Fujitsu develops technology to predict biochemical reactions, clarifying the mechanisms of genetic disorders

October 31 2018

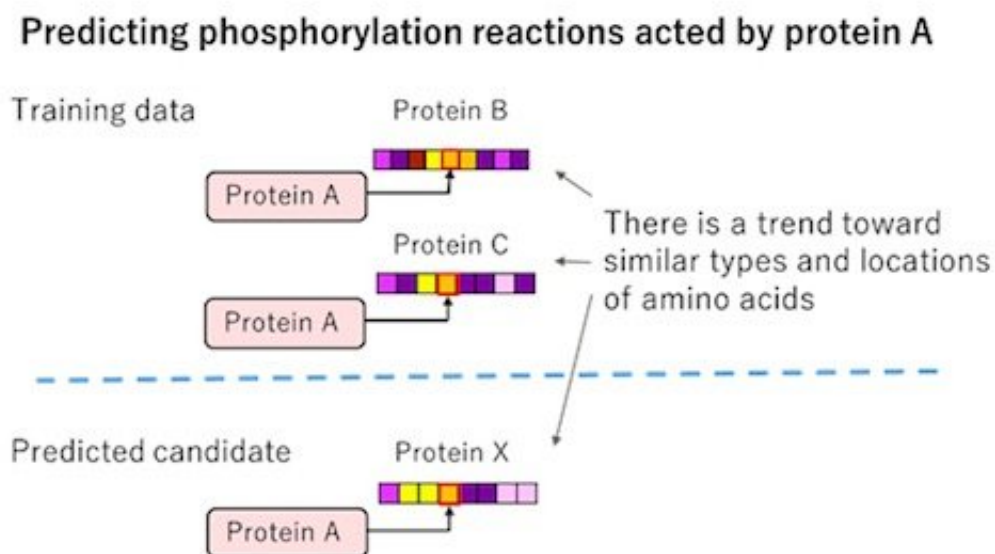Predicting phosphorylation reactions acted by protein A

Figure 1: Conventional AI Technology. Credit: Fujitsu

Fujitsu Laboratories Ltd., the Insight Centre for Data Analytics, a data analytics research institution based in Ireland, and Fujitsu (Ireland) Limited today announced the development of a technology that makes it possible to predict large volumes of unknown chemical reactions, about twice as many as the conventional procedure. In serious diseases, including cancer, it is common for there to be abnormalities in phosphorylation reactions, which are chemical reactions that occur

between proteins. Accordingly, there are high expectations that clarifying phosphorylation reactions will lead to effective treatments. At present, however, because only a few phosphorylation reactions have been identified, there has been a problem in predicting large volumes of phosphorylation reactions caused by combinations of unknown proteins. Now, by building a knowledge graph that can encompass an overview of the interrelations between proteins, it is possible to check the relationship between new proteins where phosphorylation reactions can be predicted. In this way, this technology will contribute to the advancement of medicine, as it can be expected to be useful on the front lines of drug discovery research, and have customized applications in the field of precision medicine.

Development Background

Biological systems within the body are maintained by exchanges of information through the chemical reactions of various proteins within cells. In recent years, science has come to understand that many serious diseases, such as cancer, are partially caused by abnormalities in phosphorylation reactions, which are representative of the chemical reactions between proteins. If pharmaceuticals that repaired abnormal phosphorylation reactions could be developed, that would enable more effective treatments. At present, however, only a few phosphorylation reactions are well understood, so there is a need for the discovery of unknown phosphorylation reactions, and to enrich the data on phosphorylation reactions.

Issues

Phosphorylation reactions are chemical reactions in which a protein attaches a phosphoryl group to the amino acids that make up another protein. In order to discover them, it is necessary to check the

