QSAR modeling of growth inhibition

Egon Willighagen

2011-08-12

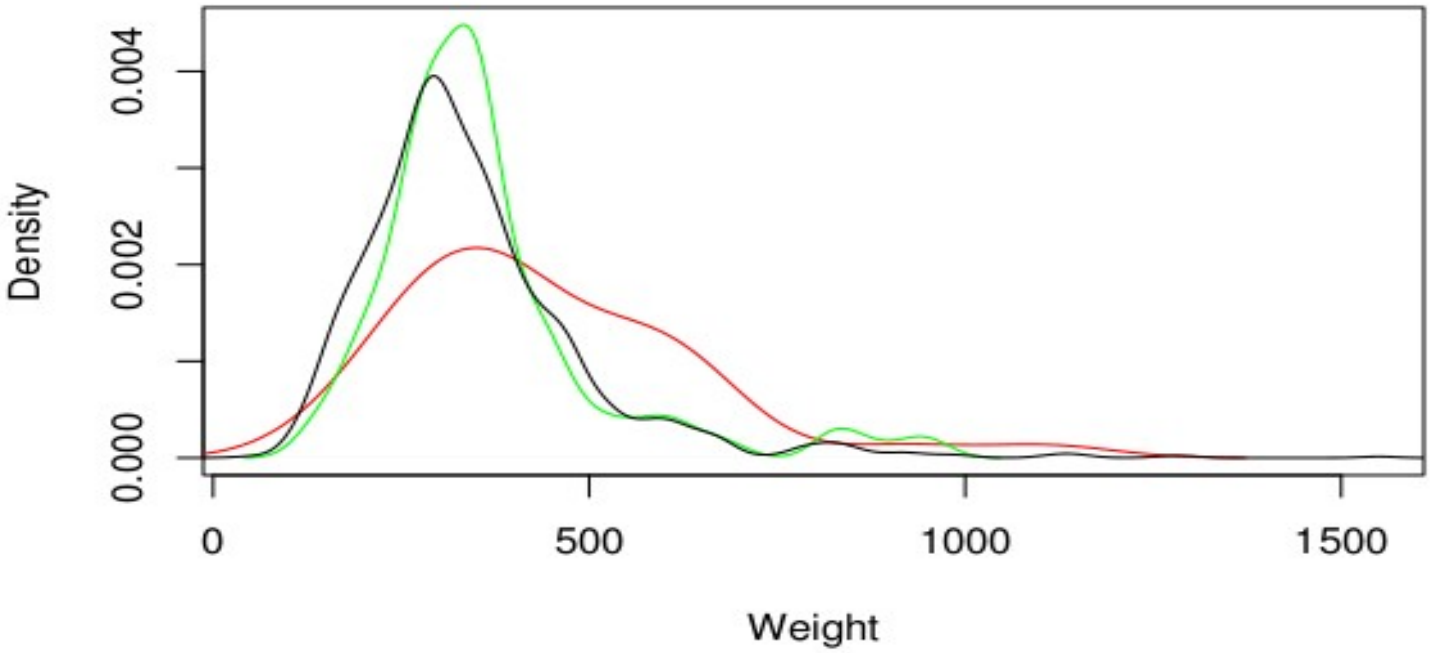
Data and Goal

- **Compounds**
  - 230 drugs → molecular descriptors
    - logP, number of acidic groups, etc
- **Activities**
  - log GI_{50} values (between -7 and -4), where GI_{50} is in molar
    - GI_{50} is the dose where the growth is inhibited 50%
  - Three cell lines: HL60, PC3, MCF7
- **Toxic molecules**

