Sensitizer potency prediction:
Parameters from in vitro tests related to potency and their combination in ITS

Andreas Natsch, Givaudan Schweiz AG

Givaudan

engage your senses

Agenda

- Summary of the study performed by Givaudan:
  - Natsch, A., et al., Predicting Skin Sensitizer Potency Based on In Vitro Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment. Toxicol Sci, 2015.
- Summary of the study performed by Joanna Jaworska, P&G:
  - Jaworska, J.S., et al., Bayesian Integrated Testing Strategy (ITS) for skin sensitization potency assessment – a decision support system for quantitative weight of evidence and adaptive testing strategy. Archives of Toxicology, 2015
- Some learnings from the different study on importance and interrelationships of parameters measured in *in vitro* assays
  - My Personal conclusions !

Parameters

- Peptide reactivity (Key event 1): DPRA
  - Cys- and Lys-depletion
- Peptide reactivity (Key event 1): LC-MS assay
  - LC-MS evaluation of direct peptide modification (MW of adduct to interpret possible reaction mechanism)
  - **Peptide depletion after 24 h**
  - Dose-response of peptide depletion at earlier time-points
  - **Kinetic rate constant** derived from the multiple depletion values
- KeratinoSens™ (Key event 2, Keratinocyte activation):
  - Positive/negative rating according prediction model
  - $EC1.5_{KS}$ /$EC2_{KS}$ / $EC3_{KS}$ concentration for 1.5/2/3-fold **luciferase gene induction**
  - $IC50_{KS}$ concentration for 50% **reduction in viability**
- hClat (Key event 3, dendritic cell activation):
  - Positive/negative rating according prediction model
  - EC150 /EC200: concentration for 1.5/2-fold **CD86/CD54 induction**
  - CV75 concentration for 25% **reduction in viability**
- Physicochemical parameters:
  - cLogP, **Vapor pressure**

Correlation of individual parameters

- **LLNA EC3 best available parameter for *in vivo* potency**
  - linearized by Log transformation = pEC3
- **Quantitative *in vitro* data partly correlate to LLNA potency**
  - Data on 244 chemicals
  - dose response in KeratinoSens™
  - rate constant in peptide reactivity
    - All data can be linearized by Log transformation

- Best single parameter for global correlation is rate constant from peptide reactivity (better than fixed depletion value due to higher dynamic range)
- Both luciferase induction and cytotoxicity from KeratinoSens correlate to potency

Parameter	**R^2 adjusted (%)**	***p* value**
Peptide reactivity kinetic: K_{max}	51.7	< 0.0005
Peptide reactivity: $K_{24\ h\ depletion}$	43.6	< 0.0005
Luciferase $EC1.5_{KS}$	42.5	< 0.0005
Luciferase $EC2_{KS}$	44.8	< 0.0005
Cytotoxicity $IC50_{KS}$	33.5	< 0.0005

Single parameters – alternative dataset

- 191 chemicals with hClat, KeratinoSens, and reactivity data

Parameter	R^2 adjusted (%)	*p* value
K_{max}	43.2	< 0.0005
$K_{24\ h\ depletion}$	35.6	< 0.0005
Adduct TIMES (in silico reactivity)	27.0	< 0.0005
$K_{24\ h\ CYS\ DPRA}$*	22.9	< 0.0005
$K_{24\ h\ CYS\ DPRA}$*	0.1	n.s.
$EC1.5_{KS}$	31.9	< 0.0005
$EC3_{KS}$	40.6	< 0.0005
$IC50_{KS}$	34.1	< 0.0005
$EC150_{\ hClat}$	24.9	< 0.0005
$EC200_{\ hClat}$	27.9	< 0.0005
$CV75_{\ hClat}$	38.0	< 0.0005
$MIT_{\ hClat}$	46.1	< 0.0005

* Smaller dataset

Correlation to multiple parameters – multiple regression

- **Multiple regression uses most predictive combination of linear parameters**
  - Treats all chemicals equal
  - Fixed coefficients over whole potency range
- **Global model**:
  - Reactivity has strongest weight
  - Followed by luciferase from KeratinoSens
  - Significant impact also for cytotoxicity and vapor pressure
    - Without hClat, with hClat see below
    - Regression equation can the be used to make predictions

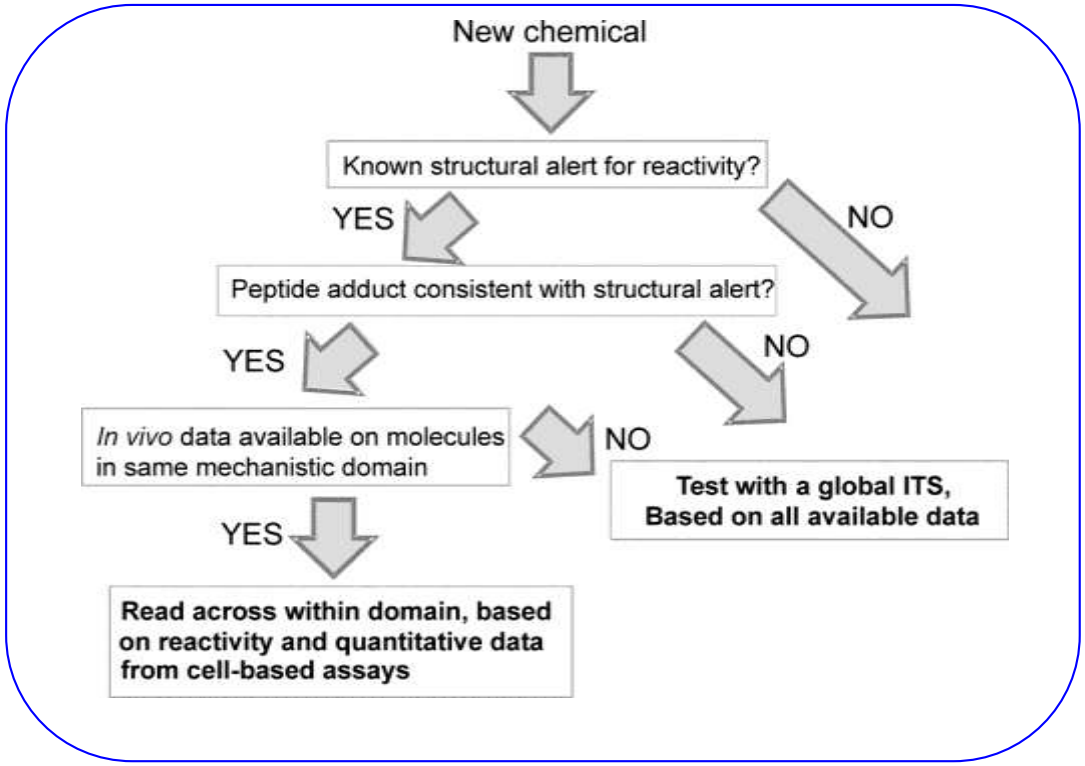
Equation 1: A global regression analysis on prediction of $EC3_{LLNA}$ by *in vitro* and *in chemico* data

