iSafeRat® HA-QSARs vs. commonly used predictive models: A statistical comparison

Faizan SAHIGARA, Pascal BICHEREL, Paul THOMAS

KREATiS, Knowledge & Research in Environment And Toxicology in Silico

KREATiS SAS, Bât. Corvette, 80 Rue Condorcet, 38090 Vaulx Milieu, France

www.kreatis.eu

Replacing Experimentation

Introduction

As an alternative to animal testing, provisions for Quantitative Structure Activity Relationship (QSAR) predictions for regulatory purposes are well-discussed and documented within the framework of REACH regulations for the safe use of chemicals[1]. As a result, several QSAR models have been made available freely/commercially to the market over the last few years and this has generated great interest from the chemical industry competent authorities and academia. Despite efforts to integrate state-of-the-art modelling approaches within a user-friendly interface and to make published data available, the majority of existing QSAR tools has been created for screening purposes and tend to lack the precision required to replace experimental studies. The iSafeRat® toolbox v1.1 was developed to provide high accuracy QSAR (HA-QSAR) predictions supported by appropriate documentation for their regulatory acceptance. This work provides the results derived from iSafeRat® Toolbox[2] using a dataset of 15 well defined substances for which high quality experimental data for physicochemical and ecotoxicological endpoints are readily available. The iSafeRat results are compared with those derived from the same dataset using two commonly used predictive tools – EPISUITE v4.11[3] and OECD QSAR Toolbox v3.0 [4]. This comparative study is supported by recognised statistical methods to quantify the results.

Materials and Methods

Selection of dataset and endpoints: A dataset of 15 well defined chemicals were randomly selected to represent diversity in terms of chemical classes (eg. Alkanes, Alkenes, Aliphatic and Olifenic Alcohols, PAH, Ketones, Aldehydes, Ethers and Esters).

The experimental data for log octanol-water partition coefficient (Log Kow), water solubility (both being physicochemical endpoints) and Acute toxicity to Fish – LC50 96h (ecotoxicological endpoint) were collected for this work and the studies validated using Klimisch scores[5]. None of the substances were part of the training set for any of the reviewed iSafeRat models and all fell within the applicability domain of all the models.

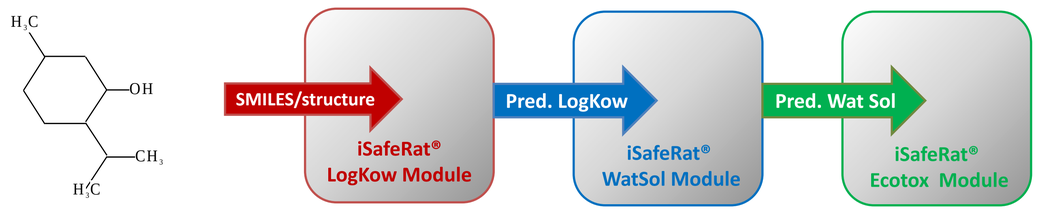

Figure 1: A schematic overview of iSafeRat® approach

iSafeRat® Holistic approach to predict physicochemical and ecotoxicological endpoints:

The iSafeRat® Toolbox uses a holisitic approach where a series of models interrelated by the laws of phase-equilibrium thermodynamics can be predicted in a cascade approach. The basic structural input from the user generates prediction for log Kow, which is then used as input (prediction for first endpoint) to generate prediction for solubility and then ecotoxicity (Fig 1) by means of a linear regression model for three trophic levels (Fish, Algae and invertebrates). By default, the Log Kow of the target chemical structure is predicted using the fragment-based approach.

EPISUITE and OECD QSAR Toolbox predictions: The Log Kow, water solubility and the acute toxicity to Fish were determined using KOWWIN v1.68, WSKOWWIN v1.42 and ECOSAR v1.11 modules, respectively integrated into the EPISUITE package. For OECD QSAR Toolbox, the Log Kow and Water Solubility were determined using a read across approach (category based on structural similarity of 75 percent thresholds with tanimoto distances). The acute toxicity to Fish on the other hand was also predicted via read across but defining a category based on acute aquatic toxicity classification by Verhaar *et al*.[6] and recent updates to this scheme. In all the cases, each predicted value was based on 5 nearest neighbours, the default setting in OECD toolbox.

Comparison of accuracy of the methods

For each method the results were classified as “accurate” [green]; “borderline” [orange] and “unacceptable” [red] as explained in the table footnote.

Results and Discussion

Table 1: Predictions derived for the three selected endpoints using iSafeRat® Toolbox and other commonly used predictive models.

