ORCID ID:0000-0002-5401-9706

An evaluation of selected in silico approaches for the assessment of skin sensitization potential – performance and practical utility considerations

Jeremy Fitzpatrick, Ph.D.
National Center for Computational Toxicology (NCCT), US EPA

Overview

- Previous Work
- Models Evaluated
- Test Data
- Models and Performance
- Combining Their Predictions

Overview

- **Previous Work**
- Models Evaluated
- Test Data
- Models and Performance
- Combining Their Predictions

Previous Work with Predictive Models for Skin Sensitization

- Previous work compared the results of Derek Nexus, Toxtree, OECD Toolbox, Topkat, Case Ultra, Vega, TIMES
  - Success was hampered by limited available data
- Current work expanded available data

ORCID ID:0000-0002-5401-9706

Regulatory Toxicology and Pharmacology 67 (2013) 468–485

Contents lists available at ScienceDirect

ELSEVIER

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Computer models versus reality: How well do *in silico* models currently predict the sensitization potential of a substance

Wera Teubner [a,*], Anette Mehling [b], Paul Xaver Schuster [c], Katharina Guth [c], Andrew Worth [d], Julien Burton [d], Bennard van Ravenzwaay [c], Robert Landsiedel [c]

[a] BASF Schweiz AG, Basel, Switzerland
[b] BASF Personal Care and Nutrition GmbH, Düsseldorf, Germany
[c] BASF SE, Ludwigshafen, Germany
[d] European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Ispra, Italy

ARTICLE INFO

Article history:
Received 18 April 2013
Available online 30 September 2013

Keywords:
Case Ultra
GHS
OECD toolbox
Adverse outcome pathway
(Q)SAR
Skin sensitization
TIMES-SS
TOPKAT
Toxtree
Vega

ABSTRACT

National legislations for the assessment of the skin sensitization potential of chemicals are increasingly based on the globally harmonized system (GHS). In this study, experimental data on 55 non-sensitizing and 45 sensitizing chemicals were evaluated according to GHS criteria and used to test the performance of computer (*in silico*) models for the prediction of skin sensitization. Statistic models (Vega, Case Ultra, TOPKAT), mechanistic models (Toxtree, OECD (Q)SAR toolbox, DEREK) or a hybrid model (TIMES-SS) were evaluated. Between three and nine of the substances evaluated were found in the individual training sets of various models. Mechanism based models performed better than statistical models and gave better predictivities depending on the stringency of the domain definition. Best performance was achieved by TIMES-SS, with a perfect prediction, whereby only 16% of the substances were within its reliability domain. Some models offer modules for potency; however predictions did not correlate well with the GHS sensitization subcategory derived from the experimental data. In conclusion, although mechanistic models can be used to a certain degree under well-defined conditions, at the present, the *in silico* models are not sufficiently accurate for broad application to predict skin sensitization potentials.

Overview

- Previous Work
- **Models Evaluated**
- Test Data
- Models and Performance
- Combining Their Predictions

