Name: Srayaa S

Mega Bioinformatics Internship – Bversity – 2024 – Project Report

1. Problem statement

Identification of an existing drug similar in structure and function to that of the drug of context (Regorafenib) against colorectal cancer through drug repurposing studies in order to overcome the effects of the existing drug.

2. Disease and drug

According to WHO statistics, colorectal cancer is the third most commonly caused cancer, for which various FDA-approved medications are available. One of the approved drugs is regorafenib, which is categorised under the class of small molecule kinase inhibitors, and this exerts its action by targeting the metastatic state of colorectal cancer.

2.1. Mechanism of action of regorafenib in colorectal cancer

Regorafenib targets tumor-associated processes through targeting specific proteins, in spite of the presence of mutations in patients with respect to *RAS* and *BRAF* genes. The targeted mechanisms include:

a. Downregulating angiogenesis by targeting TIE2 (angiopoietin) and VEGFR
b. Retarding cancer cell proliferation through the inhibition of RAF, BRAF, and c-KIT
c. Controlling metastasis through PDGFR retardation
d. CSF1R inhibition leading to the alleviation of immunosuppressive effects (Arai et al., 2019).

The drug's effects on immune-related mechanisms are as follows:

a. The release of MHC class I polypeptide-related sequence A (soluble form) from cancer cells is halted by the downregulation of ADAM 9 and 10 due to regorafenib's action. This in turn restores the activity of the NKG2D receptor on natural killer cells to act against the cancer cells.
b. The role of cancer-associated fibroblasts (CAFs): CD73 and CXCL2 release from regulatory T cells are stimulated by CAF; also, CAF releases TGF-beta and complement factor (C5a), which signals the infiltration of tumor associated macrophages. These together retard the function of effector T cells in tumor

microenvironment. But CAF are directed to apoptotic state by the drug's action through Bcl-2 and Bax (Liu et al., 2022).

2.2. Retrieval of drug

The search term used in PubChem was 'Regorafenib' and the drug with both the 2D and the 3D structures available was chosen, since 3D structure is required for drug similarity analysis. Both the 2D **(Figure 1)** and the 3D structures were downloaded in SDF (Structure Data Format) format. 3D SDF file was converted into PDB format using OpenBabel and visualised using Biovia Discovery Studio **(Figure 2)**.

