Internship Report

Name: Vinu Madavan

Organization: Bversity

Period of Internship: 26.11.2024 to 30.11.2024

Mode: Online

Overview:

The internship was based on the application of bioinformatics in drug discovery and development. It also delved into the uses of artificial intelligence and machine learning in the field of bioinformatics and cheminformatics to facilitate faster and efficient production of drugs and compounds without compromising safety or efficacy. The program used was KNIME which is free and open-source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining "Building Blocks of Analytics" concept. It can be used to develop workflows which are used to analyze data. The internship also covered the basics of the molecule to market process, which is the process of identifying a potential drug and commercializing it in the market.

Problem Statement:

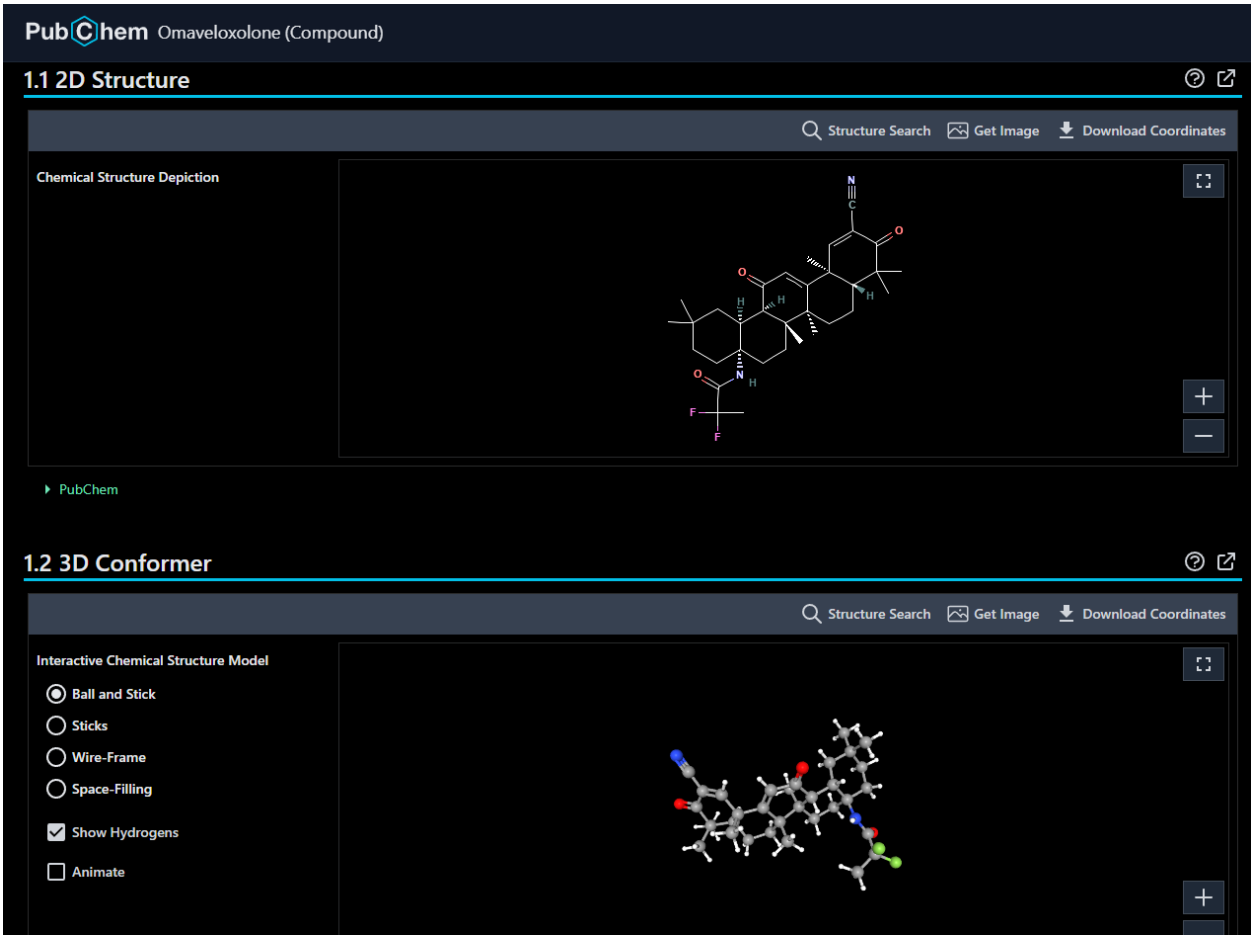
A company, XYZ, wants to develop a new drug for a disease since the existing drug in the market is expensive and not affordable to all segments of the population. The company wants the new drug to be developed with similar efficacy to the existing one, but at a lesser price.

