Representation of Protein Aggregates in an Ontological Framework

Lauren Wishnie[1], Alexander Cox[1], Alexander Diehl [3]

Department of Biomedical Informatics, University at Buffalo, Buffalo, NY, USA

Abstract

The Protein Ontology is a valuable tool for the organization, representation, and retrieval of data related to proteins and protein-related entities. The Complex Portal organizes macromolecular complexes, including those containing proteins, from a number of model organisms. Together, these two resources serve the scientific community by curating, organizing, and representing proteins and protein complexes and placing those entities in an ontological context. However, there is a dearth of information between these two databases pertaining to macromolecule and, more specifically, protein aggregates, which are distinct from complexes. A macromolecule aggregate has primary constituents that are two or more macromolecules. Their compositions and conformations can vary between instances, and they can contain other constituents like small molecules or ions. A protein aggregate, then, is a macromolecule aggregate whose primary constituents are proteins, but can be configured in various ways from instance to instance. Protein aggregates are important in diverse biological processes and pathologies, the most well-known of which are neurodegenerative diseases such as Alzheimer's and Parkinson's. However, it is a misconception that the presence of protein aggregates is always pathological in nature; the immune response uses antibody aggregates to detain pathogens and mark them for destruction, and the clotting cascade in response to vascular injury relies upon fibrinogen aggregation. Despite these important functions of protein aggregation in biological processes, there is no unified protein aggregate ontology. Based on this, we propose the construction of a protein aggregate ontology as an extension of the Protein Ontology, which would allow for the generation of a framework to organize and represent protein aggregates as a complement to the representation of proteins and protein complexes in Protein Ontology and the Complex Portal. We have performed preliminary curation of terms for this protein aggregate ontology, and are structuring the ontology using the Protege ontology editor. The availability of a protein aggregate ontology will support researchers in categorizing types of protein aggregates, understanding both new and existing processes of protein aggregation, as well as contextualizing protein aggregates in relation to physiological and pathological processes and disease. Thus, a protein aggregate ontology will promote an understanding of how protein aggregates fit structurally and functionally into the larger web of biological molecules.

Keywords:

Ontology, protein, protein aggregate, macromolecules