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 <u>info@genexplain.com</u> 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

Vever		0 8 1
50 file with structures is not opened.		Name of Activity: Rat IV LD50 log10(annob/hg)
Activity Selection		Characteristics of Concensus model from 50 models. N = 520 R2 = 0.732 F = 5.964 SD = 0.524 Q2 = 0.656 V = 138
Sort: Unsoled Selection:		Disease was constrol from the SAR base.
Predictable Activity	Number of Models	
Rat IV LDSD log10(mmo0kg)	50	3
Rat SC LD50 log30(mmol/kg)	7	
Rat IP LD50 Log10(mmol/kg)	6	2
Rat Onal LD50 log10(mmol/kg)	5	state.
Unused Adjuity	Number of Hodels	
		3 amount and a second and a s
		2 4
	- 1	2
Lond Seve Ok Cards	·	
Selected Activity Types: 4		
-		4
		4
		Observed Values

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.

Selected Model Descriptors Number R2 Q2 Fisher SD V L20%Ou Model 2 MNA,L,V 2480 0.559 0.469 12.977 0.578 344 0.365 Model 3 MNA,L,V 2480 0.556 0.465 12.954 0.587 333 0.363 Model 1 MNA,L,V 2480 0.556 0.467 13.111 0.580 329 0.362 Model 4 MNA,L,V 2480 0.557 0.466 13.108 0.579 335 0.353 Model 5 QNA,L,V 2480 0.451 0.396 14.192 0.629 221 0.305							-	ice wing	a previous ou	0.01
Model 2 MNA,L,V 2480 0.359 0.469 12.977 0.578 344 0.365 Model 3 MNA,L,V 2480 0.544 0.456 12.977 0.578 343 0.363 Model 1 MNA,L,V 2480 0.556 0.467 13.111 0.580 329 0.362 Model 4 MNA,L,V 2480 0.557 0.466 13.108 0.579 335 0.362 Model 5 QNA,L,V 2480 0.451 0.396 14.192 0.629 2211 0.305 Model 4 MNA,L,V 2480 0.451 0.396 14.192 0.629 0.355	L20%Out	v	SD	Fisher	Q2	R2	Number	Descriptors	Selected Model	
Model 3 MNA,L,V 2480 0.544 0.456 12.954 0.587 333 0.363 Model 1 MNA,L,V 2480 0.556 0.467 13.111 0.580 329 0.362 Model 4 MNA,L,V 2480 0.557 0.466 13.108 0.579 335 0.355 Model 5 QNA,L,V 2480 0.461 0.396 14.192 0.629 0.305	 0.365	344	0.578	12.977	0.469	0.559	2480	MNA,L,V	Model 2	
Model 1 MNA,L,V 2480 0.556 0.467 13.111 0.580 329 0.362 Model 4 MNA,L,V 2480 0.557 0.466 13.108 0.579 335 0.355 Model 5 QNA,L,V 2480 0.461 0.396 14.192 0.629 221 0.304	0.363	333	0.587	12.954	0.456	0.544	2480	MNA,L,V	Model 3	
Model 4 MNA,L,V 2480 0.557 0.466 13.108 0.579 335 0.355 Model 5 QNA,L,V 2480 0.461 0.396 14.192 0.629 221 0.304 Model 5 QNA,L,V 2480 0.461 0.396 14.192 0.629 221 0.304	0.362	329	0.580	13.111	0.467	0.556	2480	MNA,L,V	Model 1	_
Model S QNA,L,V 2480 0.461 0.396 14.192 0.629 221 0.304	 0.355	335	0.579	13.108	0.466	0.557	2480	MNA,L,V	Model 4	
	 0.304	221	0.629	14.192	0.396	0.461	2480	QNA,L,V	Model S	
Model 6 Qrat, V 2460 0.525 0.454 12.757 0.546 300 0.501	0.301	300	0.598	12.757	0.434	0.523	2480	QNA,L,V	Model 6	
Unused Model Descriptors Number R2 Q2 Fisher SD V L20%Ou	L20%Out	v	SD	Fisher	Q2	R2	Number	Descriptors	Unused Model	

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; \mathbf{R}^2 = coefficient of determination; \mathbf{Q}^2 = cross-validated R²; **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	N	Nmo	odels	D2	O ²	D 2	DMCE	
Administration	INtrain	Ntest	QNA	MNA	K ⁻train			RIVIJEtest	Coverage [%]
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:



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.

