

**科学研究費助成事業 研究成果報告書**

平成 27 年 5 月 26 日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2011～2014

課題番号：23770174

研究課題名(和文)フラグメントアセンブリ法の逆発想による新規フォールド予測法の開発

研究課題名(英文) Development of a new De Novo protein structure prediction method based on a different approach of the fragment assembly method

研究代表者

千見寺 浄慈 (George, Chikenji)

名古屋大学・工学(系)研究科(研究院)・助教

研究者番号：10420366

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：タンパク質立体構造予測問題は生物物理における最重要問題の一つである。本研究はフラグメントアセンブリ法というこれまでで最も優れた構造予測手法の本質と限界を見極め、その弱点を補うべく、既知構造の安定な二次構造の順列変異によって新規フォールドを生成し立体構造予測する手法を開発した。構造予測手法の開発に際し必要となった、構造アラインメント手法やループクロス判定法など多数の構造解析プログラムも開発した。構造予測ベンチマークテストの結果、従来法では全く予測できなかったものに対して優れた予測が出来た例もあり、従来法と全く異なるアプローチによる立体構造予測法の確立に向けて確かな進歩を遂げることができた。

研究成果の概要(英文)：Protein structure prediction is one of the most important problems in biophysics. Although the fragment assembly method has been known as the most successful method, it has clear limitation. In this research project, we developed a new De Novo protein structure prediction method based on a different approach of the fragment assembly method. In this method, new fold structures are generated by permutation of secondary structures, and model qualities are assessed by physic-chemical energy function. We also developed several algorithms of structural bioinformatics, such as a non-sequential structure alignment algorithm and a program for detecting loop crossing. We performed protein structure prediction benchmark tests using the new method and found that our method outperformed the fragment assembly method for some targets.

研究分野：生物物理

キーワード：タンパク質 立体構造予測 フラグメントアセンブリ法 構造アラインメント non-sequential

## 1. 研究開始当初の背景

アミノ酸配列の情報だけからタンパク質の立体構造を予測することは、遺伝子型の情報から表現型への情報への変換過程を理解するという観点で興味深いばかりでなく、この情報変換過程が翻訳時のリボソームのような大掛かりな装置を必要とせず、それ単独で自発的に行われるという物理的観点からも興味深い。また、このメカニズムが真に理解できれば、望みの形を持った人工タンパク質の設計も可能となり、応用研究の可能性を大きく広げることができる。このように立体構造予測の研究は基礎的観点からも応用的観点からも極めて重要であるといえる。

この問題に対して 1960 年代に発表されたアンフィンゼンのドグマ以来、多くの第一原理的立体構造予測に関する研究が行われてきた。しかしその結果を一言でまとめれば、20 世紀の終わりまでは大きな進歩はなかったといえる。ところが 1997 年に発表された D.Baker らによるフラグメントアセンブリ法の登場により状況は一変した。具体的には、1998 年に行われたタンパク質の立体構造予測コンテスト (CASP3) の新規フォールド部門において、D.Baker らのフラグメントアセンブリ法を用いた予測によって、世界で初めて第一原理的な構造予測の成功例が報告されたのである (Proteins, suppl.3,171(1999))。

フラグメントアセンブリ法とは、まず予測したいタンパク質の配列を約 10 残基程度の局所配列に分割し、その局所配列にとって自由エネルギー的に安定な構造を PDB から複数予測し、その後、そのフラグメント構造の組み合わせ方を変えながら全体構造のエネルギーが安定な構造を探索しようとするものである。実はこのフラグメントアセンブリ法の成功の背後には、郷が提唱した「コンシステンシー原理」(Go, Adv.Biophys. 18,149, 1984)と密接な関係があるということを見出した (Chikenji, PNAS 2006)。コンシステンシー原理は「局所構造も非局所構造も無矛盾な構造が天然構造である」ということを主張するものであったが、フラグメントアセンブリ法は、PDB からとってきた局所構造は自由エネルギー的に安定な (無矛盾な) 構造である、という仮定のもとに、「局所的には無矛盾な構造を組みあせて全体構造を作ることによって、非局所的にも無矛盾な構造を探す」という手続きを具現化している方法論である。

このようにフラグメントアセンブリ法は予測能力と構築原理の両方の観点から興味深い方法論であるが、そのセンセーショナルな登場以降現在に至るまで、予測精度の観点からは本質的な進歩がない状態が続いている。例えば、最新の立体構造予測コンテスト CASP8 での新規フォールド部門の結果によると、予測に成功した例は極めて限定的 (9