Models to Evaluate

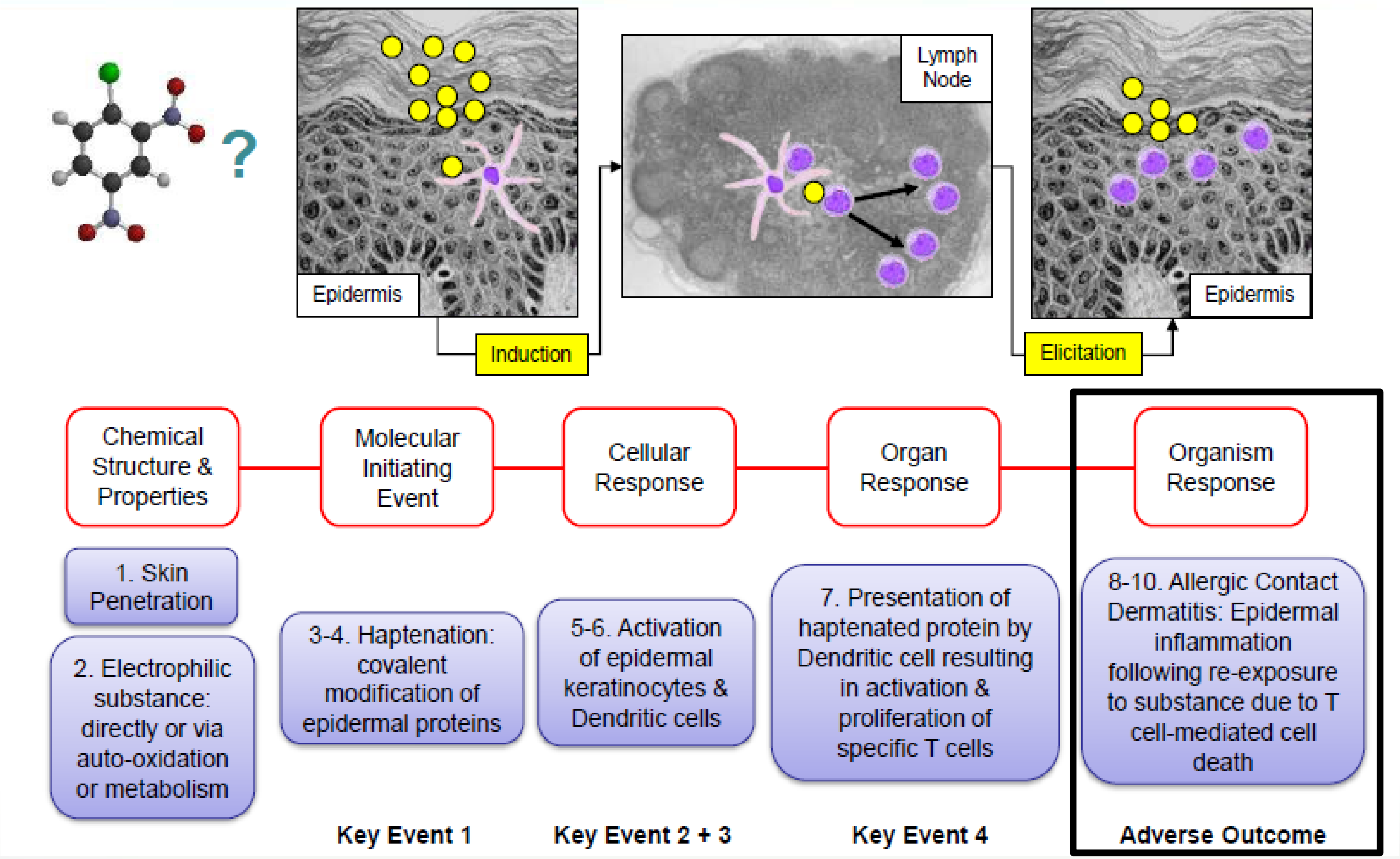
Model the Adverse Outcome

- TIMES (skin sensitization with autoxidation v. 20.24)
- VEGA (skin sensitization model CAESAR v. 2.1.3)
- MCASE A33 (skin sensitization Danish EPA DB in OECD Toolbox)

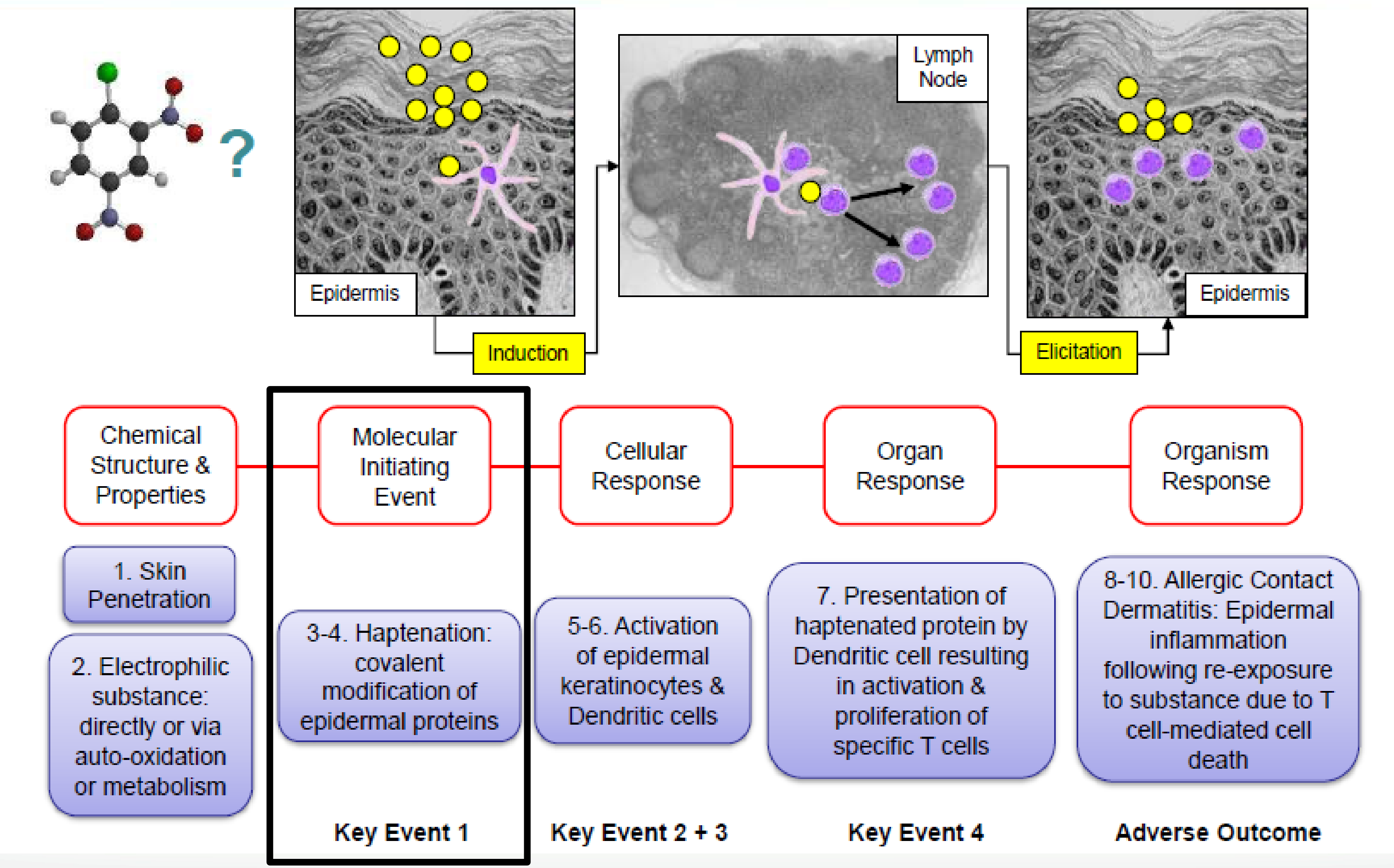
Protein Binding Domains (Prediction of the MIE)

