**Project Title**

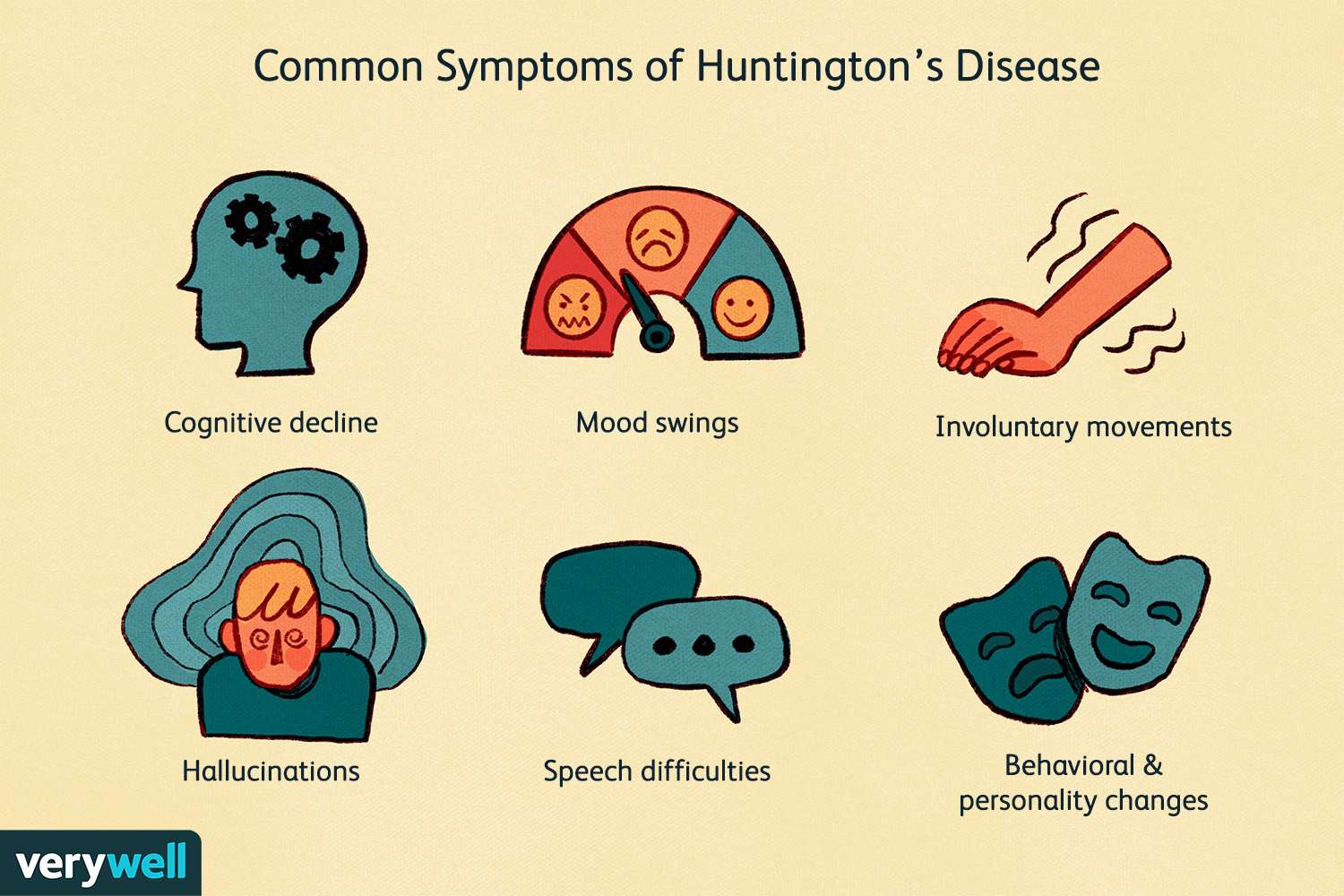
**2D and 3D Structural Similarity Analysis of Tetrabenazine Using KNIME Analytics Platform**

