Library Design

February 7, 2006

Guha Jayachandran
(guha@stanford.edu), CS379A

Two Cases

- Choose what existing known compounds to screen
  - Basically an early stage of screening in the way we've talked about before
- Choose reagents to build a combinatorial library

Two Divergent Aims

- Diversity
  - This is combinatorial chemistry's original motivation
  - Can be measured in various ways
  - Cluster or partition compounds and take samples from each group
- Focus
  - Especially if you have information on structure
  - If protein structure, then can do docking. If know active ligands, then can use similarity searching, etc.

Balance between two depends on amount of information available

Combinatorial Libraries

- Monomers/reagents combine together to form compounds
  - Some set of reagents allowed at each position of variability along scaffold. Make all possible combinations.
- Two main computational ways for selecting reagents
  - Reagent based - choose reagents from each pool separately (whether based on diversity, focus, etc.)
  - Product based - choose reagents based on the products yielded

Reagent Based and Product Based

- Product based selection is
  - Better because doesn't assume that variation sites on scaffold are independent
  - Slower because many more combinations when considering reagent combinations across variation sites together
  - More effective when aiming for some property across library (diversity, focus, etc.)
- Both have been used

Product Based Selection Steps

- Input: Reagents for each position
- Output: Library of reagents to use

1. Enumerate compounds
2. Determine descriptors for compounds
3. Screen compounds (perhaps with docking)
4. Select reagents from previous step's hits, optimizing some criteria (diversity, etc.)
   - Either respecting combinatorial constraint (every reagent chosen for a position can react with the reagents chosen for other positions) for the sake of synthetic efficiency, or not (termed "cherry picking")

Multiobjective Library Design

- Optimize multiple properties at once (like cost, physiochemical properties, etc.)
- Multiobjective Genetic Algorithm (MOGA) is one method
  - Yields Pareto optimal solutions (where no other solutions exist that are better in all objectives)

Readings

- Combinatorial Library Design Using a Multiobjective Genetic Algorithm (Gillet, et. al.)
- Novel Dihydrofolate Reductase Inhibitors. Structure-Based versus Diversity-Based Library Design and High-Throughput Synthesis and Screening (Wyss, et. al.)
- Luddite: An Information-Theoretic Library Design Tool (Miller, et. al.)