- Toxtree (skin sensitization reactivity domains)
- OASIS (protein binding alerts for skin sensitization v1.3 in OECD Toolbox)

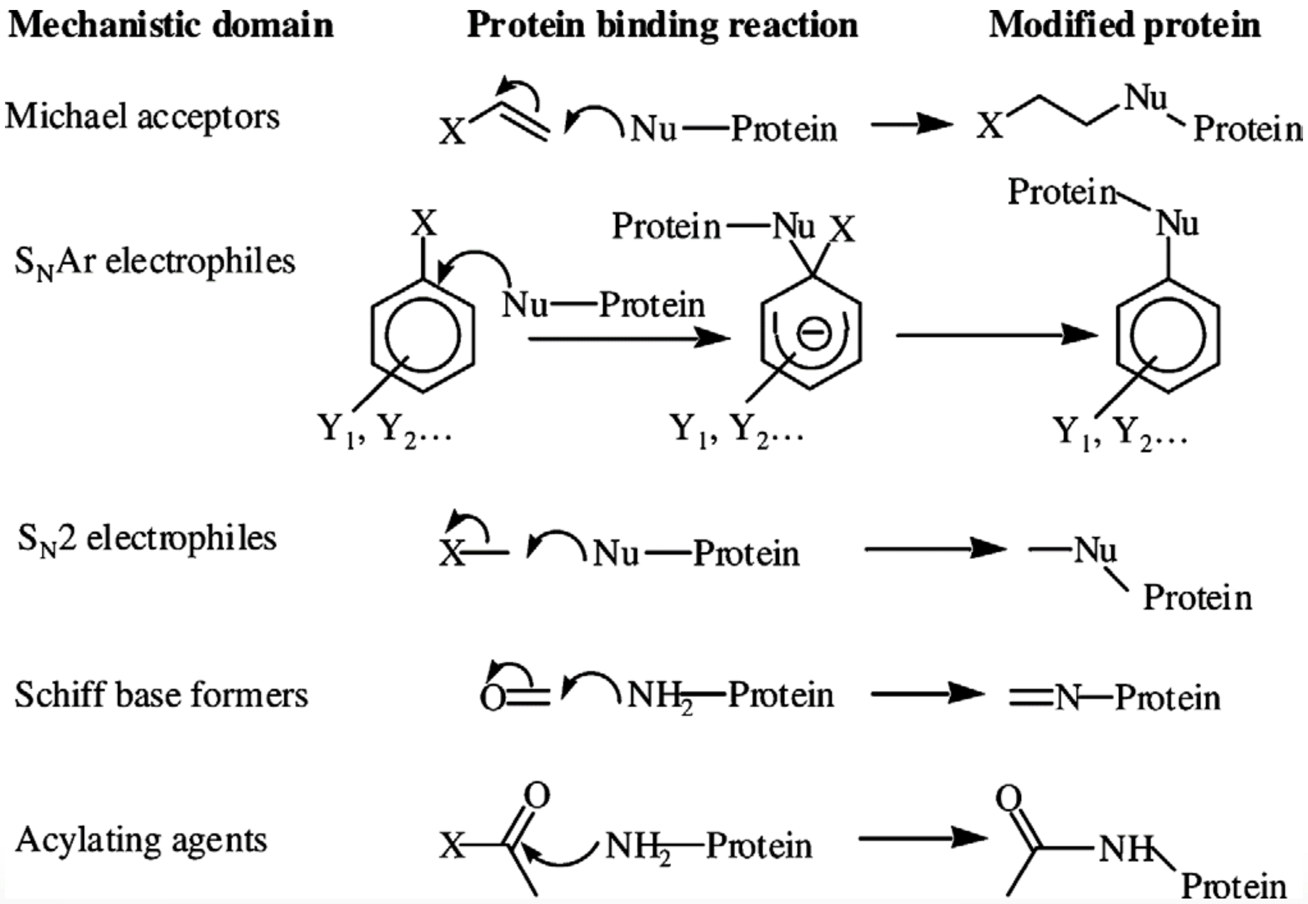
AOP for Skin Sensitization (OECD, 2012)



AOP for Skin Sensitization (OECD, 2012)



