Appendix 10

Reactivity Domain and Physicochemical Property of EpiSensA Dataset

Chemical reactivity domain of the tested chemicals

The chemical reactivity domains covered by the chemicals in the EpiSensA dataset were determined with the OECD QSAR Toolbox 4.4 using the profiler "Protein binding alerts for skin sensitisation by OASIS". In addition, the results were compared to DASS database used to the 2o3 (reactive domains of the overall DASS database were not indicated at the Supporting document to the OECD guideline 497). The 136 tested chemicals with reference classifications cover the following reactivity domains:

Table 1. List of chemical reactivity domains found for the chemicals tested using the EpiSensA and 2o3 DASS.

Chemical reactivity domain	EpiSensA	2o3 DASS[a]
	n=136	n=168
Acylation	10*	14
Michael addition	13*	29
Nucleophilic addition	3	1
SN2	12*	15
SNAr	4*	3
Schiff base formation	11	28
SNVinyl	0	1
Quinone oxide structure	2*	2
Nitroquinone, naphthoquinone(s)/imines	2*	2
No alert	36	65
No alert (alert at autoxidation/metabolism simulator)	49	15
Total	142	177

[a]: Supporting document to the OECD guideline 497 on defined approach for skin sensitisation, 2021.

*: Benzyl salicylate was categorised as Acylation and SN2; Fluorescein isothiocyanate was categorised as Acylation, Quinoide oxime structure, and Nitroquinones, naphthoquinone(s)/imines; Brandowski's base was categorised as Michael Addition, QuiQuinoide oxime structure, and Nitroquinones, naphthoquinone(s)/imine; Chlorpyrifos was categorised as SN2 and SNAr. Therefore, 136 chemicals were classified in 142 categories as there are 6 chemicals classified in more than one category which were counted in each category

As a result of reactivity domain analysis, the largest group correspond to the “No alert”, which are the chemicals for which the OECD QSAR Toolbox profiler could not find any alert in the parent structure (Table 1). Of these 85 chemicals, 49 chemicals show alerts if the autoxidation or skin metabolism simulators are used. The second largest groups of domains are the Michael addition (n=13), Schiff base formation (n=11), SN2 (n=12) and Acylation (n=10). The rest of the domains contain 3 or less chemicals each. When comparing to 2o3 DASS dataset described in the Supporting document, the chemical reactivity domain covered by the EpiSensA dataset was comparable to that of 2o3 DASS dataset except for the chemicals which was “No alert” in the parent structure but alert if the two simulators were used (49 chemicals for EpiSensA and 15 chemicals for 2o3 DASS) (Table 1). This fact indicates that the EpiSensA dataset contains more pre/pro-haptens than that of 2o3 DASS dataset.

Physicochemical properties of the tested chemicals

The physicochemical property values used to describe the chemical space covered by the tested chemicals were obtained from the OPERA model which has been used at Supporting document to the OECD guideline 497.

