

平成 21 年 6 月 26 日現在

研究種目：基盤研究（C）

研究期間：2007～2008

課題番号：19510202

研究課題名（和文）残基間平均距離統計を用いたタンパク構造形成部位の予測

研究課題名（英文）Prediction of protein folding sites based on the interresidue average distance statistics

研究代表者

菊地 武司（KIKUCHI TAKESHI）

立命館大学・生命科学部・教授

研究者番号：90195206

研究成果の概要：本研究においてはタンパクの立体構造の精密予測を目標として、まずタンパク立体構造形成部位の位置や構造形成にかかわるアミノ酸の予測を試みるものである。研究期間において、タンパクにおけるアミノ酸間平均距離統計に基づくコンタクトマップ（平均距離図、Average Distance Map）などを用い、グロビンや免疫グロブリン結合タンパクといったタンパクの実験結果と予測結果を詳細に比較・精査を行い本研究の有効性を示すことができた。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,900,000	570,000	2,470,000
2008年度	1,600,000	480,000	2,080,000
総計	3,500,000	1,050,000	4,550,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム情報科学

キーワード：バイオインフォマティクス、タンパク構造予測

1. 研究開始当初の背景

(1) 研究分野について

タンパクの立体構造をアミノ酸配列情報のみから予測するという問題、いわゆるab-initio予測は古くから論じられている問題ではあるが、今なお確立された方法というものは存在しない。それでもかなりの部分ab-initio予測は可能になってきているが、ドラッグデザインなどの実用的に有用なレベルには達していない。現在提案されている予測法を少しずつ改善していけば、実用に耐える予測能力を有する方法が開発される可能性もないではないが、やはりタンパク立

体構造精密予測のための別の戦略が必要であると思われるのが現状であろう。本研究では今まで提案されている戦略とは異なったタンパクの立体構造予測法の確立を目指し、第一段階の方法論の構築を図る。

(2) 本研究の準備状況

応募者は、これまでタンパクにおけるアミノ酸間平均距離統計に基づくコンタクトマップ（平均距離図、Average Distance Map, ADM）を用いて、タンパク構造ドメインの位置など様々なタンパク構造情報の予測を試みてきた。特に最近、タンパクの立体構造形成過程

(フォールディング過程)における初期構造形成部位が、ADM による予測サブドメイン領域とよく一致することが見出された。従って、これらの成果に基づいていくことにより本研究の遂行が可能である。

2. 研究の目的

(1)本研究の目標設定における基礎的考え方
本研究の究極の目的は、タンパクの立体構造の実用的に有用な精密予測である。まず、タンパクの精密予測をタンパク構造を二分割することから始まる。つまり、原子レベルで精密に立体構造を予測すべき部分と、比較的ラフに予測してよい部分である。まずタンパクの一部を精密に予測しておいて、その部分を基礎として残りの部分をモデリングするわけである。それでは、精密に予測すべき部分とはどのような領域であるか。本研究の研究期間における第一の目的は、この部分を明らかにすることである。すなわち、

タンパク立体構造予測を行う際、予測を緻密に行うべき領域についてきちんと定義すること

そのような領域の立体構造に規則性が存在するか明らかにすることである。

(2)研究期間における目標

研究期間においては、まず平均距離統計に基づく解析を行い、タンパクアミノ酸配列を情報として構造形成部位(フォールディング部位)の位置や構造形成に関わるアミノ酸残基の予測法を開発することである。それが確立されれば、それに基づいてタンパクの構造形成遷移状態の立体構造の予測を試みる。すなわちこの部分が、精密に予測すべき構造部分であり、それが可能であることを示すのが本研究の目的となる。

3. 研究の方法

(1)平均距離統計に基づくコンタクトマップ(ADM)

報告者は、これまでタンパクにおけるアミノ酸間平均距離統計に基づくコンタクトマップ(平均距離図、Average Distance Map, ADM)を用いて、タンパク構造ドメインの位置など様々なタンパク構造情報の予測を試みてきた(T. Kikuchi, *et al.*, J. Protein Chem., 7, 427 (1988))。特に最近、タンパクの立体構造形成過程(フォールディング過程)における初期構造形成部位が、ADM による予測サブドメイン領域とよく一致することが見出された。(T. Ichimaru and T. Kikuchi, Proteins, 51, 515 (2003) および C. Nishimura *et al.*, Nature Structural Biology, 7, 679 (2000))このことは、ADM による予測領域が初期構造形成部位を予測