**Introduction**

Huntington's disease (HD) is a progressive neurodegenerative disorder caused by the expansion of a CAG trinucleotide repeat in the gene encoding the huntingtin protein. This mutation results in an abnormally long polyglutamine (polyQ) tract, which leads to protein misfolding, aggregation, and neuronal death. The symptoms of HD include motor dysfunction, cognitive decline, and psychiatric issues, with chorea being a hallmark feature. Tetrabenazine, a drug that depletes monoamine neurotransmitters such as dopamine, is commonly used to manage chorea in HD patients. However, it is often associated with severe side effects like depression and motor issues, which limit its use in the long term.

Due to these drawbacks, there is an ongoing search for alternative treatments that can effectively manage HD symptoms with fewer adverse effects. Structural similarity analysis plays an essential role in drug discovery by identifying new drug candidates that share similar chemical and spatial characteristics to existing drugs. This project combines **2D fingerprint-based analysis** and **3D conformational analysis** to explore potential analogs of tetrabenazine using a library of polyQ aggregation inhibitors. The aim is to identify compounds that could serve as alternatives for HD treatment.

The 2D analysis focuses on comparing chemical structures by evaluating molecular fingerprints, while the 3D analysis examines the spatial orientation of molecules. Together, these methods provide a more comprehensive understanding of how a compound’s structure may influence its function, making them invaluable tools in drug discovery.



**Materials and Methods**

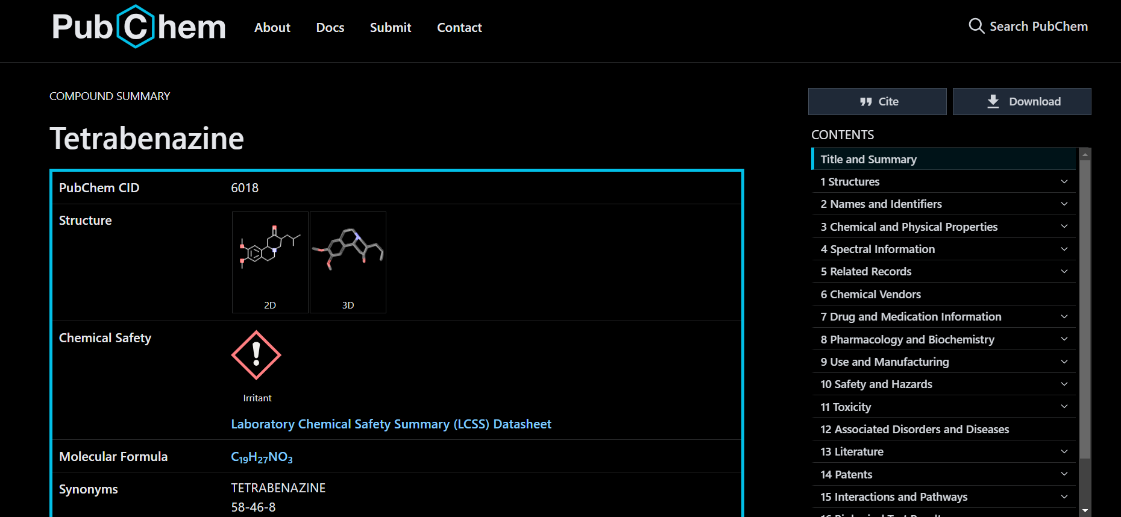
**Focus Area and Drug Selection**

The research centers on **Huntington's disease**, a disorder characterized by the expansion of the polyglutamine tract in the huntingtin protein.

