Population-based de novo Molecule Generation, Using Grammatical Evolution

Naruki Yoshikawa*[1], Kei Terayama[1, 2, 3], Masato Sumita[2], Teruki Homma[4], Kenta Oono[5], and Koji Tsuda[1, 2, 6]

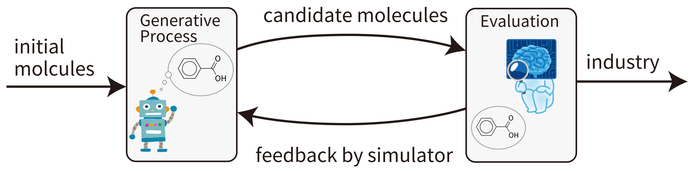
1. The University of Tokyo, 2. RIKEN Center for Advanced Intelligence Project, 3. Kyoto University,
4. RIKEN Systems and Structural Biology Center, 5. Preferred Networks, Inc., 6. National Institute for Materials Science

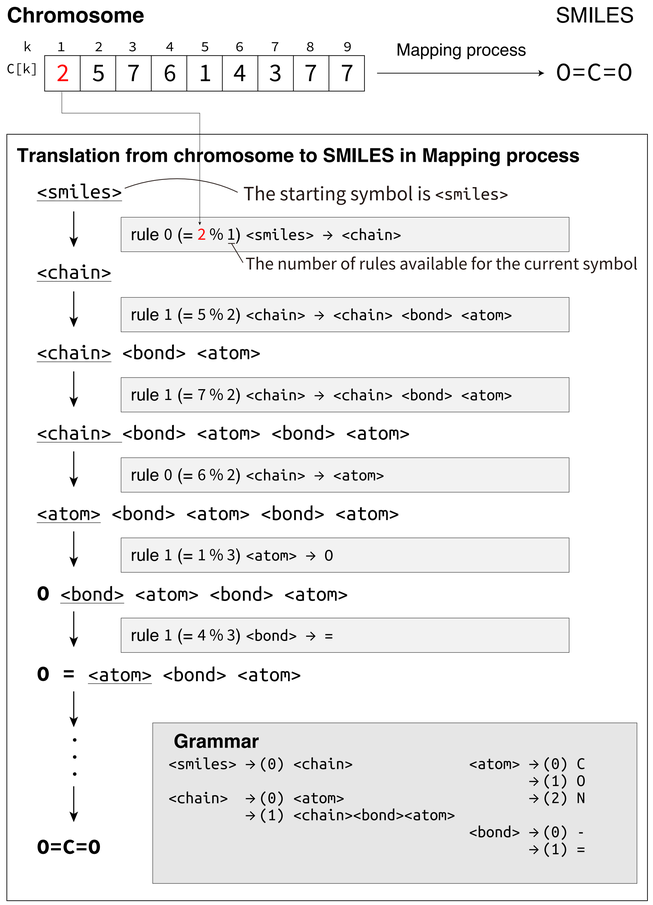
* E-mail: naruki-yoshikawa@g.ecc.u-tokyo.ac.jp

Introduction

■ Automatic Molecular Design

Automatic molecular design is conducted by a combination of molecule design by generative process and evaluation by simulators or machine learning models.





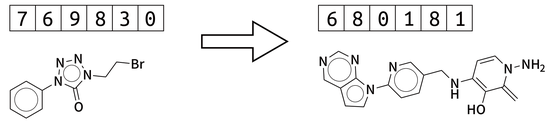
Method

■ Grammatical Evolution

In grammatical evolution, a molecule is doubly encoded:

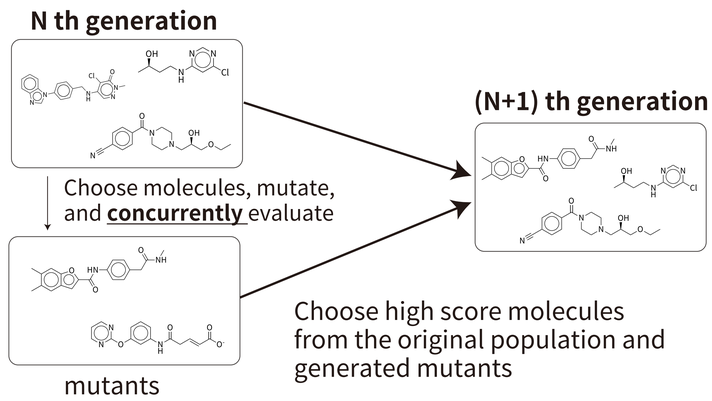
- chromosome — an array of integers
- SMILES — a string representing molecule

SMILES is obtained from a chromosome through the mapping process. Molecules are optimized by optimizing chromosomes using an evolutionary strategy.



■ Evolutionary Strategy

A population of molecules is optimized by the following evolutionary strategy.

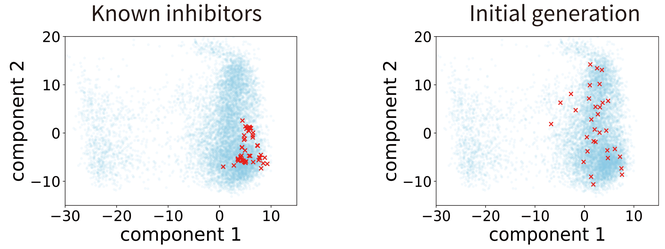


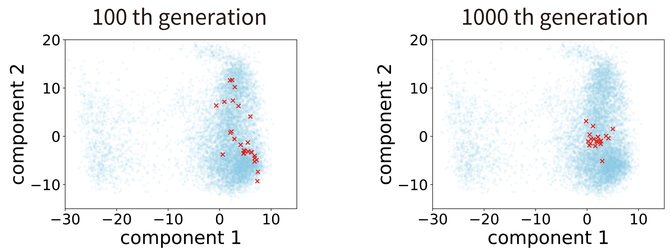
■ Mapping Process

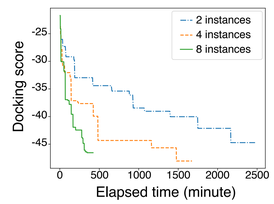
SMILES string is generated by applying grammar rules specified by integers in a chromosome to the leftmost remaining non-terminal symbol. The left figure illustrates how this translation is conducted.

Results

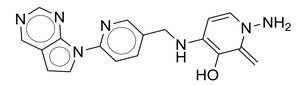
- We optimized the sum of docking score (representing interaction with thymidine kinase, calculated by rDock) and the synthetic accessibility score.
- We found new molecules whose scores are better than known inhibitors. We used isomap to visualize molecules in 2D space.







Acceleration by concurrent simulation



An example of molecules in the 1000 th generation.

Conclusion

We developed a new molecular generator using grammatical evolution. This work demonstrated that molecule generation is possible without costly deep learning and showed a new direction for research.

Our paper is available at https://arxiv.org/abs/1804.02134