問中 3 問しかなかった) であり、このような状況はコンテストの審査員によると 4 年前と比べても本質的な進歩は認められなかったと結論づけられている (Proteins, suppl.9,50(2009))。我々はこの状況を打破するために、フラグメントアセンブリ法の長所とその限界に関して丁寧に調べてきた。その結果、特に複雑なトポロジーを持つタンパク質に顕著であるのだが、PDB に存在するフラグメント構造で新規フォールドを表現することに限界があることを見出した。これはコンシステンシー原理が破綻しているという原理的な困難ではなく、方法論のテクニカルな問題である。以上から、第一原理的立体構造予測のブレークスルーのためには、フラグメントアセンブリ法の成功の本質である「コンシステンシー原理」の精神は保ちつつ、タンパク質の立体構造を効率よく表現できる新たな構造探索法が必要であるという結論に達した。

## 2. 研究の目的

本研究の目的は、第一原理的タンパク質立体構造予測の新規手法を開発し、構造予測能力の劇的な向上を目指すとともに、タンパク質の立体構造構築原理の理解を深めることにある。これを達成するための基本的アイデアは、現在最も強力な第一原理的予測手法として知られているフラグメントアセンブリ法の成功の本質である「郷によるコンシステンシー原理を具現化したアルゴリズムである」という側面は残しつつ、FA 法の問題点であるフラグメント構造の精度の問題を回避するために FA 法とは逆の考え方を使ったアルゴリズムを開発する、というものである。より具体的には、配列にそって非局所構造が安定な構造、すなわち、物理化学的に安定な二次構造パッキングを多数作成し、それらが配列にそって局所的にも物理学的に安定な構造でつなげることができるかどうかを評価する構造生成手法を開発する。

## 3. 研究の方法

フラグメントアセンブリ法とは異なるアプローチでコンシステンシー原理に基づく立体構造予測法の開発を行う。その具現化のために以下のような手法開発および解析を行う。

- ① 鎖のつながり方に依存しない構造アラインメント法を開発する。これは、安定な二次構造とは何か、といった問題や、二次構造パッキングの様式は有限なのか否かといった問題を明らかにするために必要である。
- ② ①で開発した鎖のつながり方に依存しない構造アラインメント手法を用いて、タンパク質立体構造データベースの網羅的解析を行い、安定な二次構造パッキングとは何か、また、それらの数は有限に限られているのか否かを、明らかにする。

- ③ 安定な二次構造パッキングを利用した、新規フォールドをもつタンパク質に対しても予測可能な、立体構造予測手法の開発を行う。具体的には、安定な二次構造パッキングに対して、網羅的に二次構造の順列変化を行い、パッキングとしては既知構造であるが、フォールドとしては新しい構造を多数作成する手法を開発する。この際、それらをエネルギー関数および、タンパク質鎖の局所的な性質として好ましい構造か否かを判定する手法を開発し、コンシステンシー原理を具現化した立体構造予測手法の開発を行う。作成後はベンチマークテストを行い、この手法の有効性と限界を批判的に検証する。

#### 4. 研究成果

##### 4. 1. 鎖のつながり方に依存しないタンパク立体構造比較手法の開発

安定な二次構造パッキングとは何か、といった問題を調べるためにはタンパク質鎖のつながりかたをあえて無視した立体構造比較手法を用いてデータベース解析する必要がある。我々は高速に、かつ正確な、鎖のつながり方に依存しない立体構造比較手法MICANを開発した。この手法は、タンパク質の二次構造を複数のベクトルで表現し、ベクトルで表現された二つのタンパク質構造を、画像認識の分野で開発された幾何学ハッシングを用いて検出するというものである。開発した手法の有効性を検証するために、構造バイオインフォマティクスの専門家が生物学的知識に基づき、手作業によって作成した構造アラインメントのデータセット、MALISAM, MALIDUP に対して、どれくらいMICANが正解アラインメントを出力できるかを計算し、既存方法と比較した。その結果、MICANは既存手法と比べると、正解アラインメントを出力する能力が高いことがわかった。また、MICANは鎖のつながり方に依存しない方法として開発されたが、少しの変更で、鎖のつながり方に依存した構造比較方法に変更することができた。既存のつながり方に依存した構造比較方法は膨大な数にのぼるが、そのなかでも、MALISAM, MALIDUPを再現できる能力としては、MICANは最も優れた部類のプログラムであることが示された。MICANの開発によって、生物学的、物理化学的に正しい構造比較が可能になったといえる。

