Analysis of Chemical Modification Patterns Extracted from KEGG DRUG Structure Maps

Daichi Shigemizu[1], Michihiro Araki[2], Susumu Goto[1], Minoru Kanehisa[1 3]

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan,
[2]Education Unit for Global Leaders in Advanced Engineering and Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan, and
[3]Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan

Currently available drugs have been mostly derived from prototypic compounds found through empirical screening. In the history of drug development, medicinal chemists have continuously introduced modifications around core chemical structures found in the prototypic compounds to improve efficacy or to apply to different therapeutic categories. Chemical and biological functions of drugs highly depend on the combinations of limited numbers of conserved core structures and diverse fragments around them. From this perspective, it would be useful to collect and computerize the knowledge of medicinal chemists on the chemical modifications in the drug development process. As part of the KEGG project, the KEGG DRUG structure maps have been developed in order to capture this knowledge. In the previous work (Shigemizu et al., J. Chem. Inf. Model. Vol.49, Num.4, 2009), we extracted 125 chemical modification patterns from 255 drug pairs in the structure maps. Here, we increased the number of drug pairs by selecting matching fragment pattern changes in the entire KEGG DRUG database, and developed a computational tool to generate modified structures. This tool was applied to representative 27 starting compounds in the structure maps, producing 123 drug candidates on average per starting compound. We then examined whether at least one of the drug candidates is included in each structure map. As the result we found that 19 out of 27 starting compounds were correctly converted to the existing drugs. This tool was further applied to the reconstruction of structure maps by applying multiple steps of conversions. It is expected that the computational tool presented here will be applicable to future drug development.