About

The formulation development stage in drug discovery is often time-intensive and resource-demanding due to the limited knowledge of the physicochemical properties of novel molecules. Traditionally, this process relies on iterative trial-and-error experiments to identify the optimal formulation, which can significantly delay timelines. Our platform addresses this challenge by enabling users to search using an API name, formulation name, or SMILES representation and instantly retrieve similar APIs or formulations. This functionality provides a scientifically informed starting point for developing novel molecules, reducing experimental redundancy and accelerating R&D workflows.

API surrogate search engine

Following is the stepwise process followed to identify, annotate & extract the information of interest:

Approved drugs, formulation, and safety data

Data was collected from well known drug regulatory agencies like FDA. This data provides information of each approved formulation till 21 April 2025.

1. DATA COLLECTION:
    - Formulation related data collected from the DailyMed an U.S. Food and Drug Administration's repository.
    - Data related to other experimental properties like melting point, pKa, etc. were collected/ extracted from multiple public sources
2. NATURE OF DATA:
    - Data in dailymed was present in the form of Structured product label (SPL). Total of 1532 SPL catalog files recovered. Each SPL catalog file consists of the brand name, along with a description of the formulation like API, inactive ingredients, indication, dosage form, etc and a "SETID" is provided to identify the SPL file with maximum 100 SETIDs in each SPL catalog file.
    - The experimental properties were available either as web-page, database engine, publications, datasets, csv's, etc.
3. SEARCH ALGORITHM:
    - Current version works with;
        - 'text' based search, users can search with the name of API or by brand name.
        - 'SMILES' search, users can search with SMILES (canonical/isomeric). SMILES based search uses tanimoto similarity