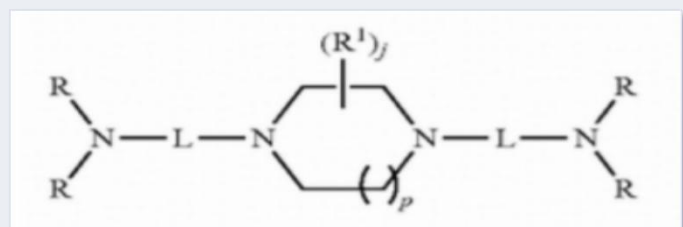




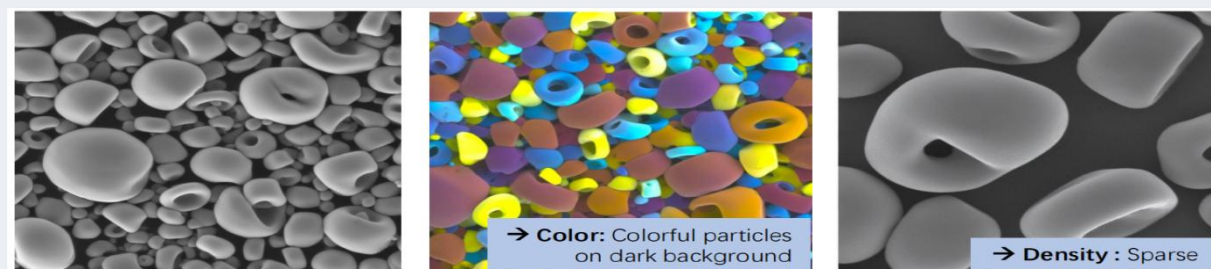
# Synthetic Generation

e.g. E-SMILES format representation



```
*N(*)*N1CCN(*N(*)*)CC1<sep>
<a>0:R</a><a>2:R</a>
<a>3:L</a><a>6:CH[2]?p</a>
<a>8:L</a><a>10:R</a>
<a>11:R</a><r>0:R[1]?j</r>
```

e.g. Generate electron micrograph image



Filter

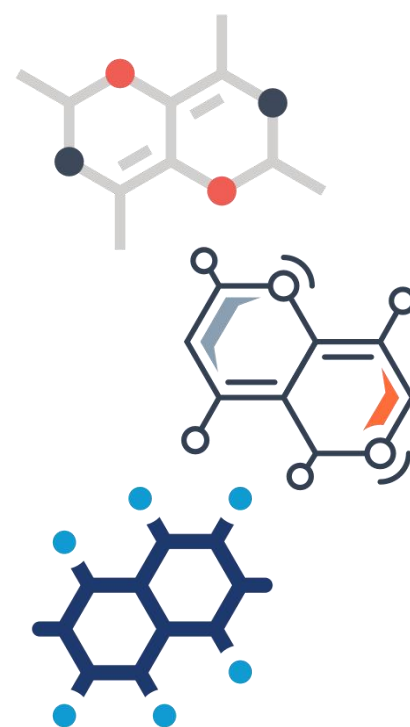
**For EM representation:**

1. Ensure chemical validity.
2. Check high-value complex molecules.

**For image:**

1. Verify image quality.

# Refining Loop



Scientist in the loop

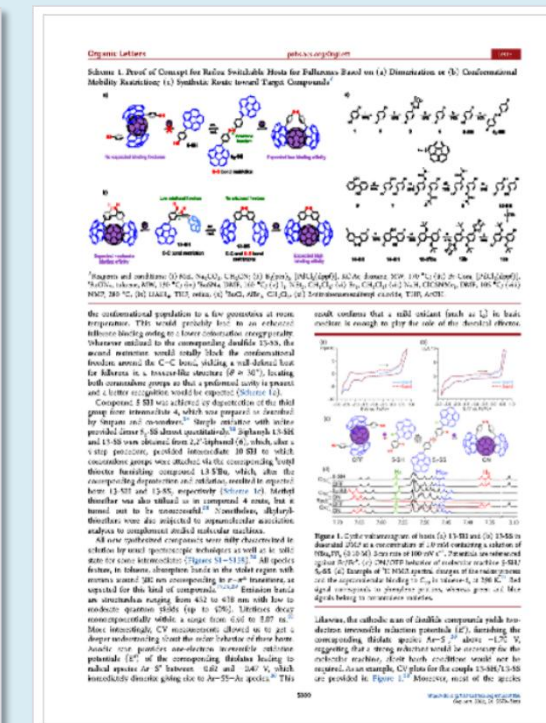
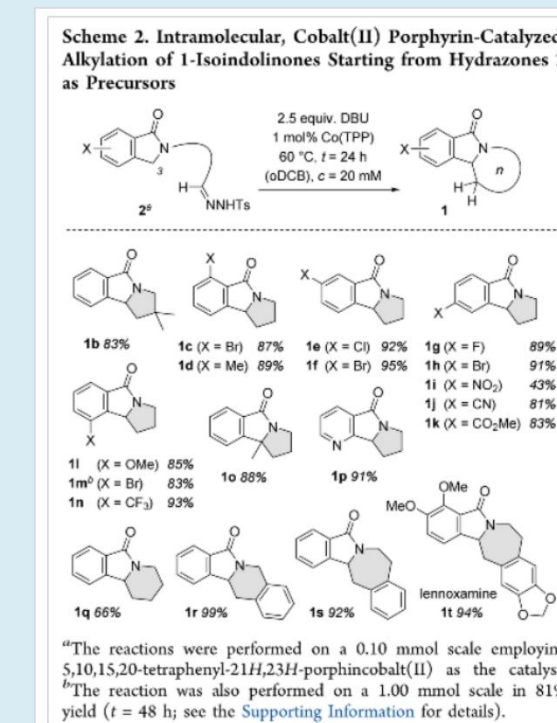


# Real-World Sources

Papers

EM images

Patents



Generate Questions

**For questions:**

1. Review and edit.
2. Create hard negatives.