Supporting read-across predictions of chemical toxicity using high-throughput text-mining

Nancy C. Baker, PhD
Leidos, contractor to the US EPA
ACS National Meeting, San Francisco
April 6, 2017

DISCLAIMER: The views expressed in this presentation are those of the presenter and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Acknowledgements

- Grace Patlewicz
- Thomas Knudsen
- Kevin Crofton
- Rusty Thomas
- NCCT Team

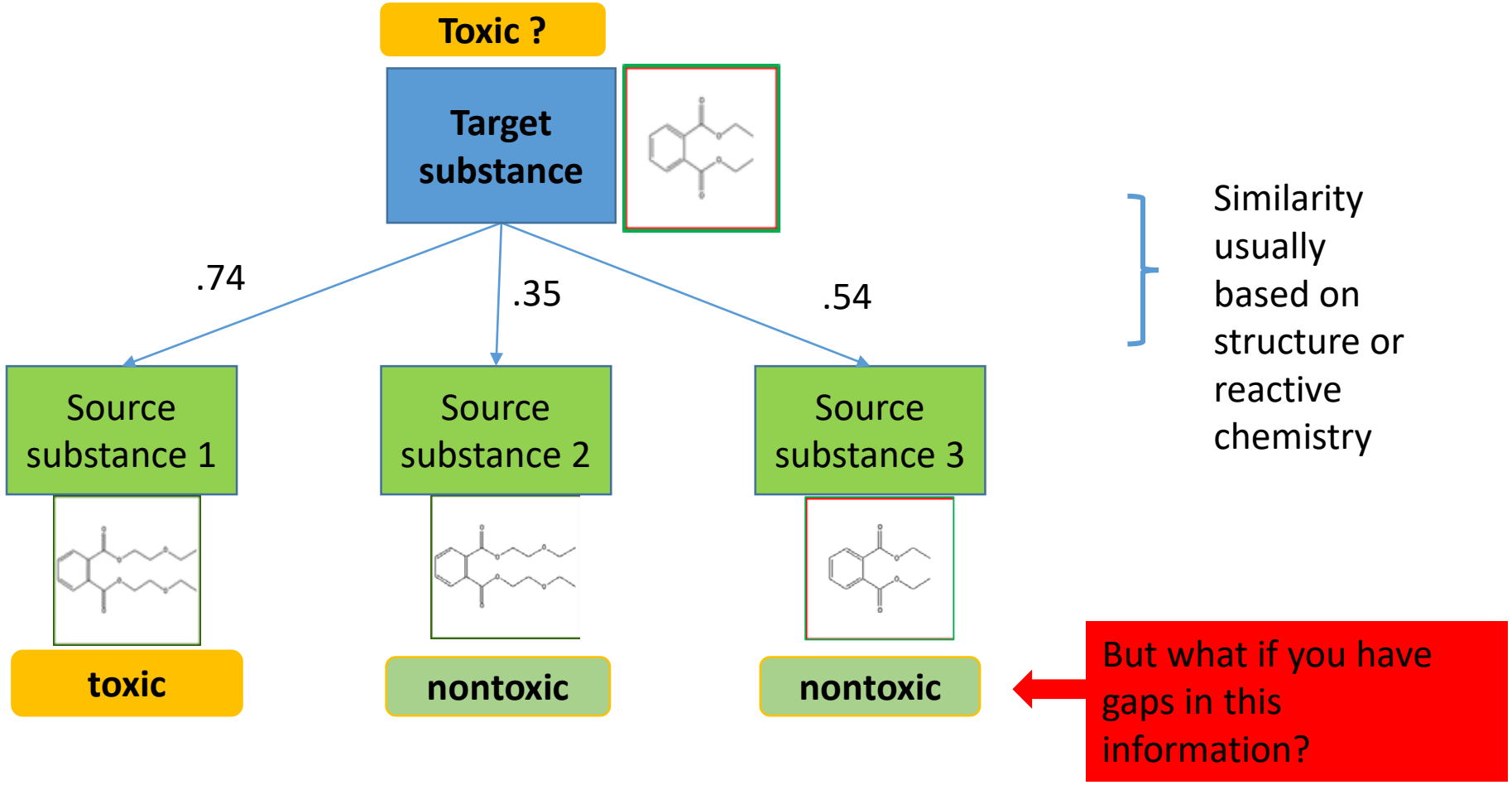
Contact: baker.nancy@epa.gov

Outline

- Read-across overview
- What is the problem we're trying to solve?
- Why literature mining?
- Methods to gather and process information from the biomedical literature
- Examples of how literature mining can be applied to read-across

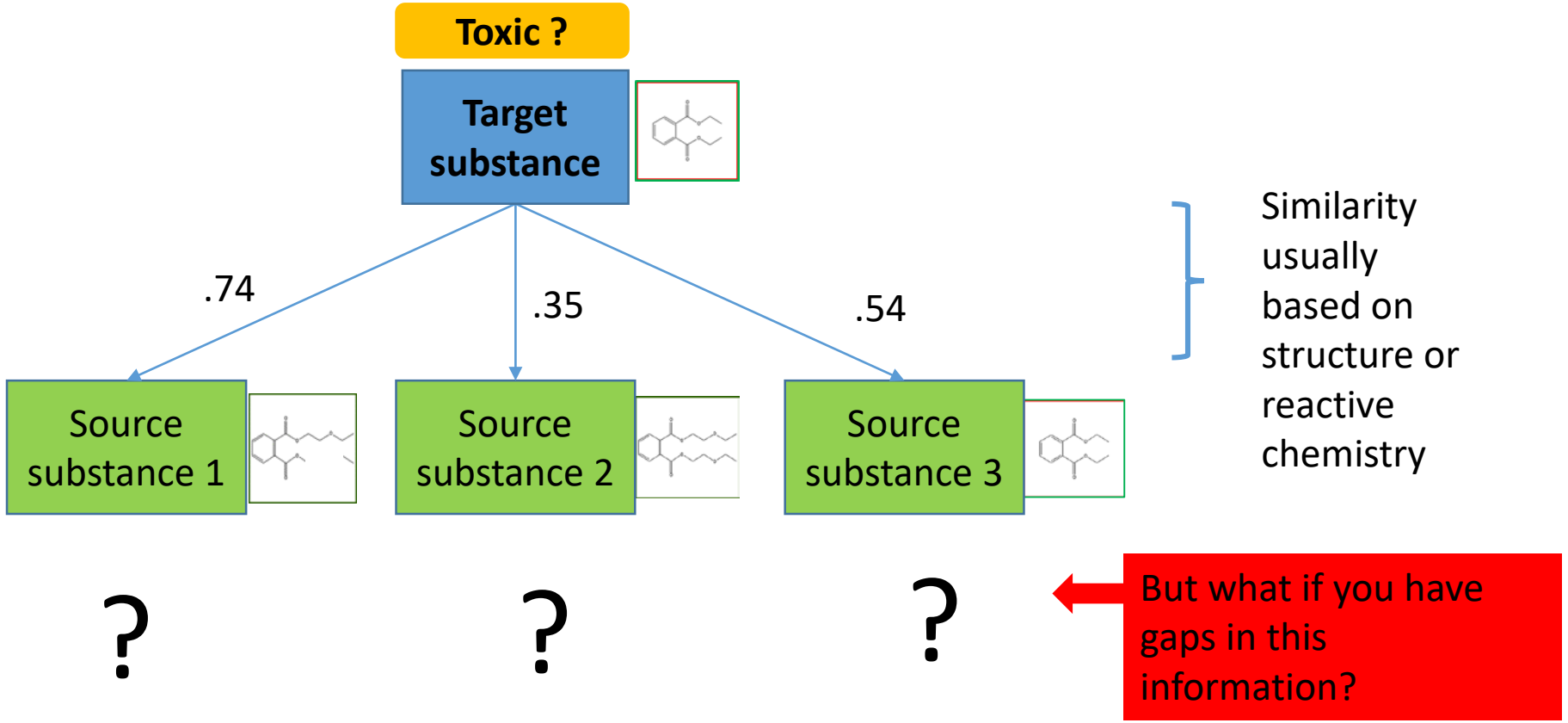
Read-across: technique for filling data gaps