NO.	NAME	Log Kow				Water Solubility (mg/L)				LC50 Fish - 96h (mg/L)			
		Measured	iSafeRat® v1.1	KOWWIN v1.68	QSAR Toolbox v3.0	Measured	iSafeRat® v1.1	WSKOWWIN v1.42	QSAR Toolbox v3.0	Measured	iSafeRat® v1.1	ECOSAR v1.11	QSAR Toolbox v3.0
1	n-pentane	3.39	3.45	2.80	2.88	38.50	34.71	49.76	32.30	4.26	1.85	11.41	10.20
2	1-dodecanol	5.13	5.27	4.77	4.95	1.90	2.49	6.90	2.63	1.01	0.35	0.50	2.50
3	menthol	3.3	3.16	3.38	3.26	456.00	337.36	434.50	415.00	18.90	15.83	7.38	2.43
4	sanderol	4.3	4.43	5.14	No RA	20.00	24.56	5.26	NN issue	1.10	1.91	1.37	1.37
5	nerolidol	4.7	4.62	5.68	4.50	14.10	16.10	1.53	18.00	1.80	1.43	1.21	2.32
6	terpineol	2.60	2.38	3.33	3.13	2540.00	3758.48	371.70	710.00	62.00	68.61	8.07	4.09
7	ebanol	4.2	4.25	4.93	NN issue	34.10	39.58	7.84	No RA	2.30	2.71	0.40	NN issue
8	2-methyl-2-butane	2.67	2.81	2.64	NN issue	193.00	225.29	206.10	NN issue	4.99	7.14	15.40	13.80
9	biphenyl	4.01	4.03	3.76	NN issue	7.50	8.17	27.35	NN issue	3.00	1.70	3.35	3.49
10	naphthalene	3.40	3.31	3.17	3.92	31.00	39.92	142.10	134.00	6.08	6.25	9.39	5.68
11	acetophenone	1.61	1.54	1.67	NN issue	6130.00	13381.78	4484.00	NN issue	162.00	161.52	193.85	5.74
12	2-methylcyclohexyl acetate	2.75	2.78	3.06	No RA	440.00	730.07	173.40	No RA	14.00	13.07	5.02	2.90
13	4-propylcyclohexanone	2.62	2.31	2.53	NN issue	1960.00	2157.42	1610.00	NN issue	43.00	44.64	38.80	17.30
14	diisopropylether	1.52	2.08	1.88	NN issue	3110.00	2839.55	5800.00	NN issue	91.70	50.15	106.61	27.30
15	2-ethylhexanal	3.07	2.98	2.71	No RA	450.00	362.87	455.60	No RA	5.50	4.24	3.12	36.80
		Avg. Err	*0.14*	*0.40*	*1.56*	*Avg. folds*	*1.31*	*3.16*	*2.14*	*Avg. folds*	*1.47*	*2.48*	*5.78*

Colour code: In case of LogKow, values with prediction error of 0-0.2 log units are highlighted in green, 0.2-0.5 log units in yellow and >0.5 log units in red. For water solubility and LC50 96h in Fish, values with prediction error of up to 2 folds are highlighted in green, 2-3 folds in yellow and >3 folds in red. ‘No RA’ (in grey) indicates that Read-Across was not feasible as no chemicals within the database satisfied the threshold criteria. ‘NN issue’ (in grey) indicates that Read Across was not feasible due to lack of nearest neighbours required.

The decision criteria for “accurate” results was based on repeatability criteria (i.e. the difference between two sets of results on the same substance which could be reasonably expected from the same laboratory, using the same method, performed two weeks in a row). From Table 1, iSafeRat predictions, 12/15 Log Kow values (80%) were associated with an absolute prediction error of <0.2 log units, only one solubility prediction and two LC50 values were beyond 2 folds error. Prediction accuracy for all three endpoints derived from EPISUITE and OECD QSAR Toolbox was lower. The average prediction error (Avg. Err) and average folds error (Avg. folds) were in the order iSafeRat®<KOWWIN<OECD Toolbox for log Kow and acute fish toxicity and iSafeRat®<OECD Toolbox <KOWWIN for solubility. Moreover, for OECD QSAR Toolbox, the read-across was not possible for 9/15 (60%) of the substances due to lack of structurally similar substances necessary to define a chemical category.

Conclusions

- This work demonstrates that the iSafeRat® predictions were associated with low prediction errors making them suitable for use in REACH.
- The average absolute error for Log Kow and average folds error for solubility and acute toxicity to fish were considerably lower with iSafeRat® as compared to the other models reviewed.

References

[1]ECHA – Practical guide 5: How to report (Q)SARs. Link:http://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf
[2]iSafeRat® – in Silico Algorithms For Environmental Risk And Toxicity version 1.1
[3]EPI Suite v4.11- U.S. Environmental Protection Agency. http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm
[4]OECD QSAR Toolbox v3.0 for grouping chemicals into Categories - http://www.qsartoolbox.org/.
[5]Klimisch H.J., Andreae M., Tillmann U. (1997) A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regul Toxicol Pharmacol. 25(1), 1-5.
[6]Verhaar H.J.M., Van Leeuven C., Hermens J.L.M.,Classifying Environmental Pollutants. 1: Structure-Activity Relationships for Prediction of Aquatic Toxicity, Chemosphere, Vol.25, No.4, pp.471-491, 1992."

Poster No. WE232 / Abstr. No. 20743

This poster is presented at SETAC Europe 24th annual meeting 2014, Basel, Switzerland.