Table Host Version (student version does not have the text that is in red)

2018 SOT *In Vitro* Lecture

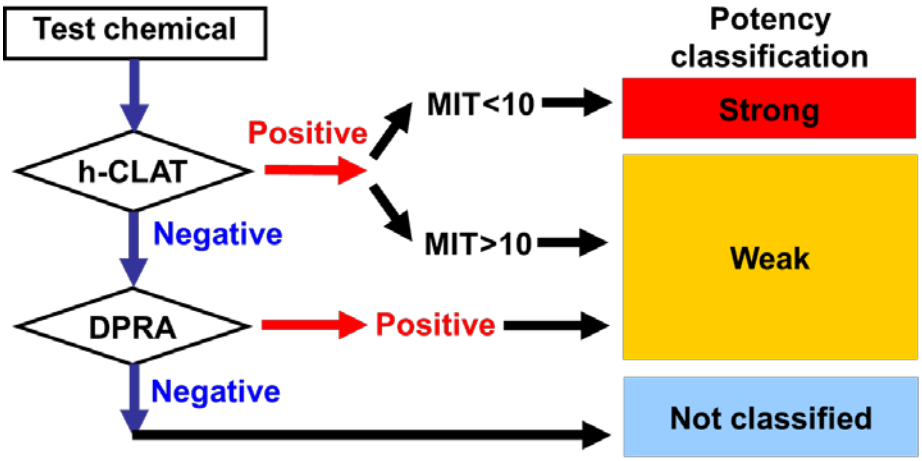
Case Study: Evaluating a Non-Animal Defined Approach for Skin Sensitization

Is this non-animal defined approach ready to replace the LLNA?

Consider the example non-animal defined approach below for replacement of the LLNA. Seven criteria categories are listed below with questions to guide your evaluation. *[the skin sensitization AOP will be projected during this discussion]*

Example: Sequential Testing Strategy (STS) Defined Approach (DA)

The figure below summarizes a DA that is currently being considered as an alternative approach to replace the LLNA. The DA's output is a potency classification (i.e., chemical is categorized for likelihood as a strong or weak skin sensitizer, or is not classified as a skin sensitizer). The minimum induction threshold (MIT) is the quantitative readout of the h-CLAT assay, and determines the potency for chemicals that are positive in the h-CLAT. Chemicals that are negative in the h-CLAT are then run in the DPRA to either confirm their lack of sensitization activity, or classify them as weak sensitizers.



Adapted from Takenouchi et al. 2015

1. Structure: Components, Information Provided

- This DA is a simple decision tree
- Outputs are hazard and potency predictions
- This DA depends on two OECD Test Guideline methods: hCLAT, DPRA

➢ **What are advantages/disadvantages to this STS approach?**

- Advantage is the simplicity: only have to run two assays to get an outcome (financially feasible and less demanding on lab resources). Also, consider how complex a DA could be that integrates multiple data sources; a simple decision tree is quite simple and easy to do.
- Limitation could be how clear results are from input assays (is it easy to decide the positive vs. negative outcome from each assay)?

2. Relevance: Mechanistic Coverage

- This DA integrates data from two OECD Test Guideline assays representing key events (including molecular initiating event) of the AOP

➢ **Would this DA be acceptable with other assays if they addressed the same key events in the AOP (i.e., same framework but using assays other than h-CLAT or DPRA)? If so, under what conditions?**

- The LLNA test represents the fourth key event in the AOP. Inputs for this DA are from the earlier key events that lead to this one. Are there acceptable conditions warranting use of other assays that inform on the same biology or do you feel that using the inputs as defined is best to achieve relevance?
- If other assays are proposed as substitutions, how should they be validated? Do they need to have OECD Test Guidelines associated with them, or is the level of validation required up to the end-user (e.g., the regulatory agency)?

3. Predictive Accuracy: Performance Compared to Reference Data

Test Method:	*Non-animal: STS DA*	*Animal: LLNA*
Accuracy:*	80%	74%

*Accuracy was assessed vs. human data (n = 128)

➢ **In the absence of human data, would the reproducibility of the animal test be a reasonable threshold for predictive performance of a DA?**

- Having human data is unique for skin sensitization, so we can evaluate accuracy relative to human outcomes. However, if we didn't have human data (most other toxicities), would comparing to another animal tests be effective to evaluate predictive accuracy?
- Limitations are that animal doesn't always predict human well to begin with (as is seen here where the LLNA animal test is only 74% accurate). But when there is nothing else we must use animal as reference data – so should we have a lower threshold for performance? Is 74% acceptable for achieving human safety? The reproducibility of the LLNA (concordance of independent studies on the same chemicals) is also around 75% for hazard outcomes.

4. Reliability: Reproducibility

- Decision tree is rule-based, i.e., 100% reproducible
- Depends on the reproducibility of the assays input (>80% for *in vitro* assays)

➢ **What else does the reliability of this DA depend on?**

- Test methods must be clearly defined, easily reproduced, and commercially available from multiple sources
- Decision tree is clear, and assays used as input are well defined
- Is interpretation of data from the assays (i.e., positive vs. negative) clear so that independent groups would always get the same results?

5. Applicability: Technical Limitations, Chemical Space

- Depends on the technical limitations of the data sources (assays).

➢ **How would you characterize the applicability domain of a DA that depends on multiple assays, with different limitations?**

- There are limits when using alternative assays systems. For *in vitro* assays this may be a restriction to only being able to assay chemicals that are soluble in DMSO or water-soluble (i.e., volatile chemicals from sprays would be difficult to test). Or if the assays have a colorimetric readout, then any colorants could not be reliably tested.
- Realizing that the h-CLAT and DPRA may have different restrictions, the overall DA would only be applicable toward chemicals that fall within the domain of both *in vitro* assays.

6. Complexity: Data Interpretation Procedure

- Simple decision tree, straightforward to apply.

➢ **Consider more complex DAs. For example, would a machine learning algorithm be an acceptable replacement? If so, under what conditions?**

- As long as interpretation of outputs from the *in vitro* assays is straightforward, the STS should be easy to apply and interpret.
- For perspective, note that this simplicity is not always the case. Other methods may be more complex. Some DAs are based on computational algorithms such as machine learning but we don't necessarily explicitly know how the computer decided on the logic for categorization. If that were the case would such computationally complex approaches be acceptable? Computational methods may perform better, but what level of complexity is acceptable and how does the group feel about drawing a line on complexity? Should there be a line?
- Is complexity a subjective enough aspect that the acceptability of the DA should be determined by the end-user?

7. Transparency: Proprietary Elements

- Fully transparent, third-party assessment performed (Kleinstreuer et al. 2018).

➢ **Do the information sources/assays in the DA need to be widely available, or can they be proprietary? What conditions would apply to each?**

- Consider whether there is sufficient documentation available to be able to apply the DA yourself. Since the STS uses OECD validated and defined test guidelines, the large amount of documentation for the h-CLAT and DPRA makes it easy to set this up in your lab and take the outputs and put it into this framework.
- But some other approaches require computational models that require software to which you may not have access. This makes the DA under consideration for the case study appealing because no proprietary software is needed and you have access to documentation for the input assays. If you had to go to a company to run an assay or buy software, there may be limits to transparency. Would that still be acceptable for a potential DA?