Rationale: similarity principle



Read-across: technique for filling data gaps

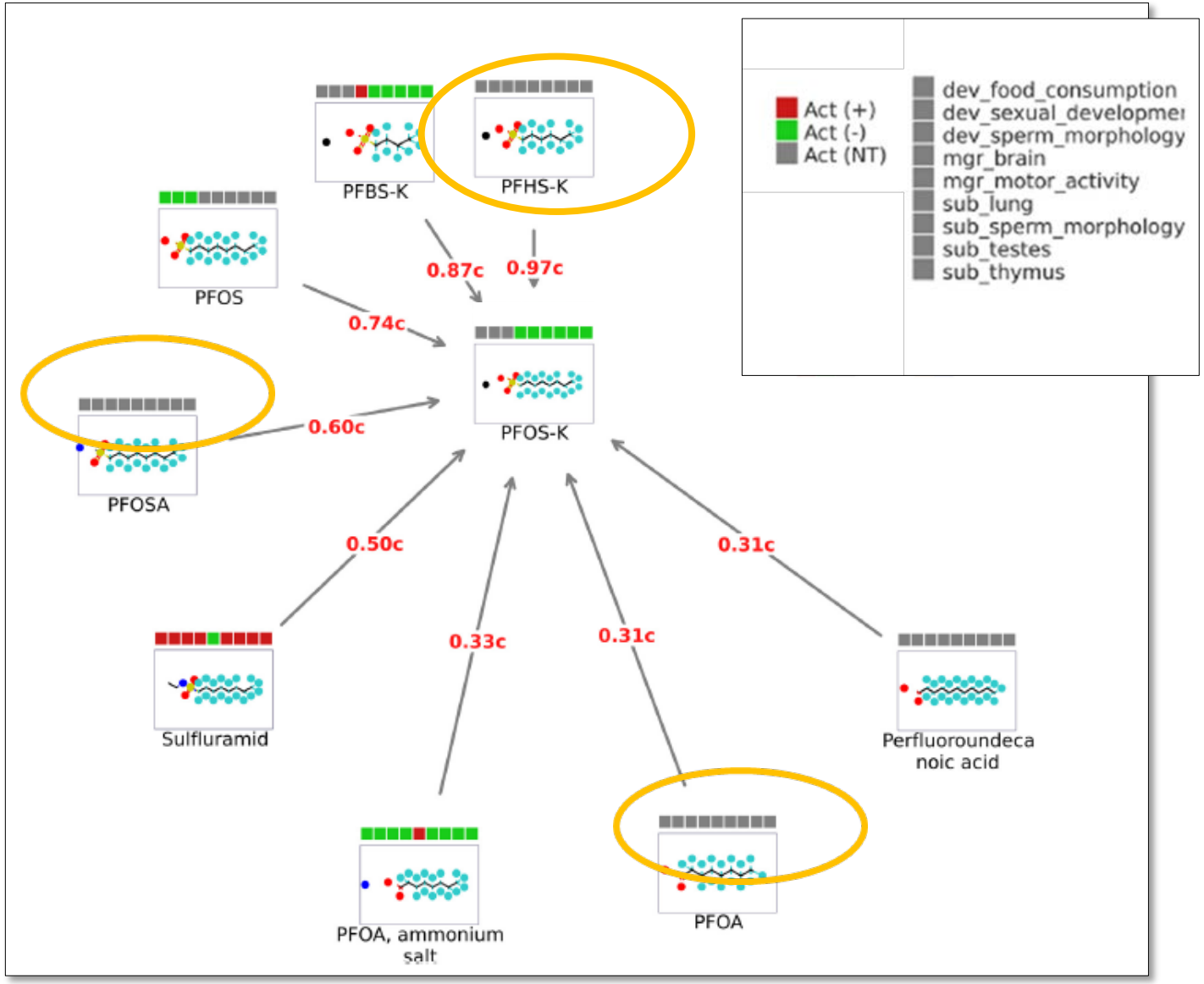
Rationale: similarity principle



The more information a researcher can bring to bear on the decision, the better the decision is likely to be.