$$pEC3_{LLNA} = 0.04 + 0.38 \times \text{Log } K_{norm} + 0.25 \times \text{Log } EC1.5_{norm} + 0.25 \times \text{Log } IC50_{norm} - 0.19 \times \text{Log } VP_{norm}$$

Constant	T = 0.51, p = 0.612	**Log K_{norm}**	T = 9.55, ***p* < 0.0005**
Log $EC1.5_{norm}$	T = 4.06, ***p* < 0.0005**	Log $IC50_{norm}$	T = 3.05, p = 0.003
Log VP_{norm}	T = - 3.39, p = 0.001	**R^2 (adj) = 62.3%**	

Global vs. mechanistic domain models

- The concept of grouping of chemicals is widely accepted (e.g. used in OECD toolbox)
- Chemicals should be predicted in domains if:
  - They can be grouped in domains with related chemicals
  - Related chemicals have been tested *in vitro* and *in vivo*



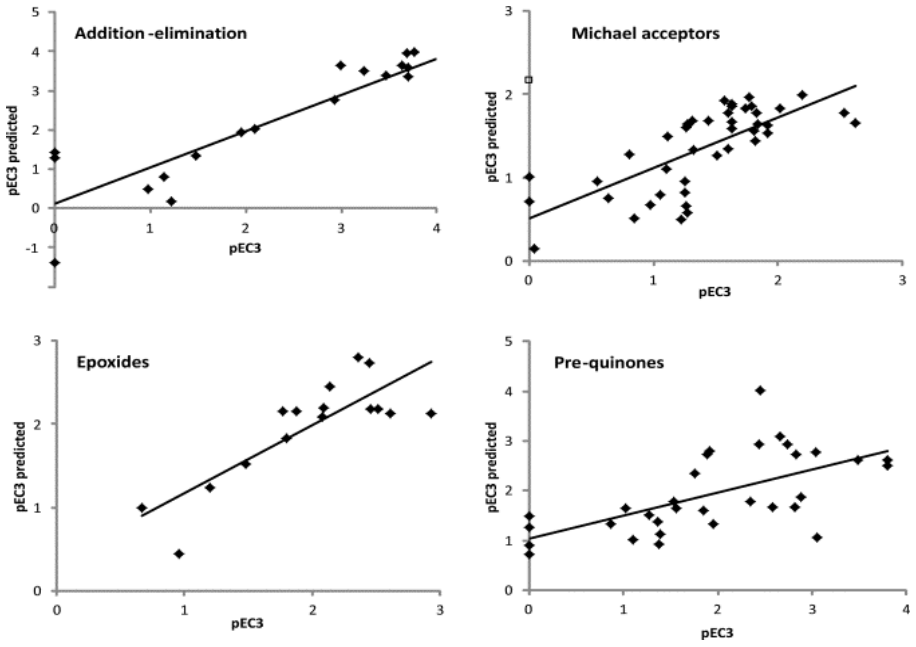
Local models – predictive capacity

Domain models – leave one-out analysis.

Each chemical is predicted with the remaining chemicals in dataset as training set, avoids bias due to too small groups

Domain models allow fold misprediction of 2 – 3 fold for many chemicals

This may be more useful as point of departure in risk assessment as compared to 10-fold potency classes

