Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2021)12

Unclassified English - Or. English

8 July 2021

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Validation Study Report pf the Kinetic Direct Peptide Reactivity Assay (kDPRA) in Test Guideline 442C

Series on Testing and Assessment, No. 337

JT03479325

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

SERIES ON TESTING AND ASSESSMENT

NO. 337

Validation Study Report of the Kinetic Direct Peptide Reactivity Assay (kDPRA) in Test Guideline 442C

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2021

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 36 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Testing and Assessment link

For this and many other Environment,
Health and Safety publications, consult the OECD's
World Wide Web site (www.oecd.org/chemicalsafety/)

or contact:

OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

© OECD 2021
Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France
OECD Environment, Health and Safety Publications

FOREWORD

This document contains the Validation report of the kinetic Direct Peptide Reactivity Assay (kDRPA). It supports the development of this test method for inclusion in Test Guideline 442C for in chemico skin sensitisation assays, addressing the Adverse Outcome Pathway Key Event on covalent binding to proteins.

The kDPRA, proposed by Germany and Switzerland, is a modification of the DPRA, included in **TG 442C** in 2015. The kDPRA uses kinetic rates of cysteine- peptide depletion to distinguish between two levels of skin sensitization potency, i.e. to discriminate between GHS sub-categories 1A and 1B.

The Peer review report and the Validation report of the kDPRA were made available in July and December 2020, as supporting documents during the two submissions for comments of the updated TG 442C to the Working Party of the National Coordinators of the Test Guidelines Programme (WNT).

The WNT endorsed the peer review report at its 33rd meeting in April 2021. The WNT also endorsed the Peer review report of the kDPRA. This report is published under the responsibility of the Chemicals and Biotechnology Committee.

Kinetic Direct Peptide Reactivity Assay (kDPRA): Validation Study Report

Lead Laboratories and Study Management:

BASF SE

Britta Wareing
Phone: +49 621 60-58107, Email: britta.wareing@basf.com
Dr. Susanne Kolle
Phone: +49 621 60-56731, Email: susanne.kolle@basf.com,
Experimental Toxicology and Ecology
Postal Address: BASF SE, Carl-Bosch-Strasse 38, 67056 Ludwigshafen am Rhein, Germany

Responsibilities:
Protocol authorship
Coordination between laboratories
Organisation of test chemical procurement

Givaudan Schweiz AG

Dr. Andreas Natsch
Phone: +41 44 824 21 05, Email: andreas.natsch@givaudan.com;
Tina Haupt
Phone: +41 44 824 25 15, Email: tina.haupt@givaudan.com
Postal Address: Department of in vitro molecular screening, Givaudan Schweiz AG, Kemptpark 50, CH-8310 Kemptthal, Switzerland

Responsibilities:
Test chemical selection
Optimisation of rate constant calculation in evaluation sheet and cut-off refinement
Biostatistics

Naïve Test Laboratories:

Institute for In Vitro Sciences (IIVS), 30 West Watkins Mill Road Suite 100, Gaithersburg, MD 20878, USA: Contact persons: Erin H Hill (ehill@iivs.org), Hans Raabe (hraabe@iivs.org), Rishil Kathawala (RKathawala@iivs.org)

Procter & Gamble, Mason Business Center, DS3-812, 8700 Mason-Montgomery Road, Mason, OH, 45040, USA: Contact persons: Cindy Ryan (ryan.ca@pg.com) and Petra Kern (kern.ps@pg.com)

Charles River Laboratories Den Bosch BV, Hambakenwetering 7, 5231 DD 's-Hertogenbosch, The Netherlands: Contact persons: Walter Westerink (Walter.Westerink@crl.com), Sjoerd Verkaart (Sjoerd.Verkaart@crl.com)

National Institute of Public Health, Srobarova 48, 10042 Praha, Czech Republic:

Contact persons: Marian Rucki (marian.rucki@szu.cz), Dvořáková Markéta; (marketa.dvorakova@szu.cz), Kejlová Kristina (kristina.kejlova@szu.cz)

L'Oréal Research & Innovation, Bâtiment 25-101S, 1 Avenue Eugène Schueller, 93600 Aulnay sous Bois, France:
Contact persons: Nathalie Alepée (nathalie.alepee@rd.loreal.com) Fleur Tourneix (fleur.tourneix@rd.loreal.com)

Test chemical procurement and blinding

BioTeSys GmbH, Schelztorstrasse 54-56, 73728 Esslingen, Germany
Contact persons: Dr. Dirk Dressler (d.dressler@biotesys.de)

Following completion of the validation study, conducted between spring 2018 and summer 2019, the kDPRA test method and its validation study underwent an independent peer-review between August 2019 and February 2020 by an international panel (1).

Appendices 5 and 6 to this validation report have been developed and added to the validation report after completion of the peer review.

(1) OECD (2021), Independent Peer-Review Panel Report on the scientific validity of the kinetic Direct Peptide Reactivity Assay (kDPRA). Series on testing and Assessment n°338. Organisation for Economic Co-operation and Development, Paris.

1. Contents

1. Contents 3
2. Abbreviations 6
3. Introduction 9
3.1. Background 9
3.2. Principle of the test 10
3.3. Log k_{max} Calculation Example DNCB 10
4. Goals set for the validation study 13
5. Work packages 13
5.1. WP 3.1.: Preparation phase / protocol refinement 15
5.2. WP3.2. Transfer phase of the refined protocol in experienced labs 16
5.2.1. Procedure 16
5.2.2. Results for the positive chemicals (Set A) 17
5.2.3. Results for the positive control 18
5.2.4. Results for the chemicals of Set B 19
5.2.5. Conclusions testing in experienced labs 22
5.3. WP 3.3. Transfer phase to naïve labs (phase I) 22
5.3.1. Goals and set-up 22
5.3.2. Setting up the test in different labs 23
5.3.3. Reproducibility of log k_{max} 24
5.3.4. Reproducibility of predicted GHS categories 28
5.3.5. Results for positive control 28
5.3.6. Discussion on transfer phase to naïve labs 29
5.4. WP 3.4 Blind-coded Testing – Phase II 31
5.4.1. Goals and set-up 31
5.4.2. Rationale for test chemical selection for Phase II 32
5.4.3. Overview of the different data evaluations performed: 34
5.4.4. Intra-laboratory reproducibility of log k_{max} 36
5.4.5. Inter-laboratory reproducibility of log k_{max} 39
5.4.6. Comparison of intra-laboratory and inter-laboratory reproducibility of log k_{max} 41
5.4.7. Prediction of GHS Categories 43
5.4.8. Intra-laboratory reproducibility of predicted GHS categories 43
5.4.9. Inter-laboratory reproducibility of predicted GHS categories 44

5.4.10. Reproducibility of positive control ... 48
5.4.11. Interference from autofluorescence and potential fluorescence quenching ... 50
5.4.12. Intra-laboratory reproducibility of 24 h / 5 mM depletion value ... 51
5.4.13. Inter-laboratory reproducibility of 24 h / 5 mM depletion value ... 51
5.4.14. Comparison of kDPRA (24 h / 5 mM depletion value) to the classical DPRA... 53
5.4.15. Congruency of solvent-decision ... 55
5.4.16. Intra-laboratory congruency of solvent-decision ... 55
5.4.17. Inter-laboratory congruency of solvent-decision ... 55
5.4.18. Conclusions on intra-and inter-laboratory reproducibility ... 58
5.5. WP 3.5. Evaluation of the predictive capacity for the dataset of Phase I and Phase II 58
5.6. WP 3.6. Building the database and testing the prediction cut-off ... 61
5.6.1. Deriving the optimal cut-off to discriminate GHS Cat 1A ... 62
6. General discussion ... 70
6.1. Test definition ... 70
6.2. Transferability ... 70
6.3. Intra-laboratory Reproducibility ... 71
6.4. Inter-laboratory Reproducibility ... 71
6.5. Predictive capacity to identify GHS Cat1A ... 71
6.6. Applicability domain ... 72
6.6.1. Technical limitations ... 72
6.6.2. Predictive limitations ... 73
6.7. Relationship to DPRA 442C and advantages over the classical DPRA ... 74
6.8. Potential use in tiered testing strategy ... 74
6.9. Potential use of numeric log k_{max} in defined approaches ... 75
6.10. Introduction as a test method into OECD 442C ... 76
6.11. Potential improvements ... 76
7. References ... 76
Appendix 1: Detailed result tables of the blind-coded inter-laboratory testing (Phase II) . 77
Appendix 2: Result interpretation in case of fluorescence quenching and potential control experiments ... 87
Non-linear peptide depletion observed for other chemicals when establishing the database ... 89
Appendix 3. Extended dataset to evaluate predictivity and to determine optimal GHS 1A vs. GHS 1B/NC cut-off ... 92

Appendix 4: Statistical analysis for the contribution of continuous parameters from validated in vitro assays to predict LLNA EC3: kDPRA vs. existing validated in vitro tests (KeratinoSens and h-CLAT)... 98

Appendix 5: Follow-up to the public consultation on the validation report (circulated for comments on 7 July 2020). 106

Appendix 6: The predictivity of the kDPRA to identify 1A sensitizers vs. the consolidated OECD database 110

2. Abbreviations

A	=	alanine
Ac	=	acetyl group
ACN	=	acetonitrile
approx..	=	approximately
BA	=	Benzylidene acetone
BI	=	Benzoisiothiazolinone
BOU	=	Bourgeonal
C	=	cysteine
CA	=	cinnamic aldehyde
CAR	=	Carvone
CB	=	Chlorobenzene
CLP	=	Classification, Labelling and Packaging of substances and mixtures
CMI	=	(Chloro)methylisothiazolinone
conc.	=	concentration
COOH	=	carboxyl group
CV	=	coefficient of variation
d	=	day
dp	=	depletion
DAM	=	δ-Damascone
DEM	=	Diethylmaleate
DHC	=	Dihydrocoumarin
DPRA	=	Direct Peptide Reactivity Assay
EGDMA	=	ethylene glycol dimethacrylate
F	=	phenylalanine
GHS	=	Globally Harmonized System of Classification, Labelling and Packaging of Chemicals
GLY	=	Glyoxal

h	=	hour
HC	=	Hydroxycitronellal
HEX	=	trans-2-hexenal
HPLC	=	high performance liquid chromatography
I	=	Isoeugenol
IU	=	Imidazolidinyl urea
k	=	kinetic rate constant
kDPRA	=	kinetic direct peptide reactivity assay
l	=	liquid
LOEL	=	lowest observed effect level
M	=	mol / liter
MAP	=	4-methoxy-acetophenone
mBrB	=	monobromobimane
mg	=	milligram
µg	=	microgram
MHD	=	Methylhexanedione
MI	=	Methylisothiazolinone
MIE	=	Molecular Initiating Event
min	=	minute(s)
µL	=	microliter
mL	=	milliliter
mM	=	millimol / liter
µm	=	micrometer
MOY	=	Methyl-2-octynoate
MPD	=	mean peptide depletion
MW	=	molecular weight
n	=	number
NC	=	negative control
No.; no.	=	number

nm	=	nanometer
OECD	=	Organisation for Economic Co-operation and Development
p.a.	=	per analysis
PA	=	Perillaaldehyde
PB	=	Phenyl benzoate
PC	=	positive control
PPA	=	Phenylpropionaldehyde
PPD	=	4-phenylenediamine
R	=	arginine
s	=	solid
SD	=	standard deviation
SC	=	substance control
TCS	=	Tetrachlorsalicylanilide
TMD	=	Tetramethyldiuram disulfide
UV	=	ultraviolet
WP	=	work package

3. Introduction

3.1. Background

Reaction of electrophilic chemicals with nucleophilic residues in skin proteins is the Molecular Initiating Event (MIE) in skin sensitization. The modification of the proteins creates novel immunogenic epitopes. These epitopes ultimately trigger activation of epitope-specific T-cells which are the effector cells finally elliciting allergic contact dermatitis. While additional steps are involved in the acquisition of skin sensitization (epitope processing, danger signal formation, activation of dendritic cells and keratinocytes, cell migration), the MIE is of predominant importance. To characterize the potency of a skin sensitizer, it is therefore important to characterize the reactivity of the test chemicals with skin proteins or surrogate nucleophilic residues with similar reactivity.

Reactivity in chemical terms is ideally expressed as a rate constant, which indicates how fast two chemicals react with each other, or more precisely how much reaction product is formed from a given amount of chemicals in a given unit of time.

The only OECD approved methods (OECD 442C) to characterize reactivity of chemicals to determine their skin sensitization potential are the Direct Peptide Reactivity Assay (DPRA) and the ADRA, the latter being a modification of the DPRA testing at a lower test concentration with more easily detectable amino acid derivatives as surrogate nucleophiles. In the classical DPRA, reactivity with two test peptides as surrogate nucleophiles is measured, one containing a reactive cysteine and the other containing a lysine residue. Reactivity is quantified as the relative depletion of the peptide signal (detected with HPLC-UV) after incubation with the test chemical, assuming that all peptide depletion comes from the formation of reaction products between the peptide and the test chemical. The prediction model of the DPRA is based on the mean of lysine- and cysteine-peptide depletion at a single time point measurement (≥ 24 h) and at one fixed test material concentration (5 mM for the Cys-peptide). A very similar setup is used for the ADRA, with the exception that the chosen fixed concentration is lower (1 mM test material stock concentration vs. 100 mM in the DPRA). A late and single detection time point and one fixed concentration limit the dynamic range of these assays: Chemicals of differing potency may lead to a complete reaction (i.e. 100% depletion) or chemicals with widely differing initial reaction rates may produce similar final depletion.

The kinetic direct peptide reactivity assay (kDPRA) is a modification of the DPRA (OECD TG 442C, DB-ALM protocol 154). The kDPRA uses kinetic rates of cysteine- peptide depletion to distinguish between two levels of skin sensitization potency, i.e. to discriminate between CLP/GHS sub-categories 1A and 1B. In addition, kinetic rates generated with this method have a strong quantitative correlation to sensitizing potency and can therefore be used in defined approaches (DA) with a quantitative data integration procedure (DIP) for skin sensitization potency assessment.

In the kDPRA, the reaction kinetics of a test substance towards the same synthetic cysteine (C)-containing peptide as used in the DPRA is evaluated. For this purpose, several concentrations of the test substance are incubated with the synthetic peptide for several incubation times at 25°C. After the respective reaction time, the reaction is stopped by addition of the fluorescent dye monobromobimane (mBrB). Highly reactive and non-fluorescent mBrB rapidly reacts with unbound cysteine moieties of the model peptide to form a fluorescent complex. The remaining non-depleted peptide concentration is determined thereafter by fluorescence measurement. By stopping the reaction with mBrB, the extent of peptide depletion can be determined at precisely defined time points.

The analysis of different databases on Cys- and Lys-depletion by skin sensitizers have shown that the vast majority of skin sensitizers predominantly react with - or at least more strongly react with - the Cys-peptide. Therefore, the kDPRA only addresses Cys-peptide reactivity. In this context it should be mentioned that also the classical DPRA

includes a Cys-only prediction model to characterize reactivity, e.g. in absence of Lys-peptide depletion values due to technical reasons. This Cys-only prediction model has a similarly high predictivity.

The basic method of the kDPRA has been described for the first time in [1]. The approach to use it for rate constant determinations was described in [2] and later applied to more chemicals from specific domains [3]. Lately, it was applied to a more diverse set of chemicals and a tentative prediction model was proposed with a cut-off of log k_{max} = -1.1 to discriminate between GHS categories 1A and 1B (referred to as the 'published cut-off' in the following) [4].[1]

The accuracy and reproducibility of these rate constants needs to be ascertained in order to use them in a DA and DIP. If rated successful, this protocol could be integrated into OECD test guideline 442C, to amend the DPRA for potency assessment of chemicals rated as sensitizers by either the DPRA or any binary ITS for hazard assessment.

Note: The unit of the kinetic rate constants as calculated throughout this report and according to the SOP and evaluation template is $M^{-1}s^{-1}$. The majority of calculations are made with the logarithmic value of the rate constants. These logarithmic values are given unit-free and always refer to the logarithm of the rate constant measured in $M^{-1}s^{-1}$.

3.2. Principle of the test

The kDPRA is a modification of the in chemico test method DPRA (described in Appendix I of OECD TG 442C). The kDPRA uses the cysteine containing test peptide (Ac-RFAACAA-COOH) also used in the DPRA, while it does not use a lysine containing peptide. The final concentration of the test peptide (0.5 mM) and the reaction medium (25% acetonitrile in phosphate buffer) is identical in the kDPRA as in the DPRA. While the DPRA measures only at one concentration of the test chemical (5 mM for the cysteine containing peptide) and at one time point (≥ 24 h), the kDPRA is performed as parallel reactions at five concentrations (5, 2.5, 1.25, 0.625 and 0.3125 mM) and at six time-points (30, 90, 1250, 210 and 1440 min) at 25ºC. Residual concentration of the cysteine peptide after the respective reaction time is measured by stopping the reaction by the addition of monobromobimane. The highly reactive and non-fluorescent monobromobimane rapidly reacts with unbound cysteine moieties of the model peptide to form a fluorescent complex. The remaining non-depleted peptide concentration can thus be determined. If the depletion of the top concentration (5 mM test chemical) surpasses the threshold of 13.89% (cut-off used in the DPRA for positivity) and this depletion is statistically significant vs. controls with peptide only, further calculations are performed: The natural logarithm of the non-depleted peptide concentrations is plotted vs. the concentration of the test chemical at each time point. If a linear relationship is observed (correlation co-efficient >0.9), the slope of this curve is determined and divided by the incubation time to calculate the rate constant in [$min^{-1}mM^{-1}$]. This value is transformed to the rate constant in [$s^{-1}M^{-1}$] and the logarithm is taken. The maximal value observed at any time point is taken as the log k_{max}, and this maximal rate constant is the primary read-out of the test. It gives a quantification of the maximal kinetic rate of the reaction of the test chemical with the test peptide. Kinetic reaction rates of the Cys-peptide depletion are then used to discriminate GHS subcategory 1A sensitiser from GHS subcategory 1B sensitisers/ not classified. The quantitative rate constant can also be used in other defined approaches for sensitization potency assessment.

3.3. Log k_{max} Calculation Example DNCB

[1] In the original paper [4] k_{max} = -1.73 is reported as cut-off. This value was derived using stock concentrations of the substances for k calculation. In the present ring trial all calculations were performed using the actual final concentrations of the substances and hence the value originally reported in [4] is given transferred to this calculation in this report.

Detailed technical information including instructions for calculations are available in the SOP. We provide here an example calculation for the ease of understanding.

A depletion matrix summarises the Cys-peptide depletions (in %) determined at the various final test chemical concentrations as well as exposure times. The following is the depletion matrix for DNCB

	Concentration [mM]	5	2.5	1.25	0.625	0.3125
t [min]	10	49.09	31.50	14.83	6.60	7.62
	30	81.53	61.32	40.70	26.58	15.91
	90	99.00	90.46	70.01	43.04	23.40
	150	99.00	98.07	84.98	60.77	39.46
	210	99.00	99.00	92.00	70.20	45.66
	1440	99.00	99.00	99.00	99.00	80.83

Only if at a given exposure time the depletion of the 5 mM (final test chemical) concentration surpasses the threshold of 13.89% (cut-off used in the DPRA for positivity in the Cys-only prediction model) and this depletion is statistically significant (t-test, 2 sided, unequal variance, $p < 0.05$ of three replicates vs. 12 wells with controls with peptide only), further calculations are performed for that given exposure time.

The values are plotted as ln(100-dp) with dp being the relative depleted peptide concentrations (Figure 1). From each of the curves (one per exposure time), the slope is calculated. For this calculation, points at saturation (i.e. several consecutive test chemical concentrations with >95% depletion) will be omitted, as this would affect the linearity and in those cases reaction went to completion and no more peptide is available for the reaction. This gives the following matrix for slope calculations with the associated slopes:

		ln(100 - dp) at concentration [mM]					Slope	Correlation
		5.0	2.5	1.25	0.625	0.3125		
t [min]	10	3.93	4.23	4.44	4.54	4.53	0.1343	0.9947
	30	2.92	3.66	4.08	4.30	4.43	0.3207	0.9988
	90	0.00	2.26	3.40	4.04	4.34	0.9239	0.9998
	150		0.66	2.71	3.67	4.10	1.5839	0.9996
	210		0.00	2.08	3.39	4.00	1.8262	0.9985
	1440				0.00	2.95	9.4510	1.0000

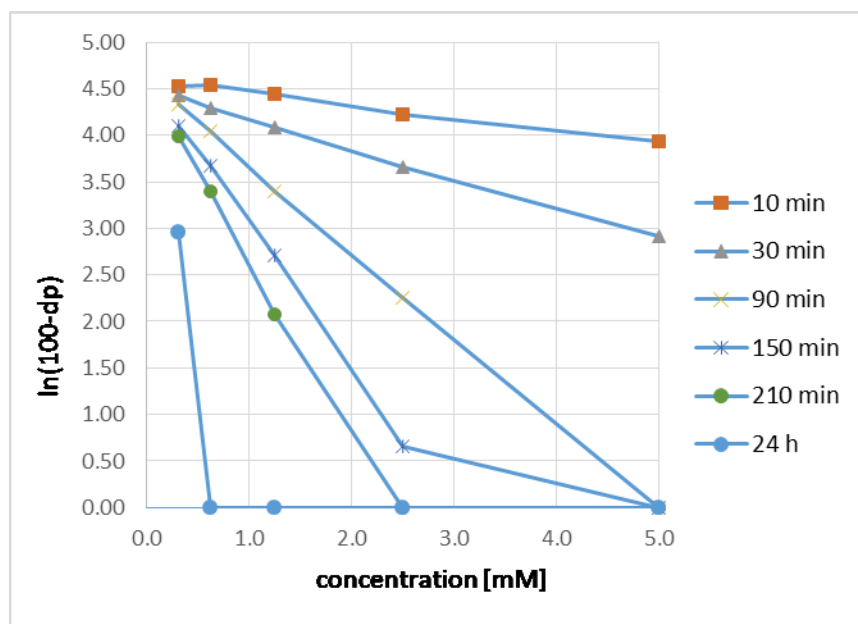


Figure 1. Depletion matrix for DNCB. For each exposure time a separate curve is provided. On the x-axis the concentration is plotted and on the y-axis the term Ln(100-depletion) is plotted. The slope from these plots gives rate constants (k_{obs}) at different time points, which can be divided by the time to arrive at the rate constant calculated by the evaluation sheet.

The slopes [mM^{-1}] are then divided by the time [min], transformed to [$s^{-1}M^{-1}$] and logarithm is taken.

This gives the following result:

	t [min]	Slope		Correlation	**log k**
calculation of k for each time point	**10**	0.13429	k (mM^-1)	0.9947	
		0.22382	k (s^-1 M^-1)		**-0.65**
	30	0.32069	k (mM^-1)	0.9988	
		0.17816	k (s^-1 M^-1)		**-0.75**
	90	0.92386	k (mM^-1)	0.9998	
		0.17109	k (s^-1 M^-1)		**-0.77**
	150	1.58391	k (mM^-1)	0.9996	
		0.17599	k (s^-1 M^-1)		**-0.75**
	210	1.82623	k (mM^-1)	0.9985	
		0.14494	k (s^-1 M^-1)		**-0.84**
	1440	9.45103	k (mM^-1)	1.0000	
		0.10939	k (s^-1 M^-1)		**-0.96**

As can be seen for such an "ideal chemical" like DNCB, reaction is linear over time, and a similar rate is derived from time points up to 210 min, but then it becomes less accurate as only the lowest concentrations can be considered after 24 h and the higher concentrations are out range (reaction went to completion).

The automated Excel spreadsheet then reports as a final result the k_{max} of -0.65 observed at 10 min.

The maximum reactivity (k_{max}) was at:	10	min
with a log k_{max}:	**-0.65**	

4. Goals set for the validation study

This study is based on the modular approach to test validation as defined by Hartung *et al.* [5], which includes the modules (i) Test definition, (ii) Transferability, (iii) Within-laboratory variability, (iv) Between-laboratory variability, (v) Predictive capacity, (vi) evaluation of applicability domain, and (vii) eventually definition of performance standards.

Thus, the specific goals of this study were:

- Standardization of the kDPRA protocol and the rate constant calculations and submission of the method as DB-ALM protocol (Test definition)
- Test transferability of the method (Phase I)
- Test intra- and inter-laboratory reproducibility of the method (Phase II):
    - Evaluation of quantitative accuracy of the continuous variable, i.e. log-transformed rate constants (variability in terms of standard deviation; both within / between laboratories)
    - Evaluation of classification reproducibility with the published tentative cut-off to discriminate GHS Cat 1A and GHS Cat 1B [4] (within / between laboratories) (published cut-off log k_{max} = -1.1; based on final test substance concentrations)
    - Evaluation of classification reproducibility with a refined cut-off which is based on the evaluation of a broader dataset to discriminate GHS Cat 1A and GHS Cat 1B

- Test predictivity of the assay both with the published and the refined cut-off

The validation plan is summarized in a separate document "**2018-04-06 Kinetic DPRA_Validation outline.doc**". This document had been shared with and was reviewed by ECVAM prior to initiation of the study.

5. Work packages

The study included the following work packages:

WP 3.1. Preparation phase / protocol refinement: The detailed first version of the SOP was developed by the two lead labs.

WP 3.2. Transfer phase of the refined protocol in experienced labs: The lead labs tested 6 reactive proficiency chemicals from the DPRA and 6 weakly / non-reactive chemicals. The latter chemical set was chosen to optimize the data evaluation template.

WP 3.3. Transfer phase to naïve labs (Phase I): The protocol was transferred to 5 naïve labs and tested with the positive control and 6 reactive proficiency chemicals from the DPRA.

WP 3.4. Blind-coded Testing – Phase II: 24 chemicals were tested for inter-laboratory variability in 7 labs and for intra-laboratory variability in 3 – 4 labs.

WP 3.5. Predictive capacity for the dataset of Phase I and Phase II: Predictivity with published and with the refined cut-off was evaluated on the chemical set tested in multiple labs. [The 'refined cut-off' is based on the ROC analysis conducted using the data of WP3.6]

WP 3.6. Building the database, testing the prediction cut-off and evaluate predictive capacity on broader set of chemicals. A larger kDPRA database was created based on 1) literature data, 2) the validation study data and 3) by testing more chemicals in single laboratories. This extended database was used to determine the optimal rate constant cut-off.

In the different modules, also a direct comparison to historical Cys- depletion as measured in the classical DPRA was made in order to compare the two methods.

5.1. WP 3.1.: Preparation phase / protocol refinement

The principle of the protocol was not changed from the previous publications [2-4], however a fully standardized SOP was written, which includes a fixed plate-setup to test three chemicals in a run at multiple time points and at five concentrations in triplicate.

A standardized EXCEL spreadsheet for rate constant calculation (including cut-offs and outlier treatment) was developed. This template forms part of the SOP. Both the SOP and the template will be freely publicly available.

The refinement history of the SOP and spreadsheet is summarized in table 1.

Table 1. Overview of SOP and Excel spreadsheet versions used for each phase of the ring trial.

SOP Revision	Date:	Description of change:	Used for:	Associated Excel Template:
Version 1	16 Mar 2018		Phase I (experienced labs) (Runs 1 and 2)	Version of 07 Mar 2018 (later transferred to version of 17 Apr 2018)
Version 2	17 Apr 2018	- PC was changed from EGDMA to cinnamic aldehyde - 5 min time point was discarded from combined measurements - Positivity and linearity criteria were added to the evaluation	Phase I (experienced labs, remaining runs)	Version of 17 Apr 2018
Version 3	30 May 2018	- Positivity cut-off was explained - Validity criteria was added	Phase I (naïve labs)	Version of 30 May 2018
Version 4	18 Sep 2018 (including 1 amendment)	- Additional source for mBrB and buffer salts added - Test chemical solubilisation was clarified - mBrB preparation clarified - Alternative preparation for water-soluble substances was added - Validity criteria added - Measurement time points were confirmed for phase II - Explanation how to treat special cases in evaluation (Annex I) was added	Phase II (all labs)	Version of 18 Sep 2018

1) In the original version of the SOP used in [4], EGDMA was used as PC, however for all the inter-laboratory testing cinnamic aldehyde was the positive control. The reason for this selection is that (i) Cinnamic aldehyde is also the PC in DPRA and (ii) it reacts significantly already at early time points, while EGDMA ony reacts late (iii) it is a GHS 1A sensitizer

5.2. WP3.2. Transfer phase of the refined protocol in experienced labs

5.2.1. Procedure

- The refined protocol and rate constant calculation sheet were tested in **2 experienced labs** (Phase I, BASF and Givaudan).
- Six chemicals (positive proficiency substances and positive control from the DPRA) were selected (Set A) (Table 2).
- In addition 6 chemicals with weak reactivity in the DPRA were tested (Set B) (Table 2).
  - This was required to optimize the evaluation sheet to avoid the calculation of reaction rates in the case of non-significant peptide depletion, which would lead to meaningless constants calculated from noise in the data.
  - The calculation sheet was refined based on these results.
  - The goal was to automate data evaluation as much as possible to avoid any bias in data evaluation
- All 12 chemicals (see Table 2) were tested in two repetitions and data were shared between the two labs.
  - Rate constant variability and consistency in calculations were checked.
  - Refinements needed for the calculation sheet were made as required.
  - Minimal requirements (see section 5.3.1 5.3.1 Goals and set-up) for the Set A chemicals (positive proficiency substances and positive control from the DPRA) were defined, which the naïve labs had to fulfill in the transfer phase in order to progress to the blind-coded phase.

Table 2. Chemicals for transfer phase of the refined protocol in two experienced labs.

Name	CAS-Number	LLNA based UN GHS category	Rationale for selection	Cys-depletion [%] [2)]	Lys-depletion [%][2)]
Set A: Positive chemicals for transfer phase					
2,4-Dinitrochlorobenzene	97-00-7	1A	DPRA proficiency substance, positive	100.00	14.67
Oxazolone	15646-46-5	1A	DPRA proficiency substance, positive	75.50	49.64
Formaldehyde	50-00-0	1A	DPRA proficiency substance, positive	60.41	11.20
Ethylene glycol dimethacrylate (EGDMA)	97-90-5	1B	DPRA positive, originally used as positive control in Wareing et al. [1)]	87.28	12.38
Benzylideneacetone	122-57-6	1B	DPRA proficiency substance, positive	94.68	1.50
2,3-Butanedione	431-03-8	1B	DPRA proficiency substance, positive	79.04	27.00

Set B: No or minimal to low-reactive chemicals to optimize calculations					
3-Propylideneph-thalide	17369-59-4	1B	Low Cys-reactivity in DPRA	14.30	30.60
Cinnamic alcohol	104-54-1	1B	No or minimal Cys-reactivity in DPRA	0.00	15.10
Ethylenediamine	107-15-3	1B	No or minimal Cys-reactivity in DPRA	3.40	0.00
Eugenol	97-53-0	1B	No or minimal Cys-reactivity in DPRA	9.20	19.20
Pentachlorophenol	87-86-5	2	No or minimal Cys-reactivity in DPRA	0.00	14.50
Xylene	1330-20-7	1B	No or minimal Cys-reactivity in DPRA	0.00	0.73

Note: This list includes five DPRA proficiency substances and the positive control from the DPRA. Farnesal, a positive proficiency substance from the DPRA is excluded, since it leads mainly to peptide oxidation – which is a variable, partly stochastic process, and thus this chemical is considered not ideal for a transferability study. The Set A chemicals were also used in the transferability phase to the naïve labs (WP3.2).

[1)] Cinnamic aldehyde is now routinely used as positive control in the kDPRA since it is also the positive control of the DPRA

[2)] Historical data from Urbisch et al. (2015).

5.2.2. Results for the positive chemicals (Set A)

The six positive reference chemicals were successfully tested in both lead labs with the refined protocol. Rate constants were calculated with the automated evaluation sheet, and consistent results were obtained. The results for log k_{max} values are shown in Table 3.

The logarithmic standard deviations are between 0.06 (for 2,4-dinitrochlorobenzene) and 0.22 (ethylene glycol dimethacrylate). Standard deviations tend to be higher for chemicals with lower reaction rates. A logarithmic standard deviation of 0.06 corresponds to a variation by a factor of 1.15-fold for linear data, while 0.22 corresponds to 1.65-fold. This is the fold variation of the rate constant around the geometric mean (geometric standard deviations are factors). For these chemicals the rate constants range from -0.56 - -2.78, i.e. two orders of magnitude.