Sor	t: Model name a	ascending	• Sele	ction:		王周)			
	Selected Model	Descriptors	Number	Method	R2	Q2	Fisher	SD	V	1.
1	Model 1	MNA,L,V,PhysChem	9621	RBF-SCR	0.979	0.555	28.308	0.470	417	
2	Model 2	MNA,L,V,PhysChem	9621	RBF-SCR	0.979	0.557	26.215	0.469	453	
3	Model 3	MNA,L,V,PhysChem	9621	RBF-SCR	0.978	0.558	28.991	0.467	414	
4	Model 4	MNA,L,V,PhysChem	9621	RBF-SCR	0.978	0.553	29.690	0.470	396	
5	Model 5	MNA,L,V,PhysChen	9621	RBF-SCR	0.978	0.558	30.040	0.469	398	
	Unused Model	Descriptors	Number	Method	R2	Q2	Fisher	SD	v	1
	Save	ok Ca	ancel							

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; \mathbf{R}^2 = coefficient of determination; \mathbf{Q}^2 = cross-validated R²; 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	м	Nm	odels	D 2	O ²	D2	DMCE
Administration	INtrain	Ntest	QNA	MNA	™ train	₩ [−] train	™ [−] test	RIVIJEtest
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.



- IC₅₀ (50% of the inhibitory concentration)

- K_i (inhibition constant)
- Kact (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).

iort	: Model name	ascending	•	Select	ion:		- X			
	Selected Model	Descriptors	Number	R2	Q2	Fisher	SD	V	L10%Out	Т
	Model 1	QNA,L,V	297	0.769	0.720	28.551	0.617	31	0.644	
	Model 2	QNA,L,V	297	0.769	0.715	28.549	0.623	31	0.602	
	Model 3	QNA,L,V	297	0.782	0.727	26.843	0.611	35	0.608	_
	Model 4	MNA,L,V	297	0.804	0.758	33.930	0.578	32	0.656	_
	Model 5	MNA,L,V	297	0.754	0.708	29.439	0.633	28	0.629	_
	Model 6	MNA,L,V	297	0.761	0.707	26.429	0.629	32	0.609	_
	Model 7	MNA,L,V	297	0.805	0.756	29.081	0.581	37	0.616	
	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	Т

Example: models for 5-hydroxytryptamine 1B receptor antagonist IC50. **QNA/MNA, L, V** – QNA/MNA, topological length and volume of a molecule descriptors; \mathbf{R}^2 = coefficient of determination; \mathbf{Q}^2 = cross-validated R²; **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	NI /NI	Nmodels		R ² turnin	O^{2}	D 2	Coverage [%]	
Activity	INtrain/INtest	QNA	MNA	K ⁻train	Q [−] train	R ⁻test	Coverage [%]	
5-hydroxytryptamine 1B receptor antagonist IC50	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 Ki	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 Ki	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 IC50	186/71	4	5	0.80	0.75	0.72	100.0	
amine oxidase [flavin-containing] A inhibitor Ki	60/15	2	3	0.73	0.62	0.64	100.0	
androgen receptor antagonist IC50	116/29	2	6	0.79	0.73	0.67	100.0	
carbonic anhydrase I activator Kact	108/27	4	8	0.98	0.97	0.93	100.0	
carbonic anhydrase I inhibitor Ki	935/234	4	7	0.91	0.86	0.86	98.3	
carbonic anhydrase II activator Kact	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 Ki	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 Ki	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 Ki	255/68	2	11	0.76	0.71	0.70	100.0	
kappa-type opioid receptor antagonist Ki	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 Ki	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	