Questions

1. Formulate a PUG-REST request URL to perform the following tasks.

   (a) Get the full record (in XML) for CID 60823 that are derived from its 3-D structure.

   (b) Get the small-size 2-D image of CID 2244.

   (c) Get the XLogP and TPSA values of CID 1983, 2244, and 3672 in the CSV format.

   (d) Get the list of CIDs whose name contains the string “crestor” in the TXT format.

   (e) Get the list of CIDs for compounds that are similar to CID 3348 in terms of 2-D similarity (at the threshold = 95)

   (f) Get the full record (in XML) of CIDs that have the substructure “C3=NC1=C(C=NC2=C1C=NC=C2)[N]3”

   (g) Get a list of AIDs (in TXT) that target the protein encoded by the KCNH2 gene.

   (h) Get a list of CIDs (in TXT) for compounds tested to be active in AID 376.

   (i) Get the bioactivity data (in CSV format) for AID 376.

   (j) Get the sources of the boiling point annotations in PubChem (Hint: see **Module 5**).

2. This question is designed to test whether you can access PubChem data from a spread sheet program.

   (a) Make a Google Sheet like this template (http://bit.ly/2o0pzoT) and auto-populate it with the CID, molecular formula, heavy atom count, and XLogP values for alkanes and alcohols (with 1 through 10 carbon atoms). Provide the link to the completed google sheet so that your instructor can access it.

   (b) Compare the XLogP values of alkanes with those of alcohols. Which one is greater, and explain why.

3. Modify the example python script 4 (https://trinket.io/python/3437cdefb0) to retrieve (in a CSV format) the molecular formula, molecular weight, hydrogen bond donor count, hydrogen bond acceptor count, and XLogP values of compounds returned from a 2-D similarity search using the following three smiles strings:

   - C1CCNCC1
   - C1CCC(CC1)(CC(=O)O)CN
   - C1=CC=C(C=C1)NC2=CC=CC=C2C(=O)O

   [Please create an account for trinket.io, save your work, and provide your instructor with the link to it.]