

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Development of a protein sequence alignment method for accurate homology modeling
著者(和文)	牧垣秀一朗
Author(English)	Shuichiro Makigaki
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11681号, 授与年月日:2020年12月31日, 学位の種別:課程博士, 審査員:石田 貢士,秋山 泰,岡崎 直観,村田 剛志,関嶋 政和
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11681号, Conferred date:2020/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	牧垣 秀一朗	
論文審査 審査員		氏名	職名	氏名	職名
	主査	石田 貴士	准教授	関嶋 政和	准教授
	審査員	秋山 泰	教授		
		岡崎 直観	教授		
村田 剛志		教授			

論文審査の要旨 (2000 字程度)

本論文は「Development of a protein sequence alignment method for accurate homology modeling」と題し、英文7章より構成されている。

第1章「Introduction」では、本研究の背景と目的について述べている。生体高分子であるタンパク質の立体構造の予測技術の中で、ホモロジーモデリング法は現在最も実用的であり、本研究ではその予測精度を改善することを目的とすると述べている。ホモロジーモデリング法では、予測対象と配列相同性のある別のタンパク質の立体構造情報をテンプレートとして用いることで予測を行うが、そこで用いられるテンプレートタンパク質と、予測対象タンパク質との配列アライメントが低品質であると、その立体構造予測精度が低下してしまう問題があることを指摘している。そこで、本研究では教師あり機械学習や中間的な相同配列を利用して、より良質の配列アライメントを生成する技術を開発することで、立体構造予測精度を改善することを目指すとして述べている。

第2章「Homology modeling」では、まずテンプレートタンパク質検索を行い、得られたテンプレートタンパク質と予測対象タンパク質との配列アライメントを生成し、テンプレートタンパク質の立体構造情報を利用して予測構造のモデリングを行う、というホモロジーモデリング法の流れを概説し、それぞれのプロセスについて既存の技術を紹介している。また、過去の研究の問題点として、多くの研究がテンプレートタンパク質検索の検索感度の向上に注力しており、ホモロジーモデリング法において、配列アライメントの質が予測精度に大きく影響を与えるにも関わらず、配列アライメントの品質の向上については研究が不足している点を指摘している。

第3章「Sequence alignment generation by substitution score prediction using machine learning」では、動的計画法による配列アライメント生成の計算において、一般的に用いられるアミノ酸置換スコア行列に換えて、各アミノ酸位置に対して機械学習モデルによって推定された一致スコアを用いる手法を提案している。提案手法では、一致を判定するアミノ酸位置の、周辺の情報を含めた配列プロフィールを入力とし、立体構造既知のタンパク質間での構造アライメントのパス上のアミノ酸位置を教師あり機械学習の正例、それ以外のアミノ酸位置とギャップを負例として予測モデルを訓練することで、ホモロジーモデリング法において理想的である構造アライメントに近い配列アライメントを生成する一致スコアを推定している。評価実験では、生成された配列アライメントを用いて、実際に立体構造の予測を行い、正解である天然構造と比較することで手法の性能を評価している。提案手法に基づく予測は、テンプレートタンパク質検索で最も良い性能を持つ手法の一つである HHsearch による配列アライメントを用いた場合と比較して、より良い立体構造予測精度を示し、アミノ酸置換スコア行列と Smith-Waterman 法を使用した場合に比べても、天然構造と類似した領域の割合を示す TM-score が向上し、良い立体構造モデルが予測可能であることを示している。また、提案手法の機械学習モデルにおける精度の向上の要因について解析を行い、配列プロフィール上のアミノ酸出現頻度における多様性とその判定に影響を与えていることを示唆している。

第4章「Sequence alignment generation for protein remote homologs」では、第3章で述べた手法では精度が不十分であった遠縁のテンプレートタンパク質との配列アライメントについて、intermediate sequence search で得られる中間的な相同配列の情報を用いることで、より高品質な配列アライメントを生成する手法を提案している。Intermediate sequence search は Park らによって提案された高感度なテンプレートタンパク質検索手法で、立体構造情報が決定されていないタンパク質を含む、より大規模なタンパク質配列データベースに対して複数の配列相同性検索を行い、そこで得られた予測対象とテンプレートタンパク質と双方に類似した中間的な相同配列を仲介として遠縁のテンプレートタンパク質検索を行うが、その際に立体構造予測に必要な配列アライメントが生成されない問題を指摘している。提案手法では、予測対象タンパク質配列、テンプレートタンパク質配列、中

間的な相同配列のそれぞれの間のペアワイズ配列アライメントを統合し、立体構造予測に利用可能な高品質な配列アライメントを生成している。評価実験を行い、提案手法は、遠縁のテンプレートタンパク質に対して、中間的な相同配列を利用しない手法に比べ、配列アライメントの品質を向上し、タンパク質立体構造予測の精度を改善することを示している。また、得られた中間的な相同配列のうちどの配列アライメントを利用するかを組み合わせても検討しており、可能な全ての組み合わせから配列アライメント長が最長のものを選択することで、その予測精度の向上が可能となることを示している。

第5章「Discussion」では、提案手法による立体構造予測精度の改善の生物学的な応用における重要性について、タンパク質間ドッキングを例にとり議論している。また、第3章と第4章で提案された2つの手法を組み合わせた予測パイプラインを提案し、僅かではあるがさらなる予測精度の向上が得られることを示している。

第6章「Conclusion」では本論文の結論を総括している。

第7章「Future work」では提案手法の限界について議論し、今後の課題について述べている。

以上で述べたように、本論文は現在最も実用的とされるホモロジーモデリング法によるタンパク質立体構造予測について、配列アライメントの品質の向上という観点から改良を行い、その予測精度の向上を達成している。その有用性は高く、その成果は工学上貢献するところが大きい。よって我々は、本論文が博士（工学）の学位論文として十分価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。