Bigdata Analysis: Outcome of the 2nd AMES/QSAR International Challenge Project

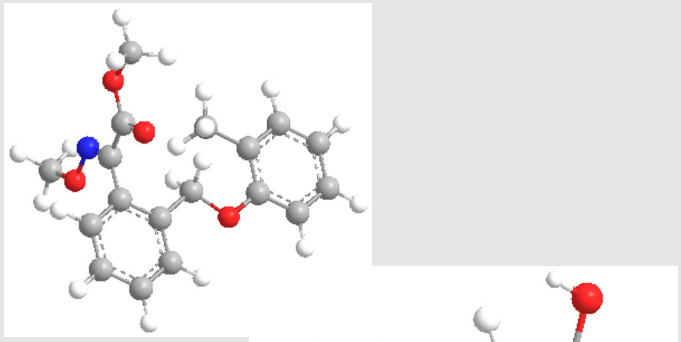
Ayako Furuhama, Ph.D.

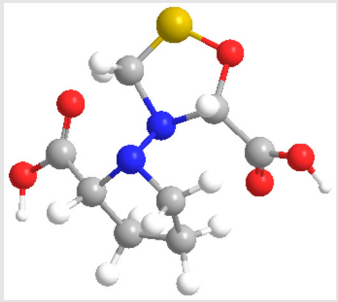
Division of Genetics and Mutagenesis (DGM)
Center for Biological Safety and Research (CBSR)
National Institute of Health Sciences (NIHS), Japan

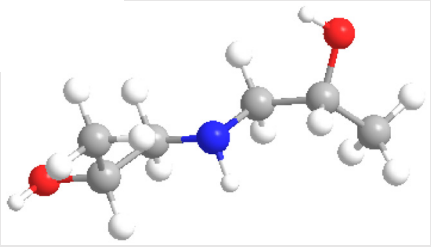
11th Annual Global Summit on Regulatory Science (GSRS21)
October 5 2021

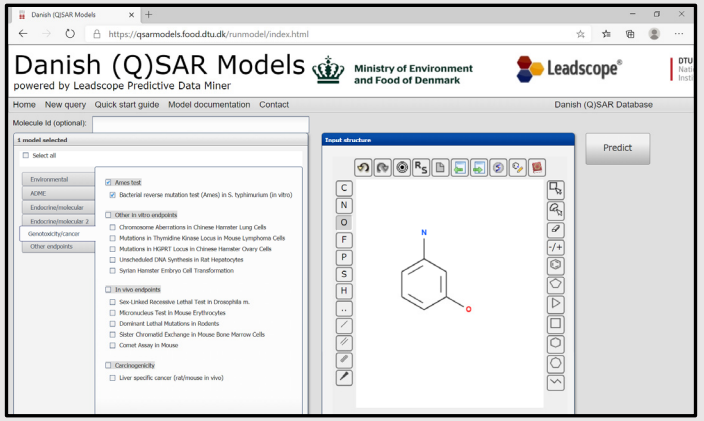
The opinions in this presentation are my own and do not necessarily reflect the views and policies of NIHS and Ministry of Health, Labour and Welfare of Japan (MHLW) or else.

Chemical screening and assessment







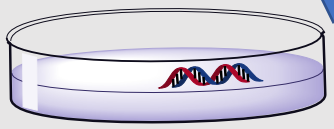


https://qsarmodels.food.dtu.dk/index.html

In silico

QSAR (quantitative structure-activity relationship) used instead of *in vitro* Ames testing