Table 3. Six positive transfer phase chemicals (Set A) repeated twice in two experienced labs: log k_{max} values

	BASF – Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2	Average	SD
2,4-Dinitrochlorobenzene	-0.52	-0.51	-0.58	-0.65	-0.56	0.06
Oxazolone	-0.16	-0.20	-0.06	-0.10	-0.13	0.06
Formaldehyde	-0.84	-0.63	-0.64	-0.53	-0.66	0.13
Ethylene glycol dimethacrylate	-2.49	-2.13	-2.60	-2.58	-2.45	0.22
Benzylideneacetone	-1.82	-1.71	-1.92	-1.78	-1.81	0.09
2,3-Butanedione	-3.08	-2.74	-2.61	-2.68	-2.78	0.21

Table 4 indicates the classification based on the published prediction model, while Table 5 indicates the prediction by the refined prediction model. Consistent predictions in the four runs conducted in the two labs were noted. Benzylideneacetone was assessed as GHS Cat 1A by the refined prediction model, which is the human classification.

Table 4. Six positive transfer phase chemicals (Set A) repeated twice in two experienced labs: Prediction by the published prediction cut-off (log k_{max} -1.1).

	BASF - Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2
2,4-Dinitrochlorobenzene	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Oxazolone	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Formaldehyde	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Ethylene glycol dimethacrylate	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B
Benzylideneacetone	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B
2,3-Butanedione	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B

Table 5. Six positive transfer phase chemicals (Set A) repeated twice in two experienced labs: Prediction by refined prediction cut-off (log k_{max} -2.0).

	BASF - Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2
2,4-Dinitrochlorobenzene	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Oxazolone	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Formaldehyde	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
Ethylene glycol dimethacrylate	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B
Benzylideneacetone	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A	GHS Cat 1A
2,3-Butanedione	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B	GHS Cat 1B

5.2.3. Results for the positive control

Table 6 summarizes the results obtained for the positive control in this test phase. For the values recorded at 90 and 150 min, very similar results between the two labs and similarly low standard deviations were recorded. BASF observed statistically significant reactivity at earlier time points in four repetitions (runs 2, 3, 4 and 6) - hence in these cases log k_{max} was lower than the one measured at Givaudan, which also led to a different classification with the published prediction model. However, the values are close to the cut-off. With the refined cut-off, all runs rated cinnamic aldehyde as a 1A chemical.

Based on these data (see table 6), acceptance criteria for the positive control for phase I with naïve laboratories and for the SOP were defined. These cover the range observed in this set of experiments (mean of four experiments ± 2 × standard deviation of phase I data of Givaudan and BASF). This is a preliminary range strongly dependent on the observed standard deviation from testing in these two labs.

- Acceptance criteria for the positive control:
  - the log k of the PC at **90 min** should be within the following range: **-1.70 $M^{-1}s^{-1}$ to -1.45 $M^{-1}s^{-1}$.**
  - If no log k is obtained at 90 min, the value at **150 min** can be taken into account and should lie in the following range: **-1.83 $M^{-1}s^{-1}$ to -1.50 $M^{-1}s^{-1}$.**

Based on the results for the positive chemicals in Set A and the positive control, the protocol was deemed stable and ready for transfer. However, to make sure that the rate constant calculations using the Excel spreadsheet are robust, the tests on Set B were conducted as planned.

Table 6. Results of all runs (conducted on Set A and Set B) for the positive control cinnamic aldehyde in the two experienced labs

	log k_{90}	log k_{150}	log k_{max}	Classification published cut-off	Classification refined cut-off
			BASF		
Run 1	No data available (-EGDMA used as positive control)				
Run 2	-1.73	-1.60	-1.09	GHS Cat 1A	GHS Cat 1A
Run 3	-1.52	-1.60	-0.94	GHS Cat 1A	GHS Cat 1A
Run 4	-1.59	-1.59	-1.08	GHS Cat 1A	GHS Cat 1A
Run 5	-1.65	-1.60	-1.23	GHS Cat 1B	GHS Cat 1A
Run 6	-1.70	-1.65	-1.00	GHS Cat 1A	GHS Cat 1A
Run 7	-1.56	-1.61	-1.56	GHS Cat 1B	GHS Cat 1A
Run 8	-1.54	-1.63	-1.54	GHS Cat 1B	GHS Cat 1A
			GIV		
Run 1	-1.55	-1.92	-1.21	GHS Cat 1B	GHS Cat 1A
Run 2	-1.44	-1.66	-1.44	GHS Cat 1B	GHS Cat 1A
Run 3	-1.53	-1.66	-1.17	GHS Cat 1B	GHS Cat 1A
Run 4	-1.54	-1.71	-1.54	GHS Cat 1B	GHS Cat 1A
Run 5	-1.59	-1.68	-1.29	GHS Cat 1B	GHS Cat 1A
Run 6	-1.66	-1.67	-1.66	GHS Cat 1B	GHS Cat 1A
Run 7	-1.60	-1.70	-1.57	GHS Cat 1B	GHS Cat 1A
Run 8	-1.61	-1.68	-1.41	GHS Cat 1B	GHS Cat 1A
BASF mean	-1.61	-1.61	-1.20		
GIV mean	-1.57	-1.71	-1.41		
Overall mean	-1.59	-1.66	-1.31		
SD	0.07	0.08	0.23		
2 SD	0.15	0.16	0.47		
3 SD	0.22	0.24	0.70		
min	-1.73	-1.92	-1.66		
max	-1.44	-1.59	-0.94		

5.2.4. Results for the chemicals of Set B

Table 7 list the results for log k_{max} for the chemicals of Set B. Eugenol was rated weakly positive by all four experiments with a consistent rate calculated, indicating proper rate calculation also for chemicals with maximal 20 – 30% depletion. 3-propylidenephthalide, cinnamic alcohol, ethylenediamine, and xylene were basically non-reactive, with no rate constants calculated in most instances. In addition, peptide depletion at the concentration of 5 mM at 24 hours (i.e. DPRA conditions) was below the cut-off for positivity of the Cys-only model (> 13.89% depletion as described in OECD TG 442C) for these four chemicals (Table 8).

Table 7. Six minimally reactive substances (Set B) repeated twice in two experienced labs: log k_{max} values

	BASF - Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2	Average	SDSD
3-propylidenephthalide	n-r	n-r	-2.38	n-r	n-r	
cinnamic alcohol	n-r	-2.24	n-r	n-r	n-r	
ethylenediamine	n-r	-3.44	n-r	n-r	n-r	
eugenol	-2.50	-2.22	-2.49	-3.32	-2.64	0.48
pentachlorphenol	-1.14	-1.01	-0.96	-1.08	-1.05	0.08
xylene	n-r	n-r	n-r	n-r	n-r	

n-r = non-reactive

Table 8. Six minimally reactive substances (Set B) repeated twice in two experienced labs: Depletion at 5 mM after 24 h (equivalent to 100 mM stock concentration as in OECD TG 442C used to estimate reactivity based on Cys-only)

	BASF - Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2	Average	SDSD
3-propylidenephthalide	11.54	6.58	8.01	3.71	7.46	3.26
cinnamic alcohol	1.00	12.82	5.42	1.00	5.06	5.58
ethylenediamine	10.19	17.35	1.00	1.00	7.38	7.93
eugenol	20.67	22.96	28.46	19.24	22.83	4.05
pentachlorphenol	29.11	20.76	24.65	34.93	27.36	6.09
xylene	11.18	5.67	6.95	1.00	6.20	4.19

However, for 3-propylidenphthalide, ethylene diamine and cinnamic alcohol, a rate constant was calculated in one run in one lab and only based on one time point, i.e. at this time point the chemical did pass the Cys depletion cut-off for positivity, but this was not consistent over time (Table 9). Based on this observation, an alert was defined in the revised evaluation sheet, which appears as **'time course interrupted'** if indeed the cut-off is passed at an early timepoint but the result is not consistent over time. In the revised SOP for phase II it was defined that in such instances the test needs to be repeated, and the reaction rate calculated is only accepted if this reaction at an early time point only is verified in an additional independent experiment.

Table 9. Time course of kinetic rates automatically determined for 3-propylidenphthalide (3-PP) and cinnamic alcohol (CA)

Run	Test sub-stance	log k_{Max}	reaction time [min]						Cys-depletion 5 mM at 24 h [%]
			10	30	90	150	210	1440	
BASF 1	3-PP	n-r							11.54
BASF 2	3-PP	n-r							6.58
GIV 1	3-PP	-2.38			-2.38				8.01
GIV 2	3-PP	n-r							3.71
BASF 1	CA	n-r							1.00
BASF 2	CA	-2.24			-2.24				12.82
GIV 1	CA	n-r							5.42
GIV 2	CA	n-r							1.00

n-r = not reactive; note the observed reaction above cut-off at one time point only

In addition, it was observed that in the case of pentachlorophenol a reaction rate was already calculated at a very early time point, due to ca. 20% depletion. This (average) depletion did not significantly change over time (Table 10). Such a kinetic behavior is possible in case of a reversible reaction rapidly reaching equilibrium, and indeed it

can also be observed in the case of formaldehyde. For formaldehyde this can chemically be explained by the formation of a rapidly reversible thiohemiacetal adduct which reaches equilibrium quickly. However, in the case of pentachlorophenol this was further investigated, as this chemical is negative in the classical DPRA (OECD TG442C) and has no alert for reversible adduct formation. When looking at the fluorescence values of the solvent controls with chemical alone, in absence of the peptide, it is noted, that the values decrease at increasing test chemical concentrations (Figure 2), hence the residual fluorescence of non-reacted mBrB is quenched by the test chemical. While this quenching phenomenon is no direct cause of quenching of the fluorescence of the peptide- mBrB adduct finally measured, it is an indication that such a quenching at the wavelengths used for fluorescence detection may occur. A reduction of the fluorescence by optical interference with the chemical would lead to the calculation of an apparent peptide depletion, which would be just based on fluorescence quenching / absorbance rather than true depletion. Based on these observations two additional alerts were programmed into the evaluation sheet: **'depletion not time dependent'** is highlighted in case an early depletion is observed which is not significantly increasing over time, i.e. no kinetic reaction is observed. An alert **'potential quenching'** was added in case a reduction of the residual fluorescence is observed for the wells with test chemicals only and no peptide. If both these alerts are triggered, then a third alert is triggered **depletion ev. from quenching => 'TRUE'**. See also Appendix 2 for more discussion on potential quenching effects.

In parallel, test chemicals could also exert autofluorescence, i.e. enhanced fluorescence by the chemical alone. Since this fluorescence is subtracted from the value in presence of the chemical and the peptide, a strong autofluorescence may also lead to an apparent peptide depletion, which is not time dependent. (This is the case if the fluorescence of the peptide- mBrB adduct in presence of the test chemical is smaller than the theoretical sum of the peptide- mBrB adduct fluorescence and the autofluorescence of the test chemical).

Therefore, two further alerts were added: If increased fluorescence is observed for the test chemical in absence of the peptide, the alert **'Autofluorescence may influence result'** is triggered, if this coincides with peptide depletion which is not time-dependent, the alert **depletion ev. from autofluorescence => 'TRUE'** is triggered. These additional alerts do not change the way the reaction rates are calculated but give a warning to the user where artefacts may be observed, which need to be recorded and further explored when evaluating the data (similar to potential coelution of the chemical and the peptide in HPLC in the classical DPRA).

Table 10. Depletion [%] over time in the presence of 5 mM pentachlorophenol (%)

Time	BASF - Run 1	BASF - Run 2	GIV - Run 1	GIV - Run 2	Average
10	19.2	26.4	29.7	23.0	24.6
30	16.7	12.7	31.6	23.0	21.0
90	37.6	17.7	27.5	24.0	26.7
150	15.5	7.6	27.0	23.1	18.3
210	11.7	12.0	29.4	27.9	20.3
1440	29.1	20.8	24.7	34.9	27.4

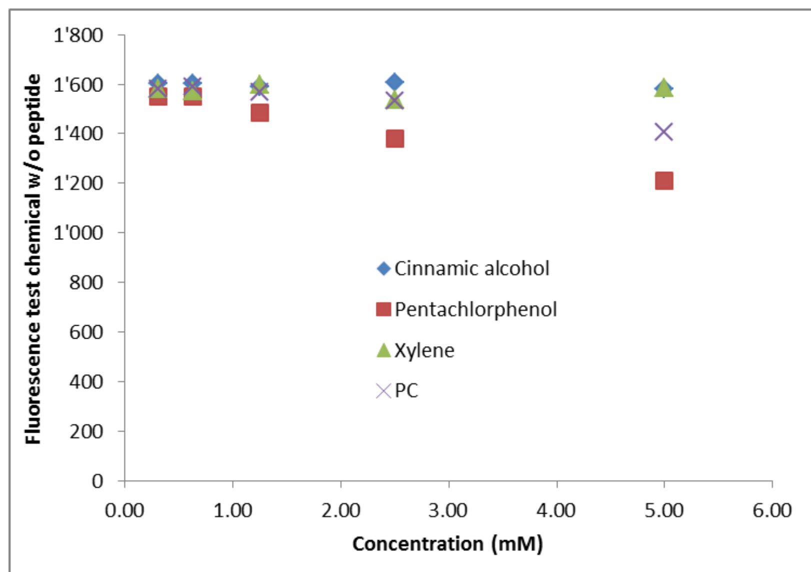


Figure 2. Fluorescence of control wells with test chemical only (no peptide) after mBrB addition.

(Note the reduced fluorescence with increasing pentachlorophenol concentration; a slight depletion is also observed for the positive control, but this is below the threshold of positivity)

5.2.5. Conclusions testing in experienced labs

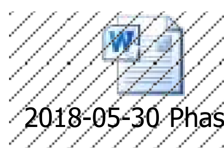
- With the SOP, reproducible results for both DPRA positive chemicals (set A) and the PC are obtained with low standard deviations.
- The Excel spreadsheet template for automatic rate constant calculation for these chemicals works well.
- Additional alerts for chemicals with very weak reactivity at single time point and chemicals with potential autofluorescence or fluorescence quenching further improve the Excel spreadsheet template and alert the user to cases where there may be test substance specific interference which needs to be further explored.

5.3. WP 3.3. Transfer phase to naïve labs (phase I)

5.3.1. Goals and set-up

- The refined protocol and rate constant calculation sheet were next tested in **5 naïve labs**, using the six positive proficiency substances and positive control from the DPRA mentioned above (Set A). At least two repetitions were performed. If data fulfilled the acceptance criteria and were considered consistent by the conducting naïve lab, they were sent to the lead labs.
- Data were evaluated vs. the minimal requirements set above (chapter 5.2.3. Results for the positive control) and, if needed, further repetitions were performed by the labs to be ready to progress to reproducibility phase (phase II).
- Test chemicals for this phase were purchased from Sigma Aldrich and sent by BASF and were not blind coded.
- For this phase, the labs worked with the peptide quality they routinely use for the DPRA (according to OECD TG 442C).

- The naïve laboratories received the SOP, the evaluation sheet and below word document, containing the substances to be tested and the span of results considered proficient as well as the validity criteria for a test run and further instruction on how to proceed in phase I (Note: The file can be viewed by double clicking).
- The labs received an empirical range of log k_{max} observed in the lead labs as guidance (see WP 3.2). They had to list all runs performed, in order to track the amount of training they needed to reach proficiency.
- Two telephone conferences were held to clarify all questions after reading the SOP, but no hands-on training was performed.



5.3.2. Setting up the test in different labs

All labs started the testing of chemicals in set A and recorded each run performed. Tests with the positive control within the acceptance range and with consistent results in the target range for the test chemicals were accepted, but all data were sent to the lead labs. Table 11 lists the number of tests conducted in the 5 naïve labs and Table 12 specifies this for the different chemicals.

Table 11. Training phase for naive labs: number of runs conducted.

Lab	Number of Runs conducted [1)]	PC criterion fullfilled	PC failed	Number of chemicals in range	Not in range[2)]
Lab D	11	8	3[3)]	5	2,3-Butanedione [4)]
Lab F	7	6	1	5	Benzylideneacetone[4)]
Lab C	5 (in one test run the PC was tested, only)	5	0	5	Oxazolone4[)]
Lab E	6	6	0	3	2,4-Dinitrochlorobenzene Formaldehyde 2,3-Butanedione [5)]
Lab G	7	7	0	6	-

[1)] One run means an experiment with one master plate, normally containing three chemicals and the PC. A minimum of 4 runs thus is needed to test the 6 chemicals in set A in two repetitions.
[2)] The ranges were based on the Phase I data of the lead labs and corresponds to the mean of total five experiments ± 2 × standard deviation.
[3)] Two of the runs with the PC acceptable range not met were conducted rather earlier during the implementation of the assay. The third one missing the AC was late during the ring trial, but was close to the (narrow) AC range set.
[4)] The obtained result was only slightly out of the pre-defined range and can principally considered to be proficient.
[45)] For two chemicals (DNCB and 2,3-Butanedione) the obtained result was only slightly out of the pre-defined range and can principally considered to be proficient.

Table 12. Training phase for naive labs: number of runs needed to achieve at least 2 valid and proficient runs for the different chemicals.

Chemical	Lab D	Lab F	Lab C	Lab E	Lab G
2,4-Dinitrochlorobenzene	6	5	2	2 proficient runs not obtained in 3 total runs conducted with this chemical[1)]	4
Oxazolone	5	5	2 proficient runs not obtained in 2 total runs conducted with this chemical [1)]	2	2
Formaldehyde	5	2	2	2 proficient runs not obtained in 3 total runs conducted with this chemical	3
Ethylene glycol dimethacrylate	5	2	2	2	2
Benzylideneacetone	4	2 proficient runs not obtained in 5 total runs conducted with this chemical [1)]	2	2	2
2,3-Butanedione	2 proficient runs not obtained in 6 total runs conducted with this chemical [1)]	2	2	2 proficient runs not obtained in 3 total runs conducted with this chemical [1)]	2

[1)] The obtained result was only slightly out of the pre-defined range and can principally considered to be proficient, see detailed results below.

5.3.3. Reproducibility of log k_{max}

Table 13 lists the log k_{max} values obtained for the two final runs (three runs reported by Lab E) for the chemicals of Set A. Overall very similar results were obtained for these chemicals as reported for the experienced labs in Table 3. The only exception is formaldehyde which was much less reactive in all three repetitions performed at Lab E.

Figure 3 shows a comparison of the results obtained for the two runs in experienced labs vs. the two final runs in naïve labs. The data are shown with the outlier formaldehyde at Lab E omitted.

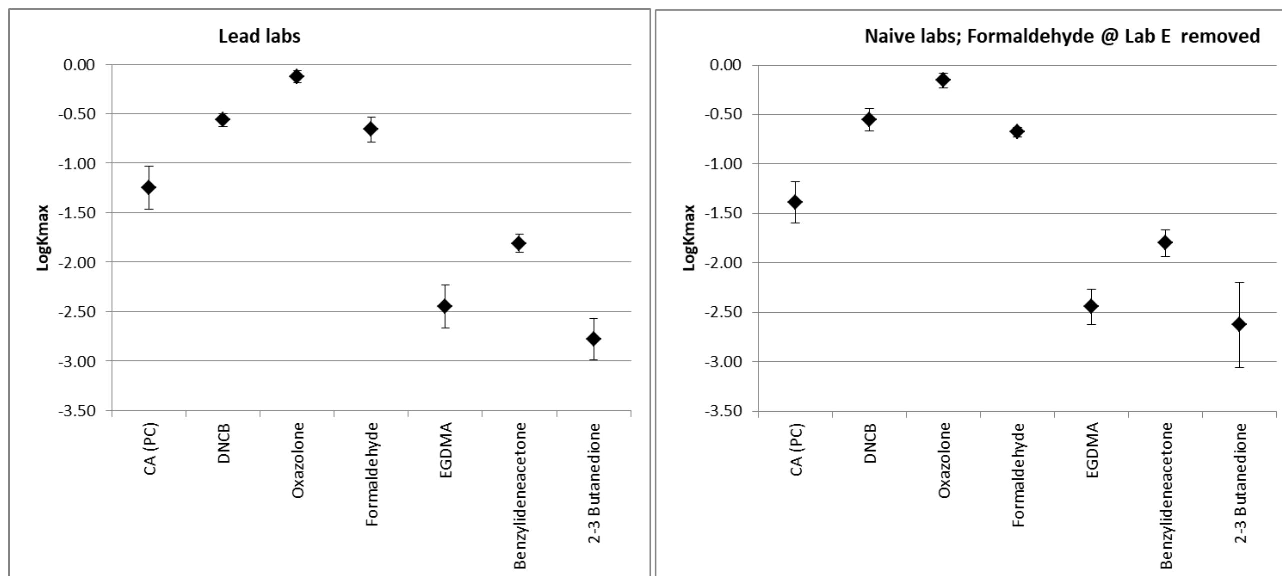


Figure 3. Reproducibility of log k_{max} values for PC and Set A tested in the lead labs and in the naive labs. Shown are averages and standard deviations of 4 runs (lead lab) labs, Table 3) and 11 runs (naïve labs, Table 13).

For illustration of the kinetic matrix of depletion values obtained, Figure 4 shows the results of all seven labs for the chemical DNCB. On the x-axis the concentration is plotted and on the y-axis the term Ln(100-depletion) is plotted. The slope from these plots gives rate constants (k_{obs}) at different time points (see Figure 1), which can be divided by the time to arrive at the rate constant calculated by the evaluation sheet. For each lab the data from the second valid repetition is shown. Overall, the matrix shows that the test at different time points is well reproducible, with some variations observed for Lab E. In this specific case, the k_{max} is calculated for all laboratories from the data recorded at 10 min, as this time-point already gave significant depletion over the cut-off and a highly correlated slope vs. concentration in all labs.

Table 13. Six positive transfer phase chemicals (Set A) tested in five naïve labs: log k_{max} values. Values written in red were not in the proficiency range defined by the experienced labs. Values written in blue were only slightly outside the proficiency range and can generally considered to be acceptable. * The range for proficiency for the chemicals was based on the phase I data of Givaudan and BASF and cor-responds to the mean of four experiments ± 2 × standard deviation.

Transfer Labs	Pre-de-fined pro-ficiency range*	Lab D Run 1	Lab D Run 2	Lab F Run 1	Lab F Run 2	Lab C Run 1	Lab C Run 2	Lab E Run 1	Lab E Run 2	Lab E Run 3	Lab G Run 1	Lab G Run 2	Average	SD
2,4-Dinitrochloro-benzene	(-0.67) – (-0.44)	-0.57	-0.52	-0.54	-0.64	-0.78	-0.60	-0.41	-0.41	-0.41	-0.55	-0.67	-0.55	0.12
Oxazolone	(-0.23) – (-0.02)	-0.03	-0.13	-0.08	-0.20	-0.14	-0.28	-0.13	-0.15	-0.14	-0.13	-0.17	-0.14	0.07
Formaldehyde	(-0.92) – (-0.44)	-0.60	-0.73	-0.69	-0.63	-0.64	-0.69	-2.35	-2.37	-2.38	-0.71	-0.73	-1.14	0.79
Ethylene glycol di-methacrylate	(-2.79) – (-2.20)	-2.55	-2.40	-2.60	-2.38	-2.71	-2.53	-2.20	-2.20	-2.20	-2.52	-2.56	-2.44	0.18
Benzylideneacetone	(-1.99) – (-1.66)	-1.86	-1.67	-1.57	-2.05	-1.91	-1.83	-1.88	-1.75	-1.76	-1.83	-1.70	-1.80	0.13
2,3-Butanedione	(-3.20) – (-2.20)	-2.07	-3.19	-2.51	-2.92	-3.02	-3.02	-2.15	-2.15	-2.14	-2.14	-2.84	-2.56	0.44
Formaldehyde w/o Lab E		-0.60	-0.73	-0.69	-0.63	-0.64	-0.69				-0.71	-0.73	-0.68	0.05

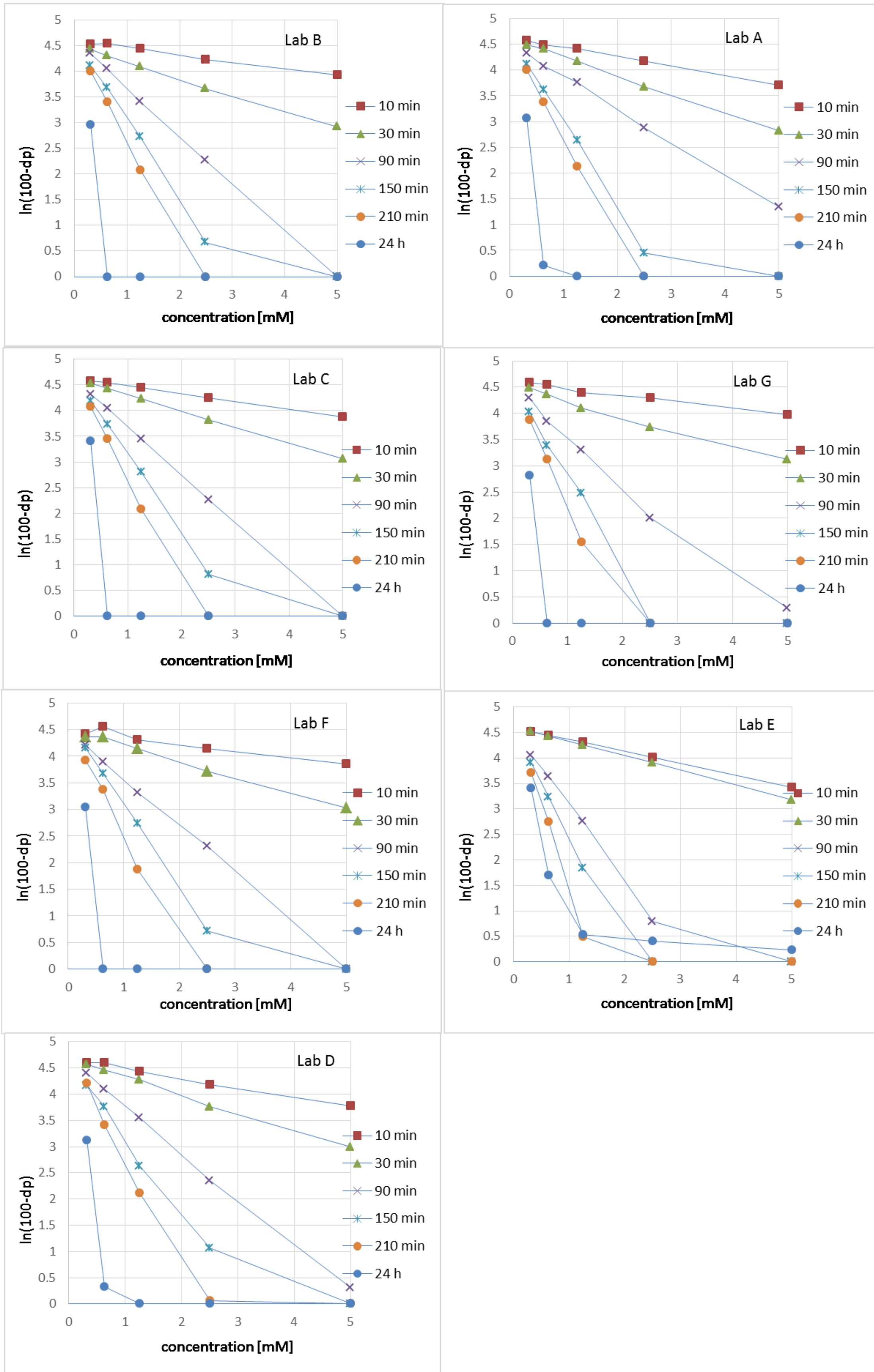


Figure 4. Depletion matrix for DNCB in the second valid run from all seven laboratories. The natural logarithm of the non-depleted peptide concentrations is plotted vs. the concentration of the test chemical at each time point.

The range for proficiency for the chemicals was based on the phase I data of Givaudan and BASF and corresponds to the mean of four experiments ± 2 × standard deviation. Hence, they were defined based on a small data base and the range defined is strongly dependent on how close the repetitions in these two labs were. The minimal deviations, marked in blue color in table 13, are attributed to this limited data available and the low variability observed at the lead labs which led to definition of a very narrow range at this moment in time and these runs marked in blue are thus considered still to be suitable to demonstrate proficiency of the respective labs.

The highly congruent data between the proficient labs and the naïve labs is also illustrated by the very similar average kinetic rates and standard deviations for testing set A in lead labs and naïve labs in Figure 3.

5.3.4. Reproducibility of predicted GHS categories

Tables 14 and 15 show the same results as Table 13 expressed as predicted GHS categories using the published and the refined cut-off, respectively. Consistent classifications were obtained and congruent with data from the experienced labs. The only exception is again formaldehyde which is placed in category 1B in the case of Lab E. As for the lead labs, benzylideneacetone moved to Cat 1A with the refined prediction model in all repetitions in the different labs with the exception of one repetition in Lab F.

Table 14. Six positive transfer phase chemicals (Set A) tested in five naïve labs: Prediction by the published prediction cut-off (log k_{max} -1.1). Values written in blue were only slightly outside the proficiency range and can generally considered to be acceptable.

	Lab D Run 1	Lab D Run 2	Lab F Run 1	Lab F Run 2	Lab C Run 1	Lab C Run 2	Lab E Run 1	Lab E Run 2	Lab E Run 3	Lab G Run 1	Lab G Run 2
2,4-Dinitrochlorobenzene	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A
Oxazolone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A
Formaldehyde	1A	1A	1A	1A	1A	1A	1B	1B	1B	1A	1A
Ethylene glycol dimethacrylate	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B
Benzylideneacetone	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B
2,3-Butanedione	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B

Table 15. Six positive transfer phase chemicals (Set A) tested in five naïve labs: Prediction by the refined prediction cut-off (log k_{max} -2.0). Values written in blue were only slightly outside the proficiency range and can generally considered to be acceptable.

	Lab D Run 1	Lab D Run 2	Lab F Run 1	Lab F Run 2	Lab C Run 1	Lab C Run 2	Lab E Run 1	Lab E Run 2	Lab E Run 3	Lab G Run 1	Lab G Run 2
2,4-Dinitrochlorobenzene	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A
Oxazolone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A
Formaldehyde	1A	1A	1A	1A	1A	1A	1B	1B	1B	1A	1A
Ethylene glycol dimethacrylate	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B
Benzylideneacetone	1A	1A	1A	1B	1A	1A	1A	1A	1A	1A	1A
2,3-Butanedione	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B

5.3.5. Results for positive control

Table 16 summarizes the PC results in the valid runs in the different labs. The results demonstrate a low standard deviation for the values determined at 90 min or 150 min. The 90 min value is used to conclude on validity of the run for the PC, as it is in the linear range for cinnamic aldehyde. The 150 min value can be used if linearity and statistical criteria for rate constant calculations are not fulfilled at 90 min. Overall, very similar and almost identical summary results for all runs and standard deviations were observed as reported in Table 6 for the lead labs.

5.3.6. Discussion on transfer phase to naïve labs

The test was easily transferable to five naïve labs without a hands-on training. Four of the five laboratories rapidly were proficient and no or only one experiment was not proficient based on the data for the positive control. The variability of the raw data in the kinetic profiling were low enough to automatically calculate reproducible rate constants. This point has to be emphasized, as in this test a complex data-matrix is automatically reduced by an automatic Excel sheet to a single k_{max} value. Noise in the data or variability between replicates and non-linear concentration-responses preclude these automatic calculations to be reliable. The data from the transfer phase indicate that this was not a major hurdle for the chemicals tested here.
Figure 3 indicates that the 6 test chemicals first tested in the two lead laboratories gave very similar results when tested in the 5 naïve labs. Moreover, the standard deviation from the four experiments in the two lead labs is similar to the standard deviation in the 11 experiments in the 5 naïve labs, and thus the variability was not significantly increased by moving from the lead labs to the naïve labs.

Based on these observations, namely:

- Rapid adoption without hands-on training
- Successful automatic data evaluation to derive k_{max} values
- Reproducible absolute k_{max} values obtained by lead and naïve labs
- No increased variability and noise in the data in the naïve labs
- Reproducible quantitative results for positive control in all experiments

We can conclude that the kDPRA is a readily transferable method which can be easily adopted by new laboratories based on the SOP. Hence the blinded phase was immediately started after completion of this transfer phase.

The only changes made were further changes to the evaluation sheet to control for variability of the solvent control and to encode alerts in case fluorescence quenching or autofluorescence is observed, or in case the depletion is not linearly progressing over time and data gaps in the linearity are observed. No changes to the experimental manipulation / experimental protocol were made.

Table 16. Results for the positive control of the valid runs in the transfer phase to naïve labs.