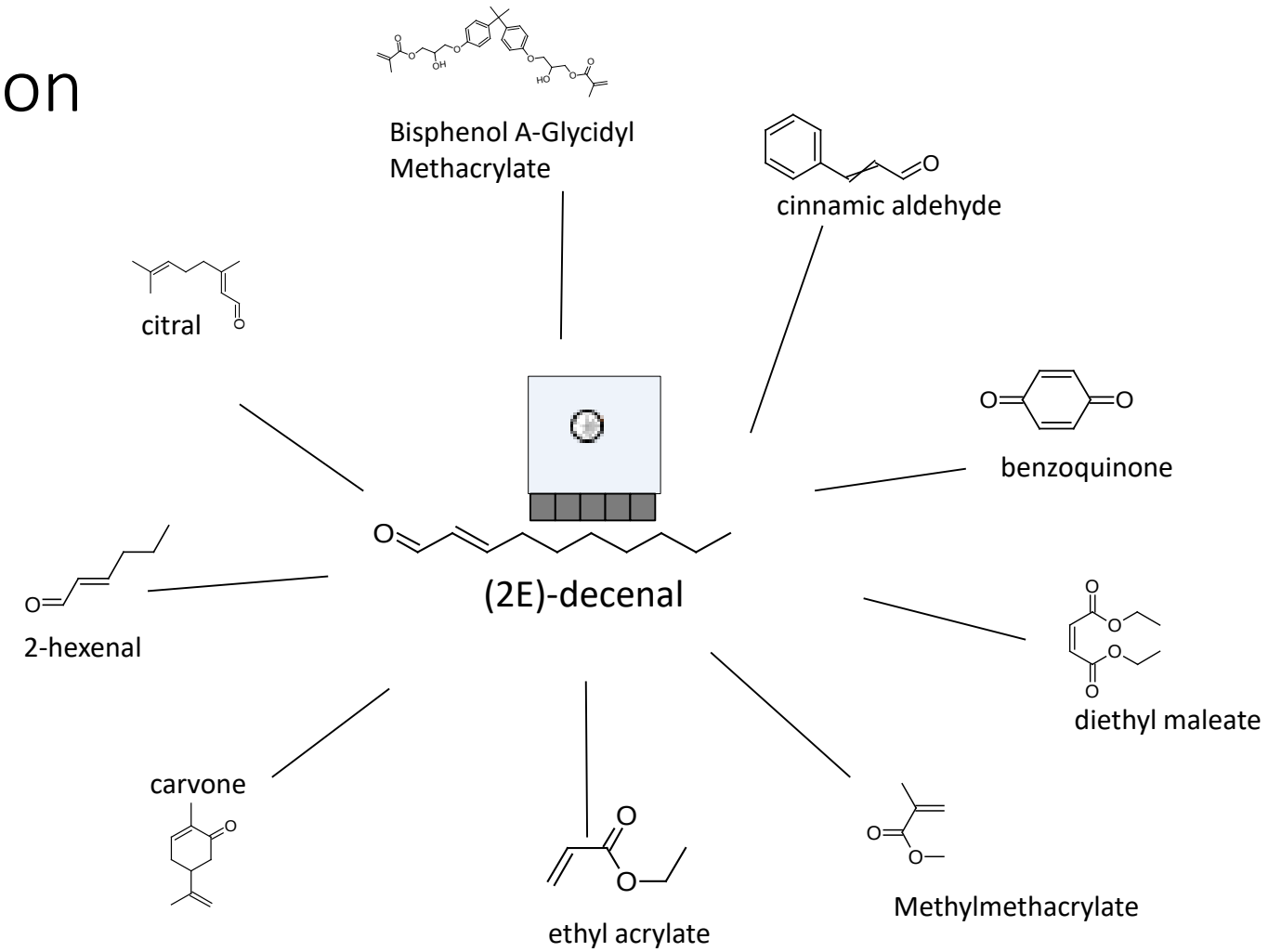
Example 1: Filling the information gap with ToxRefDB animal assay data

Problem: ToxRefDB data is not available for all chemicals or for all endpoints.



Modified from Shah, I. et al. Regulatory Toxicology and Pharmacology, Volume 79, 2016,12-24

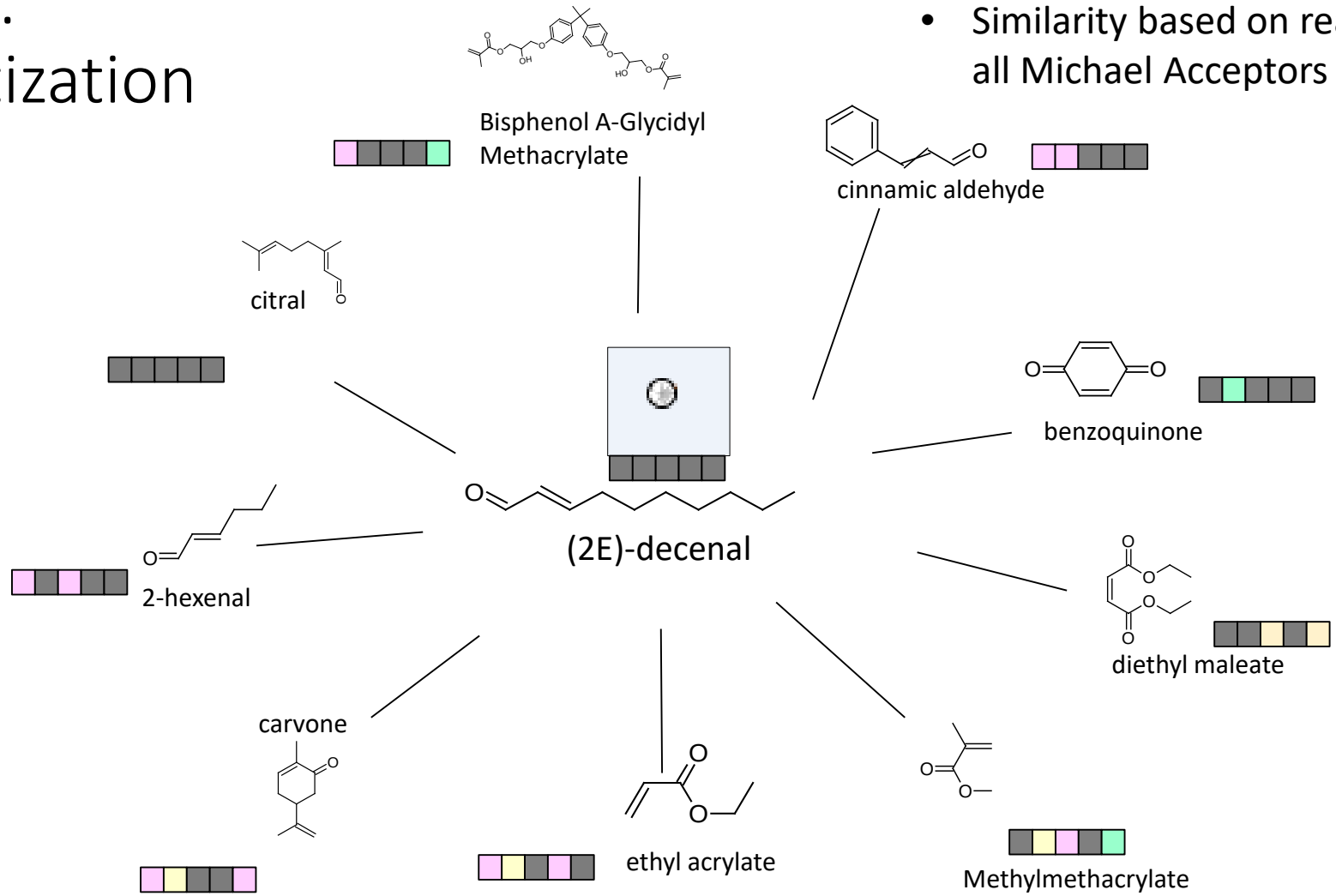
Example 2:
Skin sensitization



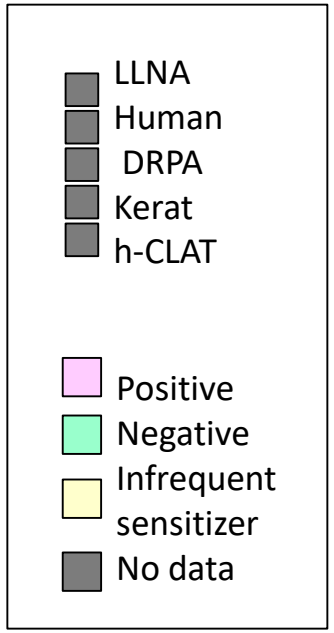
- Similarity based on reactive chemistry - all Michael Acceptors

Example 2: Skin sensitization

- Similarity based on reactive chemistry - all Michael Acceptors



In vivo / in vitro key



Problem: In vitro and animal data is not available for all chemicals or for all endpoints.

