The use of *in silico* tools to support expert review under ICH M7

Alex Cayley

Alex.cayley@lhasalimited.org

ICH M7 guidance

Table 1: Impurities Classification with Respect to Mutagenic and Carcinogenic Potential and Resulting Control Actions

Class	Definition	Proposed action for control (details in Section 7 and 8)
1	Known mutagenic carcinogens	Control at or below compound-specific acceptable limit
2	Known mutagens with unknown carcinogenic potential (bacterial mutagenicity positive*, no rodent carcinogenicity data)	Control at or below acceptable limits (appropriate TTC)
3	Alerting structure, unrelated to the structure of the drug substance; no mutagenicity data	Control at or below acceptable limits (appropriate TTC) or conduct bacterial mutagenicity assay; If non-mutagenic = Class 5 If mutagenic = Class 2
4	Alerting structure, same alert in drug substance or compounds related to the drug substance (e.g., process intermediates) which have been tested and are non-mutagenic	Treat as non-mutagenic impurity
5	No structural alerts, or alerting structure with sufficient data to demonstrate lack of mutagenicity or carcinogenicity	Treat as non-mutagenic impurity

*Or other relevant positive mutagenicity data indicative of DNA-reactivity related induction of gene mutations (e.g., positive findings in *in vivo* gene mutation studies)

A computational toxicology assessment should be performed using (Q)SAR methodologies that predict the outcome of a bacterial mutagenicity assay (Ref. 6). Two (Q)SAR prediction methodologies that complement each other should be applied. One methodology should be expert rule-based and the second methodology should be statistical-based. (Q)SAR models utilizing these prediction methodologies should follow

If warranted, the outcome of any computer system-based analysis can be reviewed with the use of expert knowledge in order to provide additional supportive evidence on relevance of any positive, negative, conflicting or inconclusive prediction and provide a rationale to support the final conclusion.

What are we aspiring to with *in silico* predictions?

Predictivity

Transparency

Super Expert Scientist

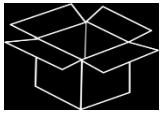
“right” = Every time
Explanation = Full and Reasoned

“right” = Sometimes
Explanation = None

What are we aspiring to with *in silico* predictions?

Predictivity

Transparency



“right” = Most Times
Explanation = None

Expert Scientist

“right” = Most Times
Explanation = Full and Reasoned

Super Expert Scientist

“right” = Every time
Explanation = Full and Reasoned

“right” = Sometimes
Explanation = None

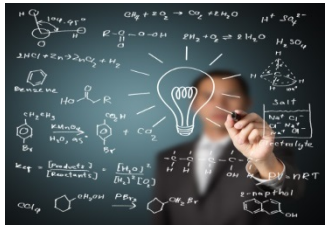
“right”= Sometimes
Explanation = Some

What are we aspiring to with *in silico* predictions?

Validation

	Test Set	Performance Stats BA	SEN	SPEC
1	Pharma A	72	68	76
2	Pharma B	63	38	89
3	Pharma C	72	64	80
4	Pharma D	75	65	85
⋮				
14	Pharma N	xx	xx	xx

Expert Interpretation



Super Expert Scientist

“right” = Every time
Explanation = Full and Reasoned

Using expert interpretation to improve predictions

Table 5
(Q)SAR vs. Q(SAR) plus expert evaluation. Summary of concordance.

Structural assessment	Predicted Ames result	Number of compounds[a]	Number of Ames positive[b] (%)	Number of Ames negative[c] (%)
(Q)SAR	Negative	566	35 (6)	531 (94)
(Q)SAR + expert evaluation	Negative	408	5 (1)	403 (99)

[a] Total number of compounds predicted negative.
[b] The number of compounds predicted negative that produced a positive response in the Ames assay.
[c] The number of compounds predicted negative that produced a negative response in the Ames assay.

Table 2
Use of recent versions of DEREK alone (push-button).

DEREK (version)	A (v12 + 13)	B (v13)	C (v13)	D (v9 – Nexus 2.0)	E (Nexus 2.0)
Sensitivity (true pos/Ames pos)	44% (67/153)	72% (28/39)	97% (36/37)	65% (37/57)	85% (61/72)
Specificity (true neg/Ames neg)	78% (355/455)	70% (161/230)	6% (5/82)	50% (70/140)	81% (149/184)
Negative predictivity (true neg/pred neg)	80% (355/441)	94% (161/172)	83% (5/6)	78% (70/90)	93% (149/160)
Positive predictivity (true pos/pred pos)	40% (67/167)	29% (28/97)	32% (36/113)	35% (37/107)	64% (61/96)
Concordance (correct pred/total)	69% (422/608)	70% (189/269)	34% (41/119)	54% (107/197)	82% (210/256)