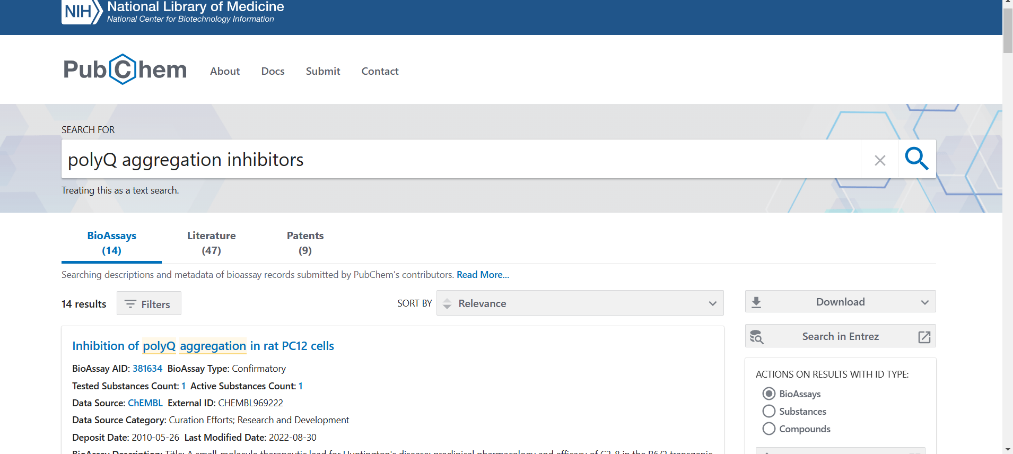
* **Primary Drug:** Tetrabenazine, a medication used to alleviate chorea in HD patients, was selected for its relevance in treating HD symptoms, despite its side effects.
* **Drug Library:** The study also explores **polyQ aggregation inhibitors**, which are compounds known to target polyglutamine aggregation—a critical mechanism in HD progression.

**Data Collection**

1. **Tetrabenazine Structure:**  
   The 2D and 3D structures of tetrabenazine were sourced from **PubChem** in **SDF format**, which provides detailed information on molecular structures and atomic connections.



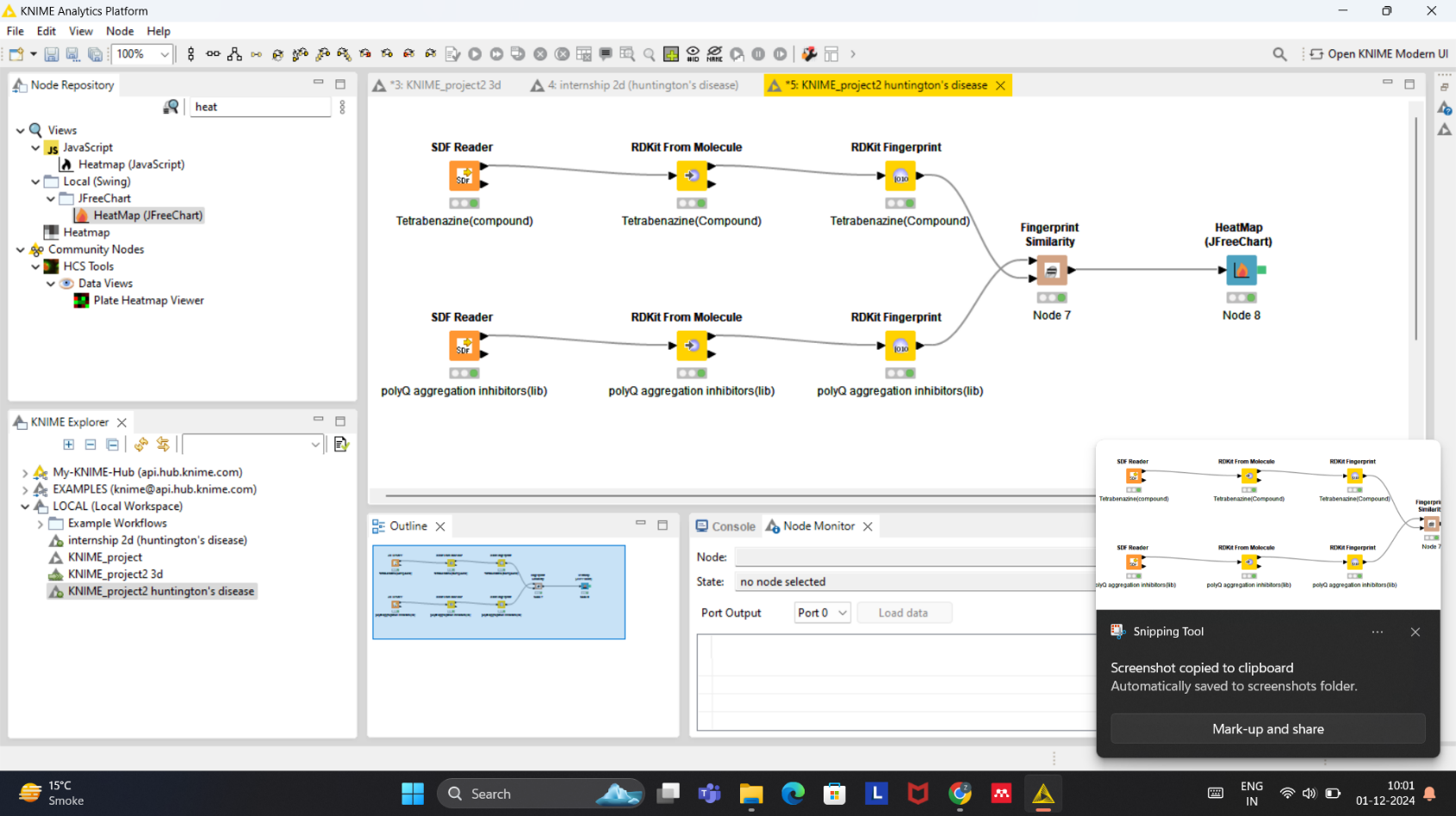
1. **PolyQ Aggregation Inhibitors:**  
   A dataset of known polyQ aggregation inhibitors was compiled from literature and PubChem, focusing on compounds that have been shown to interfere with polyQ aggregation, a key pathological feature of HD.



