Instruction for designing ionizable lipids generative model with MCTS

Overview

- Number of initial head building blocks: **200K**
- Number of initial tail building blocks: **2K**

Ionizable lipid generation rules

- Number of tails: **2–4**
- Maximum number of distinct tail types: **2**
- Preference for **identical tails**
- Preference for **symmetric structures**

General Workflow

1. **Level 1**
   Randomly select head building blocks.
   Molecules: `{head A}`

   (Since the Synthemol building blocks input file does not distinguish head and tail building blocks and contains them in a single file, either two separate input files for head and tail building blocks should be created, or head and tail building blocks should be distinguished via indexing within a single file.)

2. **Level 2**
   Load 2K tail building blocks and, based on reaction templates, extract only the tail building blocks that can react with the head selected at Level 1.

   (As in Synthemol, it may be convenient to pre-filter head and tail building blocks that can react with a specific reaction ID in `reaction_to_building_blocks.pkl`.)

3. **Level 2 (continued)**
   Randomly select one tail building block (tail A) from the filtered set and combine it with the selected head building block using the reaction template.
   Molecules: `{head A, tail A}`
4. **Level 3**
   Generate intermediate products with one tail attached.
   Assume up to four possible products can be generated (Product A, B, C, D).
   Molecules: `{Product A}`
5. **Level 4**
   Using the reaction template, extract only the tail building blocks that can react with Product A, then randomly select a tail (tail A').
   Assign a higher weight when `tail A' = tail A` (preference for identical tails).
   Molecules: `{Product A, tail A'}`

6. **Level 5**
   Combine the intermediate product from Level 3 (Product A) with tail A’ to generate an intermediate product with two tails attached.
   Molecules: `{Product A’}`
7. **Level 6**
   - If tail A (selected at Level 2) and tail A’ (selected at Level 4) are different, select either tail A or tail A’.
     Molecules: `{Product A’, tail A or tail A’}`
   - If tail A and tail A’ are identical, use the reaction template to extract tail building blocks (tail A’’) that can be attached to Product A’, then randomly select one.
     Molecules: `{Product A’, tail A’’}`
8. **Level 7**
   Generate an intermediate product with three tails attached.
   Molecules: `{Product A’’}`
9. **Level 8**
   - If the three previously selected tails consist of two distinct tail types, select one of the two types.
     Molecules: `{Product A’’, tail A or tail A’ / tail A or tail A’’}`
   - If all three previously selected tails are identical, use the reaction template to extract tail building blocks (tail A’’’) that can be attached to Product A’’, then randomly select one.
     Molecules: `{Product A’’, tail A’’’}`
10. **Scoring**
    Input the final molecules (Product A’’’) obtained at Level 8 into the prediction model to calculate scores.
    Molecules: `{Product A’’’}`
11. **Node Selection (Synthemol-style UCB)**
    If following the Synthemol UCB (node) calculation method, the prediction model score obtained from intermediate products can be used as the `P(node)` value for node selection.

Version with Fully Identical Tails Attached to the Head

1. **Level 1**
   Randomly select head building blocks.
   Molecules: `{head A}`

   (As the Synthemol building blocks input file does not distinguish head and tail building blocks, either two separate input files should be created or head and tail building blocks should be distinguished by indexing within a single file.)

2. **Level 2**
   Load 2K tail building blocks and, based on reaction templates, extract only tail building blocks that can react with the head selected at Level 1.

   (As in Synthemol, pre-filtering head and tail building blocks for each reaction ID in `reaction_to_building_blocks.pkl` may be useful.)

3. **Level 2 (continued)**
   Randomly select one tail building block (tail A) and combine it with the selected head building block using the reaction template.
   Molecules: `{head A, tail A}`
4. **Level 3**
   Generate an intermediate product with one tail A attached.
   Molecules: `{Product A}`
5. **Level 4**
   Generate an intermediate product with two tail A units attached.
   Molecules: `{Product A'}`
6. **Level 5**
   Generate an intermediate product with three tail A units attached.
   Molecules: `{Product A''}`
7. **Level 6**
   Generate an intermediate product with four tail A units attached.
   Molecules: `{Product A'''}`
8. **Scoring**
   Input the final molecules (Product A''') obtained at Level 6 into the prediction model to calculate scores.
   Molecules: `{Product A'''}`