Haptenation: Mechanisms of Reaction Domains



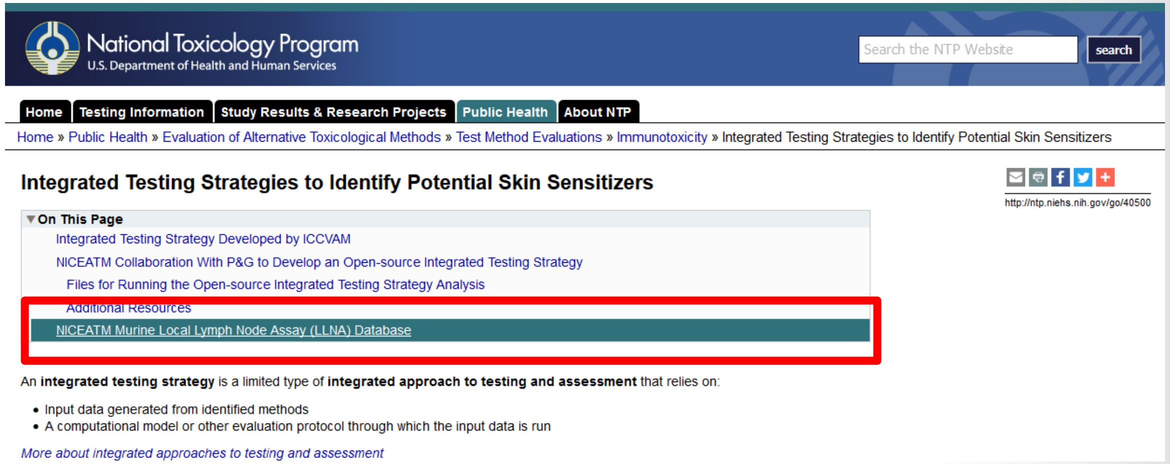
Aptula A., Patlewicz G., Roberts D., Skin Sensitization: Reaction Mechanistic Applicability Domains for Structure−Activity Relationships. Chem. Res. Toxicol. 2005; 18, 9: 1420-1426

Overview

- Previous Work
- Models Evaluated
- **Test Data**
- Models and Performance
- Combining Their Predictions

Test Data Source

- NICEATM LLNA database
- 515 compounds with structures and LLNA results, including EC3 values
- 186 non-sensitizers, 329 sensitizers



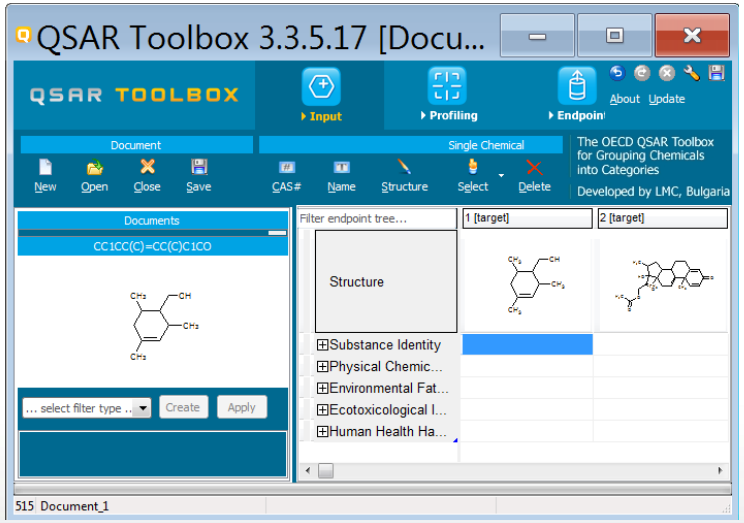
http://ntp.niehs.nih.gov/iccvam/methods/immunotox/niceatm-llnadatabase-23dec2013.xls
http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/immunotoxicity/nonanimal/index.html

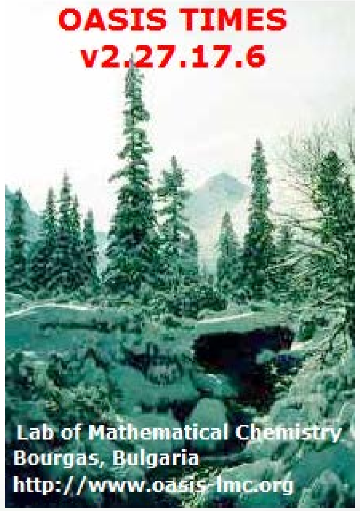
Overview

- Previous Work
- Models Evaluated
- Test Data
- **Models and Performance**
- Combining Their Predictions

Endpoint Models

- Turn the excel spreadsheet into a SMILES file
- Since the different models generate different prediction outcomes, we retrieve the binary outcomes i.e. sensitizer or non-sensitizer





