Calculation of 1843 molecular descriptors (by SCBDD, ACD and TEST)

843 pesticides and their RTs (Wang et al., 2019)

Data set cleaning

No *a priori* ideas on which molecular descriptors to select

Literature review

16 molecular descriptors

7 strategies to select 16 MD:

Lasso, Pearson, RFE, PCA1 to PCA4

Dataset split into Internal and External set (3:1)

External dataset

(n = 198)

Internal dataset

(n = 594)

Training set (n = 445) & Test set (n = 148)

Optimized QSRR

applied to

Standardization

& Tuning the hyper parameters (cv = 5)

Performances

R² training set - RMSE training set

R² test set - RMSE test set

Performances

R² external set - RMSE external set

Feature importance analysis

**Figure 1** QSRR model development and evaluation of performances