

平成 22 年 5 月 1 日現在

研究種目： 若手研究(B)
 研究期間： 2008 ～ 2009
 課題番号： 20770120
 研究課題名(和文) 局所配列の物理化学に着目したタンパク質の立体構造予測の研究
 研究課題名(英文) A protein structure prediction study focusing on physico-chemical properties of local sequences.
 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

研究成果の概要(和文)：タンパク質の立体構造構築原理を理解するために、また、構造予測の精度向上のために、タンパク質の局所構造に着目した研究を行った。特に、タンパク質の局所構造を表現する新しいモデルとしてゴーストウォーターモデルという陰溶媒モデルを導入し、そのモデルの開発と改良、構造予測のベンチマークテストを行った。大規模ベンチマークの結果、タンパク質の局所構造はそれ単独で取り出しても比較的安定であることが多いことがわかった。

研究成果の概要(英文)：In order to understand principle of protein structure and improve protein structure prediction accuracy, we have developed a De Novo protein structure prediction method focusing on physico-chemical properties of local sequence. Especially, a new implicit solvent model called "ghost water model" has been developed for the purpose of evaluating energies of local structures. We performed large scale structure prediction benchmark test and found that local structures of the native are generally stable even if these are excised from a entire structure.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	2,800,000	840,000	3,640,000
2009 年度	600,000	180,000	780,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：生物学

科研費の分科・細目：生物科学・生物物理学

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、水素結合、フォールディング、アロステリック

1. 研究開始当初の背景

1960 年代に Anfinsen が「タンパク質の天然構造は自由エネルギー最小の状態である」

ということを実験的に示して以来、アミノ酸配列の情報のみからタンパク質の立体構造を理論的に予測するという、いわゆる「タン

タンパク質の折り畳み問題」は多くの研究者を惹きつけてきた。この問題の歴史は古く、かつ極めて多くの研究者が挑戦してきたにもかかわらず、ほとんど進展がみられないまま 20 世紀を終わろうとしていた。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 (J.Mol.Biol.,268,209 (1997))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 (CASP4) 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,45,98(2002))。ここで、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、その局所配列のとりうると思われる構造 (フラグメント構造) を配列検索によって複数用意する。その後、そのフラグメント構造を組み合わせることによって全体構造を構築しようとするものである。その後 D.Baker や我々のグループをはじめとする多くのグループがフラグメントアセンブリ法の改良などに取り組み、予測精度を上げる努力がなされてきた。我々は「なぜフラグメントアセンブリ法がうまくいくのか？」という根本的な問題を問い、それを理解することによって立体構造予測の精度向上とともに、立体構造構築原理を理解しようとする研究を行ってきた。

1998 年以降の立体構造予測コンテスト (CASP) の結果を眺めると、フラグメントアセンブリ法を用いると、確かに新規フォールドを予測できる場合もあるが、必ずしも全ての問題に対して成功するわけではなく、むしろ成功例の方が少ない、ということがわかってきた。我々はどのような場合に立体構造予測が成功し、どのような場合に失敗するのかを丁寧に調べてきた。その結果、フラグメントアセンブリ法が成功するのは、予測したいタンパク質のフラグメント構造が正しく用意できた場合に限られており、また、(特に新規フォールドでは顕著に) 配列検索では正しいフラグメント構造を用意できないことが明らかとなった。逆に、正しいフラグメント構造を用意できれば極めて高い確率で立体構造を予測することも明らかとなった。(PNAS,103, 3141 (2006)) これらの事実から、タンパク質の局所構造の物理化学的性質を理解し、物理化学的手法によってフラグメント構造を合理的に予測できるようにすることが、タンパク質立体構造予測にとって最重要課題であるということが明らかとなった。

2. 研究の目的

本研究の目的はタンパク質研究の中心的課題であるアミノ酸配列という 1 次元情報から立体構造という 3 次元情報への『情報の変換過程の一般論』を物理化学的な観点から明らかにすることである。具体的には、実験やシミュレーションなどから学んだタンパク質のフォールディング過程を可能な限り人間にとってわかりやすい概念としてアルゴリズム化し、それを立体構造予測プログラムとして実装し、精度よく立体構造予測できるようにする、ということである。数値的な目標としては 150 残基以下の任意のタンパク質について、天然構造との RMSD が 3 Å 以下の精度で予測できるようになることを目指す。