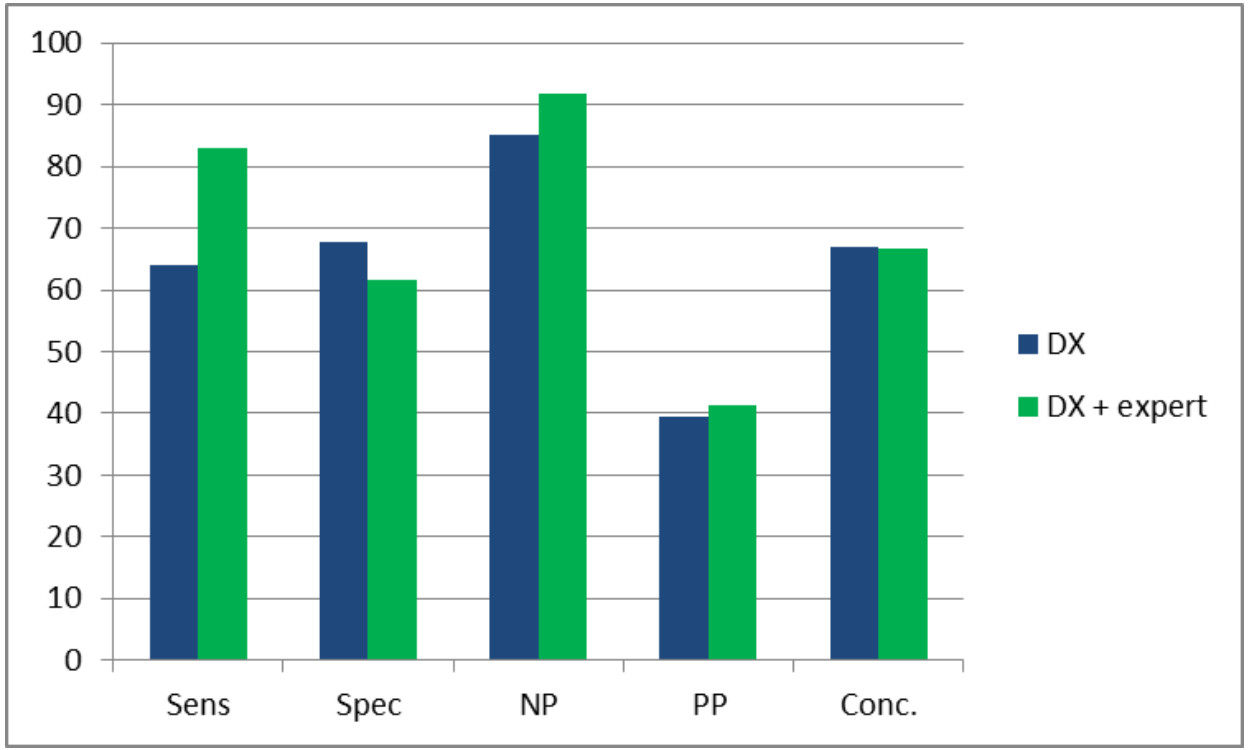
Table 5
Use of DEREK combined with expert knowledge.

DEREK (version)	A (v12 + 13)	B (v13)[a]		C (v13)	D (v9 – Nexus 2.0)	E (Nexus 2.0)[b]
		In-house LSMA model	Public and in-house databases			
Sensitivity (true pos/Ames pos)	77% (117/153)	85% (33/39)	79% (31/39)	97% (36/37)	80% (46/57)	93% (67/72)
Specificity (true neg/Ames neg)	63% (287/455)	66% (151/230)	72% (165/230)	5% (4/82)	47% (66/140)	81% (149/184)
Negative predictivity (true neg/pred neg)	89% (287/323)	96% (151/157)	95% (165/173)	80% (4/5)	86% (66/77)	97% (149/154)
Positive predictivity (true pos/pred pos)	41% (117/285)	30% (33/112)	32% (31/96)	31% (35/114)	38% (46/120)	66% (67/102)
Concordance (correct pred/total)	66% (404/608)	68% (184/269)	73% (196/269)	34% (40/119)	57% (112/197)	85% (216/254)

[a] Expert knowledge = Internal LSMA model only or public and in-house database search only.
[b] Proprietary data used to create QSAR system: Internal LSMA model.

Dobo et al.; Reg. Tox. Pharm.; 62; 449-455 (2012), Sutter et al.;Reg. Tox. Pharm.; 67; 39-52 (2013), Powley et al.; Reg. Tox. Pharm.; 71; 295-300 (2015)

Using expert interpretation to improve predictions



Dobo et al.; Reg. Tox. Pharm.; 62; 449-455 (2012), Sutter et al.;Reg. Tox. Pharm.; 67; 39-52 (2013), Powley et al.; Reg. Tox. Pharm.; 71; 295-300 (2015)

Supplying the information required

SX Hypotheses

DX negative predictions

DX comments and examples

DX Alert

Read Across

SX nearest neighbours

Reasoning behind prediction

Strain profile

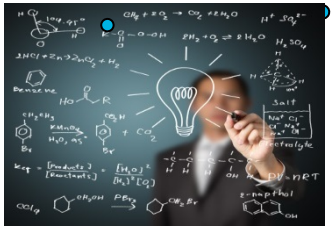
Data sources

Data quality

DX references

SX references in VX

Special protocols



Key: DX = Derek Nexus, SX = Sarah Nexus, VX = Vitic Nexus

Conclusions

- A good baseline predictivity is important for an *in silico* system to be used as part of ICH M7
- This baseline performance can be improved by expert interpretation of the information provided
- QSAR providers should give information that is useful for expert analysis and in a way that makes this as easy as possible
- Is there anything we are missing or could do better to supply the information you need to make your expert analysis?

Barber, C. et al.(2015). Establishing best practise in the application of expert review of mutagenicity under ICH M7. *Regulatory Toxicology and Pharmacology*, *73*(1), 367–377. http://doi.org/10.1016/j.yrtph.2015.07.018
Amberg, A. et al (2016). Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses. *Regulatory Toxicology and Pharmacology*. http://doi.org/doi:10.1016/j.yrtph.2016.02.004
Powley, M. W. (2015). (Q) SAR assessments of potentially mutagenic impurities : A regulatory perspective on the utility of expert knowledge and data submission. *Regulatory Toxicology and Pharmacology*, *71*(2), 295–300. http://doi.org/10.1016/j.yrtph.2014.12.012

Acknowledgements

- Chris Barber
- Richard Williams
- All of my colleagues at Lhasa Limited
- All of our members

Work in progress disclaimer

This document is intended to outline our general product direction and is for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon. The development, release, and timing of any features or functionality described for Lhasa Limited's products remains at the sole discretion of Lhasa Limited.