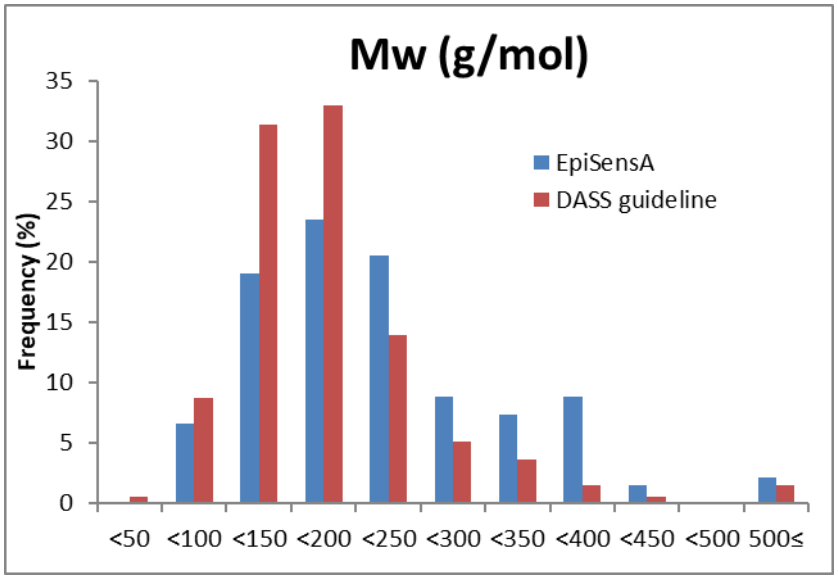
The summary of the physicochemical property ranges that describe the chemical space of the chemicals tested by EpiSensA are shown in Table 2, and the distributions were described as histograms in Figure 1. Note that 132 out of the 136 chemicals were used to determine the ranges, since fluorescein isothiocyanate, cobalt chloride, dinocap, and xylene were excluded as the physicochemical properties did not be calculated by OPERA model. In addition, the physicochemical property ranges were compared to those of DASS database containing 194 chemicals (kanamycin and kathon CG were excluded because the physicochemical properties were not available at Annex2 of the Supporting document to the OECD guideline 497).The range of six physicochemical properties were obviously similar between two datasets (Table 2). In addition, regarding the ranges covered by the two datasets, the distributions of six properties were also similar except for log WS (the range of EpiSensA dataset is broader than that of DASS dataset and cover chemicals with low water solubility) (Figure 1). Furthermore, regarding the peaks of the distributions, there are a little difference between two datasets at log P (the peak of EpiSensA dataset is $3 \leq \log P < 4$ but that of DASS dataset is $1 \leq \log P < 2$) and log WS (the peak of EpiSensA dataset is $-4 \leq \log WS < -3$ but that of DASS dataset is $-3 \leq \log WS < -2$). These facts indicate that EpiSensA dataset contains more lipophilic chemicals than that of DASS dataset.

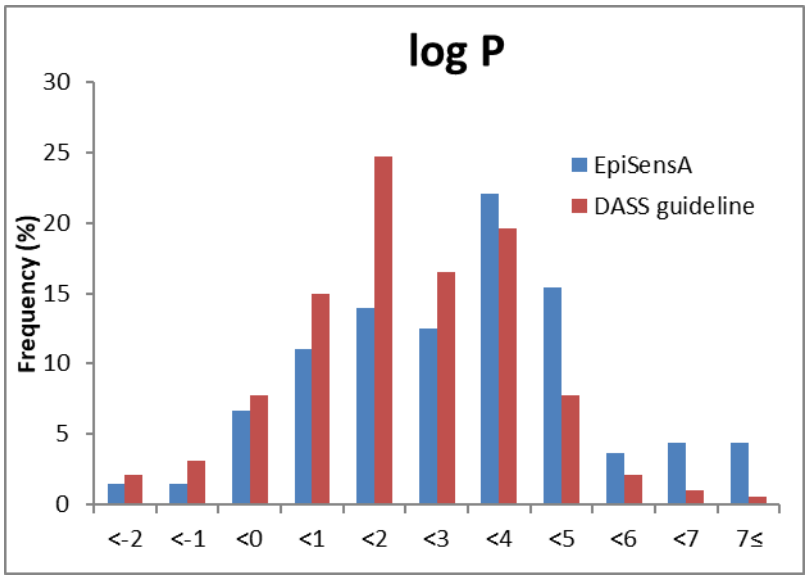
Table 2. Summary of the physicochemical property ranges that describe the chemical space of the chemicals tested in the EpiSensA and DASS.