特に本研究課題においては、以下の点に重点をおく事で上記目的を達成しようとする。現時点で最も有効な立体構造予測方法はフラグメントアセンブリ法であるが、この方法の問題点のひとつとして、フラグメント構造の選択に配列検索などの経験的方法を使っているという点が挙げられる。この方法論では、特に新規フォールドを予測する際、正解構造に近いフラグメント構造を用意することが困難な場合が多く、それが原因で結果的に全体構造の予測に失敗する、ということを我々は明らかにしてきた (PNAS,103,3141,(2006))。また、配列検索に頼っているために、情報の変換過程の物理化学的理解にはつながらない、という決定的な欠点がある。そこで本研究では全原子モデルを用いて局所配列のシミュレーションを行い、天然構造になりうる局所構造の条件とは何か?といった問題や、局所配列に刻み込まれている 3 次元情報をどうやって読み取るか?といった問題を常に意識しながら、物理化学的にフラグメント構造を予測する手法を開発する。それを用いてフラグメントアセンブリを行い、高精度の立体構造予測法の確立と同時に情報の変換過程の一般論の構築を目指す。

3. 研究の方法

典型的なタンパク質の折れ畳み時間のスケールはミリ秒から秒の時間スケールであるが、通常的全原子レベルの分子動力学計算で追跡できるのはマイクロ秒の程度である。さらに、仮に全原子レベルで十分に長い時間スケールのシミュレーションが可能であり、それによって立体構造予測ができるとしても、アミノ酸配列という 1 次元情報から立体構造という 3 次元情報への「情報の変換過程」を理解したことにはならない。本研究では、タンパク質フォールディングの研究等から得られた描像をもとに、大胆な粗視化モデルと折れ畳みアルゴリズムを開発し、それを用

いてタンパク質立体構造を構築することによって、タンパク質の立体構造構築原理を理解することを目指す。具体的にはフラグメントアセンブリ法におけるフラグメント構造予測をゴーストウォーターモデルと繰り込み的手法という新しい手法を開発することによって行った。

(1) 新しい溶媒モデルの開発 (ゴーストウォーターモデル)

本研究では10残基程度の局所構造予測に全原子モデルによるシミュレーションを採用する。ただし、溶媒分子は陽には含めない新しいモデルを開発する。水溶媒の重要な効果として1つ挙げられるのは水素結合である。タンパク質の立体構造データベースを丁寧に調査すると、タンパク質の原子団の中で極性をもったものはほとんど100%、タンパク質に属する他の極性原子、もしくは溶媒分子のいずれかと水素結合していることがわかる。この観察を原理に、タンパク質内の極性原子はタンパク質内部の他の極性原子、もしくは溶媒分子と常に水素結合していなければならないようなモデルを構築する。これを実現するための新しい手法としてゴーストウォーターを導入する。ゴーストウォーターとは、タンパク質の極性原子と常に水素結合している水分子であるが、そのタンパク質極性原子が他のタンパク質極性原子と水素結合できる状態になると消える、というものである。本研究ではまずこのモデルの開発、改良と大規模なベンチマークテストを行いこの手法の妥当性を批判的に検証する。上記のモデルを用いて10残基程度の短いフラグメント構造の予測を行うが、ここでの予測とは、10残基程度のペプチドの構造をユニークに決めることではなく、むしろ天然構造になりうる構造候補を複数用意することである。開発した手法の大規模ベンチマークを行うことによって検証する。

(2) CASP8 への参加

開発された方法を基に第8回タンパク質立体構造予測コンテスト(CASP8)に参加し、新しい手法が国際的にどの水準にあるのかを確認すると同時に、改善点を徹底的に洗い出す作業を行う。

(3) 様々な観点からのベンチマーク

Protein GのC末端の β ターン、src SH3のdiverging turnなど、その一部分を切り出してきても、それ単独で天然構造と同じ構造をとることが実験的にわかっているペプチド配列を対象にして、我々の手法でもそれらの配列が最低エネルギー状態として認識できるようなレベルになることを目指す。さらに、予測した局所構造群が局所配列のとりうる

構造空間のリーズナブルな代表構造になっていることを確認するため、予測された局所構造を用いてフラグメントアセンブリを行い、多数の構造予測ベンチマークを行うのはもちろんのこと、自由エネルギー曲面の解析からフォールディング経路の計算を行い、どれくらい ϕ 値解析や重水素交換などの実験データを再現できるか、を検証することにより手法の妥当性を批判的に検討する。さらに、アロステリック反応など、機能発現時に大きな構造変化を伴うものの記述も試みる。

