Supplemental Materials

QSAR models of human data can enrich or replace LLNA testing for human skin sensitization

Vinicius M. Alves,[a,b] Stephen J. Capuzzi,[a] Eugene Muratov,[a,c] Rodolpho C. Braga,[b] Thomas Thornton,[a] Denis Fourches,[d] Judy Strickland,[e] Nicole Kleinstreuer,[f] Carolina H. Andrade,[b] and Alexander Tropsha.[a,]*

[a] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.

[b] Laboratory for Molecular Modeling and Design, Faculty of Pharmacy, Federal University of Goias, Goiania, GO, 74605-170, Brazil.

[c] Department of Chemical Technology, Odessa National Polytechnic University, Odessa, 65000, Ukraine.

[d] Department of Chemistry, Bioinformatics Research Center, North Carolina State University, Raleigh, NC, 27695, USA.

[e] Integrated Laboratory Systems, Inc., P.O. Box 13501, Research Triangle Park, NC, 27709, USA.

[f] National Institutes of Environmental Health Sciences, Research Triangle Park, NC, 27709, USA.

Corresponding Author

* Address for correspondence: 100K Beard Hall, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA; Telephone: (919) 966-2955; FAX: (919) 966-0204; E-mail: alex_tropsha@unc.edu.

Table of Contents

Cluster analysis

In Cluster 1, there are four sensitizers and four non-sensitizers. All sensitizers were predicted correctly by QSAR. Similar to outliers in previous sections, benzyl alcohol (sensitizer) has a high DSA_{05} of 48.67 μg/cm^2 (45.06 mol/m^2) and a NOEL of 5,906 μg/cm^2 (5.47 mol/m^2). The very low skin sensitization potency may be the reason for the wrong prediction by our QSAR model. Propylidene phthalate was also mispredicted. The other two sensitizers (2-mercaptobenzothiazole and benzisothiazole) were correctly predicted. Human non-sensitizers benzoic acid, salicylic acid, methyl salicylate, and resorcinol were correctly predicted by QSAR. Of those, the only compound correctly predicted by LLNA was benzoic acid. In a recent study [1], resorcinol is labeled as a human sensitizer, but no NOEL is available. It is used at high levels in hair dyes and skin preparations, but is not considered to be dangerous, since it has a low frequency of human sensitization.

Pyridine, one of the outliers detected in the previous section, is in the cluster of aromatic amines (Cluster 2). Within this cluster, sulfanilamide (human sensitizer) was the only compound mispredicted by the LLNA. The two non-sensitizers (p-aminobenzoic acid and sulfanilic acid) were mispredicted by QSAR model. Sulfanilamide has a similar structure to sulfanilic acid, with Tanimoto coefficient equal to 0.83. The substitution of a sulfo group or a carboxyl in the *para*-position of aniline decreases sensitization potency, while another amine preserves it.

Phenyl benzoate was also identified as an outlier, due to the high DSA_{05}, which shows that this compound is safe at low concentrations in most of the tested population. All of the non-sensitizers in this cluster (Cluster 3) were mispredicted by the LLNA and correctly predicted by QSAR: benzyl cinnamate, benzyl benzoate, benzyl salicylate, and hexyl salicylate. The sensitizers

benzoyl peroxide and phenyl benzoate were mispredicted and correctly predicted, respectively. Both compounds were correctly predicted by the LLNA.

In Cluster 8, α-amylcinnamyl alcohol is the only human sensitizer. This compound was predicted as a non-sensitizer by LLNA and QSAR. All other compounds (α-amylcinnamic aldehyde, hexyl cinnamic aldehyde, and hexyl salicylate) were human non-sensitizers and were mispredicted by LLNA correctly predicted by QSAR. In a recent publication [1], hexyl salicylate and α-amylcinnamic aldehyde are labeled as human sensitizers. The first has a NOELs as 35,433 (15.96 mol/m^2), which indicates low sensitization rates in the tested populations at relatively high doses, while the second is in fact a strong sensitizer, with DSA_{05} of 23.622 (μg/cm^2) 0.01 (mol/m^2).

