

Abstract of Doctoral Dissertation

Title : Prediction of the folding processes for immunoglobulin-like beta-sandwich and beta-trefoil proteins

Doctoral Program in Advanced Life Sciences
Graduate School of Life Sciences
Ritsumeikan University

アウンプチン パンヤブット
AUMPUCHIN Panyavut

Describing the folding mechanism of a protein is one of the goals in molecular bioinformatics study. Nowadays, it is enigmatic and difficult to extract folding information using standard bioinformatics techniques or even experimental protocols which can be time consuming. However, protein folding mechanisms have been intensively investigated with experimental as well as simulation techniques. Since a protein folds into its specific 3D structure from a unique amino acid sequence, it is interesting to extract as much information as possible from the amino acid sequence of a protein. The mainly purpose of this thesis is to extract the folding mechanisms of variety target proteins, especially the structure formation in an initial state of folding, from its amino acid sequence by means of inter-residue average distance statistics methods Average Distance Map (ADM) and contact frequency analysis (F-value). Moreover, a coarse-grained Gō-model simulation was applied for some proteins to simulate the whole story of protein folding. The proteins used in this thesis are immunoglobulin (Ig)-like beta sandwich fold proteins and β -trefoil fold proteins.

Ig-like beta sandwich fold protein is one of the most common structural families that play key roles in the immune system. In this study, the PDB codes 1TIT and 1TEN are used to represent the Ig domain and FN3 domain, respectively. The β strand 3 and 5 are suggested to be the significant strands for the initial folding processes in both proteins. This common feature coincides well with the experimental results and underscores the significance of the beta-sandwich proteins' common structure, namely, the key strands for folding and the Greek-key motif, which is located in the central region. The results show good correspondence to experimental data, ϕ -value and protection factor from H-D exchange experiments. The significance of conserved hydrophobic residues near F-value peaks for structural stability by using hydrophobic packing is confirmed. It is worth noting that the sequence-based techniques were able to predict the initial folding event just next to the denatured state. The long-range contacts formed in an early event of folding derived from Gō-model simulation also corresponded well to the experimental data and the conserved hydrophobic residues near the peaks of F-value.

The β -trefoil fold proteins are known to have the same 3D scaffold, namely, a three-fold symmetric scaffold, despite the proteins' low sequence identity among superfamilies. In this study, 26 high symmetric β -trefoil proteins and four irregular

structure β -trefoil proteins from six superfamilies were analyzed by the same manner as Ig-like beta sandwich fold protein. The ADM and F-value analyses predict that the N-terminal and C-terminal regions are conserved to form a compact region in an initial state of folding and that the hydrophobic residues at the central region can be regarded as an interaction center with other residues. These results coincide well with the experimental data obtained so far for folding of some of the β -trefoil proteins. In the case of irregular β -trefoil proteins which an experimental data has not been available so far, the 3D-based G \ddot{o} -model simulations were performed to investigate the protein folding processes. It is interesting that the conserved hydrophobic residues near the peaks of F-value plot also observed to form the long-range interaction in the contact frequency map derived from a G \ddot{o} -model simulation. Moreover, the results indicate that the irregular parts in the β -trefoil proteins do not hinder the protein formation. Conserved hydrophobic residues on the β 5 strand are always the interaction center of packing between the conserved hydrophobic residues in both regular and irregular β -trefoil proteins. Suggesting that β strand 5 plays an important role in β -trefoil protein structure construction.

In summary, it can be confirmed that the average distance statistics methods, ADM and F-value analysis, were able to predict the initial folding event just next to the denatured state by using only amino acid sequence information and that a 3D-based G \ddot{o} -model simulation can be used to investigate the whole process of protein folding.

