Chemoinformatic Approaches to Quantifying Structural Complexity and Diversity

Timothy A. Flood, Jr. and Paul J. Hergenrother

High-throughput screening (HTS) is the primary method for the discovery of lead compounds in the development of novel drug therapeutics. However, there is a growing concern in the pharmaceutical research community surrounding the content of standard screening collections and the implications for future discovery efforts. Most compounds identified through HTS efforts possess high aromatic/sp^2 content and few, if any, stereogenic centers. While such compounds lead to effective drugs against a certain portion of targets (e.g. kinases), they are chemically unsuitable for modulation of the majority of protein targets. In an effort to address this deficiency, we have launched a new synthetic approach to drug discovery aimed at the rapid generation of complex and diverse scaffolds.

As we began this project we quickly encountered a philosophical problem: how exactly does one quantify complexity? And how do we assign a numerical value to structural differences that most organic chemists simply know intuitively? To answer these questions, chemoinformatic methods were employed: first, it was necessary to assess whether the compounds generated were sufficiently *complex* relative to standard commercial screening collections; at the same time, it was necessary to determine whether significant *diversity* of scaffolds could be rapidly achieved. Using a combination of molecular property distribution profiling and Tanimoto similarity analyses, it was demonstrated that the structures created by our methods occupy regions of chemical space distinct from current collections while still possessing high internal variability.