Contribution ID: **18** Type: **not specified**

Discovering Novel Fluorescent Molecules by Combining Machine-learning and Global Optimization

Thursday, 7 November 2019 13:50 (20 minutes)

Fluorescent molecules are widely used for bio-imaging. They are attached to specific cell organelles and/or proteins, enabling observation of detailed structure and dynamics in the cell. Efficient fluorescent molecules must have a high quantum yield for effective bio-imaging. Here, we present a systematic approach to discovering novel fluorescent molecules that combines machine-learning and global optimization algorithms. We recast the problem of discovering novel fluorescent molecules with high-intensity emission light into a global optimization problem by using the oscillator strength of a molecule as an objective function for optimization.

A statistical machine that predicts excitation energies and associated oscillator strengths, the probability of absorption or emission of light in transitions between different energy states, of a molecule were trained using the random forest algorithm. The Pub-chemQC database, which contains TD-DFT calculation results of 3.8 million known molecules, was used as a training set. The extended connectivity fingerprints of molecules were used as input vectors. To optimize the oscillator strength of a molecule, a highly efficient global optimization algorithm called CSA was used. For CSA global optimization, SMILES representation of a molecule was mapped to a 200-dimensional integer vector by using Natural Language Toolkit. After CSA global optimization calculation converged, we assessed the validity of our approach by performing quantum mechanical calculations. TD-DFT calculations were carried out to verify whether novel molecules obtained by this procedure actually have high oscillator strength.

Presenter: Prof. LEE, Juyong (Kangwon National University)

Session Classification: Invited talks - Chemistry 1