Disease and Drug Overview:

Friedreich ataxia (FA) is a rare, inherited disorder that causes progressive damage to the nervous system. This can cause movement and sensory symptoms and trouble with walking and gait. In FA, nerve fibers in the spinal cord and peripheral nerves break down, becoming thinner. In the brain, the cerebellum, part of the brain that coordinates balance and movement, is most affected.

Omaveloxolone, sold under the brand name Skyclarys, is a medication used for the treatment of Friedreich's ataxia. It is taken by mouth. The mechanism of action of Omaveloxolone and its related compounds has been demonstrated to be through a

combination of activation of the antioxidative transcription factor Nrf2 and inhibition of the pro-inflammatory transcription factor NF-κB.



Approach:

- To tackle this task, a disease was chosen for analysis.
- The disease chosen in this case was “Friedreich’s Ataxia” and the existing drug in the market for it is “Omaveloxolone” (Skyclarys).
- The molecular properties of the drug can be studied in public databases such as PubChem.
- The mechanism of action of the drug is studied to understand the target sites of the drug and its clinical pathway.
- Then based on its action mechanism, a drug library is selected for comparison.
- The drug library in this case is a “NrF2 Activator” library.
- The drug library consists of several drugs to be compared with our drug of interest. The drug with the most similarity is then used to develop the new drug.
- The software then runs the program to search for similarities between the drug and the drug library. Then, the result is obtained in the form of visual chart.

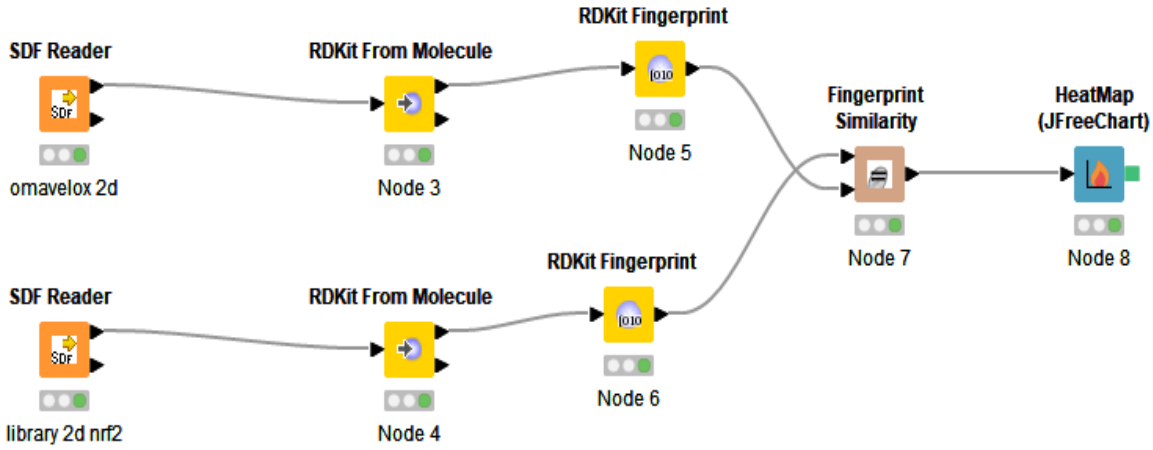
KNIME:

The entire process of the comparison of drug comparison and analysis is carried out using the KNIME software. The process is run through protocols called “workflows”. The workflows are made up numerous connections of nodes. The nodes are small blocks of codes that can be connected to perform certain functions. It is a functionality/small script that performs a certain function.

Principle: The main logic behind this process is inputting the drug of interest and drug library in two different slots and using machine learning process to find out which drug from the drug library matches the drug of interest.