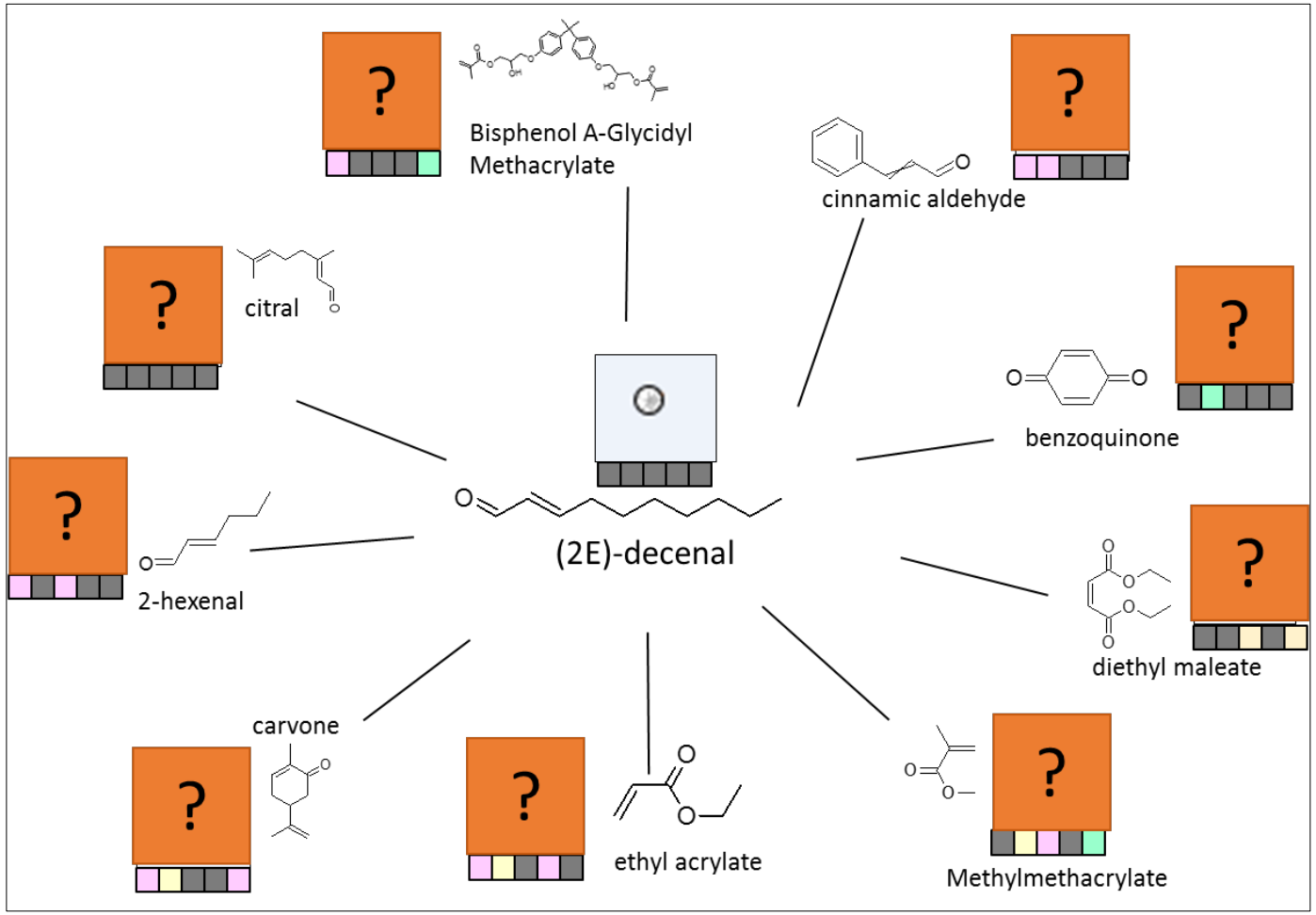
Literature can inform read-across

- Why literature?
  - Large and growing source
    - > 26 million articles; > 12 million about chemicals
    - Encompasses all sorts of toxicity
- Challenges
  - Literature is unstructured
  - Limitations of literature mining; e.g., publication bias and granularity
  - Large and growing source – challenges for human cognition of big data
  - Curation and validation
- Goal: gather and condense literature information into a signature that can be visualized for read-across.

Goal

?

- Construct a literature signature or fingerprint or descriptors
- Visualize



Methodology for literature mining to inform read-across

1. Defining and organizing toxicity type
2. Gathering and extracting the literature information
3. Condensing and strengthening signal → signature
4. Visualizing

