



### 3. Übung zur Vorlesung Grundlagen der Chemieinformatik

*Abgabe als pdf bis 08.11.2013 8:00h an [gci-uebung@zbh.uni-hamburg.de](mailto:gci-uebung@zbh.uni-hamburg.de)*

#### Aufgabe 1: MONA (A)

(12 Punkte)

In dieser Übung werden Sie sich mit einem Molekülverarbeitungstool mit dem Namen Mona auseinandersetzen. Das Tool wurde hier am ZBH von Matthias Hilbig entwickelt und eignet sich um durch große Molekülmengen zu browsen, diese zu filtern, einzelne Sets zu bilden und für das Konvertieren in verschiedene Moleküldateiformate. Sie werden in dieser Übung das Konzept von Molekülinstanzen kennenlernen, Moleküle filtern und verschiedene Moleküldateiformate ansehen.

Das Executable für Mona finden Sie hier im ZBH unter  
`/home/schomburg/GCI/mona_linux32`

Kopieren Sie es sich in einen Ordner in ihrem Verzeichnis und starten Sie das Programm.

Als erstes wollen Sie Moleküle einlesen. Öffnen Sie dafür den Reiter *Molecule Sources*. Klicken Sie nun auf *Add Directories* und wählen Sie den Ordner  
`/home/schomburg/GCI/dud_ligands/`.

Markieren Sie alle Moleküle und laden Sie diese in ein Set. Klicken Sie nun auf den Reiter *Visual Selection* und sie sehen die Strukturdiagramme ihrer geladenen Moleküle. Ein Doppelklick auf eine Molekülzelle öffnet rechts das *Instance View*-Tab. Monas Konzept der Instanzen betrachtet Liganden als gleich, wenn ihre Topologie identisch ist. In diesem Fall werden Sie nur einmal gezeigt und nur die Liste im *Instance View*-Tab zeigt auf, in wie vielen verschiedenen Instanzen ein Molekül in dem geladenen Datensatz auftritt. Suchen Sie ein Molekül, bei dem Sie hier mehr als einen Eintrag sehen:

- Aus welchen Dateien stammen die Einträge?
- Was können Sie noch sehen?
- In welcher Hinsicht können sich zwei Instanzen unterscheiden?

Gehen Sie nun auf den Reiter *Filters*. Wenden Sie den schon vorhandenen Lipinski-Filter auf die Moleküle an.

- Wie viele Moleküle bleiben nach dem Filter übrig?

Klicken Sie nun auf *Create Filtered Set* (unten links) und erzeugen so ein Set mit den übrigen Molekülen. Benutzen Sie die Set-Operationen die durch die bunten Kreis-Icons oben in der Leiste kenntlich gemacht sind, um herauszufinden, welche Moleküle herausgefiltert wurden. Dazu müssen Sie die Differenz vom ersten und zweiten Set bilden. Schauen Sie sich die Moleküle vom gefilterten Set an. Die Eigenschaften der Moleküle können Sie sich unter dem Reiter *Histogramm* anschauen. Markieren Sie mit der Maus das Set der Lipinski-gefilterten Moleküle und das Differenz-Set. Schauen Sie sich die Eigenschaften *Donors*, *Acceptors*, *MW* und *Rings* an:

- Beschreiben Sie, was Ihnen auffällt.

Laden Sie nun die Moleküle aus dem Ordner  
`/home/schomburg/GCI/dud_decoys/`  
in ein Set. Dies kann aufgrund der Molekülmenge einen Moment dauern.

- Wieviele Moleküle hat das erstellte Set?

Vereinigen Sie nun dieses Set mit dem Set was Sie als allererstes erstellt haben. Als nächstes sollen Sie selbst Filterregeln aus der großen Menge verfügbarer Filtermöglichkeiten zusammenstellen. Nutzen Sie auf dem neuen vereinigten Set die Filter so, dass Sie am Ende nur noch maximal 10 Moleküle (mindestens aber 1) übrig haben.

- Welche Filter mit welchen Parametern haben Sie genutzt?

Erstellen Sie ein gefiltertes Set mit Ihren gewählten Filtern. Exportieren Sie nun das Set was sie gefiltert haben mit einem Rechtsklick und *Export Set*. Geben Sie einen Dateinamen an und wählen Sie einen Speicherort in Ihrem Verzeichnis. Exportieren Sie die Datei einmal als SDF, einmal als SMILES und einmal als MOL2-Datei. Öffnen Sie die Dateien dann mit einem Texteditor.

- Beschreiben Sie knapp die Unterschiede der Dateiformate und nennen Sie zu jedem Format einen Vorteil.

Schauen Sie sich ebenfalls die Datei `/home/schomburg/GCI/testmolecule.moe` mit einem Texteditor an.

- Was für Informationen sind in dieser Datei noch abgelegt?

Sie können sich das Molekül mit dem Befehl  
`/usr/local/zbhtools/moe2012.10/bin/moe /home/schomburg/GCI/testmolecule.moe`  
anzeigen lassen. (Für das Programm moe gibt es im ZBH nur Lizenzen für drei parallele User. Eventuell öffnet das Programm nicht, wenn alle Lizenzen gerade benutzt werden. Bitte schließen Sie daher auch das Programm sofort wieder, wenn Sie mit der Benutzung fertig sind).

Finden Sie nun selbst heraus, ob es Moleküle gibt, die bei der DrugBank sowohl als *approved* als auch als *withdrawn* klassifiziert werden. Dafür müssen Sie sich die Moleküldaten selbst aus der DrugBank ([www.drugbank.ca](http://www.drugbank.ca)) herunterladen und in Mona öffnen.

- Wie viele Moleküle sind in beiden Klassifizierungen zu finden?
- Was fällt Ihnen Unerwartetes auf?
- Suchen Sie sich eines der Moleküle heraus und beschreiben Sie, warum es vom Markt genommen wurde.