GUSAR Models

Additionally to the GUSAR program, we provide ready-trained GUSAR models to predict certain biological activities. These are SAR bases that can be used with the GUSAR software for predictions on acute rat toxicity, acute mouse toxicity or antitargets (off-targets).

The acute rat or mouse toxicity SAR bases can be used for in silico prediction of LD50 values for rats or mouse with four types of administration. The training sets were created on the basis of the SYMYX MDL Toxicity Database and data from RTECS and ChemIDPlus.

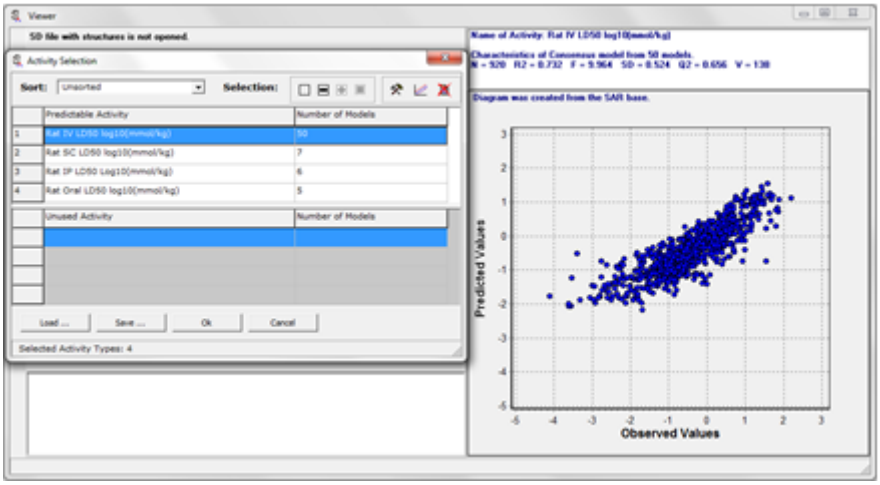
A quantitative prediction of antitarget interaction for chemical compounds can be done with the other SAR base. The QSAR models for the set of 32 activities (using IC_{50}, K_i or K_{act} values) includes data on about 4,000 chemical compounds interacting with 18 antitarget proteins (13 receptors, 2 enzymes and 3 transporters).

If you want to use these models, please send an e-mail to info@genexplain.com with your name, specifying which model you would like to use.

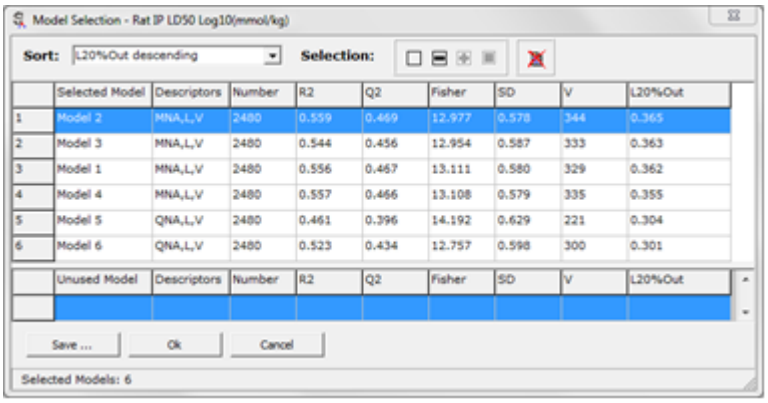
Acute Rat Toxicity

The SAR base on acute rat toxicity contains information about approximately 12,000 chemical structures with data on acute rat toxicity represented by LD50 (Lethal Dose, 50%) values for four types of administration:

- intravenous
- subcutaneous
- intraperitoneal
- oral



With this SAR base, you can make predictions about what concentration of a substance in focus is required to lead to LD50 in rats. If the log10 of this concentration is below zero, the substance is considered as toxic using the given way of administration.



As with SAR bases you created yourself, you can exclude whole activities and single QNA/MNA models that do not suit your needs with the "Selection" option. Thus, you are able to only use the most predictive models.

GUSAR can create a consensus model to predict one activity, or you can do batch predictions for all chosen activities at once.

Example: models for Rat IP LD50 Log10(mmol/kg).
MNA/QNA, L, V – MNA/QNA, topological length and volume of a molecule descriptors; **R^2** = coefficient of determination; **Q^2** = cross-validated R^2; **Fisher** = value of Fisher's statistics; **SD** = standard deviation; **V** = number of variables in the final regression equation; **L20%Out** = results of leave 20% out cross-validation.

