Searching in synthon-based chemical space using shape similarity as the ranking criteria with Shape-Aware Synthon Search (SASS)

Chen Cheng[1]

[1]*Genentech, Inc, 1 DNA Way, South San Francisco, CA 94080*

Summary:

- Searching in unenumerated library space using shape and chemical feature similarity
- Wide variety of queries, including cyclic molecules
- Increases search efficiency by around 1000-fold

Product Keywords: OMEGA, ROCS

Abstract:

Accessible virtual libraries continue to expand rapidly; at the time of writing Enamine and WuXi's collections both contain over 10 billion molecules, and they will only continue to grow. Simply enumerating and storing collections of this size is challenging, while brute-force searching such large spaces is becoming prohibitively expensive and time-consuming. Therefore, there is a growing interest in search methods that operate in unenumerated or fragment space and several approaches to this problem have been published recently.

Shape-Aware Synthon Search (SASS) is Genentech's solution for searching unenumerated spaces at the fragment or synthon level, utilizing OpenEye's shape similarity tool ROCS. SASS takes a query molecule, fragments it and searches a database of reagents (or synthons) for matching fragments in shape and chemical features using customized definitions of important chemical features. The approach can handle both acyclic and cyclic molecules equally well, while most existing methods cannot process cyclic molecules, greatly expanding the types of query that can be used with SASS.
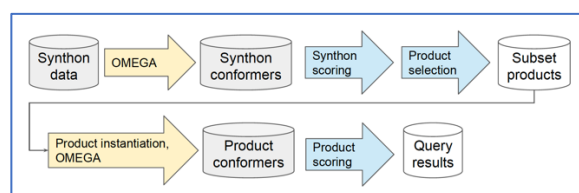
Searching ultra-large unenumerated chemical spaces is of increasing interest in pharmaceutical discovery. Shape-Aware Synthon Search (SASS) is a ROCS-based search tool developed at Genentech for searching unenumerated multi-billion molecule spaces up to 3 orders of magnitude faster than brute-force search.

Comparison of brute-force searching and SASS shows that SASS is able to identify 70-90% of the hits found by searching the complete library while searching less than 0.5% of that library, resulting in substantial reduction in run-time; for a 230M library brute-force searching requires around 43,000 CPU hours, while SASS requires only 50. Also, SASS has been scaled up to libraries of 5B molecules, indicating that searching libraries of greater than 10^{10} is now within reach.

Source: OpenEye miniCUP Presentation, October 2023, San Francisco

info@eyesopen.com