**Machine Learning for MOFs**

Descriptor definitions:

PLD – Pore Limiting Diameter

LCD – Largest Cavity Diameter

Density – g/cm3

VSA – Volumetric SA

GSA – Gravimetric SA

VF – Void Fraction

PV – Pore Volume

K0 – Henry Constants

DC – Diffusion Coefficients

P – Permeability

Qst – Heat of adsorption

Dataset definitions:

**MOF\_data.csv**

All 6768 MOFs

All 21 descriptors

**MOF\_data\_reduced.csv**

GSA, PV, K0s and Ps were discarded due to similarity to other descriptors

PLD, LCD and DCs were changed to a logarithmic scale

MIGKIW was removed due to infeasible structure

6767 MOFs

12 descriptors

**MOF\_data\_reduced2.csv**

MOF database further screened to remove infeasible structures

6638 MOFs

12 descriptors

**100absoluteloading.xlsx**

Selectivity of 99 MOFs, including top 8 from previous screen and 91 selected at random

**Inital\_screen\_data.xlsx**

Data for 34 MOFs from the initial screen, the 8 best MOFs are included in 100absoluteloading.xlsx

**6638\_cifs.csv**

Identity of the 6638 MOFs after infeasible structures were removed