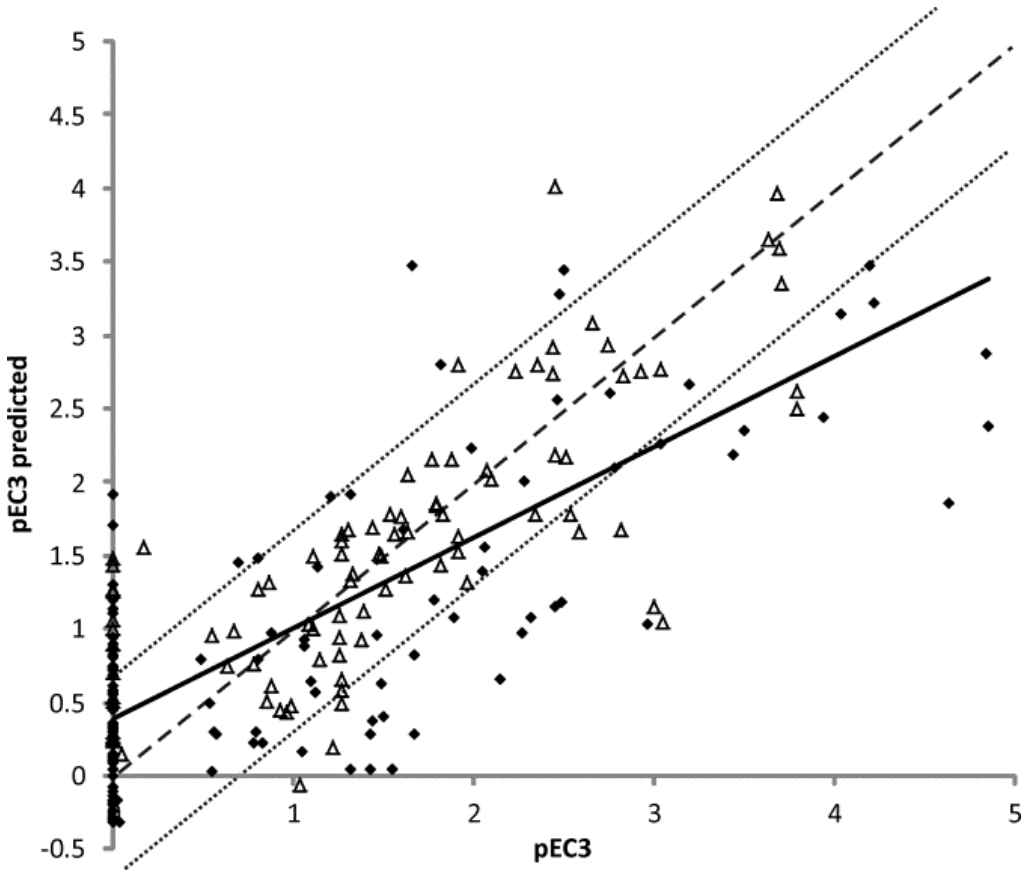


- In general prediction by global model somewhat less accurate as compared to local model

Domain [1)]	N	R^2-adj. of best model (*p*-value)	Fold-misprediction domain model	Fold- misprediction global model
Michael acceptors	44	58.4% (< 0.0005)	**2.26**	3.22
Addition-elimination	19	85.9% (< 0.0005)	**2.60**	3.43
Epoxides	16	81.2% (< 0.0005)	**1.97**	2.88
Aldehydes	28	43% (0.001)	**3.16**	3.26
pre-quinone-domain	32	48.2% (< 0.0005)	**4.54**	6.45

Predictive capacity – local and global models combined

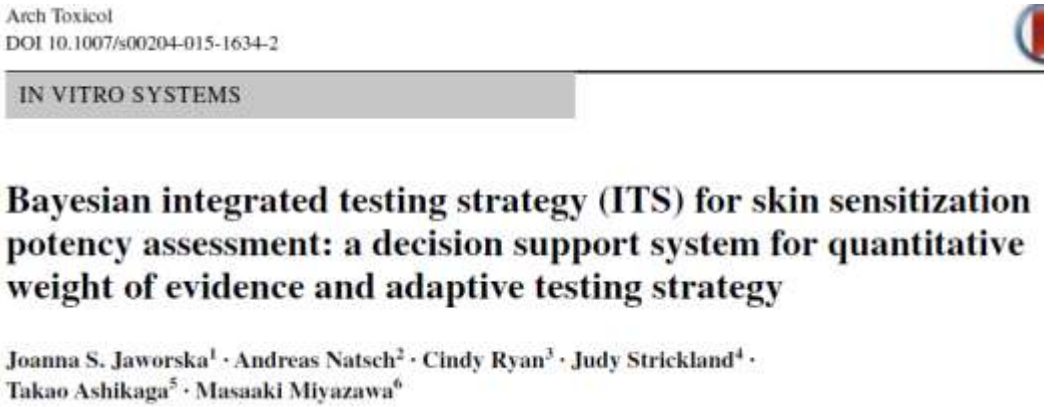
- Combined view of predictions with domain models (open triangels) and global predictions according (closed diamonds).
  - Chemicals attributable to domain predicted by domain model.
  - Remaining chemicals predicted by global model.
- Solid line indicates regression line
- dashed line indicates line of identity
- dotted lines indicate the area of chemicals with ≤ 5 fold misprediction.



Conclusions

- Quantitative readouts from Peptide reactivity and Nrf2-induction can partly explain sensitization potency
- Predictions are most accurate within domains of chemicals reacting with similar mechanism
- Within several domains, predictions with an average 2-fold misprediction are possible
  - Working on a continous scale may be more useful as point of departure in risk assessment as compared to predicting 10-fold potency classes
- There is also a correlation to human data (not shown here, see paper)
  - However, prediction of human data by in vitro data and LLNA is limited, which may be partly due to the very heterogeneous nature of the available human data.

Integration of multiple parameters – Bayesian net



Arch Toxicol
DOI 10.1007/s00204-015-1634-2

IN VITRO SYSTEMS

Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy

Joanna S. Jaworska[1] · Andreas Natsch[2] · Cindy Ryan[3] · Judy Strickland[4] · Takao Ashikaga[5] · Masaaki Miyazawa[6]

- Advantages
  - Probabilistic, no fixed coefficient over scale, each new information refines probabilty distribution
  - Informs about robustness of prediction
  - Can handle very different inputs
  - Can work with data gaps