Run	log k_{90}	log k_{150}	log k_{max}	Classification published cut-off	Classification refined cut-off
LAB D-2	-1.46	-1.62	-1.43	GHS Cat 1B	GHS Cat 1A
LAB D-4	-1.60	-1.74	-1.60	GHS Cat 1B	GHS Cat 1A
LAB D-5	n/a	-1.80	-1.47	GHS Cat 1B	GHS Cat 1A
LAB D-6	n/a	-1.68	-0.93	GHS Cat 1A	GHS Cat 1A
LAB D-7	n/a	-1.70	-1.70	GHS Cat 1B	GHS Cat 1A
LAB D-8	-1.68	-1.58	-1.14	GHS Cat 1B	GHS Cat 1A
LAB D-9	-1.66	-1.75	-1.66	GHS Cat 1B	GHS Cat 1A
LAB D-11	-1.62	-1.70	-1.36	GHS Cat 1B	GHS Cat 1A
LAB F-1	n/a	-1.76	-1.76	GHS Cat 1B	GHS Cat 1A
LAB F-2	-1.58	-1.73	-1.58	GHS Cat 1B	GHS Cat 1A
LAB F-4	-1.56	-1.74	-1.56	GHS Cat 1B	GHS Cat 1A
LAB F-5	-1.63	-1.67	-1.43	GHS Cat 1B	GHS Cat 1A
LAB F-6	-1.59	-1.70	-1.59	GHS Cat 1B	GHS Cat 1A
LAB F-7	-1.64	-1.65	-1.52	GHS Cat 1B	GHS Cat 1A
LAB C-1	-1.61	-1.72	-1.29	GHS Cat 1B	GHS Cat 1A
LAB C-2	-1.63	-1.69	-1.28	GHS Cat 1B	GHS Cat 1A
LAB C-3	-1.67	-1.70	-1.64	GHS Cat 1B	GHS Cat 1A
LAB C-4	-1.55	-1.66	-1.19	GHS Cat 1B	GHS Cat 1A
LAB C-5	-1.54	-1.65	-1.54	GHS Cat 1B	GHS Cat 1A
LAB E-1	-1.69	-1.68	-1.18	GHS Cat 1B	GHS Cat 1A
LAB E-2	-1.53	-1.81	-1.15	GHS Cat 1B	GHS Cat 1A
LAB E-3	-1.70	-1.67	-1.15	GHS Cat 1B	GHS Cat 1A
LAB E-4	-1.67	-1.87	-1.19	GHS Cat 1B	GHS Cat 1A
LAB E-5	-1.69	-1.68	-1.18	GHS Cat 1B	GHS Cat 1A
LAB E-6	-1.68	-1.83	-1.18	GHS Cat 1B	GHS Cat 1A
LAB G-1	-1.51	-1.58	-1.51	GHS Cat 1B	GHS Cat 1A
LAB G-2	-1.60	-1.62	-1.23	GHS Cat 1B	GHS Cat 1A
LAB G-3	-1.63	-1.72	-1.34	GHS Cat 1B	GHS Cat 1A
LAB G-4	-1.53	-1.61	-1.50	GHS Cat 1B	GHS Cat 1A
LAB G-5	-1.60	-1.64	-1.37	GHS Cat 1B	GHS Cat 1A
LAB G-6	-1.61	-1.68	-1.61	GHS Cat 1B	GHS Cat 1A
LAB G-7	-1.61	-1.71	-1.33	GHS Cat 1B	GHS Cat 1A
mean	**-1.61**	**-1.70**	**-1.39**		
SD	**0.06**	**0.07**	**0.20**		
min	**-1.70**	**-1.87**	**-1.76**		
max	**-1.46**	**-1.58**	**-0.93**		

5.4. WP 3.4 Blind-coded Testing – Phase II

5.4.1. Goals and set-up

- Based on experience from phase I, several modifications to the SOP and evaluation template were made:
  - Treatment of data in specific circumstances (interrupted time course, quenching, autofluorescence)
  - Validity criteria added (maximal variance of solvent control added, acceptance range for PC was slightly expanded)
- Test chemicals for this phase were sent blinded. An external lab (BioTeSys GmbH, Schelztorstrasse 54-56, 73728 Esslingen, Germany) performed procurement, blinding and sent chemicals to the different labs.
- For this phase, all labs worked with the same peptide quality (> 95% purity, obtained from a single batch synthesized by Genscript); except Lab G and Lab F who did use their own peptide. For Lab G this choice was made as they joined the study later when this common batch was already fully distributed, while lab F found poor solubility with the peptide quality delivered, which may be due to a customs issue – the peptide traveled twice to this US lab as it was rejected by customs clearance, hence the long transit time may have affected quality. As can be seen in the results, no effect was noted whether the common batch or different quality was used.
- All seven labs tested 24 blinded chemicals in one repetition (inter-laboratory reproducibility)
- Out of these 24 chemicals a random subset of 12 chemicals were tested in additional two repetitions in three or four labs (additional intra-laboratory reproducibility). Thus, for 12 chemicals this intra-laboratory comparison was in 3 labs and for the remaining 12 chemicals in 4 labs (in total $84 = 12 \times 3 + 12 \times 4$ intra-laboratory comparisons).
- The full study setup is shown in Appendix in Table A10.
- All test chemicals for intra-laboratory repetitions were also sent as blind-coded, independent samples and thus the laboratories were not aware of the intra-laboratory repetitions, they just received a total of 48 blind-coded test chemicals.
- Note: Here is a small deviation from the original study plan: It was initially planned that the chemicals are separated into two sets of 12 and then two groups of labs would each test a set of 12. Here the blinding-lab BioTeSys introduced a further randomization by attributing the chemicals for intra-laboratory testing randomly to the laboratories and not as pre-defined sets.
- The phase study II was not run under GLP, but under GLP-like conditions. All the labs had to record all steps (like weighting of chemicals, solvent selection) in predefined reporting templates. An SOP defined according to GLP-like standards was used and data evaluation was standardized.
- After the study was completed, all data were collected in a grand summary Excel sheet which is linked directly to the original source files, hence all data can be fully tracked back to the original files provided by the labs.
- The data were sent to BioTeSys GmbH prior to de-coding of the chemicals. Thus, the data as used for the biostatistics are safely stored at an external independent site before decoding.
- Analysis of and biostatistics on the decoded data were then conducted at one of the lead laboratories (Givaudan).

5.4.2. Rationale for test chemical selection for Phase II

Chemicals for the validation (transferability and reproducibility) were mainly selected from the ICCVAM list on LLNA potency validation [6]: These 24 chemicals are presumably well characterized for potency in mice and humans. Selected chemicals were also checked for congruent rating in the compilation by Basketter et al. [7]. They include:

a. All chemicals classified as GHS Cat 1A in both mice and humans. Some chemicals from this list are already tested in the transferability phase and the remaining chemicals are tested in the reproducibility phase (n = 7).
b. 5 chemicals in ICCVAM list classified as GHS Cat 1A in humans only, but not in the LLNA. This allows additional evaluation of the assay for human potency assessment. (incongruent results of LLNA vs. human)
c. 10 chemicals classified as GHS Cat 1B in humans and in the LLNA.
d. 2 chemicals rated negative in mice and humans

These 24 chemicals include 2 correct negatives and 21 correct positives in the Cys-only DPRA according to literature data. The test set intentionally also includes 1 sensitizer known to be negative / minimally reactive in the DPRA (dihydrocoumarin). [This set also included one chemical already tested in the transfer phase (benzylidene acetone, 122-57-6): This was not intentionally planned, but rather a mistake when setting up the validation set. However, this helped to double check for one chemical whether the result obtained is different when moving to blinded assessment.]
The 24 chemicals are given in Table 17 including their chemical diversity as expressed by reaction mechanistic domain assignment. This set is strongly biased for positive chemicals (sensitizers) – this is with clear intention, as the kDPRA is intended for potency discrimination within chemicals rated reactive. Two negative chemicals are nevertheless included to also assess reproducibility for non-reactive chemicals.

Table 17. Chemicals for the reproducibility (inter- and intralab) assessment (phase II); l = liquid; s = solid; *k_{max} in this table is calculated based on final concentration, while in the original paper [4] stock concentrations were used for calculation.

Nr.	Name	Abrevia-tion	Physical form	CAS	LLNA EC 3 [%]	ICCVAM human potency (DSA 05; µg/cm²)	Consolidated GHS LLNA	Consolidated GHS Human	Cys-only DPRA prediction	k_{max} from [4]*	Consolidated Cys-depletion [%]	Mechanistic domain consolidated
1	(Chloro)methylisothiazolinone	CMI	l	26172-55-4	0.009	5	1A	1A	Correct positive	as Kathon CG - 0.12	96.3	SN_2-reaction
2	Glyoxal	GLY	l	107-22-2	1.4	345	1A	1A	Correct positive	not tested	56.5	Schiff base
13	Methylisothiazolinone	MI	s	2682-20-4	1.9	223.5	1A	1A	Correct positive	not tested	97.9	SN_2-reaction at the S-atom proposed
14	Methyl-2-octynoate	MOY	l	111-12-6	0.45	388	1A	1A	Correct positive	not tested	97.2	Michael Acceptor
3	4-phenylenediamine	PPD	s	106-50-3	0.16	30	1A	1A	Correct positive	-0.70	95.3	Quinone methide(s)/imines
15	Tetrachlorsalicylanilide	TCS	s	1154-59-2	0.04	27	1A	1A	Correct positive	not tested	36.8	Acyl Transfer
4	Isoeugenol	IE	l	97-54-1	1.8	1016	1A	1A	Correct positive	-1.10	92.6	Quinone methide(s)/imines
16	Bourgeonal	BOU	l	18127-01-0	4.3	1541	1B	1B	Correct positive,	not tested	17.7	Schiff base
5	Carvone	CAR	l	6485-40-1	13 / 10.7	19284	1B	1B	Correct positive	not tested	25.7	Michael Acceptor
17	Dihydrocoumarin	DHC	l	119-84-6	5.6	759	1B	1B	False-negative	not tested	0.0	Acyl Transfer
6	Hydroxycitronellal	HC	l	107-75-5	23	5237	1B	1B	Correct positive	-1.90	32.3	Schiff base
18	Imidazolidinyl urea	IU	s	39236-46-9	24	3846	1B	1B	Correct positive	-1.45	38.4	Acyl Transfer
7	Methylhexanedione	MHD	l	13706-86-0	25.8	3595	1B	1B	Correct positive	not tested	25.8	Schiff base
19	Perillaldehyde	PA	l	2111-75-3	8.1	1484	1B	1B	Correct positive	not tested	31.9	Michael acceptor / Schiff base
8	Phenyl benzoate	PB	s	93-99-2	17.1	52489	1B	1B	Correct positive	-2.22	50.9	Acyl Transfer
20	Phenylpropionaldehyde	PPA	l	93-53-8	6.3	692	1B	1B	Correct positive	not tested	37.4	Schiff base
9	Tetramethyldiuram disulfide	TMD	s	137-26-8	5.2	4544	1B	1B	Correct positive	-1.62	99.5	SN_2-reaction at the S-atom proposed
21	Benzisothiazolinone	BI	s	2634-33-5	2.3	50	1B	1A	Correct positive	not tested	97.7	SN_2-reaction at the S-atom proposed
10	Benzylidene acetone	BA	s	122-57-6	3.7	299	1B	1A	Correct positive	-1.77	93.5	Michael Acceptor
22	δ-Damascone	DAM	l	57378-68-4	9.6 / 0.9 / 5.2	193	1B	1A	No data	not tested	No data	Michael Acceptor
11	Diethylmaleate	DEM	l	141-05-9	2.1	400	1B	1A	Correct positive	-1.16	99.9	Michael Acceptor
23	trans-2-hexenal	HEX	l	6728-26-3	5.5	49	1B	1A	Correct positive	not tested	97.9	Michael acceptor / Schiff base
12	4-methoxy-acetophenone	MAP	s	100-06-1	>50	Non-sensitiser	NC	unknown	correct-negative	not tested	2.4	No alert
24	Chlorobenzene	CB	l	108-90-7	>25	Non-sensitiser	NC	unknown	correct-negative	not tested	0.4	No alert

5.4.3. Overview of the different data evaluations performed:

- Since the continuous variable log k_{max} may be used beyond stand-alone application for classification in a DA with a DIP, the quantitative accuracy of the derived rate constants was evaluated in detail:
  - o Average and standard deviation were calculated for log-transformed kinetic rate constants and all values are tabulated
  - o The variability within labs was compared to the variability between labs to evaluate how much of the variability comes from intrinsic test variability and how much comes from inter-laboratory variability

- Data were evaluated for reproducibility and transferability when applying a fixed cut-off for classification.
  - o This was (i) the cut-off from the literature [4] (log k_{max} = -1.1) and (ii) a refined cut-off (log $k_{max\,=}$ -2.0) coming from a more extended chemical dataset with the refined protocol (see section 3.6).

- For inter-laboratory comparisons of the log k_{max} values, averages of the three values obtained within the labs testing for intra-laboratory variability were first calculated, so that only one value was compared for each lab. Similarly, the classification prediction for each lab was made based on this average log k_{max} value for the chemicals tested in three instances when they were used for inter-laboratory comparison. Thus, seven values were compared in the inter-laboratory comparisons (see Scheme 1).

- For inter-laboratory comparison of binary GHS subcategorization, the data were treated as two independent datasets: (i) the results from the laboratories testing the chemicals only once (ii) the results from the laboratories testing the chemicals three times. This is because a classical validation study for classification of chemicals is normally performed on 3 labs and outcome of three labs is tested for congruency. Hence it does not make sense to test for congruent results in 7 labs as an indicator of reproducibility of class attribution. Splitting the data in two sets is thus justified and enhances the power with two independent evaluations. Furthermore, we do not mix the data coming from the average of three repetitions with the data coming from one repetition (see Scheme 1).

- The read out from the classical DPRA is depletion at 24 h with 5 mM test chemical concentration assessed with HPLC. The 24 h / 5 mM data point is also included in the kDPRA. While this is not a primary output of the kDPRA, we still evaluated these results and their intra- and inter-laboratory variability. This allows a direct linkage of the kDPRA to the DPRA and indicates to which extent these results (obtained by fluorescent vs. HPLC detection and in microplates vs. glass vials) are congruent with each other and how much variability of results is different between the two approaches.

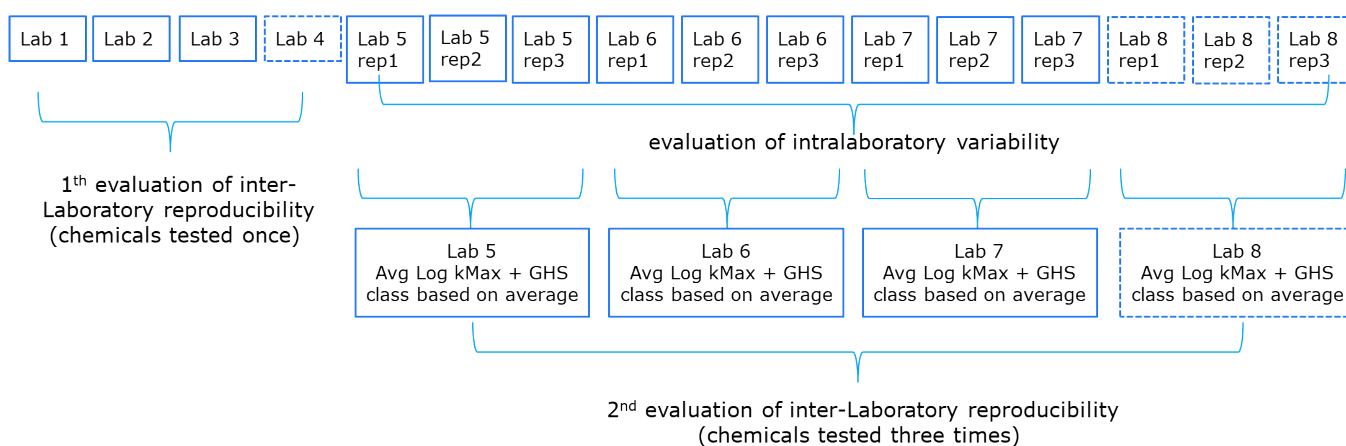
- For inter-laboratory comparison of solvent selection, the data were treated as two independent datasets as described for comparison of class attribution above. The three solvent decisions obtained within the labs testing for intra-laboratory variability were summarized to one vehicle in case of consistent intra-lab decision or to n/a in case of inconsistent intra-lab decision.

Note: 12 chemicals were tested in 4 labs for inter-laboratory reproducibility only, while 12 chemicals were tested additionally in 4 labs for intra-laboratory variability. This was fully randomized, so no laboratory received the same chemicals for intra-laboratory reproducibility. The remaining chemicals were tested in 3 labs each for intra- and inter-laboratory reproducibility.
Below all data are reported as 'Laboratory 1-4' for the inter-laboratory only experiments while the code 'laboratory 5 - 8' is assigned to the labs performing intra-laboratory experiments on that particular chemical. The value for laboratory 4 or 8 is empty accordingly for each 12 chemicals.

Hence Laboratory 1 is a different laboratory for each chemical and the codes 1 - 8 do not refer to a specific laboratory but are default numbers to label the experiments / labs per chemical.

Only where it is needed to test for the effect a specific laboratory may have on the results, Laboratory A – G are indicated, referring to specific testing laboratories. This approach to data display simplifies the complex data matrix for viewing. Certainly, all the data can be tracked back to the originating lab in the underlying Excel sheets.

The full set-up is also displayed in Table A10.



Scheme 1: Illustration of the inter-laboratory data evaluation

5.4.4. Intra-laboratory reproducibility of log k_{max}

Table A1 in Appendix 1 indicates individual log k_{max} values obtained for the different labs when retesting the same chemical repeatedly in a blinded manner, while Table A2 indicates the standard deviations in the individual labs. The average and the standard deviations are also shown for all the labs in Figure 5 (chemicals tested in 4 labs for intra-laboratory reproducibility and Figure 6 (chemicals tested in 3 labs).

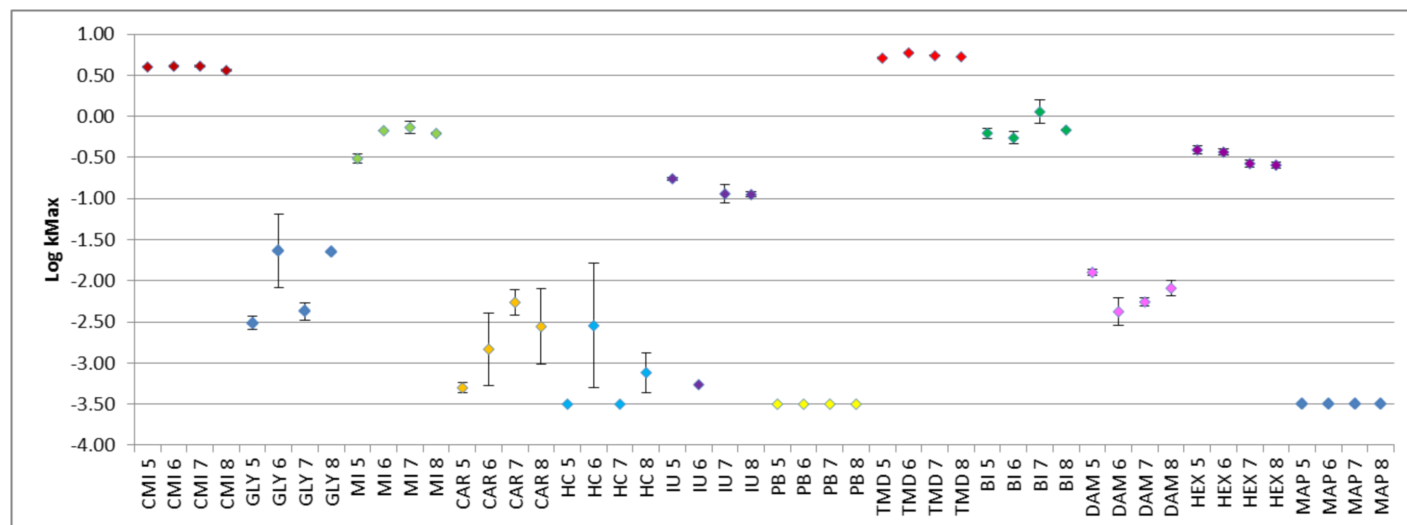


Figure 5. Intra-laboratory testing: Variability expressed as average values and standard deviation in repeated intra-laboratory testing (3 times each) in 4 labs. For chemicals not reactive (log $k_{max} < -3.46$) a default value of -3.5 was indicated to allow plotting the results. Abbreviated chemical names (see table 18) and default laboratory number are indicated on the x-axis.

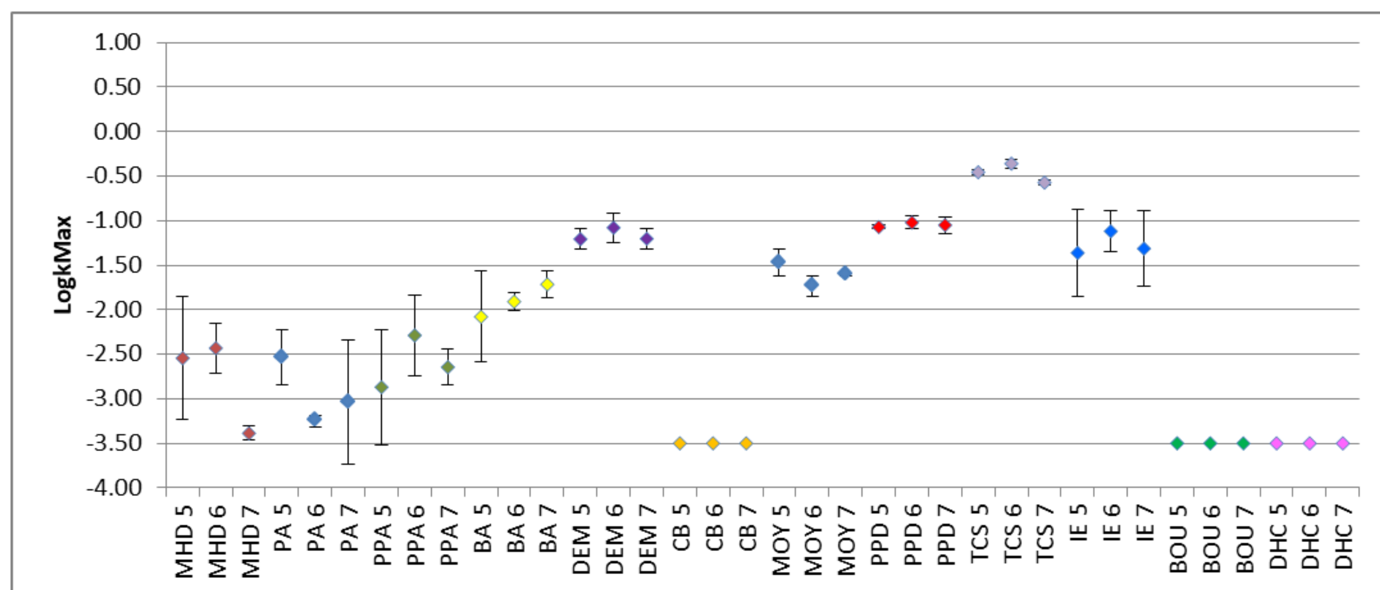


Figure 6. Intra-laboratory testing: Variability expressed as average values and standard deviation in repeated intra-laboratory testing (3 times each) in 3 labs. For chemicals not reactive (log $k_{max} < -3.46$) a default value of -3.5 was indicated to allow plotting the results. Abbreviated chemical names (see table 17) and default laboratory number are indicated on the x-axis.

To express variability in validation trials, standard deviations are often expressed as coefficient of variation (CV). On the logarithmic scale of the log k_{max} this would be meaningless. To estimate the size of the variability, it has to

be kept in mind that the dynamic range of the log k_{max} is between -3.46 (chemical with 13.89% Cys-depletion at 5 mM after 24 h) and ca. +0.75 (98% depletion already after 10 min also at the lowest concentration). Thus, the size of the standard deviation should be viewed within this dynamic range which is also given in all the graphs (plotted from -4 to +1). A standard deviation of 0.3 in the logarithmic scale indicates a two-fold difference in kinetic rate.

For most chemicals, the intra-laboratory variability is low and the standard deviations are very small (see Table A2 in Appendix and Table 18). For chemicals with very low standard deviation (below 0.3), the values reported from the different laboratories are also very close to each other. For four of the 24 chemicals (CAR, MHD, PPA IE), intra-laboratory variability is higher (between 0.35 and 0.45), but this increased variability is then also observed in multiple labs and it appears to be intrinsic to certain chemicals as will also be illustrated by the inter-laboratory analysis reported below.

The average standard deviation of the 24 chemicals in intra-laboratory testing was at 0.158, and the average was < 0.1 for 10 chemicals, with further five chemicals being non-reactive in all laboratories. These include the two non-sensitizers 4-methoxy-acetophenone and Chlorobenzene and 3,4-dihydrocoumarin, which are also non-reactive in DPRA. In addition, phenylbenzoate and bourgeonal were reproducibly non-reactive in kDPRA despite the fact that-Cys-depletion had been reported in the DPRA.

One key question is, whether some laboratories have a higher intrinsic variability as compared to other laboratories, hence Table 18 gives the standard deviations sorted by actual labs (A – G) testing the different chemicals repeatedly. Indeed, it appears that some laboratories have a higher variability (Lab A, C and F) than others (Lab B, E, G), while Lab D appears in the middle. However, this is to a large part due to the fact that labs A, C and F received chemicals which overall are prone to higher variability. Thus, lab F with the highest variability has received CAR, IE, PA and PPA for intra-laboratory testing, which are chemicals with generally higher variability. Thus, we cannot, from the randomized setup of the study, conclude that specific laboratories had a higher variability. This can also be seen form analysis of the variability of the positive control which was not particularly high in laboratories A, C and F (Tables 24 – 26).

While the exact reason for enhanced variability for specific chemicals is not known, some explanations can be given. Very slowly reacting chemicals like CAR and MHD may have higher variation, as variation over prolonged incubation time may be cumulative. Chemicals triggering peptide oxidation like PPA are known to be subject to higher variability. For PPA this had already been observed in the DPRA pre-validation study. Finally, pre-haptens spontaneously oxidizing like IE may be more variable, as autoxidation is known to be a self-catalyzed process and hence prone to more stochastic effects.)

Table 18. Standard deviation (SD) of log k_{max} in repeated testing of 24 chemicals in different laboratories sorted by actual test lab

Abbreviation	Substance	SD Lab A	SD Lab B	SD Lab C	SD Lab D	SD Lab E	SD Lab F	SD Lab G
CMI	(Chloro)methylisothiazolinone	0.006	0.014			0.011		0.013
GLY	Glyoxal			0.104		0.077	0.450	0.015
MI	Methylisothiazolinone			0.076	0.003	0.061		0.015
MOY	Methyl-2-octynoate	0.148	0.118				0.021	
PPD	4-phenylenediamine		0.026	0.093	0.074			
TCS	Tetrachlorsalicylanilide	0.030	0.051		0.030			
IE	Isoeugenol			0.418	0.232		0.485	
BOU	Bourgeonal	n-r	n-r	n-r				
CAR	Carvone				0.154	0.057	0.444	0.460
DHC	Dihydrocoumarin	n-r	n-r		n-r			
HCA	Hydroxycitronellal			n-r		n-r	0.757	0.241
IU	Imidazolidinyl urea		0.017			n-r	0.114	0.030
MHD	Methylhexanedione	0.690		0.082	0.283			
MHD	Perillaaldehyde			0.691	0.064		0.307	
PA	Phenyl benzoate	n-r	n-r			n-r		n-r
PPA	Phenylpropionaldehyde	0.649		0.207			0.449	
TMD	Tetramethyldiuram disulfide		0.008			0.011	0.004	0.001
BI	Benzosiothiazolinone	0.062			0.145	0.072		0.006
BA	Benzylidene acetone	0.512	0.101		0.154			
DAM	Delta Damascone			0.050		0.042	0.162	0.088
DEM	Diethylmaleate	0.111			0.118		0.162	
HEX	trans-2-hexenal		0.048	0.046		0.036		0.033
MAP	4-methoxy-acetophenone		n-r	n-r		n-r		n-r
CB	Chlorobenzene	n-r			n-r		n-r	
	Average all chemicals	**0.276**	**0.048**	**0.196**	**0.126**	**0.046**	**0.305**	**0.090**

n-r: non-reactive

5.4.5. Inter-laboratory reproducibility of log k_{max}

Table A3 gives all the k_{max} values for inter-laboratory testing. For Laboratories 5 – 8 (testing the particular chemical in 3 intra-laboratory repetitions) the average of the repeated testing is used for the inter-laboratory comparison. In addition, Table A3 reports also the average of all experiments, i.e. taking the individual repetitions of the Lab 5 – 8 as individual experiments. Figures 7 and 8 indicate the inter-laboratory variability from the 7 laboratories for all 24 chemicals.

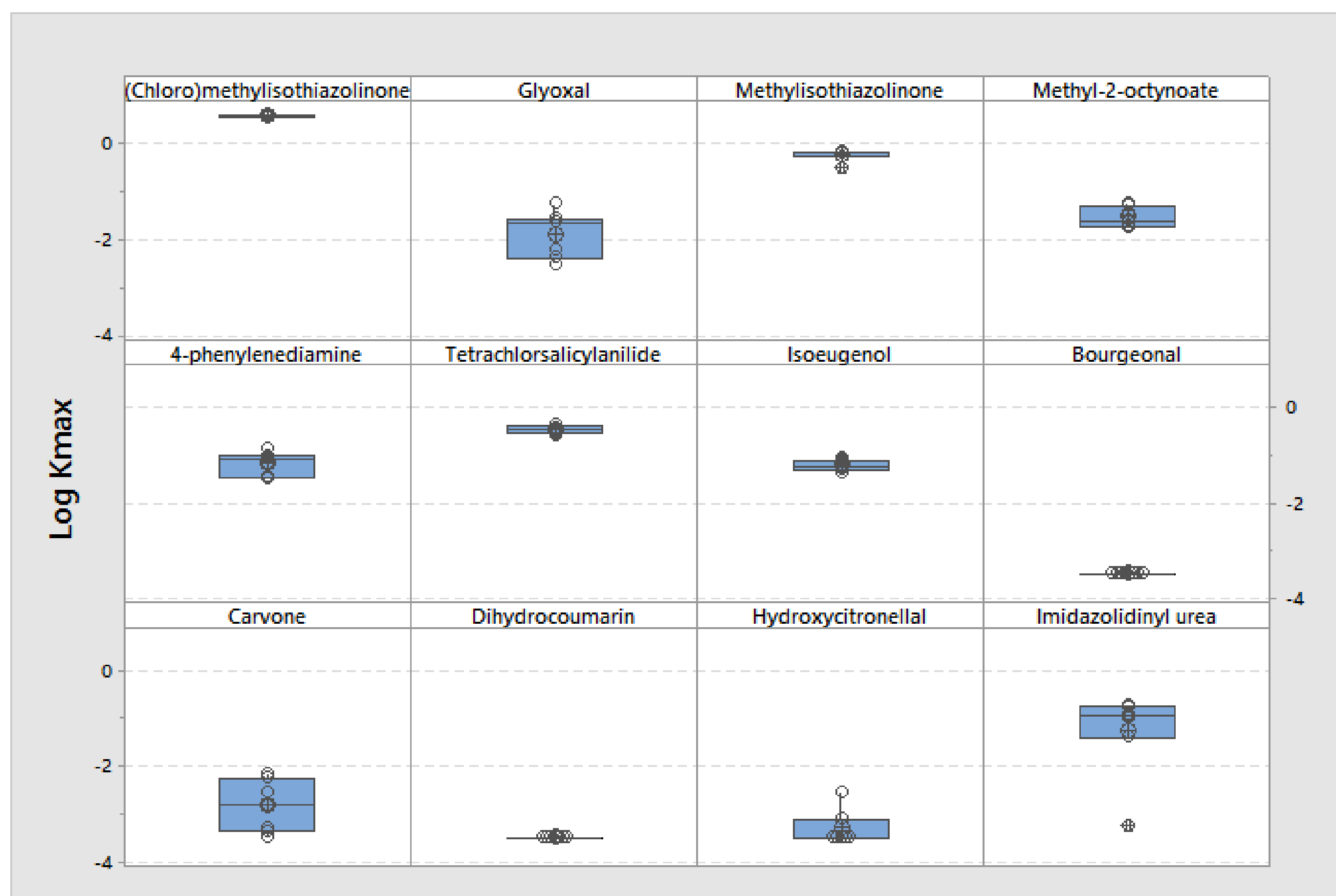


Figure 7. Log k_{max} values from inter-laboratory testing, Chemicals 1-12. For laboratory 5 – 8, the average from repeated testing is plotted. Shown are the 7 individual lab results (circles), the interquartile range box (blue) and the average (horizontal line)

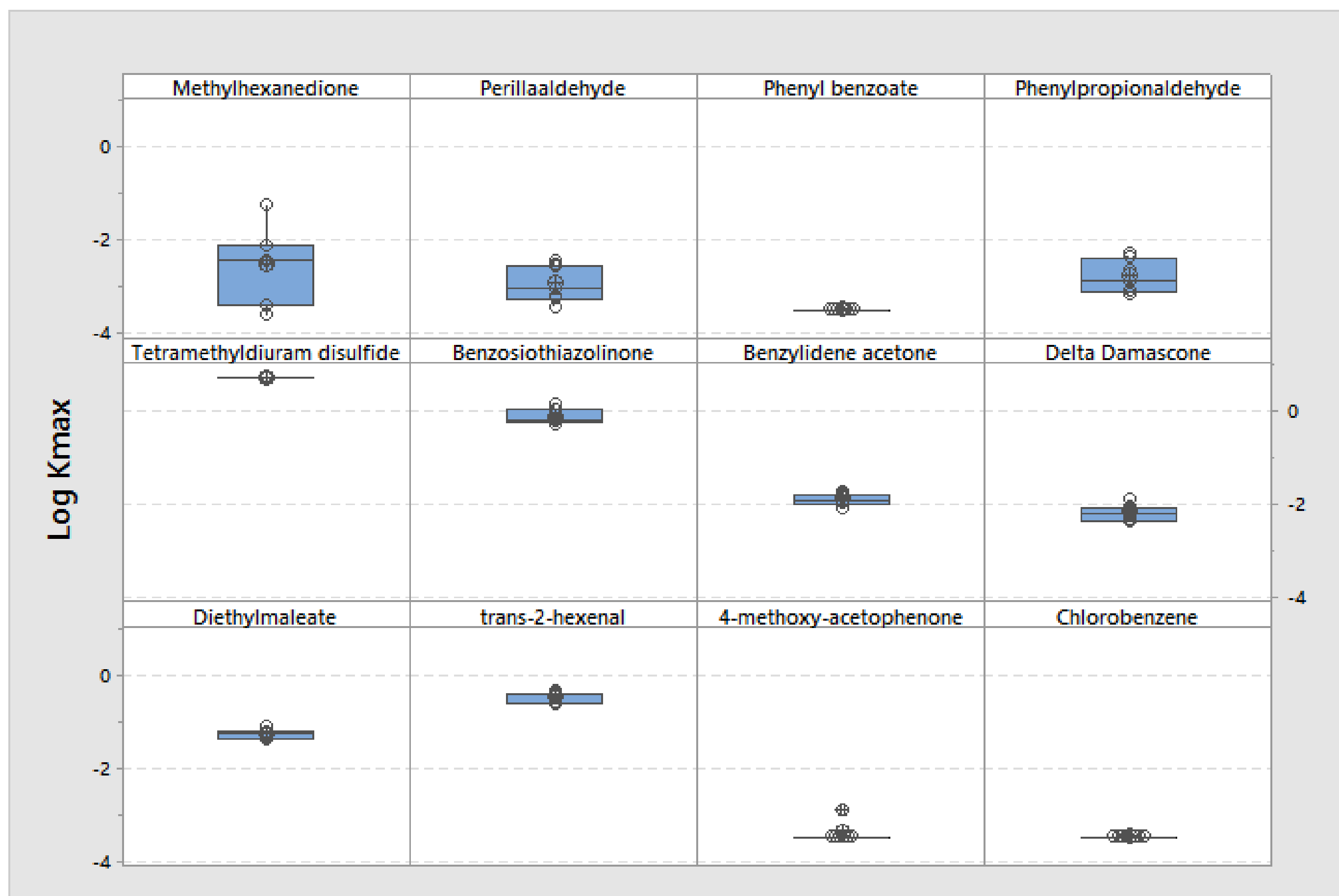


Figure 8. Log k_{max} values from inter-laboratory testing, Chemicals 13-24. For laboratory 5 – 8, the average from repeated testing is plotted. Shown are the 7 individual lab results (circles), the interquartile range box (blue) and the average (horizontal line).

