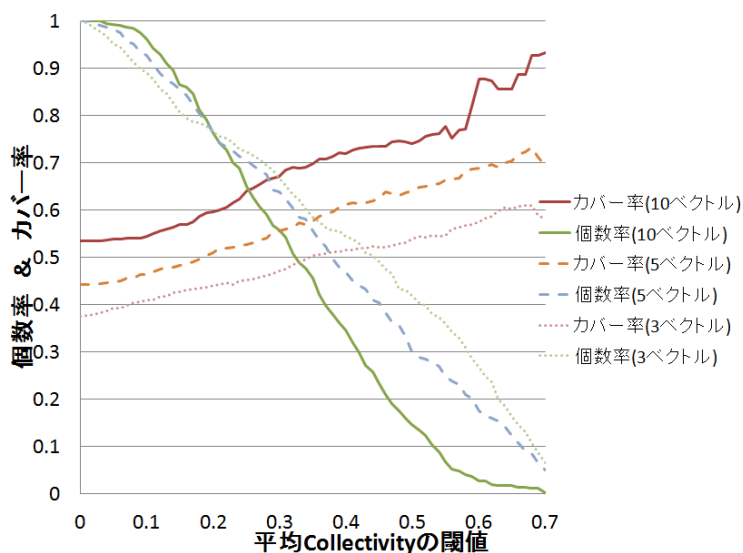


Fig3-5 Collectivity の閾値に伴うカバー率の期待値とサンプル数の変化

平均 Collectivity の低いタンパク質が基準振動解析に不向きであることから、平均 Collectivity の閾値を考え、閾値以上のタンパク質の個数の割合とカバー率の期待値を求めた。

3.4. 予測構造と実際の構造の一致率

本研究では、構造変化予測を行うタンパク質を以下のようにして決定した。以下の表をもとに、本研究では大規模な構造変化の予測を目的とするため、OPEN-CLOSE 間の RMSD が 2.0 Å 以上の 82 個のタンパク質に限定した。

<collectivity> > 0.35				
RMSD(open-close) Å	> 0	> 1	> 2	> 3
個数	189	165	82	52

次に、予測構造と実際の CLOSE 構造が一致したと判定する条件を検討した。以下の表をもとに、本研究では、予測構造と CLOSE 構造との RMSD が 1.5 Å 未満である場合を構造一致と判定することにした。この値は、内積の振幅を各基準振動ベクトルに割り振った際に 52 個のタンパク質で構造一致と判定される値である。

<collectivity> > 0.35 & RMSD(OPEN-CLOSE) > 2 Å				
RMSD(予測-close) Å	< 1.0	< 1.5	< 2.0	< 2.5
個数	32	52	59	65

(2-19)式を用いて構造変化後のタンパク質構造を発生させた。

内積を用いて振幅を各ベクトルに割り振った際に CLOSE 構造と予測構造が一致するタンパク質の数は、予測に用いるベクトル数 $M=3,5,10$ に対して、以下の表のようになった。

	RMSD(OPEN-CLOSE) > 2.0
ベクトル数	構造一致タンパク質数 / 82 個
M=3	38
M=5	42
M=10	52

一方、乱数を用いて振幅を割り振る施行を 100 回および 1000 回行い、一致構造が含まれたタンパク質の個数、および一致した構造数は以下の表のようになった。

	<collectivity> > 0.35 & RMSD(OPEN-CLOSE) > 2.0			
	100 構造ずつ予測		1000 構造ずつ予測	
ベクトル数	一致タンパク質数/82	一致構造数	一致タンパク質数/82	一致構造数
M=3	16±3 (18%)	47±5	25±1 (30%)	512±6
M=5	12±2 (15%)	29±4	23±1 (26%)	317±16
M=10	8±1 (11%)	20±4	20±1 (24%)	175±12

0.35 以上の 10 ベクトル平均 Collectivity を持つタンパク質に対して無作為に 10 ベクトルに対する振幅を設定した際の一致率をプロットした。

使用するベクトルの数が 3 個、5 個、10 個と増加するに従って乱数による探索の範囲が広がるため、同じ 1000 回ずつの構造生成の施行回数であれば一致率は減少する。また、構造変化の RMSD が大きく平均 Collectivity が高いタンパク質であっても Induced Fit の影響で構造が大きく歪んでいるタンパク質は、この手法では構造変化が予測できなかった。

3.5. 今後の課題

現時点では、予測される構造と実際の構造との一致率は十分に高くはない。これは、基準振動ベクトルの振幅の決定が難しいことが原因である。構造変化を予測する際の振幅を適切に選択する手法を検討することで、予測構造の一致率を上げることが可能と考えられる。一例として、結合する分子の位置や大きさを振幅の基準として加味することで適切な振幅が得られる可能性が有る。