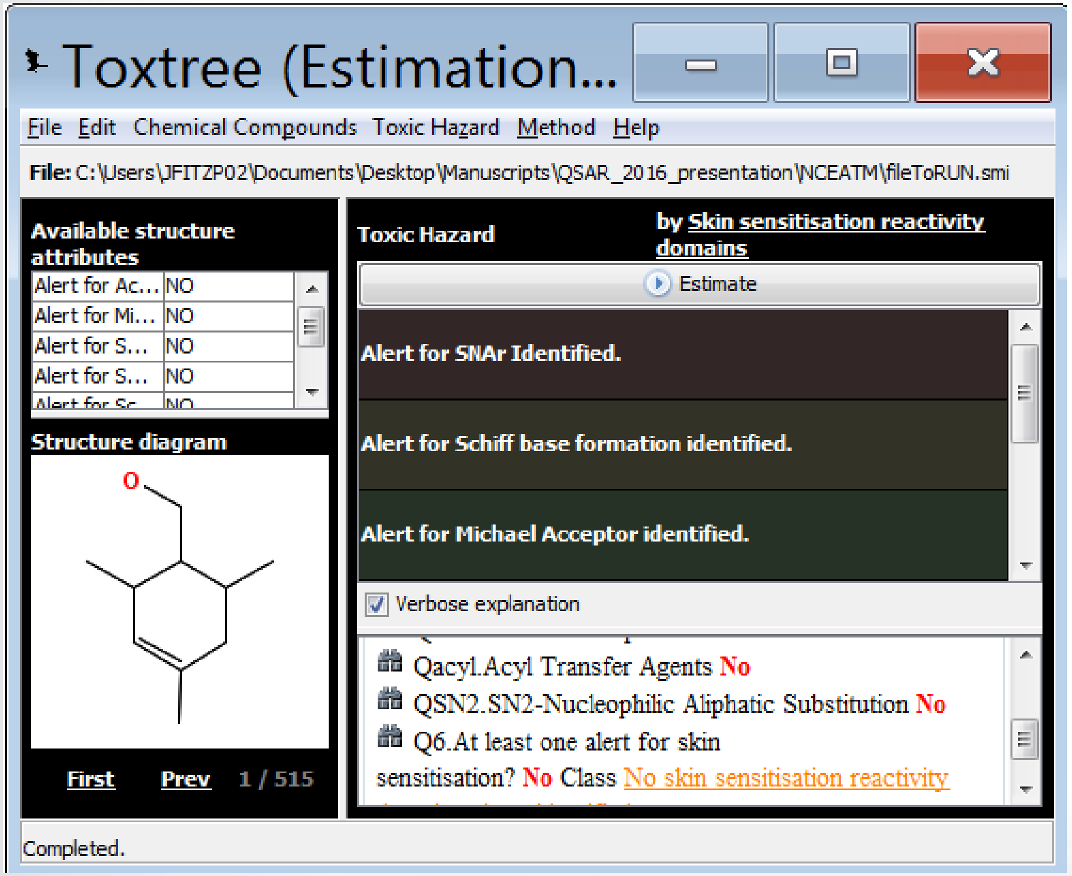
Predictivity based on Endpoint Models

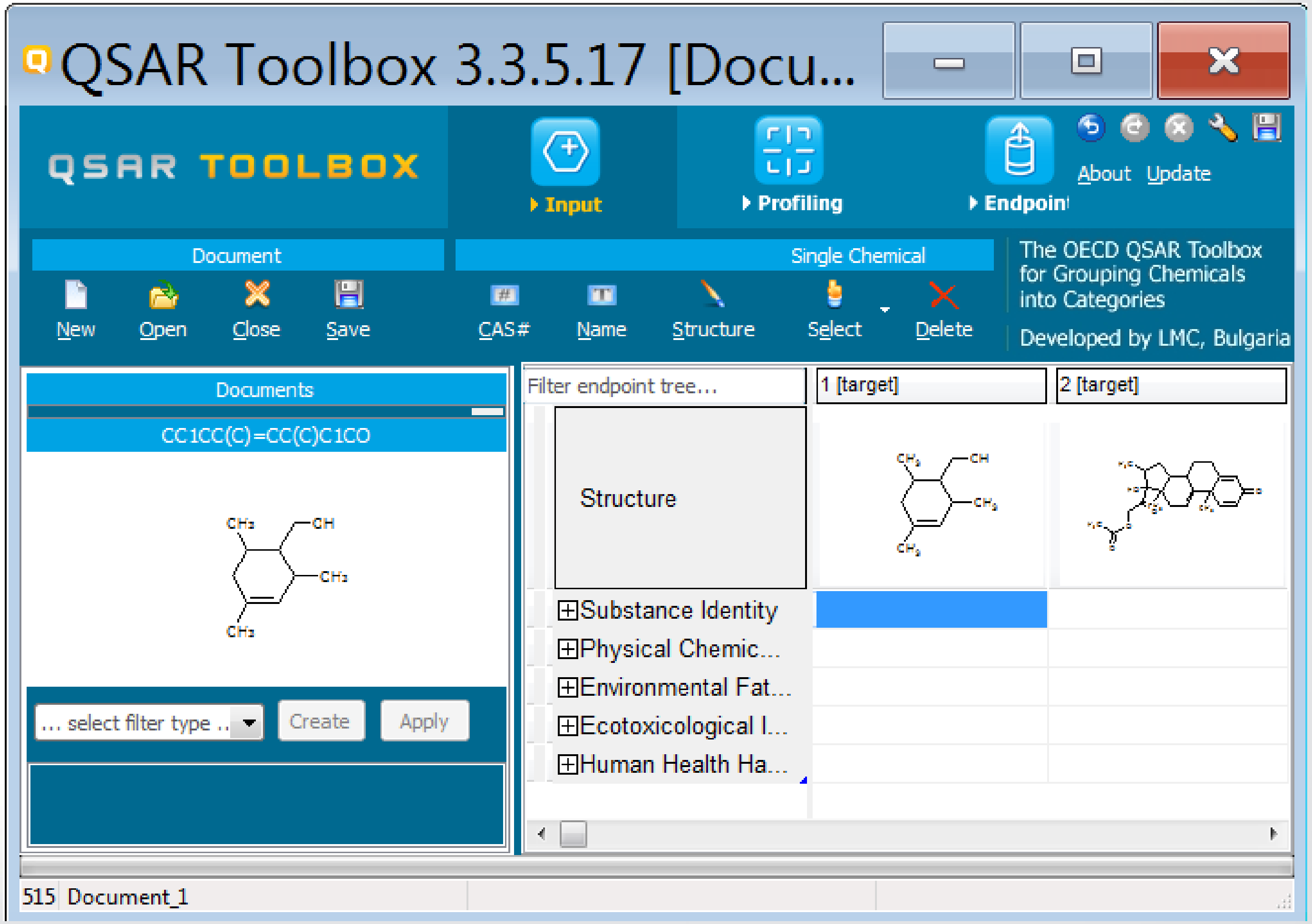
	Positive Predictivity (# of compounds tested)	Negative Predictivity (# of compounds tested)	Overall Predictivity (# of compounds tested)
Overall VEGA	84% (201)	36% (151)	64% (352)
Overall MCASE A33	61% (156)	61% (70)	61% (226)
Overall TIMES	76% (148)	45% (106)	63% (254)
Overlapping TIMES	69% (122)	44% (101)	57% (221)
Overlapping VEGA	80% (122)	40% (101)	62% (221)