As can be seen in the summary Table A3 and in Figures 7 and 8, for most chemicals the inter-laboratory reproducibility is high (little spread of the values around the mean), and the average standard deviation for inter-laboratory comparison is at 0.244. In general, similar chemicals which had higher variability in intra-laboratory testing also exhibited higher inter-laboratory variability (GLY, CAR, MHD, PPA), indicating that it is an intrinsic property of the chemical and not experimental variability due to the experimental procedures.

There is one significant outlier in the whole data-set: Lab E reported very low / no reactivity (in repeated intra-lab testing) for IU (table 18), which is a known formaldehyde releaser. Interestingly this lab had reported a strong outlier (very weak reactivity) for formaldehyde in the transfer phase – thus for some unknown reason this lab obtains different results for chemicals related to formaldehyde. This outlier is very consistent for IU and formaldehyde and appears to be linked to the chemistry. Formaldehyde does form a reversible peptide-adduct, and for unknown reasons the reaction must have been reversed prior to reaction / during the reaction with mBrB in that particular laboratory.

Benzylidene acetone was tested both in the transfer phase and in the blind-coded phase. The log k_{max} value from 7 labs in the blind coded phase was at -1.89 ± 0.13, while it was at -1.80 ± 0.12 in the transfer phase, hence a very similar result and similar variability is observed in both phases.

5.4.6. Comparison of intra-laboratory and inter-laboratory reproducibility of log k_{max}

One key question in a validation study is, to what extent the variability is inherent to the test and how much variability is contributed by the testing laboratory. Thus, we compared the variability of intra-laboratory testing to the variability of inter-laboratory testing.

For each chemical, the intra-laboratory standard deviation from each of lab 5 - 8 was compared to the inter-laboratory standard deviation of lab 1-4. Thus, it is compared what is the variability if one would test one chemical three times in one lab vs. the variability if it is tested once in 3 – 4 independent and different labs. In each case we calculate the ratio:

$$Ratio = \frac{Stdev\ (\ interlab\ 1{-}4)}{Stdev\ (\ intralab\ n)}$$

Where n is either of 5 – 8. Thus, for each chemical we have 3 – 4 comparisons for intra- vs. inter-laboratory variability. These values cannot be directly averaged, but the geometric mean has to be taken to equally weight cases with higher intra-laboratory or higher inter-laboratory variability.

As can be expected, for the majority of chemicals the ratio is > 1 (see Table 19), indicating that the variability increases when testing in multiple independent labs. The geometric mean for all chemicals is 2.01, thus indicating, that the variability increases by a factor of two when moving from repeated testing in the same lab to testing in independent labs.

However, this value is influenced by the fact that for some chemicals the intra-laboratory variability is extremely low, in these cases also a very modest inter-laboratory variability leads to a relatively high ratio (e.g. CMI, MOY MI and BI have a ratio of 2.5 – 4.4 despite a low inter-laboratory standard deviation (max. of 0.2).

Table 19. Ratio between inter-laboratory and intra-laboratory standard deviation

	Ratio inter-lab 1-4 vs intra 5	Ratio inter-lab 1-4 vs intra 6	Ratio inter-lab 1-4 vs intra 7	Ratio inter-lab 1-4 vs intra 8	GeoMean ratio inter - intra
(Chloro)methylisothiazolinone	7.008	2.899	3.586	2.965	**3.834**
Glyoxal	6.479	1.109	4.789	34.069	**5.852**
Methylisothiazolinone	0.898	18.099	0.721	3.565	**2.543**
Methyl-2-octynoate	1.620	2.033	11.398		**3.348**
4-phenylenediamine	11.665	4.018	3.205		**5.316**
Tetrachlorsalicylanilide	2.838	1.653	2.830		**2.368**
Isoeugenol	0.147	0.307	0.171		**0.198**
Bourgeonal	n-r	n-r	n-r		**n-r**
Carvone	14.855	1.919	5.537	1.853	**4.135**
Dihydrocoumarin	n-r	n-r	n-r		**n-r**
Hydroxycitronellal	n-r	n-r	n-r	n-r	**n-r**
Imidazolidinyl urea	19.834	n-r	2.986	11.374	**8.766**
Methylhexanedione	1.429	3.484	12.031		**3.913**
Perillaaldehyde	1.571	7.500	0.698		**2.018**
Phenyl benzoate	n-r	n-r	n-r	n-r	**n-r**
Phenylpropionaldehyde	0.561	0.811	1.760		**0.929**
Tetramethyldiuram disulfide	1.158	0.851	2.470	10.391	**2.242**
Benzosiothiazolinone	3.172	2.742	1.365	33.744	**4.474**
Benzylidene acetone	0.186	0.940	0.618		**0.476**
Delta Damascone	3.388	0.872	2.836	1.601	**1.914**
Diethylmaleate	0.560	0.384	0.529		**0.484**
trans-2-hexenal	0.431	0.579	0.451	0.629	**0.516**
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	**n-r**
Chlorobenzene	n-r	n-r	n-r	n-r	**n-r**
Geometric mean all					**2.011**

n-r: non-reactive, no standard deviation can be calculated

5.4.7. Prediction of GHS Categories

All k_{max} values were transformed to classifications with (i) the published tentative cut-off and (ii) with the refined cut-off set based on ROC analysis reported under chapter 3.6.

Note: When attributing chemicals to different GHS categories, Table A4 and A5 and Tables 20 -23 report three classes: (i) “n-r” for non-reactive chemicals negative according to the Cys prediction model of the classical DPRA, (ii) 1A and (iii) 1B for the positive chemicals, discriminated according to the published and the refined cut-offs, respectively. As one potential key application of the kDPRA is to attribute chemicals to Cat 1A after hazard ID by a DA, reproducibility of this Cat 1A attribution is of major interest. Hence, all tables also list consistency of a 2 class prediction model 1A vs 1B/NC. This indicates reliability of GHS Cat 1A attribution in repeated intra- or inter-laboratory testing.

5.4.8. Intra-laboratory reproducibility of predicted GHS categories

a) Published cut-off

Table A4 in the appendix shows the individual classifications for each chemical in three independent intra-laboratory repetitions.

For 15 chemicals, all laboratories reported consistent results with the three-class classifications from 3 repetitions. For 9 chemicals at least one lab found inconsistent ratings over the three experiments when attributing chemicals to three classes.

Overall, of the 84 chemical – lab combinations in intra-laboratory testing (4 × 12 + 3 × 12), consistent results were observed in 72 instances, thus intra-laboratory reproducibility with the published prediction model for three-class attribution (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified) is at 86%.

Using the two class model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified), attribution to GHS Cat 1A was consistent in 79 of 84 instances, hence intra-laboratory reproducibility with the published prediction model for identifying 1A chemicals is at 94%.

Only for isoeugenol all labs reported varied GHS categories. The log k_{max} for this chemical is at -1.21, while the published prediction cut-off is at -1.1, thus the limited reproducibility for class determination is clearly linked to the fact that this chemical is so close to the particular prediction cut-off (borderline result).

b) Refined cut-off

Data for predicted classes according to the refined prediction cut-off are shown in Table A5. For 16 chemicals, all laboratories reported consistent results with the three-class classifications from 3 repetitions. For 8 chemicals at least one lab found inconsistent ratings over the three experiments.

Overall, of the 84 chemical – lab combinations in intra-laboratory testing, consistent results were observed in 74 instances, thus intra-laboratory reproducibility with the refined prediction model for three-class attribution (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified) is at 88%.

Using the two class model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified), attribution to GHS Cat 1A was consistent in 81 of 84 instances, hence intra-laboratory reproducibility with the refined prediction model for identifying 1A chemicals is at 96%.

5.4.9. Inter-laboratory reproducibility of predicted GHS categories

Two sets of results were separately evaluated for inter-laboratory reproducibility of the predicted GHS categories: (i) the data for the 3 – 4 labs testing the chemicals once (Lab 1 - 4, Tables 20 and 22) and (ii) for the 3 – 4 labs testing the chemicals three times (Lab 5 – 8, Tables 21 and 23) (See Scheme 1). The latter evaluation was done by calculating the average log k_{max} values from repeated intra-laboratory testing and then assigning the GHS category according the prediction model (an alternative way would be to make a '2 out of 3' assessment based on the three values, but this was not considered here).

These two evaluations give an independent dataset of inter-laboratory reproducibility for all chemicals, with the second set theoretically a bit more robust as it is based on three repetitions of the experiment in each laboratory.

a) Published cut-off

When applying the published prediction cut-off (log k_{max} = -1.1), for 19 chemicals a consistent result with the three-class classifications (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified) was obtained in Lab 1 – 4 (Table 20) while for 21 chemicals the result is consistent with the three-class classifications in Laboratory 5 – 8 (Table 21). This then gives an inter-laboratory reproducibility for class assignment of 79% and 88%, respectively, with an average for the two independent evaluations of 83%.

Using the two class model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified), for 21 chemicals a consistent result was obtained in Lab 1 – 4 (Table 20) while for 22 chemicals the result is consistent in Laboratory 5 – 8 (Table 21). This then gives an inter-laboratory reproducibility for identifying 1A chemicals of 88% and 92%, respectively, with an average for the two independent evaluations of 90%.

b) Refined cut-off

When applying the refined prediction cut-off (log k_{max} = -2.0), for 20 chemicals a consistent result with the three-class classifications (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified) was obtained in Lab 1 – 4 (Table 22) while for 19 chemicals the result is consistent with the three-class classifications in Laboratory 5 – 8 (Table 23). This then gives an inter-laboratory reproducibility for class assignment of 83% and 79%, respectively, with an average for the two independent evaluations of 81%

Using the two class model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified), for 22 chemicals a consistent result was obtained in Lab 1 – 4 (Table 22) while for 20 chemicals the result is consistent in Laboratory 5 – 8 (Table 23). This then gives an inter-laboratory reproducibility for identifying 1A chemicals of 92% and 83%, respectively, with an average for the two independent evaluations of 88%.

When evaluating these values vs. other published validation studies where each chemical was tested three times, this analysis is a bit more stringent as 50% of the comparisons are made in 4 labs, and the random chance of congruent results falls from 25% to 13% with testing in 4 labs instead of 3 labs (i.e. the more labs the higher the chance of producing one deviating result).

Table 20. Inter-laboratory reproducibility for GHS category determination based on the published log k_{max} cut-off of -1.1. Data for Laboratory 1 – 4 (testing the chemicals once).

Substance	Lab 1	Lab 2	Lab 3	Lab 4	Consistent 3 classes[1)]	Consistent 1A vs. 1B/NC[2)]
(Chloro)methylisothiazolinone	1A	1A	1A		YES	YES
Glyoxal	1B	1B	1B		YES	YES
Methylisothiazolinone	1A	1A	1A		YES	YES
Methyl-2-octynoate	1B	1B	1B	1B	YES	YES
4-phenylenediamine	1B	1B	1A	1B	NO	NO
Tetrachlorsalicylanilide	1A	1A	1A	1A	YES	YES
Isoeugenol	1A	1B	1B	1B	NO	NO
Bourgeonal	n-r	n-r	n-r	n-r	YES	YES
Carvone	1B	n-r	1B		NO	YES
Dihydrocoumarin	n-r	n-r	n-r	n-r	YES	YES
Hydroxycitronellal	n-r	n-r	n-r		YES	YES
Imidazolidinyl urea	1A	1A	1B		NO	NO
Methylhexanedione	1B	1B	1B	1B	YES	YES
Perillaaldehyde	1B	1B	1B	1B	YES	YES
Phenyl benzoate	n-r	n-r	n-r		YES	YES
Phenylpropionaldehyde	1B	1B	1B	1B	YES	YES
Tetramethyldiuram disulfide	1A	1A	1A		YES	YES
Benzosiothiazolinone	1A	1A	1A		YES	YES
Benzylidene acetone	1B	1B	1B	1B	YES	YES
Delta Damascone	1B	1B	1B		YES	YES
Diethylmaleate	1B	1B	1B	1B	YES	YES
trans-2-hexenal	1A	1A	1A		YES	YES
4-methoxy-acetophenone	n-r	1B	n-r		NO	YES
Chlorobenzene	n-r	n-r	n-r	n-r	YES	YES
n consistent					**19**	**21**
Reproducibility (%)					**79%**	**88%**

[1)] 3 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified)

[2)] 2 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

Table 21. Inter-laboratory reproducibility for GHS category determination based on the published log k_{max} cut-off of -1.1. Data for Laboratory 5 – 8 (testing the chemicals three times).

Substance	Lab 5	Lab 6	Lab 7	Lab 8	Consistent 3 classes[2)]	Consistent 1A vs. 1B/NC[3)]
(Chloro)methylisothiazolinone	1A [1)]	1A	1A	1A	YES	YES
Glyoxal	1B	1B	1B	1B	YES	YES
Methylisothiazolinone	1A	1A	1A	1A	YES	YES
Methyl-2-octynoate	1B	1B	1B		YES	YES
4-phenylenediamine	1A	1A	1A		YES	YES
Tetrachlorsalicylanilide	1A	1A	1A		YES	YES
Isoeugenol	1B	1B	1B		YES	YES
Bourgeonal	n-r	n-r	n-r		YES	YES
Carvone	1B	1B	1B	1B	YES	YES
Dihydrocoumarin	n-r	n-r	n-r		YES	YES
Hydroxycitronellal	n-r	1B	n-r	1B	NO	YES
Imidazolidinyl urea	1A	1B	1A	1A	NO	NO
Methylhexanedione	1B	1B	1B		YES	YES
Perillaaldehyde	1B	1B	1B		YES	YES
Phenyl benzoate	n-r	n-r	n-r	n-r	YES	YES
Phenylpropionaldehyde	1B	1B	1B		YES	YES
Tetramethyldiuram disulfide	1A	1A	1A	1A	YES	YES
Benzosiothiazolinone	1A	1A	1A	1A	YES	YES
Benzylidene acetone	1B	1B	1B		YES	YES
Delta Damascone	1B	1B	1B	1B	YES	YES
Diethylmaleate	1B	1A	1B		NO	NO
trans-2-hexenal	1A	1A	1A	1A	YES	YES
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	YES	YES
Chlorobenzene	n-r	n-r	n-r		YES	YES
n consistent					**21**	**22**
Reproducibility (%)					**88%**	**92%**

[1)]The average log k_{max} from three intra-laboratory experiments was taken to make the prediction

[2)]3 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified)

[3)]2 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

Table 22. Inter-laboratory reproducibility for GHS category determination based on the refined log k_{max} cut-off of -2.0. Data for Laboratory 1 – 4 (testing the chemicals once).

Substance	Lab 1	Lab 2	Lab 3	Lab 4	Consistent 3 classes1)	Consistent 1A vs. 1B/NC2)
(Chloro)methylisothiazolinone	1A	1A	1A		YES	YES
Glyoxal	1A	1B	1A		NO	NO
Methylisothiazolinone	1A	1A	1A		YES	YES
Methyl-2-octynoate	1A	1A	1A	1A	YES	YES
4-phenylenediamine	1A	1A	1A	1A	YES	YES
Tetrachlorsalicylanilide	1A	1A	1A	1A	YES	YES
Isoeugenol	1A	1A	1A	1A	YES	YES
Bourgeonal	n-r	n-r	n-r	n-r	YES	YES
Carvone	1B	n-r	1B		NO	YES
Dihydrocoumarin	n-r	n-r	n-r	n-r	YES	YES
Hydroxycitronellal	n-r	n-r	n-r		YES	YES
Imidazolidinyl urea	1A	1A	1A		YES	YES
Methylhexanedione	1B	1B	1A	1B	NO	NO
Perillaaldehyde	1B	1B	1B	1B	YES	YES
Phenyl benzoate	n-r	n-r	n-r		YES	YES
Phenylpropionaldehyde	1B	1B	1B	1B	YES	YES
Tetramethyldiuram disulfide	1A	1A	1A		YES	YES
Benzosiothiazolinone	1A	1A	1A		YES	YES
Benzylidene acetone	1A	1A	1A	1A	YES	YES
Delta Damascone	1B	1B	1B		YES	YES
Diethylmaleate	1A	1A	1A	1A	YES	YES
trans-2-hexenal	1A	1A	1A		YES	YES
4-methoxy-acetophenone	n-r	1B	n-r		NO	YES
Chlorobenzene	n-r	n-r	n-r	n-r	YES	YES
n consistent					**20**	**22**
Reproducibility (%)					83%	92%

[1)]3 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified)

[2)]2 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

Table 23. Inter-laboratory reproducibility for GHS category determination based on the refined log k_{Max} cut-off of -2.0. Data for Laboratory 5 – 8 (testing the chemicals three times).

Substance	Lab 5	Lab 6	Lab 7	Lab 8	Consistent 3 classes[2)]	Consistent 1A vs. 1B/NC[3)]
(Chloro)methylisothiazolinone	1A [1)]	1A	1A	1A	YES	YES
Glyoxal	1B	1A	1B	1A	NO	NO
Methylisothiazolinone	1A	1A	1A	1A	YES	YES
Methyl-2-octynoate	1A	1A	1A		YES	YES
4-phenylenediamine	1A	1A	1A		YES	YES
Tetrachlorsalicylanilide	1A	1A	1A		YES	YES
Isoeugenol	1A	1A	1A		YES	YES
Bourgeonal	n-r	n-r	n-r		YES	YES
Carvone	1B	1B	1B	1B	YES	YES
Dihydrocoumarin	n-r	n-r	n-r		YES	YES
Hydroxycitronellal	n-r	1B	n-r	1B	NO	YES
Imidazolidinyl urea	1A	1B	1A	1A	NO	NO
Methylhexanedione	1B	1B	1B		YES	YES
Perillaaldehyde	1B	1B	1B		YES	YES
Phenyl benzoate	n-r	n-r	n-r	n-r	YES	YES
Phenylpropionaldehyde	1B	1B	1B		YES	YES
Tetramethyldiuram disulfide	1A	1A	1A	1A	YES	YES
Benzosiothiazolinone	1A	1A	1A	1A	YES	YES
Benzylidene acetone	1B	1A	1A		NO	NO
Delta Damascone	1A	1B	1B	1B	NO	NO
Diethylmaleate	1A	1A	1A		YES	YES
trans-2-hexenal	1A	1A	1A	1A	YES	YES
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	YES	YES
Chlorobenzene	n-r	n-r	n-r		YES	YES
n consistent					19	20
Reproducibility (%)					79%	83%

1)The average log k_{max} from three intra-laboratory experiments was taken to make the prediction
2)3 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified)
3)2 class prediction model (kDPRA prediction GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

5.4.10. Reproducibility of positive control

For the positive control, the log k_{max} values and the rate constants at a fixed time (90 min and 150 min) are reported from each experiment. (These time points are defined for acceptance of the PC in an experiment.) Table 24 – Table 26 indicate the results from all valid runs, while Table 27 indicates how many runs were performed for each lab and how many of those were invalid.

The average log k_{max} of all valid runs is at -1.35, and this value is identical to the value obtained as average value of all labs in the transfer phase, thus very comparable results were obtained in both phases of this study. The average log k_{max} value was between -1.15 and -1.51 for the 7 labs (Table 24). The standard deviation for intra-laboratory

reproducibility of the log k_{max} was between 0.14 and 0.23, similar to the average standard deviation obtained for all chemicals in the intra-laboratory reproducibility (0.158).

The inter-laboratory variability in the PC was even lower for the rate constant derived at 90 or 150 min, and log $k_{90\ min}$ varies between -1.53 and -1.64 for the 7 labs (Table 25), while the intra-laboratory standard deviation is between 0.04 and 0.08. The overall standard deviation for all runs is at 0.04.

Log $k_{150\ min}$ varies between -1.62 and -1.75 for the 7 labs (Table 26), while the intra-laboratory standard deviation is between 0.03 and 0.09. The overall standard deviation for all runs is at 0.04.

The acceptance criterion for the PC in phase II was as follows: The log k of the PC at 90 min should be within the following range: -1.75 $M^{-1}s^{-1}$ to -1.40 $M^{-1}s^{-1}$. If no log k is obtained at 90 min, the value at 150 min can be taken into account and should lie in the following range: -1.90 $M^{-1}s^{-1}$ to -1.45 $M^{-1}s^{-1}$.

Only in few instances in phase II was the PC not fulfilling the acceptance criteria (Table 27). In a total of 148 runs, the acceptance criterion for the PC was missed 7-times. This corresponds to a 5% failure rate.

Table 24. Reproducibility of positive control: log k_{max} values [$M^{-1}s^{-1}$]

	Average	SD	Min	Max
All labs Phase II	**-1.35**	**0.13**	**-1.51**	**-1.15**
Lab A	-1.37	0.16	-1.59	-0.87
Lab B	-1.44	0.14	-1.63	-1.16
Lab E	-1.15	0.23	-1.52	-0.73
Lab F	-1.51	0.18	-1.70	-1.09
Lab D	-1.46	0.23	-1.88	-0.81
Lab C	-1.32	0.19	-1.59	-0.99
Lab G	-1.22	0.15	-1.54	-1.01
All labs Phase I [1)]	*-1.35*	*0.22*	*-1.76*	*-0.94*

[1)]Shown for comparison

Table 25. Reproducibility of positive control: log $k_{90\ min}$ values [$M^{-1}s^{-1}$]

	Average	SD	Min	Max
All labs Phase II	**-1.58**	**0.04**	**-1.64**	**-1.53**
Lab A	-1.53	0.07	-1.66	-1.43
Lab B	-1.60	0.04	-1.68	-1.53
Lab E	-1.62	0.08	-1.75	-1.48
Lab F	-1.64	0.07	-1.75	-1.52
Lab D	-1.60	0.08	-1.75	-1.48
Lab C	-1.54	0.07	-1.66	-1.37
Lab G	-1.56	0.04	-1.64	-1.47
All labs Phase I [1)]	*-1.60*	*0.08*	*-1.73*	*-1.38*

[1)]Shown for comparison

Note: The k_{90min} value is the rate used to decide on acceptability of an experiment. In case no rate is calculated at 90 min (reaction not linear or not statistically significant) then k_{150min} can be considered. Only in 4 of 148 runs in the Phase 2 testing labs had to report 150 min value instead of 90 min value.

Table 26. Reproducibility of positive control: log $k_{150\ min}$ values [$M^{-1}s^{-1}$]

	Average	SD	Min	Max
All labs Phase II	-1.66	0.04	-1.75	-1.62
Lab A	-1.62	0.05	-1.71	-1.50
Lab B	-1.67	0.03	-1.71	-1.57
Lab E	-1.65	0.06	-1.73	-1.53
Lab F	-1.69	0.04	-1.77	-1.65
Lab D	-1.75	0.07	-1.88	-1.62
Lab C	-1.63	0.06	-1.74	-1.51
Lab G	-1.64	0.09	-1.71	-1.26
All labs Phase I [1)]	*-1.68*	*0.09*	*-1.92*	*-1.45*

[1)] *Shown for comparison*

Table 27. Number of experiments performed with valid and invalid positive control.

	N runs with valid PC	N runs with invalid PC
All labs	**148**	**7**
Lab A	21	0
Lab B	18	0
Lab E[1)]	15	0
Lab F	25	2
Lab D	23	3
Lab C	22	2
Lab G	24	0

[1)] For lab E the number of runs is lower, as this lab performed 4 chemicals in parallel with one PC, using two sets of test plates. Although the procedure described in the SOP suggests a PC on every set of test plates, this deviation was considered to be acceptable, as the two sets of test plates were conducted in parallel on the same test days. The other labs performed a maximum of 3 chemicals in parallel with one PC, as suggested in the SOP.

5.4.11. Interference from autofluorescence and potential fluorescence quenching

The alert "autofluorescence may influence results" was triggered in 14 instances, 13 of which were when testing the chemical tetrachlorsalicylanilide (TCS). These were all the tests performed on this chemical (9 repeated testing in 3 labs and 4 inter-laboratory tests), thus this alert is reproducibly triggered. Indeed, as seen in Figure 9, strong background fluorescence is observed in absence of the peptide for this chemical, and in parallel an apparent depletion is observed already at 10 min which does not increase over time. To properly evaluate this chemical, an alternative fluorescent probe would be needed as this result does not reflect the true reactivity.

There was one isolated additional event, where the alert was triggered by the automatic calculation, however, this was due to high background levels of the solvent control and no enhanced autofluorescence of the substance in absence of the peptide was noted.

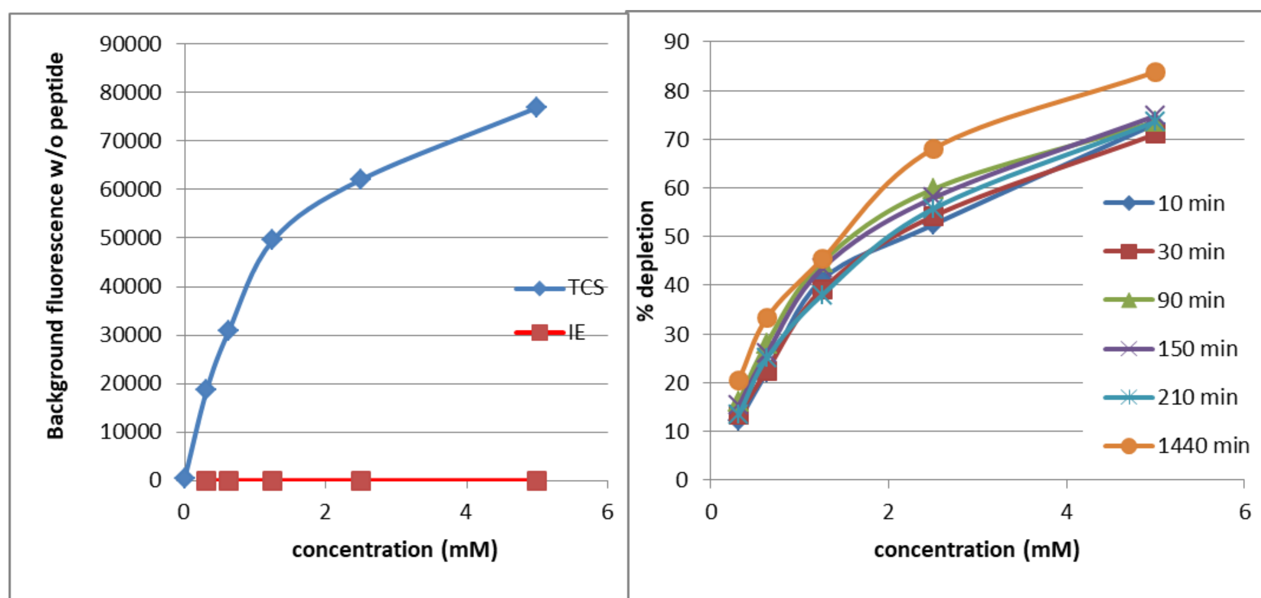


Figure 9. Autofluorescence effect of TCS. Autofluorescence of the test chemical in absence of the peptide in comparison to a non-fluorescent chemical (IE) (left) and apparent peptide depletion calculated based on subtraction of the control (right), which is not time dependent and appears to come from the auto-fluorescence effect. One curve is provided per exposure time (right panel, time in [min]).

The alert 'depletion may be due to quenching' is triggered more often – in 23 instances in total. It was repeatedly triggered for isoeugenol (5 instances), PPD (9 instances), BI (7 instances) and TMD (2 instances).

Details how quenching / absorbance by test chemicals may affect result for these chemicals and how such results should be interpreted are discussed in detail in Appendix 2.

5.4.12. Intra-laboratory reproducibility of 24 h / 5 mM depletion value

The classical DPRA in OECD 442C reports the 24 h depletion value at 5 mM final test chemical concentration. To be able to compare the kDPRA with the DPRA results (please note that the molar ratios of the test chemical and peptide are the same in DPRA and kDPRA) and with the intra- and inter-laboratory variability of the DPRA, the 24 h / 5 mM depletion value from the kDPRA is here reported as an additional read-out. Table A6 shows the intra-laboratory variability for these data. The average standard deviation of all intra-laboratory comparisons (n = 84) of this depletion value in % is at 4.6%.

In the DPRA validation study conducted by ECVAM [8, 9], 15 chemicals were tested each three times in three laboratories (n = 45). The average standard deviation of these intra-laboratory comparisons of the 24 h / 5 mM depletion value was at 4.5%, thus very close to the value observed in this study, although different chemicals were tested in the DPRA validation study [8, 9]. The ECVAM study is the basis of the DPRA validation, but the single depletion values are not published and were directly received from ECVAM for this analysis (personal communication, S. Casati).

5.4.13. Inter-laboratory reproducibility of 24 h / 5 mM depletion value

The 24 h / 5 mM depletion values for all 7 labs are reported in Table A7 and in Figures 10 and 11. Overall, with few exceptions, these values are also highly reproducible between labs. The average standard deviation between

labs for all 24 chemicals is 7%. In the DPRA validation study conducted by ECVAM , 24 chemicals were tested in three laboratories and the average inter-laboratory standard deviation was at 3.4%, thus a bit lower than observed in the current study, but this difference may partly be specific to the chemical test-set, as different chemicals were tested in the two studies. (Note: the background data on individual depletion values in the DPRA study used to calculate this comparison are not published and were directly received from ECVAM).