In the Cluster 11, lilial was the only compound correctly predicted by QSAR models. This compound is labeled as a sensitizer; however, it has a high DSA_{05}, since at the LOEL of 29.53 μg/cm^2 it was positive in only one out of 225 people. Cyclamen aldehyde (non-sensitizer) was mispredicted, while bourgeonal (sensitizer) and majantal (non-sensitizer) were predicted as sensitizers. All the compounds in this cluster, except cyclamen aldehyde, were correctly predicted by LLNA.

Table S1. Number of records, assay outcome, and QSAR prediction for the 62 substances with multiple records present in the human skin sensitization dataset (Dataset A).

#	Compound name	CASRN	No. of records	No. of sensitizers	No. of non-sensitizers	QSAR
1	(Chloro)methylisothia-zolinone (Kathon)	26172-55-4	14	4	10	Not predicted
2	Hydroxycitronellal	107-75-5	12	8	4	Sensitizer
3	Cinnamyl alcohol	104-54-1	10	6	4	Sensitizer
4	Cinnamic aldehyde	104-55-2	8	5	3	Sensitizer
5	Citral	5392-40-5	7	5	2	Non-sensitizer
6	Streptomycin	3810-74-0	6	6	0	Sensitizer
7	Phenylacetaldehyde	122-78-1	6	5	1	Non-sensitizer
8	Geraniol	106-24-1	6	2	4	Sensitizer
9	Benzoyl peroxide	94-36-0	5	5	0	Non-sensitizer
10	Neomycin sulfate	1405-10-3	5	5	0	Sensitizer
11	Penicillin G	61-33-6	5	4	1	Sensitizer
12	Benzocaine	94-09-7	5	4	1	Sensitizer
13	Ethyl acrylate	140-88-5	5	3	2	Sensitizer
14	Methylisothiazolinone	2682-20-4	5	2	3	Sensitizer
15	Coumarin	91-64-5	5	2	3	Sensitizer
16	Eugenol	97-53-0	5	1	4	Sensitizer
17	*dl*-Citronellol	26489-01-0	5	1	4	Non-sensitizer
18	4-Phenylenediamine	106-50-3	4	4	0	Sensitizer
19	Methylhexanedione	13706-86-0	4	4	0	Non-sensitizer
20	Tetrachlorosalicylanilide	1154-59-2	4	4	0	Sensitizer
21	Isoeugenol	97-54-1	4	2	2	Non-sensitizer
22	Lilial	80-54-6	4	1	3	Sensitizer
23	Cinnamyl nitrile	4360-47-8	4	1	3	Sensitizer
24	Potassium dichromate	7778-50-9	3	3	0	Not predicted
25	Thioglycerol	96-27-5	3	3	0	Non-sensitizer
26	Nickel (II) salts	7718-54-9; 7786-81-4	3	3	0	Not predicted
27	Formaldehyde	50-00-0	3	2	1	Sensitizer
28	Tetramethylthiruam-	137-26-8	3	2	1	Sensitizer