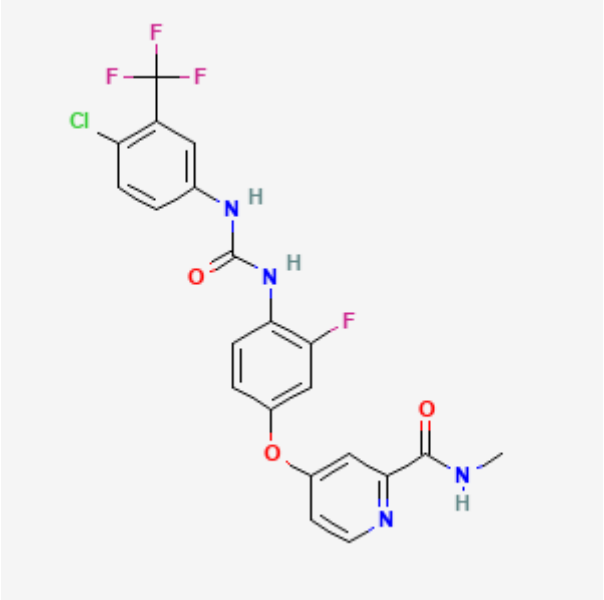

Figure 1 Regorafenib 2D structure

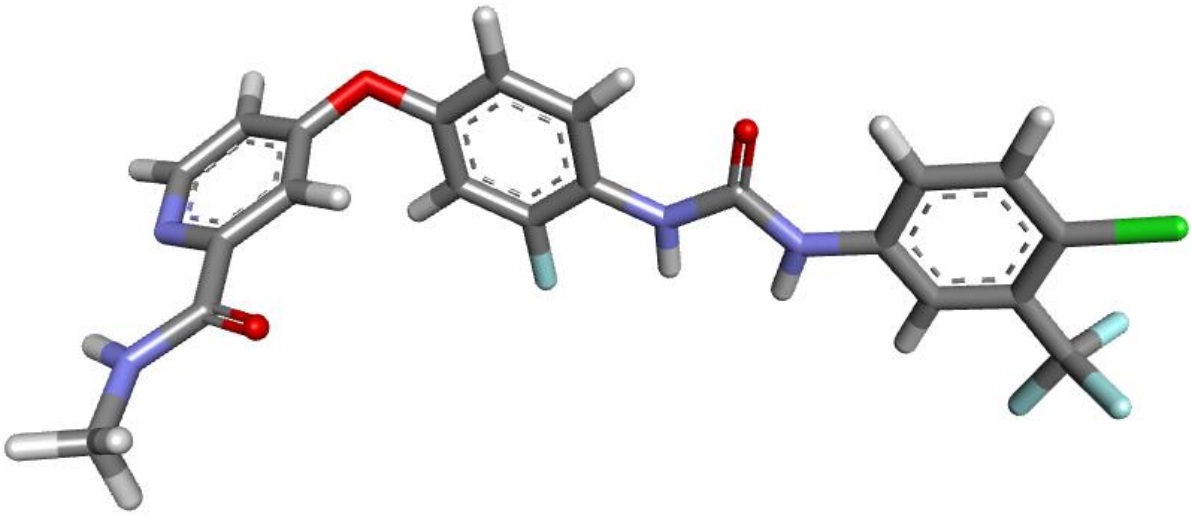

Figure 2 Regorafenib 3D structure

2.3. Retrieval of drug library

Regorafenib is categorised under the 'tyrosine kinase inhibitors' class as well, and so this search term was used in PubChem to retrieve the drug library with existing drugs (library with a total of 410 drug molecules). All the drug molecules were retrieved in both 2D and 3D formats.

3. KNIME methodology

KNIME Analytics Platform was employed for performing the drug similarity analysis to identify a drug molecule from the retrieved drug library that is similar to that of regorafenib. There are various nodes in KNIME available from the node repository, each node showing three different colours on modification. It shows red on dropping into the workspace, shows yellow on configuring it with data, and shows green when executed.

3.1. KNIME workflow with drug molecules of 2D structure

A new KNIME workflow was created, and various nodes were employed from the node repository. The nodes used and the workflow are as follows: SDF reader is a node to read the SDF file into KNIME. 2 of these nodes were incorporated (for reading the library and regorafenib in 2D structures). Using the 'configure' option, the files were read into KNIME and executed. The RDKit From Molecule node was employed to use the SMILES format of drug molecules to create a molecule column. RDKit Fingerprint was used to create fingerprints for both regorafenib and the drug library. This was performed by reading the molecular features of drug molecules through two fingerprints: MACCS **(Figure 3)** and Morgan **(Figure 4)**. MACCS is based on an array format, and Morgan is based on spatial structural format. The Fingerprint Similarity node performed the similarity analysis between one reference drug molecule and 410 test drug molecules. And the outputs were visualised as a heatmap using the Heatmap (JFreeChart) node. All the nodes were connected to one another using input and output ports.