##### 4. 2. 安定な二次構造パッキングパターンの解析

MICANを用いてタンパク質立体構造の網羅的比較を行った。まず、異なるフォールド同士は、どれくらい他のフォールドと同じ二次構造パッキングを有しているかを調べた。その結果、約80%のフォールド、他のフォールドと少なくとも一つは同じ二次構造パッ

キングを有していることがわかった。このことは、配列順序に依存しない構造をテンプレートとして利用することの有用性を示している。また、改良されたアラインメント手法を用いて、進化的関連性はないが構造類似性を示すタンパク質ペアの特徴を調べた。その結果、疎水親水のパターンも大きく破られており、タンパク質のサイズも大きく異なっていること等が確認された。このことから、二次構造のつながりかえによって新規フォールドを予測する際は、アミノ酸配列の類似性に頼ることはできず、構造と配列の適合性を評価する方法が必要であることがわかった。

##### 4. 2.

上記解析等から得られた見地に基づき、既知構造の二次構造を順列変異することによって新規構造を生成する手法を開発した。この際、素朴な順列変異では左巻きの $\beta$ - $\alpha$ - $\beta$ 構造や、ループクロスといった物理的に好ましくない構造が生成されてしまうという問題が判明したため、このような構造を正しく判定し、フィルタリングする手法もあわせて開発し、予測手法に実装した。具体的には、左巻き $\beta$ - $\alpha$ - $\beta$ の検出には、 $\beta$ シートから作られる面に対して $\alpha$ ヘリックスどちら向きにあるかを判定することによって判定する手法を開発し、ループクロス判定のアルゴリズムは、ループ内の残基から作られる三つのC $\alpha$ 原子から三角形をつくり、三角形と三角形の衝突判定によってループクロスを検出する方法を開発した。順列変異によって作成された構造は、拡張したコンタクトナンバーを用いてスレッディングにより構造-配列アラインメントを行い、その構造に対して、SimFold エネルギー関数で予測構造の品質を評価する、といった手法を開発した。いくつかのタンパク質で構造予測テストを行ったが、答えの構造のコンタクトオーダーが大きいものに関しては、フラグメントアセンブリ法よりもよい成績をおさめることができた。原点では二次構造の数が7つまでのものしか計算不可能であることが問題だが、フラグメントアセンブリ法と全く異なるアプローチによる立体構造予測法の確立に向けて確かな進歩を遂げることができた。

##### 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2件)

- ① Shintaro Minami, Kengo Sawada, George Chikenji, How a Spatial Arrangement of Secondary Structure Elements Is Dispersed in the Universe of Protein Folds, PLoS ONE, 査読有, Vol. 9, No.9, pp. 1-14, doi:

10.1371/journal.pone.0107959

- ② Shintaro Minami, Kenago Sawada, George Chikenji, MICAN:a protein structure alignment algorithm that can handle Multiple-chains, Inverse alignments, C  $\alpha$  only models, Alternative alignments, and Non-sequential alignments, BMC Bioinformatics, 査読有, Vol. 14, No. 24, 2013, pp.1-22, doi:10.1186/1471-2105-14-24

[学会発表] (計 33 件)