4. 研究成果

現時点で最も有力なタンパク質立体構造予測法として知られているフラグメントアセンブリ法に関する基礎的、応用的の両方の観点から研究を行った。

基礎的観点からの研究としては10残基程度のフラグメント構造を物理化学的に予測するための新しい溶媒モデル—ゴーストウォーターモデル—を構築した。このモデルでは、タンパク質内部の原子の極性原子は90%以上水素結合をしている、という観察をもとに、タンパク質内部の極性原子は可能なかぎり水素結合をしている、という効果を直接反映しつつ、計算コストを抑えるために陽に多数の溶媒分子を含まないように工夫したモデルである。具体的には、タンパク質内部の極性原子から理想的な配置にいつも仮想的な水分子(ゴーストウォーター)を配置して、タンパク質極性原子は常にこの水分子と水素結合を形成しているが、タンパク質の極性分子同士が水素結合を形成できるような場合はこのゴーストウォーターは消える、というモデルである。このゴーストウォーターを全原子モデルの一つであるCHARMM19に加える形で実装をした。パラメータの調整はPark & Levittのデコイセットを使って、天然構造が安定になるように行った。

開発したモデルの性能を評価するために、天然構造から抜き出した10~20残基程度の長さのフラグメント構造のエネルギーの安定性評価を行った。天然構造の安定性評価のために構造データベースからフラグメント構造を抽出し、それをクラスタリングすることによってフラグメントの構造空間(デコイセット)を表現した。多数のタンパク質の局所構造とデコイセットのエネルギーを計算し、天然構造の安定性をZ-scoreで評価した。その結果、天然構造は平均的にはZ-scoreで-1、すなわち、平均値から標準偏差一つ分安定であるという結果を得た。これは、タンパク質の部分構造は、それ単独で存在したとしても、ある程度安定であるものが多いことを示している。また、ゴーストウォーターの効果を確認するためにゴーストウ

ウォーターを考慮した場合としない場合を比較したところ、ゴーストウォーターの導入によって著しく天然構造が安定化していることがわかった。このことは、タンパク質の一部を切り出してきたフラグメント構造の安定性を評価する際に、ゴーストウォーターを導入することによってうまくフラグメント構造の環境を表現できていることを示唆している。

この方法のさらなるテストとして、2008年に行われたタンパク質の立体構造予測コンテスト CASP8 に参加した。ここでは我々は主に2つの点に注目してテストを行った。一つ目は、ゴーストウォーターモデルを用いたフラグメント構造予測を行い、それを用いてフラグメントアセンブリ法を行った場合のパフォーマンスである。二つ目は様々な構造の中に対して、ゴーストウォーターモデルを用いてエネルギー計算を行い安定性を評価することによって天然構造に近い構造を選べるか? という問題である。1つ目の着目点、すなわち、フラグメント構造をゴーストウォーターモデルでシミュレーションすることによって予測し、さらにその構造をもとに不フラグメントアセンブリを行い、全体構造を予測する、というテストであるが、残念ながらあまり良好な予測結果は得られなかった。CASP8 終了後にこの問題点を洗い出したところ、たしかに、我々のベンチマークどおりに天然構造に近い構造はゴーストウォーターモデルで比較的安定なエネルギーを持っているものが多かったが、それよりも安定な非天然様構造の数が多かったこと、また、数は少ないながらも天然構造の、特にループ領域で、極めてエネルギー的に不安定な構造が存在し、そのような領域に対してフラグメント予測が完全に失敗していることがわかった。これはフラグメントアセンブリ法の限界を示しており、新しい方向性を考えるきっかけを与えてくれていると思われる。2番目の着目点である、ゴーストウォーターモデルを用いた構造安定性の評価によって天然構造に近い構造を選べるか? という問題であるが、これは極めて良好な結果が得られた。CASP においては他のグループ (サーバーグループ) の予測構造を利用できるタイミングがあるのだが、我々はゴーストウォーターモデルを用いてこのサーバー構造群に対してエネルギー計算を行い、エネルギーの値でモデル選択を行った。いくつかのターゲット (高精度ホモロジーモデリングのターゲット) に対して実際にこの方法で我々の予測構造を提出したが、全参加グループの中で、トップ5に入った構造がいくつかあった。中でも特筆すべき例として T0498 と T0499 という2つのターゲットが挙げられる。この2つの問題は共に 56 残基からなるタンパク質でお互いの配