1. Defining and organizing toxicity type

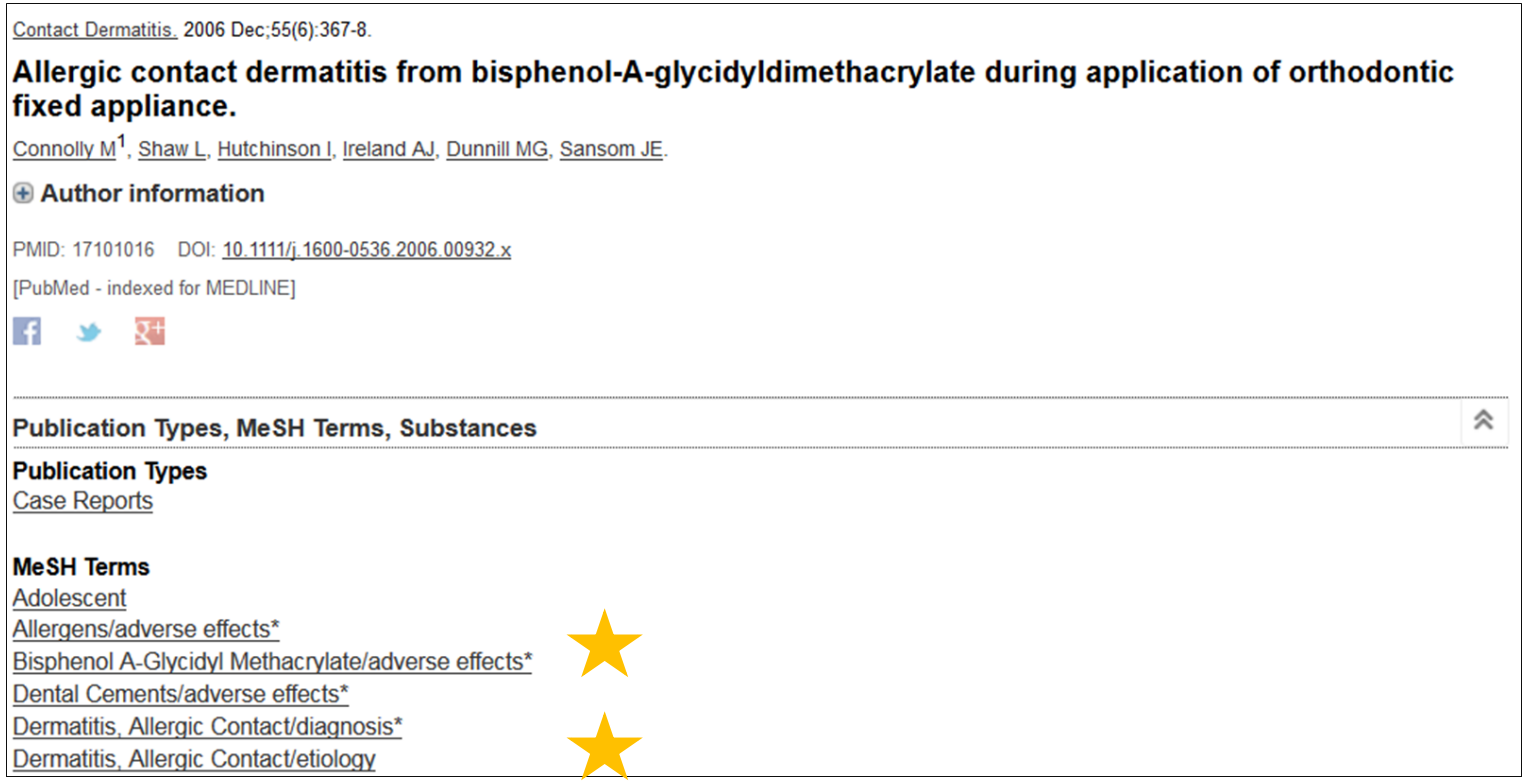
Toxicity Type
Skin sensitization
GeneTox

Category	MeSH Term	Qualifier	Score
Dermatitis Contact/Atopic	Dermatitis		1
Dermatitis Contact/Atopic	Dermatitis	Chemically induced OR etiology	3
Immune Processes	Cross Reactions		1
Immune Processes	Cross Reactions	Drug effects	3
Cell	Lymphocytes		1
Cell	Lymphocytes	Drug effects	3
Chemical mediators	Cytokines		2

Category	MeSH Term	Qualifier	Score
DNA Damage/Repair	DNA Damage		1
DNA Damage/Repair	DNA Damage	Drug effects	3
Mutagens	Mutagens		2
Genetic structures	DNA	Drug effects	2

Subject matter experts advise on the terms and categories for each Toxicity Type.

2. Gathering and extracting



Contact Dermatitis. 2006 Dec;55(6):367-8.

Allergic contact dermatitis from bisphenol-A-glycidyldimethacrylate during application of orthodontic fixed appliance.

Connolly M[1], Shaw L, Hutchinson I, Ireland AJ, Dunnill MG, Sansom JE.

Author information

PMID: 17101016 DOI: 10.1111/j.1600-0536.2006.00932.x

[PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

Publication Types

Case Reports

MeSH Terms

Adolescent

Allergens/adverse effects*

Bisphenol A-Glycidyl Methacrylate/adverse effects*

Dental Cements/adverse effects*

Dermatitis, Allergic Contact/diagnosis*

Dermatitis, Allergic Contact/etiology

2. Gathering and extracting

Br Dent J. 1997 Oct 25;183(8):297-8.

Allergic contact dermatitis to bisphenol-A-glycidyldimethacrylate (BIS-GMA) dental resin associated with sensitivity to epoxy resin.

Carmichael AJ[1], Gibson JJ, Walls AW.

Author information

Abstract

A patient presented with recurrent facial dermatitis associated with dental work. Dermatology referral identified the cause as allergic contact dermatitis to the epoxy acrylate BIS-GMA. Occupationally-induced allergic contact dermatitis to epoxy resin was also demonstrated. A structurally distinct aliphatic acrylate was successfully substituted. Contact sensitivity to BIS-GMA is reviewed and the potential for epoxy sensitive patients to cross-react to BIS-GMA is discussed.

MeSH terms

- Bisphenol A-Glycidyl Methacrylate/adverse effects*
- Chronic Disease
- Cross Reactions
- Dermatitis, Allergic Contact/diagnosis
- Dermatitis, Allergic Contact/etiology*
- Epoxy Resins/adverse effects*
- Facial Dermatoses/chemically induced*

Arch Toxicol. 2011 Nov;85(11):1453-61. doi: 10.1007/s00204-010-0593-x. Epub 2010 Sep 29.

Bisphenol A-glycidyl methacrylate induces a broad spectrum of DNA damage in human lymphocytes.

Drozdz K[1], Wysokinski D, Krupa R, Wozniak K.

MeSH Terms

- Bisphenol A-Glycidyl Methacrylate/toxicity*
- Cell Cycle/drug effects
- Cell Line, Tumor
- Comet Assay
- DNA Breaks, Double-Stranded/drug effects*
- DNA Repair
- DNA Repair Enzymes/metabolism
- Humans
- Lymphocytes/cytology
- Lymphocytes/drug effects*

2. Organizing literature results using toxicity type and categories – getting numbers

Example : Bisphenol A-Glycidyl Methacrylate

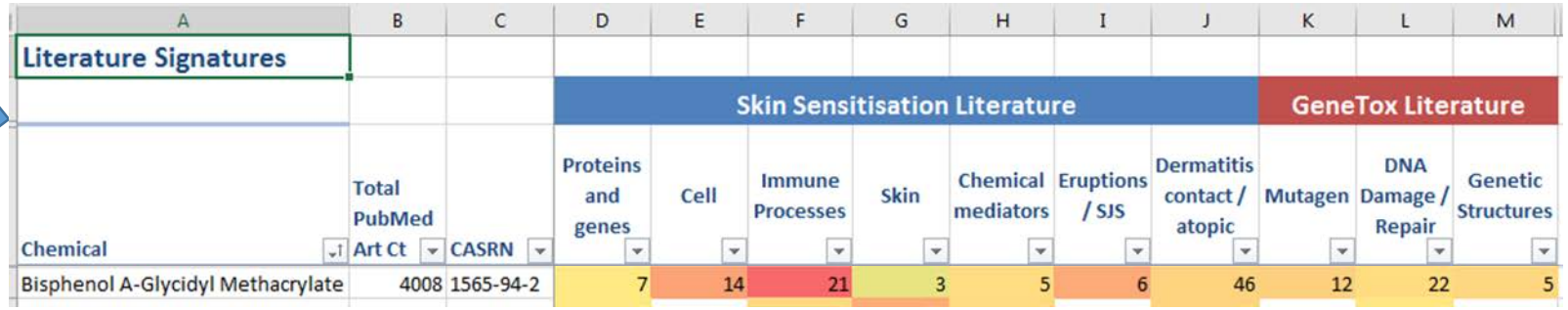
MeSH heading	Score	ToxType	Category	Category Score
Contact Dermatitis	2	Skin sensitization	Dermatitis contact / atopic	
Dermatoses	2	Skin sensitization	Dermatitis contact / atopic	4
Lymphocyte	3	Skin sensitization	Cell	3
Cross Reactions	1	Skin sensitization	Immune Processes	2
DNA Repair	2	GeneTox	DNA Damage / Repair	2
DNA / drug effects	1	GeneTox	Genetic Structures	2

Next step: 3. Condensing into a signature.

