A Semantic Design for the Biological Processes Associated with Intrinsically Disordered Proteins

Atsuko Yamaguchi[1,*,†], Yumiko Kado[2], Shigetaka Sakamoto[3], Satoshi Fukuchi[4] and Motonori Ota[2]

[1]*Tokyo City University, Tokyo 158-8557, Japan*

[2]*Nagoya University, Nagoya 464-8601, Japan*

[3]*HOLONICS Corporation, Numazu 411-0803, Japan*

[4]*Maebashi Institute of Technology, Maebashi 371-0816, Japan*

Abstract

Intrinsically disordered proteins (IDPs) challenge the traditional view of protein structure, lacking a stable tertiary structure yet playing pivotal roles in various biological processes. To address the growing need for a comprehensive resource, we developed IDEAL (https://ideal-db.org/), a curated database of IDPs. IDEAL leverages semantic web technologies like RDF and SPARQL to enhance accessibility and interoperability. Here, we introduce the design of the newly constructed RDF dataset that represents the biological processes associated with IDPs.

Keywords

Semantic Web, Resource Description Framework, Intrinsically disordered proteins

1. Introduction

Intrinsically disordered proteins (IDPs) form a dynamic protein class lacking a fixed three-dimensional structure in the isolated state, allowing diverse conformations upon interacting with other molecules. Vital in cellular processes like signaling and regulation, IDEAL (http://www.ideal-db.org/) is a significant repository for experimentally verified IDPs and intrinsically disordered regions (IDRs) [1]. Our contribution to IDEAL includes a semantic web-based resource with RDF and SPARQL technologies, providing both a user-friendly website and downloadable XML data. Recognizing IDPs' growing importance, we have expanded our RDF dataset to cover associated processes. Leveraging knowledge graphs for dynamic and procedural representation, this poster paper introduces our knowledge graph design for biological processes involving IDPs.

2. Representation of IDP-Associated Processes

In order to depict processes associated with IDPs, we have introduced two primary classes: "State" and "BiologicalProcess." The "State" class encompasses components such as proteins,

nucleic acids, etc., including their complexes. In the case of a component being a complex, the subunits are categorized by the "Subcomponent" class. The "BiologicalProcess" class is linked to "ProcessType," with instances including "catalyze," "associate," "dissociate," "translocate," among others. Furthermore, the "BiologicalProcess" class is associated with the "Object" class, which identifies the facilitator of the given biological process.

Figure 1 illustrates two instances of these processes. The left panel demonstrates a process where a complex is formed from the IDP APP and APBB1. APP and APBB1 are classified by the "Component" class in the first "State," connected to a "BiologicalProcess" with "ProcessType" as "associate." This "BiologicalProcess" is linked to the second "State," which contains a complex of APP and APBB1 as its components. Similarly, the catalyze process of phosphorylation facilitated by MAPK10 can be represented using the "State" and "BiologicalProcess" classes.
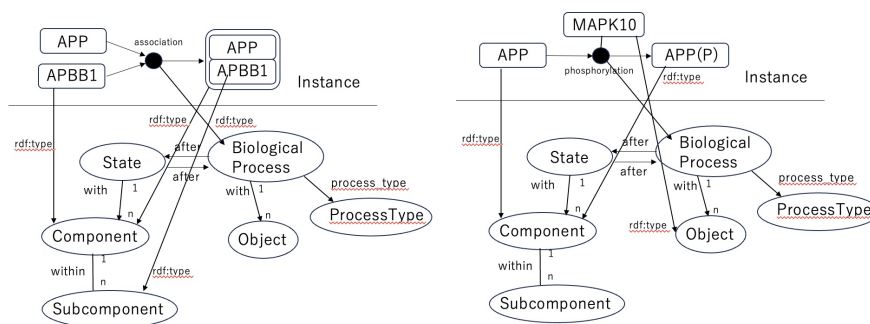

Figure 1: Instances of biological processes associated with an IDP.

3. ProcessType Verification

In addition to relying on the notes of ProcessType in XML and RDF, we can verify it by examining the relationship between the two connected "States" and the corresponding "BiologicalProcess." For instance, if the first "State" comprises two protein components, and the second "State" includes a complex, the inferred "ProcessType" would be "association." In cases where the first "State" involves a protein and the second "State" involves a phosphorylated protein, the determined "ProcessType" would be "catalyze." We have designed such rules for verifying the "ProcessType" based on the characteristics of the associated "States."

4. Acknowledgments

References

[1] S. Fukuchi, T. Amemiya, S. Sakamoto, Y. Nobe, K. Hosoda, Y. Kado, S. D. Murakami, R. Koike, H. Hiroaki, M. Ota, IDEAL in 2014 illustrates interaction networks composed of intrinsically disordered proteins and their binding partners, Nucleic Acids Research 42 (2014).