**Response to referee 1**

**Although the meaning of the descriptors listed in tables 4 and 5 is given in the supplementary material, it is necessary to give a brief description of them in the core of the text. The meaning of the figures in parentheses has also to be provided.**

Please provide some description on the PCA descriptors for the main text.

**The discussion looks like a conclusion and as a result, a true conclusion is missing.**

Changed heading of Section 4.

**Response to referee 2**

**+ What was the S. thyphimurium strain for the second data set of 508 chemicals?**

For the 508 dataset the response variable is 0-1 mutagenicity status found by Ames mutagenicity test. We have made this clear (pg 5).

**+ Description of principal component analysis does not follow the classical matrix notation.**

Changed notation (pg 6).

**+ PCR and PLS were used as predictive tools but the numbers of PCs and latent variables retained in the models are not given. In the same way, Random Forest and Gradient Boosting Machine were used but the number of trees used in each approach is not given. In addition, the parameter settings leading to each model are not provided. The characteristics of the SCAD and Lasso models are also not given.**

Added subsection 3.1 and table 2 giving all information on the model settings.

**+ What is the mechanistic information supported by the models? Are these collections of descriptor allow us to better explain mutagenicity?**

Please give your inputs.

Finally, we have also corrected all the editorial issues mentioned.