**Lab Report for Lab 2: Chemoinformatics II**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Objective:**

**Data and Results**:

1. How many Molecules were there with IC50 data?
2. Plot of from module **2A** validating Solubility forecast index.
3. Box Plots arranged 5x2
4. Box plots arranged 10x1
5. 3 plots from module **3A**
6. Screenshot of your code for your Lipinski filter and your PCA.
7. List of 3 smiles strings and IC50’s from module 4.

**Discussion:**

1. What is a principal component analysis?
2. Convert your SMILES you wrote down into a chemical structure. This can be done by pasting into ChemDraw or by typing it in at the following link: <https://pubchem.ncbi.nlm.nih.gov//edit3/index.html>
3. The following three molecules clustered very near each other on the PCA yet have vastly different IC50 values. What structural feature(s) do you think account for this and why?



**Figure S12.** Example molecules used to test student understanding of dimensionality reduction algorithms in a molecular context.