	disulfide					
29	Propylidene phthalate	17369-59-4	3	1	2	Non-sensitizer
30	α-amylcinnamyl alcohol	101-85-9	3	1	2	Non-sensitizer
31	Benzyl alcohol	100-51-6	3	1	2	Non-sensitizer
32	Oakmoss	68917-10-2	3	1	2	Not predicted
33	Treemoss	68648-41-9	3	1	2	Not predicted
34	Linalool	78-70-6	3	0	3	Non-sensitizer
35	Clove oil (bud, leaf, stem)	8000-34-8	3	0	3	Not predicted
36	2,4-Dinitrochlorobenzene	97-00-7	2	2	0	Sensitizer
37	2-Mercaptobenzothiazole	149-30-4	2	2	0	Sensitizer
38	Benzylidene acetone	122-57-6	2	2	0	Non-sensitizer
39	Diethylmaleate	141-05-9	2	2	0	Sensitizer
40	Dihydrocoumarin	119-84-6	2	2	0	Non-sensitizer
41	Cobalt (II) salts	7646-79-9; 10124-43-3	2	2	0	Not predicted
42	Mercuric (II) chloride	7487-94-7	2	2	0	Not predicted
43	Benzoisothiazolione	2634-33-5	2	1	1	Sensitizer
44	Farnesol	4602-84-0	2	1	1	Sensitizer
45	Glutaraldehyde	111-30-8	2	1	1	Sensitizer
46	Imidazolidinyl urea	39236-46-9	2	1	1	Sensitizer
47	Isocyclogeraniol	68527-77-5	2	1	1	Non-sensitizer
48	Methyl 2-nonynoate	111-80-8	2	1	1	Sensitizer
49	Methyl 2-octynoate	111-12-6	2	1	1	Sensitizer
50	*p*-methylhydrocinnamic aldehyde	5406-12-2	2	1	1	Sensitizer
51	*t*-2-Hexenal	6728-26-3	2	1	1	Sensitizer
52	Ylang Ylang	8006-81-3; 68606-83-7; 83863-30-3	2	1	1	Not predicted
53	α-methyl cinnamic aldehyde	101-39-3	2	0	2	Non-sensitizer
54	Benzyl cinnamate	103-41-3	2	0	2	Non-sensitizer

55	Benzyl salicylate	118-58-1	2	0	2	Non-sensitizer
56	Benzylbenzoate	120-51-4	2	0	2	Non-sensitizer
57	*d*-Limonene	5989-27-5	2	0	2	Sensitizer
58	Isocyclocitral	1335-66-6	2	0	2	Non-sensitizer
59	Lyral	31906-04-4	2	0	2	Sensitizer
60	Propylene glycol	57-55-6	2	0	2	Sensitizer
61	β-damascone	23726-91-2	2	0	2	Non-sensitizer
62	Benzalkonium chloride	8001-54-5	2	0	2	Not predicted

Table S2. Number of records and the data outcome for the 19 substances with different annotations between the records present in the murine skin sensitization dataset (Dataset B).

#	Compound name	CASRN	No. of records	No. of sensitizers	No. of non-sensitizers
1	Hexyl cinnamic aldehyde	101-86-0	44	42	2
2	Eugenol	97-53-0	31	30	1
3	Benzocaine	94-09-7	24	7	17
4	Nickel (II) salts	7718-54-9; 7786-81-4	16	5	11
5	Methyl salicylate	119-36-8	14	2	12
6	Sodium lauryl sulfate	151-21-3	11	10	1
7	Aniline	62-53-3	11	6	5
8	Potassium dichromate	7778-50-9	10	9	1
9	2-Mercaptobenzothiazole	149-30-4	8	6	2
10	Geraniol	106-24-1	7	6	1
11	Tetramethylthiuram disulfide	137-26-8	7	6	1
12	Streptomycin	3810-74-0	6	2	4
13	Coumarin	91-64-5	4	2	2
14	Ethyl acrylate	140-88-5	3	2	1
15	Zinc sulfate	7733-02-0	3	2	1
16	Resorcinol	108-46-3	3	1	2
17	Benzyl benzoate	120-51-4	2	1	1
18	Ethylenediamine	107-15-3	2	1	1
19	Salicylic acid	69-72-7	2	1	1

Table S3. List of chemical compounds predicted as sensitizers and confirmed in the literature.