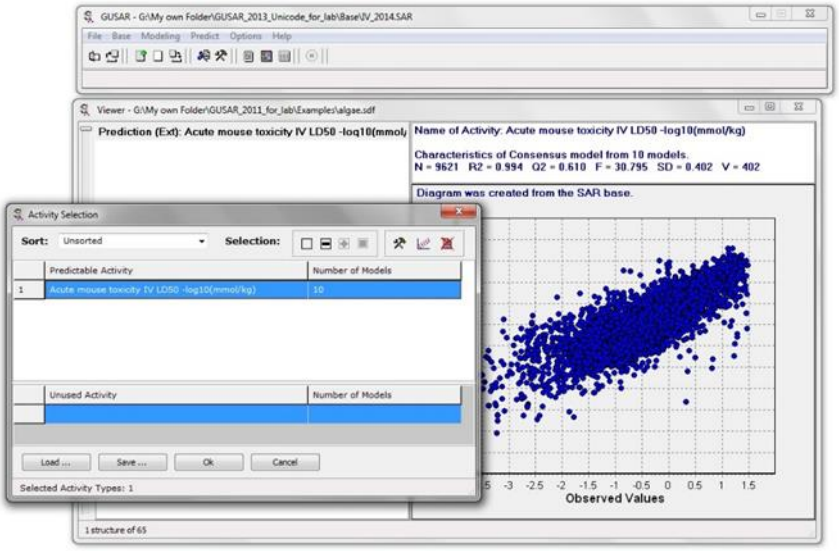
Characteristics of QSAR models for prediction of rat LD50 values (log10 (mmol/kg)) and their validation.
$N_{train/test}$ – number of compounds in the training/test set; **N_{models}** – number of models; **$R^2_{train/test}$** – average R^2 of the models calculated for the appropriate training/test set; **Q^2** – average Q^2 of the models calculated for the appropriate training set; **$RMSE_{test}$** – root mean square error of the test set; **Coverage** – % of compounds from the test set in Applicability Domain.

Administration	N_{train}	N_{test}	Nmodels		R^2_{train}	Q^2_{train}	R^2_{test}	$RMSE_{test}$	Coverage [%]
			QNA	MNA					
Oral	6280	2692	2	3	0.62	0.57	0.59	0.57	97.5
Intraperitoneal	2480	1065	2	4	0.63	0.54	0.57	0.57	96.1
Intravenous	920	394	10	40	0.73	0.66	0.63	0.62	99.2
Subcutaneous	759	325	2	5	0.69	0.59	0.50	0.69	92.0

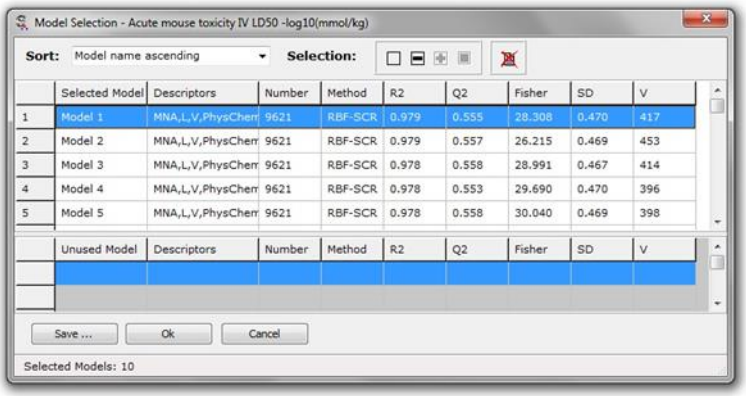
Acute Mouse Toxicity

The SAR base on acute mouse toxicity contains information about approximately 35,000 chemical structures with data on acute mouse toxicity represented by LD50 (Lethal Dose, 50%) values for four types of administration:

- intravenous
- subcutaneous
- intraperitoneal
- oral



With this SAR base, you can make predictions about what concentration of a substance in focus is required to lead to LD50 in rats. If the log10 of this concentration is below zero, the substance is considered as toxic using the given way of administration.



As with SAR bases you created yourself, you can exclude whole activities and single QNA/MNA models that do not suit your needs with the "Selection" option. Thus, you are able to only use the most predictive models.

GUSAR can create a consensus model to predict one activity, or you can do batch predictions for all chosen activities at once.

Example: models for Mouse IV LD50 Log10(mmol/kg).
MNA/QNA, L, V, PhysChem – MNA/QNA, topological length and volume of a molecule, and physicochemical properties;
R^2 = coefficient of determination; **Q^2** = cross-validated R^2; **Fisher** = value of Fisher's statistics; **SD** = standard deviation; **V** = number of variables in the final regression equation.

Characteristics of QSAR models for prediction of mouse LD50 values (log10 (mmol/kg)) and their validation.