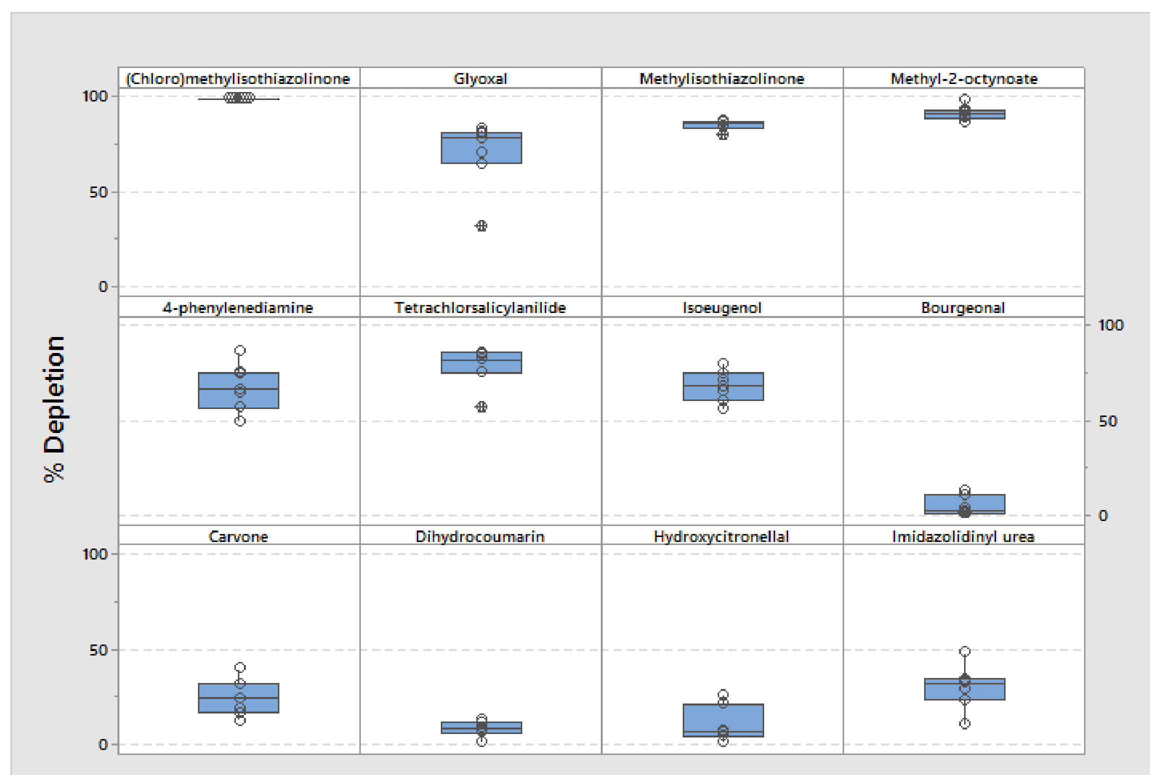


Figure 10. 24 h / 5 mM depletion values from inter-laboratory testing, chemicals 1-12. For laboratory 5 – 8, the average from repeated testing is plotted. Shown are the 7 individual values (circles), the interquartile range box (blue) and the average (horizontal line)

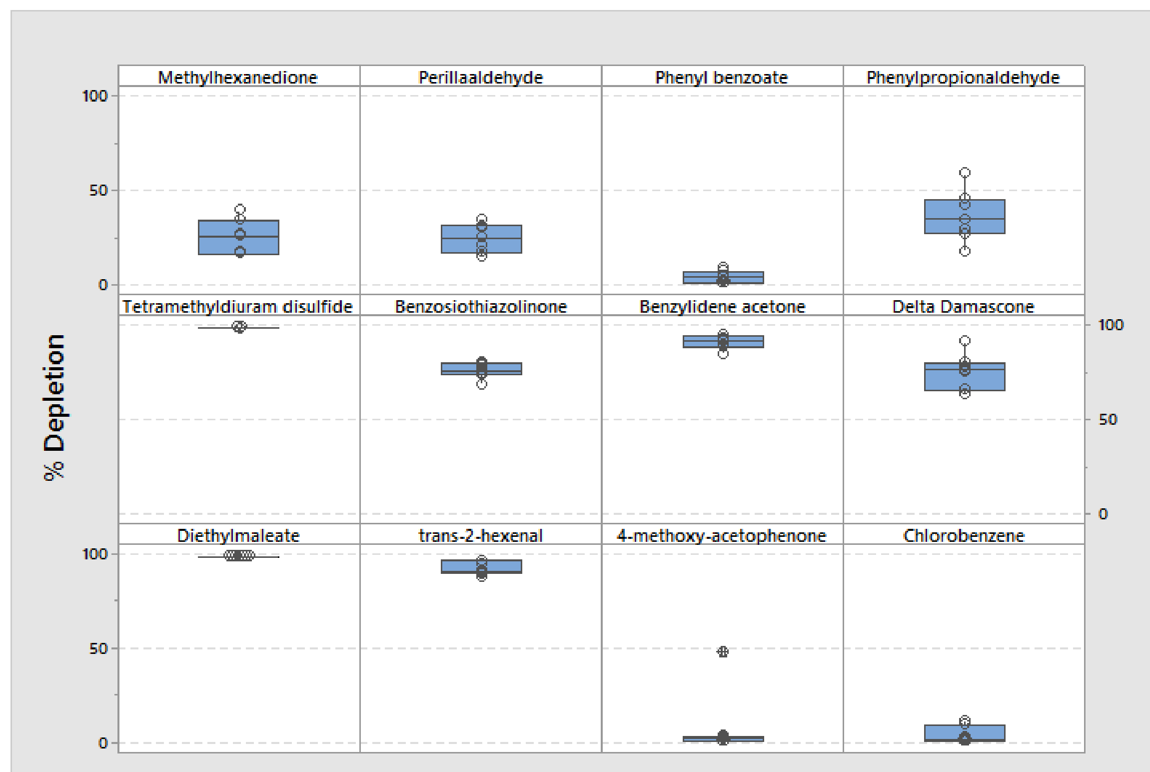


Figure 11. 24 h / 5 mM depletion values from inter-laboratory testing, chemicals 13-24. For laboratory 5 – 8, the average from repeated testing is plotted. Shown are the 7 individual values (circles), the interquartile range box (blue) and the average (horizontal line).

For 16 chemicals tested in phase I and phase II, there are also data from the DPRA pre-validation study conducted in four laboratories available. This prevalidation study had been done in four labs and was coordinated by P&G. The detailed data are not published and they were directly obtained from P&G to analyze in this context. For this subset of 16 chemicals, the inter-laboratory standard deviation of the 24 h / 5 mM depletion values in the kDPRA is at 8.9%. For the same chemicals the standard deviation from 4 labs in the DPRA for Cys-depletion in the pre-validation study was at 10.2% (Table A8). Thus, the inter-laboratory variability of the two approaches is rather similar when assessed on the common set of 16 chemicals.

5.4.14. Comparison of kDPRA (24 h / 5 mM depletion value) to the classical DPRA

Figure 12 shows the correlation between the 24 h / 5 mM depletion values in the kDPRA and the classical DPRA for (a) published reference values [10]and (b) data from the pre-validation study (P&G, personal communication). The correlation coefficient for these comparisons is at $R^2 = 0.85$ and 0.81 respectively.

If the DPRA values from the pre-validation studies is compared to the independent reference DPRA values [10](i.e. the two DPRA values on the y-axis of the two figures in Figure 12), the correlation is at $R^2 = 0.83$(F-value = 75.37, $p < 0.0005$). Thus, the repeated testing in the DPRA gives a similar variability as when comparing kDPRA data vs DPRA data (Table A8 and Figure 12).

The 24 h / 5 mM data were also compared between the extended data-set on the kDPRA reported under chapter 3.6 below and published DPRA values (data not shown). For 155 chemicals with both kDPRA and DPRA results, the correlation between the 24 h / 5 mM values is at $R^2 = 0.81$, thus again similar to the consistency observed in Figure 12 for the results from the kDPRA validation study data-set with historical DPRA data and with the comparisons of DPRA to an independently repeated DPRA.

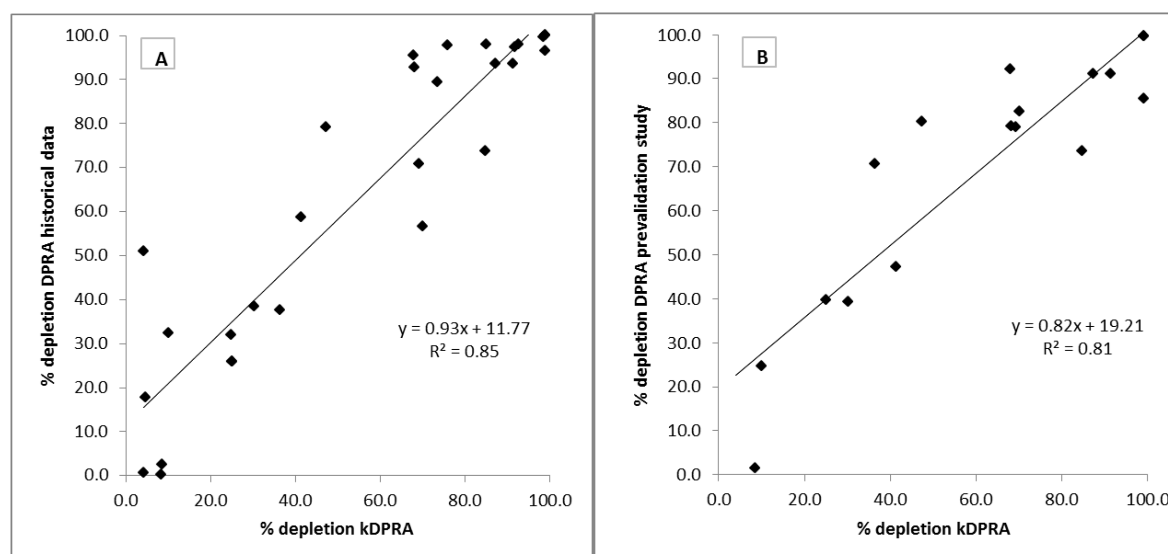


Figure 12. Correlation between 24 h/ 5 mM values recorded in the kDPRA and in the DPRA for (a) published reference values and (b) data from the pre-validation study.

Regression statistics for regression A: F-value = 156.9, $p < 0.0005$

Regression statistics for regression B: F-value = 64.3, $p < 0.0005$

5.4.15. Congruency of solvent-decision

All chemicals were sent as blinded, independent samples and thus each laboratory received a total of 48 vials. The labs were not aware of the intra-laboratory repetitions; hence an independent solubility assessment was performed for each of the 48 samples. Figure 13 indicates the log k_{max} values obtained with different solvents for the four chemicals with differing solvent choice. For IU the laboratory with the outlier result was not depicted, as this result is clearly not related to the chosen vehicle since the laboratory chose one-time ACN and two times H2O as the vehicle but got consistent results showing no reac-tivity or a very low log k_{max} value independently of the solvent used. The average SD of the log k_{max} values of these four substances over all labs was calculated to be 0.21 (excluding IU values of the outlier lab) while the average SD of all other substances over all labs was 0.29, demonstrating that solvent selection was not a driving factor for variance.

Note: For water-soluble substances the pH 7.5 phosphate buffer was defined as the vehicle in the SOP. However, 2 labs dissolved these chemicals in H_2O instead. This deviation is considered to be minimal and without impact on the results of the tests; as depicted in Figure 13, log k_{max} values did not differ between chemicals dissolved in water and buffer. Further, water was used for chemical preparation only, and the buffer was used for sample preparation in the kDPRA assay in all cases. In consequence, phosphate buffer and water are counted as “consistent” in the following evaluation.

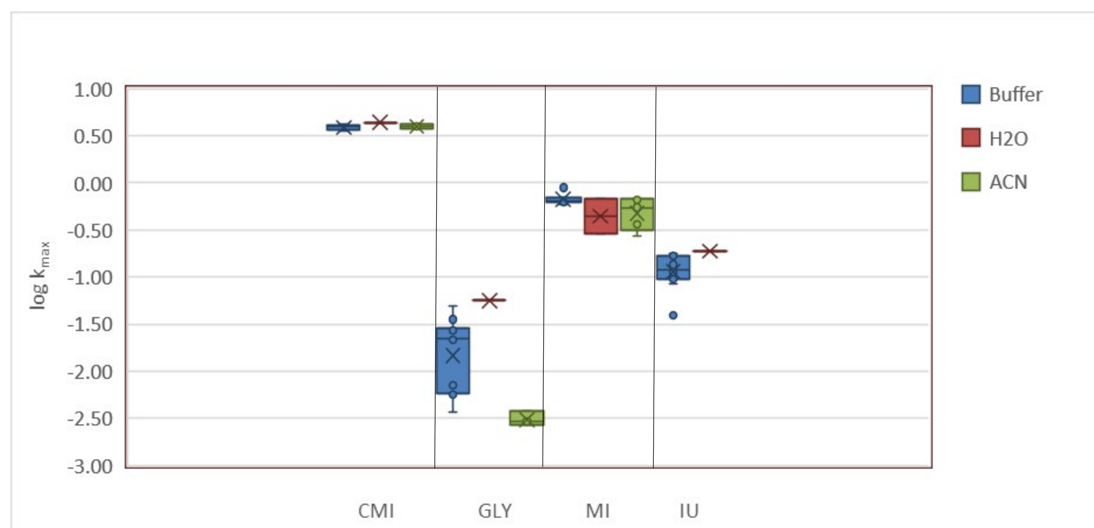


Figure 13: log k_{max} [$M^{-1}s^{-1}$] values per vehicle for the 4 chemicals, for which different vehicles were chosen by some par-ticipating labs..

5.4.16. Intra-laboratory congruency of solvent-decision

Table A9 in the appendix shows the individual solvent decisions for each chemical in three independent repetitions. For 20 chemicals, all laboratories reported consistent solvent selection of acetonitrile. For 4 chemicals at least one laboratory made inconsistent solvent selection over the three experiments. Overall, of the solvent decisions of the 84 chemical-lab combinations (4 x 12 + 3 x 12), consistent decisions were observed in 79 instances, thus intra-laboratory reproducibility of solvent decision is at 94%.

5.4.17. Inter-laboratory congruency of solvent-decision

Two sets of results were separately evaluated for inter-laboratory congruency of solvent selection: (i) the data for the 3 – 4 labs testing the chemicals once (Lab 1-4, table 28) and (ii) for the other 3 – 4 labs testing the chemicals three times (Lab 5 – 8, table 29). The later evaluation was done by giving a WoE decision for solvent selection for each lab, which was either the selected solvent, when consistently chosen in all three instances, or “n/a”, when inconsistent solvent selection was performed. In both sets, water and buffer are counted as “consistent” decision.

In the 3 – 4 labs testing the chemical once, only one inconsistent solvent decision was made. This was for the chemical methylisothiazolinone (MI), where one lab chose acetonitrile as the vehicle, while the other labs chose an aqueous solvent. Hence, 23 of 24 chemicals were consistent in the inter-laboratory comparison of the labs testing the chemical once, corresponding to 96% (table 28).

In the 3 – 4 labs testing the chemicals three times, four inconsistent solvent decisions were made. These were for the chemicals (chloro)methylisothiazolinone (CMI), glyoxal (GLY), MI and imidazolidinyl urea (IU). Hence, comparing inter-laboratory consistency, 20 of 24 decisions were concordant, corresponding to 83%.

All four chemicals with inconsistent solvent selection have a good or very good water solubility (table 30), indicating that the selection of an aqueous solvent was justifiable. Also, all labs reported clear solutions of these chemicals, independent whether they chose water, buffer or ACN as the vehicle. Further, it has to be noted that the final samples in the assay plates, though prepared in different ways, are exactly the same, independent of the solvent used for substance formulation.

Table 28. Inter-laboratory reproducibility for solvent-decision. Data for Laboratory 1 – 4 (testing the chemicals once).

Substance	Lab 1	Lab 2	Lab 3	Lab 4	Consistent
(Chloro)methylisothiazolinone	Buffer	H_2O	Buffer		YES
Glyoxal	Buffer	Buffer	H2O		YES
Methylisothiazolinone	ACN	Buffer	Buffer		NO
Methyl-2-octynoate	ACN	ACN	ACN	ACN	YES
4-phenylenediamine	ACN	ACN	ACN	ACN	YES
Tetrachlorsalicylanilide	ACN	ACN	ACN	ACN	YES
Isoeugenol	ACN	ACN	ACN	ACN	YES
Bourgeonal	ACN	ACN	ACN	ACN	YES
Carvone	ACN	ACN	ACN		YES
Dihydrocoumarin	ACN	ACN	ACN	ACN	YES
Hydroxycitronellal	ACN	ACN	ACN		YES
Imidazolidinyl urea	Buffer	H_2O	Buffer		YES
Methylhexanedione	ACN	ACN	ACN	ACN	YES
Perillaaldehyde	ACN	ACN	ACN	ACN	YES
Phenyl benzoate	ACN	ACN	ACN		YES
Phenylpropionaldehyde	ACN	ACN	ACN	ACN	YES
Tetramethyldiuram disulfide	ACN	ACN	ACN	ACN	YES
Benzosiothiazolinone	ACN	ACN	ACN		YES
Benzylidene acetone	ACN	ACN	ACN	ACN	YES
Delta Damascone	ACN	ACN	ACN		YES
Diethylmaleate	ACN	ACN	ACN	ACN	YES
trans-2-hexenal	ACN	ACN	ACN		YES
4-methoxy-acetophenone	ACN	ACN	ACN		YES
Chlorobenzene	ACN	ACN	ACN	ACN	YES
n consistent					**23**
Reproducibility [%}					**96**

Table 29. Inter-laboratory reproducibility for solvent-decision. Data for Laboratory 5 – 8 (testing the chemicals three times).

Substance	Lab 5	Lab 6	Lab 7	Lab 8	Consistent
(Chloro)methylisothiazolinone	Buffer	Buffer	ACN	n/a	NO
Glyoxal	ACN	Buffer	n/a	Buffer	NO
Methylisothiazolinone	n/a	n/a	Buffer	Buffer	NO
Methyl-2-octynoate	ACN	ACN	ACN		YES
4-phenylenediamine	ACN	ACN	ACN		YES
Tetrachlorsalicylanilide	ACN	ACN	ACN		YES
Isoeugenol	ACN	ACN	ACN		YES
Bourgeonal	ACN	ACN	ACN		YES
Carvone	ACN	ACN	ACN	ACN	YES
Dihydrocoumarin	ACN	ACN	ACN		YES
Hydroxycitronellal	ACN	ACN	ACN	ACN	YES
Imidazolidinyl urea	Buffer	n/a	Buffer	Buffer	NO
Methylhexanedione	ACN	ACN	ACN		YES
Perillaaldehyde	ACN	ACN	ACN		YES
Phenyl benzoate	ACN	ACN	ACN	ACN	YES
Phenylpropionaldehyde	ACN	ACN	ACN		YES
Tetramethyldiuram disulfide	ACN	ACN	ACN	ACN	YES
Benzosiothiazolinone	ACN	ACN	ACN	ACN	YES
Benzylidene acetone	ACN	ACN	ACN		YES
Delta Damascone	ACN	ACN	ACN	ACN	YES
Diethylmaleate	ACN	ACN	ACN		YES
trans-2-hexenal	ACN	ACN	ACN	ACN	YES
4-methoxy-acetophenone	ACN	ACN	ACN	ACN	YES
Chlorobenzene	ACN	ACN	ACN		YES
n consistent					**20**
Reproducibility [%}					**83.3%**

n/a indicates that the laboratory chose different vehicles in independent intra-laboratory repetitions

Table 30. Water solubility and log Kow of the four chemicals with inconsistent solvent selection. Source: QSAR Toolbox, calculated values.

Name	Abbreviation	water solubility [g/L]	log Kow
(Chloro)methylisothiazolinone	CMI	149	-0.34
Glyoxal	GLY	1000	-1.66
Methylisothiazolinone	MI	537	-0.83
Imidazolidinyl urea	IU	1000	-8.28

5.4.18. Conclusions on intra-and inter-laboratory reproducibility

- The intra-laboratory reproducibility of the quantitative log k_{max} determination is very high with an overall geometric standard deviation of 0.158 in a logarithmic scale between -3.5 and + 0.75. This corresponds to a 1.46-fold variation in the linear scale.
- The intra-laboratory reproducibility for GHS category attribution is 86% (3-classes, [non-reactive, GHS 1B, GHS 1A]) and 94% (2-classes, [non-reactive/GHS 1B, GHS 1A]) with the published prediction model and 88% (3-classes) and 96% (2-classes) with the refined prediction model. Variable predictions are mostly observed for chemicals with a log k_{max} very close to the classification cut-off.
- The inter-laboratory variability of the quantitative log k_{max} is by a factor two higher than the intra-laboratory variability, but overall still low with an average inter-laboratory geometric standard deviation of 0.244 which corresponds to a variation of 1.75-fold on the linear scale.
- Reproducibility of class attribution with the published prediction model is at 83% (3-classes) and 90% (2-classes).
- Reproducibility of class attribution with the refined prediction model is at 81% (3-classes) and 88% (2-classes).
- Again, variable predictions are mostly observed for chemicals with a log k_{max} very close to the classification cut-off.
- Highly reproducible results for the PC were obtained, especially for the log k_{90min} and log k_{150min} value used as proficiency criteria. Identical values as in the transfer phase were obtained for the log k_{max}.
- Highly consistent solvent selection was performed. In cases of inconsistent solvent selection, both choices were suitable to dissolve the chemical and the results obtained are within the generally observed standard deviation.
- Good reproducibility was also observed for the 24 h / 5 mM depletion value, and variability was similar when compared to the pre-validation (personal communication) and validation studies [8, 9] on the DPRA. Also, the values measured in the DPRA with a different detection system are in general in good agreement with the depletion values measured with the fluorescent derivatization in the kinetic profiling, indicating that the kDPRA is indeed an extension of the DPRA.

5.5. WP 3.5. Evaluation of the predictive capacity for the dataset of Phase I and Phase II

While the predictive capacity can best be estimated with a large dataset, and hence is explored in detail in WP 3.6, the data from both phase I and phase II including PC were compared vs. the human (n = 27) and LLNA (n = 31) reference data, Table 31 lists the overall results and predictive capacity, while Table 33 gives the individual results for the different chemicals. The average log k_{max} from all 7 labs was used for class attribution and to calculate predictivity in Tables 31 – 33.

A low sensitivity was observed for the published tentative cut-off, which is also observed on the extended dataset in 3.6 and which is why a more conservative cut-off was now introduced as 'refined cut-off' based on ROC analysis. A sensitivity of 100% for LLNA data and 94% for human data is observed for the refined cut-off if the reported log k_{max} values are used to identify GHS Cat 1A sensitizers. Specificity is at 65% for LLNA data and 82% for human data. The reduced specificity against LLNA data is not surprising, as the validation set intentionally included GHS Cat 1A chemicals from the ICCVAM list, a number of which are under-predicted by the LLNA. Hence predictivity of human data by the LLNA for this data-set is far from perfect (63% sensitivity and 100% specificity).

However, it has to be noted that for phenylenediamine the apparent log k_{max} is an overestimation of the kinetic rate due to partial fluorescence quenching / absorbance (see Appendix 2). The refined kinetic rate led to phenylenediamine being rated as a Cat 1B chemical. In addition, tetrachlorsalicylanilide reactivity is not accurately reflected by the data due to autofluorescence. Predictivity was also recalculated omitting TCS and using the adjusted apparent kinetic rate for PPD. This led to a slightly inferior predictive capacity (Table 32).

Table 31. Predictive capacity for all chemicals of phase I and phase II. Evaluation performed using 2-classes (non-reactive/GHS 1B, GHS 1A).

	Prediction LLNA published cut-off	Prediction LLNA refined cut-off	Prediction human published cut-off	Prediction human refined cut-off
CP	6	11	7	15
FP	4	7	2	2
CN	16	13	9	9
FN	5	0	9	1
Sum	31	31	27	27
Sensitivity	55%	100%	44%	94%
Specificity	80%	65%	82%	82%
Accuracy	71%	77%	59%	89%
Balance Accuracy	67%	83%	63%	88%

CN, correct negative CP, correct positive; FN, false negative, FP, false positive; accuracy [%] = (CP+ CN)/(CP + FP +CN + FN) ∗ 100; balanced accuracy [%] = (sensitivity [%]+ specificity[%] / 2; sensitivity [%] = $CP_{1A\ as\ 1A}/(CP_{1A\ as\ 1A}+FN_{1A\ as\ 1B/NC})$ ∗ 100; specificity [%] = $CN_{1B\ as\ 1B}/(CN_{1B\ as\ 1B}+FP_{1B/NC\ as\ 1A})$

Table 32. Predictive capacity for all chemicals of phase I and phase II adjusted for the quenching effect observed for PPD and omitting TCS. Evaluation performed using 2-classes (non-reactive/GHS 1B, GHS 1A)

	Prediction LLNA published cut-off	Prediction LLNA refined cut-off	Prediction human published cut-off	Prediction human refined cut-off
Sensitivity	50%	90%	40%	87%
Specificity	80%	65%	82%	82%
Accuracy	70%	73%	58%	85%
Balanced Accuracy	65%	78%	61%	84%

CN, correct negative CP, correct positive; FN, false negative, FP, false positive; accuracy [%] = (CP+ CN)/(CP + FP +CN + FN) ∗ 100; balanced accuracy [%] = (sensitivity [%]+ specificity[%] / 2; sensitivity [%] = $CP_{1A\ as\ 1A}/(CP_{1A\ as\ 1A}+FN_{1A\ as\ 1B/NC})$ ∗ 100; specificity [%] = $CN_{1B\ as\ 1B}/(CN_{1B\ as\ 1B}+FP_{1B/NC\ as\ 1A})$

Table 33. Predicitve capacity for all chemicals of phase I and phase II: Individual predictions.

Name	CAS	LLNA Cat	Human Cat	log k_{max} [2] $[M^{-1}s^{-1}]$	Prediction published cut-off	Prediction refined cut-off
(Chloro)methylisothiazolinone	26172-55-4	1A	1A	0.60	1A	1A
Glyoxal	107-22-2	1A	1A	-1.89	1B	1A
Methylisothiazolinone	2682-20-4	1A	1A	-0.23	1A	1A
Methyl-2-octynoate	111-12-6	1A	1A	-1.52	1B	1A
4-phenylenediamine	106-50-3	1A	1A	-1.16	1B	1A
4-phenylenediamine adjusted[1]	106-50-3	1A	1A	-2.81	1B	1B
Tetrachlorsalicylanilide *(affected by autofluorescence)*	1154-59-2	1A	1A	-0.46	1A	1A
Isoeugenol	97-54-1	1A	1A	-1.21	1B	1A
Bourgeonal	18127-01-0	1B	1B	n-r	n-r	n-r
Carvone	6485-40-1	1B	1B	-2.75	1B	1B
Dihydrocoumarin	119-84-6	1B	1B	n-r	n-r	n-r
Hydroxycitronellal	107-75-5	1B	1B	-2.82	1B	1B
Imidazolidinyl urea	39236-46-9	1B	1B	-0.97	1A	1A
Methylhexanedione	13706-86-0	1B	1B	-2.54	1B	1B
Perillaaldehyde	2111-75-3	1B	1B	-2.92	1B	1B
Phenyl benzoate	93-99-2	1B	1B	n-r	n-r	n-r
Phenylpropionaldehyde	93-53-8	1B	1B	-2.76	1B	1B
Tetramethyldiuram disulfide	137-26-8	1B	1B	0.74	1A	1A
Benzosiothiazolinone	2634-33-5	1B	1A	-0.10	1A	1A
Benzylidene acetone	122-57-6	1B	1A	-1.89	1B	1A
Delta Damascone	57378-68-4	1B	1A	-2.17	1B	1B
Diethylmaleate	141-05-9	1B	1A	-1.26	1B	1A
trans-2-hexenal	6728-26-3	1B	1A	-0.44	1A	1A
4-methoxy-acetophenone	100-06-1	NC		n-r	n-r	n-r
Chlorobenzene	108-90-7	NC		n-r	n-r	n-r
2,4-Dinitrochlorobenzene	97-00-7	1A	1A	-0.56	1A	1A
Oxazolone	15646-46-5	1A		-0.13	1A	1A
Formaldehyde	50-00-0	1A	1A	-0.66	1A	1A
Ethyleneglycol dimethacrylate	97-90-5	1B	1B	-2.45	1B	1B
Benzylideneacetone	122-57-6	1B	1A	-1.81	1B	1A
2,3-Butanedione	431-03-8	1B		-2.78	1B	1B
Cinnamic aldehyde	104-55-2	1A	1A	-1.31	1B	1A

[1] 4-phenylenediamine triggered a quenching alert, and as described in Appendix 2, indeed this could lead to an overestimation of the kinetic rate. Hence an adjusted rate was calculated and is shown here. This also leads to slightly different predictive capacities as shown in Table 33.

[2] The average log k_{max} from all 7 labs was used to calculate predictivity in Tables 31 – 33.

5.6. WP 3.6. Building the database and testing the prediction cut-off

In parallel to the multi-laboratory validation study as described above, more data were generated with the refined protocol to enlarge the database. This database work will be separately reported in detail and does not form part of the validation study, but it is an important contribution to evaluate predictivity of the assay both as stand-alone method and as part of a DA. Hence, this work is also summarized here.

- Additional chemicals were tested in single runs in single proficient lab(s).
- Chemical selection was optimized to prepare a database of kinetic rate constants on reference chemicals which were previously evaluated in various other alternative methods and with well-known *in vivo* reference data, e.g. those in [6, 10, 11]. They were thus selected from the following lists:

  a) Additional chemicals from the ICCVAM list on LLNA potency [6] if not already in validation set described in Tables 2 and 17.
  b) Additional (if not already in above lists) performance standards substances selected for OECD 442D (Nrf-2 luciferase test), OECD 442C (DPRA) and OECD 429 (LLNA): These are well characterized chemicals used to validate methods for binary predictions of sensitization hazard
  c) Additional chemicals from the recent Cosmetic Europe database [11]: These chemicals were all tested in multiple *in vitro* tests and complementarity of the kDPRA for improved potency assessment in ITS / DA can in the future be easily assessed when combining the data with the established database of those methods. In addition, for most of these chemicals some human evidence, at least from clinical data and expert judgment, is available [7].
  d) Additional chemicals from the compilation by Urbisch et al.[10]
  e) Data from published papers on the kDPRA were also included [2-4]

Metals and complex mixtures such as essential oils included in above datasets were not included, as they are considered outside of the applicability domain.

This complete database contains log k_{max} data on a total of 186 chemicals with LLNA and 123 chemicals with human potency category attribution. Six chemicals were excluded from evaluation due to strong interference (quenching or autofluorescence or reaction with mBrB, Table 38).

- The resulting database was used to examine best cut-offs / cut-off values for GHS sub-categorization into categories 1A and 1B when tested against larger datasets
  - The optimal cut-off was calculated for both human and mouse data using ROC analysis.
  - The intra- and inter-laboratory reproducibility for classification was then recalculated by applying the refined cut-off value to the continuous data generated in the validation study (see above).

- Correlation analysis of the rate constants towards LLNA potency data on a continuous scale (and not only class attribution) was performed separately and is reported in Appendix 4. Combination of log k_{max} values with other *in vitro* parameters for potency assessment was also evaluated in that Appendix. Application of the log k_{max} in different defined approaches will be tested in a separate evaluation.

The database including the human and LLNA reference data is summarized in Appendix 3.

5.6.1. Deriving the optimal cut-off to discriminate GHS Cat 1A

One key potential application of the kDPRA is to discriminate Cat 1A sensitizers from other sensitizers in a tiered strategy whereby the sensitizer hazard is first identified based on a defined approach and the kDPRA result is then used to identify Cat 1A sensitizers. Thus a most predictive cut-off is needed which discriminates GHS Cat 1A sensitizers from GHS Cat 1B and non-classified (GHS 2) chemicals in a binary prediction model.

Thus, all the data from the extended database were separated into GHS Cat 1A vs. GHS Cat 1B / non-classified chemicals and the sensitivity and specificity for identifying 1A chemicals was calculated vs. LLNA data or, separately, vs. human data.

The *in vivo* reference data for the LLNA were taken from the different reference lists cited above. If LLNA data in the Cosmetics Europe database were available, these were given precedence, as they are derived from multiple LLNA studies if available, unless there is evidence that more reliable data were reported elsewhere. This LLNA dataset is currently under review by the OECD expert group (DA SS) and this analysis may eventually be updated, but it is not expected that many chemicals would be attributed to a different GHS Category based on this data review.

The human data come from the assessment by ICCVAM when evaluating the LLNA for potency discrimination [6], the RIFM database and largely also from the Cosmetics Europe database and the corresponding potency assessment by Basketter et al. [7]. As described in that paper, human potency classes 1 and 2 fall in GHS Cat 1A while 3 – 6 are considered GHS Cat 1B/NC. However, it should be noted that part of that assessment is based on clinical data, and not on predictive human tests. Thus, not all chemicals labeled as Cat 1A in that assessment would be rated as Cat 1A based on human predictive tests (HRIPT and HMT).

In total 186 chemicals were tested. Six could not be evaluated due to quenching, autofluorescence or reaction with mBrB, thus the ROC analysis is based on 180 chemicals.

