【Grant-in-Aid for Transformative Research Areas（A）】

Latent Chemical Space Based on Diverse Natural Products for Bio-active Molecular Design (Latent Chemical Space)

Head Investigator	Osaka University, Graduate School of Engineering, Professor KIKUCHI Kazuya　　Researcher Number:70292951
Research Area Information	Number of Research Area : 23A204　Project Period (FY) : 2023-2027 Keywords : Chemical Biology, Natural Products, Latent Space, Bio-informatics, Organic Synthesis

Purpose and Background of the Research

●Outline of the Research

The discovery and identification of biologically active molecules using two typical compound resources, natural products (first) and synthetic compound libraries (second), has been a driving force in promoting chemical biology research, a field that integrates chemistry and biology. In this research area, we propose a third resource to follow these two. This third resource is virtually generated from the Latent Chemical Space, which is constructed by deep learning technology based on bioactivity data of natural products, and is realized in real space using robust organic synthesis. The Latent Chemical Space created by the fusion of natural products and informatics research will bring about a paradigm shift in data-driven chemical biology research and revolutionize the design of biologically active molecules. To realize this, we will launch the "Cyber Bioactive Molecule Design Lab" consisting of three groups: Chemical Biology, Informatics, and Organic Synthesis. The goal is to establish a new science of bioactive molecule design that can develop innovative molecules that lead to the clarification of new biological functions and to the seeds for pharmaceuticals and agrochemicals, starting from the compounds created from this third resource.

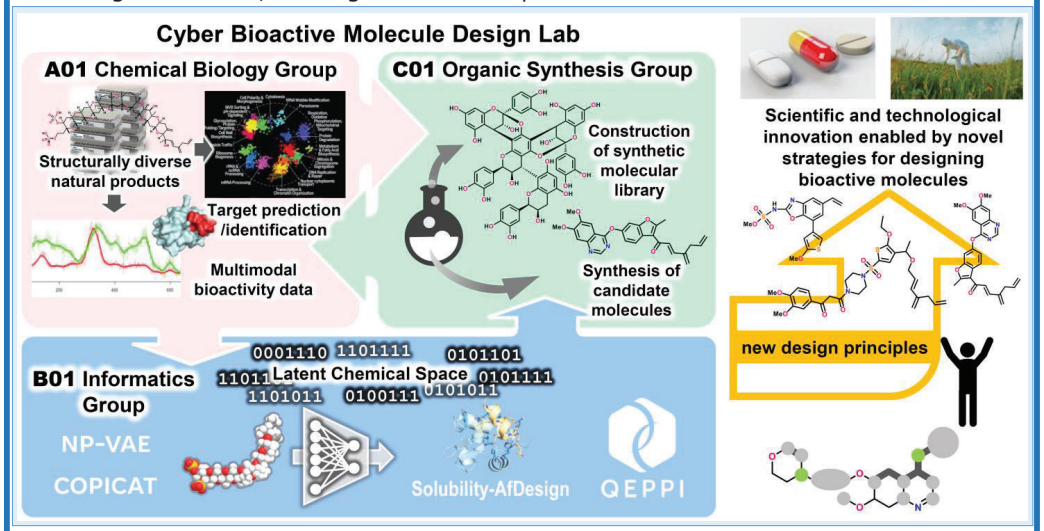

Figure 1. Cyber Bioactive Molecule Design Lab

The relationship between chemical structure and biological activity of a compound is called Structure-Activity Relationship (SAR). In modern chemical biology and drug discovery research, the established methods for enhancing the potency of bioactive molecules involve initially acquiring hit compounds from a vast array of library molecules through random exploratory studies. Subsequently, SAR was elucidated through empirical sturctural modifications. However, even if the number of compounds in the library is increased, this search method, which finds functional molecules at random, has limitations in structural design of diverse molecules. Therefore, there is a strong need for systematic strategies to design potent bioactive molecules in a logical and efficient manner. In this area, we built a plan to revolutionize chemical biology research by creating a third resource by combining complex natural products and informatics such as deep learning to construct compound latent chemical space and design new bioactive molecules.

Expected Research Achievements

In this field, we will develop a fusion of chemical biology and informatics based on natural product chemistry in order to realize innovative speed-up and labor-saving processes for both the simplification of complex structures and the search for new skeletal compounds with similar activity in the design of bioactive molecules. The compound latent space constructed by deep learning based on bioactivity data of natural products will create a third compound resource. By utilizing this molecular design lab, our goal is to identify the structural patterns required for bioactive molecules and to find universal laws of bioactive molecular design. For a new research field to emerge, it is essential to align objectives as well as technical support for new research methods. In the conventional organic chemistry of natural products, "complex molecular structures" have been the most important factor, and the pursuit of scientific theories to improve their functions has been a challenging research field. From this perspective, this research area aims to raise awareness of the scientific principles underlying molecular design through joint research and information exchange, mainly among young researchers, and at the same time, to share knowledge on research methods and to improve the bottom line of the entire field.

Group A (Chemical Biology): Select molecules that interfere with protein-protein interactions, and other molecular interactions to cause biological activities from traditional natural product resources in Japan, and collect comprehensive data on the mechanisms of action and target molecules. SAR information of the derivatives and various omics data will be provided for domain studies. In other words, it will play a role in connecting latent space and real space with activity data.

Group B (Information Analysis): The multimodal structure and activity data provided by Group A will be quantified by information analysis by Group B and used as input data for original multiple algorithms. Specifically, we will extract knowledge from data by Group A, using COPICAT, a method for predicting molecular interactions, using QEPPI, a small molecule evaluation function for prediction of modification in protein-protein interaction, and using Solubility-AfDesign, a deep learning method for solving peptide sequence design problems. In addition to above methods latent space is constructed by NP-VAE. The molecular structures designed by the latent space will be presented to Group A and Group C. At the same time, we aim to revitalize the field of chemoinformatics through the development of highly original methods.

Group C (Organic Synthesis): The candidate bioactive molecules with novel structures predicted by the compound potential space are synthesized in real space by Group C using its own chemical technologies, and after activity evaluation by Group A, the results are compared with those predicted by Group B. Since the chemical structures of the candidate molecules are expected to be diverse, Group C will select its own unique and diverse synthetic technologies. The prediction accuracy of the potential space will be improved, and the bioactive molecules ultimately created will be developed into tools for chemical biology research and drug discovery seeds.

Homepage Address, etc.	https://latent.chemical.space　@latentchem