Prediction results are given for compounds not in the underlying training set of the model.

Assigning Reaction Domains

- Reaction domains were assigned using Toxtree and OASIS (within the OECD Toolbox)





Reaction Domain Assignments

Tool	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Alert
OASIS v1.3	51	55	41	75	18	278
Toxtree	78	123	81	87	21	174
Matching	40	49	35	58	18	156
Disagree	49	80	52	46	3	140

Reaction Domain Assignments

Tool	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Alert
OASIS v1.3	51	55	41	75	18	278
Toxtree	78	123	81	87	21	174
Matching	40	49	35	58	18	156
Disagree	49	80	52	46	3	140

Overall Results

219 compounds showed some alert in both tools

156 compounds showed no alert in both tools

140 compounds had conflicting results

Predictivity based on Reaction Domain

	Positive Predictivity (# of compounds tested)	Negative Predictivity (# of compounds tested)	Overall Predictivity (# of compounds tested)
Toxtree	77% (328)	54% (186)	69% (514)
OASIS	58% (328)	74% (186)	64% (514)

Predictivity for Reaction Domains and Endpoint Models

	Positive Predictivity (# of compounds tested)	Negative Predictivity (# of compounds tested)	Overall Predictivity (# of compounds tested)
Overall VEGA	84% (201)	36% (151)	64% (352)
Overall MCASE A33	61% (156)	61% (70)	61% (226)
Overall TIMES	76% (148)	45% (106)	63% (254)
Toxtree	77% (328)	54% (186)	69% (514)
OASIS	58% (328)	74% (186)	64% (514)

Overview

- Previous Work
- Models Evaluated
- Test Data
- Models and Performance
- **Combining Their Predictions**

Combining the Predictions

- Combined VEGA, ToxTree, and OASIS results into a consensus prediction model
  - To exploit broad chemical coverage
  - All programs are freely available

	Positive Predictivity (# of compounds tested)	Negative Predictivity (# of compounds tested)	Overall Predictivity (# of compounds tested)
Consensus Prediction	69% (200)	64% (151)	67% (351)

Conclusions

- All models with the exception of MCASE A33 are more likely to generate false positive over false negatives
- Combining the results does not improve the prediction performance significantly for this dataset evaluated in this study

Acknowledgements

- Grace Patlewicz (US EPA)
- Chris Grulke (US EPA)
- Nicole Kleinstreuer (NICEATM)

Thank you for your attention

Questions?

Extra backup slides

Overview

- Previous Work
- Models to Evaluate
- Test Dataset
- How They Perform?
- Combining Their Predictions
- **Local Performance**
- Conclusions/Acknowledgments

Predictions Grouped by Toxtree Assignments

	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Positive Predictivity						
TIMES	43% (30)	57% (37)	64% (22)	54% (28)	56% (16)	60% (40)
VEGA	50% (20)	65% (26)	72% (18)	73% (26)	60% (10)	55% (76)
Negative Predictivity						
TIMES	71% (14)	0% (1)	43% (7)	29% (7)	0	54% (56)
VEGA	71% (24)	58% (12)	55% (11)	89% (9)	50% (6)	60% (20)
Overall Predictivity						
TIMES	52% (44)	55% (38)	59% (29)	49% (35)	56% (16)	56% (96)
VEGA	61% (44)	63% (38)	66% (29)	77% (35)	56% (16)	56% (96)

All overlapping compounds

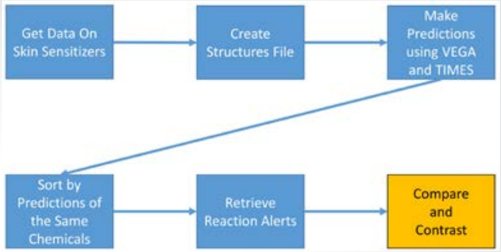
Which preforms best overall?