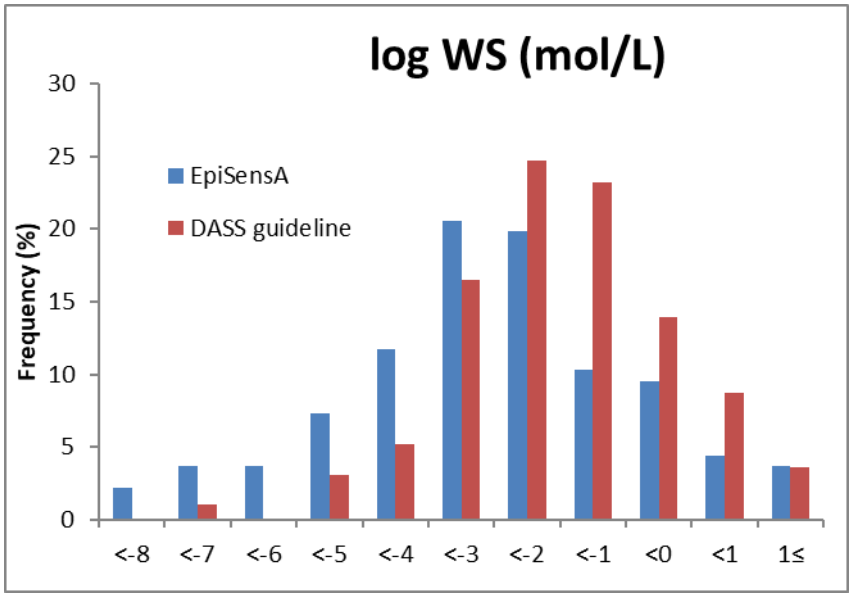
	EpiSensA		DASS database	
	n=132		n=194	
	Min	Max	Min	Max
MW (g/mol)	58.04	719.76	30.0	614.3
log P	-2.82	9.45	-2.4	9.4
log WS (mol/L)	-8.23	1.24	-7.6	1.2
MP (ºC)	-112.40	252.67	-122.5	252.7
BP (ºC)	50.40	512.00	-19.1*	445.3
log VP (Pa)	-10.27	2.41	-10.3	2.9

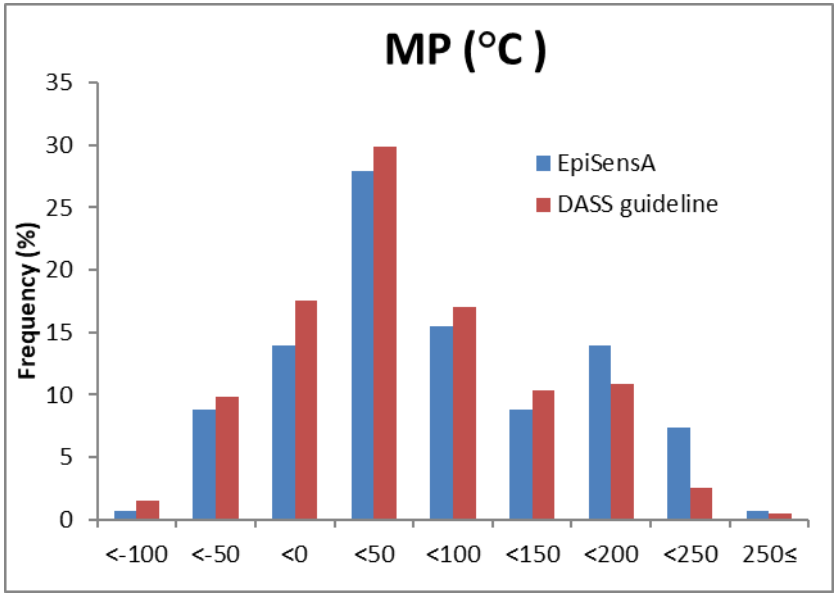
*: -19.1 ºC is the boiling point of formaldehyde, and the second smallest one is 31.6 ºC of 1,1-dichloroethene.