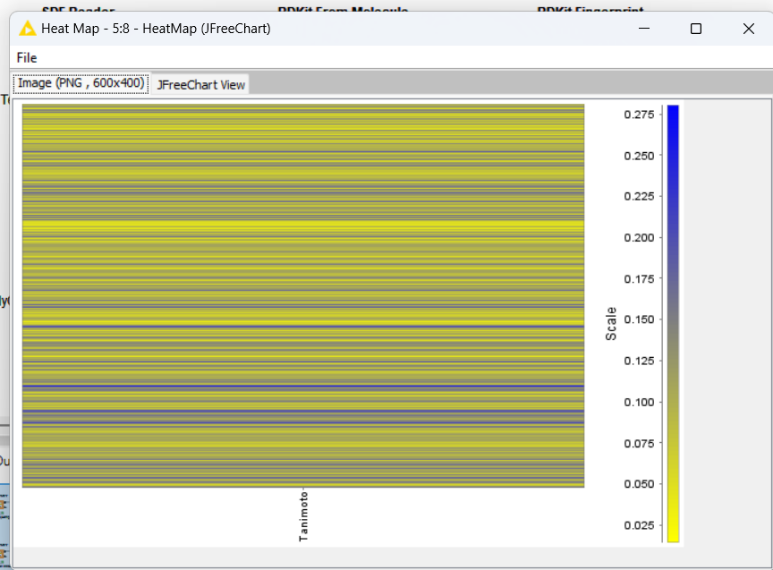
**Workflow Steps**

**2D Similarity Analysis**

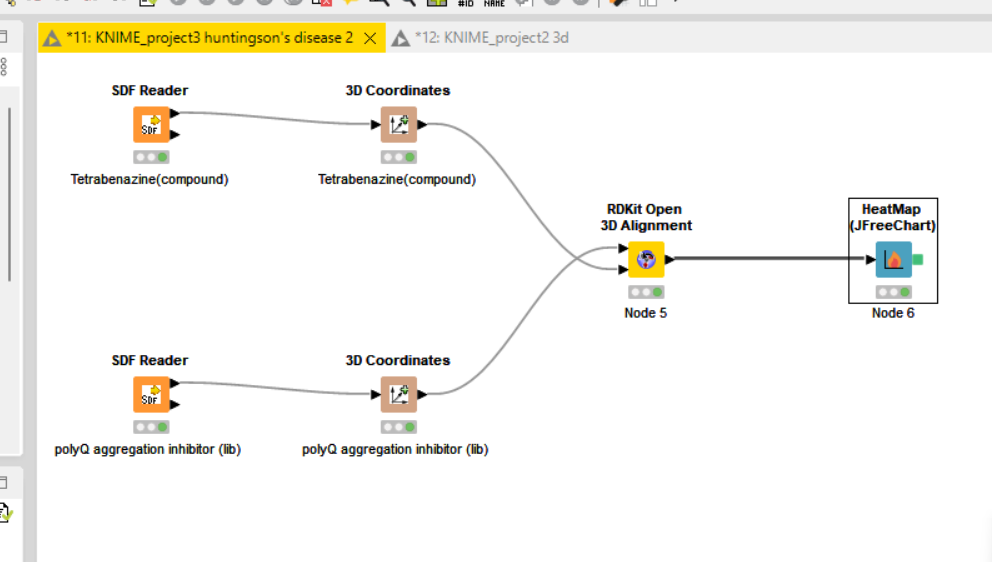
1. **Data Input:**  
   The **SDF files** of tetrabenazine and the polyQ inhibitors were imported into **KNIME Analytics Platform** using the **SDF Reader node**. This step prepares the molecular structures for further processing.