列の一致度は 98%にものぼる。したがって、ほぼ同じ構造をとると考えられるが、T0498 の答えの構造は3ヘリックスバンドルで、T0499 は α/β 型の構造であった。ほとんどのグループが両方とも α/β の予測を提出していたが、我々のゴーストウォーターモデルを用いた計算の予測では、両方のターゲット共に正しいトポロジーの構造を予測することができた。このことは全体構造の安定性を評価する上でも、ゴーストウォーターモデルは優れた方法であり、水素結合の重要性を示唆するものである。

CASP 終了後に今までの問題点を洗い出したところ、フラグメントアセンブリ法は確かに強力な方法ではあるが、現状のフラグメントアセンブリ法のままでは明らかな限界があることも見えてきた。フラグメントアセンブリ法ではローカルな構造の重要性ばかりが強調されてきた。確かにそれが重要である場合が多いことは明らかであるが、必ずしも全ての局所相互作用が重要であるわけではないことも浮き彫りになってきた。このことから、郷によるコンシステンシー原理、すなわち、「タンパク質の天然構造は局所的にも非局所的にも無矛盾な構造である」という主張をより正しく反映した立体構造予測法を新たに開発すべきであるという結論に達した。すなわち、フラグメントアセンブリ法の長所は残しつつも、安定な (疎水性相互作用などの) 非局所相互作用が見つかった場合は、先にその非局所的相互作用を形成して、それにコンシステントな局所構造を作る、という考え方である。実際にこのような考え方の予測方法を実装し、予備的な計算を行ったところ、特に β シートタンパク質では既存の方法よりも優れた結果を出す傾向が見られた。このように、本研究を通して既存の方法の問題点と、更なる予測精度向上と立体構造構築原理へ向けての新しい方向性が見えてきた。

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1件)

- ① Naoto Hori, George Chikenji, R. Stephen Berry, and Shoji Takada, Folding energy landscape and network dynamics of small globular proteins, Proceedings of the National Academy of Sciences USA, vol. 106, 73-78, (2009), 査読有

[学会発表] (計9件)

- ① 南慎太郎, 千見寺浄慈, 新規フォールドに対

- する立体構造予測に向けて一フラグメントアセンブリ法による予測精度のライブラリ依存性一、生物物理学会、2009年11月1日、アステイトくしま
- ② 澤田賢吾、千見寺浄慈、フラグメントアセンブリ法による大規模蛋白質リファインメント、生物物理学会、2009年11月1日、アステイトくしま
- ③ 澤藤俊平、千見寺浄慈、“ゴーストウォーター”を含む全原子モデルによる蛋白質構造予測、生物物理学会、2009年11月1日、アステイトくしま
- ④ **George Chikenji**, **Template-based modeling and free-modeling by fragment assembly with SimFold energy function**、Asian Workshop for Protein Structure Prediction Methodology, 2009.09.18, Korea Institute for Advanced Study, Seoul, Korea
- ⑤ 千見寺浄慈、タンパク質立体構造予測の現状とフリーモデリング的のテンプレートモデリングの試み、蛋白研セミナー、2009年7月31日、大阪大学蛋白質研究所
- ⑥ 澤藤俊平、千見寺浄慈、仮想水分子を含んだ水素結合ポテンシャルを用いたタンパク質構造の安定性評価、蛋白質科学会、2009年5月22日、熊本全日空ホテルニュースカイ
- ⑦ 澤田賢吾、千見寺浄慈、フラグメントアセンブリによるテンプレートを用いた立体構造予測、蛋白質科学会、2009年5月22日、熊本全日空ホテルニュースカイ
- ⑧ 南慎太郎、千見寺浄慈、カルモジュリンの構造変化：構造予測の手法およびMDシミュレーションを相補的に用いた研究、蛋白質科学会、2009年5月21日、熊本全日空ホテルニュースカイ
- ⑨ 澤田賢吾、千見寺浄慈、**Template based protein structure prediction by Fragment assembly in CASP8**、第9回日本バイオインフォマティクス学会創薬インフォマティクス研究会「CASP8 タンパク質立体構造予測の最前線」、2009年1月19日、東京大学医科学研究所

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況 (計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]
なし

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