博士論文要旨

論文題名：免疫グロブリン様ベータ - サンドイッチタンパク質およびベータ - トレフォイルタンパク質のフォールディング過程の予測

立命館大学大学院生命科学研究科
生命科学専攻博士課程後期課程

アウンプチン パンヤブット
AUMPUCHIN Panyavut

タンパク質に立体構造形成（フォールディング）機構を詳細に記述することはバイオインフォマティクスの目標の一つである。しかしながら、近年においてもフォールディング情報を標準的な生命情報学的手法により抽出することはまだまだわからないことも多く困難であり、実験的手法を用いるにしても多大な労力と時間を必要とするためやはり困難である。それでも、タンパク質フォールディング機構の研究は、実験的手法あるいはシミュレーションを用いて精力的に行われている。固有のアミノ酸配列を持つタンパク質が、対応する特異的立体構造を形成（フォールド）するので、アミノ酸配列からできるだけ多くの情報を抽出することは非常に興味ある問題である。本論文の主な目的は、様々な対象タンパク質のフォールディング機構、特に初期構造形成部位、の情報を残基間平均距離統計に基づく方法（Average Distance Map(ADM)）と残基対コンタクト頻度解析(F 値)を用いてアミノ酸配列情報から抽出することである。さらに、本研究ではいくつかのタンパク質に対しては粗視化 Gō モデルを用いてフォールディング過程全体のシミュレーションを行った。本研究では Ig 様 β サンドイッチフォールドタンパク質及び β -トレフォイルタンパク質を取り扱う。

本研究ではタンパク質立体構造分類データベースである SCOPe の分類における免疫グロブリン様 β サンドイッチフォールドタンパク質を取り上げる。その中の免疫グロブリン(Ig)及びフィブロネクチンタイプ 3(FN3)スーパーファミリーのタンパク質タイティン(titin)とテナシン(tenascin)を代表タンパク質として選んだ。それぞれのタンパク質のタンパク質データベース (PDB)コードは 1TIT、1TEN である。本研究では両タンパク質においても β 鎖 3, 5 がフォールディング過程において重要であることが示唆された。この 2 つのタンパク質に共通の特徴は実験結果ともよく一致し、 β サンドイッチタンパク質に共通に見られるタンパク質の中央部分に位置するキーストランドやグリークキー構造の形成に重要であることが示された。これらの結果はこれまで発表された、 β 値解析、H-D 交換実験における保護因子といった実験結果とよく対応する。F 値ピーク近

傍の保存疎水残基が構造安定性に重要であることが天然構造のパッキングに吟味からも確認された。ここで用いられているアミノ酸配列に基づく方法は変性状態に近い状態での初期フォールディング現象であることは注意するべきであろう。また G δ モデルによるフォールディングシミュレーションで見られる初期フォールディング過程における、F 値ピーク近傍の保存疎水残基による長距離コンタクト形成も実験結果とよく対応することも確認された。

β -トレフォイルフォールドタンパク質は、それぞれ異なるスーパーファミリーに属するタンパク質の間ではアミノ酸配列相同性が低いにもかかわらず、ほぼ同じ構造、つまり 3 つの構造単位からなる、三回対称性を持つタンパク質として知られている。本研究では、6 つのスーパーファミリーの中から 26 個の高い対称性をもつ β -トレフォイルタンパク質と 4 つの不規則構造を含む β -トレフォイルタンパク質を選び、Ig 様 β サンドイッチフォールドタンパク質に対して用いた方法と同じ手法を適用し解析を行った。ADM 及び F 値解析の結果はフォールディングの初期過程において N 末端と C 末端領域が共通して形成され、中央部分に存在する疎水残基が他の残基を繋ぐ相互作用中心であることを予測した。この予測は β -トレフォイルタンパク質のフォールディングに関する実験結果とよく一致する。不規則構造を含む β -トレフォイルタンパク質については、これらのタンパク質に対してはフォールディング実験はなされていないので、立体構造に基づいた G δ モデルシミュレーションを行い、フォールディング過程を調べた。興味深いことに、G δ モデルシミュレーションにおいても F 値ピーク近傍の保存疎水残基が、シミュレーション途中のコンタクト頻度マップにおいて長距離コンタクトを形成することが示された。さらに、 β -トレフォイルタンパク質の不規則部分はフォールディングを阻害しないことが予測された。 β 5 鎖上の保存疎水残基は、対称的な β -トレフォイルタンパク質においても不規則構造を含む β -トレフォイルタンパク質において常に保存疎水残基の相互作用中心を形成していた。このことは β 5 鎖が β -トレフォイルタンパク質の構造形成において重要な役割を担っていることを示している。

以上をまとめると、本研究では、Ig 様 β サンドイッチフォールドタンパク質と β -トレフォイルタンパク質に対し、平均距離統計法、ADM・F 値解析、は変性状態に近いフォールディング初期過程をアミノ酸配列情報のみから予測することが可能であり、立体構造に基づいた G δ モデルシミュレーションはタンパク質フォールディングの全過程を調べるのに有用であることが示された。