combinations of proteins that cause phosphorylation reactions through biological experiments. Nonetheless, as there are more than about 800,000 possible combinations just with proteins, and because significant costs and time are required for biological experiments, it is necessary to right from the start predict high-probability combinations. It is known that whether a phosphorylation reaction will occur depends on the structure of the amino acid sequence that makes up the protein. AI technology is therefore already being used to predict new phosphorylation reactions by training the AI on the structure of amino acid sequences that are already known to cause phosphorylation reactions. While this technology can predict reactions in which the structure of the amino acid sequence is similar to those that are known to cause phosphorylation reactions, it has not been capable of predicting those in which the structure of the amino acid sequence is significantly different from the already known phosphorylation reactions.
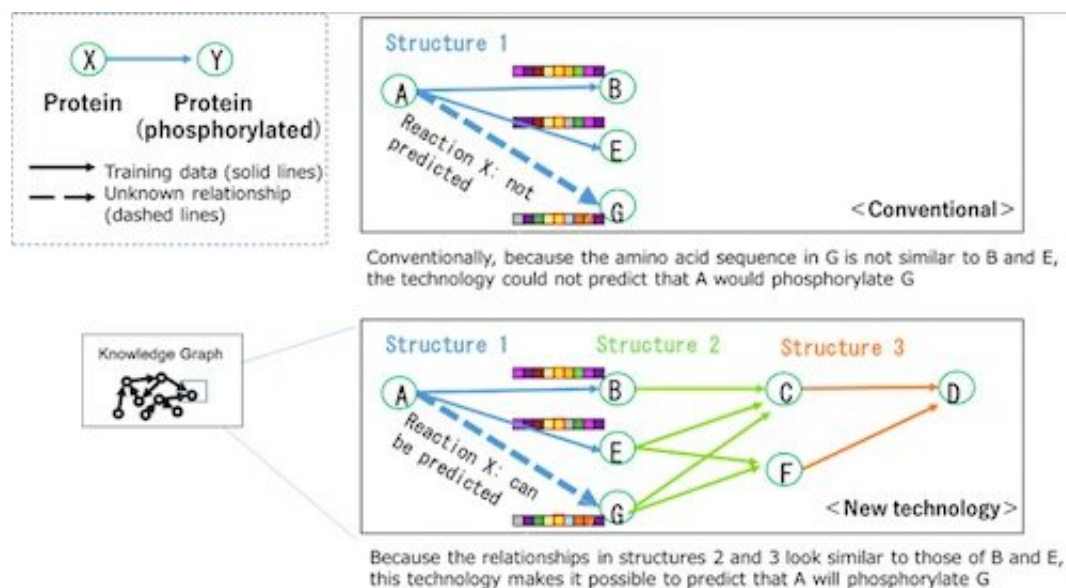

Figure 2: Example of predicting phosphorylation reactions using knowledge graphs. Credit: Fujitsu

According to recent medical research, there is a phenomenon in which proteins that have undergone reactions may phosphorylate other proteins in a chain reaction (chained information), and this may be the key to predicting new, unknown phosphorylation reactions related to that phenomenon. Based on such research, Fujitsu Laboratories, the Insight Centre, and Fujitsu Ireland have now included not only structural information about amino acid sequences in the knowledge graph, but also chained information. The organizations have developed a technology (patent pending) to represent the complex patterns of chemical reactions as optimized attributes, which are attached to the lines in the knowledge graph. As these attributes were tailored to the sophisticated construction by the knowledge graph, they can lead to highly accurate prediction results. Conventionally, the relationship between proteins could only be checked through a single link in the chain. Yet by comprehensively displaying the relationship between proteins as connections of phosphorylation reactions (chained information), it becomes possible to clarify the positioning of the various proteins from a holistic perspective, and to predict unknown relationships.

Effects

When this technology was tested using evaluation data, the model was trained on phosphorylation reactions (9,802 reactions), and predicted 11,581,940 new phosphorylation reactions. This showed its capability in predicting about twice as many phosphorylation reactions compared to conventional technology that trained AI on the structure of amino acid sequences, without significant change to prediction accuracy. In addition, in order to test whether phosphorylation reactions predicted using this technology could actually occur within a living being, tests were conducted by Systems Biology Ireland, an Irish biological research institution and a joint research partner, using mass spectrometry equipment and antibodies. In this test, experts in biology selected and

tested a few phosphorylation reaction prediction results for proteins related to cancer, and were able to confirm nine phosphorylation reactions, of which eight were reactions that could not have been predicted with conventional technology. Systems Biology Ireland (SBI) director Walter Kolch, a world leading authority on systems biology research, said about these results "Combining Fujitsu's knowledge graph technology with SBI's understanding of biological networks, we have developed a new computational method that can predict which kinase phosphorylates which substrates. The method is accurate and could discover previously unknown phosphorylation sites, a major step forward for new drug development and more focused precision medicine."

By combining data on new phosphorylation reactions predicted by this technology with other biomedical data, it is expected to connect the chemical reactions from the causes of a disease (abnormalities in phosphorylation reactions) to the disease's symptoms, which can then be provided to those on the front lines of research as useful information in drug discovery. The effectiveness of treatments for diseases such as cancer can vary widely between patients. This technology, however, is expected to clarify the individual variation in the effects of treatments, contributing to the promotion of medicine tailored to individual patients. Fujitsu Laboratories, the Insight Centre, and Fujitsu Ireland will continue to further improve the accuracy of this technology to process biomedical data with knowledge graphs, extending the technology to biomedical projects at Fujitsu Limited in fiscal 2018. Moreover, by incorporating this technology into Fujitsu's AI technology, including Fujitsu Human Centric AI Zinrai, the organizations plan to accelerate the biomedical business.

Provided by Fujitsu