

蛋白質のアミノ酸配列から構造そして機能の
予測へ向けて; 配列と構造間のアライメント

(課題番号 12680651)

2000年度～2001年度 科学研究費補助金 (基盤研究 (C)(2))
研究成果報告書

2002年3月

研究代表者 宮澤 三造

(群馬大学工学部助教授)

目 次

はしがき	1-4
Identifying sequence-structure pairs undetected by sequence alignments	
Miyazawa, S. and Jernigan, R. L.	5-21
<i>Protein Engineering</i> , 13 , 459-475, 2000.	
Errata for "Identifying sequence-structure pairs undetected by sequence alignments"	22-25
Protein sequence-structure alignment based on site-alignment probabilities	
Miyazawa, S.	26-35
<i>Genome Informatics</i> , 11 , 141-150, 2000.	

はしがき

ヒトゲノムもその DNA 配列の 90% は解明された。ポストゲノムは、DNA/ 蛋白質の配列情報から、機能情報を如何にして読み取るかが主題である。機能が既知の配列との配列比較では類似配列が検出されない蛋白質の機能の予測が求められている。蛋白質の機能はその立体構造と不可分であり、構造に関する知見を得ることは機能予測に継る。また構造は配列より保存され易いという利点を持つ。

しかし、アミノ酸配列から配列固有の蛋白質構造を予測する方法はまだ確立していない。蛋白質のコンフォメーションの記述には多数の座標変数が必要とされ、またそのエネルギー面も複雑である。現在、平均的な蛋白質の分子動力学計算で記述できる時間スケールはたかだかナノ秒のオーダーである。一方、蛋白質の折り畳みはミリ秒から秒のオーダーの現象である。第一原理からの構造予測が困難である所以である。幸いにも、これまで解析された 13000 を越える多数の蛋白質ドメイン構造の解析を通じ、自然界にある蛋白質の(骨格)構造には、(多分進化の産物として)、限られた数(650 から 7900 程度)のプロトタイプがあることが判明した。すなわち十分な数のプロトタイプの構造が判明すれば、ある配列に対しどのプロトタイプの骨格構造が最適であるかを判定することにより骨格構造の予測が可能となる。そこで、配列からの構造予測に向けて、プロトタイプの構造を全て 10 年以内に解析しようとする計画が日本及び米国で発足した。解析せねばならないプロトタイプの数、ある配列に対しどの骨格構造が最適であるかの判定の精度に依存する。ある配列に対しどの骨格構造が最適であるかを判定するには、配列と骨格構造の関数であるフリーエネルギーポテンシャルが必要となる。しかし、通常半経験的原子間ポテンシャルで溶媒効果を考慮した骨格構造のフリーエネルギーの計算はデータベース検索の目的には計算時間がかかり過ぎる。配列と構造との適合度を評価するデータベース検索に適した高速な方法が求められている。

これまでに我々は、配列と構造間での適合性を計るメジャーとして利用可能な残基レベルでの経験的ポテンシャルとして、アミノ酸残基間の 2 体の接触ポテンシャル (Macromolecule, **18**, 534-552, 1985; J. Mol. Biol., **256**, 623-644, 1996; Proteins, **34**, 49-68, 1999) および残基の 2 次構造ポテンシャル (Proteins, **36**, 347-356, 1999) を、既知の蛋白質立体構造におけるアミノ酸残基の隣接頻度分布と 2 次構造の頻度分布から評価した。残基の挿入欠失を許さない条件の下での配列の構造へのはめ込み法では、これらのポテンシャルは配列と構造の適合性を適切に評価することが確かめられた。(J. Mol. Biol., **256**, 623-644, 1996; Proteins, **36**, 357-369, 1999)。この経験

的ポテンシャルを用い、残基の挿入欠失を許す配列と構造間のアライメント作成のアルゴリズムを開発し、配列と構造の適合性の同定に関する簡便で高速な方法を確立することがこの研究の目的である。

配列アライメントとは異なり、この評価関数は2体ポテンシャルを含むため、エネルギー最小のアライメントは動的計画法で計算することができずNP 困難の問題となる。我々は、2体ポテンシャルを平均場近似により残基毎のアライメント (対合) 確率から評価し、繰り返し法により self-consistent な残基毎の対合確率を転移行列法により計算した。そのようにして評価した2体ポテンシャルの下、エネルギー最小のアライメントと対合確率の順に残基を対合する確率アライメントの2つの方法によりアライメントを計算した。残基の挿入欠失のためのペナルティは、立体構造上内側の座位程挿入欠失が生じにくいよう設定した。蛋白質構造はそのコア部分ほど類似性が高く、類似性検出の精度を上げるためコア構造のデータベースを構築するような方法もあるが、我々の方法では類似性の強弱を対合確率として予測する。統計力学的に可能な全てのアライメントを考慮し、残基の対合に関する信頼度 (残基毎のアライメント確率) を提供するので、信頼度に基づき段階的に骨格構造をリファインメントできるという利点を持つ。

SCOP 蛋白質構造分類データベースに含まれる類似蛋白質対と非類似蛋白質対を用い類似性検出能力を評価したところ、平均的には配列比較と同程度であるが、配列比較では同定できない類似蛋白質対を多数検出できることから、配列比較と相補的に使用できることが明らかになった。また、確率アライメントにおいて、高い対合確率を持つ残基対はより類似した構造部分に対応することが示された。このような特性を持つ確率アライメントは、エネルギー最小 / 適合度最大のアライメントに比べ、アライメントの信頼度 (対合確率) を明示するので、ホモロジー構造予測において特に有用である。一方、1次配列上で隔たった残基間の相互作用はフォールドの認識に、2次構造ポテンシャルは正確な残基対合を得る上で欠かせないことが示唆された。

配列比較では同定できないような遠縁の蛋白質間の類似性検出に関し、既知の立体構造データにおけるアミノ酸残基の統計分布から評価した経験的相互作用ポテンシャルの配列・構造間適合度評価関数としての有効性が確認された。類似性 (同一残基) が20% 以下の蛋白質間であっても、配列と構造の比較からその類似性が検出できる経験的ポテンシャルとアライメント作成法は、骨格構造と側鎖構造のリファインメントのた

めの初期骨格構造を提供し、配列からの構造予測において有用であろう。

2002年3月

研究代表者： 宮澤 三造 (群馬大学工学部助教授)

交付決定額 (配分額)

(金額単位: 千円)

	直接経費	間接経費	合計
2000年度	1600	0	1600
2001年度	500	0	500
総計	2100	0	2100

研究発表

(1) 学会誌等

1. Miyazawa, S. and Jernigan, R. L.: Identifying sequence–structure pairs undetected by sequence alignments. *Protein Engineering*, **13**, 459–475, 2000.
2. Miyazawa, S.: Protein sequence–structure alignment based on site–alignment probabilities. *Genome Informatics*, **11**, 141–150, 2000.

(2) 口頭発表

1. Miyazawa, S.: An energy potential and alignment method for identifying protein sequence–structure compatibilities. International Conference on Structural Genomics 2000, Yokohama, Japan, November 2–5, 2000.
2. Miyazawa, S.: Protein sequence–structure alignment based on site–alignment probabilities. GIW 2000 (The Eleventh Workshop on Genome Informatics), Tokyo, December 18–19, 2000.