していることを示唆している。さらに、構造形成残基を特定することのできるいわゆる値解析により実験的に明らかにされたフォールディング遷移状態構造を形成する残基位置とADM により予測された部位がよく対応することも最近見出されている。この方法を本研究にも採用する。

(2)残基間コンタクト頻度の解析

報告者は独自に、残基間平均距離統計に基づく変性状態における残基間コンタクト頻度の予測・解析法を提出している。(T. Kikuchi, Amino Acids, 5, 541-549 (2008))これはタンパクをビーズモデルで近似し(図1)ADMの際に用いた残基間平均距離統計から、残基間有効ポテンシャルを定義し、モンテカルロシミュレーションを行い残基間コンタクトを計算するものである。

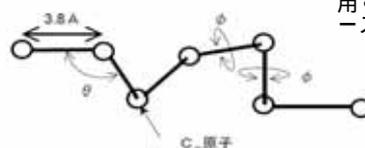


図1. F値解析に使用するタンパクのビーズモデルの概略図

その予測結果は実験における値解析の結果によく対応する。これを今F値解析と呼ぶことにする。上記タンパクに対しF値解析を行い、構造形成部分(フォールディング遷移状態形成部位)の予測の精密化を図る。

(3)典型的なタンパクのADM解析・F値解析
フォールディング実験がしっかりと行われている典型的なタンパクを選び、それぞれについてADMを作成し解析を行う。本研究では、グロビンファミリー、IgG結合タンパクファミリー、サンドイッチ型タンパクを選んだ。

(4)フォールディング部位の配列比較

同じファミリーに属するタンパクの予測構造形成部位の配列を比較し、構造形成部位のモチーフを定義する。もしこれが定義できれば、このモチーフを用いて構造既知タンパク配列を検索することにより構造形成テンプレートを定義することが可能である。このテンプレートは立体構造予測に用いることができる。

4. 研究成果

(1)典型的なタンパクのADM解析・F値解析

グロビンフォールド

グロビンフォールドがA-Hヘリックスから構成されているのはよく知られているが、特に植物ヘモグロビンについてはE-Hヘリック

ス部分を中心としてフォールドすることが ADM 解析により予測され (T. Ichimaru and T. Kikuchi, *Proteins*, 51, 515 (2003))、大豆レグヘモグロビンにおいては NMR により検証もされている。(Nishimura *et. al.*, *Nature Str. Biol.*, 7, 679 (2000)) 図 2 にグロビンフォールドの E-H ヘリックス部分を示す。

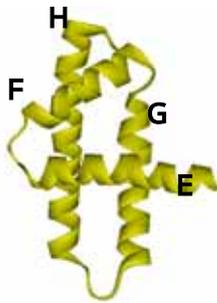


図 2 .グロビンフォールドの E - Hヘリックス部分

本研究では、公開されている DALI データベース (<http://www.ebi.ac.uk/dali/>) を利用して、レグヘモグロビン E-H ヘリックスの 3 次元座標と類似するタンパクを検索し、それらの ADM 解析を行い、確かに E-Hヘリックスに対応する部分が安定構造として存在することを確かめた。検索されたタンパクはフィコシアニン、ESCR-I、RSBR、サーカディアンクロックタンパクである。このうちサーカディアンクロックタンパクはグロビンフォールドとは異なるフォールド (KaiA/RbsU domain) であることに留意すべきである。また F 値解析から構造形成の際に重要な残基はヘリックス E と G の存在することも予測されるが、これは NMR の結果と一致する。(Nishimura *et. al.*, *Nature Str. Biol.*, 7, 679 (2000))