また、よりも小さなカットオフ距離を用いて基準振動解析を行うことで、タンパク質の局所的な構造変化が検出できることが期待される。タンパク質の局所的な構造変化に対応する基準振動ベクトルは低い **Collectivity** と高い寄与率を持つことが多いため、**Collectivity** の低い構造変化に対しても予測の範囲を拡大できる可能性がある。また、本研究では **OPEN** 構造のタンパク質に対する基準振動解析しか行っていないが、今後は **CLOSE** 構造に対しても同様の研究を行う必要性も考えられる。

結論

本研究では、PDB に登録されている単量体タンパク質の立体構造データをもとに、タンパク質の集団的構造変化を予測する手法を検討した。そのために、まず 415 種類のタンパク質ベンチマークに対して基準振動解析を行った。そして、集団的構造変化の指標の一つとして Collectivity を計算し、先行研究が為されているドメイン運動との相関を確認した。

次に、集団的構造変化の有無を判定する基準を検討した。低振動数 10 ベクトルの平均 Collectivity が 0.35 以上のタンパク質を集団的構造変化の基準とすることで、集団的構造変化をカバー率 70%以上の期待値で予測することが可能となった。

最後に、変化後の構造を予測して実際の構と比較する実験を行ったものの、基準振動ベクトルの振幅の決定が困難なため、十分な一致率での構造予測はできていない。

謝辞

本研究の遂行にあたって、熱心な指導とアドバイスをいただきました、北尾彰朗准教授ならびに、城地保昌助教授には心よりお礼を申し上げます。

研究室のメンバーである竹村和浩博士、Yang LeeWei 博士、Raghunadha Reddy Burri 博士、西原泰孝博士、Choon-Peng Chng 博士、二島渉博士、桜庭俊博士、大森聡博士、田口淳子さん、原田隆平さん、山守優さん、丁月曦さん、若井信彦さん、原田和幸さんには、セミナーや日常会話を通して助けていただきました。また、秘書の岩佐千里さん、柴山秀美さんにも、研究生活を支えていただきました。

皆様に心よりお礼を申し上げます。

参考文献

- [1] Steven Hayward, Akio Kitao, and Herman J.C. Berendsen (1997) Model-Free Methods of Analyzing Domain Motions in Proteins From Simulation: A Comparison of Normal Mode Analysis and Molecular Dynamics Simulation of Lysozyme . *PROTEINS: Structure, Function, and Genetics* ., 27, 425-437
- [2] Hayward,S. (1999) Structural principles governing domain motions in proteins. *Proteins.*, 36, 425-435
- [3] Hayward S, Berendsen HJ (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* ., 30, 144-154
- [4] Richard A. Lee, Hayward.S (2003) The DynDom database of protein domain motions. *bioinformatics* ., 19, 1290-1291
- [5] Dobbins, S. E., V. I. Lesk, et al. (2008). "Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking." *Proc Natl Acad Sci U S A* 105(30): 10390-10395.
- [6] Qiang Cui, Ivet Bahar (2006) *Normal Mode Analysis: Theory And Applications to Biological And Chemical Systems* . (Chapman & Hall, London).

- [7] May A, Zacharias M (2008) Energy Minimization in low-frequency modes to efficiently allow for global flexibility during systematic protein-protein docking. *Proteins* 70,794-809
- [8] Atilgan, A. R., S. R. Durell, et al. (2001). "Anisotropy of fluctuation dynamics of proteins with an elastic network model." *Biophys J* 80(1): 505-515.
- [9] Eran Eyal, Lee-Wei Yang(2006) Anisotropic Network Model: systematic evaluation and new web interface. *Bioinformatics* 22, 2619-2627
- [10] Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Engineering* ., 14, 1-6
- [11] B.L.de Groot, D.M.F. van Aalten (1997) Prediction of Protein Conformational Freedom From Distance Constraints. *Proteins* 29, 240-251
- [12] Geometry-Based Sampling of Conformational Transitions in Proteins
Daniel Seeliger,¹ Jürgen Haas,² and Bert L. de Groot
- [13] Wataru Nishima, Guoying Qi, Steven Hayward, and Akio Kitao (2009) DTA: dihedral transition analysis for characterization of the effects of large main-chain dihedral changes in proteins, *bioinformatics*, 25, 628-635
- [14] Qi, G. *et al.* (2005) A comprehensive and non-redundant database of protein domain movements. *Bioinformatics*, 21, 2832–2838.