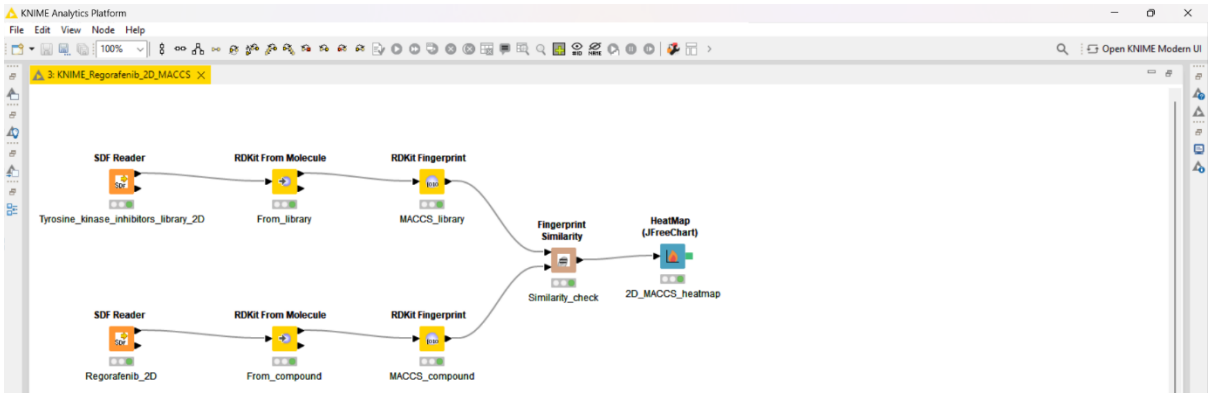

Figure 3 2D drug molecules' structure – MACCS workflow

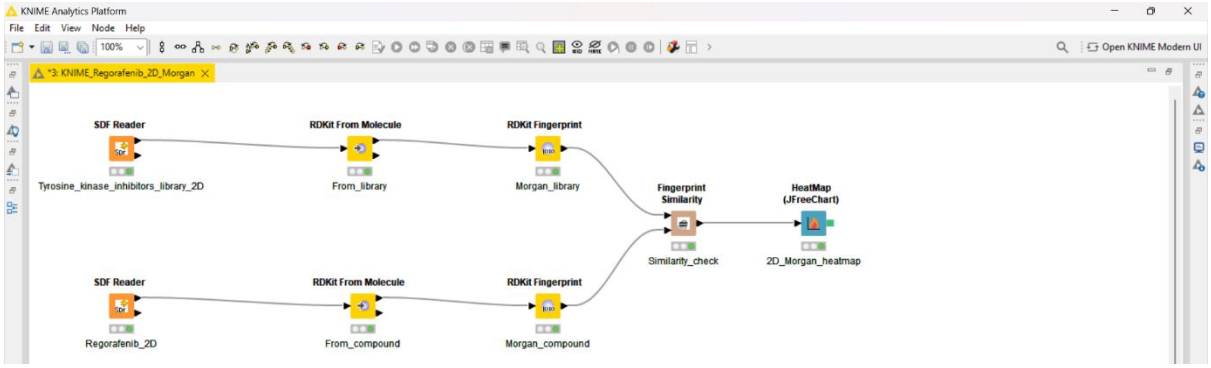

Figure 4 2D drug molecules' structure – Morgan workflow

3.2. KNIME workflow with drug molecules of 3D structure

Using SDF reader, the drug molecules were read into KNIME in 3D SDF format. The analysis was performed for a total of 306 molecules in the drug library (It means out of 410 downloaded drug molecules, only 306 molecules were available with 3D structure). The 3D coordinates node was employed to generate 3D coordinates for all atoms in the drug molecules' structure, including hydrogen, and the resultant drug molecules in the library were reduced to 296 (only for these were the proper coordinates able to be applied). The RDKit Open 3D Alignment node was used to align the molecules in 3D structure. This aligns each drug molecule from the library to regorafenib (reference drug molecule) and generates the aligned structures. And the outputs were visualised as a heatmap using the Heatmap (JFreeChart) node, with all the nodes connected.

Another node called 'Rank' was used to rank the molecules based on RMSD (in ascending order) and score (in descending order) to identify the highly similar drug molecule to that of regorafenib. The ranked data table listed the drug molecules in the order of high to low score

and low to high RMSD. And the same was visualised as a heatmap, in which an additional column 'rank' was incorporated.