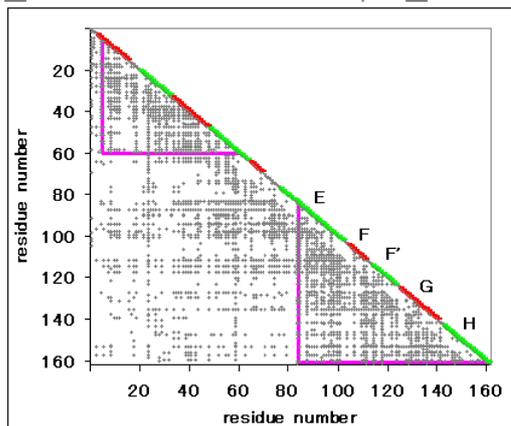


図 2 . フィコシアニンの ADM E - Hヘリックス単位が構造形成部分であることが予測される。

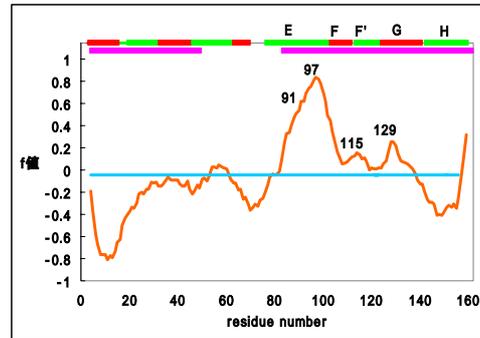


図 3 . フィコシアニンの F 値解析結果
ヘリックス E と G の残基のコンタクト頻度が高く、構造形成残基であることを示唆する。91 番目と 129 番目の残基は Tyr である。

以上の結果より、E-H ヘリックス部分はグロビンフォールド以外にも普遍的に存在し、ADM 解析の結果 E-H ヘリックス部分が構造として安定であることが予測される。すなわち、E-H ヘリックス単位がグロビンフォールドの構造形成部分であり、かつ普遍的に安定構造として存在することが示唆される。

免疫グロブリン結合タンパク

免疫グロブリン結合タンパクの一種であるプロテイン A とプロテイン G は 3 ヘリックス束 (3 束) 構造と + 構造をもつことはよく知られている。これらの関連タンパクで配列相同性が極めて高く (~ 60 - 90%) 常識的には同じ構造を持つと予想されるが、実際は異なる構造 (3 ヘリックス束と + 構造) をとるものが存在することが最近見出されている。(Alexander *et al.*, *Biochemistry*, 44, 14045 (2005))

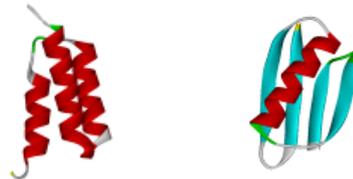


図 4 3 束構造 (左) と + 構造 (右)

本研究においてはこれらの類縁タンパクを取り上げ、フォールディング機構について検討した。プロテイン A219 とプロテイン G311 である。前者は 3 束構造、後者は + 構造をとる。配列類似性は 60% である。これらについて ADM 解析および F 値解析を行った結果、どちらの解析においても 3 束の場合は 3 つのヘリックスが疎水パッキングを形成してフォールドするのに対し、+ 構造の場合は中央のヘリックスと 3 番目と 4 番目の ストランドが構造形成に重要であることが示唆された。図 5 にプロテイン A219、図 6 にプロテイン G311 の F 値解析の結果を示す。60% の配列相同性があるにもかかわらず F 値のピークの位置が異なる。この知見は配列情報からフォールディング機構の予測可能性を示すのみならず、本方法がこのよ

