

Задача подструктурного поиска химического соединения

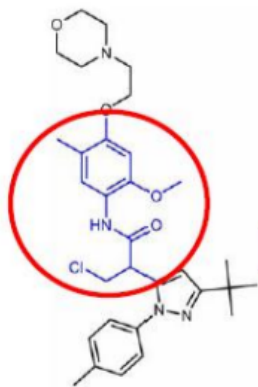
1 Введение

Достижения в синтетической химии приводят к тому, что молекулы, синтезируемые в настоящее время, состоят из более сложных объектов с механическими связями и более обширных каркасов. Все более важным становится вопрос о том, как пользователи могут эффективно искать такие структуры в больших базах данных.

Задача поиска химических соединений, содержащих заданный фрагмент, в больших базах данных также является одной из задач в процессе разработки лекарств и позволяет решать конкретные проблемы при разработке новых лекарств. Рассмотрим на конкретном примере.

Эта молекула является активной, но не может быть использована по причине:

- фармакологических проблем (ADMET - absorption, distribution, metabolism, excretion and toxicity)
- защищенности патентом

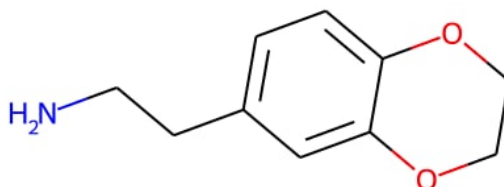


Мы знаем что активность данной молекулы обусловлена выделенным фрагментом.

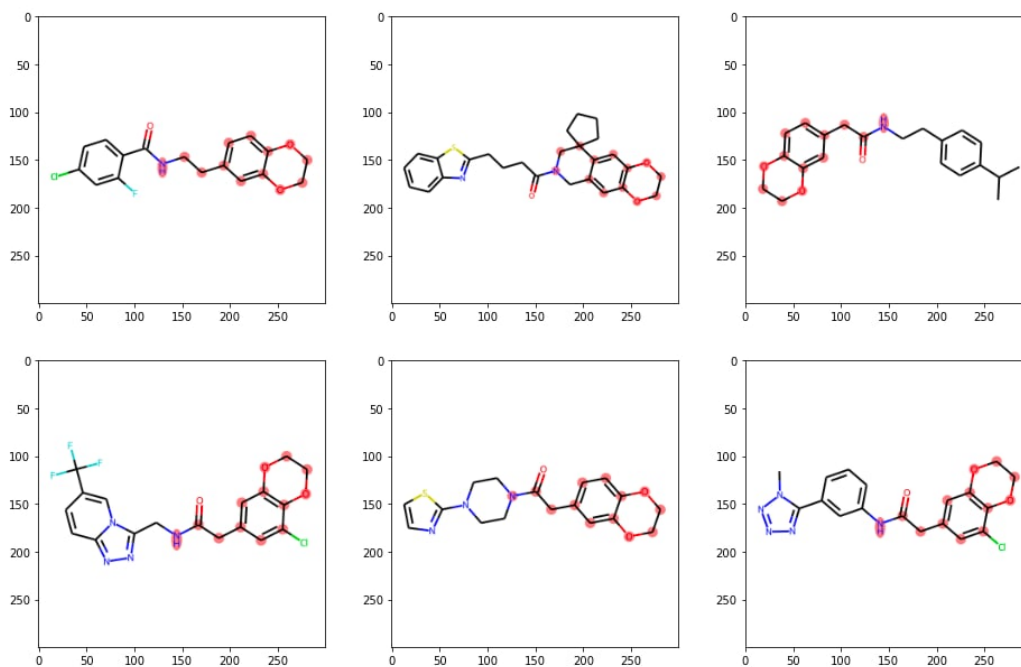
Значит существует вероятность найти активные молекулы среди ее производных, содержащих данный фрагмент.

2 Постановка задачи

Рассмотрим химическое соединение C :



Требуется разработать алгоритм, который будет находить все такие химические соединения C' , в которых C является подструктурой C' :



На данный момент база данных состоит из 10^9 химических соединений. Ограничение на время поиска 1 секунда.