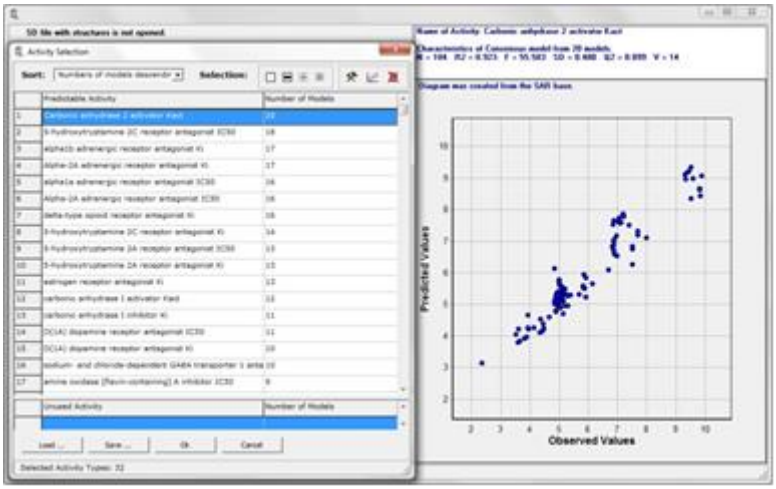
$N_{train/test}$ – number of compounds in the training/test set; **N_{models}** – number of models; **$R^2_{train/test}$** – average R^2 of the models calculated for the appropriate training/test set; **Q^2** – average Q^2 of the models calculated for the appropriate training set; **$RMSE_{test}$** – root mean square error of the test set.

Administration	N_{train}	N_{test}	N_{models}		R^2_{train}	Q^2_{train}	R^2_{test}	$RMSE_{test}$
			QNA	MNA				
Oral	18189	2020	1	1	0.99	0.58	0.57	0.40
Intraperitoneal	25086	2787	1	1	0.99	0.61	0.61	0.41
Intravenous	9621	1069	5	5	0.99	0.61	0.63	0.39
Subcutaneous	3215	357	5	5	0.99	0.63	0.63	0.53

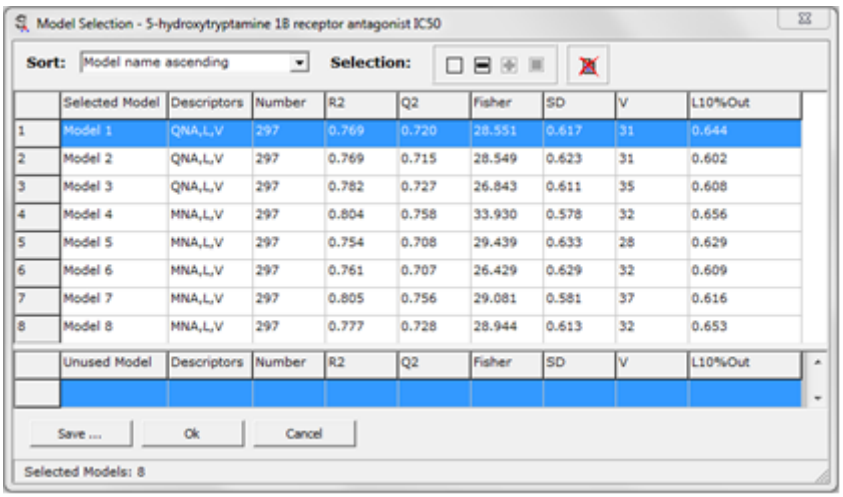
Antitargets (off-targets)

This is a SAR base on the affinity of substances to certain antitargets, containing 32 activities. The interactions of substances in focus with antitargets might be the cause of adverse or toxic effects.

- IC_{50} (50% of the inhibitory concentration)
- K_i (inhibition constant)
- K_{act} (activation constant)



With this SAR base, you can make predictions about what concentration of a substance is required to lead to an interaction (inhibition or activation) with one of the 18 antitarget proteins included (13 receptors, 3 transporters, 2 enzymes).



	Selected Model	Descriptors	Number	R2	Q2	Fisher	SD	V	L10%Out
1	Model 1	QNA,L,V	297	0.769	0.720	28.551	0.617	31	0.644
2	Model 2	QNA,L,V	297	0.769	0.715	28.549	0.623	31	0.602
3	Model 3	QNA,L,V	297	0.782	0.727	26.843	0.611	35	0.608
4	Model 4	MNA,L,V	297	0.804	0.758	33.930	0.578	32	0.656
5	Model 5	MNA,L,V	297	0.754	0.708	29.439	0.633	28	0.629
6	Model 6	MNA,L,V	297	0.761	0.707	26.429	0.629	32	0.609
7	Model 7	MNA,L,V	297	0.805	0.756	29.081	0.581	37	0.616
8	Model 8	MNA,L,V	297	0.777	0.728	28.944	0.613	32	0.653

	Unused Model	Descriptors	Number	R2	Q2	Fisher	SD	V	L10%Out

Example: models for 5-hydroxytryptamine 1B receptor antagonist IC50. **QNA/MNA, L, V** – QNA/MNA, topological length and volume of a molecule descriptors; **R^2** = coefficient of determination; **Q^2** = cross-validated R^2; **Fisher** = value of Fisher's statistics; **SD** = standard deviation; **V** = number of variables in the final regression equation; **L10%Out** = results of leave 10% out cross-validation.

