QSAR model for *in vivo* gene mutation study in somatic cells (comet assay) (v1.0)

GenoITS

GenoITS is a computational workflow focused on the prediction of genotoxicity using the Integrated Testing Strategy proposed by REACH. GenoITS uses 5 different QSAR models to perform the complete workflow, one per each kind of study demanded by REACH (gene mutation in bacteria; *in vitro* cytogenicity; *in vitro* gene mutation; *in vivo* cytogenicity; *in vivo* gene mutation). ProtoITS also allows the users to supply their own experimental data.

GenoITS was developed as a part of the GenoQSAR research project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 101030422.

Endpoint

Human health effects: Mutagenicity/Genotoxicity. *In vivo* comet assay in mouse.

The comet assay is a method for measuring DNA strand breaks in eukaryotic cells. Single cells/nuclei embedded in agarose on a slide are lysed with detergent and high salt concentration. This lysis step digests the cellular and nuclear membranes and allows the release of coiled DNA loops generally called nucleoids and DNA fragments. Electrophoresis at high pH results in structures resembling comets, which, by using appropriate fluorescent stains, can be observed by fluorescence microscopy; DNA fragments migrate away from the "head" into the "tail" based on their size, and the intensity of the comet tail relative to the total intensity (head plus tail) reflects the amount of DNA breakage.

Metrics

Training set

Experimental values	QSAR predictions	
	non-mutagenic	mutagenic
non-mutagenic	90	10
mutagenic	5	80

Validation set

Experimental values	QSAR predictions	
	non-mutagenic	mutagenic
non-mutagenic	31	7
mutagenic	6	36

Parameters	Training	Validation
Accuracy	0.92	0.84
Sensitivity / recall	0.94	0.86
Specificity	0.90	0.82
Precision	0.89	0.84
Negative predictive value	0.95	0.84
F-score	0.91	0.85
Matthews Correlation Coefficient	0.84	0.67
Critical Success Index	0.84	0.73
Area under the ROC	0.92	0.84