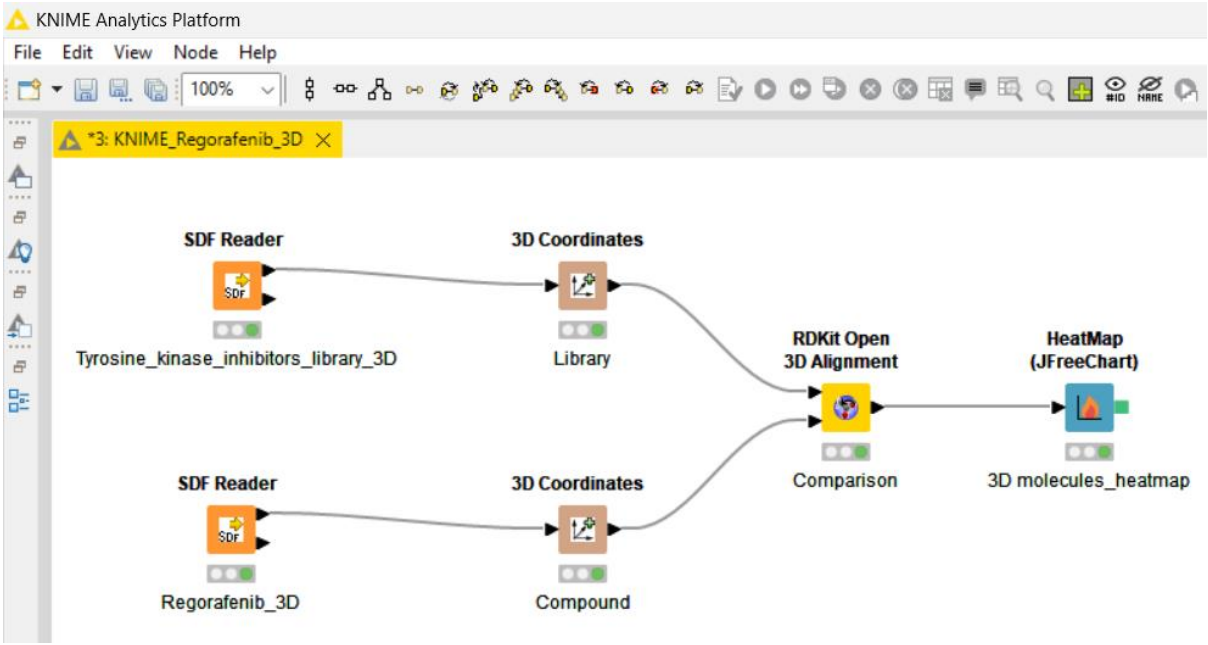

Figure 5 3D drug molecules' structure – Workflow

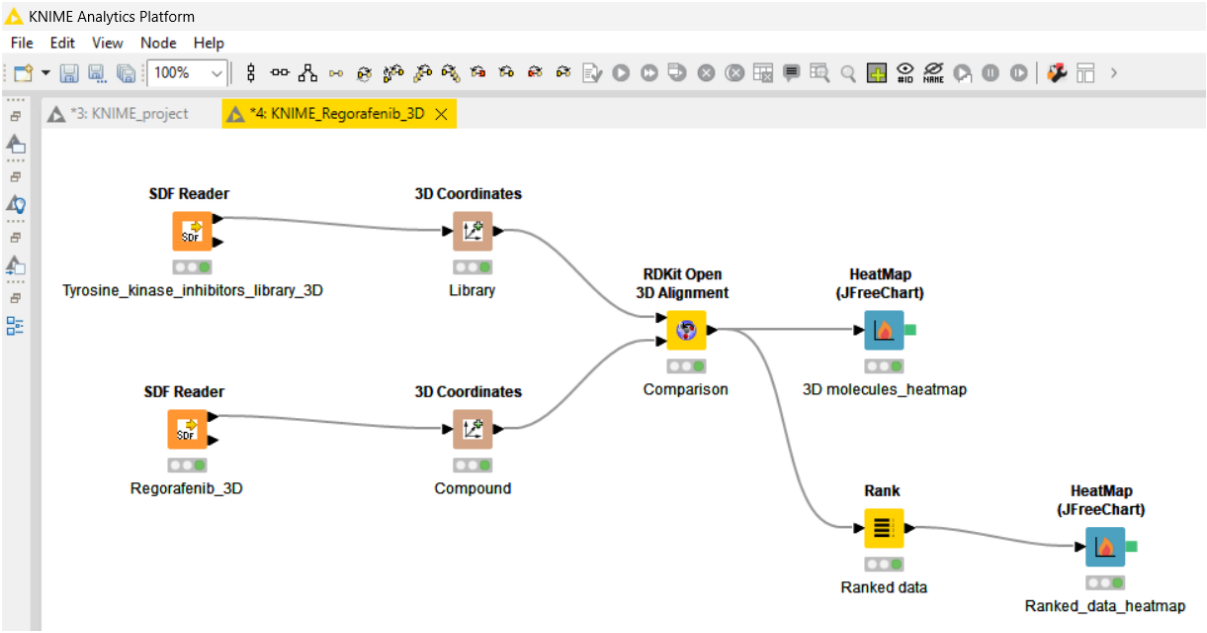

Figure 6 3D drug molecules' structure – Ranked data workflow

4. Results and inference

The heatmap obtained using the MACCS fingerprint **(Figure 8)** is with a scale of Tanimoto similarity, ranging from 0.1 (less similarity) to 1 (high similarity). Each row represents each

