TOXIN MEDICAL COUNTERMEASURES - DEVELOPMENT OF NOVEL, BROAD-SPECTRUM COUNTERMEASURES FOR TOXIN EXPOSURE

How Do Autoencoders Help Explore The Conformational Space Of Md Simulations Of Cyclic Peptides?

Leonie Windeln Univeraity of Southampton **Charlie Holdship** University of Southampton **Steven Turner** University of Southampton **Jeremy Frey** University of Southampton **Andrew Jamieson** Unversity of Glasgow **Jonathan Essex** University of Southampton

α-Conotoxins are a class of disulfide-rich cyclic peptides produced by marine cone snails that target human nicotinic acetylcholine receptors (nAChRs). The specificity of these toxins against different isoforms of nAChRs makes them attractive pharmacophore candidates but also imposes a concern for potential misuse in bioterrorism.

To develop anti-toxins, the solution conformations must be determined before modelling their interactions with those newly developed binders and the receptor. We have carried out molecular dynamics (MD) simulations of five conotoxins using enhanced sampling methods. In MD simulations, large datasets with high dimensionality (many variables) are generated. These variables are the cartesian coordinates of each atom for each time-step of the simulation. From this, other variables such as the backbone torsion angles can be derived. To extract meaningful movement of the system over time and thus obtain stable solution structures, the dimensionality of these data must be reduced significantly.

A popular method for dimensionality reduction is principal component analysis (PCA) which uses a linear combination of all input variables to calculate the orthogonal collective variables which maximally capture the data covariance. In the case of our conotoxin simulations, the first two principal components of the PCA captured less than 50 % of the variance. We compare the use of PCA to analyze the conformational space of conotoxins to using Machine Learning autoencoders for the same task. Standard autoencoders have a symmetric neural network architecture with a bottleneck in the center (encoder layer) whilst the network is trained to reduce the error between input and output (decoder) layer. We were able to perform reduction of the simulation data to two dimensions (2 encoder nodes) and extract diverse conformations with better separation than for the PCA workflow. Furthermore, we achieved high model accuracy (defined by low reproduction error) throughout the trajectory for autoencoders, which was not the case for PCA.

Going forward autoencoders will enable us to more accurately model interactions of conotoxins and their targets. Our findings in using autoencoders for analyzing simulations of cyclic peptides will allow us to expand their use in structural biology and help identify more biologically relevant conformations of other cyclic peptides.

All simulation was performed using Iridis 5 HPC. We would like to thank Dr Jack Sawdon for helpful discussions.