# Protocolo seguido para generar el modelo de CYP3A4

## Endpoint

Inhibition of CYP3A4 (Classification)

## Origen de los datos

Los datos vienen del repositorio disponible en la web de Deep-PK:

* <https://biosig.lab.uq.edu.au/deeppk/data>

## Tratamiento de los datos

Del Excel original nos quedamos con las columnas “SMILES\_standarized” y “label”.

**Interpretation:**A compound is considered to be a cytochrome P450 inhibitor (Class 1) if the concentration required to inhibit P450 activity by 50% is less than 10 uM.

Después, y antes de pasar por Hygieia, se usó el script de Ágata (Elbow\_Undersampling.py) para quitar los outliers según peso molecular.

## Transformación de la “y”

No se ha llevado a cabo

## Train/test ratio

Se ha realizado una partición del 70% para el train y del 30% para el test. Durante el proceso de partición se han mantenido los compuestos señalados como *quite dissimilar*.

## Scaler

Se ha usado el *Standard Scaler*