REVEALING CYTOTOXIC SUBSTRUCTURES IN MOLECULES USING DEEP LEARNING

Henry E. Webel, Talia B. Kimber, Silke Radetzki,
Martin Neuenschwander, Marc Nazaré, Andrea Volkamer

Supplementary Material

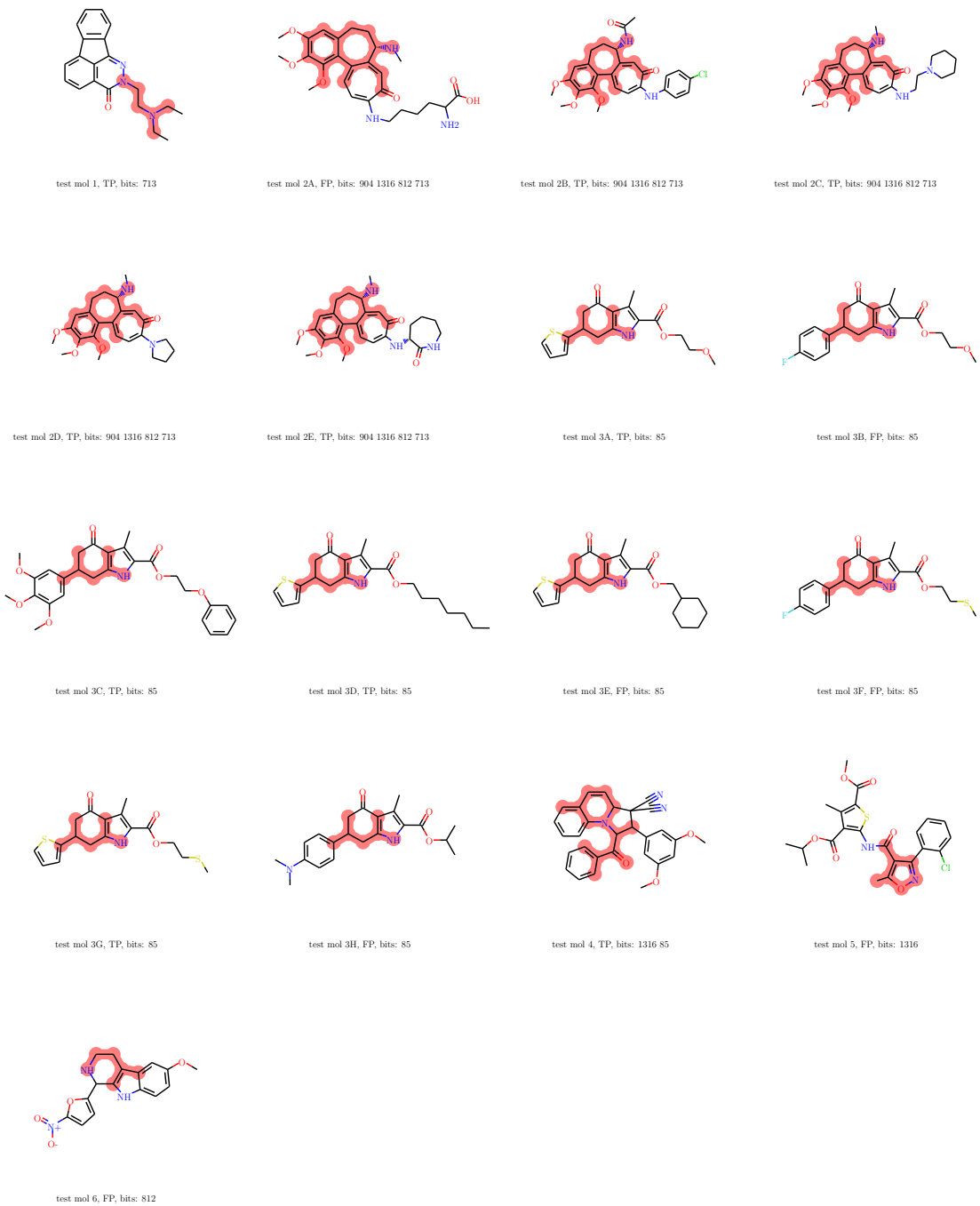

FIGURE S1. The Deep Taylor Decomposition was used on a feed-forward neural network to assign a relevance score to each atom environment, which can be used to identify potential toxicophores. The five highest scores were selected and are associated to bits 85, 713, 812, 904 and 1316. The figure shows molecules in the test set which contain at least one of the five atom environments, highlighted in red. The label for each molecule specifies its name, weather it was correctly predicted cytotoxic (TP) by the model or not (FP: False Positive) and lastly the bit(s) it contains.

TABLE S2. The table shows the IDs of the decomposable molecules in the test set sorted by decreasing order with respect to the FNN model prediction probability, the true value experimentally determined (1: toxic, 0: non-toxic) and the value predicted by the model (TP: true positive, FP: false positive).

Molecule ID	FNN Score	True Value	Predicted Value
2E	0.91	1	TP
2C	0.90	1	TP
2B	0.89	1	TP
2D	0.89	1	TP
2A	0.86	0	FP
4	0.83	1	TP
3C	0.79	1	TP
3D	0.78	1	TP
3A	0.78	1	TP
1	0.75	1	TP
3G	0.72	1	TP
6	0.69	0	FP
3B	0.68	0	FP
3H	0.67	0	FP
3E	0.61	0	FP
3F	0.61	0	FP
5	0.58	0	FP