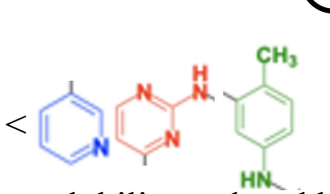
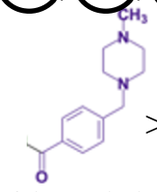
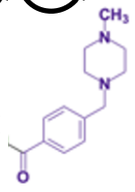
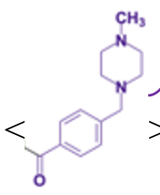
Modular Chemical-Language Model

In imatinib   , the group  acts as activation loop binder and promotes solubility and oral bioactivity.

3. Reasoning with Function & Synthesis Awareness

Functional Group-Informed Reasoning



[Methyl, **Piperazine**, Benzene, Carbonyl]

Functional group detection

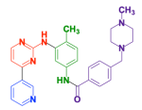
the group < > … ***promotes solubility*** …

Property-functional group mining

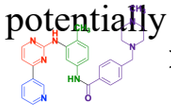
[Hydroxyl, Carboxyl, Amino, ..., **Piperazine**, Pyrrole, …]

2. Training Corpus

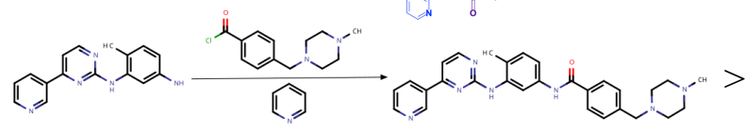
1. Molecule property/function description

Imatinib  is a tyrosine kinase inhibitor that specifically targets the BCR-ABL protein, effectively blocking cell proliferation in certain cancers, particularly chronic myeloid leukemia.

2. Molecule proposal

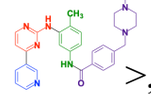
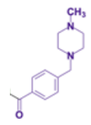
For improved selectivity and potency against resistant cancer cell lines, potentially enhancing efficacy in cases of drug-resistant leukemia, we can use the following molecule: <  >

3. Synthesis prediction

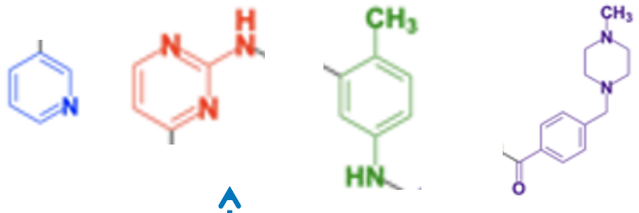
A last-step synthesis reaction of imatinib is: <  >

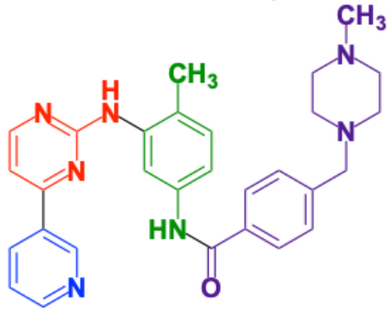
4. Function group description (few)

…

In imatinib <  >, 4-(N-Methylpiperazinyl)methyl benzoic acid group <  > acts as activation loop binder and promotes solubility and oral bioactivity.

1. Input Representation



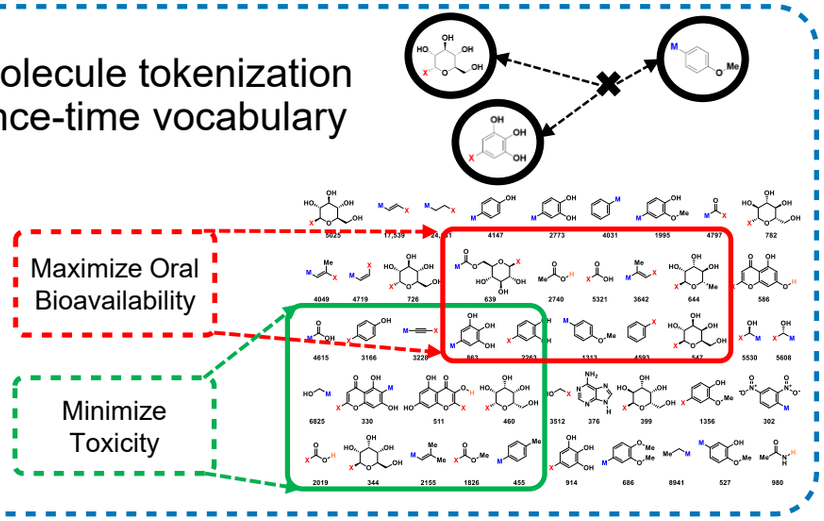


Retrieval-based molecule tokenization for flexible, inference-time vocabulary

Genetic algorithm for function-aware vocabulary selection

Maximize Oral Bioavailability

Minimize Toxicity



[19, 18825, 45, 6189, 53546, 79, 29, …]

“4“ “-(” “N” “-M” “ethyl” “p” “iper” “azin”…

text tokenization

4-(N-Methylpiperazinyl)methyl benzoic acid group acts as activation loop binder and promotes solubility and oral bioactivity in imatinib.