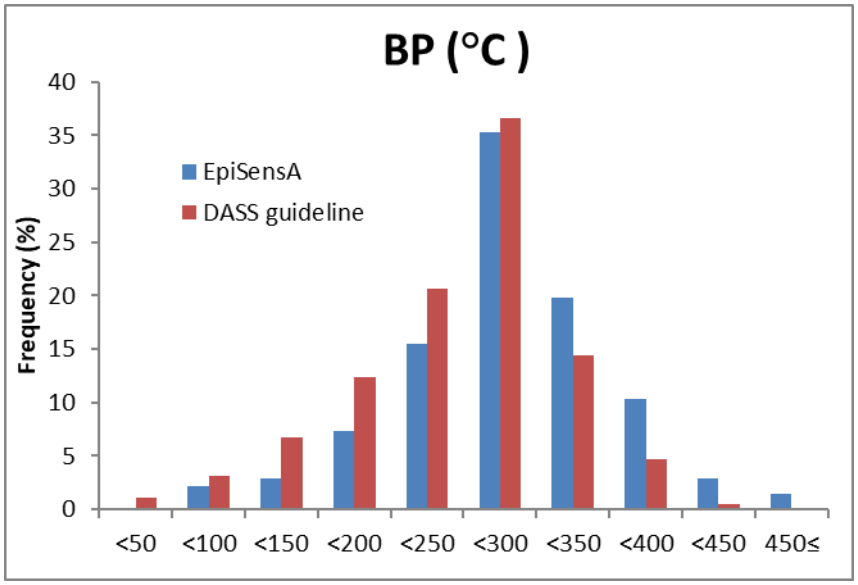
Figure 1. Distribution of the physicochemical properties of the EpiSensA dataset and DASS database.

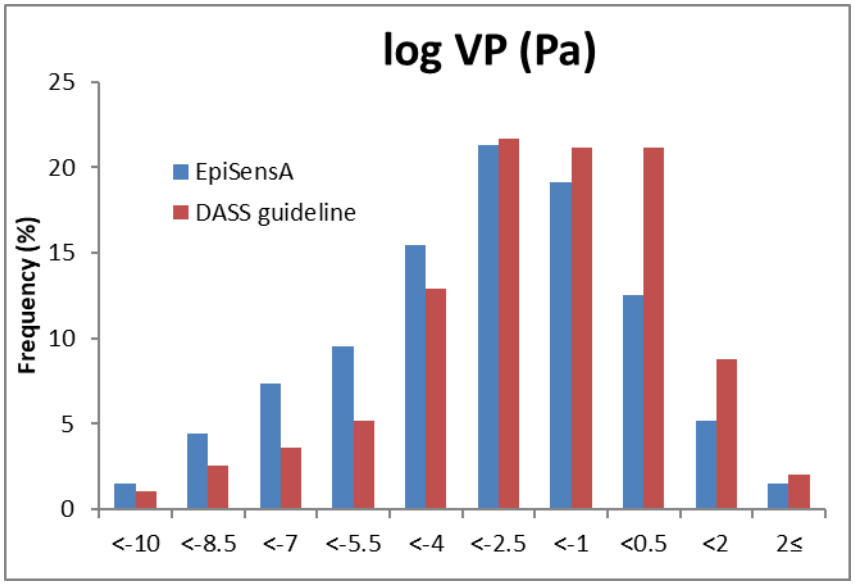





log WS (mol/L)
Frequency (%)
EpiSensA
DASS guideline
<-8
<-7
<-6
<-5
<-4
<-3
<-2
<-1
<0
<1
1≤


MP (°C)
Frequency (%)
EpiSensA
DASS guideline
<-100
<-50
<0
<50
<100
<150
<200
<250
250≤


BP (°C)
Frequency (%)
EpiSensA
DASS guideline
<50
<100
<150
<200
<250
<300
<350
<400
<450
450≤


log VP (Pa)
Frequency (%)
EpiSensA
DASS guideline
<-10
<-8.5
<-7
<-5.5
<-4
<-2.5
<-1
<0.5
<2
2≤