Cheminformatics at geneXplain GmbH

Introducing GUSAR

geneXplain GmbH

We can explain your genes and put bricks together...

• Gene regulation • Biomarkers • Drug activity • Network modeling •

Products

▸ bioinformatics and systems biology: geneXplain platform

▸ bioinformatics and NGS: IMC and GenomeTraveler

▸ **cheminformatics:** PASS, PharmaExpert, **GUSAR**

Bioinformatics, Systems Biology, NGS

▸ geneXplain platform

  ◦ collection, storage and analysis of experimental data

  ◦ network clustering and search for master regulators

  ◦ features: graphical programming of workflows and the possibility to write new scripts and add-ons

▸ IMC

  ◦ handling of whole genome data

  ◦ supports: feature mapping and annotation

  ◦ PCR primer calculation

▸ GenomeTraveler

  ◦ handling of next generation sequencing (NGS) data

  ◦ genome analysis

  ◦ all functions of IMC

Contents

GUSAR

>> General Unrestricted Structure-Activity Relationships

General Information

▸ GUSAR…

　◦ creates models on quantitative structure-activity relationships.

　◦ uses 2D chemical structures from SD files for model building and activity prediction.

　◦ can work with large data sets containing up to 30,000 chemical compounds.

　◦ lets you select the most predictive models.

authorship and copyright by Zakharov A.V., Filimonov D.A., Poroikov V.V., Lagunin A.A., Moscow, Russia

distributed by the geneXplain GmbH

Key Data

▸ ## self-consistent regression

  ◦ unique algorithm based on the statistical regularization of ill-posed problems

  ◦ allows the selection of the best set of descriptors for robust and reliable QSAR models

▸ ## „Star Track" approach

  ◦ also unique to GUSAR

  ◦ represents any molecule as a set of points (atoms) in a 2D QNA descriptor space

▸ ## MNA and QNA descriptors

Filimonov D.A., Akimov D.V., Poroikov V.V. (2004) *Pharmaceutical Chemistry Journal* 38:21-24.
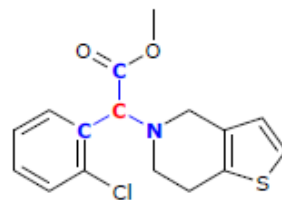Filimonov D.A., Zakharov A.V., Lagunin A.A., Poroikov V.V. (2009) *SAR QSAR Environ. Res.* 20:679-709.
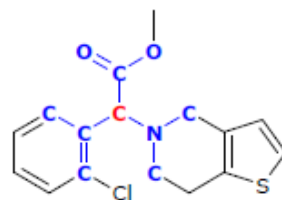
MNA Descriptors

Multilevel Neighborhood of Atoms
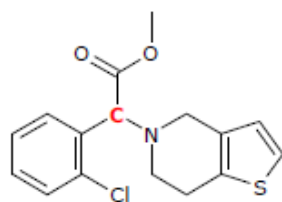
MNA/0: -C

MNA/1: -C(CN-H-C)
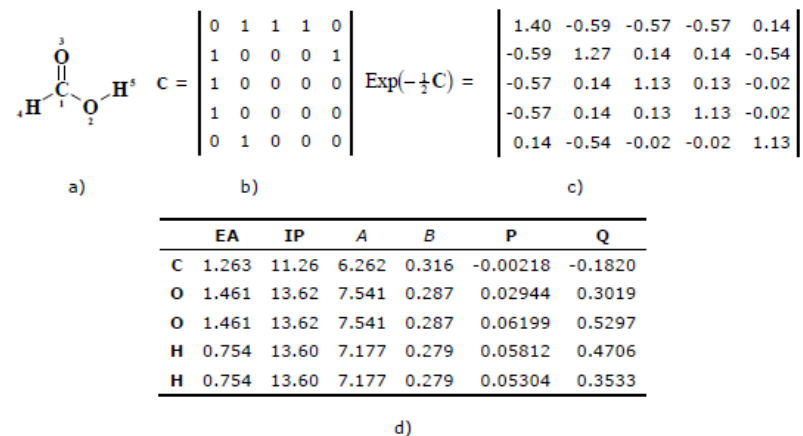
MNA/2: -C(C(CC-C)N(CC-C)-H(-C)-C(-C-O-O))

Filimonov D.A. et al. (1999) *J. Chem. Inform. Computer Sci.*, 39, 666.
Filimonov D.A., Poroikov V.V. (2008) *Chemoinformatics Approaches to Virtual Screening. RSC Publ.*, p.182-216.