  → low(er) log GI_{50} values. That is log $GI_{50} < -5$

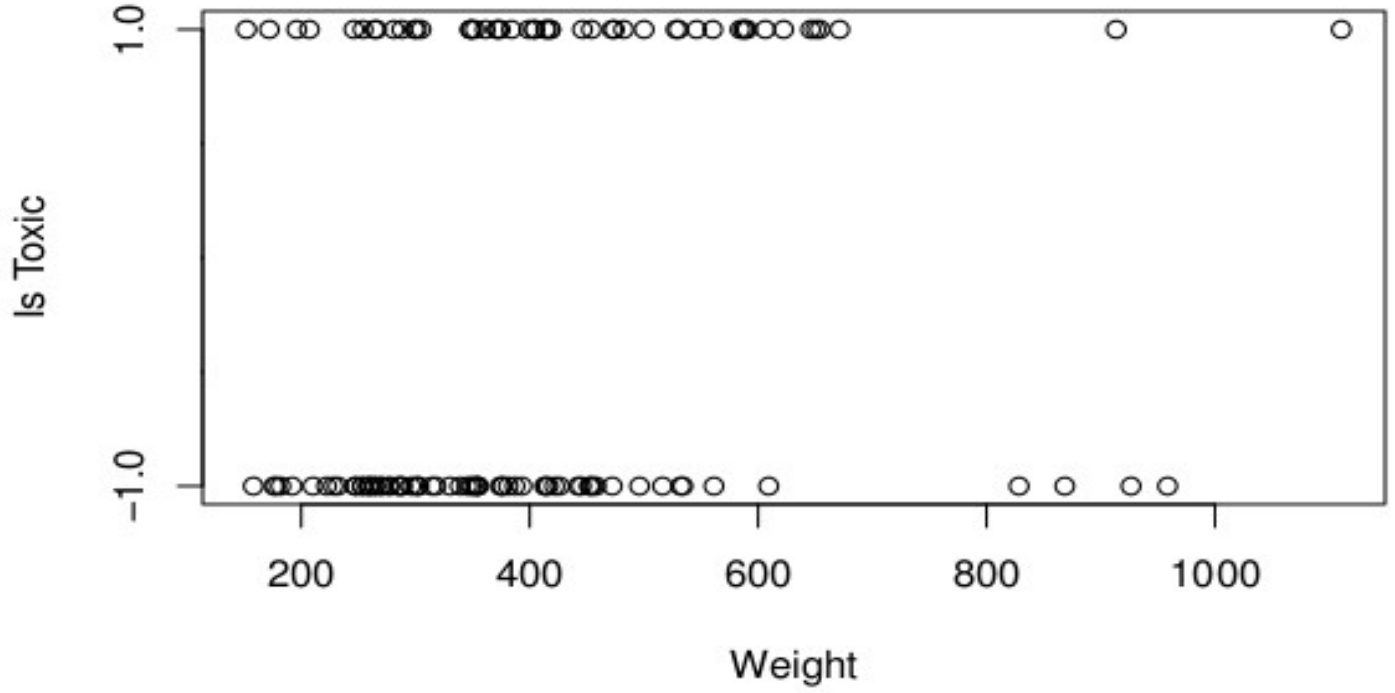
  → Toxic: ~20%, non-toxic: ~80%

Toxic vs non-toxic



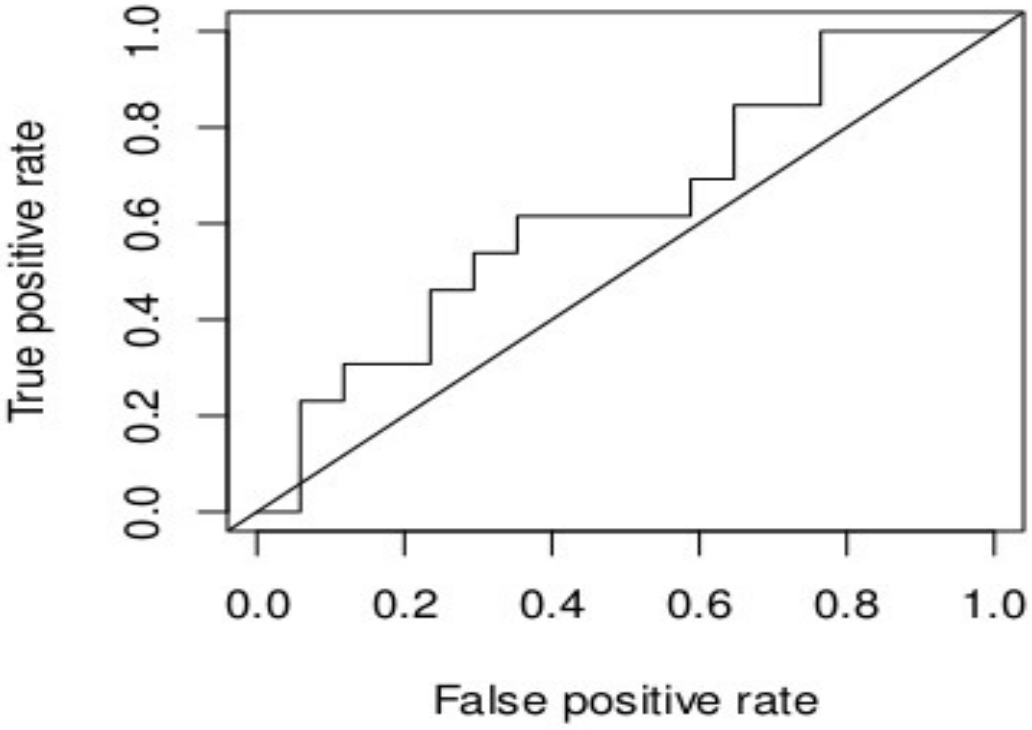
Density plots for all (black), toxic (red), and non-toxic (green) Compounds (relative).