Figure 14 shows results of the receiver operating characteristic (ROC) analysis for LLNA data and Figure 15 presents the ROC analysis vs human data plotting sensitivity versus 1-specificity. The ROC analysis is summarized in table 34.

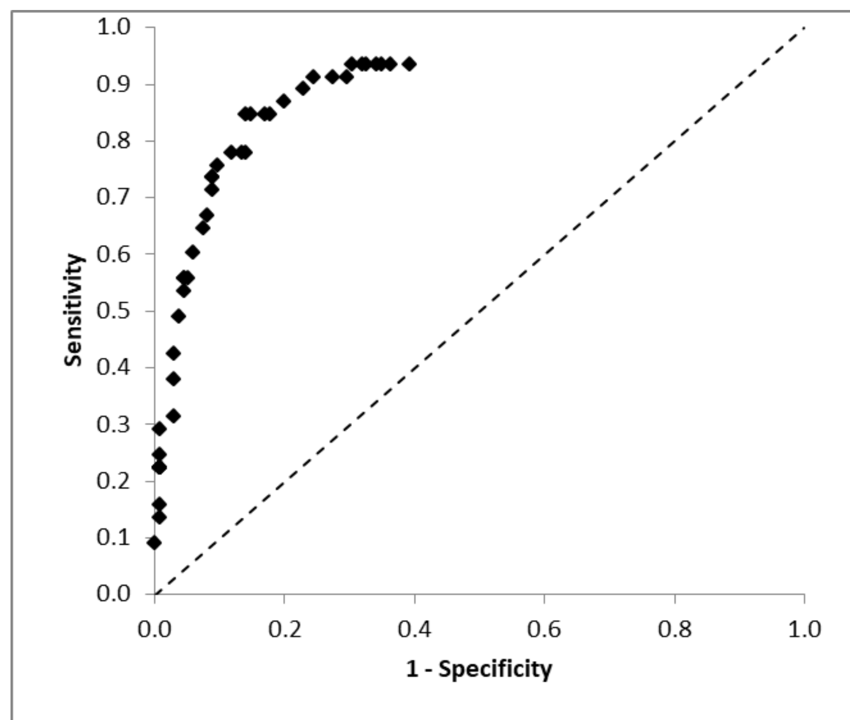


Figure 14. ROC analysis for different log k_{max} cut-off values to predict GHS Cat 1A vs. LLNA data.

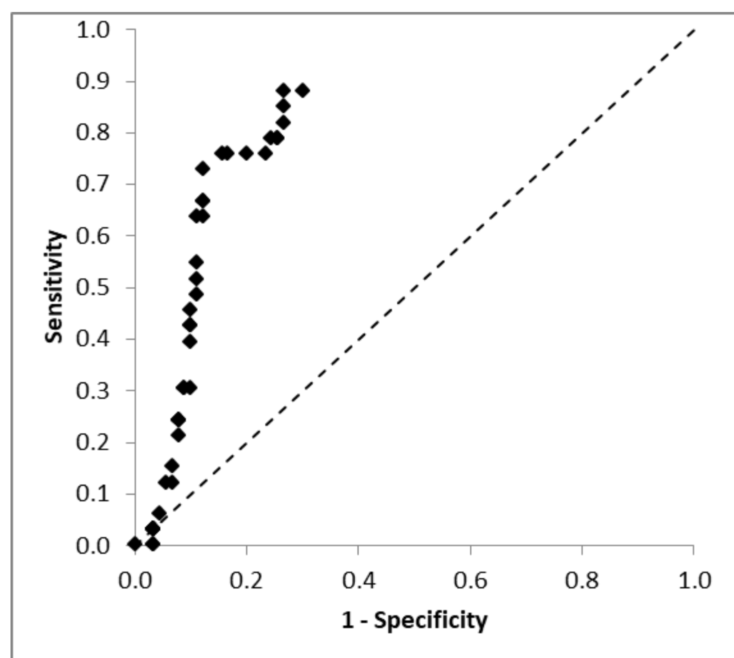


Figure 15. ROC analysis for different log k_{max} cut-off values to predict GHS Cat 1A vs. Human data.

Table 34. ROC analysis for different log k_{max} cut-off values to predict GHS Cat 1A vs. LLNA and human data. The range from log k_{max} = -1.0 - -3.0 is shown.

Cut-off	LLNA			Human		
	Sensitivity	Specificity	Balanced accuracy	Sensitivity	Specificity	Balanced accuracy
-1	60.0%	94.1%	77.0%	30.3%	91.1%	60.7%
-1.1	64.4%	92.6%	78.5%	30.3%	91.1%	60.7%
-1.2	66.7%	91.9%	79.3%	30.3%	90.0%	60.2%
-1.3	71.1%	91.1%	81.1%	39.4%	90.0%	64.7%
-1.4	73.3%	91.1%	82.2%	42.4%	90.0%	66.2%
-1.5	73.3%	91.1%	82.2%	42.4%	90.0%	66.2%
-1.6	75.6%	90.4%	83.0%	45.5%	90.0%	67.7%
-1.7	77.8%	88.1%	83.0%	48.5%	88.9%	68.7%
-1.8	77.8%	86.7%	82.2%	51.5%	88.9%	70.2%
-1.9	77.8%	85.9%	81.9%	54.5%	88.9%	71.7%
-2	84.4%	85.9%	85.2%	63.6%	88.9%	76.3%
-2.1	84.4%	85.2%	84.8%	63.6%	87.8%	75.7%
-2.2	84.4%	83.0%	83.7%	66.7%	87.8%	77.2%
-2.3	84.4%	82.2%	83.3%	66.7%	87.8%	77.2%
-2.4	86.7%	80.0%	83.3%	72.7%	87.8%	80.3%
-2.5	88.9%	77.0%	83.0%	75.8%	84.4%	80.1%
-2.6	91.1%	75.6%	83.3%	75.8%	83.3%	79.5%
-2.7	91.1%	72.6%	81.9%	75.8%	80.0%	77.9%
-2.8	91.1%	70.4%	80.7%	75.8%	76.7%	76.2%
-2.9	93.3%	69.6%	81.5%	78.8%	75.6%	77.2%
-3	93.3%	68.1%	80.7%	78.8%	74.4%	76.6%

Table 34 reports the predictivity of different cut-offs for human and LLNA data as plotted in the ROC analysis, and Table 35 summarizes the overall predictivity for the previous and the refined cut-off. The previous tentative cut-off is highlighted in purple, the new cut-off with maximal balanced accuracy for LLNA is in green, and the maximal balanced accuracy for the human dataset is in light blue.

The ROC analysis vs. LLNA data shows a clear gain in sensitivity (64% to 84%) from the published cut-off (log k_{max} = -1.1) up to a cut-off of a log k_{max} = -2.0 with a minor loss in specificity (93 % to 86%) (Table 34). Thus, based on the LLNA data, a refined cut-off of log k_{max} = -2.0 yields maximized predictivity with a balanced accuracy of 85 %.

The ROC analysis vs. human data shows a gain in sensitivity (30% to 64%) from the published cut-off (log k_{max} = -1.1) up to a cut-off of -2.0, too, with a minor loss in specificity (91% to 89%). Thus, based on the human data, a refined cut-off of log k_{max} = -2.0 yields a balanced accuracy of 76%. A further gain in sensitivity (73%) and balanced accuracy (80%) would be achieved with a cut-off of log k_{max} = -2.4, however this would then reduce the accuracy vs. LLNA data (83% instead of 85% balanced accuracy).

To evaluate whether this predictivity for human data is sufficiently good, a comparison vs. the LLNA can be made: The LLNA has a sensitivity of 54.5% and a balanced accuracy of 72.8% for this dataset when predicting the human potency categorization for the same chemicals and based on the same reference data. Thus also with the cut-off of log k_{max} = -2.0, the kDPRA is slightly superior to the LLNA in predicting human potency (Table 35) in this data set.

The kDPRA has (for this sub-set of data with human data available) a sensitivity of 76.9%, specificity of 88.7% and balanced accuracy of 82.6% to predict the LLNA outcome (Table 35 b).

Table 36 lists the false-negative human GHS Cat1A sensitizers based on a cut-off of log k_{max} = -2.0, it is worthy to discuss these chemicals to interrogate the predictivity and to check whether this cut-off is sufficiently protective:

- For creosol the Basketter *et al.* compilation attributed a Cat 1A classification based on a low NOEL value only (i.e. from a study showing no sensitization reactions). Thus this class attribution is due to the arbitrarily chosen low test concentration, and it is highly likely that this chemical would only sensitize at much higher doses as similar molecules like eugenol or dihydroeugenol are weak to moderate sensitizers falling into Cat 1B. Thus, it is probably an incorrect assignment.
- Lyral was classified as Cat 1A based on clinical observations, while predictive human tests had not found sensitization reactions. Thus, also human predictive testing would probably not have led to this class attribution.
- The false-positives include two clear pro-haptens (diethylenetriamine and 3-dimethylaminopropylamine) and two pre-haptens (4-phenylenediamine, which reacts more slowly in the kinetic assay as it requires abiotic oxidation and 2-aminophenol).
- They include also three Michael acceptor chemicals with LOEL values close to the human cut-off for Cat 1A chemicals of 500 µg/cm^2, and an extrapolated DSA05 (extrapolated value leading to induction of sensitization in 5% of the panelist) therefore closely below the cut-off. (δ-Damascone (human LOEL = 500 µg/cm^2), 2-hexylidene cyclopentanone (human LOEL= 500 µg/cm^2), Methylanisylidene acetone (human LOEL= 550 µg/cm^2) which are also 1B in LLNA, so these are clearly borderline chemicals. Phenylacetaldehyde was rated 1B by the Basketter et al. compilation, but it is 1A based on the ICCVAM evaluation of the RIFM data and here included in 1A.

Finally, 9 of 12 of these under-predicted chemicals are rated as Cat 1B by the LLNA, too.

Thus, overall, only a limited number of important and clear-cut human 1A sensitizers are missed by this refined cut-off (4-phenylenediamine, Diethylenetriamine, Glutaric aldehyde, 2-aminophenol, 3-Dimethylaminopropylamine, 6-Methyl-3,5-heptadien-2-one). Based on all these evaluations, a refined cut-off of log k_{max} = -2.0 appears as an optimal prediction model to balance accuracy for LLNA and human data.

In terms of specific domains that tend to be underpredicted it is obvious, that phenolic pre-haptens (chemicals which can oxidized to reactive catechols or quinones, e.g. phenylenediamine and 2-aminophenol), pro-haptens which can be transformed to potent allergens eracting by Schiff-base formation (diethylenetriamine and 3-dimethylaminopropylamine) and potent Schiff-base forming chemicals (Glutaric aldehyde) may be underpredicted for their human sensitization potency.

Table 35a. Summary of the predictivity with the published and the refined cut-off.

	Sensitivity [%]	Specificity [%]	Balanced accuracy [%]	n tested
Published cut-off				
kDPRA vs LLNA	64.4	92.6	78.5	180
kDPRA vs Human	30.3	91.1	60.7	123
Refined Cut-off				
kDPRA vs LLNA	**84.4**	**85.9**	**85.2**	**180**
kDPRA vs Human	**63.6**	**88.9**	**76.3**	**123**
LLNA vs Human	*54.5*	*91.1*	*72.8*	*123*

Table 36b. Detailed contingency tables of the predictivity with the refined cut-off.

		LLNA (all)				
		1A	1B	1B/NC	NC	sum
kDPRA	1A	38	19	0	0	57
	1B/NC	7	77	2	37	123
	ex.	3	2	0	1	6
	sum	45	96	2	37	180

kDPRA vs. LLNA (all)		
sensitivity	84.4%	38/45
specificity	85.9%	116/135
accuracy	85.6%	154/180
balanced accuracy	85.2%	
n (exluding autofluor. /quen-ching)		180

		human				
		1A	1B	1B/NC	NC	sum
kDPRA	1A	21	10	0	0	31
	1B/NC	12	42	2	36	92
	ex.	0	0	0	0	0
	sum	33	52	2	36	123

kDPRA vs. human		
sensitivity	63.6%	21/33
specificity	88.9%	80/90
accuracy	82.1%	101/123
balanced accuracy	76.3%	
n (exluding autofluor. /quen-ching)		123

		LLNA				
		1A	1B	1B/NC	NC	sum
human	1A	18	14	0	1	33
	1B	8	38	1	5	52
	1B/NC	0	1	1	0	2
	NC	0	15	0	21	36
	ex.	22	30	0	11	63
	sum	26	68	2	27	123

LLNA vs. Human		
sensitivity	54.5%	18/33
specificity	91.1%	82/90
accuracy	81.3%	100/123
balanced accuracy	72.8%	
n (exluding autofluor. /quen-ching)		123

	LLNA (for substances w/ human data)					
		1A	1B	1B/NC	NC	sum
kDPRA	1A	20	11	0	0	31
	1B/NC	6	57	2	27	92
	ex.	0	0	0	0	0
	sum	26	68	2	27	123

	kDPRA vs. LLNA (w/human data)	
sensitivity	76.9%	20/26
specificity	88.7%	86/97
accuracy	86.2%	106/123
balanced accuracy	82.8%	
n (exluding autofluor. /quen-ching)		123

Table 37. Human GHS Cat 1A sensitizers underpredicted by applying a cut-off of log k_{max} = -2.0

Name	CAS	Log k_{max}	LLNA GHS Cat
4-phenylenediamine	106-50-3	-2.81	1A
δ-Damascone	57378-68-4	-2.16	1B
Diethylenetriamine	111-40-0	-3.50	1B
2-hexylidene cyclopentanone	17373-89-6	-2.36	1B
Methylanisylidene acetone	104-27-8	-3.10	1B
Phenylacetaldehyde	122-78-1	-2.36	1B
Glutaric aldehyde	111-30-8	-3.50	1A
2-aminophenol	95-55-6	-2.46	1A
3-Dimethylaminopropylamine	109-55-7	-3.50	1B
Lyral	31906-04-4	-3.31	1B
6-Methyl-3,5-heptadien-2-one	1604-28-0	-3.29	NC
2-Methoxy-4-methylphenol	93-51-6	-3.50	1B

Table 38. LLNA GHS Cat 1A sensitizers underpredicted by applying a cut-off of log k_{max} = -2.0

Name	CAS	Log k_{max}	LLNA EC 3 [%]	Human GHS Cat
4-phenylenediamine	106-50-3	-2.81	0.15	1A
Glutaric aldehyde	111-30-8	-3.50	0.09	1A
2-aminophenol	95-55-6	-2.46	0.45	1A

Hexyl salicylate	6259-76-3	-3.50	0.18	(1B / NC)[1)]
Bisphenol A-diglycidyl ether	1675-54-3	-2.53	1.5	1B
2,4-Dinitrobenzenesulfonic acid	885-62-1	-2.30	1.9	no data
Chlorpromazine	69-09-0	not reactive	1.0	1B

[1)] Attributed to 1B in Basketter compilation, but human HRIPT indicates no sensitization when tested up to 35'400 µg/cm^2 and no indication from the clinic indicating it to be a sensitizer – thus no evidence that it is a human sensitizer

Table 37 lists the seven LLNA Cat 1A chemicals underpredicted by the refined cut-off, two of them are pre-haptens (4-phenylenediamine and 2-aminophenol) as discussed above. There is no evidence that hexyl salicylate is a human sensitizer (tested up to 35'400 μg/cm^2 in human predictive tests) and two chemicals (Bisphenol A-diglycidyl ether, 2,4-Dinitrobenzenesulfonic acid) are very close to the prediction cut-off. Glutaric aldehyde is a predominately Lysine reactive Schiff-base former which is clearly underpredicted by the kDPRA.

Table 38 lists the larger list of false-positives. It is interesting that 11 of the 19 cases are in the LLNA EC3 range of 2 – 5%, thus close to the cut-off. Furthermore, the list contains a number of chemicals with clear sensitization risks from clinical data (imidazolidinyl urea, 2-mercapotobenzothiazole and tetramethyldiuram disulfide) and 6 human Cat 1A chemicals. Thus, these false-positives are clearly reactive molecules with a significant human sensitization risk.

The list also includes two strongly reactive acrylates, which are highly volatile (ethyl acrylate and butyl acrylate), and these rapidly evaporate under LLNA conditions, but are likely to be strong sensitizers when applied under occluded conditions.

Table 39. LLNA Cat 1B sensitizers overpredicted by applying a cut-off of log k_{max} = -2.0

Name	CAS	lLog k_{max}	LLNA EC 3 [%]	Human GHS Cat
Methyl 2-nonynoate	111-80-8	-1.66	2.5	1A
2-Decenal	3913-71-1	-1.03	2.5	
2-mercapotobenzothiazole	149-30-4	-0.15	2.6	1B
Tetramethyldiuram disulfide	137-26-8	0.74	2.93	1B
α-damascone	24720-09-0	-1.64	3.3	
4-vinylcyclohex-1-ene-carbaldehyde	1049017-68-6	-1.77	3.4	
Benzylidene acetone	122-57-6	-1.85	3.7	1A
2,4-Heptadienal	5910-85-0	-1.52	4	
trans-2-hexenal	6728-26-3	-0.47	4.05	1A
Bergamal	22418-66-2	-1.64	4.5	
Diethylmaleate	141-05-9	-1.21	4.7	1A
Benzosiothiazolinone	2634-33-5	-0.12	4.8	1A
Safranal	116-26-7	-1.74	7.5	1A
1,2,4-Benzenetricarboxylic anhydride	552-30-7	-0.13	9.2	
Abietic acid	514-10-3	-0.55	11	1B
Oxalic acid anhydrous	144-62-7	-1.01	15	
Butyl acrylate	141-32-2	-0.83	20	
Imidazolidinyl urea	39236-46-9	-1.11	24	1B
Ethyl acrylate	140-88-5	-0.97	32.75	1B

6. General discussion

This study is based on the modular approach to test validation as defined by Hartung *et al.* [5], which includes the modules (i) Test definition (ii) Transferability (iii) Within-laboratory variability (iv) Between-laboratory variability (v) Predictive capacity (vi) evaluation of applicability domain and (vii) eventually definition of performance standards. Below the different modules are briefly summarized.

As multiple tests on the sensitization endpoints already exist, we also discuss the relationship to the current DPRA protocol and the potential application of the kDPRA in a tiered testing strategy.

6.1. Test definition

The test is addressing the molecular initiating event (MIE) of sensitization, which most likely is the rate limiting step for acquisition of skin sensitization. The kinetic rate of the reaction of a chemical with the skin proteins is, based on theoretical chemical understanding, the most direct quantitative measure correlating to sensitizing potency as it will indicate which amount of sensitizer is needed to yield a given number of new epitopes in a period of time. Of course, other factors, such as stability of the generated epitopes, eventually the ability for protein cross-linking epitopes to be formed and their immunogenicity may also be contributors to potency at the level of MIE.

Nevertheless, in an analysis of multiple quantitative parameters from *in vitro* assays, the strongest correlation to LLNA potency was found for the kinetic rate constant measured with a protocol closely related to the kDPRA (same experimental principle, different test peptide) [12] and correlation to potency of kDPRA data was shown in multiple studies [2-4]. Hence, mechanistically the test is strongly supported by our knowledge of the skin sensitization AOP and by experimental data.

Technically, the test is identical in the constituents of the reaction mixture to the current DPRA, so it is already very well established. The SOP fully defines the assay set-up and plate format for high-throughput testing and it comes along with a standardized Excel template to which the raw data can be pasted. Thus, the assay is defined and standardized to a very detailed technical level.

The Excel spreadsheet template allows automatic calculation of apparent maximal rate constants log k_{max} for peptide depletion. Additional information to treat special cases and to safeguard against fluorescence interference effects is provided to guarantee most meaningful data evaluation. This will also define the data evaluation and treatment to a high degree, although to evaluate special cases a certain degree of expert knowledge will always help to interpret the data correctly as in any test.

6.2. Transferability

The transferability study indicated easy transferability of the technical aspects of the assay to multiple laboratories. Most laboratories needed only 2 – 3 training runs to then arrive at a status where they were able to deliver valid runs. The five naïve labs reported very similar quantitative results to the lead laboratories.

No hands-on training was required, indicating that the SOP is largely self-explanatory. This will facilitate global adoption of the test. No proprietary elements are used to conduct the assay, hence any laboratory equipped with a 96-well fluorometer is in principle able to add the test to their test battery.

While evaluation of clearly reactive or non-reactive chemicals is straightforward with the automatic calculation sheet, the most critical part (like in many toxicological tests) remains the evaluation of borderline results and results from chemicals exhibiting non-linear kinetics or interference based on fluorescence quenching or autofluorescence.

To facilitate data evaluation for such chemicals, more information has been added to the SOP and calculations were added to the evaluation template after the transfer phase. Some more information on data evaluation will be added to the final SOP which was gathered from questions arising in the ring trial and during testing for expansion of the database (See Appendix 2).

6.3. Intra-laboratory Reproducibility

The reproducibility of the log k_{max} values was very high, and larger variation was only observed for specific chemicals, indicating that the assay *per se* is highly reproducible and most variation observed is chemical specific. This good reproducibility of the quantitative data has to be emphasized, as many other validation studies on *in* vitro tests for sensitization did solely focus on class attribution and did not assess in detail the reproducibility of the quantitative parameters (such as concentration-response data in cellular tests).

Intra-laboratory reproducibility of GHS subcategory attribution is also very good – which certainly follows from the reproducibility of the rate constants. Thus, using the refined cut-off, 88% intra-laboratory reproducibility for the 3-class-attribution (GHS Cat1A vs. GHS Cat 1B vs. GHS not classified) and 96 % reproducibility of identifying GHS Cat 1A chemicals in a 2 class model (Cat 1A vs. Cat 1B/NC) was found. Moreover, intra-laboratory variability of the 24 h / 5 mM depletion value was similar in the ECVAM validation of the DPRA and in the current study on the kDPRA indicating that the new assay set-up is not prone to higher variability as compared to the validated DPRA.

6.4. Inter-laboratory Reproducibility

The reproducibility of the log k_{max} between labs was good and again mostly chemical-specific variation was observed. The average geometric standard deviation is 0.244, which corresponds to a variation of 1.75-fold around the geometric mean on a linear scale. This then also led to high predictive accuracy for class attribution, both with the previously published and the refined prediction cut-off. The few instances of discordant results mostly occurred for chemicals with log k_{max} values close to the prediction cut-off. Only in few cases was significant variability of class attribution observed (esp. for phenylpropionic aldehyde). This may be due to the fact that such aldehydes lead to depletion of the test peptide mostly by peptide oxidation rather than adduct formation [1], and this process appears to be more prone to random variation as compared to direct chemical reactivity. Interestingly this chemical also had the highest inter-laboratory standard deviation in the DPRA pre-validation study (P&G, personal communication).

Thus, using the refined cut-off, 81% inter-laboratory reproducibility for the 3-class-attribution (GHS Cat 1A vs. GHS Cat 1B. vs. GHS not classified) and 88 % reproducibility of identifying GHS Cat 1A chemicals in a 2 class model (Cat 1A vs. Cat 1B/NC) was found.

The inter-laboratory reproducibility of the 24 h / 5 mM depletion value was similar to the reproducibility observed for the DPRA when the same chemicals were tested in the DPRA pre-validation study, indicating that the different assay format (160 µL instead of 1 mL; microtiter plates instead of glass vials) and different detection system did not change the reproducibility of the reaction under classical DPRA reaction conditions.

6.5. Predictive capacity to identify GHS Cat1A

For the set of chemicals (n=31) tested in multiple laboratories, a balanced prediction accuracy for the LLNA Cat 1A attribution of 83% was observed. For human data, predictivity is slightly higher, with a sensitivity of 94% and a balanced accuracy of 88%.

As one chemical (4-phenylenediamine) may lead to an overestimation of the rate constant due to fluorescence interference, accuracy was also calculated using a corrected rate constant and omitting the value for tetrachlorsalicylanilide which triggered autofluorescence. Taking these corrections into account, balanced accuracy of LLNA Cat 1A attribution is 78% and predictivity for human data gives sensitivity of 87% and a balanced accuracy of 84%.

This predictivity is still higher as compared to the LLNA, which has a sensitivity of 63% and a balanced accuracy of 81% for this dataset.

For the large dataset (n=180) used to evaluate predictivity in more detail, the predictivity with the refined cut-off is similar to this small dataset with a sensitivity of 84.4% and a balanced accuracy of 85.2% vs. LLNA data.

6.6. Applicability domain

6.6.1. Technical limitations

In principle the same issues regarding solubility of the test item as in the DPRA have to be considered, as the chemicals are dissolved at equal concentrations in the same matrix. The fact that some chemicals are not fully dissolved must not limit the generation of a valid result, as also chemicals in a supersaturated solution may still be sufficiently available for the reaction. Solubility will, however, affect the linearity of the reaction over time as described before [2], but since we consider the earliest and strongest reactivity in the k_{max} calculations, a changing dissolved concentration over time must not exclude generation of meaningful results. Still, testing chemicals above a log K_{ow} of 5 will be prone to some limitations due to inadequate solubility – however this statement is made from a conceptual point of view and we cannot, from the data, conclude that Cat 1A chemicals of a certain clogP are missed by the test. Regardless of the log K_{ow} of a test substance, a solubility experiment should be performed to determine the technical applicability of the kDPRA.

Other technical limitations are related to the specific way the results are measured with the fluorescence of a derivatizing agent.

- Thus, chemicals with primary SH-group (thiols) cannot be tested as the thiol group will react with the mBrB and lead to enhanced fluorescence. Such an example is thioglycerol (CAS 96-27-5). Furthermore, chemicals decomposing under the conditions of the assay (neutral, aqueous conditions) thereby releasing a free SH-group will be prone to the same limitations.
- Chemicals with an autofluorescence at the wavelength used for measurements cannot be tested with the current set-up. So far this was only observed for tetrachlorsalicylanilide. In a follow-up study, alternative fluorescent probes will be tested absorbing at higher wavelength to allow testing of these rare cases of molecules with autofluorescence in the desired wavelength range.
- Chemicals leading to fluorescence quenching may lead to an overestimation of apparent peptide depletion. How this is detected and handled is described in Appendix 2.

Despite above mentioned limitations, only 6 chemicals of 182 could not be tested in the database work reported under 3.6. These chemicals are listed in Table 39.

Precipitation of the test chemical at higher test concentration has been discussed as a potential reason for false-negative (Yamamoto et al; [13]) or non-linear (Roberts and Natsch [2]) results in the DPRA and kinetic DPRA. Thus precipitation, on theoretical grounds, could be another technical limitation.

At the test concentration (5 mM) in the DPRA, precipitation was observed for 16 of 82 test chemicals under the Cys-peptide conditions, i.e. at the top concentration of the kDPRA by Yamamoto et al.. Precipitation indicates that the test chemical concentration is above saturation and that actually a lower concentration of the chemical is available for reaction. In case the molecule reacts with the peptide, one would expect that more chemicals will re-dissolve and be able to participate in the reaction, but the molar ratio test chemical to peptide will not be as expected and not constant. In the study of Yamamoto et al., however, this observation had no effect on the prediction accuracy, actually for the chemicals for which precipitation was observed, there were five false-negatives with the DPRA and seven false negatives with the ADRA, the latter being conducted at non-precipitating concentrations.

Precipitation at top concentrations only could have an additional impact in the kDPRA: Within the concentration response one may observe a non-linear effect (Roberts and Natsch,[2]) – reactivity would not increase with increasing concentration, as the dissolved concentration would not increase linearly. Users of the kDPRA would then note significant peptide depletion, but no positive result based on the standard slope calculation method in the evaluation sheet, as the regression line used for rate calculation would not be linear and not be accepted for log kmax calculations by the evaluation sheet. However, in that case, the EXCEL evaluation sheet does perform an alternative calculation as detailed in ANNEX1, section b) of the SOP. Finally, if precipitation leads to significant reduction of the background fluorescence in the control wells with substance only due to light scattering (i.e. the conditions investigated in the Yamamoto et al. publication), then the alert 'potential quenching is triggered', which should alert user to investigate what is going on in more detail.

Table 40. Chemicals which could not be evaluated based on technical limitations.

Name	CAS	Observed problem
2-Nitro-1,4-phenylendiamine	5307-14-2	fluorescence quenching
Vanillin	121-33-5	fluorescence quenching
Tropolone	533-75-5	fluorescence quenching
Thioglycerol	96-27-5	reaction with fluorescent probe
Tetrachlorsalicylanilide	1154-59-2	autofluorescence
Bandrowski's Base (N,N-bis(4-aminophenyl)-2,5-diamino-1,4-quinone-diimine)	20048-27-5	fluorescence quenching

6.6.2. Predictive limitations

The assay only measures reactivity with the Cys-peptide. There are chemicals with exlcusive reactivity to lysine, such as some acyl-halides, phenol-esters or aldehydes. This may limit applicability / predictivity for strong sensitizers with an exclusive lysine-reactivity. However, as can be derived from the decent predictivity for detecting GHS Cat 1A sensitizers, there appears to exist only a small minority of Cat 1A sensitizers able to react exclusively with lysine residues. Glutaric aldehyde is certainly such a case. Considering exclusive strong Lys-reactivity from the DPRA in a tiered strategy is certainly an option to reduce this uncertainty. Besides glutaraldehyde, only few Cat 1A sensitizers such as atranol and chloratranol are currently known with specific selective lysine reactivity.

Pre-haptens leading to strong haptens such as 4-phenylenediamine are further cases which potentially may be underestimated – the time needed for them to oxidize will reduce the apparent kinetic rate of the reaction with the test peptide. Thus, chemicals spontaneously, but not instantly being transformed to very reactive species may lead to some underestimation of the sensitization potential if the lag-period for oxidation is in the range of hours.

As discussed widely before, the DPRA does not contain a metabolic system and thus prediction of pro-haptens requiring exclusively metabolic activation (i.e. not acting as direct haptens nor pre-haptens) in theory is a significant limitation. From the database evaluated so far, two compounds were identified which are clearly false-negative

against human GHS classification (3-dimethylaminopropylamine and diethylenetriamine) and are potential pro-haptens. However, these are rated as 1B in the LLNA. It has also been shown that in vitro skin sensitization test methods are able to detect most pre- and pro-haptens and negative predictions are considered acceptable. In vitro investigations (Urbisch et al., 2016 and Patlewicz et al., 2016) using compounds requiring molecular transformation to attain a sensitizing potential have shown that pre-haptens can readily be detected in the DPRA, many of which involve autoxidation processes. Moreover, many pro-haptens are also activated by non-enzymatic oxidation (and therefore are pre- and pro-haptens).

Interestingly, prediction accuracy to detect GHS Cat 1A sensitizers is apparently quite good, and most limitations in accuracy are observed for chemicals close to the cut-off – this observed decent predictivity indicates that the limitations for pro-haptens is not a dramatic shortcoming of the assay. Is there a scientific explanation for this observation? Indeed, most *bona fide* pro-haptens, for which activation by metabolic systems is well described and understood are weak to moderate sensitizers (e.g. dihydroeugenol, eugenol, cinnamic alcohol, ethylene diamine, geraniol). As the skin is a rather poor metabolic organ, metabolic activation may often be a rate-limiting step for sensitization by pro-haptens, thus rendering them less potent allergens. This may explain why an assay without metabolic activation may recognize most strong sensitizers, which will not be the case when it comes to detect weak allergens.

6.7. Relationship to DPRA 442C and advantages over the classical DPRA

A detailed comparison of both the actual 24 h / 5 mM depletion values and their reproducibility between the classical DPRA and the kDPRA was performed. Similar peptide depletion is observed in the classical assay performed in HPLC vials and the microplate assay and when applying the different detection systems (fluorescence derivatisation and HPLC analysis). This analysis therefore shows that, the assay format of the kDPRA nicely recapitulates the DPRA with the Cys-peptide. While in most cases a kDPRA would be considered after a positive DPRA result according to OECD TG 442C is obtained, in principle – for chemicals not exhibiting any fluorescence interference – the assay could be able to replace the DPRA when run according to the Cys-only prediction model. Especially chemicals predicted to be reactive (by *in silico* or cell-based assays) could directly be tested in the kDPRA. A classical DPRA including the Lys-peptide may then only be needed in case of non-reactive chemicals according to the kDPRA to ensure that no Lys-selective reactivity is missed.

The clear advantage over the classical DPRA is the fine granularity given by the kinetic rate, which grades reactivity on a logarithmic scale spanning almost 4 order of magnitude. As shown elsewhere, this does clearly improve the input into any DA which tries to exactly model an LLNA EC3 value as point of departure for quantitative risk assessment of chemicals. Moreover, as first shown in [4] and here based on refined cut-off on a much larger dataset, the kDPRA has the potential for GHS subcategory discrimination which cannot be achieved with the DPRA alone.

The kDPRA also presents a technical advantage over the classical DPRA that is conducted using a UV-HPLC analysis. The HPLC analytical equipment needed is often not available to laboratories primarily conducting cell-based work while a fluorescence plate reader is accessible at the majority of those institutes.