	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Predictivity OASIS Domains						
TIMES	43% (21)	53% (19)	88% (8)	70% (23)	50% (14)	57% (140)
VEGA	52% (21)	63% (19)	75% (8)	70% (23)	57% (14)	62% (140)
Predictivity Toxtree Domains						
TIMES	52% (44)	55% (38)	59% (29)	49% (35)	56% (16)	56% (96)
VEGA	61% (44)	63% (38)	66% (29)	77% (35)	56% (16)	56% (96)

VEGA performs better for compounds with a Acylation of Michael Addition Domain

Predictions Grouped by Toxtree Assignments

Compare And Contrast

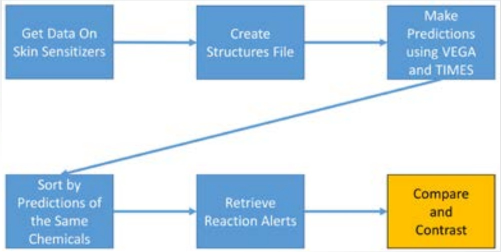


	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
<u>Positive Predictivity</u>						
TIMES	48% (33)	69% (52)	70% (30)	58% (31)	56% (16)	62% (42)
VEGA	55% (29)	81% (62)	77% (31)	78% (37)	67% (12)	47% (116)
<u>Negative Predictivity</u>						
TIMES	71% (14)	0% (1)	50% (8)	38% (8)	0	55% (58)
VEGA	63% (27)	47% (15)	47% (15)	91% (11)	50% (6)	75% (32)
<u>Overall Predictivity</u>						
TIMES	55% (47)	68% (53)	66% (38)	54% (39)	56% (16)	58% (100)
VEGA	59% (56)	74% (77)	67% (46)	81% (48)	61% (18)	53% (148)

All compounds not in a programs training set

Best positive predictivity?

Compare
And
Contrast

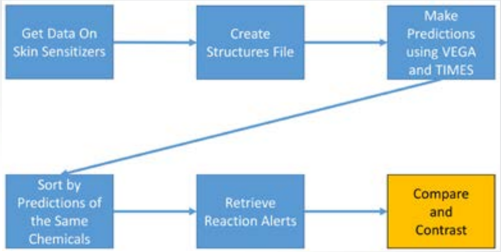


	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Predictivity OASIS Domains						
TIMES	43% (30)	57% (37)	64% (22)	54% (28)	56% (16)	60% (40)
VEGA	50% (20)	65% (26)	72% (18)	73% (26)	60% (10)	55% (76)
Predictivity Toxtree Domains						
TIMES	43% (21)	53% (19)	88% (8)	70% (23)	50% (14)	62% (58)
VEGA	45% (11)	62% (13)	100% (5)	80% (15)	56% (9)	60% (105)

TIMES performs best for compounds with no domain

Best negative predictivity?

Compare And Contrast



	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Predictivity OASIS Domains						
TIMES	0	0	0	0	0	54% (82)
VEGA	60% (10)	67% (6)	33% (3)	50% (8)	60% (5)	69% (35)
Predictivity Toxtree Domains						
TIMES	71% (14)	0% (1)	43% (7)	29% (7)	0	54% (56)
VEGA	71% (24)	58% (12)	55% (11)	89% (9)	50% (6)	60% (20)

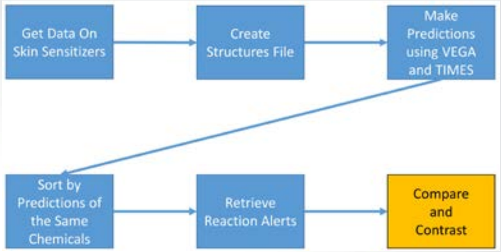
TIMES appears not to make negative predictions for most compounds with a reaction domain

Future Directions

- A more in depth analysis using Chemotypes
- Get more data from eChemportal
- Possibly evaluate other programs

Predictions Grouped by OASIS Assignments

Compare And Contrast

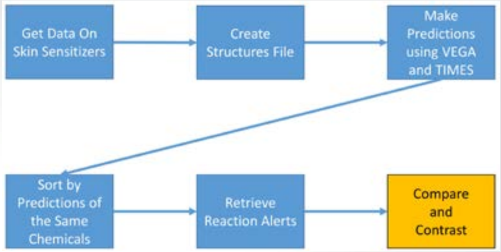


	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Positive Predictivity						
TIMES	43% (21)	53% (19)	88% (8)	70% (23)	50% (14)	62% (58)
VEGA	45% (11)	62% (13)	100% (5)	80% (15)	56% (9)	60% (105)
Negative Predictivity						
TIMES	0	0	0	0	0	54% (82)
VEGA	60% (10)	67% (6)	33% (3)	50% (8)	60% (5)	69% (35)
Overall Predictivity						
TIMES	43% (21)	53% (19)	88% (8)	70% (23)	50% (14)	57% (140)
VEGA	52% (21)	63% (19)	75% (8)	70% (23)	57% (14)	62% (140)