Larger molecules more toxic?



Toxic (1) and non-toxic (-1) as function of the molecule weight. A small shoulder is visible, but not enough to get regression (see next slide).

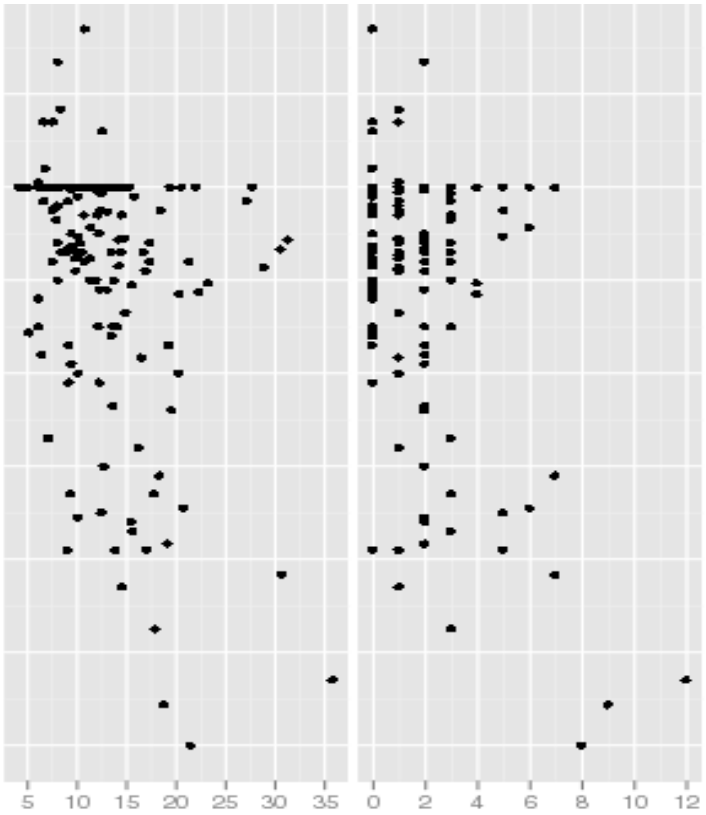
Toxicity cannot be predicted for our molecular structures



Receiver Operator curve plot, which should show a steeply ascending curve.

Classification methods cannot predict if a compound is toxic.

The “best” descriptors...



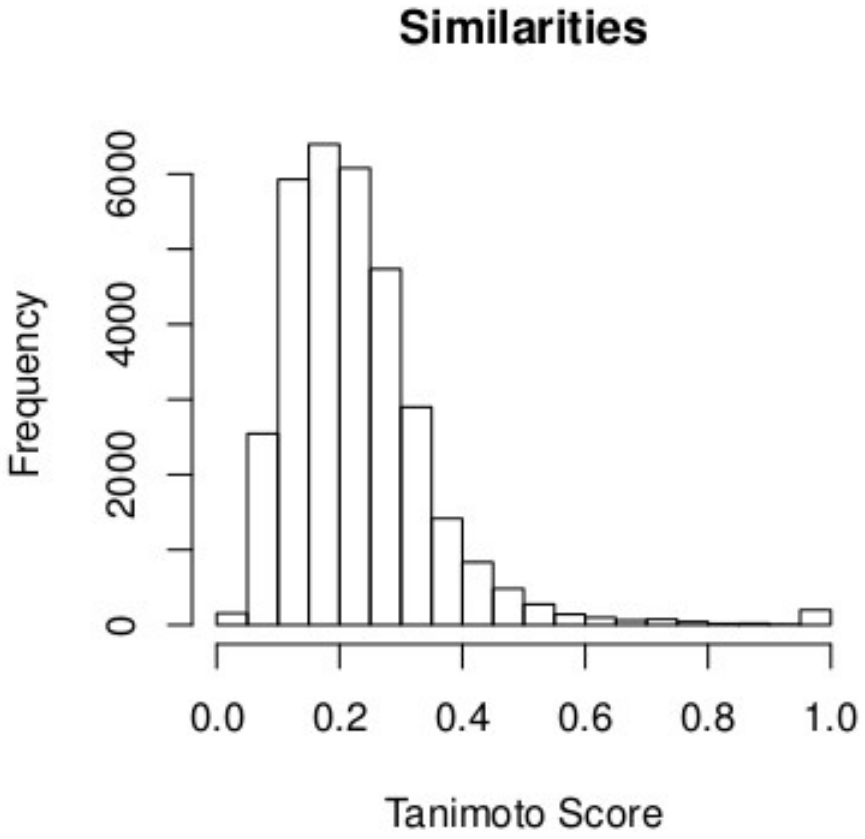
Skeletal variation

Number of double bonded carbons

Even the best QSAR descriptors show little correlation (*0.38* and *0.4*).

