Abstract

Amyloids are a group of peptides and proteins that fold into assemblies of insoluble fibrils of very regular and tightly packed β-cross structures, which resemble a steric zipper. They are associated with many civilization diseases, such as type 2 diabetes, and wide range of neurodegenerative diseases. However, due to their self-assembly properties and unique physicochemical characteristics, as well as favorable mechanical and optical features, amyloids can and are being used in different fields of material and life sciences.

Amyloids are very sensitive to various experimental conditions in which their aggregation process is studied. Moreover, experimental conditions have a great influence on the amyloid characteristics and classification. In many cases, data collected from experiments using different techniques and performed by various scientific groups, do not include all the information about the specifics regarding experimental conditions. This has a strong impact on the ability, for instance of bioinformatics methods, to predict the amyloid propensity or the structural models of aggregation.

This dissertation focuses on the influence of selected internal and external factors affecting the aggregation process of amyloidogenic peptides. In the study, we consider in particular the length of the amino acid sequence, the effect of mutations, the influence of the solvent including, type and salt concentration, the interactions with other peptides, as well as the proximity and interaction with a cell wall (lipid environment). This examination concerned selected pathological and functional amyloids and their most relevant fragments, and relied on results from experimental and computational investigations.

First, we set a reference protocol to study the aggregation properties of short peptides, necessary to identify possible amyloidogenic amino acid sequences. Hence, we selected hexapeptides using AmyloGram, a well-established bioinformatics prediction tool. Then we examined whether and to what extent, poorly annotated training data can influence the accuracy of bioinformatics methods. Finally, we investigated whether longer sequences can be treated in the same way as short ones.

Second, we examined experimentally long amino acid sequences (up to 23 amino acids) from the functional amyloid CsgA protein (consisting of R1–R5 imperfect fragments) in order to check their amyloid propensity. We investigated whether such peptides are more sensitive than pathological amyloids to selected internal and external factors, such as: point mutations, solvent, type of ions, ions' concentration. To do so we compared the effects of these parameters on the aggregation properties of homologous repeats from two bacteria *Escherichia coli* and

Salmonella enterica. This study included the influence of mutations, which has been validated by computational and theoretical studies.

We finally conducted a study on pathological amyloids (sequences up to 42 amino acids), namely Aβ42 and hIAPP. To mimic physiological conditions, experiments and simulations were carried out in the presence of a lipid membrane. Therefore, we investigated the influence of lipid membrane and interaction with another peptide on aggregation of these amyloids.

The results of the research present that it is important to collect coherent data from experiments, which could be disrupted by non-identical conditions. Additionally, different experimental methods may deliver somehow different results regarding the same peptides. This effect is more significant in case of some sequences.

We found that short hexapeptide sequences exhibit distinct aggregation propensities in response to external factors, e.g., the solvent used, compared to longer sequences (up to 23 amino acid). The flexibility of six-amino acid sequences allows them to adopt specific conformations. Moreover, the symmetry-breaking transitions phenomenon plays a crucial role in this process.

Additionally, we showed that the choice of solvent influences on the aggregation process. The usage of deuterium oxide might alter the classification, as observed for the R2 fragment of *S. enterica*. Fibrils were detected in the presence of heavy water but not in phosphate buffered saline. The dominant peptide conformation in D_2O was attributed to intermolecular aggregates, a signature typical of amyloid structures. Based on these findings, the R2 fragment's classification as amyloid or non-amyloid depends on the conditions applied.

Next, we showed that R4 fragment from *S. enterica* (SR4) has a larger tendency to aggregate compared to the R4 fragment from *E. coli* (ER4). Theoretical sequence analysis revealed that SR4 has higher hydrophobicity and electric charge values than ER4. Those factors are known to increase protein aggregation. Additionally, SR4 has more significant variations in gatekeeper residues, which play a crucial role in regulating amyloid formation. The results indicate that specific amino acid substitutions can significantly affect the propensity of functional amyloid proteins to form amyloid structures, with potential functional consequences.

The research highlighted the influence of the phosphate buffer ions on the peptides' morphology, affecting the local electrostatic interactions involving the polypeptide chains. The molecular dynamics simulations revealed that the interactions of positively charged amino acids with negatively charged phosphate moieties in the buffer determined the morphology. Non-specific ion enrichment may occur in the proximity of protein moieties that bear charges

opposite to those of the ions. Hence, the dimers structures observed in the simulation may have contributed to the snail-like conformation of the M4 peptide.

In the last phase of the study, the stability of Aβ42 in a lipid membrane was examined through molecular dynamics simulations and atomic force microscopy. The results suggested that Aβ42 remains embedded in the lipid membrane, indicating that peptide aggregation may occur within the membrane. It was also found that Aβ42 interacting with hIAPP exhibited a higher β-sheet content when in close proximity, compared to the analysis of Aβ42 alone. Atomic force microscopy investigations showed that supported lipid bilayers were more affected in the presence of both Aβ42 and hIAPP.

Overall, the study sheds light on the potential challenges arising from the ambiguity of experimental outcomes in the context of amyloid investigations. Our research has shown that even small changes in experimental conditions can alter the properties of amyloid peptides and proteins, which can be the cause of obtaining incorrect models and predictions of bioinformatics tools, based on incompatible learning data. However, as shown in the study, minor deviations are not detrimental to these tools. Bioinformatics methods are fairly robust to incompatibilities in the data, but their performance can be disrupted if the influence of experimental conditions and seemingly insignificant differences in sequences of homological sequences are not taken into the consideration. The results described above, confirm that in the process of planning experimental research is important to choose proper conditions according to the studied object. Based on that choice, we can effectively modulate the peptide tendency to aggregate.

23.04.2023 Natalia Szulc