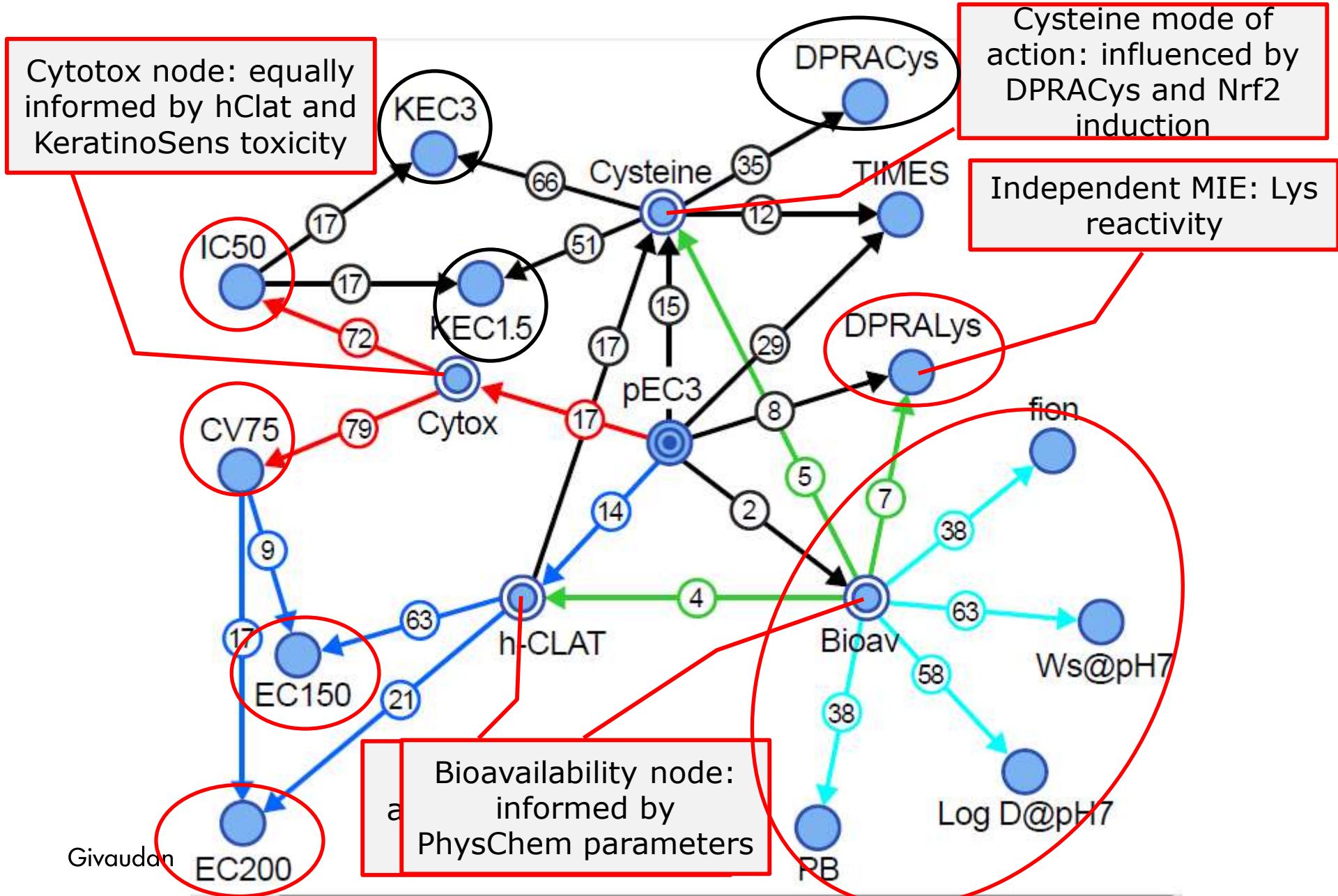
- Disadvantages
  - Data are binned into classes – information loss with continous data
  - Output is a likely class attribution – not a concrete point of departure value
    - **But probability distribution can be recalculated to become a concrete value, see paper!**

Input parameters

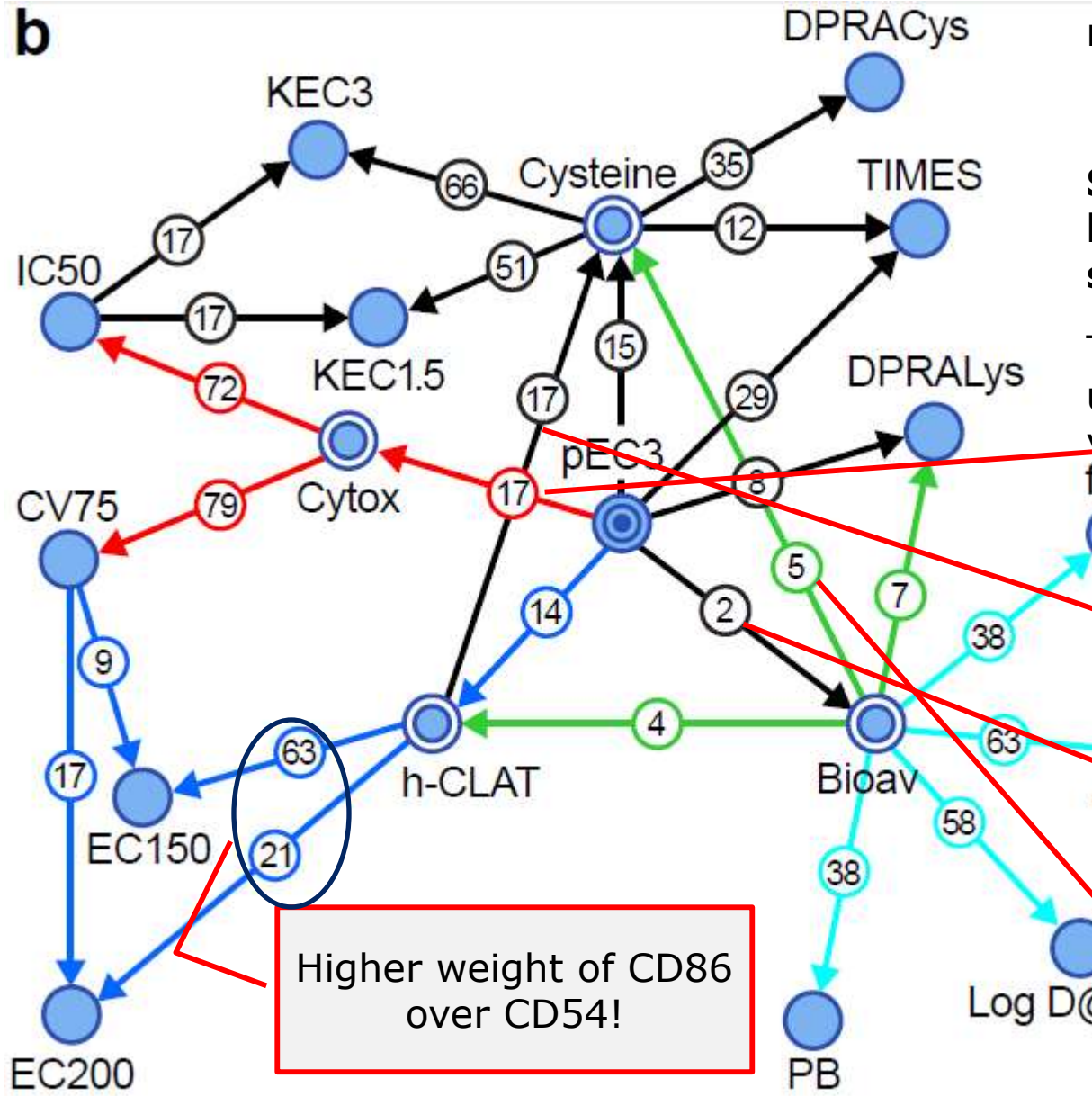
For TIMES, the **stand-alone prediction models (3 classes) have to be used** – there is prior information on relationship of alerts to LLNA classes