We saw the same for Molecular weight earlier.

Structural diversity is too high?

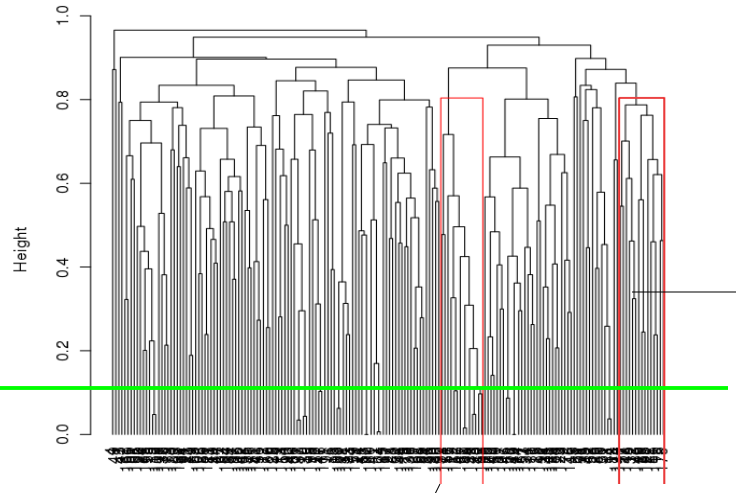


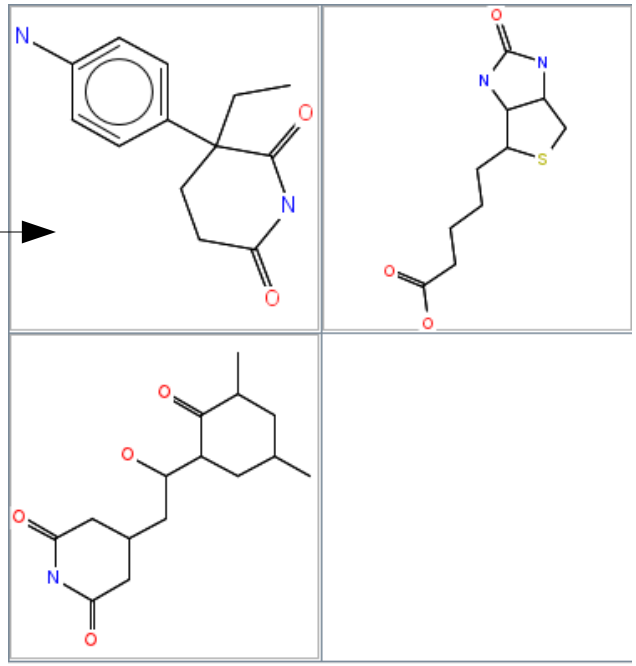
Structure are likely too dissimilar that there are no structure activity patterns.

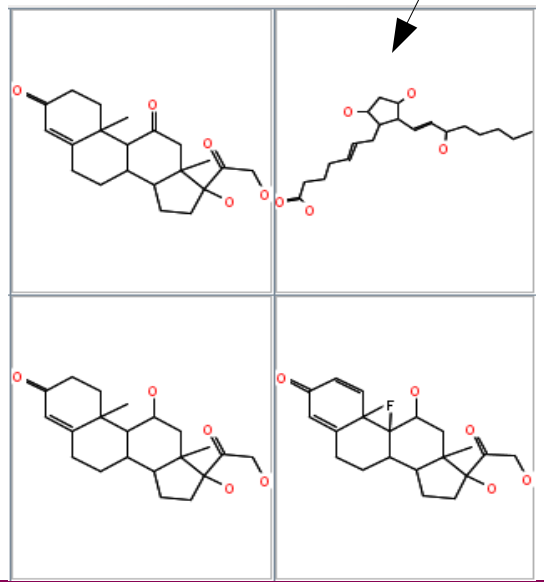
Structure diversty

Dissimilar compounds

Similar compounds







The ***high diversity*** of molecules makes it harder for the QSAR modeling to find patterns.