Abstract

Amyloids are insoluble, fibrilar protein aggregates, commonly known for their role in the development of neurodegenerative disorders. However, more recent studies show that such structures can be utilized by a variety of organisms to perform physiological functions including biofilm formation, hormone storage and signaling. Furthermore, it was shown that both functional and pathological amyloids can interact in a number of ways. Such interactions can lead to a significant increase in aggregation rates or inhibition of fibril formation. Plenty of experimental methods have been proposed to identify and characterize amyloidogenic proteins, however, they all share the same problem of being expensive and time-consuming. Therefore, their use is still limited to small-scale studies. To overcome this problem several computational methods have been proposed. Unfortunately, their accuracy is still limited. This is especially true in the case of functional amyloids which are severely underrepresented in amyloid databases. Furthermore, there are no tools dedicated to the prediction of amyloid cross-interactions. This relatively recently discovered phenomenon can play a pivotal role in our understanding of the comorbidity of amyloid-related disorders. In this work, I present novel computational tools and methods which can improve the identification of amyloid-prone regions as well as their cross-interactions.

During my PhD, I developed a new method for the identification of aggregation-prone regions in proteins - PATH (Prediction of Amyloidogenicity by THreading). The method combines structural modeling with machine learning. The proposed method allows for the accurate identification of amyloidogenic fragments and enables the user to infer the most probable structural class of the resulting amyloid core. Our results showed that PATH, as well as some other bioinformatics methods, is robust against misannotated training data. All tested methods have problems with the identification of functional amyloids which are severely underrepresented in available databases. To better understand functional amyloids, a detailed characterization of CsgA proteins from Escherichia coli and Salmonella enterica was performed. We also investigated amyloids in the Caenorhabditis elegans proteome. Experiments by our co-workers showed congo red binding structures in the pharynx of this organism which suggested the presence of amyloids. Therefore, we scanned the whole proteome for possible amyloidogenic proteins, using PATH and AmyloiGram tools. Despite using two different methods, our analysis showed many false positives, which shows potential problems with applying computational methods on a proteome scale.

Discussed difficulties with large-scale identification of amyloids encouraged us to try a different method in the next project. We aimed to better understand the fungal NLR system. The peculiar property of this system is that it uses amyloid aggregation to propagate the signal between the receptor and effector protein. It involves the so-called Amyloid Signaling Motifs (ASM) which are usually located in N or C-terminals of proteins. Using de novo motif detection tools as well as natural language processing models we identified a new type of ASM - PUASM which stands for Pnp_Udp Amyloid Signaling Motif. We also significantly improve the annotation of NLR-related protein domains. Finally, I developed PACT - the first method for the prediction of amyloid cross-interactions. PACT not only achieved good accuracy on the task of interaction prediction but can also be used to identify novel amyloid-prone regions. The method was then used to identify which region is most likely involved in interactions of CsgA with human Islet Amyloid Polypeptide.

26-06-2023