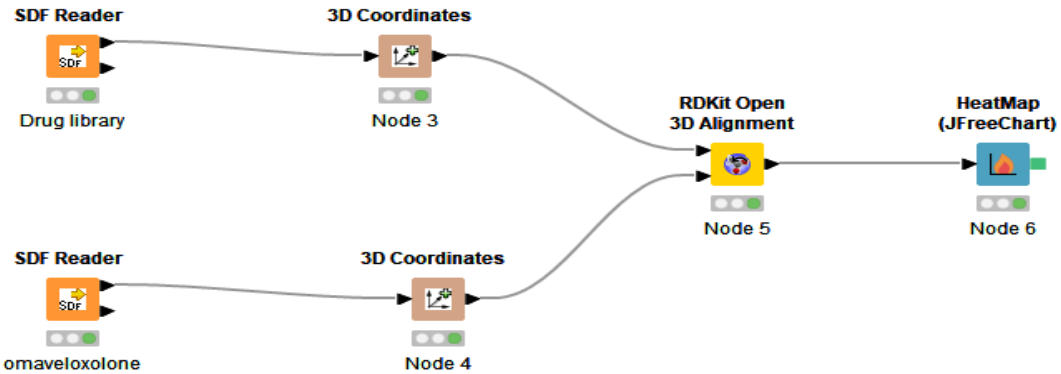
Workflow: (Using 2D Conformation)

- Initially the SDF files containing the 2D molecular structure of the drug and the drug library are downloaded from a public database.
- Then the files are configured into a “**SDF Reader**” node and executed.
- To compare the two inputs, they must be converted to fingerprints. Molecular fingerprints are binary codes that determine the individuality of a molecule and make it easier to identify and compare.
- The nodes used for this conversion are” **RDKit from molecule**” and “**RDKit Fingerprint**”.
- There are two types of fingerprint arrangements which are **MAACS (Array Format)** and **Morgan (Spatial Arrangement)**. These must be configured in the RDKit Fingerprint node.
- The two nodes are then connected to a “**Fingerprint Similarity**” node which compares both the fingerprints.
- The result is then obtained from a “**Heatmap**” node which presents the data with the similarity represented by Tanimoto coefficient and differing shades based on the coefficient. The closer the value is to “1”, the more the similarity.



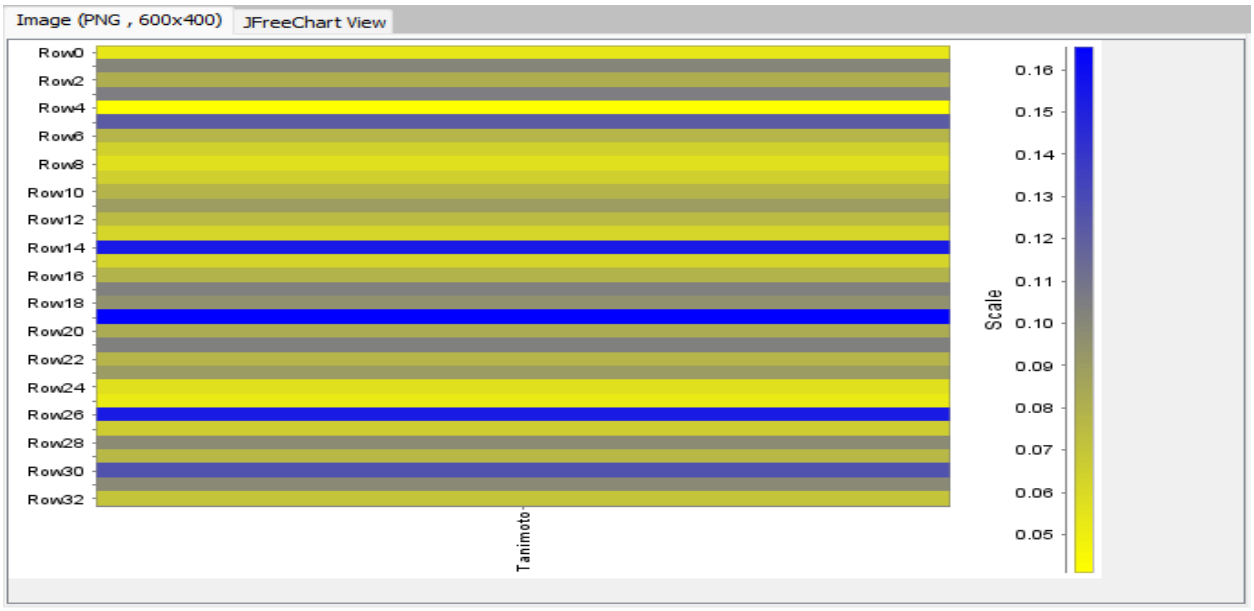
Workflow: (Using 3D Conformation)

