様式３

論　文　内　容　の　要　旨

氏　名　（　小山恭平　）	
論文題名	Transformer encoders for predicting T cell receptor-peptide binding can associate attention weights with interpretable protein structural properties （T細胞受容体とペプチド結合予測のためのトランスフォーマーエンコーダーにおけるアテンションとタンパク質構造に基づく解釈可能性に関する研究）

This dissertation presents a novel exploration into T cell receptors (TCRs), crucial components of the immune system that interact with ligand peptides. These peptides, presented by Major Histocompatibility Complex (MHC) molecules, are recognized by TCRs to initiate immune responses. The interaction between a TCR and its corresponding peptide-MHC complex (pMHC) is fundamental to initiating protective immune reactions against pathogens and malfunctioning cells. Understanding TCR-pMHC interactions is, therefore, vital for analyzing immune system mechanics, designing vaccines, and developing targeted immunotherapies. However, current experimental methodologies for studying TCR-pMHC interactions are resource-intensive and time-consuming, with computational models limited to retrospective data analysis and lacking interpretability. Additionally, the prediction of the TCR-pMHC interaction is difficult due to the massive combination patterns of TCRs and peptides. Addressing these challenges, this work introduces a novel approach using a machine learning model with a modified Transformer encoder, employing a source-target-attention neural network, or cross-attention layer. Central to this research is the development of a model that predicts TCR-pMHC interactions from amino acid sequences of the TCR's complementarity-determining region 3 (CDR3) and peptides. Unique to this study is the utilization of an external prospective dataset and the Transformer encoder layer to examine TCR-pMHC structural properties through attention weights. The model demonstrates superior performance on benchmark test sets and external datasets, surpassing other models in the average precision score, although the score limitation of the model is revealed by visualizing the data distribution difference. A detailed analysis links neural network attention weights to protein structural properties, classifying residues into attended groups to identify statistically significant properties, such as hydrogen bonds within CDR3, not between CDR3s and peptides. Chapters 2 and 3 of the dissertation delve into the cross-attention mechanism's predictive power and its interpretability in TCR-pMHC interactions, with Chapter 3 specifically comparing the efficacy of cross-attention and standard-attention mechanisms. The findings affirm the cross-attention model's superiority in revealing interaction dynamics at the molecular level, thus confirming more interpretability. In summary, this dissertation contributes substantially to bioinformatics and immunological studies using the Transformer-based attention's interpretability, providing a pathway toward more effective and interpretable computational tools. The insights acquired hold significant implications for the prediction of TCR-pMHC interactions, and this research not only enhances our understanding of molecular recognition but also lays the groundwork for developing new therapeutic approaches.

様式 7

論文審査の結果の要旨及び担当者

氏　名　（　小山恭平　）			
	（職）		氏　名
論文審査担当者	主 査	招へい教授	水口　賢司
	副 査	教授	上田　昌宏
	副 査	教授	山﨑　晶
	副 査	招へい教授	泰地　真弘人

論文審査の結果の要旨

小山恭平氏は、T細胞受容体(TCR)とペプチド結合予測のためのニューラルネットワークの研究において、タンパク質のアミノ酸配列のみから特定のTCRと抗原ペプチドが結合するかどうかを高い精度で予測する計算モデルの構築に成功した。この特異的な結合の有無を予測できれば、ワクチン開発やがん免疫療法などへの大きな貢献が期待されるが、現状は実用に足る予測方法は存在していない。申請者は、新たな深層学習手法を適用することにより、既存研究よりも高い精度の予測が可能であることを客観的なベンチマークテストを用いて実証した。また、トランスフォーマーモデルのアテンションと呼ばれるパラメーターの値をタンパク質立体構造データと組み合わせて分析することで、モデルのアテンションが大きい（即ち予測に重要な）アミノ酸残基は、TCR内での水素結合形成に関わっているという結果を統計的に結論づけた。これは、本解析が分子設計についての有用な指針を与え得る結果を生み出したと評価できる。このように、本研究は、将来的にワクチン開発などに寄与し得る分子間相互作用予測モデルを提供するのみならず、タンパク質の構造・機能の観点からモデルに解釈性を与えるもので、博士（理学）の学位論文に値するものと認める。なお、チェックツール“iThenticate”を使用し、剽窃、引用漏れ、二重投稿等のチェックを終えていることを申し添えます。