Integrative Approaches of Bioassay and Computational Analysis for Discovering Potential Bioactive Compounds and Predictive Toxicity

Hideko SONE1, Xian-Yang QIN2, Kohsuke HAYAMIZU3, Wataru FUJIBUCHI* and Yoichi NAKAO5

1 Environmental Health and Prevention Research Unit, Yokohama University of Pharmacy, Yokohama 245–0066, Japan
2 Laboratory for Cellular Function Conversion Technology, RIKEN Center for Integrative Medical Sciences, Yokohama 230–0045, Japan
3 Department of Food Sciences, Yokohama University of Pharmacy, Yokohama 245–0066, Japan
4 Department of Life Science Frontiers, Center for iPS Cell Research and Application, Kyoto University, Kyoto 606–8507, Japan
5 Graduate School of Advanced Science and Engineering, Waseda University, Tokyo 169–8555, Japan

Summary Changes in eating habits are brought about by drastic changes in lifestyle and environment, and, it has been pointed out, are strongly involved in the increase in neurological diseases and onset of cancer in younger adult ages. There is a wide variety of chemical substances in food, and there is a need to analyze the effects of complex exposures on complex mechanisms of action and to develop methods for evaluating and predicting them. The power of molecular nutrition needs to create an integrated approach to human nutrition in line with the grand social challenges of diet-related illnesses. The current article aims to explore some of these areas where integration is appropriate. Therefore, in this symposium, we will introduce the contents of four performers who are conducting cutting-edge research. 1) Chemoprevention by vitamin A and its derivatives, 2) Toxicity prediction of natural compounds from a developing database of bioactive gradients from Kampo medicine, 3) Toxicity prediction of chemicals using pluripotent stem cells, 4) Detection of bioactive compounds in “Aji” or “Umami” in food. By detecting and predicting the biological activity and toxicity of chemical substances such as nutrients in foods, it will be possible to provide better molecular information on dietary components. In addition, we will introduce next-generation health and prevention methods.

Key Words integrative approaches, bioassay, computational analysis, bioactive compounds, predictive toxicity

Introduction The power of molecular nutrition needs to create an integrated approach to human nutrition in line with the grand social challenges of diet-related illnesses. Changes in eating habits are brought about by drastic changes in lifestyle and environment, and, it has been pointed out, are strongly involved in the increase in neurological diseases and onset of cancer in younger adult ages. There is a wide variety of chemical substances in food, and there is a need to analyze the effects of complex exposures on complex mechanisms of action and to develop methods for evaluating and predicting them. Computational toxicology is a next-generation toxicity assessment method that comprehensively utilizes various methods and technologies in computational toxicology to evaluate the safety of pharmaceutical compounds and environmental substances. There is growing interest in it as a technique for speeding up and improving the accuracy of in silico drug discovery screening. Currently, in one field of toxicology, the use of technology and knowledge in various research fields such as chemistry, medicine, computer science, mathematics, statistics, and informatics is becoming active. In other words, computational toxicology is a fusion of wet and dry studies to estimate the hazards or effects of exposure to chemicals. The computational toxicology method uses a wide variety of analytical methods, but is often machine learning. To perform a computational toxicity assessment using the chemical structure of the components contained in foods is the same as ascertaining a quantitative structure-activity relationship (QSAR) using the information of the chemical structural formula as a parameter. Evaluation by computational toxicology is energetically underway for environmentally hazardous substances and drug candidate substances (1).

However, usefulness, side-effect information or toxicity information on nutritional components in the daily diet has not been comprehensively established. The reason for this is that it is a complex system with multiple components. We consider that this problem can be solved quickly and exactly by positively using the in-
tigated approaches with wet and dry testing. In this symposium, we will introduce the contents of four examples in these integrative approaches. Through detecting and predicting the biological activity and toxicity of chemical substances such as nutrients in foods, it will be possible to provide better molecular information on dietary components.