Example: ICH M7 guideline

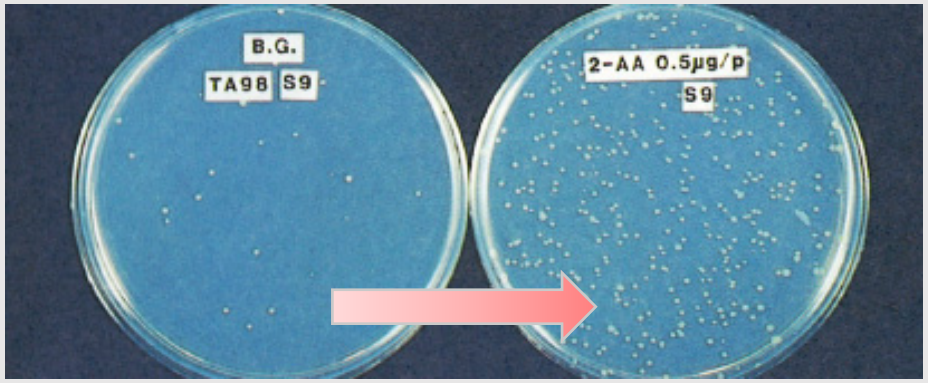


In vitro testing

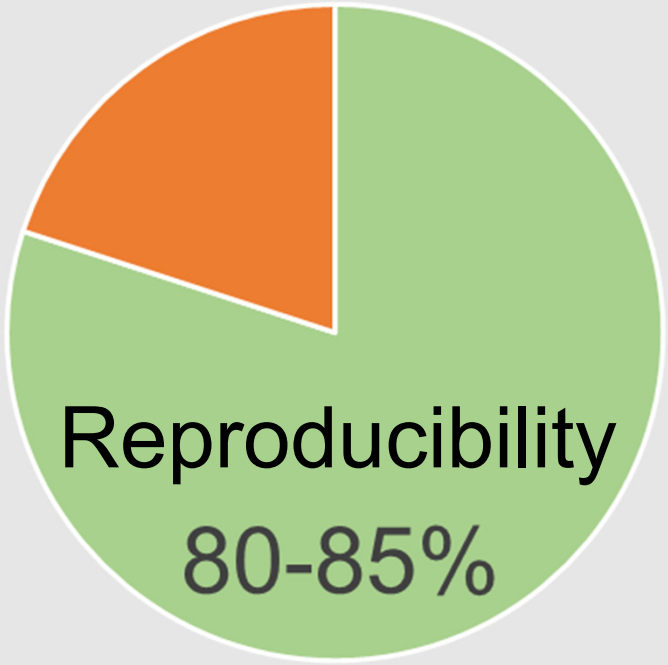
In vivo testing

Ames test

- *In vitro* genotoxicity testing
- High reproducibility
- Sensitive to detect carcinogens and non-carcinogens compared with other tests.



Positive



→*In silico* Ames QSAR is effective for regulation.

Regulatory perspective QSAR

- High sensitivity
- High negative predictivity
  - Minimum false-negative
- Wide coverage
  - Max chemical space

→ Need high quality data

Bigdata used in the 1st and 2nd Ames/QSAR International Challenge Projects

Proprietary dataset by DGM/NIHS

1. Ames database with >10,000 new chemicals.
2. The origin of the Ames test reports is ANEI-HOU, MHLW
3. The reports were originally undisclosed but the outcomes (positive or negative) were made available for validation, development and improvement of QSAR tools.

High quality Ames test data

1. ANEI-HOU test guideline, similar to OECD TG 471
2. Five strains with/without metabolic activation

MHLW: the Ministry of Health Labour and Welfare of Japan
ANEI-HOU: Industrial Safety and Health Act; To secure safety and health in the workplace, new chemicals in >100 kg/year require Ames test.
Honma et al., *Mutagenesis* 34, 2-16, 2019 modified.