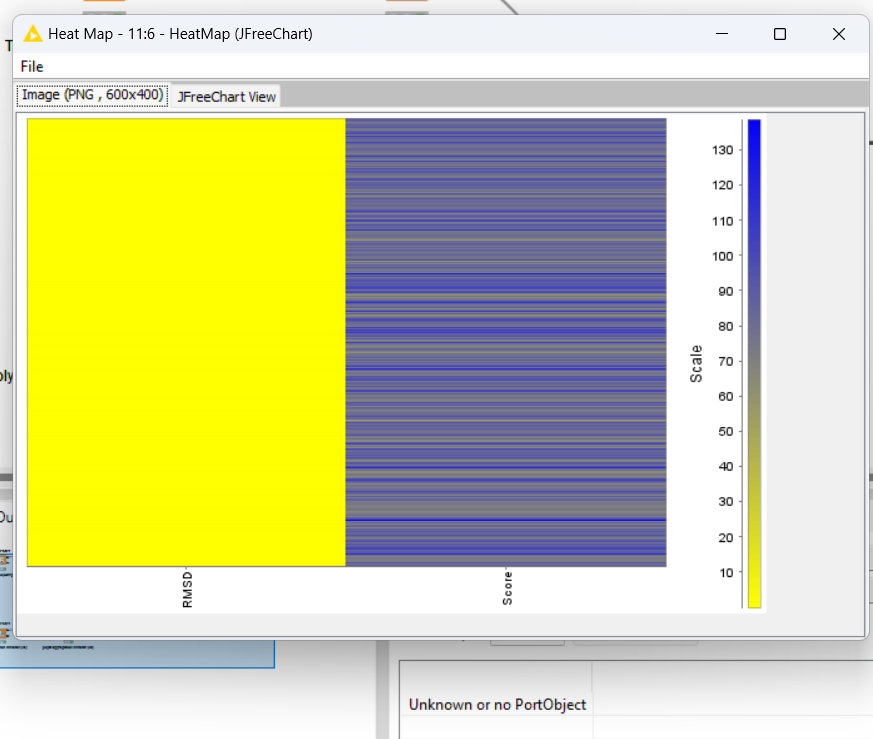


1. **Fingerprint Generation:**  
   **Molecular fingerprints** represent the key chemical features of a molecule, including functional groups and bond types. The **RDKit Fingerprint node** in KNIME was used to create fingerprints for both tetrabenazine and the polyQ aggregation inhibitors. These fingerprints allow for efficient comparisons between compounds based on their chemical characteristics.
2. **Tanimoto Similarity Calculation:**  
   The **Tanimoto coefficient** is a measure of how similar two compounds are, based on the overlap of their molecular features. The **Fingerprint Similarity node** in KNIME was used to calculate Tanimoto scores between tetrabenazine and each polyQ inhibitor. A score closer to 1 indicates a high degree of similarity.
3. **Heatmap Visualization:**  
   A **heatmap** was created using the **JFreeChart node** to visualize the Tanimoto similarity scores. The heatmap provides a clear graphical representation of which compounds are most structurally similar to tetrabenazine, with darker indicating higher similarity.



**3D Similarity Analysis**

1. **3D Coordinate Generation:**  
   The **RDKit Generate 3D Coordinates node** was used to generate 3D molecular structures for tetrabenazine and the polyQ inhibitors. The process also involved energy minimization to ensure that the molecules adopted their most stable conformations. 
2. **3D Alignment and RMSD Calculation:**  
   The **RDKit Open 3D Alignment node** was employed to align the 3D structures of the polyQ aggregation inhibitors with tetrabenazine, calculating the **Root Mean Square Deviation (RMSD)**. Lower RMSD values indicate that the molecules have a similar spatial arrangement, which is critical for predicting how they might interact biologically.
3. **Heatmap Visualization:**  
   The results from the 3D analysis were visualized through a heatmap showing **RMSD values**, where lower RMSD values are represented by blue regions, indicating better alignment, and higher values are shown in yellow or red.



**Integration of 2D and 3D Analyses**

By integrating the results from both the 2D and 3D analyses, the study provides a more complete understanding of the structural similarities between tetrabenazine and the polyQ aggregation inhibitors. This integrated approach combines the speed of 2D similarity screening with the depth provided by 3D spatial analysis.