Having a score that can be summarized is key.

3. Condensing into a signature

MeSH heading	Score	ToxType	Category	Category Score
Contact Dermatitis	2	Skin sensitization	Dermatitis contact / atopic	
Dermatoses	2	Skin sensitization	Dermatitis contact / atopic	4
Lymphocyte	3	Skin sensitization	Cell	3
Cross Reactions	1	Skin sensitization	Immune Processes	2
DNA Repair	2	GeneTox	DNA Damage / Repair	2
DNA / drug effects	1	GeneTox	Genetic Structures	2



Literature Signatures			Skin Sensitisation Literature							GeneTox Literature		
Chemical	Total PubMed Art Ct	CASRN	Proteins and genes	Cell	Immune Processes	Skin	Chemical mediators	Eruptions / SJS	Dermatitis contact / atopic	Mutagen	DNA Damage / Repair	Genetic Structures
Bisphenol A-Glycidyl Methacrylate	4008	1565-94-2	7	14	21	3	5	6	46	12	22	5

Having a score that can be summarized is key.

3. Condensing into signature

Each numeric value is the total score for that chemical for the corresponding toxicity category / type.

Literature Signatures

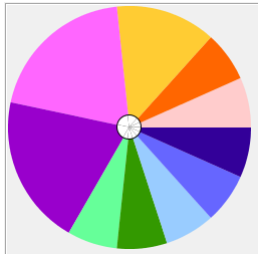
			Skin Sensitisation Literature							GeneTox Literature		
Chemical	Total PubMed Art Ct	CASRN	Proteins and genes	Cell	Immune Processes	Skin	Chemical mediators	Eruptions / SJS	Dermatitis contact / atopic	Mutagen	DNA Damage / Repair	Genetic Structures
(2E)-decenal	16	3913-81-3	0	0	0	0	0	0	0	0	0	1
2-ethylhexyl acrylate	27	103-11-7	3	0	1	4	0	0	14	0	0	0
2-hexenal	103	6728-26-3	9	0	0	3	7	0	0	8	56	7
2-hydroxyethyl acrylate	40	818-61-1	0	0	0	4	0	0	22	0	0	0
2-hydroxypropyl methacrylate	14	923-26-2	0	0	0	0	0	0	31	0	0	0
2-nonynoic acid methyl ester	2	111-80-8	0	0	0	0	0	0	8	0	0	0
4-vinylpyridine	139	100-43-6	0	0	0	0	0	0	23	4	4	1
benzoquinone	1434	106-51-4	73	16	21	5	44	0	28	57	150	32
benzyl cinnamate	14	103-41-3	0	0	0	4	0	0	0	4	0	0
benzylideneacetone	56	1896-62-4	0	0	0	4	0	0	0	0	5	0
Bisphenol A-Glycidyl Methacrylate	4008	1565-94-2	7	14	21	3	5	6	46	12	22	5
carvone	226	99-49-0	7	0	3	22	6	0	45	0	9	0
cinnamic aldehyde	717	104-55-2	54	25	7	42	20	12	198	14	57	4
cinnamonitrile	6	4360-47-8	1	0	0	0	0	0	0	0	0	0
citral	373	5392-40-5	11	2	2	31	6	0	82	15	13	5
diethyl maleate	709	141-05-9	149	4	1	0	27	0	0	4	34	8
Dimethyl Fumarate	355	624-49-7	11	9	0	0	6	0	0	0	0	0
ethyl acrylate	112	140-88-5	10	4	0	8	0	0	9	54	18	4
ethylene dimethacrylate	477	97-90-5	4	0	4	0	2	0	53	0	6	3
hexyl cinnamic aldehyde	17	101-86-0	0	14	0	4	0	0	16	0	0	0
methyl acrylate	125	96-33-3	2	0	0	4	0	0	10	12	4	0
methyl heptine carbonate	5	111-12-6	0	0	0	1	0	0	13	0	0	0
Methylmethacrylate	1727	80-62-6	16	6	5	8	2	4	98	4	13	15
n-butyl acrylate	121	141-32-2	4	0	0	2	0	0	0	16	14	0
perillaldehyde	29	713-95-1	6	0	0	0	3	0	0	0	0	0
Tropolone	592	533-75-5	14	9	10	13	8	0	6	0	10	3

Methodology for literature mining to inform read-across

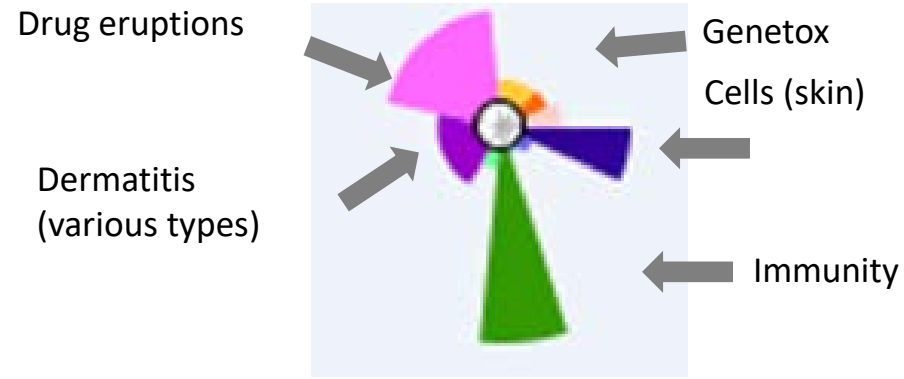
1. Gathering and extracting the literature information
2. Organizing by toxicity type
3. Condensing and strengthening signal → signature
4. Visualizing - use benzenes as example

4. Visualization: new concept- LitToxPI

Chemical	Proteins and genes	Cell	Immune Processes	Skin	Chemical mediators	Eruptions / SJS	Dermatitis contact / atopic	Mutagen	DNA Damage / Repair	Genetic Structures
Bisphenol A-Glycidyl Methacrylate	7	14	21	3	5	6	46	12	22	5