The 1st project achievements

Mutagenesis Special Issue, **34** (2019)

Mutagenesis, 2019, 34, 3–16
doi:10.1093/mutage/gey031
Original Manuscript

OXFORD

Original Manuscript

Improvement of quantitative structure–activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project

Masamitsu Honma*, Airi Kitazawa, Alex Cayley[1], Richard V. Williams[1], Chris Barber[1], Thierry Hanser[1], Roustem Saiakhov[2], Suman Chakravarti[2], Glenn J. Myatt[3], Kevin P. Cross[3], Emilio Benfenati[4], Giuseppa Raitano[4], Ovanes Mekenyan[5], Petko Petkov[5], Cecilia Bossa[6], Romualdo Benigni[6,7], Chiara Laura Battistelli[6], Alessandro Giuliani[6], Olga Tcheremenskaia[6], Christine DeMeo[8], Ulf Norinder[9,10], Hiromi Koga[11], Ciloy Jose[11], Nina Jeliazkova[12], Nikolay Kochev[12,13], Vesselina Paskaleva[13], Chihae Yang[14], Pankaj R. Daga[15], Robert D. Clark[15] and James Rathman[14,16]

Division of Genetics and Mutagenesis, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kanagawa 210-9501, Japan, [1]Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, UK, [2]MultiCASE Inc., 23811 Chagrin Blvd Ste 305, Beachwood, OH 44122, USA, [3]Leadscope, Inc., 1393 Dublin Road, Columbus, OH 43215, USA, [4]Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via G. La Masa19 Milano, Italy, [5]Laboratory of Mathematical Chemistry, As. Zlatarov University, Bourgas, Bulgaria, [6]Istituto Superiore di Sanita', Viale Regina Elena, 299 00161 Rome, Italy, [7]Alpha-Pretox, Via G. Pascoli 1, 00184 Rome, Italy, [8]Prous Institute, Rambla de Catalunya, 135, 3-2, Barcelona 08008, Spain, [9]Swetox, Karolinska Institutet, Unit of Toxicology Sciences, Södertälje 15136, Sweden, [10]Department of Computer and Systems Sciences, Stockholm University, Box 7003, SE-164 07 Kista, Sweden, [11]Fujitsu Kyushu Systems Limited, 1-5-13 Higashihie, Hakata-ku, Fukuoka 812-0007, Japan, [12]IdeaConsult Ltd., 4 A. Kanchev str., Sofia 1000, Bulgaria, [13]Department of Analytical Chemistry and Computer Chemistry, University of Plovdiv, 24 Tsar Assen St., Plovdiv 4000, Bulgaria, [14]Molecular Networks GmbH and Altamira LLC, Neumeyerstrasse 28 90411 Nürnberg, Germany and 1455 Candlewood Drive, Columbus, OH 43235, USA, [15]Simulations Plus, Inc., 42505 10th Street West Lancaster, CA 93534, USA and [16]Chemical and Biomolecular Engineering, The Ohio State University, 151 W. Woodruff Ave. Columbus, OH 43210, USA

This paper received best paper award of *Mutagenesis* in 2019

The 2nd Ames/QSAR project

Started in 2020

https://www.nihs.go.jp/dgm/2nd_amesqsar.html

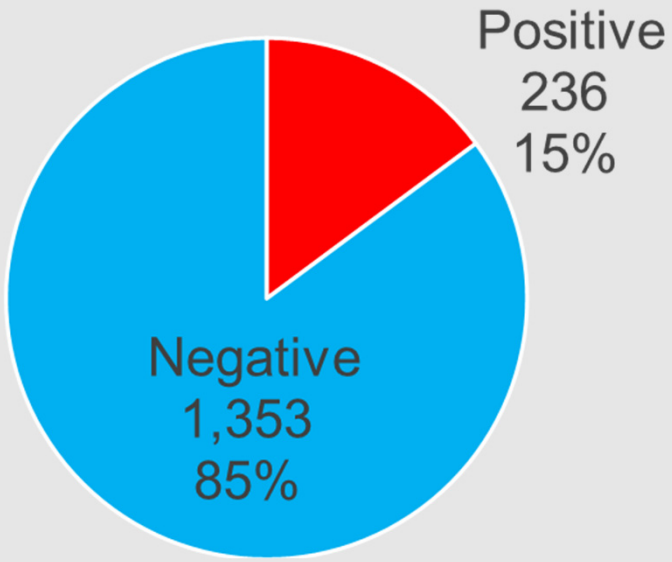
Chemical Hazards Control Division, Industrial Safety and Health Department, MHLW of Japan for providing the ANEI-HOU Ames dataset and allowing us to use the data in the projects.

