INNOVATING CROSS-DOMAIN SOLUTIONS TO DETECT EMERGING BIOLOGICAL THREATS

Melding Experimental And Ai Approaches For Threat Sensing And Characterization

Brian Lee MIT Lincoln Laboratory **Ta-Hsuan Ong** MIT Lincoln Laboratory **Phil Tomezsko** MIT Lincoln Laboratory **Rafael Jaimes** MIT Lincoln Laboratory **Paul Ippoliti** MIT Lincoln Laboratory **Jeff Werlich** MIT Lincoln Laboratory **Avery Meyer** MIT Lincoln Laboratory **Meghan Ramsey** MIT Lincoln Laboratory **Brian Lee** MIT Lincoln Laboratory **Josh Dettman** MIT Lincoln Laboratory

Rapid technological advances in AI foundational models and specialized design tools, bioengineering and biomanufacturing, and high-throughput automated production facilities have the potential to both create a significantly more diverse and complex CB threat space as well as provide means to mitigate known and novel threats. Proactively meeting upcoming detection, identification, characterization, exploitation, medical countermeasure, and attribution defense challenges will require re-thinking the current list-based paradigm where a set number of specified threat agents are controlled and prepared for. MITLL envisions a future where any small molecule chemical, protein toxin, virus, and bacterium can be detected and characterized without prior knowledge of the agent. To achieve this vision, a highly integrated system of experimental characterization with predictive AI algorithms, as well as in vitro and wearable host-response-based sensing, will be needed.

This presentation will focus on new MITLL efforts to design and build an agent-agnostic protein detection and characterization pipeline, with emphasis on protein toxins. This system combines bottom-up proteomics and multiple computational biology pipelines. Technical objectives include untargeted proteomics measurement of a trace and complex mixture, protein structure prediction without potentially time-consuming crystallography, and rapid function prediction for any protein sequence of interest.

Current results from our lab show protein detection sensitivities at nanogram levels, including for operationally-relevant complex mixtures, using a multi-'omic capable methods. The measured protein sequences are further analyzed using multiple computational biology pipelines to predict important properties, including from uncharacterized proteins. One pipeline, "Pipeline for Investigating Protein-Protein Interaction Networks" (PIPPIN), uses AlphaFold2[1] followed by a custom combination of analysis scripts to predict the structure of protein-protein interactions and provide an assessment of protein-host interactions and hypothesize function. A second pipeline, "Protein Enzyme Prediction Pipeline Application" (PEPPA) merges and prioritizes results from ProteInfer and CLEAN algorithms[2],[3] to predict protein functions (through enzyme commission numbers, protein family numbers, and gene ontology terms.) Current work includes developing an AI classifier for conotoxin pharmacological class and robustness (aggregation propensity and thermal stability) with integration of the computational biology tools with other software tools to allow technology transition to operations. The combination of these tools across experimental and computational disciplines will enable agent-agnostic protein toxin characterization to be a routine part of national security strategy against biological threats. Combined with the development of an architecture for agent agnostic DNA/RNA detection and characterization, this work provides a foundation for CB agent agnostic characterization and future plans to advance this capability will be discussed.

[1] Jumper, J. et al., Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596 (7873), 583-589.
[2] Sanderson, T. et al., ProteInfer, deep neural networks for protein functional inference. eLife 2023, 12, e80942.
[3] Yu, T. et al., Enzyme function prediction using contrastive learning. Science 2023, 379 (6639), 1358-1363.>