Bioavailability	**Ws**-Water solubili… **Log D**- Distributio… **PB**- Plasma protein… **Fion**: Fraction ioniz…
TIMES In silico prediction of potency in vivo:	1. Mechanistic **alert** (in… direct Michael Acceptor) and auto-oxidation 2. Prediction of **3 classes**
Key Event 1:	**DPRACys**, **DPRALys**
Key Event 2:	KeratinoSens™ : 1.5-fold (**KEC1.5**); 3-fold (**KEC3**) induction of luciferase; **IC50** for cytotoxicity
Key Event 3:	h-CLAT: **EC150** (CD86), **EC200** (CD54), **CV75** (Cytotoxicity)

The network structure



The network arcs



Most arcs are defined by mechanistic knowledge

Some arcs are added based on empirical statistical reasons:

These may indicate underlying relationships not yet fu… fion

Relatively high weight of cytotoxicity

Strong arc between Cys reactivity and dendritic cell activation

Low weight of in vivo bioavailabilty

Green arcs: Importance of in vitro bioavailability

Higher weight of CD86 over CD54!

Application of BN-ITS-3

- Clearly defined process to derive prediction
- Checks for completeness of evidence
- Integrates check for applicability domain of individual in vitro tests
  - Only applicable tests are considered
- Correction for Michael acceptor alert
- **Finally gives Bayes factors (B): Which class is reliable is this prediction**

Stron

Strong evidence for

Predicted strong, but weak evidence / flat probability distribution

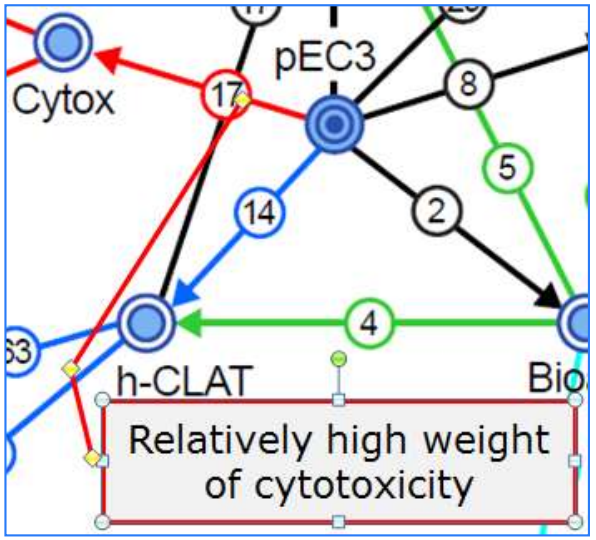
	B („NS“)	B („weak“)	B („moderate“)	B(„srong/ extreme“)
Octannitrile	**129.1**	0.1	0.0	0.0
2-methyl-4H-3,1-benzoxazin-4-one	1.1	0.5	0.0	**5.1**
benzo(a)pyrene	0.11	1.11	1.60	1.75

Some key learnings from the different projects

- Learning 1: Cytotoxicity has a high weight when predicting LLNA data
- Learning 2: Parameters related to bioavilability have little impact on potency
- Learning 3: Different parameters have different weight in different mechanistic domains
- Learning 4: Different parameters have different weight in different potency classes
- Learning 5: Significant redundancy between different *in vitro* parameters!
- Learning 6: Caveat - All these learning are highly affected by the training set: they can, but must not be true for the chemical universe!

Learning 1: Weight of cytotoxicity: Is it a key potency determinant next to reactivity?

