Abstract

Peptides are promising drug development frameworks thanks to their high target selectivity, tolerability, and relatively low production cost. However, despite the fact that several thousand potentially therapeutic peptides have been reported, until 2018 only about sixty were approved in the United States, Europe and Japan. Toxic effects, including hemolysis have been identified as the primary challenge that hinder the development of promising peptide drugs. To overcome this obstacle, we propose a novel approach based on complex network science, interactive data mining and multi-query similarity searching (MQSS) to gain a better understanding of the chemical space of hemolytic peptides. By leveraging these techniques, we aim to design more effective peptide drugs with reduced hemolytic activity, thus facilitating the development of safer and more efficient therapies. Metadata networks (METNs) were used to systematically identify and characterize general patterns that are commonly associated to hemolytic peptides. In addition, Half-Space Proximal Networks (HSPNs), were constructed using five different two-way dissimilarity measures, with the aim of effectively represent the hemolytic peptide space. Then, the best candidate HSPNs were used to extract various scaffolds that capture information of almost all the chemical space avoiding peptide overrepresentation. Such scaffolds were used to develop MQSS models for predicting hemolytic toxicity from peptide sequences. Our best model outperformed *state-of-the-art* Machine Learning (ML)-based models, achieving a MCC equal to 0.99. This model was used to characterize the prevalence of hemolytic toxicity on therapeutic peptides. We found that the number of reported hemolytic peptides might be 3.9-fold lower than the actual number. Finally, by means of an alignment-free approach, we reported 47 putative hemolytic motifs, which might provide hints about the mechanisms of hemolysis and can also be used as toxic signatures when developing novel peptide-based drugs.