Name	Type	Weight	Color
GeneTox_mutagen	genetox	1	
GeneTox_structures	genetox	1	
GeneTox_DNAdamage	genetox	2	
Skin_eruption	skin	3	
Skin_dermatitis	skin	3	
Skin_mediators	skin	1	
Skin_immunity	skin	1	
Skin_protGene	skin	1	
Skin_organ	skin	1	
Skin_cell	skin	1	

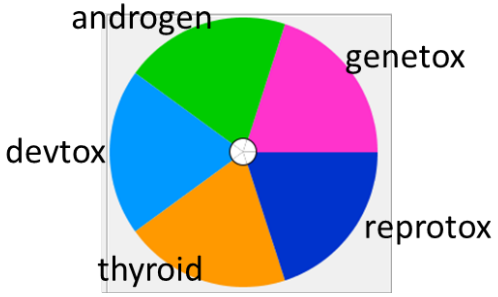


Bisphenol A-glycidyl Methacrylate

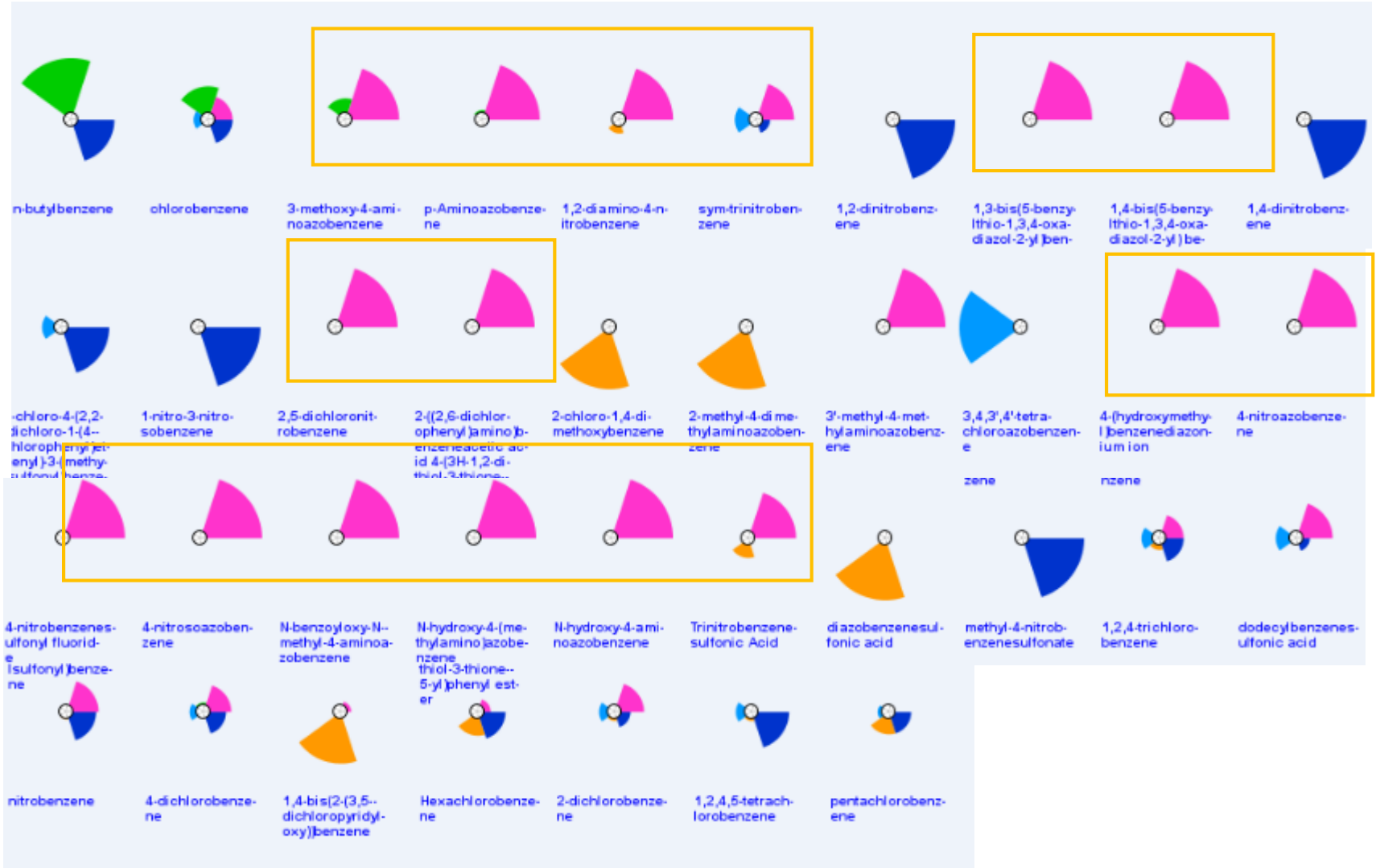
Reif DM, Martin MT, Tan SW, et al. *Environmental Health Perspectives*. 2010;118(12):1714-1720