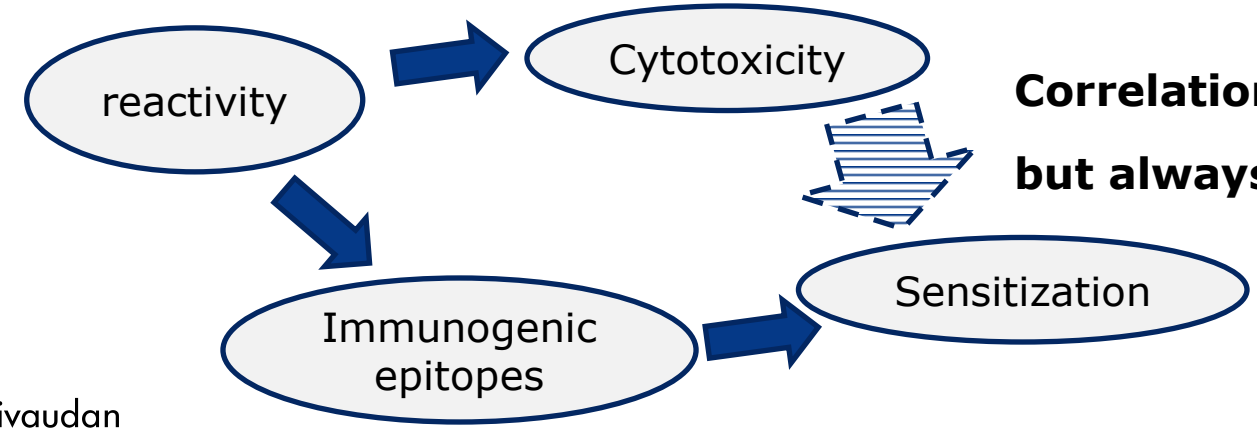
- **High weight in BN**
- Significant weight in global regression model for LLNA potency
- Limited weight in correlation to human data
- Different weight in different mechanistic domains!
- Different weight in different potency classes!



- **Importance of cytotoxicity also reported from**
  - IL-18 / epidermal equivalent assay (SensItIV)
  - SENSIS assay
  - GARD assay
  - VitoSens

Weight of cytotoxicity: Is it a key potency determinant? – some considerations

- **Database caveat** – Broadly used LLNA database contains inflated number of non-sensitizers with low MW and very low cytotoxicity (e.g. butanol, propylene glycol, glycerol ….)
- **LLNA situation**: In LLNA no adjuvans is given – **Molecule must provide danger signal and reactive, immunogenic modifications** (Difference from maximisation tests and some *in vivo* uses!)
  - Danger signal = local trauma, ATP release triggered by cytotox.
- 2nd caveat: Cytotoxicity correlates to irritancy – may trigger **false-positives in LLNA** – when training against LLNA we recapitulate that
- **Cys-Reactivity triggers cytotoxicity** – Cytotoxicity is an epiphenomen of strong reactivity!!



Correlation yes,

but always mechanistic link??

Cytotoxicity – Database caveat

- Most non-sensitizers in 'Silver list' have very low cytotoxicity!

		KeratinoSens results			
Name	LLNA EC 3	ARE I_{max}	ARE EC1.5	Reps. positive[b]	ARE IC50
Non-sensitizers					
Sodium lauryl sulfate	var.[e]	1.2	n.i.	0/2	44.7
Salicylic acid	var.[e]	1.1	n.i.	0/2	>2000
Methyl salicylate	var.[e]	1.2	n.i.	0/2	>2000
Sulfanilamide	NC[f]	1.4	n.i.	0/2	>2000
Diethyl phthalate	>100%	1.1	n.i.	0/2	>2000
Glycerol	>100%	1.2	n.i.	0/4	>2000
Propylene glycol	>100%	1.2	n.i.	0/2	>2000
Benzoic acid	>20%	1.1	n.i.	0/2	>2000
1-Butanol	>20%	1.1	n.i.	0/2	>2000
4-Hydroxybenzoic acid	>25%	1.1	n.i.	0/2	>2000
Sulfanilic acid	>25%	1.3	n.i.	0/2	>1000
Tartaric acid	>25%	1.2	n.i.	0/2	>2000
Propylparaben	>25%	**9.7**	14.5	2/2[f]	813.1
Ethyl vanillin	>50%	**5.4**	161.7	2/2[g]	>2000
Isopropanol	>50%	1.2	n.i.	0/2	>2000
Benzyl alcohol	>50%	1.2	n.i.	0/2	>2000
Dimethylisophtalate	NC[h]	**2.1**	694.9	3/4	>2000
Dextran	NC[h]	1.5	n.i.	0/2	>2000
Tween 80	NC[h]	**2.7**	19.3	2/2	399.8
Chlorobenzene	Neg.[i]	1.2	n.i.	0/2	>2000
Lactic acid	Neg.[i]	1.3	n.i.	1/4	>2000
Phenol	Neg.[i]	1.3	n.i.	0/2	>2000
Benzaldehyde	>25	**2.3**	443.1	2/2[g]	>2000
Octanoic acid	>50	1.1	n.i.	0/2	>2000

Cytotoxicity and the LLNA situation

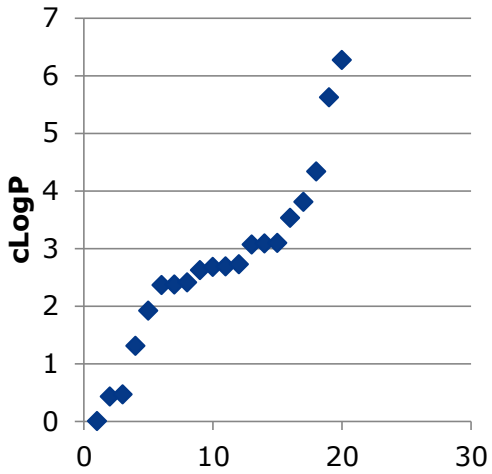
- LLNA is gold standard for potency
- Ability of chemical to provide danger signal is key for positive / more potent LLNA result
- BUT: From fragrance application viewpoint, **danger signal will very rarely be provided by the critical allergen itself**
  - Molecule applied typically at <0.1% in complex product
  - Danger signal normally comes form product excipients or preinflammed skin or co-applied products...
- If we try to best mimic LLNA – allowing for cytotoxicity as key parameter- then we may not always train our system towards the most critical application situation

- Chemicals with equal reactivity but widely differing cytotoxicity will be predicted different (see example epoxides below)

Learning 2: Parameters related to bioavailability have very little impact on potency

- No statistical effect of cLog P in global model
- Example Bayesian net – low impact of bioavailability parameters on EC3! (shown above)
- Example addition-elimination domain:
  - Highly variable logP
  - LogP has **no statistical weight for potency**
  - LogP considered key determinant in skin disposition – but potency driven by reactivity