As with SAR bases you created yourself, you can exclude whole activities and single QNA/MNA models that do not suit your needs with the "Selection" option. Thus, you are able to only use the most predictive models.

GUSAR can create a consensus model to predict one activity, or you can do batch predictions for all chosen activities at once.

Characteristics of QSAR models for antitarget sets.
$N_{train/test}$ – number of compounds in the training/test set; N_{models} – number of models; $R^2_{train/test}$ - average R^2 of the models calculated for the appropriate training/test set; Q^2 - average Q^2 of the models calculated for the appropriate training set; $RMSE_{test}$ – root mean square error of the test set; **Coverage** - % of compounds from the test set in Applicability Domain.

Activity	N_{train}/N_{test}	N_{models}		R^2_{train}	Q^2_{train}	R^2_{test}	Coverage [%]
		QNA	MNA				
5-hydroxytryptamine 1B receptor antagonist IC_{50}	297/74	3	5	0.83	0.79	0.67	100.0
5-hydroxytryptamine 1B receptor antagonist K_i	266/66	3	4	0.73	0.66	0.72	100.0
5-hydroxytryptamine 2A receptor antagonist IC_{50}	555/143	6	7	0.83	0.78	0.71	98.6
5-hydroxytryptamine 2A receptor antagonist K_i	1010/252	3	10	0.72	0.65	0.59	99.6
5-hydroxytryptamine 2C receptor antagonist IC_{50}	128/32	3	15	0.77	0.73	0.58	100.0
5-hydroxytryptamine 2C receptor antagonist K_i	487/121	6	8	0.74	0.66	0.62	99.2
alpha1a adrenergic receptor antagonist IC_{50}	438/111	6	10	0.79	0.73	0.72	98.2
alpha1a adrenergic receptor antagonist K_i	1366/344	1	4	0.83	0.79	0.80	97.0
alpha1b adrenergic receptor antagonist K_i	410/102	5	12	0.73	0.66	0.63	100.0
alpha-2A adrenergic receptor antagonist IC_{50}	109/207	5	11	0.88	0.84	0.75	100.0
alpha-2A adrenergic receptor antagonist K_i	525/131	4	13	0.84	0.79	0.77	99.2
amine oxidase [flavin-containing] A inhibitor IC_{50}	186/71	4	5	0.80	0.75	0.72	100.0
amine oxidase [flavin-containing] A inhibitor K_i	60/15	2	3	0.73	0.62	0.64	100.0
androgen receptor antagonist IC_{50}	116/29	2	6	0.79	0.73	0.67	100.0
carbonic anhydrase I activator K_{act}	108/27	4	8	0.98	0.97	0.93	100.0
carbonic anhydrase I inhibitor K_i	935/234	4	7	0.91	0.86	0.86	98.3
carbonic anhydrase II activator K_{act}	104/26	6	14	0.92	0.90	0.91	100.0
carbonic anhydrase II inhibitor K_i	866/217	1	6	0.87	0.79	0.76	98.6
d(1A) dopamine receptor antagonist IC_{50}	126/31	2	9	0.76	0.72	0.80	100.0
d(1A) dopamine receptor antagonist K_i	291/73	4	6	0.72	0.66	0.57	100.0
d3 dopamine receptor antagonist K_i	822/206	3	6	0.73	0.66	0.62	98.0
delta-type opioid receptor antagonist K_i	1044/261	3	13	0.75	0.70	0.65	98.5
estrogen receptor antagonist IC_{50}	402/100	2	2	0.66	0.61	0.70	97.0
estrogen receptor antagonist K_i	255/68	2	11	0.76	0.71	0.70	100.0
kappa-type opioid receptor antagonist K_i	844/221	3	4	0.74	0.67	0.65	100.0
mu-type opioid receptor antagonist IC_{50}	545/136	2	5	0.67	0.61	0.70	97.8
mu-type opioid receptor antagonist K_i	1354/338	1	3	0.69	0.62	0.60	96.7
sodium- and chloride-dependent GABA transporter 1 antagonist IC_{50}	79/19	2	8	0.90	0.86	0.89	100.0
sodium-dependent dopamine transporter antagonist IC_{50}	920/230	3	2	0.70	0.65	0.67	98.3
sodium-dependent dopamine transporter antagonist K_i	655/164	3	4	0.77	0.69	0.64	100.0
sodium-dependent serotonin transporter antagonist IC_{50}	796/199	5	2	0.80	0.75	0.69	97.5
sodium-dependent serotonin transporter antagonist K_i	823/206	1	1	0.72	0.65	0.61	95.6