6.8. Potential use in tiered testing strategy

The most straightforward implementation of the kDPRA into a tiered testing strategy is to use:

(i) a defined approach (DA) for hazard identification as being in advanced validation at the OECD (OECD Draft Guideline On Defined Approaches For Skin Sensitisation)

(ii) in case a chemical is predicted as a sensitizer then to perform the kDPRA and use the refined prediction cut-off to attribute chemicals with a log k_{max} > -2.0 to the GHS Cat 1A, and those below the cut-off to GHS Cat 1B.

(iii) Optionally: Chemicals with exclusive lysine reactivity (only Lys-peptide depletion not due to co-elution in absence of Cys-reactivity in DPRA) in Tier 1 could then also be attributed to GHS Cat 1A in case reactivity surpasses a yet to defined cut-off of Lys-peptide depletion which is not due to co-elution.

6.9. Potential use of numeric log k_{max} in defined approaches

It has previously been shown how kinetic rate constants can be applied to the prediction of the skin sensitization potency. It is beyond the scope of this report to propose the optimal ITS /defined approach incorporating kDPRA rates reported here to predict potency, but we have built the database in a way that for many chemicals now data are available from tests according to OECD 442D and 442E along with the kDPRA kinetic rate, which will facilitate exploration of such approaches. The high-throughput nature, ease of transferability and assay availability will also facilitate the further completion of these databases.

Still, as a first attempt we have here, in response to questions addressed during the peer-review, in Appendix 4 added an analysis of how individual continuous parameters from *in vitro* assays contribute to LLNA potency prediction in a multiple regression analysis. This analysis shows that the kinetic rates have the strongest contribution to potency predictions, and that a model only based on the kinetic rates is sufficient for potency prediction for the strong sensitizers further indicating that the kinetic rates can be used as a stand-alone parameter to identify GHS1A sensitizers, while a more complex model is required to predict sensitization potency on a continuous scale also for weaker sensitizers.

6.10. Introduction as a test method into OECD 442C

As the kDPRA addresses an unmet regulatory need (GHS Cat 1A vs. GHS Cat 1B subcategorization), it is proposed to include this new protocol as a further appendix into the key event based OECD test guideline 442C in the form of a further Appendix next to the classical DPRA and the ADRA. A tiered testing strategy such as the one discussed above could then later, after proper evaluation, become part of the test guideline on defined approaches currently under development.

6.11. Potential improvements

While in the classical DPRA limitations may come from co-elution problems in HPLC analysis, fluorescence interference from quenching / absorbance and (much more rarely fluorescence) may occur, which can esp. affect the interpretation of weak depletion at early time points of the assay. This will be evaluated by testing an alternative fluorescent probe. Such an additional detection opportunity will not require separate validation, as it is a minor technical change and which will only be used for very few chemicals.

One observation is that pre-haptens which oxidize within hours may be underestimated. It is possible to test such chemicals by pre-incubating them in the medium for a couple of hours and then start the peptide reaction, to determine the reaction rate of the formed hapten. This is another possibility to further optimize the assay for chemicals underpredicted like 2-aminophenol or phenylenediamine, but it is not part of this study.

7. References

1. Natsch, A., et al., *Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro.* Toxicol. in Vitro, 2007. **21**(7): p. 1220-1226.
2. Roberts, D.W. and A. Natsch, *High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles.* Chem. Res. Toxicol., 2009. **22**(3): p. 592-603.
3. Natsch, A., T. Haupt, and H. Laue, *Relating skin sensitizing potency to chemical reactivity: reactive Michael acceptors inhibit NF-kappaB signaling and are less sensitizing than S(N)Ar- and S(N)2- reactive chemicals.* Chem. Res. Toxicol., 2011. **24**(11): p. 2018-27.
4. Wareing, B., et al., *Prediction of skin sensitization potency sub-categories using peptide reactivity data.* Toxicol In Vitro, 2017. **45**(Pt 1): p. 134-145.
5. Hartung, T., et al., *A modular approach to the ECVAM principles on test validity.* Altern Lab Anim, 2004. **32**(5): p. 467-72.
6. ICCVAM, *ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans.* Downloaded 23.1.2014 from <http://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/LLNA-pot/TMER.pdf#search=Test%20Method%20Evaluation%20Report:%20Usefulness%20and%20Limitations%20of%20the%20Murine%20Local%20Lymph%20Node%20Assay>, 2011. **NIH Publication Number 11-7709**.
7. Basketter, D.A., et al., *Categorization of chemicals according to their relative human skin sensitizing potency.* Dermatitis, 2014. **25**(1): p. 11-21.
8. ECVAM, *EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA)* http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra, last accessed 21.5.2014, 2014.
9. ECVAM, *Direct peptide reactivity assay (DPRA) for skin sensitization testing* DB-ALM database, 2014. **Protocol 154**: p. http://ecvam-dbalm.jrc.ec.europa.eu/protocols.cfm?id_met=1953 Last accessed 19.1.2015.
10. Urbisch, D., et al., *Assessing skin sensitization hazard in mice and men using non-animal test methods.* Regul. Toxicol. Pharmacol., 2015. **71**(2): p. 337-51.
11. Hoffmann, S., et al., *Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database().* Crit. Rev. Toxicol., 2018. **in press**: p. 1-15.
12. Natsch, A., et al., *Predicting Skin Sensitizer Potency Based on In Vitro Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment.* Toxicol. Sci., 2015. **143**(2): p. 319-32.
13. Yamamoto, Y., et al., *Precipitation of test chemicals in reaction solutions used in the amino acid derivative reactivity assay and the direct peptide reactivity assay.* J Pharmacol Toxicol Methods, 2019: p. 106624.

Appendix 1: Detailed result tables of the blind-coded inter-laboratory testing (Phase II)

Table A1: Individual log k_{max} values $[M^{-1}s^{-1}]$ from repeated intra-laboratory testing (rep) and average (AVG) values.

Substance	Lab 5 rep1	Lab 5 rep2	Lab 5 rep3	Lab 6 rep1	Lab 6 rep2	Lab 6 rep3	Lab 7 rep1	Lab 7 rep2	Lab 7 rep3	Lab 8 rep1	Lab 8 rep2	Lab 8 rep3	AVG Lab 5	AVG Lab 6	AVG Lab 7	AVG Lab 8
(Chloro)methylisothiazolinone	0.61	0.60	0.59	0.62	0.62	0.59	0.60	0.61	0.62	0.55	0.56	0.57	0.60	0.61	0.61	0.56
Glyoxal	-2.54	-2.43	-2.58	-2.15	-1.45	-1.31	-2.43	-2.25	-2.44	-1.64	-1.64	-1.66	-2.51	-1.64	-2.37	-1.65
Methylisothiazolinone	-0.56	-0.44	-0.54	-0.17	-0.17	-0.18	-0.05	-0.20	-0.16	-0.22	-0.20	-0.19	-0.51	-0.17	-0.13	-0.21
Methyl-2-octynoate	-1.32	-1.48	-1.62	-1.67	-1.66	-1.87	-1.59	-1.61	-1.63				-1.47	-1.73	-1.61	
4-phenylenediamine	-1.09	-1.04	-1.08	-0.94	-1.07	-1.06	-1.14	-0.96	-1.05				-1.07	-1.02	-1.05	
Tetrachlorsalicylanilide	-0.46	-0.42	-0.48	-0.39	-0.39	-0.30	-0.55	-0.61	-0.56				-0.46	-0.36	-0.57	
Isoeugenol	-1.79	-1.45	-0.84	-1.34	-1.15	-0.88	-1.80	-1.06	-1.08				-1.36	-1.12	-1.31	
Bourgeonal	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r	n-r
Carvone	-3.33	-3.24	-3.34	-3.13	-3.05	-2.32	n-r	-2.15	-2.37	-3.09	-2.27	-2.31	-3.30	-2.83	-2.26	-2.56
Dihydrocoumarin	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r	
Hydroxycitronellal	n-r	n-r	n-r	-2.07	-2.15	-3.42	n-r	n-r	n-r	-3.29	-2.95	-3.03	n-r	-2.55	n-r	-3.09
Imidazolidinyl urea	-0.75	-0.75	-0.78	-3.26	n-r	n-r	-0.88	-0.87	-1.07	-0.92	-0.98	-0.95	-0.76	n-r	-0.94	-0.95
Methylhexanedione	-3.34	-2.18	-2.11	n-r	-2.63	-2.23	-3.29	-3.43	-3.44				-2.54	-2.43	-3.39	
Perillaaldehyde	-2.89	-2.38	-2.34	-3.18	-3.25	-3.31	-3.49	-2.24	-3.39				-2.54	-3.25	-3.04	
Phenyl benzoate	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r
Phenylpropionaldehyde	-2.41	-3.33	n-r	-2.78	-1.90	-2.18	-2.50	-2.79	n-r				-2.87	-2.29	-2.64	
Tetramethyldiuram disulfide	0.70	0.72	0.71	0.78	0.77	0.76	0.74	0.73	0.74	0.72	0.73	0.73	0.71	0.77	0.74	0.73
Benzosiothiazolinone	-0.20	-0.27	-0.14	-0.20	-0.34	-0.25	0.07	0.20	-0.09	-0.17	-0.17	-0.16	-0.20	-0.26	0.06	-0.17
Benzylidene acetone	-1.85	-1.72	-2.67	-1.80	-1.99	-1.94	-1.88	-1.58	-1.69				-2.08	-1.91	-1.72	
Delta Damascone	-1.94	-1.90	-1.85	-2.33	-2.56	-2.24	-2.21	-2.31	-2.25	-2.19	-2.05	-2.03	-1.90	-2.38	-2.26	-2.09
Diethylmaleate	-1.34	-1.16	-1.13	-1.00	-0.98	-1.27	-1.34	-1.16	-1.11				-1.21	-1.08	-1.20	
trans-2-hexenal	-0.35	-0.44	-0.43	-0.42	-0.48	-0.41	-0.52	-0.59	-0.61	-0.57	-0.63	-0.58	-0.41	-0.43	-0.57	-0.59
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r
Chlorobenzene	-3.12	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r	n-r

n-r: Non-reactive

Table A2: Standard deviations (SD) of log k_{max} values from repeated intra-laboratory testing.

Substance	SD Lab 5	SD Lab 6	SD Lab 7	SD Lab 8	Avg SD intralab	SD Interlab [1)]
(Chloro)methylisothiazolinone	0.006	0.014	0.011	0.013	0.011	0.029
Glyoxal	0.077	0.450	0.104	0.015	0.161	0.478
Methylisothiazolinone	0.061	0.003	0.076	0.015	0.039	0.130
Methyl-2-octynoate	0.148	0.118	0.021		0.096	0.202
4-phenylenediamine	0.026	0.074	0.093		0.064	0.238
Tetrachlorsalicylanilide	0.030	0.051	0.030		0.037	0.086
Isoeugenol	0.485	0.232	0.418		0.378	0.103
Bourgeonal	n-r	n-r	n-r		n-r	n-r
Carvone	0.057	0.444	0.154	0.460	0.279	0.512
Dihydrocoumarin	n-r	n-r	n-r		n-r	n-r
Hydroxycitronellal	n-r	0.757	n-r	0.241	n.a. [2)]	n.a. [2)]
Imidazolidinyl urea	0.017	n-r	0.114	0.030	0.054	0.245
Methylhexanedione	0.690	0.283	0.082		0.352	0.795
Perillaaldehyde	0.307	0.064	0.691		0.354	0.401
Phenyl benzoate	n-r	n-r	n-r	n-r	n-r	n-r
Phenylpropionaldehyde	0.649	0.449	0.207		0.435	0.343
Tetramethyldiuram disulfide	0.008	0.011	0.004	0.001	0.006	0.020
Benzosiothiazolinone	0.062	0.072	0.145	0.006	0.071	0.161
Benzylidene acetone	0.512	0.101	0.154		0.256	0.125
Delta Damascone	0.042	0.162	0.050	0.088	0.085	0.169
Diethylmaleate	0.111	0.162	0.118		0.131	0.109
trans-2-hexenal	0.048	0.036	0.046	0.033	0.041	0.100
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	n-r	n-r
Chlorobenzene	n-r	n-r	n-r		n-r	n-r

n-r: Non-reactive

[1)] Here the standard deviation for inter-laboratory testing in Lab 1 – 4 is given to compare for the different chemicals the intrinsic variability in intra-and inter-laboratory testing

[2)] n.a. –not applicable as majority of experiments is “n-r”

Table A3: log k_{max} values [$M^{-1}s^{-1}$] from inter-laboratory testing and average values. For laboratory 5 – 8, the average (AVG) from repeated testing is shown.

Substance	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	AVG all Labs	AVG all repetitions	SD all Labs	SD all repetitions
(Chloro)methylisothiazo-linone	0.58	0.64	0.57		0.60	0.61	0.61	0.56	0.60	0.60	0.029	0.026
Glyoxal	-1.57	-2.23	-1.25		-2.51	-1.64	-2.37	-1.65	-1.89	-1.97	0.478	0.478
Methylisothiazolinone	-0.26	-0.15	-0.19		-0.51	-0.17	-0.13	-0.21	-0.23	-0.25	0.130	0.147
Methyl-2-octynoate	-1.28	-1.60	-1.71	-1.22	-1.47	-1.73	-1.61		-1.52	-1.56	0.202	0.185
4-phenylenediamine	-1.50	-1.17	-0.86	-1.47	-1.07	-1.02	-1.05		-1.16	-1.11	0.238	0.186
Tetrachlorsalicylanilide	-0.55	-0.35	-0.44	-0.47	-0.46	-0.36	-0.57		-0.46	-0.46	0.086	0.091
Isoeugenol	-1.07	-1.22	-1.23	-1.16	-1.36	-1.12	-1.31		-1.21	-1.24	0.103	0.298
Bourgeonal	n-r	n-r	n-r	n-r	n-r	n-r	n-r		n-r	n-r	n-r	n-r
Carvone	-2.16	n-r	-3.37		-3.30	-2.83	-2.26	-2.56	-2.75	-2.78	0.512	0.507
Dihydrocoumarin	n-r	n-r	n-r	n-r	n-r	n-r	n-r		n-r	n-r	n-r	n-r
Hydroxycitronellal	n-r	n-r	n-r		n-r	-2.55	n-r	-3.12	-2.83	-2.77	0.383	0.574
Imidazolidinyl urea	-1.02	-0.73	-1.41		-0.76	n-r	-0.94	-0.95	-1.30	-1.11	0.245	0.673
Methylhexanedione	-2.44	-3.61	-1.22	-2.12	-2.54	-2.43	-3.39		-2.54	-2.67	0.795	0.744
Perillaaldehyde	-3.44	-3.19	-2.56	-2.44	-2.54	-3.25	-3.04		-2.92	-2.93	0.401	0.470
Phenyl benzoate	n-r	n-r	n-r		n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r
Phenylpropionaldehyde	-3.17	-3.10	-2.37	-2.90	-2.87	-2.29	-2.64		-2.76	-2.68	0.343	0.443
Tetramethyldiuram disul-fide	0.75	0.74	0.73		0.71	0.77	0.74	0.73	0.74	0.74	0.020	0.023
Benzosiothiazolinone	-0.07	0.18	-0.21		-0.20	-0.26	0.06	-0.17	-0.10	-0.12	0.161	0.156
Benzylidene acetone	-1.94	-1.98	-1.78	-1.82	-2.08	-1.91	-1.72		-1.89	-1.90	0.125	0.261
Delta Damascone	-2.06	-2.34	-2.17		-1.90	-2.38	-2.26	-2.09	-2.17	-2.16	0.169	0.190
Diethylmaleate	-1.38	-1.25	-1.32	-1.38	-1.21	-1.08	-1.20		-1.26	-1.21	0.109	0.138
trans-2-hexenal	-0.38	-0.37	-0.34		-0.41	-0.43	-0.57	-0.59	-0.44	-0.47	0.100	0.100
4-methoxy-acetophenone	n-r	-2.91	n-r		n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r
Chlorobenzene	n-r	n-r	n-r	n-r	n-r	n-r	n-r		n-r	n-r	n-r	n-r

n-r: non-reactive

Table A4: Intra-laboratory reproducibility of the classification of chemicals with the published prediction model

Substance	Lab 5 rep1	Lab 5 rep2	Lab 5 rep3	Lab 6 rep1	Lab 6 rep2	Lab 6 rep3	Lab 7 rep1	Lab 7 rep2	Lab 7 rep3	Lab 8 rep1	Lab 8 rep2	Lab 8 rep3	Lab 5 AVG [1]	Lab 6 AVG	Lab 7 AVG	Lab 8 AVG	Labs with consistent repetitions (3 classes)[2]	Labs with consistent repetitions (2 classes)[3]
(Chloro)methylisothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Glyoxal	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	4 of 4	4 of 4
Methylisothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Methyl-2-octynoate	1B	1B	1B	1B	1B	1B	1B	1B	1B				1B	1B	1B		3 of 3	3 of 3
4-phenylenediamine	1A	1A	1A	1A	1A	1A	1B	1A	1A				1A	1A	1A		2 of 3	2 of 3
Tetrachlorsalicylanilide	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
Isoeugenol	1B	1B	1A	1B	1B	1A	1B	1A	1A				1B	1B	1B		0 of 3	0 of 3
Bourgeonal	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		3 of 3	3 of 3
Carvone	1B	1B	1B	1B	1B	1B	n-r	1B	1B	1B	1B	1B	1B	1B	1B	1B	3 of 4	4 of 4
Dihydrocoumarin	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		3 of 3	3 of 3
Hydroxycitronellal	n-r	n-r	n-r	1B	1B	1B	n-r	n-r	n-r	1B	1B	n-r	n-r	1B	n-r	1B	3 of 4	4 of 4
Imidazolidinyl urea	1A	1A	1A	1B	n-r	n-r	1A	1A	1A	1A	1A	1A	1A	1B	1A	1A	3 of 4	4 of 4
Methylhexanedione	1B	1B	1B	n-r	1B	1B	1B	1B	1B				1B	1B	1B		2 of 3	3 of 3
Perillaaldehyde	1B	1B	1B	1B	1B	1B	1B	1B	1B				1B	1B	1B		3 of 3	3 of 3
Phenyl benzoate	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	4 of 4	4 of 4
Phenylpropionaldehyde	1B	1B	n-r	1B	1B	1B	1B	1B	n-r				1B	1B	1B		1 of 3	3 of 3
Tetramethyldiuram disulfide	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Benzosiothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Benzylidene acetone	1B	1B	1B	1B	1B	1B	1B	1B	1B				1B	1B	1B		3 of 3	3 of 3
Delta Damascone	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	1B	4 of 4	4 of 4
Diethylmaleate	1B	1B	1B	1A	1A	1B	1B	1B	1B				1B	1A	1B		2 of 3	2 of 3
trans-2-hexenal	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	4 of 4	4 of 4
Chlorobenzene	1B	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		2 of 3	3 of 3
																consistent instances	72 of 84	79 of 84

n-r: non-reactive; [1]Indicates the rating of the chemical by the average log k_{max} determined from all repetitions in a particular lab, this value is used for inter-laboratory reproducibility of class prediction ; [2]3 class prediction model (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified); [3]2 class prediction model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

Table A5: Intra-laboratory reproducibility of the classification of chemicals with the refined prediction model

Substance	Lab 5 rep1	Lab 5 rep2	Lab 5 rep3	Lab 6 rep1	Lab 6 rep2	Lab 6 rep3	Lab 7 rep1	Lab 7 rep2	Lab 7 rep3	Lab 8 rep1	Lab 8 rep2	Lab 8 rep3	Lab 5 AVG1)	Lab 6 AVG	Lab 7 AVG	Lab 8 AVGoE	Labs with consistent repititions (3 classes)2)	Labs with consistent repitions (2 classes)3)
(Chloro)methylisothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Glyoxal	1B	1B	1B	1B	1A	1A	1B	1B	1B	1A	1A	1A	1B	1A	1B	1A	3 of 4	3 of 4
Methylisothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Methyl-2-octynoate	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
4-phenylenediamine	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
Tetrachlorsalicylanilide	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
Isoeugenol	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
Bourgeonal	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		3 of 3	3 of 3
Carvone	1B	1B	1B	1B	1B	1B	n-r	1B	1B	1B	1B	1B	1B	1B	1B	1B	3 of 4	4 of 4
Dihydrocoumarin	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		3 of 3	3 of 3
Hydroxycitronellal	n-r	n-r	n-r	1B	1B	1B	n-r	n-r	n-r	1B	1B	n-r	n-r	1B	n-r	1B	3 of 4	4 of 4
Imidazolidinyl urea	1A	1A	1A	1B	n-r	n-r	1A	1A	1A	1A	1A	1A	1A	1B	1A	1A	3 of 4	4 of 4
Methylhexanedione	1B	1B	1B	n-r	1B	1B	1B	1B	1B				1B	1B	1B		2 of 3	3 of 3
Perillaaldehyde	1B	1B	1B	1B	1B	1B	1B	1B	1B				1B	1B	1B		3 of 3	3 of 3
Phenyl benzoate	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	4 of 4	4 of 4
Phenylpropionaldehyde	1B	1B	n-r	1B	1A	1B	1B	1B	n-r				1B	1B	1B		0 of 3	2 of 3
Tetramethyldiuram disulfide	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Benzosiothiazolinone	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
Benzylidene acetone	1A	1A	1B	1A	1A	1A	1A	1A	1A				1B	1A	1A		2 of 3	2 of 3
Delta Damascone	1A	1A	1A	1B	1B	1B	1B	1B	1B	1B	1B	1B	1A	1B	1B	1B	4 of 4	4 of 4
Diethylmaleate	1A	1A	1A	1A	1A	1A	1A	1A	1A				1A	1A	1A		3 of 3	3 of 3
trans-2-hexenal	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	1A	4 of 4	4 of 4
4-methoxy-acetophenone	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r	4 of 4	4 of 4
Chlorobenzene	1B	n-r	n-r	n-r	n-r	n-r	n-r	n-r	n-r				n-r	n-r	n-r		2 of 3	3 of 3

n-r: non-reactive; 1)Indicates the rating of the chemical by the average log k_{max} determined from all repetitions in a particular lab, this value is used for inter-laboratory reproducibility of class prediction
2)3 class prediction model (GHS Cat 1A vs. GHS Cat 1B vs. GHS not classified); 3)2 class prediction model (GHS Cat 1A vs. GHS Cat 1B/GHS not classified)

Table A6. Intra-laboratory reproducibility of the 24 h / 5 mM depletion value [%]

Substance	Lab 5 rep1	Lab 5 rep2	Lab 5 rep3	Lab 6 rep1	Lab 6 rep2	Lab 6 rep3	Lab 7 rep1	Lab 7 rep2	Lab 7 rep3	Lab 8 rep1	Lab 8 rep2	Lab 8 rep3	AVG Lab 5	AVG Lab 6	AVG Lab 7	AVG Lab 8	SD Lab 5	SD Lab 6	SD Lab 7	SD Lab 8
(Chloro)methyliso-thiazolinone	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	0.0	0.0	0.0	0.0
Glyoxal	25.2	34.0	34.5	82.2	83.5	83.5	68.6	70.2	74.2	79.5	77.7	77.8	31.2	83.0	71.0	78.3	5.2	0.8	2.9	1.0
Methylisothiazo-linone	73.5	76.6	88.5	87.8	85.9	85.6	85.8	87.5	86.7	88.9	86.9	86.4	79.6	86.5	86.7	87.4	7.9	1.2	0.9	1.3
Methyl-2-octynoate	90.3	88.6	90.8	90.3	92.3	92.2	97.8	96.4	85.5				89.9	91.6	93.2		1.2	1.1	6.7	
4-phenylenediamine	73.8	70.0	81.6	68.1	62.6	63.3	64.7	68.5	66.4				75.1	64.7	66.5		5.9	3.0	1.9	
Tetrachlorsalicylani-lide	99.0	66.6	80.7	83.9	85.0	89.4	99.0	47.6	79.3				82.1	86.1	75.3		16.2	2.9	25.9	
Isoeugenol	71.9	62.1	70.3	60.8	54.3	81.9	72.9	78.5	73.2				68.1	65.7	74.9		5.3	14.4	3.2	
Bourgeonal	1.0	1.0	4.5	1.0	7.9	1.0	2.2	1.0	1.0				2.2	3.3	1.4		2.0	4.0	0.7	
Carvone	17.5	21.6	17.7	32.6	32.1	30.0	12.0	33.5	27.5	26.8	41.1	26.2	18.9	31.6	24.4	31.3	2.3	1.4	11.1	8.4
Dihydrocoumarin	3.9	5.9	7.5	12.4	11.7	10.4	4.9	3.7	13.4				5.8	11.5	7.4		1.8	1.0	5.3	
Hydroxycitronellal	6.7	1.0	6.6	26.1	36.0	15.2	1.0	3.4	8.7	14.9	41.7	6.7	4.8	25.8	4.4	21.1	3.3	10.4	4.0	18.3
Imidazolidinyl urea	48.2	52.2	45.2	27.6	1.3	1.0	14.4	30.2	42.0	18.8	24.7	26.2	48.5	10.0	28.9	23.2	3.5	15.2	13.8	3.9
Methylhexanedione	17.4	32.2	31.6	13.4	19.4	19.9	20.0	14.4	14.1				27.1	17.6	16.2		8.4	3.6	3.3	
Perillaaldehyde	45.8	20.6	26.8	23.5	19.8	18.5	14.0	22.2	15.6				31.1	20.6	17.3		13.1	2.6	4.4	
Phenyl benzoate	8.8	4.6	7.5	11.2	9.1	4.8	7.5	1.9	2.6	3.8	3.3	6.7	7.0	8.4	4.0	4.6	2.2	3.2	3.1	1.8
Phenylpropionalde-hyde	23.5	18.2	11.1	51.3	53.4	72.1	75.0	50.2	11.0				17.6	59.0	45.4		6.2	11.5	32.3	
Tetramethyldiuram disulfide	98.6	98.6	98.6	98.6	98.3	98.1	98.5	97.8	98.6	98.5	98.5	98.5	98.6	98.3	98.3	98.5	0.0	0.3	0.5	0.0
Benzosiothiazolinone	73.0	80.9	74.1	83.4	77.1	78.5	82.0	78.3	75.1	68.6	69.1	68.8	76.0	79.7	78.5	68.8	4.3	3.3	3.5	0.2
Benzylidene acetone	95.3	94.3	64.4	92.0	92.1	92.8	89.3	95.2	93.8				84.7	92.3	92.8		17.5	0.4	3.1	
Delta Damascone	90.3	90.9	92.9	86.6	76.3	77.9	64.9	66.0	65.7	70.9	80.5	77.7	91.4	80.3	65.5	76.4	1.3	5.5	0.6	4.9
Diethylmaleate	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.0				99.0	99.0	99.0		0.0	0.0	0.0	
trans-2-hexenal	97.8	96.3	96.3	95.8	96.9	97.5	95.1	89.3	88.9	89.8	92.3	88.7	96.8	96.8	91.1	90.2	0.8	0.9	3.5	1.8
4-methoxy-aceto-phenone	1.0	3.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	3.5	1.0	2.0	1.0	1.0	1.8	1.7	0.0	0.0	1.4
Chlorobenzene	21.2	1.0	6.5	1.0	1.0	7.4	1.0	1.0	1.3				9.5	3.1	1.1		10.4	3.7	0.2	

Table A7. Inter-laboratory reproducibility of the 24 h / 5 mM depletion value [%]. For laboratory 5 – 8, the average (AVG) from repeated testing is shown.

Substance	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	Lab 8	AVG all Labs	SD all Labs
(Chloro)methylisothiazo-linone	99.0	99.0	99.0		99.0	99.0	99.0	99.0	99.0	0.0
Glyoxal	81.0	81.2	64.7		31.2	83.0	71.0	78.3	70.1	18.3
Methylisothiazolinone	86.7	84.0	84.7		79.6	86.5	86.7	87.4	85.1	2.7
Methyl-2-octynoate	98.8	89.1	86.6	92.6	89.9	91.6	93.2		91.7	3.9
4-phenylenediamine	56.9	49.8	86.7	75.4	75.1	64.7	66.5		67.9	12.4
Tetrachlorsalicylanilide	56.7	75.3	85.2	86.0	82.1	86.1	75.3		78.1	10.5
Isoeugenol	71.6	56.1	60.8	80.0	68.1	65.7	74.9		68.2	8.2
Bourgeonal	1.0	10.6	1.0	12.9	2.2	3.3	1.4		4.6	5.0
Carvone	40.0	12.2	16.3		18.9	31.6	24.4	31.3	25.0	9.9
Dihydrocoumarin	8.6	11.2	1.0	12.6	5.8	11.5	7.4		8.3	4.0
Hydroxycitronellal	1.0	6.4	7.2		4.8	25.8	4.4	21.1	10.1	9.4
Imidazolidinyl urea	33.7	34.5	32.0		48.5	10.0	28.9	23.2	30.1	11.8
Methylhexanedione	34.0	16.4	25.5	39.0	27.1	17.6	16.2		25.1	9.0
Perillaaldehyde	15.1	24.9	30.3	34.2	31.1	20.6	17.3		24.8	7.4
Phenyl benzoate	1.0	2.3	1.0		7.0	8.4	4.0	4.6	4.0	2.9
Phenylpropionaldehyde	26.7	29.0	34.5	42.4	17.6	59.0	45.4		36.4	13.7
Tetramethyldiuram disul-fide	99.0	98.5	99.0		98.6	98.3	98.3	98.5	98.6	0.3
Benzosiothiazolinone	74.0	80.2	73.7		76.0	79.7	78.5	68.8	75.8	4.0
Benzylidene acetone	88.6	94.8	94.9	91.3	84.7	92.3	92.8		91.3	3.6
Delta Damascone	77.6	74.8	63.2		91.4	80.3	65.5	76.4	75.6	9.4
Diethylmaleate	99.0	99.0	99.0	99.0	99.0	99.0	99.0		99.0	0.0
trans-2-hexenal	90.1	94.6	88.4		96.8	96.8	91.1	90.2	92.6	3.4
4-methoxy-acetophenone	3.5	47.7	2.8		2.0	1.0	1.0	1.8	8.5	17.3
Chlorobenzene	1.0	1.0	1.7	11.6	9.5	3.1	1.1		4.2	4.5

Table A8. Comparison of 24 h/ 5 mM %-depletion values and variance with historical data for % Cys-depletion and variance in the classical DPRA.

	kDPRA		DPRA Prevalidation study		DPRA ECCVAM Validation Study		Consolidated historical results [1)]
Substance	**Average all Labs**	**SD all Labs**	**Average all Labs**	**SD all Labs**	**Average all Labs**	**SD all Labs**	
(Chloro)methylisothiazolinone	99.0	0.0	85.6	21.0			96.3
Glyoxal	70.1	18.3	82.5	8.8			56.5
Methylisothiazolinone	85.1	2.7					97.9
Methyl-2-octynoate	91.7	3.9					97.2
4-phenylenediamine	67.9	12.4	92.1	8.5	92.5	7.4	95.3
Isoeugenol	68.2	8.2	79.2	23.2			92.6
Bourgeonal	4.6	5.0					17.7
Carvone	25.0	9.9					25.7
Dihydrocoumarin	8.3	4.0					0.0
Hydroxycitronellal	10.1	9.4	24.6	15.7			32.3
Imidazolidinyl urea	30.1	11.8	39.3	11.4	54.1	5.6	38.4
Methylhexanedione	25.1	9.0	39.8	18.5			25.8
Perillaaldehyde	24.8	7.4					31.9
Phenyl benzoate	4.0	2.9					50.9
Phenylpropionaldehyde	36.4	13.7	70.7	23.3			37.4
Tetramethyldiuram disulfide	98.6	0.3					99.5
Benzosiothiazolinone	75.8	4.0					97.7
Benzylidene acetone	91.3	3.6	91.2	1.7			93.5
Delta Damascone	75.6	9.4					
Diethylmaleate	99.0	0.0	99.8	0.2			99.9
trans-2-hexenal	92.6	3.4					97.9
4-methoxy-acetophenone	8.5	17.3	1.5	3.7			2.4
Chlorobenzene	4.2	4.5					0.4
DNCB	99.0	0.1	99.7	0.4			100.0
Oxazolone	84.8	3.6	73.6	3.4			73.6
Formaldehyde	41.4	8.9	47.3	8.3	41.8	7.6	58.6
EGDMA	73.5	11.7					89.3
Benzylideneacetone (transfer phase)	87.3	13.1	91.2	1.7			93.5
2-3 Butanedione	47.3	12.1	80.3	13.9			85.9
Cinnamic aldehyde	69.3	5.7	79.1	11.9			70.6
Tetrachlorsalicylanilide	autofl. (78)	10.5					36.8
Average SD common chemicals		8.7		10.3			

1) From Urbisch et al and Hofmann et al.

Table A9. Intra-laboratory congruency of solvent-decision. WoE solvent is given only, when the decision was consistent in the lab in all three runs; otherwise n/a is given. Water and buffer are counted as "consistent" decision.

