AI-POWERED DIAGNOSTICS

A New Software For Conotoxin Identification And Classification

Lyman Monroe Los Alamos National Laboratory   **Li-Wei Hung** Los Alamos National Laboratory   **Robert Williams** Los Alamos National Laboratory   **Hau Nguyen** Los Alamos National Laboratory

Conotoxins are a diverse set of neurotoxic peptides, which target a range of receptors and ion channels, causing a multitude of physiological effects. Conotoxins are grouped into different pharmacological classes: α-, δ-, κ-, μ-, and ω- classes based on their binding receptors/targets. Each of the five classes of conotoxins bind to different target receptors. For example, α-conotoxin inhibits nicotinic acetylcholine receptors at nerves and muscles while δ-conotoxin inhibits fast inactivation of voltage-dependent sodium ion channels. Knowledge of the biological targets (receptors, binding sites, etc.) and the corresponding toxicity of these molecules is vital to understand and mitigate their impact on living organisms. Despite the critical roles of conotoxins in biodefense, our ability to identify, characterize, and determine the toxicities of these toxins is difficult, costly, labor-intensive, and time-consuming which often requires a series of diverse and complex biological, toxicological, and analytical techniques for effective characterization. Here we have developed a user-friendly, machine learning (ML)-based software for the identification and classification of conotoxins. This software app leverages our recent success in using ML to improve conotoxin and conotoxin class predictions with additional features extracted from the primary amino acid sequence that include dipeptide features, residue frequency of occurrence, chemical characteristics, as well as sequence similarity to curated conotoxin class libraries. Our app is available on mac, pc, and Linux computers as well as apple and android devices. Input is either a picture of the amino acid sequence of the unknown sample or a text string. The app then reports if the sequence input is a conotoxin, and, if yes, reports its predicted pharmacological class and a confidence score. In preliminary testing, our predictions obtain average precisions, recalls, and f1-scores of 97.0%, 96.4%, and 96.8%, respectively, across the five conotoxin classes with a processing time of a fraction of a second per sequence. Our software is not only accurate but also very fast and, since it does not require access to the internet, it is a great choice for remote location field detection and identification when a putative sequence is known. We envision this software will be adapted and further expanded to include the prediction and identification of other toxic peptides such as snake, scorpion, and spider venoms. The rapid identification and classification of conotoxins, and eventually, toxin peptides in general, will facilitate JTSO's mission to anticipate, defend, and safeguard against chemical and biological threats for the protection of our warfighter and the nation.