うな配列相同性の高いタンパクの立体構造の違いをアミノ酸配列から指摘できること

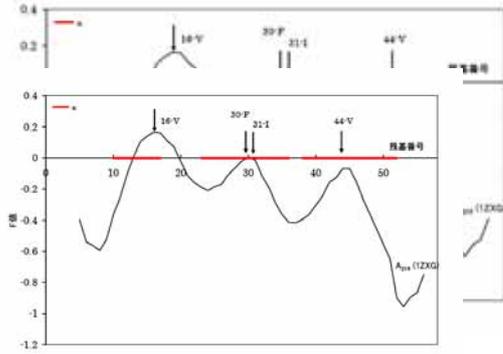


図5 プロテイン A219 の F 値解析。矢印の残基が構造形成に重要であると予測される

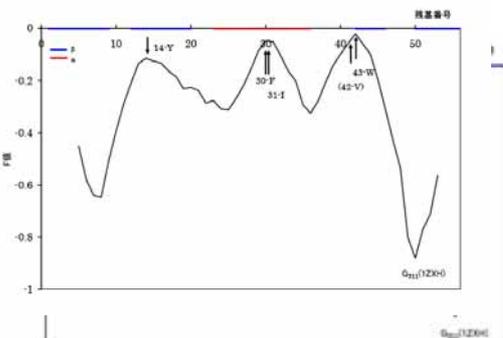


図6 プロテイン G311 の F 値解析。矢印の残基が構造形成に重要であると予測される

サンドイッチタンパク

サンドイッチタンパクは SCOP の分類において様々なスーパーファミリーにおいて共通して見られる構造であり、いわゆるスーパーフォールドである。図7に典型的な

サンドイッチタンパクであるリボソーム S 6 タンパク (PDB コード: 1RIS) の構造を示す。



図7 リボソーム S 6 タンパクの立体構造
2次構造はβ1-α1-β2-β3-α2-β4

本研究では他に同じ構造をもつタンパクとして、アシルホスファターゼ (PDB コード: 1APS)、U1 リボヌクレオタンパク (PDB コード: 1CTF)、L1/L12 リボソームタンパク (PDB コード: 1AUD)、プロカルボキシペプチダーゼ (PDB コード: 1O6X) を選んで解析した。ADM 解析と F 値解析の結果どのタンパクにおいても 1 - 1 および 3 - 2 の部分が構造形成部分であることが予測された。これは F 値解析の結果とよく一致する。(Haglund *et al.*, J. Biol. Chem., 283, 27904 (2008)) すなわち、ADM と F 値解析により サンドイッチタンパクの構造形成

部位が予測できることを示している。またこのトポロジーをもつタンパクは共通のフォールディング遷移状態構造を持つことが示唆され、この部分の立体構造が予測可能であることが推測される。

(2) 構造形成に重要な残基の普遍性

ADM および F 値解析により予測された構造形成に重要な残基が同じ構造をもつタンパクにおいて保存されているかは重要な問題である。それが保存されていれば構造形成部分のモチーフ配列として定義され、フォールディング遷移状態構造をタンパク立体構造データベースから検索することができ、フォールディング遷移状態の構造を予測することが可能である。これまでのところ、例えばグロビンフォールドにおいてはヘリックス E とヘリックス G が共通してフォールディングの中心になることは見出せ、F 値ピーク近傍に疎水残基が存在するところまでは見出せたが (図3参照) モチーフを定義するところまでは至っていない。しかしながら予備的な考察においては、サンドイッチタンパクの構造形成部位の保存残基はよく保存されていることが見出せる。図8に ADM・F 値解析で構造形成に重要と思われる2次構造の配列を比較したものを示す。星印をつけた残基は F 値解析のピークの近傍 2 ~ 3 残基以内にあり、立体構造上疎水パッキングを形成している残基である。図8より、F 値解析のピーク近傍の構造形成に關与する疎水残基はよく保存されていることが示唆される。

	β1	α1	β4	α2
1APS	KSVDVEV	FMVVAEDARKI	VTGQVQGP	EEKVHSMKSWL
1O6X	DQVLEI	HEEQIKHLLQLEAQ	AHVRVVF	HVQAVKVFLESQ
1AUD	IYENH	ELKKSIIHAI	QAFV	EVSSATNALES
1CTF	FDVILKA	KVAVIKAVHGATG	AALKEGV	SKDDAEALKKALEEAG
1RIS	RRYEVHIVH	SQALEKEITQRAIEHY	PQGYFLMYYQVEM	MDLARELR

図8 ADM・F 値解析で構造形成に重要と思われる2次構造の配列を比較したものを示す。星印をつけた残基は F 値解析のピークの近傍 2 ~ 3 残基以内にあり、立体構造上疎水パッキングを形成している残基である。

このことは、ADM・F 値解析から構造形成残基のモチーフが定義可能であることを予感させる。今後研究を展開する予定である。

(3) 一般的なタンパク構造構築形成機構とタンパク立体構造予測

本研究では、ADM 解析と F 値解析によりいくつかのタンパクにおいてアミノ酸配列からタンパク構造形成機構を予測することを試みた。また構造形成に重要な残基も予測可能であることを示した。ADM により構造形成部分と予測された領域はグロビン E-H 単位、免疫グロブリン結合タンパクではヘリックス - ヘリックス構造が - 構造と予測され、

サンドイッチタンパクにおいても - 構造が構造形成部分と予測される。さらにグロビン E-H 単位はグロビン以外でも存在することが示された。一方、ヘリックスヘリックス構造や - 構造はやはり普遍的にタンパクに見られる構造である。つまり、タンパクの構造形成部分(フォールディング遷移状態形成部分)は普遍的に見られる一般的な構造である可能性がある。言い換えれば、タンパクフォールディング遷移状態構造の数は比較的少なく、例えば PDB に存在する構造においてに頻繁に見られることが考えられる。そして、本研究はこのような領域と疎水パッキング残基の近傍領域は ADM・F 値解析により予測可能であることを示した。そのような部分の配列モチーフが定義できれば、構造形成部分の立体構造が PDB 中の構造から検索可能であると思われる。このような領域の立体構造の保存性はかなり高いと思われるので、精密な立体構造予測が可能であることが推測される。