Log P of SnAR/Sn2 chemicals



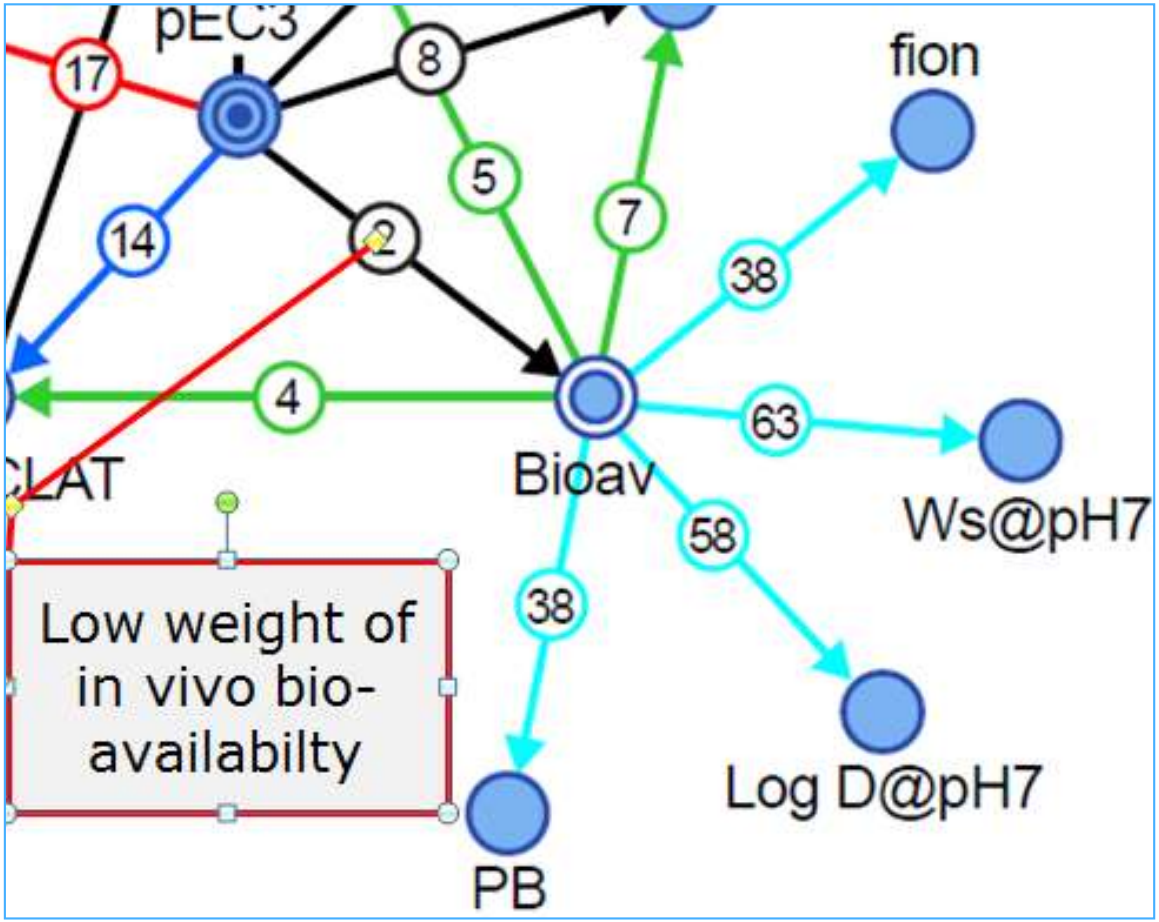
Local regression on prediction of $EC3_{LLNA}$ by *in vitro* and *in chemico* data

$$pEC3_{LLNA} = 0.304 + 0.57 \times \text{Log } K_{norm} + 0.24 \times \text{Log } IC50_{norm} - 0.66 \times \text{Log } VP_{norm} + 0.076 \times \text{cLogP}$$

Constant	T = 0.55, p = 0.590	**Log K_{norm}**	T = 4.95, ***p* < 0.0005**
cLogP	**0.68, *p* = 0.509**	Log $IC50_{norm}$	T = 1.116, p = 0.266
Log VP_{norm}	T = - 3.39, p = 0.005	R^2 (adj) = 85.4%	

Learning 2: Parameters related to bioavailability have very little impact on potency

- Recap: Situation in Bayesian net



Learning 3: Different parameters have different weight in different mechanistic domains

- Example epoxides:

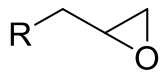
Local regression on prediction of $EC3_{LLNA}$ by *in vitro* and *in chemico* data

$$pEC3_{LLNA} = 4.57 + 0.475 \times \text{Log } IC50_{norm} - 0.66 \times \text{Log } VP_{norm}$$

Constant	T = 5.08,	$p < 0.0005$
Log $IC50_{norm}$	T = 2.38,	$p = 0.03$
Log VP_{norm}	T = - 3.96,	$p = 0.002$

R^2 (adj) = 76.3%

- Potency driven by cytotoxicity and VP!
- Reason: **very similar reactivity of the molecules addressed** - most have same reactive subunit. Difference in LLNA probably driven by different danger signal once reactivity of reactive group is almost equivalent.