4. Visualization – possible skin sensitizers

Just for fun: LitToxPIs for benzenes



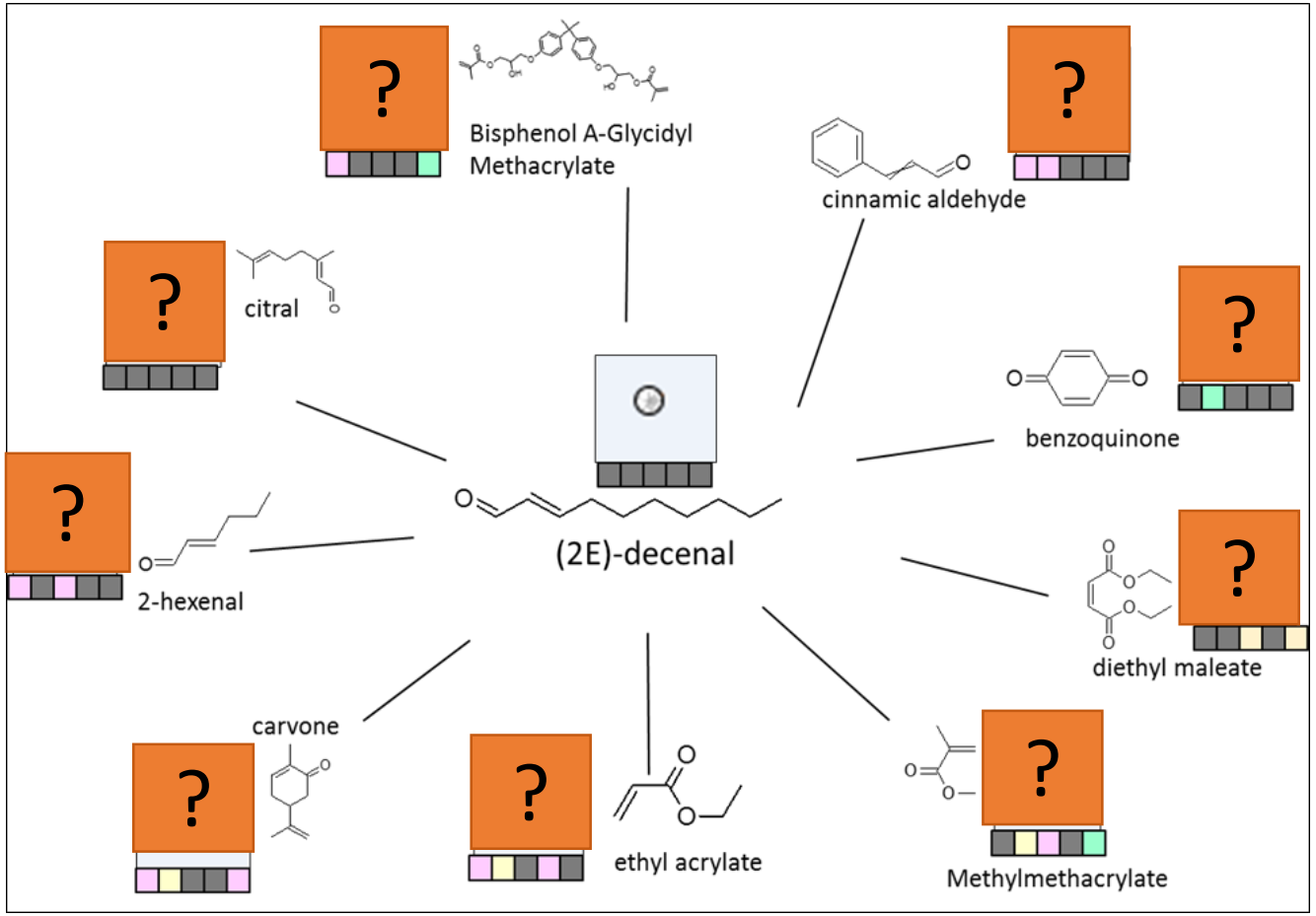
- We can use any literature corpus and boil it down to LitToxPIs.

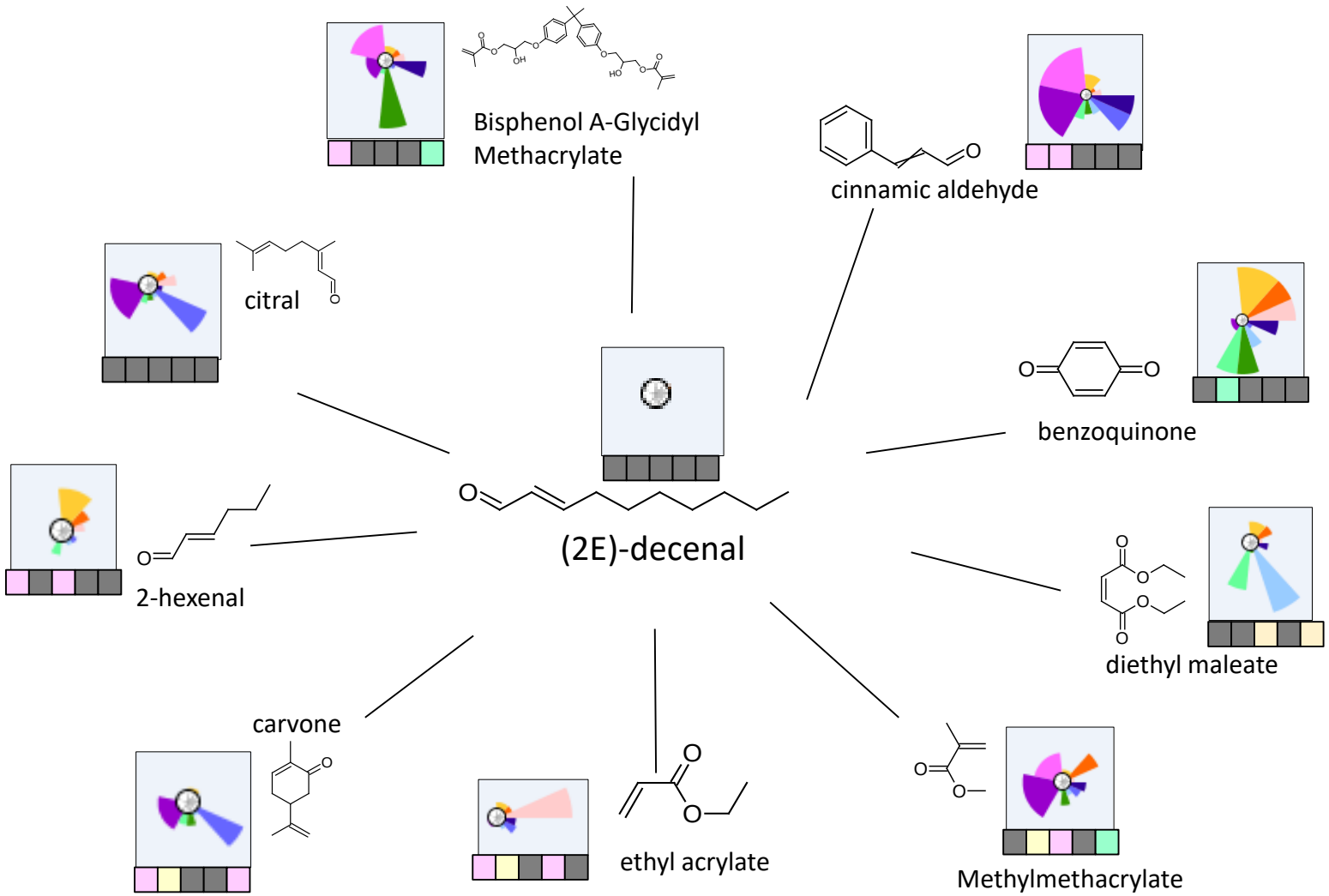


Goal

?

- Construct a literature signature or fingerprint or descriptors
- Visualize




Bisphenol A-Glycidyl Methacrylate
cinnamic aldehyde
citral
benzoquinone
2-hexenal
(2E)-decenal
diethyl maleate
carvone
ethyl acrylate
Methylmethacrylate

Next steps

- Visually integrate the LitToxPIs into Read-across application.
- Continue to refine and enhance the gathering, organizing, condensing, weighting, and normalizing steps.
- Apply this technology to other read-across areas.

Conclusion

- The laborious process of searching and evaluating the literature for chemical activity can be streamlined.
- Literature mining algorithms and processes can condense complex unstructured literature into quantitative biological descriptors or signatures.
- These literature signatures can be visualized to inform and enhance chemical read-across.

Thank you!

Questions?

Email: Baker.nancy@epa.gov