Ames/QSAR project 5w1h

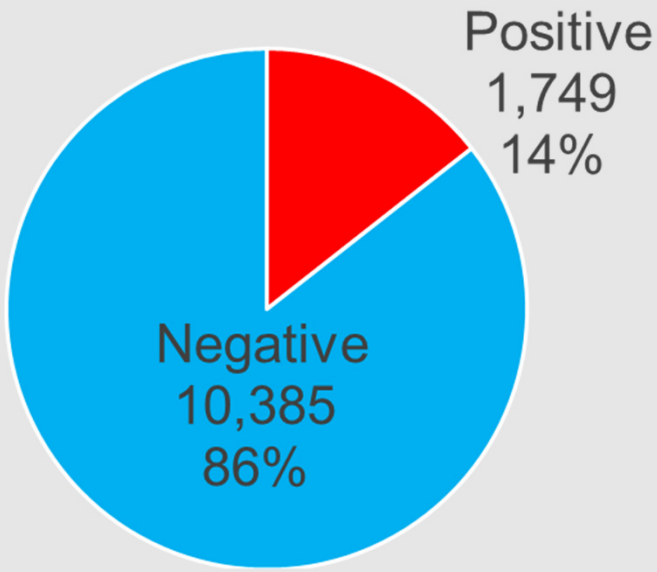
Where	1st project DGM/NIHS	2nd project DGM/NIHS
When	2014 - 2017	**2020**
Who participate	12 teams (7 countries) Mainly QSAR venders	**19 teams (11 countries)** **Academia/non-commercials**
What kind predictions	Three trials: ~4,000 chemicals/trial	**One trial:** **~1,600 chemicals**
How many training data provided by DGM	Trial I: 0 data Trial II: ~4,000 data Trial III: ~8,000 data	**>12,000 data used at the 1st project**
Why	**QSAR tool improvements**	
by using	Statistical, rule-based, its consensus models	**Machine learning models and AI based systems**

Ames results in the 2nd project

Trial chemicals



Training data provided by DGM/NIHS



Positive ~15%

Informative Ames data

Additional data provided to the participants of the 2nd project

CAS RN	Chemical Name	Structure	Result	Purity (%)	Solvent	Without metabolic activation (- S9)					With metabolic activation (+ S9)				
						TA100	TA1535	WP2 uvrA	TA98	TA1537	TA100	TA1535	WP2 uvrA	TA98	TA1537
[redacted]			++	>99	H2O	++	-	++	++	++	+	+	++	+	+
			++	99.5	DMSO	++	+	+	+	-	++	+	+	+	-

- Purity
- Solvent
- Results of each strain

++ : Strongly Positive + : Positive - : Negative

2nd Project participants

	Name	Country
1	Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences	China
2	Altox Ltd.	Brazil
3	MN-AM	Germany/USA
4	Leadscope, Inc.	USA
5	Istituto di Ricerche Farmacologiche Mario Negri - IRCCS	Italy
6	IdeaConsult Ltd.	Bulgaria
7	MultiCASE Incorporated	USA
8	Lhasa Limited	UK
9	Istituto Superiore di Sanità (ISS)	Italy
10	Gifu University	Japan
11	Massachusetts Institute of Technology	USA
12	Simulations Plus, Inc.	USA
13	Chemotargets	Spain
14	LMC - Bourgas University	Bulgaria
15	The University of Sydney	Australia
16	Meiji Pharmaceutical University	Japan
17	Liverpool John Moores University	UK
18	Evergreen AI, Inc.	Canada
19	Politecnico di Milano	Italy

The performance metrics of QSAR tools

1st project

	Phase I	Phase II	Phase III
Sensitivity (%)	56.7 (38.6–70.0)	58.0 (41.6–72.1)	57.1 (31.7–67.6)
Specificity (%)	77.7 (62.5–91.5)	84.2 (64.9–92.8)	79.9 (60.7–93.0)
Accuracy (%)	74.7 (63.6–83.9)	80.3 (65.8–87.7)	76.7 (68.0–87.3)

Honma et al., *Mutagenesis* 34, 2-16, 2019 modified.