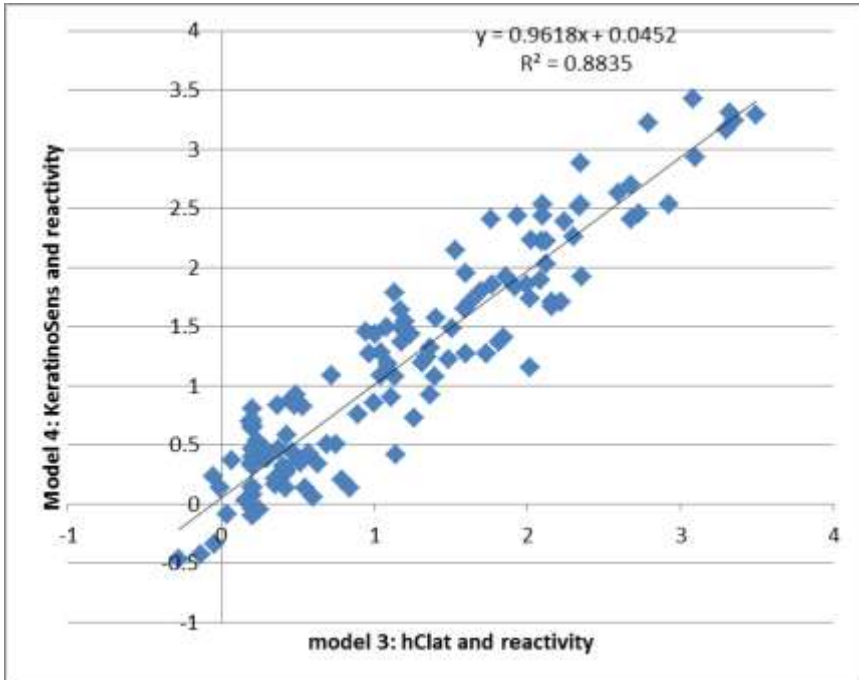
Learning 4: Different parameters have different weight in different potency classes

- Rank of importance of information source in different LLNA potency classes in bayesian net analysis

MI potency overall		**MI for „NS“**		**MI for „WEAK“**		**MI for „MODERATE“**		**MI for „STRONG/ EXTREME“**	
TIMES	28	TIMES	58	TIMES	16	TIMES	18	Cys	21
Cytox	17	Cytox	35	Cys	5.7	h-CLAT	9.6	KEC3	16
Cys	15	CV75	29	Cytox	5.4	EC150	7.4	KEC1.5	15
CV75	14	IC50	28	h-CLAT	4.6	EC200	3.4	h-CLAT	13
IC50	13	Cys	21	KEC1.5	4.5	KEC1.5	1.8	Cytox	12
h-CLAT	13	KEC1.5	20	CV75	3.9	Cytox	1.7	DPRALys	12
KEC1.5	12	KEC3	20	IC50	3.8	Cys	1.5	DPRACys	11
KEC3	12	EC200	17	KEC3	3.5	CV75	1.5	CV75	10
EC150	10	h-CLAT	17	DPRALys	3.0	IC50	1.3	IC50	10

Redundancy between tests: dataset n = 128 with hClat data

- Model with KeratinoSens and reactivity:
  - R^2 = 61.2 %, geomean **fold-misprediction = 3.22**
- Model with h-Clat and reactivity:
  - R^2 = 64.3 %, geomean **fold-misprediction = 3.12**
- Model with h-Clat, KeratinoSens and reactivity:
  - R^2 = 65.3%, geomean **fold-misprediction = 3.05**
- Generally good prediction of hClat model with KS model and vice-versa
- The in vitro models predict each other better than the in vivo response
  - Indicates data redundancy
  - Indicates a gap in coverage of relevant effects to model LLNA



Some thoughts on the way forward

- Understanding reactivity is key
  - Esp. for fragrance molecules where predicting formation of immunogenic conjugates may be more important than danger signal formation
  - Models with too strong emphasis on cytotoxicity (rather then reactivity) may model part of the LLNA response but may not be the most relevant
- Formation of reactive metabolite in skin still a key gap

- Category formation and read accross are good opportunities
- Take learnings from Bayesian net to further build a system which
  - Maximizes use of chemistry information
    - identifies alerts
    - Performs grouping of chemicals
  - Uses *in vitro* and *in chemico* data to correctly rank the new molecule in the group to derive a NESIL

Thank you

Contact

Andreas Natsch, Givaudan Schweiz AG, andreas.natsch@givaudan.com

Process to derive prediction: gathering evidence -1

- Prediction of **physico-chemical properties** of chemicals (logD, Ws@pH7, f_ion, PB)

- Prediction of **TIMES SS**:
  - Potency based on the highest potency among parent molecule and predicted metabolites
  - Assessment of potential of metabolic activations (prohapten) and autooxidation (pre-hapten)
  - reactivity alerts, direct Michael Acceptor

- Completeness of evidence on MIEs check: **Cysteine and Lysine reactivity?**

Process to derive prediction: gathering evidence -2

- Assessment of applicability domains:
  - Biological
    - Pre or prohapten DPRA , KS and hCLAT data are examined with caution. Hypothesis w/o these data is considered.
  - Chemical
    - Ionization: chemicals that are 100% ionized considered not suitable for *in vitro* assays.
    - Water solubility at pH=7 cutoffs for DPRA, KeratinoSens™, hCLAT

Ws at pH=7 [M/l]	**DPRA**	**Keratinosens**	**hCLAT**
<2.5e-08	x	x	x
2.5e-08 - 1.7e-04	ok	x	x
1.7e-04 - 2.1e-04	ok	ok	x
> 2.1e-04	ok	ok	ok

Process to derive prediction - prediction

- Integration of all the in domain evidence and prediction of the pEC3 probability distribution
- Post processing step of probability distribution correction for direct Michael acceptors
- Conversion of probability distribution to Bayes' Factors for final interpretation and decision.

$$B = \frac{P(H = x|e)/P(H = not\_x|e)}{P(H|x)/P(H = not\_x)} = \frac{posterior\ odds}{prior\ odds}$$

Bayes Factor	Strength of evidence
<1	Negative (supports alternative)
1-3	Barely worth mentioning (weak)
3-10	Substantial
>30	Strong

Jeffereys, 1961

- Conversion from pEC3 to EC3% - Estimation of EC3% : 50th and 90th percentile