(4) タンパク構造構築原理について

ADM 解析・F 値解析はアミノ酸間平均距離統計に基づく予測法である。アミノ酸間の平均距離にはどのような意味があるのだろうか。最近の研究において(T. Kikuchi, Physica A: Statistical Mechanics and its Applications, 388, 129 (2009)), 非常に単純化した 1 次元イジングスピン格子によるタンパクモデルによれば、ゼロフラストレーション系においては 2 対相互作用のみから定義したモードが変性状態における運動を記述する。変性状態における運動は最低モードに支配される。これはフォールディングファンネルに対応する現象である。天然構造への 2 状態転移は、変性状態の最低モードの運動から基底状態への転移とみなすことができる。このモデルが正しくタンパクの構造形成原理を記述しているならば、残基間平均距離統計による変性状態のコンタクト頻度による予測は最小フラストレーションの反映であると考えられる。そのため残基間平均距離統計に基づく方法は様々なタンパク構造の特徴を予測すると思われる。

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 3 件)

Takeshi Kikuchi, "Prediction of Protein Folding Properties Based on Interresidue Average Distance Statistics", the 10th International Congress on Amino Acids and Proteins (ICAAP) (August 2007, Kallithea, Greece).

Takeshi Kikuchi, "Prediction of Protein Folding Properties Based on Interresidue Average Distance Statistics" Amino Acids, (33), ppIX, (2007). 査読無

Takeshi Kikuchi, "Analysis of 3D structural differences in the IgG-binding domains based on the interresidue average-distance statistics.", Amino Acids, 5, 541-549 (2008). 査読有

Takeshi Kikuchi, "Spin correlations in a no frustrated one-dimensional spin system and the formation of the ground state as a model of protein folding." Takeshi Kikuchi, Physica A: Statistical Mechanics and its Applications, 388, 129-136 (2009). 査読有

[学会発表](計 10 件)

新江雄一、菊地武司、「自由エネルギー変分原理を用いた相対的結合自由エネルギー計算とジヒドロ葉酸還元酵素阻害剤への応用」、第 7 回蛋白質科学会講演要旨集 p92. (2007 年 5 月、仙台)

原田織江、菊地武司、「残基間平均距離統計ポテンシャルによる IgG 結合ドメイン類縁タンパクの構造の差異に関する解析」、第 7 回蛋白質科学会講演要旨集 p92 (2007 年 5 月、仙台)

Takeshi Kikuchi "Prediction of protein folding properties based on interresidue average distance statistics." International Conference of Amino acid and Proteins (August, 2007, Tessaloniki, Greece) (招待講演.)

坂本龍則、菊地武司、「自由エネルギー変分原理の基づくジヒドロ葉酸還元酵素阻害剤の相対的結合自由エネルギー計算」、日本生物物理学会第 45 回年会講演予稿集 S26. (2007 年 12 月、横浜)

菊地武司、「IgG 結合ドメインの立体構造の違いについて - 残基間平均距離に基づく解析 - 」、日本生物物理学会第 45 回年会講演予稿集 S27. (2007 年 12 月、横浜)

新江雄一、菊地武司、「自由エネルギー変分原理の基づく相対的結合自由エネルギー計算とその FKBP 結合タンパク阻害剤への応用」、日本生物物理学会第 45 回年会講演予稿集 S114. (2007 年 12 月、横浜)

南康智、菊地武司、「自由エネルギー変分原理に基づくFKBP(免疫抑制剤結合タンパク質)阻害剤の結合自由エネルギーの予測」、第8回日本蛋白質科学会年会講演要旨集p62。(2008年6月、東京)

永岡清秀、菊地武司、「グロビンフォールドヘリックスE-H単位の立体構造の普遍性について」、第8回日本蛋白質科学会年会講演要旨集p64。(2008年6月、東京)

菊地武司、「グロビンフォールドヘリックスE-H単位の立体構造の普遍性 - 残基間平均距離を用いた解析」、日本生物物理学会第46回年会講演予稿集S22。(2008年12月、福岡)

南康智、菊地武司、「自由エネルギー変分原理を用いたFKBP阻害剤の結合自由エネルギーの予測とリガンドの分類」、日本生物物理学会第46回年会講演予稿集S28。(2008年12月、福岡)

6. 研究組織

(1) 研究代表者

菊地 武司 (KIKUCHI TAKESHI)

立命館大学・生命科学部・教授

研究者番号：90195206