221 in neither programs training set

Predictions Grouped by OASIS Assignments

Compare And Contrast

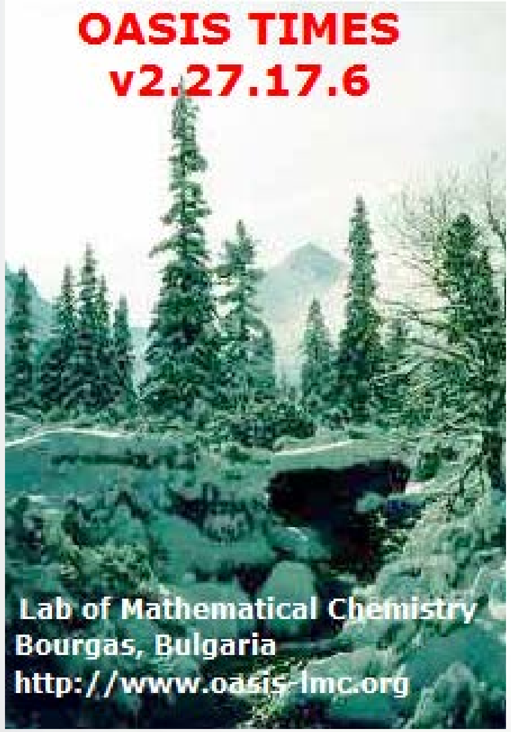


	Acylation	Michael Addition	Schiff Base	SN2	SNAr	No Domain
Positive Predictivity						
TIMES	48% (23)	70% (30)	87% (15)	73% (26)	50% (14)	66% (65)
VEGA	67% (18)	73% (30)	100% (12)	79% (28)	64% (11)	57% (165)
Negative Predictivity						
TIMES	0	0	0	0	0	56% (86)
VEGA	43% (14)	57% (7)	20% (5)	50% (10)	60% (5)	76% (49)
Overall Predictivity						
TIMES	48% (23)	70% (30)	87% (15)	73% (26)	50% (14)	60% (151)
VEGA	56% (32)	70% (37)	76% (17)	71% (38)	63% (16)	61% (214)

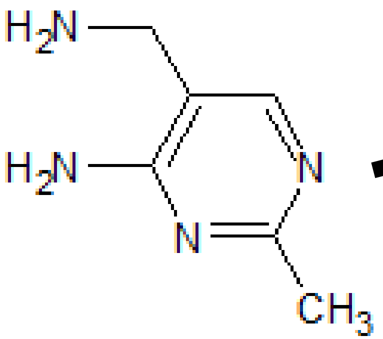
All compounds not in a programs training set

What are we trying to do?

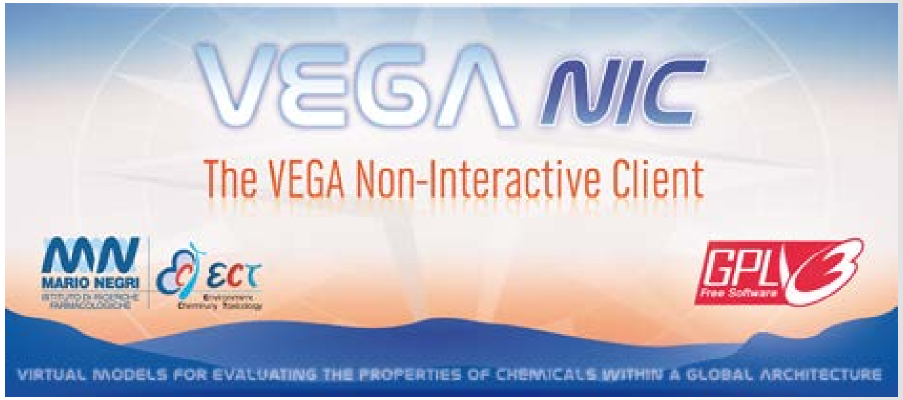
- Determine which program is most likely to predict the skin sensitization potential of a compound correctly



?

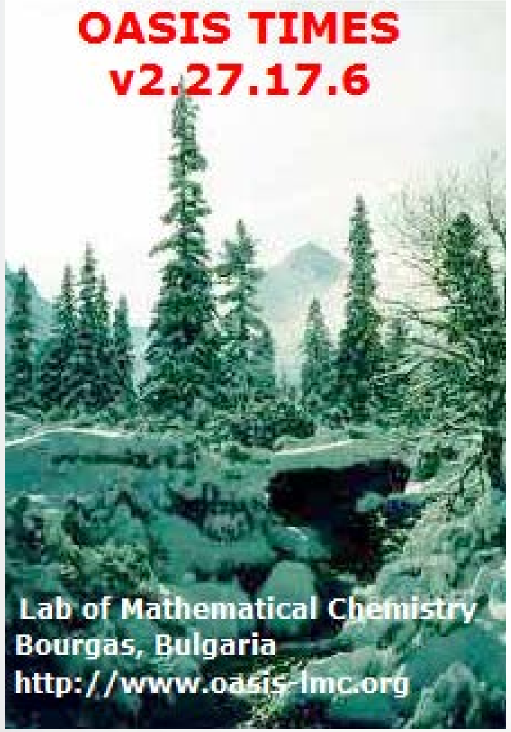


?

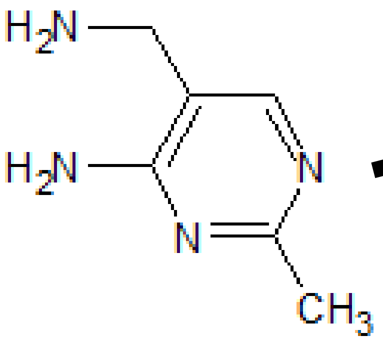


What are we trying to do?

- Determine which program is most likely to predict the skin sensitization potential of a compound correctly, **based on reaction domains from Toxtree and the OECD QSAR Toolbox.**



?



Schiff base

?