QNA Descriptors

Quantitative Neighborhood of Atoms Descriptors

- ◦ reflect the nature of intermolecular interactions

- ◦ describe every atom and the atoms' influences on each other

- ◦ are based on P and Q values calculated with a connectivity matrix (see *d)* in the example on the right)

	EA	IP	A	B	P	Q
C	1.263	11.26	6.262	0.316	-0.00218	-0.1820
O	1.461	13.62	7.541	0.287	0.02944	0.3019
O	1.461	13.62	7.541	0.287	0.06199	0.5297
H	0.754	13.60	7.177	0.279	0.05812	0.4706
H	0.754	13.60	7.177	0.279	0.05304	0.3533

d)

EA = electron affinity; IP = ionization potential; A, B = variables.

Filimonov D.A., Zakharov A.V., Lagunin A.A., Poroikov V.V. (2009) *SAR QSAR Environ. Res.* 20:679-709.

Substructures and Activity Prediction
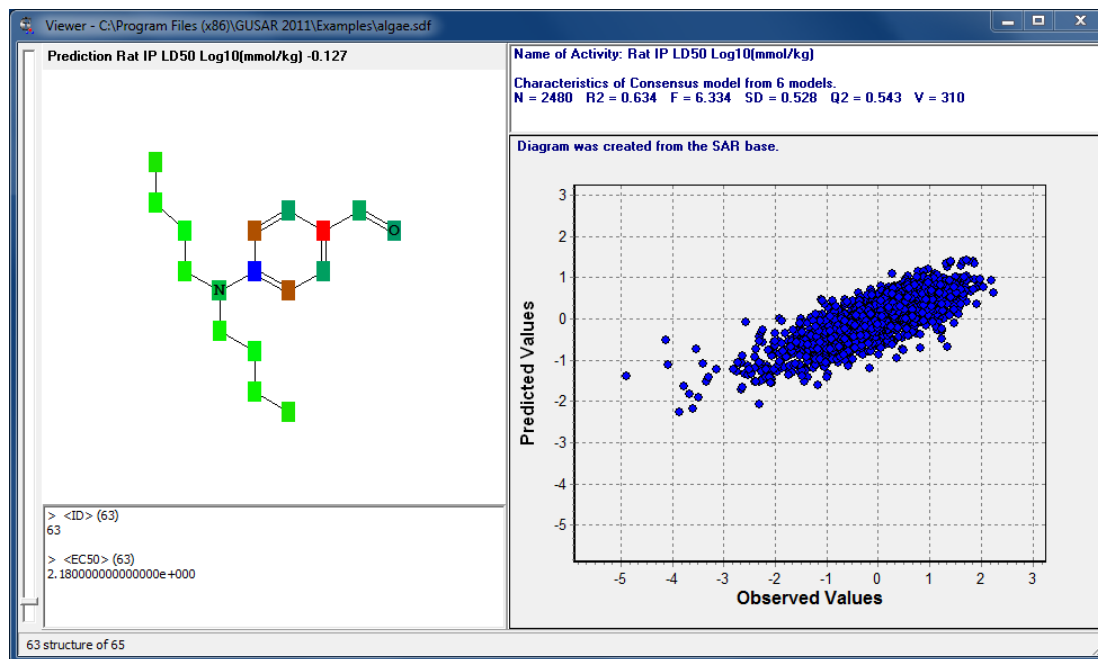
- • for green atoms, the predicted value is equal to that of the whole molecule

- • for blue atoms, the predicted value is less than that of the whole molecule

- • for red atoms, the predicted value is higher than that of the whole molecule

consensus model of 4 MNA and 2 QNA models on acute rat toxicity

- • numbers are the characteristics of the consensus model, including the number of substances used and statistical values

- • the graph shows predicted versus observed values from the SAR base

Summary

Create multiple (Q)SAR models for large data sets.

Select the most predictive QNA or MNA models.

GUSAR is a QSAR tool with unique algorithms for model creation and activity prediction.

Upload SD files for batch prediction or evaluate results using the GUSAR interface.

Predictions made with GUSAR were comparable or better than those of other QSAR methods.

