**Generation**

**User Input**

The goal of this project is to make a fully automated pipeline for generating new molecules with specific characteristics. To do this, all we should require from the user is a target characteristic and the program should generate molecules for that characteristic.

For our previously mentioned model, the input is a Vector representation of a starting molecule that is trained in a self-supervised manor, from the images generated by RDKit. This means that we need to user to provide the model with a possible starting molecule for the model and a condition with which they wish to enforce on the molecule. This however, is not ideal. Ideally, I would want the user to input a target molecule/virus etc, the model would identify multiple possible starting molecules and then generate a condition to improve upon the starting molecules and provide the user with a large amount of new novel molecules.

After this process, when the new molecules have been generated, we can then assess if they are valid, create a description profile and predict their effects. After this we can then simulate and rank the results, producing the user with a detailed analysis of potential target molecules.

This entire process is designed to be entirely automated, other than the initial user input.