- ① Koya Kato, George Chikenji, A New Virtual Screening Method for Identifying Novel Active Compounds Based on Protein Promiscuity, CBI 学会、2014年10月28~30日、タワーホール船堀 (東京都・江戸川区)
- ② Koya Kato, George Chikenji, A Ligand Based Virtual Screening method that takes into account of protein-ligand interactions, 第52回日本生物物理学会年会、2014年9月26日、札幌コンベンションセンター (北海道・札幌市)
- ③ Shunsuke Nishiyama, Tatsuo Mukai, George Chikenji, De Novo protein structure modeling by rewiring old folds, 第52回日本生物物理学会年会、2014年9月27日、札幌コンベンションセンター (北海道・札幌市)
- ④ Kodai Takagi, Tatsuo Mukai, George Chikenji, emplate based modeling utilizing an order-made template library, 第52回日本生物物理学会年会、9月27日、札幌コンベンションセンター (北海道・札幌市)
- ⑤ Takahiro Kanemitsu, Shintaro Minami, George Chikenji, Homologous protein pairs that share the same core packing but have different topology, 第52回日本生物物理学会年会、2014年9月27日、札幌コンベンションセンター (北海道・札幌市)
- ⑥ Tatsuo Mukai, George Chikenji, An algorithm for identifying loop crossing in protein structures, 第52回日本生物物理学会年会、2014年9月27日、札幌コンベンションセンター (北海道・札幌市)
- ⑦ Shintaro Minami, Motonori, George Chikenji, Non-sequential structural similarity in the protein world, 第52回日本生物物理学会年会、2014年9月27日、札幌コンベンションセンター (北海道・札幌市)
- ⑧ Determination of Free Energy Surface by Go model-optimized Multicanonical Molecular Dynamics Simulation, Mashiho Ito, George Chikenji, Shoji Takada, 第14回日本蛋白質科学会年会、2014年6月25日、ワークピア横浜/横浜産貿ホール (神奈川県・横浜市)
- ⑨ How does a protein fold share the same secondary structure packing arrangement with other folds?, Shintaro Minami, George Chikenji, 第14回日本蛋白質科学会年会、2014年6月27日、ワークピア横浜/横浜産貿ホール (神奈川県・横浜市)
- ⑩ Development of Ligand-Based Virtual Screening for Discovering Drug Candidate Compound with Novel Scaffold, Koya Kato, George Chikenji, 第14回日本蛋白質科学会年会、2014年6月27日、ワークピア横浜/横浜産貿ホール (神奈川県・横浜市)
- ⑪ An algorithm for identifying loop crossing in protein structures using triangle-triangle intersection, Tatsuo Mukai, George Chikenji, 第14回日本蛋白質科学会年会、2014年6月27日、ワークピア横浜/横浜産貿ホール (神奈川県・横浜市)
- ⑫ A new homology modeling technique that utilizes the knowledge of completeness of the PDB, Takahiro Kanemitsu, Shintaro Minami, George Chikenji, 第51回日本生物物理学会年会、2013年10月28日、国立京都国際会館 (京都府・京都市)
- ⑬ Development of Ligand Based Virtual Screening considering protein-ligand interaction, Koya Kato, George Chikenji, 第51回日本生物物理学会年会、2013年10月29日、国立京都国際会館 (京都府・京都市)
- ⑭ Some features of protein pairs which have same SSEs packing arrangement but have different topology, Tatsuo Mukai, Shintaro Minami, George Chikenji, 第51回日本生物物理学会年会、2013年10月29日、国立京都国際会館 (京都府・京都市)

