SIFF 2026
Singapore International Food Forum 2026

Learning Peptide Taste and Bioactivity from Weak Supervision

Yi Zhang Lee [a], Yijin Yang [a], Dachuan Zhang [ab], Huadong Xing [c*]

[a] *Department of Food Science and Technology, Faculty of Science, National University of Singapore, 2 Science Drive 2, Singapore 117542, Singapore*

[b] *Bezos Centre for Sustainable Protein at the National University of Singapore, 2 Science Drive 2, Singapore 117542, Singapore*

[c] *Empa−Swiss Federal Laboratories for Materials Science and Technology, Technology and Society Laboratory, CH-9014 St. Gallen, Switzerland*

Abstract

Peptides are a large and still relatively underexplored class of biomolecules with established importance for both human health and food quality. Released from plant, animal, and microbial proteins, peptides exhibit diverse taste attributes such as bitter, sweet, sour, salty, and umami, alongside a wide range of biological effects, including antibacterial, antioxidant, antihypertensive, and immunomodulatory. At the same time, peptides may display harmful effects like allergenic and hemolytic. However, traditional peptide property evaluation via experimentation is resource-intensive.

In this study, we propose a positive-unlabelled pseudo-label bootstrap framework that screens peptides across fourteen bioactivity and five sensory tasks from weak supervision. Our framework integrates curated positive peptide datasets, large unlabelled candidate peptide pools, existing task-specific teacher models, and protein-language-model-based student predictors. For each task, an ensemble of existing teacher models scores unlabelled peptides -to curate high-confidence pseudo-negative candidates, which are then combined with reported positive sequences to construct task-specific training sets. A protein language encoder is fine-tuned independently for each task, with application of contrastive learning to improve representation quality under weak supervision. Separate student models are trained for each taste and bioactivity endpoint, enabling each task to have its own positive set, pseudo-negative background set, decision threshold, and evaluation summary. In addition, we defined task-specific applicability domains in the embedding space to restrict predictions to sequences that fall within the reliable inference boundary.

Overall, our framework provides a scalable approach to prioritizing peptides for bioactivity discovery, functional food development, and safety evaluation, and thereby establishing the first computational basis for exploring the relationship between peptide bioactivity and taste properties. Moreover, the workflow can be extended to additional peptide-related tasks and support the broader discovery of functional peptides from diverse sources.