Physicochemical Signatures in Machine-Learning-Assisted Discovery of Anticancer Peptides from *Kappaphycus striatus*

Elijah V. Estante

elijah.estante@gmail.com

Research Unit, Philippine Science High School – Main Campus

Abstract

Marine macroalgae represent promising sources of bioactive peptides with potential pharmaceutical applications. This study characterized the physicochemical signatures of proteome-derived peptides from Kappaphycus striatus, an economically important carrageenan-producing red seaweed, and investigated their associations with machine-learning-based anticancer predictions. A library of 41,617 peptides previously prioritized using PeptideRanker was evaluated using two AntiCP 2.0 classification models. Seven sequence-derived physicochemical descriptors were examined across four prediction groups: concordant-positive, concordant-negative, and two discordant classifications. Group differences were assessed using Kruskal–Wallis tests and effect-size estimates, while logistic regression evaluated associations between peptide characteristics and model disagreement. Prediction-score distributions were additionally examined to assess classification-boundary sensitivity. 9,408 of the evaluated peptides were consistently classified as anticancer-positive, whereas 16,658 (40.03%) received discordant predictions. Physicochemical characterization revealed substantial between-group differences in peptide length ($\varepsilon^2 = 0.275$), acidic residue fraction ($\varepsilon^2 = 0.182$), and basic residue fraction ($\varepsilon^2 = 0.145$). Concordant-positive peptides exhibited a median length of 10 amino acids, compared with 23 among concordant-negative peptides. Logistic regression demonstrated modest discrimination of discordant predictions (AUROC = 0.6302). These findings identify distinct physicochemical patterns associated with computationally predicted anticancer activity and determine the importance of accounting for model variability in marine peptide bioprospecting. The study provides a foundation for diversity-aware candidate identification, prioritization and subsequent experimental validation of seaweed-derived peptides.

Keywords: Kappaphycus striatus, marine biotechnology, anticancer peptides, physicochemical characterization, machine learning