**Results and Discussion**

**Key Findings**

* **2D Similarity Results:**  
  The Tanimoto coefficients from the 2D analysis varied between 0.2 and 0.8. Compounds with Tanimoto scores above 0.7 showed significant chemical similarity to tetrabenazine, marking them as promising candidates for further research.
* **3D Similarity Results:**  
  The RMSD values for the 3D analysis ranged from 0.5 Å to 2.0 Å. Compounds with RMSD values lower than 1.0 Å showed excellent spatial alignment with tetrabenazine, suggesting that these inhibitors may bind similarly to tetrabenazine’s targets.
* **Integration of 2D and 3D Results:**  
  The combined analysis identified several polyQ aggregation inhibitors with both high Tanimoto scores and low RMSD values. These compounds were flagged as the most structurally and spatially similar to tetrabenazine, making them ideal candidates for further experimental testing.

**Significance**

* **The Importance of Both 2D and 3D Analysis:**  
  The 2D analysis offers a fast and computationally efficient method for initial drug screening by focusing on chemical features, while the 3D analysis provides a more detailed assessment of molecular alignment, which is crucial for understanding how drugs interact at the molecular level. Combining these methods enhances the accuracy of the predictions and helps identify candidates that are more likely to exhibit similar biological activity.
* **Potential for Drug Discovery:**  
  The polyQ aggregation inhibitors identified in this study as structurally similar to tetrabenazine could be explored further as potential alternatives to tetrabenazine. These compounds may have the potential to mitigate the symptoms of HD with fewer side effects, thus advancing the search for safer and more effective treatments.

**Conclusion**

This study illustrates how combining **2D fingerprint-based analysis** with **3D RMSD-based analysis** can be a powerful approach to identifying drug analogs. By applying these methods to tetrabenazine and a library of polyQ aggregation inhibitors, the research identified several compounds with both structural and spatial similarities to tetrabenazine. These inhibitors may serve as potential candidates for future experimental validation, offering promising alternatives for the treatment of Huntington's disease. This integrated computational strategy can accelerate drug discovery, providing a robust platform for exploring new therapies with improved safety and efficacy profiles.