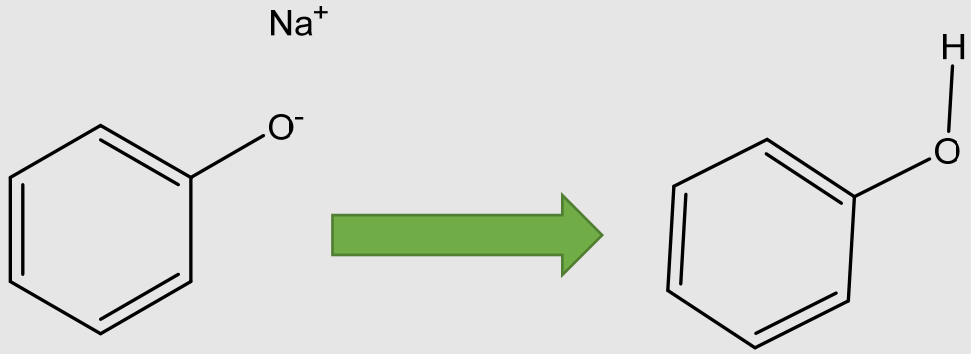
2nd project: under analysis

The models still show
-Low sensitivity
-High specificity

Sensitivity : the ability to detect mutagens
Specificity : the ability to detect non-mutagens
Accuracy: proportion of correct predictions

First overview of the 2nd project: High specificity and low sensitivity models

- Lack of data curations may enhance the data vias
  - Valance of positive/negative data (ideal 1/1)
  - Chemical structures

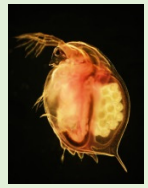


- >10,000 data: small for the Ames prediction of new chemicals
  - Experiences may cover a small chemical space

Ayako's previous study: model development for chronic ecotoxicity prediction

Daphnia toxicity + selected structural and physicochemical data brought good fish chronic toxicity perdition models

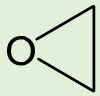
Biological data: acute daphnia toxicity

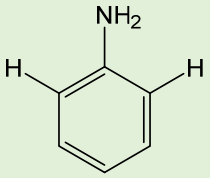


Chronic fish toxicity

Molecular substructures:

- Structural alter
- Functional group





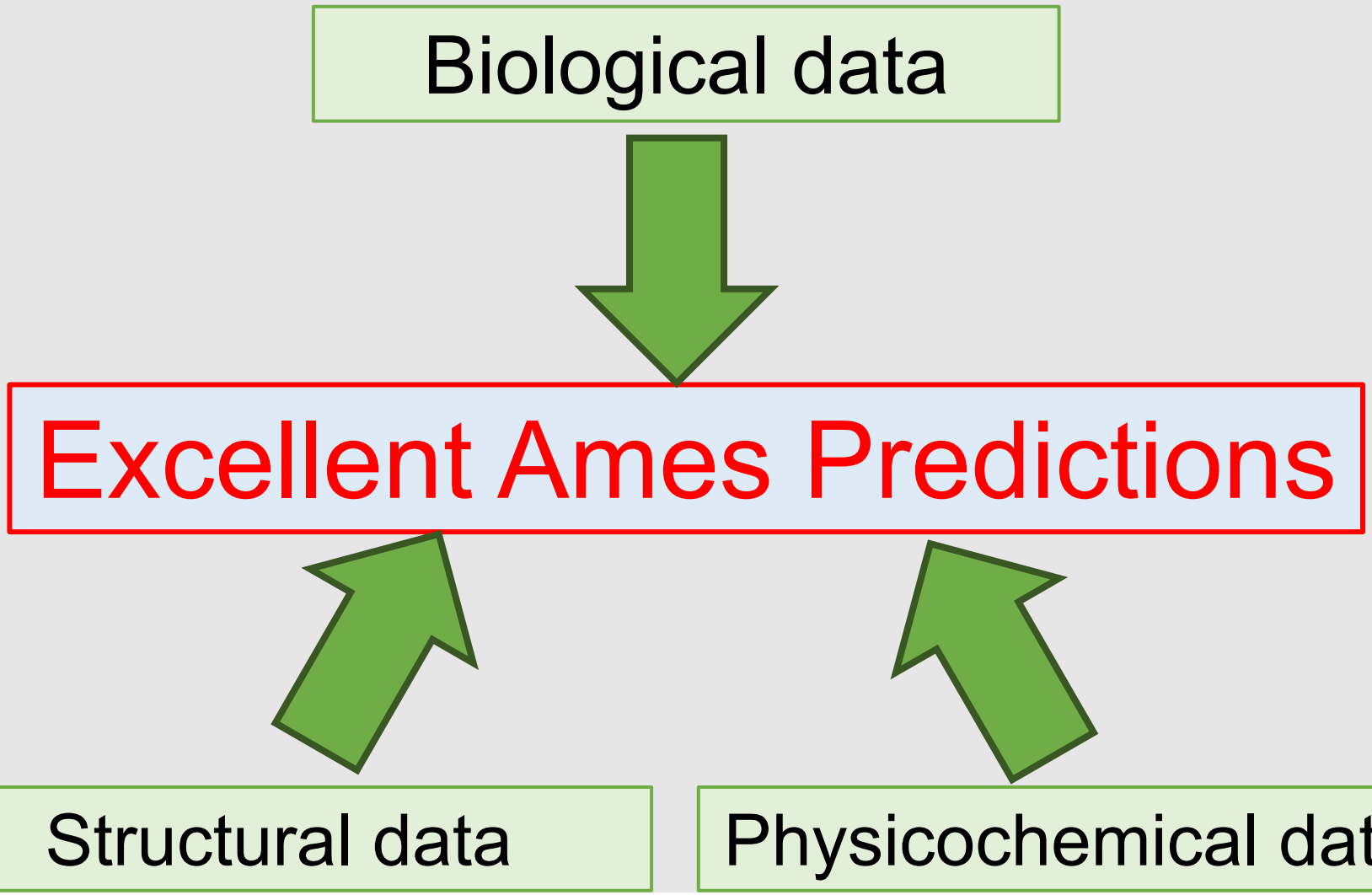
Physicochemical properties:

- Molecular weight
- Distribution coefficient
- Water solubility

Furuhama et al. (2018) *SAR QSAR Environ Res* ***29*** *725.*

My personal perspective

Selected various data required!



Summary

- The high reproductivity and amount of Ames data allow us the use of QSAR.

- DGM/NIHS started the 2nd Ames/QSAR project providing with informative >10,000 Ames data.

→ First overview: Models show high specificity and low sensitivity. >10,000 Ames data are not enough.

- For improving predictivity, experiences and various type of selected data might be important as well as the number of Ames data.

→ We need to investigate how to improve the reliable QSAR models and *in silico* predictions including machine learning and AI.

Acknowledgements and fundings

- The participants in the Ames/QSAR International Challenge

- Dr. Masamistu Honma (Deputy Director General, NIHS)
- Dr. Kei-ichi Sugiyama (DGM Director, NIHS)
- Ms. Airi Kitazawa and Dr. Toshio Kasamatsu (DGM, NIHS)

- The Chemical Hazards Control Division, Industrial Safety and Health Department, MHLW for providing the ANEI-HOU Ames dataset and allowing us to use the data in the projects.

- The Ministry of Health, Labour, and Welfare under Grants (H30-Chemistry-Destination-005, 21KD2005, and 21KA1001).