- ⑮ Generating novel protein folds from existing folds for de novo protein structure prediction, Yuki Nakagawa, George Chikenji, Shintaro Minami, 第 51 回日本生物物理学会年会、2013 年 10 月 29 日、国立京都国際会館 (京都府・京都市)
- ⑯ Non-sequential structural alignment reveals fold change by segment shuffling during evolution, Shintaro Minami, George Chikenji, Motonori Ota, 第 51 回日本生物物理学会年会、2013 年 10 月 29 日、国立京都国際会館 (京都府・京都市)
- ⑰ Concurrent use of protein structure and sequence information toward remote homology detection, Shintaro Minami, George Chikenji, 第 13 回日本蛋白質科学会年会、2013 年 6 月 13 日、とりぎん文化会館 (鳥取県・鳥取市)
- ⑱ Understanding principle of the protein structure world by non-sequential protein structure comparison, Shintaro Minami, Kengo Sawada, George Chikenji, 第 13 回日本蛋白質科学会年会、2013 年 6 月 14 日、とりぎん文化会館 (鳥取県・鳥取市)
- ⑲ Development of a de novo protein structure prediction method based on the antecedent domain segment (ADS) hypothesis, Kota Nakamori, Shintaro Minami, Kengo Sawada, George Chikenji, 第 51 回日本生物物理学会年会、2012 年 9 月 22 日～24 日、名古屋大学 東山キャンパス (愛知県・名古屋市)
- ⑳ Modeling structure of large insertion regions in homology modeling by utilizing non-sequential structure relationship, Yuki Nakagawa, Shintaro Minami, Kengo Sawada, George Chikenji, 第 51 回日本生物物理学会年会、2012 年 9 月 22 日～24 日、名古屋大学 東山キャンパス (愛知県・名古屋市)
- 21 Common structural / sequence features of proteins that share non-sequential structural similarity, Kengo Sawada, Shintaro Minami, George Chikenji, 第 51 回日本生物物理学会年会、2012 年 9 月 22 日～24 日、名古屋大学 東山キャンパス (愛知県・名古屋市)
- 22 Some characteristic features of non-sequential structural alignment: difference between sequential and non-sequential alignments, Tatsuo Mukai, Kengo Sawada, Shintaro Minami, George Chikenji, 第 51 回日本生物物理学会年会、2012 年 9 月 22 日～24 日、名古屋大学 東山キャンパス (愛知県・名古屋市)
- 23 Homologous proteins with different fold: how were they raised? Shintaro Minami, Kengo Sawada, George Chikenji, 第 51 回日本生物物理学会年会、2012 年 9 月 22 日～24 日、名古屋大学 東山キャンパス (愛知県・名古屋市)
- 24 Toward sequence order-independent sequence-structure alignment, Kengo Sawada, Shintaro Minami, George Chikenji, 第 12 回日本蛋白質科学会年会、2012 年 6 月 20 日～6 月 22 日、名古屋国際会議場 (愛知県・名古屋市)
- 25 A crowded view of protein structure space by focusing on SSE arrangement, Shintaro Minami, Kengo Sawada, George Chikenji, 第 12 回日本蛋白質科学会年会、2012 年 6 月 20 日～6 月 22 日、名古屋国際会議場 (愛知県・名古屋市)
- 26 Development of a non-sequential structure alignment algorithm consistent with intuition of human experts of structural bioinformatics, Shintaro Minami, Kengo Sawada, George Chikenji, 第 12 回日本蛋白質科学会年会、2012 年 6 月 20 日～6 月 22 日、名古屋国際会議場 (愛知県・名古屋市)
- 27 Development of a novel protein structure prediction method emphasizing non-local interaction, Kota Nakamori, Shintaro Minami, Kengo Sawada, George Chikenji, 第 50 回日本生物物理学会年会、2011 年 9 月 16 日 18 日、兵庫県立大学・姫路書写キャンパス (兵庫県・姫路市)
- 28 Simulated evolution of protein by a selective pressure on function, Masahiro Yamamoto, Takeshi N. Sasaki, Kengo Sawada, Shintaro Minami, George Chikenji, Tomoki P. Terada, Masaki Sasai, 第 50 回日本生物物理学会年会、2011 年 9 月 16 日 18 日、兵庫県立大学・姫路書写キャンパス (兵庫県・姫路市)
- 29 A related location of exon boundary and structural alignment boundary reveals a mechanism of protein fold diversity, Shintaro Minami, Kengo Sawada, George Chikenji, 第 50 回日本生物物理学会年

会、2011年9月16日18日、兵庫県立大学・姫路書写キャンパス（兵庫県・姫路市）

- 30 Structure-sequence alignment between distantly-related proteins using novel gap penalty, Kengo Sawada, Shintaro Minami, George Chikenji, 第11回日本蛋白質科学会年会、2011年6月7日～6月9日、ホテル阪急エキスポパーク（大阪市・吹田市）
- 31 A mechanism of structural diversity of proteins, Shintaro Minami, Kengo Sawada, George Chikenji, 第11回日本蛋白質科学会年会、2011年6月7日～6月9日、ホテル阪急エキスポパーク（大阪市・吹田市）
- 32 Ligand-based Virtual Screening considering conformational fluctuations and interaction patterns, Tatsuya Okuno, George Chikenji, 第11回日本蛋白質科学会年会、2011年6月7日～6月9日、ホテル阪急エキスポパーク（大阪市・吹田市）
- 33 The evolution of protein folds and De Novo protein structure prediction, Shintaro Minami, Kota Nakamori, Kengo Sawada, George Chikenji, 第11回日本蛋白質科学会年会、2011年6月7日～6月9日、ホテル阪急エキスポパーク（大阪市・吹田市）

〔その他〕

ホームページ等

<http://www.tbp.cse.nagoya-u.ac.jp/MICAN>

## 6. 研究組織

### (1) 研究代表者

千見寺浄慈 (CHIKENJI, George)

名古屋大学・大学院工学研究科・助教

研究者番号：10420366