TerraQSAR™ - *Daphnia*

Klaus L.E. Kaiser
TerraBase Inc., Hamilton, Ontario, Canada
www.terrabase-inc.com
email [see web site]

Abstract

TerraQSAR™ - *Daphnia* is a new, neural network-based computation program to predict the 48-hr LC50 of organic compounds for Daphnia magna. It is based on measured values for approximately 1000 individual chemicals.
Program Input: 2D/3D SMILES string.
Program Output: 48-hr LC50, in mg/L, 48-hr LC50 in log(1/mM), and molecular weight.

Introduction

Recent developments of neural networks (NNs) provide for the creation of NN-based QSAR programs. In particular, NNs based on the principle of estimation of the conditional average are well suited to estimate toxicological endpoints of interest.

Data

As the training set for the TerraQSAR – *Daphnia* program, measured data for approximately 1000 individual compounds have been selected from the **TerraTox™ – Explorer** database.

Methodology

TerraQSAR programs use the probabilistic neural network (PNN) methodology which is based on the **estimation of the conditional average**. Unlike most other neural network methodologies, the PNN training is 100% repeatable in every aspect of the training and application. Furthermore, and in contrast to entirely linear methodologies, such as regression methods, the PNN system uses both linear and non-linear relationships. Therefore, the PNN is particularly well suited for the computation of effects for compounds with unknown or multiple modes of action. Examples demonstrating the superiority of the PNN methodology are found at our website at www.terrabase-inc.com.

Fragments

Major fragments used in the TerraQSAR modules have been described in detail in several publications listed in the literature. An overview of the **types** of fragments used is given in Table 1. Fig. 1, gives a plot of the measured *versus* predicted data for the entire training set.

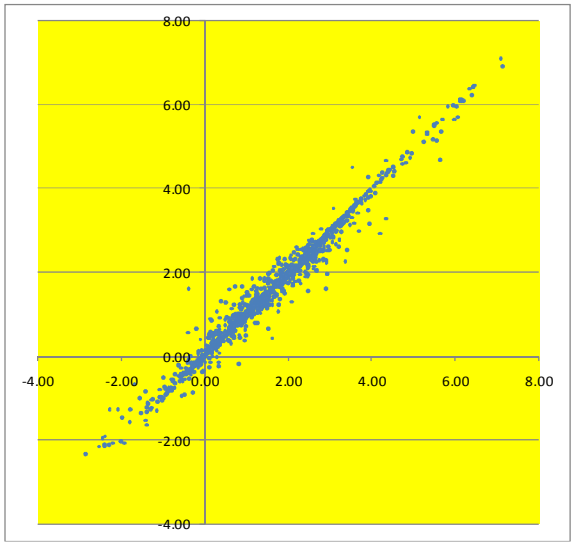

Fig. 1. Plot of measured (abscissa) *versus* predicted (ordinate) values of the ~1000 data in the TerraQSAR - *Daphnia* training set. Leave-one-out cross-validation RMSE = 0.18.

Table 1. Fragment types used.

Fragment type	Examples
Acidity fragment	C(=O)O, S(=O)(=O)O
Aliphatic ring fragment	C1CCCCC1, C1CC1
Aromatic ring fragment	c1ccccc1, c1ccccn1
Atom fragment	C, H, N, O
Bond fragment	C-C, C=C, C#C
Group fragment	C-O-H, C-O-C
Hydrophobicity fragmt.	C(C)(C)C, CCCC
Ionization fragment	[O-], [Na+]
Polarity fragment	O=N(=O)CC(O)
Reactivity fragment	C=CC=O
Stereo fragment	CC[C@H](C)N
Weight fragment	molecular weight

Example computation

The antibiotic compound Olivomycin A, CAS 11006-70-5, Fig. 2, has a predicted *Daphnia magna* 48-hr LC50 of 6.47 mg/L and a pT= log(1/LC50) of 2.16.

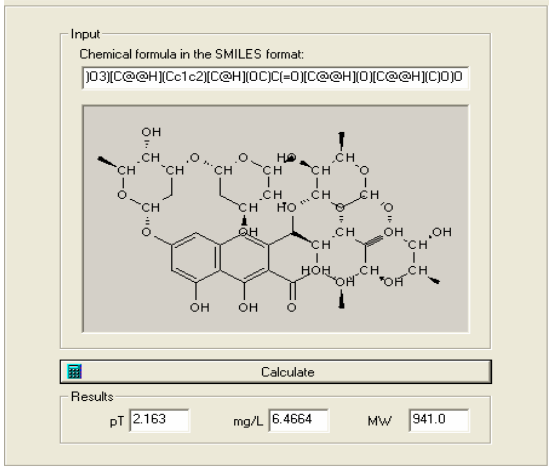

Fig. 2. The antibiotic compound Olivomycin A, CAS 11006-70-5, SMILES code
CO[C@H]1[C@@H](O)C[C@@H](O[C@H]1C)O[C@@H]1C[C@@H](O[C@@H](C)[C@@H]1O)Oc1cc2c(c(c1)O)c(c1C(=O)[C@@H](O[C@@H]3C[C@H](O[C@@H]4C[C@H](O)[C@@H](O)[C@@H](C)O4)[C@@H](O)[C@@H](C)O3)[C@@H](Cc1c2)[C@H](OC)C(=O)[C@@H](O)[C@@H](C)O)O

Invitation

Registered participants of *QSAR 2008, the 13th International Workshop on QSARs in the Environmental Sciences* are invited to submit structures for free estimation. Structures can be submitted by email in the following formats: (i) as sdf files, (ii) as text files with SMILES code, or (iii) byFAX as drawn structures.

References

Important information on both the theory and specific aspects of this software can be found in the references in the literature section of the TerraQSAR – *Daphnia* Manual, available free at www.terrabase-inc.com.

TerraBase Inc.

June 2008