drug molecule compared against regorafenib, and the lines coloured with yellow (less similar) and blue (highly similar) correspond to specific Tanimoto coefficients. The same applies for the heatmap visualised from the Morgan fingerprint as well **(Figure 8)**.

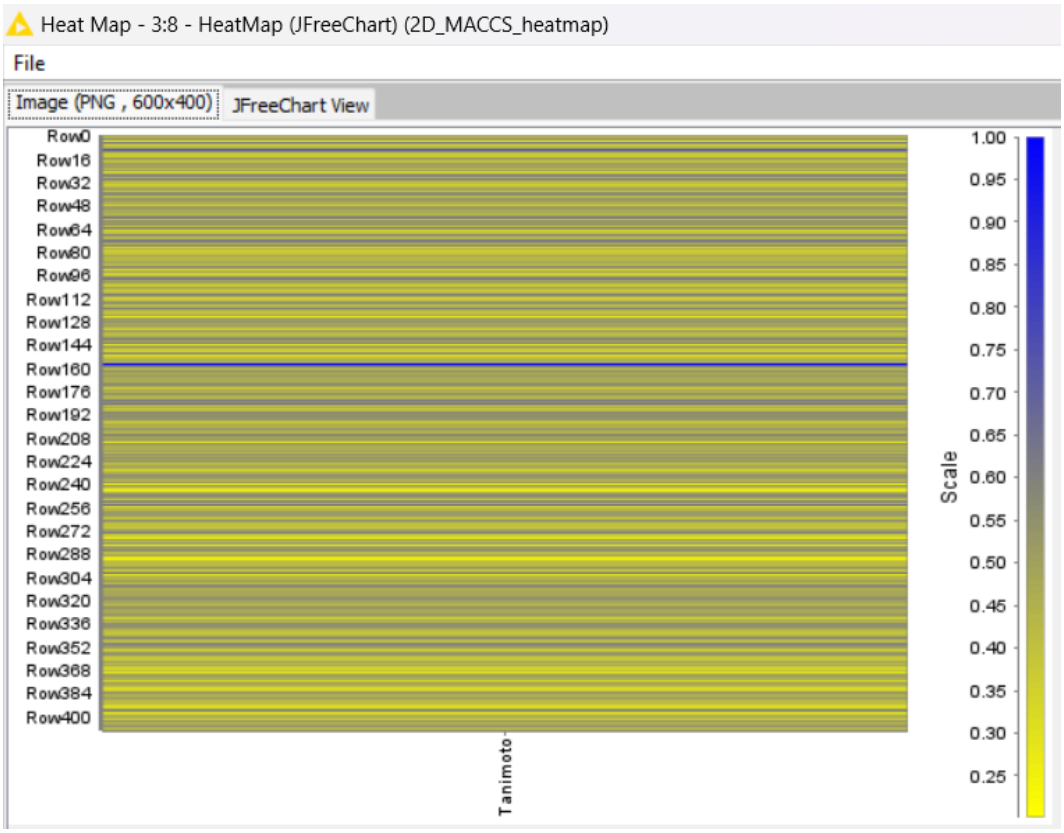

Figure 7 2D structure – MACCS fingerprint – Heatmap

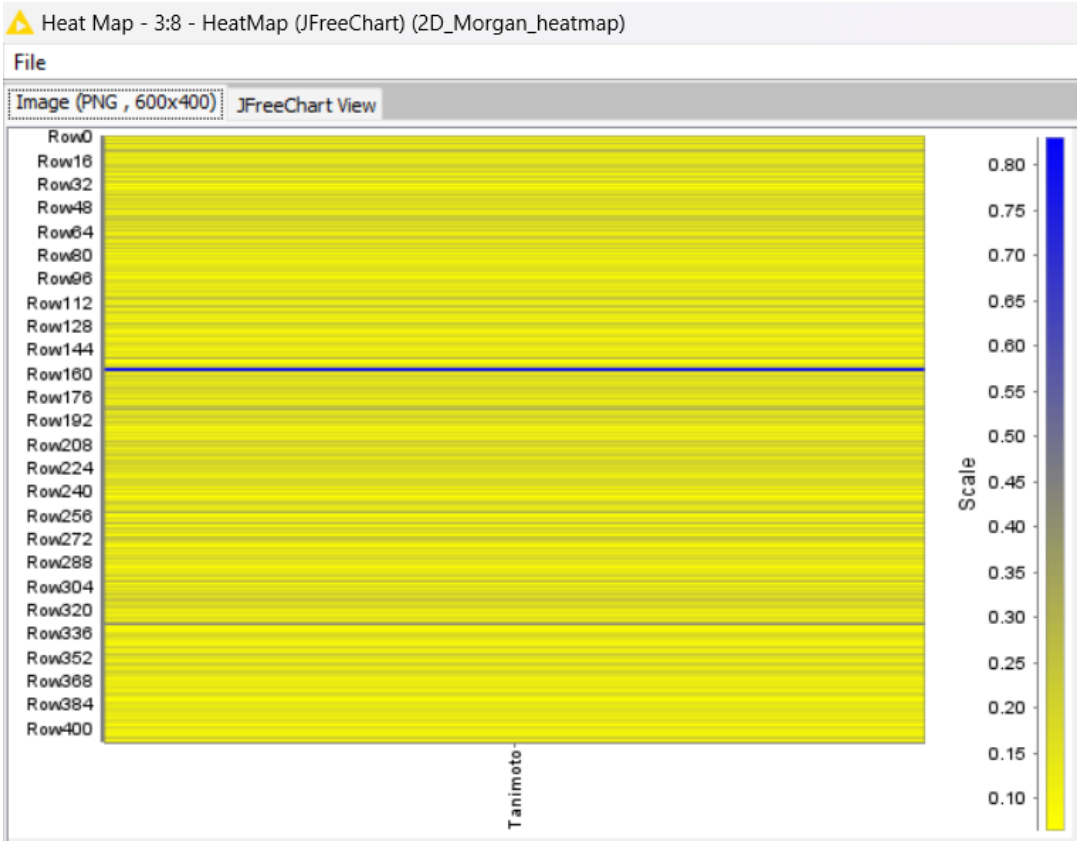

Figure 8 2D structure – Morgan fingerprint – Heatmap

The heatmap visualised for the analysis with 3D structures is with RMSD and score, represented with a scale of values **(Figure 9)**. Here, low RMSD (shown in yellow, towards 0) and high score values (shown in blue, towards 270) correspond to the highly similar drug molecules from the library to that of regorafenib.

Through the ranking of the data, it can be identified from the table **(Figure 10)** and heatmap **(Figure 11)** that row 149 (drug molecule) is highly similar to that of regorafenib. In the heatmap, lower rank (shown in yellow) represents the highly similar drug molecule.