**Chemoprevention by vitamin A and its derivatives**

Hepatocellular carcinoma (HCC) is a deadly cancer with increasing numbers of global deaths in the past 20 y. Progressive loss of total hepatic retinoid storage has been associated with the development of hepatic diseases including HCC. Acyclic retinoid (ACR) is an orally administrated vitamin A derivative that binds to cellular retinoic acid-binding protein with an affinity equal to that of all-trans-retinoic acid, selectively inducing retinoic acid receptor-β-dependent signaling. ACR is the first agent to show promising efficacy and safety in phase 2/3 trials for the prevention of HCC recurrence in patients who underwent surgical removal of the primary tumors. We performed a genome-wide transcriptome screen and identified that ACR suppressed the expression of a proto-oncogene MYCN selectively expressed in HCC cells but not in normal hepatic cells (2–4). Data mining of clinical datasets showed that MYCN expression in HCC was correlated positively with both cancer stem cell (CSC) and Wnt/β-catenin signaling markers, but negatively with mature hepatocyte markers. Proteome and metabolome analyses showed the expression of fatty acid desaturases and the content of unsaturated fatty acids were increased in MYCN high expression EpCAM+ CSC-like HCC cells. Inhibition of lipid desaturation using either the chemical inhibitor or siRNA/shRNA against stearoyl-CoA desaturase-1 (SCD1) suppressed cell proliferation as well as MYCN gene expression in HCC cells. In the cohort studies in Japan, Taiwan and Europe, MYCN gene expression was expressed at higher levels in HCC tumor than in non-tumor tissues. There was a significant positive correlation between MYCN expression and recurrence of HCC with a single tumor but not with multiple tumors. In a retrospective analysis of a phase 3 trial of ACR, serum MYCN was identified as the risk factor most associated with HCC recurrence. HCC patients with lower serum MYCN levels after a 4-wk treatment period had a significantly lower risk of recurrence in the ACR group, but not in the placebo group.

**Toxicity prediction of natural compounds from Kampo medicine**

Kampo is a general term for traditional medicine that originated in ancient China and later developed uniquely in Japan. Most Kampo formulas are composed of multiple crude drugs, so it is a complex multi-component system containing many chemical substances. It is difficult to identify the components behind the side effects of Kampo, and one solution to this problem is a computational toxicity assessment. We have comprehensively investigated information on the components of Kampo medicine and their constituent crude drugs that are frequently used in Japan. From the chemical structure information obtained, we next tried to estimate the test results related to safety by machine learning in in silico experiments (5, 6). There were 55 target crude drugs, and the final analysis targets were 920 compounds. Of the results of several in silico toxicity predictions, we specifically focused on the interaction with the human Ether-a-go-go Related Gene (hERG) K+ channel. The prediction of hERG inhibition used a classification model constructed by ANNE (Artificial Neural Network) (7). Of the 920 compounds, those predicted to be positive or negative as blockers for the hERG K+ channel were 53 and 865 compounds, respectively. Examples of predicted hERG blocker include isomagnol and obovatol contained in M. obovata. Both isomagnol and obovatol show inhibitory activities with a high probability. On the other hand, obovatal is also contained in M. obovata and it was classified as negative with a high probability. The differences in the partial structure of those three compounds are slight. However, the impact on hERG inhibition was strong, and it showed important information for the structure-activity relationship search. Eleven crude drugs containing compounds with inhibitory activities for the hERG K+ channel were indentified. Currently, we are constructing a database that integrates this crude drug information and the calculated toxicity results.

**Toxicity prediction of chemicals using pluripotent stem cells**

Human pluripotent stem cells such as embryonic stem (ES) and induced pluripotent stem (iPS) cells, combined with sophisticated bioinformatics methods, are powerful tools to predict developmental chemical toxicity. We also observed that qRT-PCR and molecular descriptors tend to contribute to specific toxicity categories. Here we describe how to apply machine learning techniques to different types of data, such as qRT-PCRs, gene networks, and molecular descriptors, for toxic chemicals, as well as how to integrate these data to predict toxicity categories. Deciphering the key mechanisms of morphogenesis during embryonic development is crucial to understanding the guiding principles of the body plan and promote applications in biomedical research fields. Although several computational tissue reconstruction methods using cellular gene expression data have been proposed, those methods are insufficient with regard to arranging cells in their correct positions in tissues or organs unless spatial information is explicitly provided. Here, we report SPRESSO, a new in silico three-dimensional (3D) tissue reconstruction method using stochastic self-organizing map (stochastic-SOM) clustering, to estimate the spatial domains of cells in tissues or organs from only their gene expression profiles (8–10). With only five gene sets defined by Gene Ontology (GO), we successfully demonstrated the reconstruction of a four-domain structure of a mid-gastrula mouse embryo (E7.0) with high reproducibility (success rate=99%). Interestingly, the five GOs contain 20 genes,
Disclosure of state of COI

No conflicts of interest to be declared.

REFERENCES