Compound name	CASRN	Function	Reference
Styrene	100-42-5	Perfuming	[2]
Benzonitrile	100-47-0	Perfuming	[3]
p-aminodiphenylamine	101-54-2	Hair dyeing	[4]
Triethylene glycol dimethacrylate	109-16-0	Nail conditioning	[5,6]
Ethanolamine	141-43-5	Buffering	[7]
Diallyl disulfide	2179-57-9	Perfuming	[8]
Diethylene glycol dimethacrylate	2358-84-1	Nail conditioning, film forming	[5]
Laureth-9, polidocanol	3055-99-0	Emulsifying	[9]
Glyceryl monothioglycolate	30618-84-9	Hair waivening or straightening	[10]
Iodopropynyl butylcarbamate	55406-53-6	Preservative	[11]
Chlorhexidine	55-56-1	Antimicrobial, oral care, preservative	[12]
C.I. Solvent Red 3	6535-42-8	Colorant	[13]
Ethyl cyanoacrylate	7085-85-0	Film forming	[6,14]
Trichloroethane	71-55-6	Solvent	[15]
Chloroacetamide	79-07-2	Preservative	[16]
Methyl methacrylate	80-62-6	Anticaking, opacifying	[6,17]
Ethyleneglycol dimethacrylate	97-90-5	Nail conditioning	[6,18]

REFERENCES

1 D. A. Basketter, N. Alépée, T. Ashikaga, J. Barroso, N. Gilmour, C. Goebel, J. Hibatallah, S. Hoffmann, P. Kern, S. Martinozzi-Teissier, G. Maxwell, K. Reisinger, H. Sakaguchi, A. Schepky, M. Tailhardat and M. Templier, *Dermat. contact, atopic, Occup. drug*, 2014, **25**, 11–21.

2 S. Sjöborg, S. Fregert and L. Trulsson, *Contact Dermatitis*, 1984, **10**, 94–96.

3 L.-F. Li, S. A. Sujan and Q. X. Li, *Contact Dermatitis*, 2004, **50**, 377–378.

4 C. Skudlik, E. Meyer, H. Allmers, E. Domagalski and S. M. John, *Hautarzt.*, 2011, **62**, 765–769.

5 L. Perale, S. De Marchi, E. Cecchin and L. A. Sechi, *Contact Dermatitis*, 2005, **53**, 181–182.

6 S. Shanmugam and M. Wilkinson, *Contact Dermatitis*, 2012, **67**, 309–310.

7 M. Bhushan, N. M. Craven and M. H. Beck, *Contact Dermatitis*, 1998, **39**, 321.

8 M. T. Bordel-Gómez and A. Miranda-Romero, *Contact Dermatitis*, 2008, **59**, 125–126.

9 R. Gallo, M. Basso, S. Voltolini and M. Guarrera, *Contact Dermatitis*, 2001, **45**, 356–357.

10 T. Leino, T. Estlander and L. Kanerva, *Contact Dermatitis*, 1998, **38**, 166–167.

11 J. P. Thyssen, K. Engkilde, M. D. Lundov, B. C. Carlsen, T. Menné and J. D. Johansen, *Contact Dermatitis*, 2010, **62**, 102–108.

12 R. Toholka and R. Nixon, *Australas. J. Dermatol.*, 2013, **54**, 303–306.

13 F. Wantke, M. Götz and R. Jarisch, *Contact Dermatitis*, 1992, **27**, 346–347.

14 R. R. Tomb, J. P. Lepoittevin, F. Durepaire and E. Grosshans, *Contact Dermatitis*, 1993, **28**, 206–208.

15 J. Mallon, M. T. Chu and H. I. Maibach, *Contact Dermatitis*, 2001, **45**, 107.

16 J. M. Taran and T. A. Delaney, *Australas. J. Dermatol.*, 1997, **38**, 95–96.

17 L. Strazzula, S. Das, V. E. Nambudiri and D. Kroshinsky, *JAMA dermatology*, 2014, **150**, 784–785.

18 K. Aalto-Korte, K. Alanko, O. Kuuliala and R. Jolanki, *Contact Dermatitis*, 2008, **58**, 340–346.