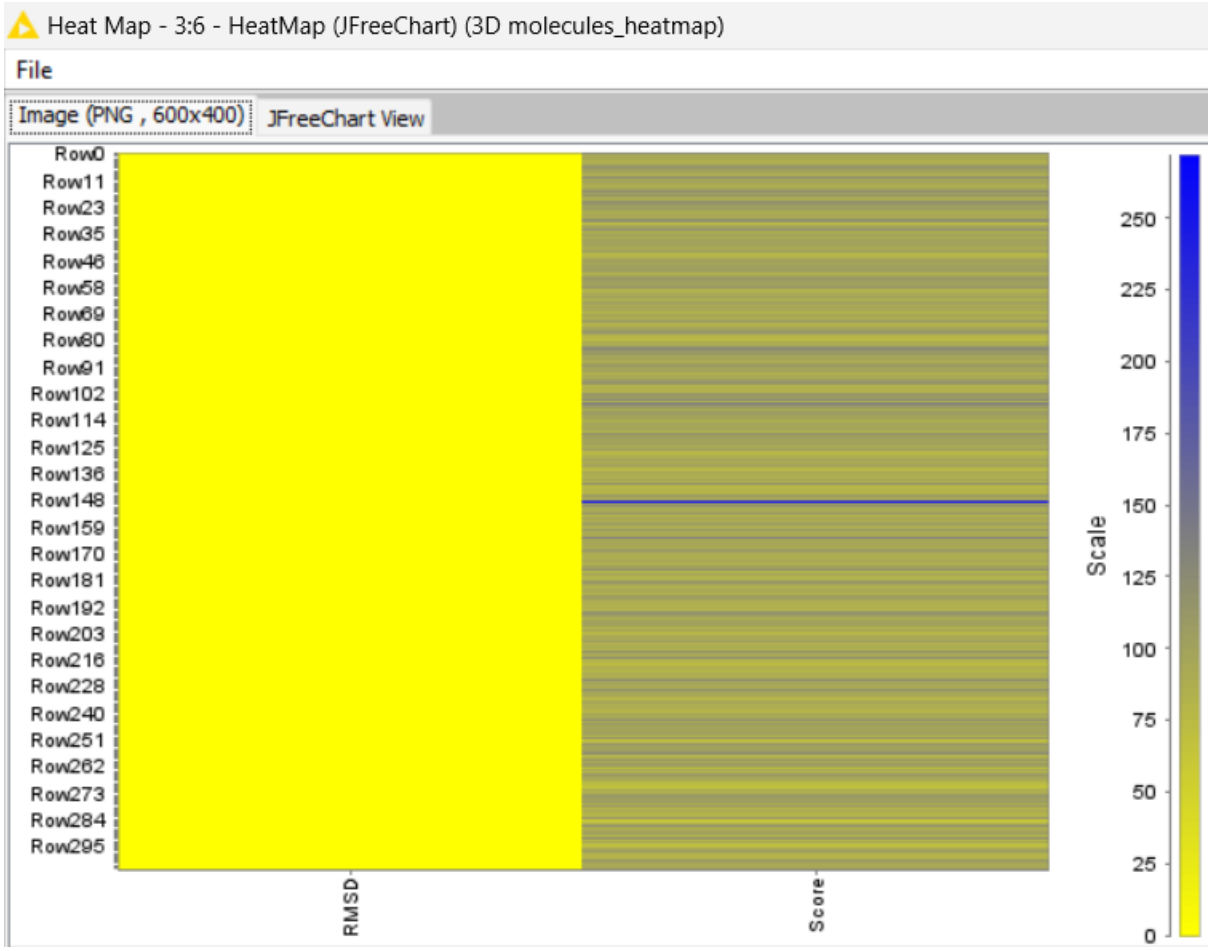

Figure 9 3D structure – Heatmap

Figure 10 3D structure – Ranked data table

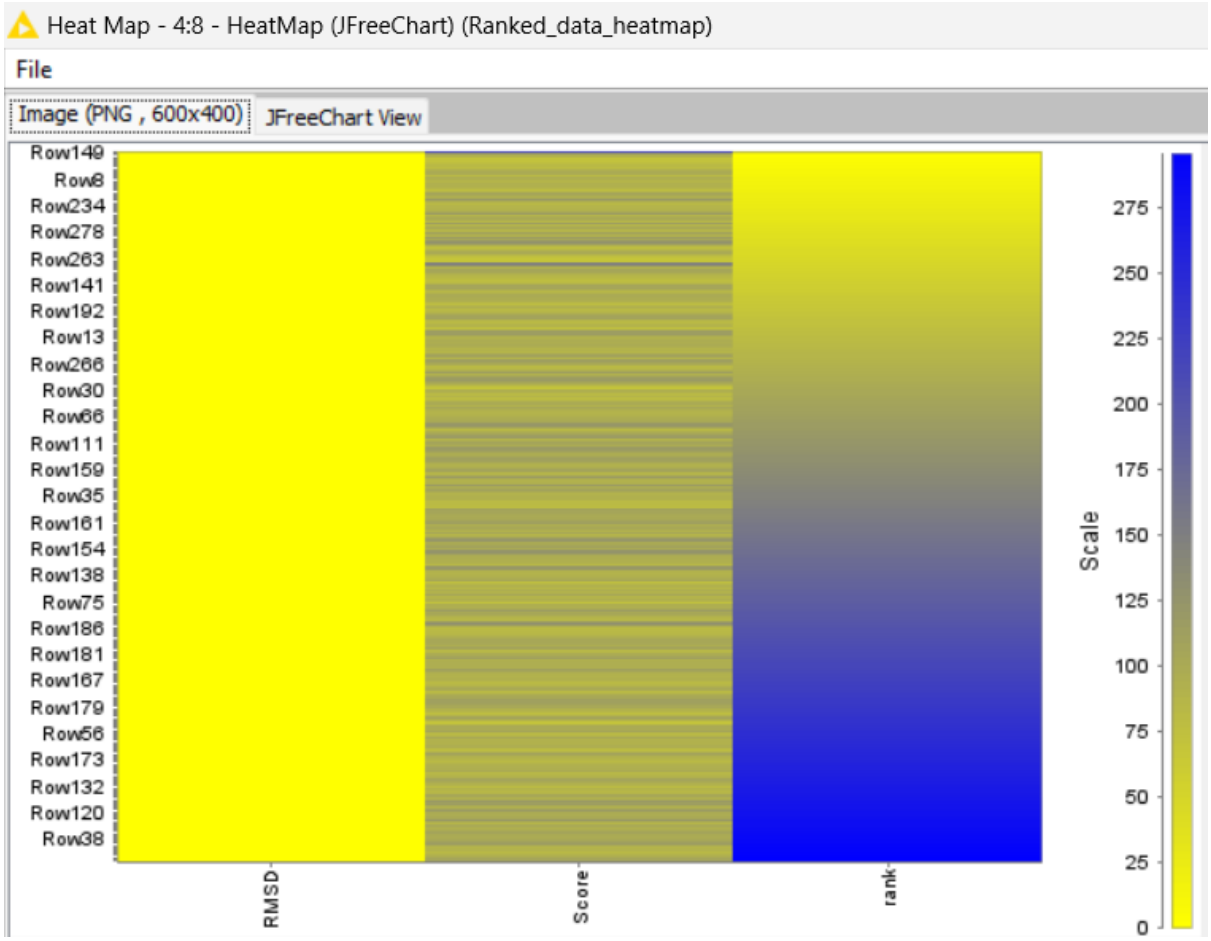

Figure 11 3D structure – Ranked data heatmap

Conclusion

Drug repurposing is applied to find an existing drug that could be similar to the drug in question, so as to apply the similar drug to treat a disease. In this case, a drug repurposing study was performed through drug similarity analysis to find an existing drug that could be similar to that of regorafenib. Based on the interpretation, specific drug molecules, which are highly similar to that of regorafenib (for example, drug molecules 149, 8, 234, 278, and 263 visualised from the heatmap), can be taken for further analysis, which involves target identification and molecular docking so as to identify a similar drug with respect to both structure and function to that of regorafenib in treating colorectal cancer.

References

Arai, H., Battaglin, F., Wang, J., Lo, J. H., Soni, S., Zhang, W., & Lenz, H. (2019). Molecular insight of regorafenib treatment for colorectal cancer. *Cancer Treatment Reviews, 81*, 101912. https://doi.org/10.1016/j.ctrv.2019.101912

Liu, J., Tao, H., Yuan, T., Li, J., Li, J., Liang, H., Huang, Z., & Zhang, E. (2022). Immunomodulatory effects of regorafenib: Enhancing the efficacy of anti-PD-1/PD-L1 therapy. *Frontiers in Immunology, 13*. https://doi.org/10.3389/fimmu.2022.992611

National Center for Biotechnology Information (2024). PubChem Compound Summary for CID 11167602, Regorafenib. Retrieved December 4, 2024 from https://pubchem.ncbi.nlm.nih.gov/compound/Regorafenib.

https://pubchem.ncbi.nlm.nih.gov/#query=tyrosine%20kinase%20inhibitors