**Generating Focussed Molecule Libraries for Drug**

**Discovery with Recurrent Neural Networks**

This is particularly

true for medicinal chemistry. However, creating

novel drugs is an extraordinarily hard and complex

problem.[2] One of the many challenges in drug

design is the sheer size of the search space for novel

molecules. It has been estimated that 1060 drug-like

molecules could possibly be synthetically accessible.[3]

2. Schneider, P.; Schneider, G. J. Med. Chem. 2016, 59, 4077–4086.

3. Reymond, J.-L.; Ruddigkeit, L.; Blum, L.; van Deursen, R. Wiley

* Creating novel drugs is hard
* A lot of possibilities (10^60)
* Expensive to do screening in the lab
* Molecular design task:
  + create molecules, two main strategies:
    1. build molecules from predefined

groups of atoms or fragments

1. virtual chemical reactions based on expert

coded rules, hoping they mimic real life

* score and filter them, and
* search for better molecules, building on the knowledge

gained in the previous steps.

* **Predictions from rule based systems sometimes fail** (Segler, M.; Waller, M. P. manuscript submitted 2016,)
  + Rule based systems can restrict
* Two main ML approached:
  1. Target prediction

classifies molecules into active and inactive

1. quantitative structure-activity relationships (QSAR)

* Standard molecular descriptors:
  1. Signature Fingerprint Extended-Connectivity (ECFP)
  2. Atom Pair Fingerprints (APFP)
* New addition to the field of molecular descriptors; CNN on graphs
* Most widely used for target prediction: Random Forests and ANN
* Inverse QSAR, from properties to molecules not well defined because molecules are inherently discrete
* Here they use a completely new approach; RNN in a generative model