Models

» Acute Rat Toxicity
  Affinity to Antitargets

Ready-Trained Models

▸ Two ready-trained QSAR model bases can be provided additionally to the software

- ◦ acute rat toxicity
- ◦ affinity to antitargets (off-targets)

Acute Rat Toxicity

▸ activities for rat toxicity in LD50
- ◦ LD50 = 50% of lethal dose

▸ 4 types of administration:
- ◦ intravenous
- ◦ subcutaneous
- ◦ intraperitoneal
- ◦ oral

▸ between 5 and 50 models using QNA or MNA descriptors

Viewer

Name of Activity: Rat IV LD50 log10[mmol/kg]

Characteristics of Consensus model from 50 models.
N = 920   R2 = 0.732   F = 9.964   SD = 0.524   Q2 = 0.656   V = 138

Diagram was created from the SAR base.
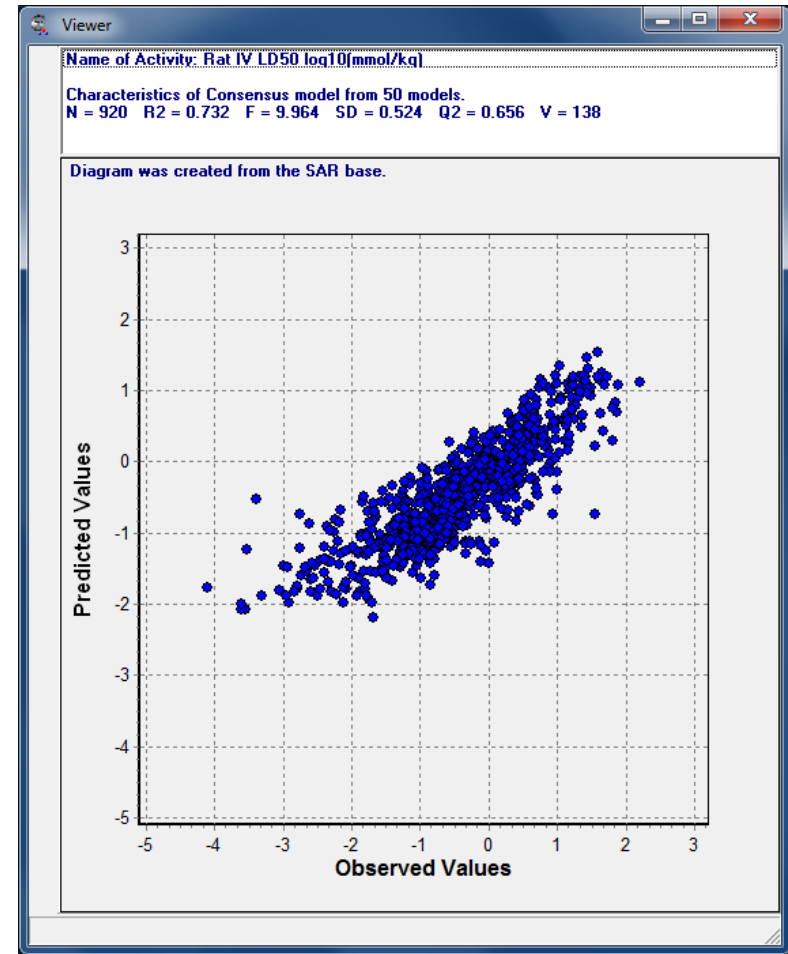
Affinity to Antitargets

- **32 activities in 3 values**
  - IC_{50} = half-maximal inhibitory concentration
  - K_i = dissociation constant
  - K_{act} = activation constant

- **18 different antitarget proteins**

- **between 2 and 20 models using QNA or MNA descriptors**

Viewer

Name of Activity: Carbonic anhydrase 2 activator Kact

Characteristics of Consensus model from 20 models.
N = 104 R2 = 0.923 F = 55.503 SD = 0.488 Q2 = 0.899 V = 14

Diagram was created from the SAR base.

Predicted Values vs Observed Values

Summary

Acute Rat Toxicity

Affinity to Antitargets

Two additional SAR bases to predict specific activities can be provided additionally.

Save the time needed to create models and SAR bases.

Explore GUSAR functions using the diverse, ready-trained models.

Why You Should Give GUSAR a Try

>> Advantages of the Software

Advantages of GUSAR

▸ Only the structural formula of a compound is required to predict its biological activity.

▸ The software is installed locally and runs on any ordinary PC.

▸ Additionally to the QNA and MNA models you create, we can provide you with ready-trained GUSAR models on

◦ acute rat toxicity (oral, intraperitoneal, intravenous and subcutaneous routes of administration) and

◦ affinity to certain antitargets that might cause adverse/toxic effects.

▸ The algorithm is fast and can handle large data sets.

▸ In comparison with a number of 2D and 3D QSAR methods, the predictivity of GUSAR was superior to that of most other methods on both heterogeneous and homogenous data sets.

Filimonov D.A., Zakharov A.V., Lagunin A.A., Poroikov V.V. (2009) *SAR QSAR Environ. Res.* 20:679-709.