- Initially the SDF files containing the 3D molecular structure of the drug and the drug library are downloaded from a public database.
- Then the files are configured into a "**SDF Reader**" node and executed.
- Since the molecules are in a 3-dimensional structure, they are formatted by using the "**3D Coordinates**" node to map the structure and make it easier for comparison.
- Then the two 3D Coordinates nodes are connected to a "**RDKit open 3D Alignment**" node which then compares the library and the drug of interest.
- To visualize the results, the node is connected to the "**Heatmap**" node which then presents the data with the similarities represented by the RMSD (Root Mean Square Deviation) values. The lesser the RMSD value, the greater the similarity and vice versa.



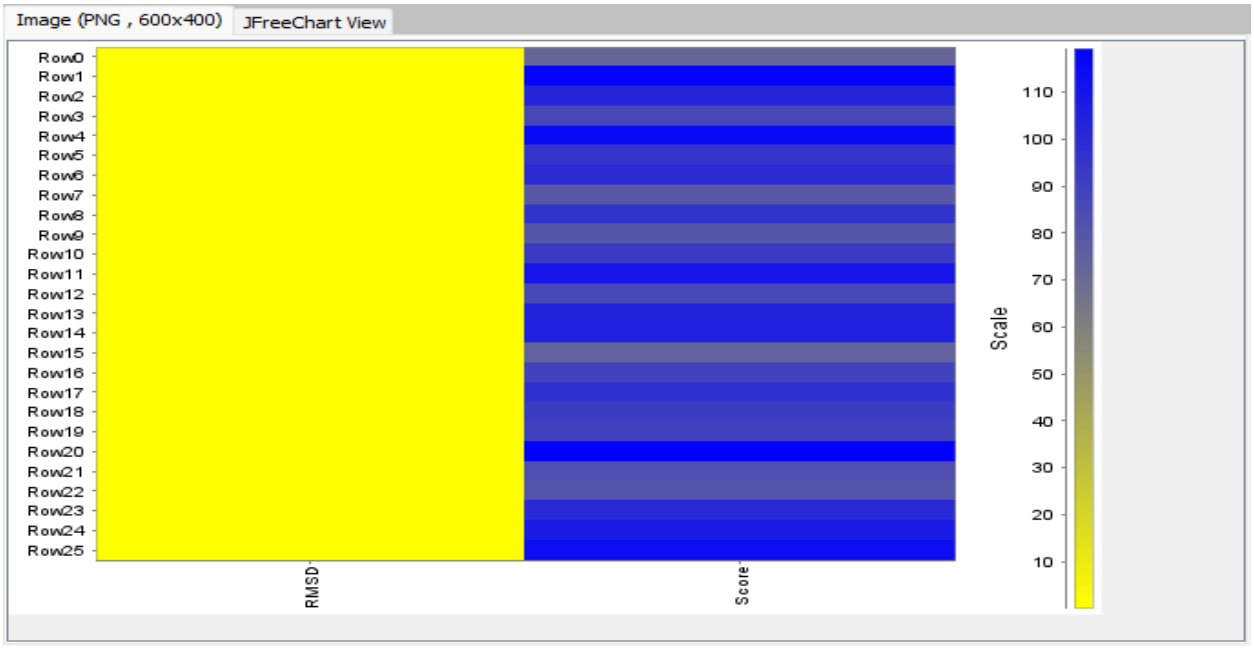
Result:

2D Conformer result:



From this result we can interpret that the dark blue represents more similarity and yellow represents less similarity since the dark blue's Tanimoto value is closer to "1". From observing the above chart, we can conclude that the drug from the drug library which has the most similarity to our drug of interest is the drug on "row 14".

3D Conformer result:



Similarly, from observing and analyzing the RMSD values from the chart, we can conclude that the drug on “row 14” has the most similarity with our drug of interest. The lesser the RMSD value, the higher the similarity and vice versa.

Conclusion:

From running all the workflows and analysis, we can come to a conclusion regarding the drug which is best suited to replace the drug already in market. This model can be used to identify potential new drugs and formulate compounds which have the same efficacy and action mechanism as the previous drug existing in market. Bioinformatics, machine learning and artificial intelligence play a huge and vital role in this process, ensuring efficiency, accuracy, and innovation in drug development.