

[SC3] Electronic-Structure Informatics Combined with Linguistic Descriptions

Manabu Sugimoto¹⁻³, Yusei Kosugi², Hitomi Kuga²

¹ Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, 860-8555, Kumamoto, Japan

² Department of Applied Chemistry and Biochemistry, Faculty of Engineering, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, 860-8555, Kumamoto, Japan

³ Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, 153-8904, Tokyo, Japan

The electronic-structure calculation is a way to describe molecules/materials numerically. The obtained numerical data are expected to be useful for analysis, discovery, design, and predictions of functional molecules and functions of molecules. Chemoinformatics based on electronic-structure calculations may be called “electronic-structure informatics”. This approach seems promising since the electronic-structure data can play a pivotal role to establish structure-property relationships. This would be reasonable because both structure and property a molecule depend on its electronic ground and excited states. Electronic-structure informatics is also expected to unveil underlying physical/chemical/biological mechanisms behind molecular properties as have been demonstrated by numerous quantum chemistry researched over many years. Versatile applications of concepts and technologies in chemoinformatics are expected to appreciably extend usefulness and possibility of electronic-structure calculations.

Here we would like to report our two studies on electronic-structure informatics: one is the identification of natural skin-whiteners contained in plants, and another is identification on smell of molecules. These projects have been carried out through electronic-structure calculations and unsupervised learning for classification on the basis of electronic similarity. In addition, we have referred to linguistic description (keywords) of the chemicals and the plants. We think that such a combination with linguistic data would also be very important for classification, molecular search, repositioning of molecules, and predictions of new molecules and new applications. This seems natural since, when we carry out computational analysis and predictions on molecular properties, we need to refer to linguistic descriptions for determining strategies for computations and obtaining conclusions after computations. The present research is an elementary trial toward applications of cognitive computing for electronic-structure informatics.