Substance	Lab 5 rep1	Lab 5 rep2	Lab 5 rep3	Lab 6 rep1	Lab 6 rep2	Lab 6 rep3	Lab 7 rep1	Lab 7 rep2	Lab 7 rep3	Lab 8 rep1	Lab 8 rep2	Lab 8 rep3	WoE Lab 5	WoE Lab 6	WoE Lab 7	WoE Lab 8	Consistency
(Chloro)methylisothiazolinone	Buffer	Buffer	Buffer	Buffer	Buffer	Buffer	ACN	ACN	ACN	Buffer	Buffer	ACN	Buffer	Buffer	ACN	n/a	3 of 4
Glyoxal	ACN	ACN	ACN	H_2O	H_2O	H_2O	Buffer	Buffer	ACN	Buffer	Buffer	Buffer	ACN	H_2O	n/a	Buffer	3 of 4
Methylisothiazolinone	ACN	ACN	H_2O	ACN	H2O	ACN	Buffer	Buffer	Buffer	Buffer	Buffer	Buffer	n/a	n/a	Buffer	Buffer	2 of 4
Methyl-2-octynoate	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
4-phenylenediamine	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Tetrachlorsalicylanilide	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Isoeugenol	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Bourgeonal	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Carvone	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Dihydrocoumarin	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Hydroxycitronellal	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Imidazolidinyl urea	Buffer	Buffer	Buffer	ACN	H_2O	H_2O	H_2O	H_2O	H_2O	Buffer	Buffer	Buffer	Buffer	n/a	H_2O	Buffer	3 of 4
Methylhexanedione	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Perillaaldehyde	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Phenyl benzoate	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Phenylpropionaldehyde	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Tetramethyldiuram disulfide	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Benzosiothiazolinone	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Benzylidene acetone	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
Delta Damascone	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Diethylmaleate	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3
trans-2-hexenal	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
4-methoxy-acetophenone	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	4 of 4
Chlorobenzene	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN	ACN				ACN	ACN	ACN		3 of 3

n/a: indicates that the laboratory chose different vehicles in intra-laboratory repetitions

Table A10. Full study setup of which lab tested which chemical in intra-and inter-lab testing and the ‘Dummy number’ 1 – 8 attributed to the experiments

Which lab was which number?	**Interlaboratory testing**					**Intralaboratory testing**											
Substance	**Lab 1**	**Lab 2**	**Lab 3**	**Lab 4**		**Lab 5 rep1**	**Lab 5 rep2**	**Lab 5 rep3**	**Lab 6 rep1**	**Lab 6 rep2**	**Lab 6 rep3**	**Lab 7 rep1**	**Lab 7 rep2**	**Lab 7 rep3**	**Lab 8 rep1**	**Lab 8 rep2**	**Lab 8 rep3**
(Chloro)methylisothiazolinone	Lab F	Lab D	Lab C			Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab G	Lab G	Lab G
Glyoxal	Lab A	Lab B	Lab D			Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
Methylisothiazolinone	Lab A	Lab B	Lab F			Lab E	Lab E	Lab E	Lab D	Lab D	Lab D	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
Methyl-2-octynoate	Lab E	Lab D	Lab C	Lab G		Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab F	Lab F	Lab F			
4-phenylenediamine	Lab A	Lab E	Lab F	Lab G		Lab B	Lab B	Lab B	Lab D	Lab D	Lab D	Lab C	Lab C	Lab C			
Tetrachlorsalicylanilide	Lab E	Lab F	Lab C	Lab G		Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab D	Lab D	Lab D			
Isoeugenol	Lab A	Lab B	Lab E	Lab G		Lab F	Lab F	Lab F	Lab D	Lab D	Lab D	Lab C	Lab C	Lab C			
Bourgeonal	Lab E	Lab F	Lab D	Lab G		Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab C	Lab C	Lab C			
Carvone	Lab A	Lab B	Lab C			Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab D	Lab D	Lab D	Lab G	Lab G	Lab G
Dihydrocoumarin	Lab E	Lab F	Lab C	Lab G		Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab D	Lab D	Lab D			
Hydroxycitronellal	Lab A	Lab B	Lab D			Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
Imidazolidinyl urea	Lab A	Lab D	Lab C			Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab G	Lab G	Lab G
Methylhexanedione	Lab B	Lab E	Lab F	Lab G		Lab A	Lab A	Lab A	Lab D	Lab D	Lab D	Lab C	Lab C	Lab C			
Perillaaldehyde	Lab A	Lab B	Lab E	Lab G		Lab F	Lab F	Lab F	Lab D	Lab D	Lab D	Lab C	Lab C	Lab C			
Phenyl benzoate	Lab F	Lab D	Lab C			Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab G	Lab G	Lab G
Phenylpropionaldehyde	Lab B	Lab E	Lab D	Lab G		Lab A	Lab A	Lab A	Lab F	Lab F	Lab F	Lab C	Lab C	Lab C			
Tetramethyldiuram disulfide	Lab A	Lab D	Lab C			Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab G	Lab G	Lab G
Benzosiothiazolinone	Lab B	Lab F	Lab C			Lab A	Lab A	Lab A	Lab E	Lab E	Lab E	Lab D	Lab D	Lab D	Lab G	Lab G	Lab G
Benzylidene acetone	Lab E	Lab F	Lab C	Lab G		Lab A	Lab A	Lab A	Lab B	Lab B	Lab B	Lab D	Lab D	Lab D			
Delta Damascone	Lab A	Lab B	Lab D			Lab E	Lab E	Lab E	Lab F	Lab F	Lab F	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
Diethylmaleate	Lab B	Lab E	Lab C	Lab G		Lab A	Lab A	Lab A	Lab F	Lab F	Lab F	Lab D	Lab D	Lab D			
trans-2-hexenal	Lab A	Lab F	Lab D			Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
4-methoxy-acetophenone	Lab A	Lab F	Lab D			Lab B	Lab B	Lab B	Lab E	Lab E	Lab E	Lab C	Lab C	Lab C	Lab G	Lab G	Lab G
Chlorobenzene	Lab B	Lab E	Lab C	Lab G		Lab A	Lab A	Lab A	Lab F	Lab F	Lab F	Lab D	Lab D	Lab D			

Appendix 2: Result interpretation in case of fluorescence quenching and potential control experiments

The alert 'depletion may be due to quenching' was repeatedly triggered – in 23 instances in total. It was repeatedly triggered for isoeugenol (5 instances), PPD (9 instances), BI (7 instances) and TMD (2 instances). This alert is triggered, if the mBrB background fluorescence decreases in control wells with test chemical only at increasing concentration of the test chemical AND if the apparent depletion over the early time points is not progressing in a time-dependent manner. In such an instance, a depletion at 10 min may be due to fluorescence interference, and it will not increase in the next measurement points as it is only due to an optical effect.

For the analysis of reproducibility, the data from these instances were taken at face value directly as they were calculated by the evaluation spreadsheet. However, for predictivity, it is important to carefully evaluate such data for artefacts which may suggest a higher reactivity which is due to a fluorescence quenching/absorbance and not really depending on peptide reactivity. Thus, we did consider this effect when discussing predictivity.

Below this effect is evaluated for the three chemicals with the alert noted in multiple instances.

a) PPD

Figure A1 shows the results from two laboratories for PPD. Both the depletion matrix and the fluorescence in the blank wells without peptide (test chemical an mBrB only) are shown. A similar effect was noted in most experiments.

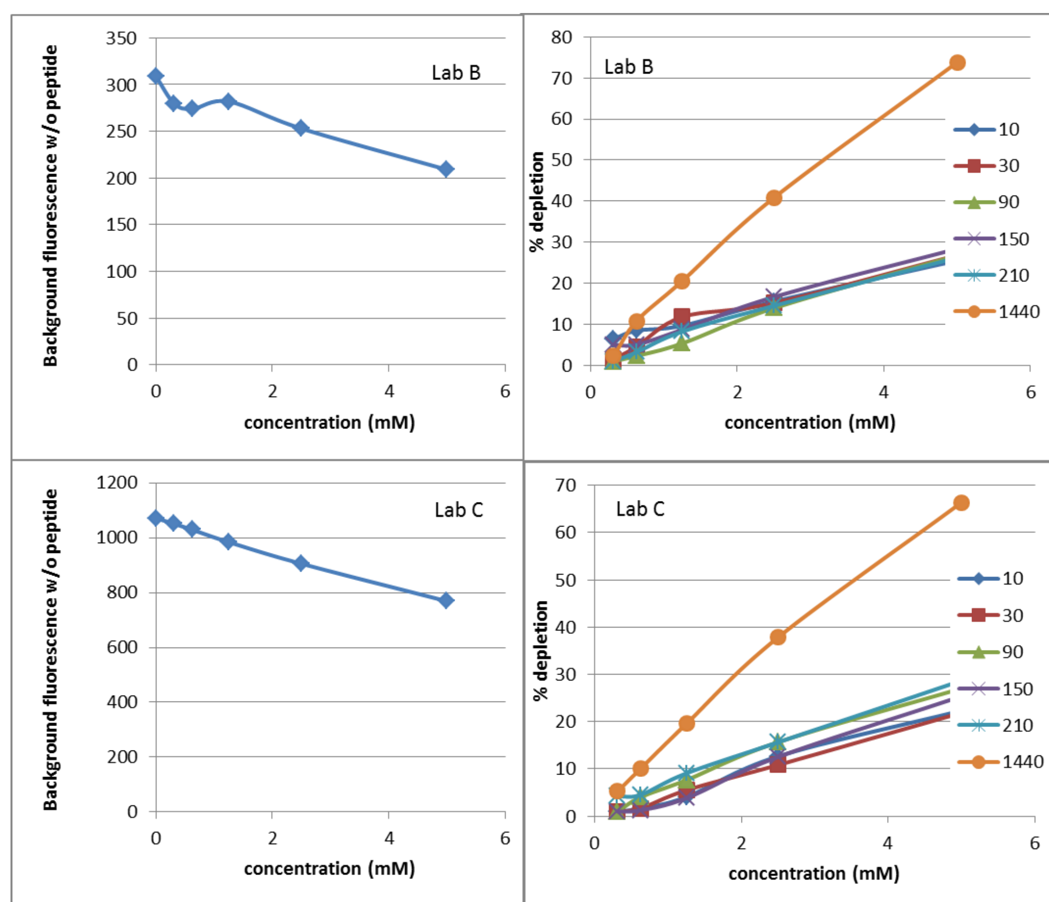


Figure A1. PPD, a chemical for which the alert "Depletion may be due to quenching" is triggered when tested in Lab B and Lab C. Left: Background fluorescence in absence of peptide, right depletion at different time and concentration)

From these data it is obvious, that instantaneous apparent 20% depletion is occurring at 10 min, which is not significantly increasing over the early time points (Figure A1, right panel), at the same time background fluorescence is 20% reduced by the highest test concentration (Figure A1, left panel). In this case therefore reactivity is apparently overestimated. To give an approximation of the true reaction rate, we thus subtracted the depletion value after 10 min from the other depletion values and calculated the rate from the residual, time-dependent depletion. This gave a clearly higher log k_{max} value (-2.81) as compared to the average value from all labs when not corrected (-1.16). Ideally, this chemical would be retested with an alternative fluorescent probe being excited at a higher wavelength to avoid this problem. But based on the data from Figure A1, the refined calculated rate may be good approximation. PPD is known to act as a pre-hapten, and from the data in Figure A1 it appears that it requires a few hours for oxidation to occur until reactivity really sets in.

b) Isoeugenol

For isoeugenol, some laboratories reported the alert 'depletion may be due to quenching' triggered in some, but not all instances, while this was not the case for others. Figure A2 shows two cases the alert was triggered and two where the alert was not triggered, each.

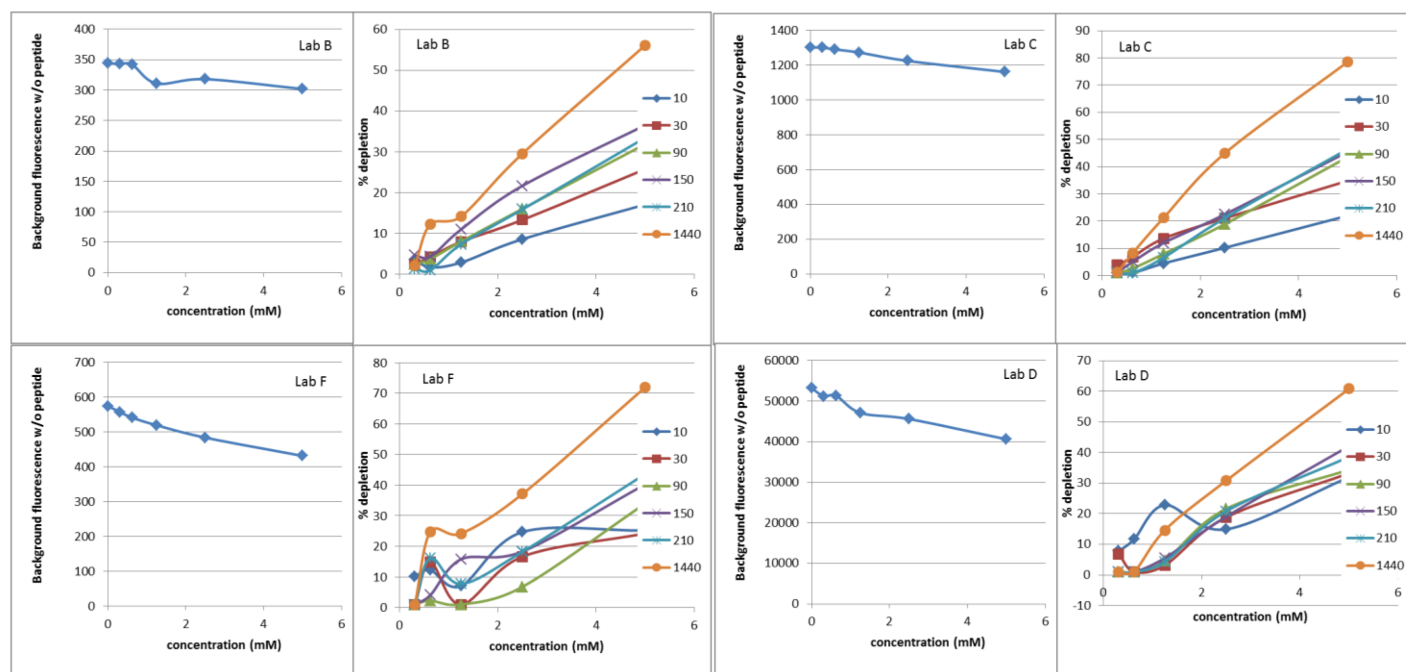


Figure A2. Isoeugenol tested in Lab B and Lab C (no quenching alert) and Lab F and D (quenching alert triggered in that particular experiment).

The results of Lab B and C indicate that isoeugenol clearly is a reactive molecule with the depletion increasing over time starting at 10 min, with little reduction of fluorescence in the blank wells. In Lab D and F, more reduction of the background fluorescence was observed, and more importantly, the kinetics of the reaction was not really time dependent over the first 210 min. With the noisy and non-linear data in Lab F one would ideally repeat such an experiment to clarify putative quenching. Nevertheless, based on the non-linearity of the data at 10 min and 30 min, the evaluation sheet ignores these data and calculated the rate constant based on the data form 90 min onwards, with still a correct GHS classification according to the refined cut-off.

c) Benzisothiazolinone

For benzisothiazolinone, reduction of background fluorescence was observed in some, but not all labs (Figure A3). At the same time a very strong depletion was observed even at low concentration of the test chemical already at 10 min, which cannot be explained by the weak or lacking fluorescence interference at those concentrations. The depletion matrix was not influence by the fact whether a lab observed the quenching alert or not, and one can safely conclude that the kinetic rate is really reflecting the actual peptide depletion.

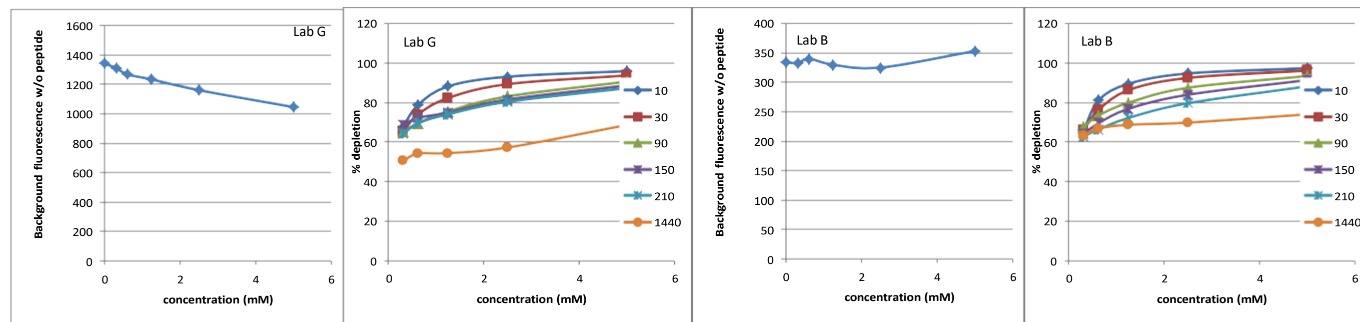


Figure A3. Benzisothiazolinone tested Lab G (quenching alert triggered in that particular experiment) and in Lab B (no quenching alert)

Non-linear peptide depletion observed for other chemicals when establishing the database

In few instances, peptide depletion above the DPRA cut-off for the cysteine peptide only (13.89% depletion) was observed at 10 min, and no time dependent increase thereafter, with absent or still very weak depletion at 24 h. In such a case this very weak and non-linear depletion could also be due to an optical interference, in such cases the alert "depletion not time dependent" is triggered.

A rapid depletion within 10 min, which does not increase over time, should thus be evaluated carefully as it could also be due to fluorescence interference, and in general such non-kinetic depletion should be scrutinized. To test for such effects, a further control experiment was devised:

Method:

- Chemicals are retested at 5 mM / 10 min only
- A parallel experiment is conducted in a second plate, whereby the mBrB is added to the test peptide before the test chemical, and after 10 min the chemical is added, and the reaction is read after another 10 min.
  ➔ In this experiment the peptide is thus reacted with mBrB and cannot further react with the test chemical. Thus, no significant peptide depletion is expected for a chemical which is reactive in the original assay. If on the second plate a very similar depletion is noted as in the first plate, this indicates that the test chemical does reduce the fluorescence of the peptide-mBrB adduct due to fluorescence quenching / absorbance, and the time-independent depletion is not resulting from true peptide reactivity.

Examples of chemicals with non-time-dependent depletion tested in this control experiment are listed in Table A10 and are discussed below:

Table A10. Chemicals tested for quenching/absorbance of the mBrB-peptide adduct vs. depletion at 10 min in the standard assay run in parallel.

	Depletion standard assay, 5 mM, 10 min	Depletion, mBrB added prior to test chemical [1)]

Chemicals with time independent depletion		
Pentachlorophenol	26.9%	23.5%
Salicylic acid	17.62%	11.78%
Vanillin	31.5%	35.7%
4-Chloroaniline	11.3%	17.3%
Tropolone	91.4%	91.7%
Oxalic acid	20.9%	1.4%
Controls: non-reactive Chemicals		
Methyl salicylate	-1.8%	-3.4%
Anethole	5.4%	6.4%
Resorcinol	5.7%	7.2%
Penicillin G	-7.1%	-4.8%
Positive chemicals: rapidly reactive chemicals		
Oxazolone	90.6%	13.1%
2,4-Dinitrochlorobenzene	50.8%	21.6%
Formaldehyde	56.2%	12.2%

1) Peptide is reacted with mBrB for 10 min, and then test chemical is added and fluorescence read 10 min later.

a) Pentachlorophenol

As shown in Table 10 (see WP 3.2.), pentachlorophenol results in time-independent depletion of 18.3 (at 150 min) – 27.4 (at 1440 min) % at 5 mM. When pentachlorophenol is added after addition of the mBrB, similar depletion is noted, indicating again that this depletion is due to interference with the fluorescence of the mBrB-peptide adduct (see Table A10) as concluded before and verified with this additional control experiment.

b) Vanillin and Tropolone

For vanillin and tropolone a time-independent depletion matrix was observed (Figure 4A). For Tropolone the quenching alert is triggered, while for vanillin only the time-independent depletion alert is triggered. Depletion is at approximately 35% and 90%, respectively, at the 5 mM concentration independently of the reaction time in the full assay. As shown in Table A10, comparable peptide depletions are observed in the control experiment independently whether the mBrB is added before or after reaction of the chemical with the peptide. For the predictivity analysis, these two chemicals were thus excluded due to fluorescence quenching / absorbance, although based on the data one may also conclude that they are non-reactive.

c) Salicylic acid and 4-chloroaniline

These two chemicals gave ca. 18% (salicylic acid) and 14% (4-chloraniline) depletion at the 5 mM concentration which again was not time-dependent (Figure 4A). As shown in Table A10, a comparably weak depletion was again noted when mBrB was added prior to the chemical, indicating that we observe a weak absorbance effect. Based on this control experiment and the non-kinetic data-matrix in Figure A4 we can conclude that these are non-reactive chemicals.

d) Oxalic acid

Oxalic acid also resulted in a time-independent peptide depletion. However, in this case the control experiment indicated no depletion if mBrB reacted first with the peptide, thus we have to conclude that the observed depletion is due to the interaction with the thiol group of the peptide and we accept the data as calculated by the spreadsheet.

e) Controls

Three reactive chemicals from the transfer phase were tested with the control experiment. For DNCB, oxazolone and formaldehyde we observe clearly strong peptide depletion in the original assay, with only residual depletion, when the peptide is first reacted with the mBrB.

Four negative controls showed no depletion whether mBrB was first reacted with the peptide or the peptide was first in contact with the test chemical (Table A10).

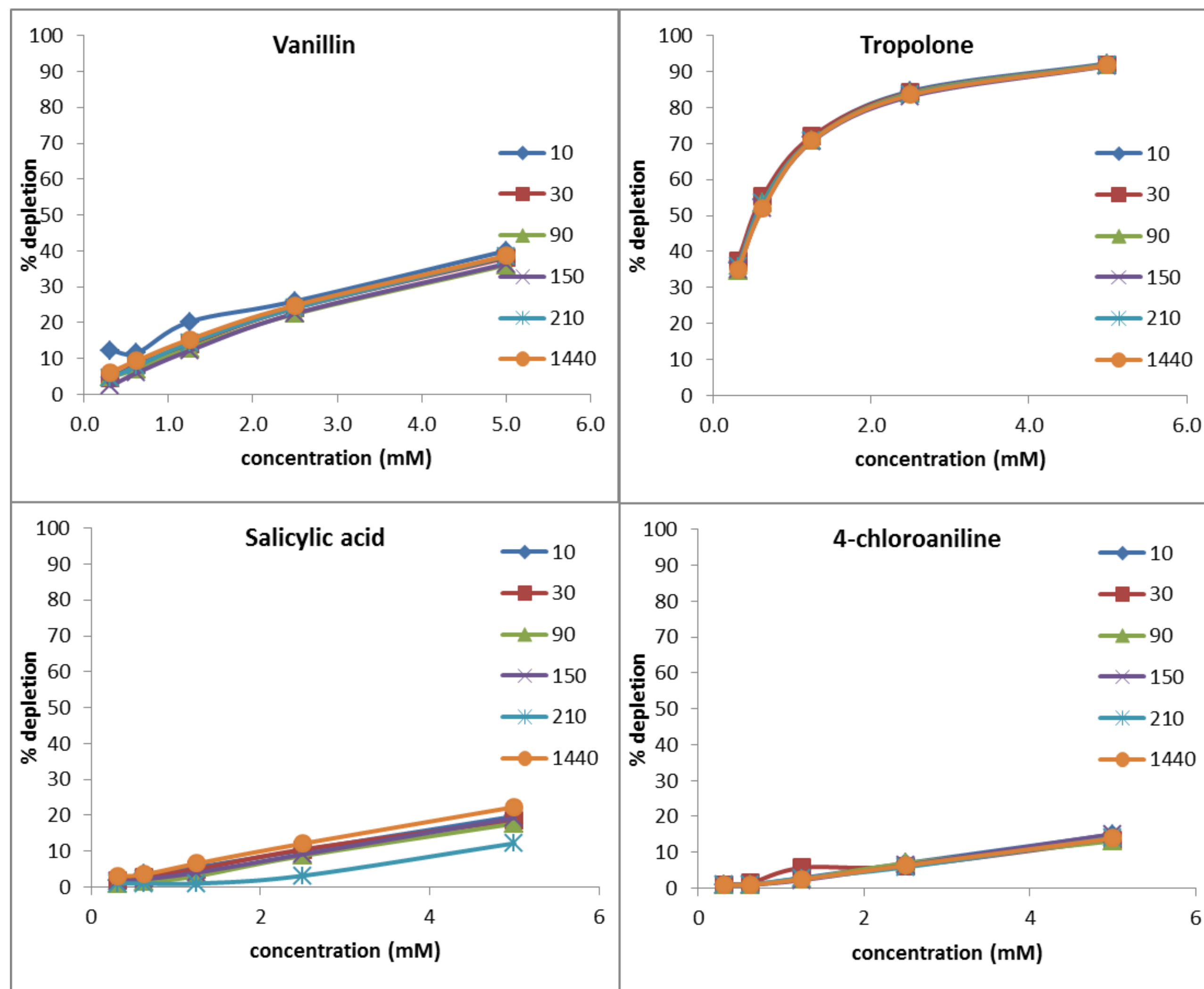


Figure 4A. Depletion matrix for 4 chemicals which had time independent depletion alert triggered.

Appendix 3. Extended dataset to evaluate predictivity and to determine optimal GHS 1A vs. GHS 1B/NC cut-off

Table A11: Extended database of 186 chemicals tested.

Name	**CAS**	**LLNA EC3 [%]**	**GHS Cat LLNA**	**GHS Cat Human**	**log k_{max} Numerical**	**GHS Potency classification kDPRA published cut-off**	**GHS Potency Classification kDPRA revised cut-off**
From ICCVAM List on validation of LLNA potency (Annex II-4 in ICCVAM report): LLNA Cat 1A and human Cat 1A							
(Chloro)methylisothiazolinone	26172-55-4	0.01	1A	1A	0.60	1A	**1A**
Cinnamic aldehyde	104-55-2	1.15	1A	1A	-1.35	1B/NC	**1A**
2,4-dinitrochlorobenzene	97-00-7	0.06	1A	1A	-0.56	1A	**1A**
Formaldehyde	50-00-0	0.85	1A	1A	-0.67	1A	**1A**
Glyoxal	107-22-2	0.60	1A	1A	-1.97	1B/NC	**1A**
Methylisothiazolinone	2682-20-4	0.40	1A	1A	-0.25	1A	**1A**
Methyl-2-octynoate	111-12-6	0.45	1A	1A	-1.56	1B/NC	**1A**
4-phenylenediamine	106-50-3	0.15	1A	1A	-2.81	1B/NC	**1B/NC**
From ICCVAM List on validation of LLNA potency (Annex II-4 in ICCVAM report): LLNA Cat 1B and human Cat 1B							
Chlorpromazine	69-09-0	1.00	1A	1B	not reactive	1B/NC	**1B/NC**
Aniline	62-53-3	89.00	1B	1B	not reactive	1B/NC	**1B/NC**
Benzocaine	94-09-7	>50	NC	1B	not reactive	1B/NC	**1B/NC**
Bourgeonal	18127-01-0	4.30	1B	1B	not reactive	1B/NC	**1B/NC**
Carvone	6485-40-1	12.90	1B	1B	-2.78	1B/NC	**1B/NC**
Cinnamyl alcohol	104-54-1	20.60	1B	1B	not reactive	1B/NC	**1B/NC**
Citral	5392-40-5	5.70	1B	1B	not reactive	1B/NC	**1B/NC**
Citronellol	106-22-9	43.50	1B/NC	1B/NC	not reactive	1B/NC	**1B/NC**
Coumarin	91-64-5	>50	1B/NC	1B	not reactive	1B/NC	**1B/NC**
Dihydrocoumarin	119-84-6	5.60	1B	1B	not reactive	1B/NC	**1B/NC**
Ethyl acrylate	140-88-5	32.75	1B	1B	-0.97	1A	**1A**
Ethylenediamine	107-15-3	3.40	1B	1B	not reactive	1B/NC	**1B/NC**
Eugenol	97-53-0	12.90	1B	1B	-2.64	1B/NC	**1B/NC**
Farnesol	4602-84-0	4.80	1B	1B	-3.41	1B/NC	**1B/NC**
Geraniol	106-24-1	23.20	1B	1B	-3.41	1B/NC	**1B/NC**
Hydroxycitronellal	107-75-5	22.20	1B	1B	-2.77	1B/NC	**1B/NC**
Imidazolidinyl urea	39236-46-9	24.00	1B	1B	-1.11	1B/NC	**1A**
Lilial	80-54-6	11.35	1B	1B	not reactive	1B/NC	**1B/NC**
2-mercapotobenzothiazole	149-30-4	2.60	1B	1B	-0.15	1A	**1A**
Methylhexanedione	13706-86-0	26.00	1B	1B	-2.67	1B/NC	**1B/NC**
Methylhydrocinnamal	5406-12-2	14.00	1B	1B	not reactive	1B/NC	**1B/NC**
Penicillin G	61-33-6	30.00	1B	1B	not reactive	1B/NC	**1B/NC**
Pentachlorophenol	87-86-5	20.00	1B	1B/NC	not reactive	1B/NC	**1B/NC**
Perillaaldehyde	2111-75-3	4.04	1B	1B	-2.93	1B/NC	**1B/NC**
Phenyl benzoate	93-99-2	18.33	1B	1B	not reactive	1B/NC	**1B/NC**
Phenylpropionaldehyde	93-53-8	6.30	1B	1B	-2.68	1B/NC	**1B/NC**
Propylidene phtalide	17369-59-4	3.40	1B	1B	not reactive	1B/NC	**1B/NC**
Tetramethyldiuram disulfide	137-26-8	2.93	1B	1B	0.74	1A	**1A**

Table A11 (cont.)

Name	CAS	LLNA EC3 [%]	GHS Cat LLNA	GHS Cat Human	log k_{max} Numerical	Potency classification kDPRA published cut-off	Potency Classification kDPRA reviseed cut-off
From ICCVAM List on validation of LLNA potency (Annex II-4 in ICCVAM report): LLNA 1B and human Cat 1A							
Benzosiothiazolinone	2634-33-5	4.80	1B	1A	-0.12	1A	**1A**
Benzylidene acetone	122-57-6	3.70	1B	1A	-1.85	1B/NC	**1A**
Butyl glycidyl ether	2426-08-6	30.90	1B	1B	-2.73	1B/NC	**1B/NC**
Delta Damascone	57378-68-4	3.55	1B	1A	-2.16	1B/NC	**1B/NC**
Diethylenetriamine	111-40-0	3.28	1B	1A	not reactive	1B/NC	**1B/NC**
Diethylmaleate	141-05-9	4.70	1B	1A	-1.21	1B/NC	**1A**
2-hexylidene cyclopentanone	17373-89-6	2.40	1B	1A	-2.36	1B/NC	**1B/NC**
Methylanisylidene acetone	104-27-8	9.30	1B	1A	-3.10	1B/NC	**1B/NC**
Phenylacetaldehyde	122-78-1	4.70	1B	1A	-2.36	1B/NC	**1B/NC**
trans-2-hexenal	6728-26-3	4.05	1B	1A	-0.47	1A	**1A**
From ICCVAM List on validation of LLNA potency (Annex II-4 in ICCVAM report): LLNA Cat 1A and human Cat 1B							
Benzoyl peroxide	94-36-0	0.06	1A	1B	0.74	1A	**1A**
Glutaric aldehyde	111-30-8	0.09	1A	1A	not reactive	1B/NC	**1B/NC**
Isoeugenol	97-54-1	1.35	1A	1A	-1.24	1B/NC	**1A**
Performance standards of LLNA or KeratinoSens, but not in above list, excluding metals							
4-methoxy-acetophenone	100-06-1	>50	NC		not reactive	1B/NC	**1B/NC**
Glycerol	56-81-5	Nd	NC	NC	not reactive	1B/NC	**1B/NC**
Isopropanol	67-63-0	>50	NC	NC	not reactive	1B/NC	**1B/NC**
Salicylic acid	69-72-7	>25	NC	NC	not reactive	1B/NC	**1B/NC**
Chlorobenzene	108-90-7	>25	NC		not reactive	1B/NC	**1B/NC**
Lactic acid	50-21-5	>25	NC	NC	not reactive	1B/NC	**1B/NC**
Methyl salicylate	119-36-8	>20	NC	NC	not reactive	1B/NC	**1B/NC**
Sulfanilamide	63-74-1	NC	NC		not reactive	1B/NC	**1B/NC**
Ethylene glycol dimethacrylate	97-90-5	35.00	1B	1B	-2.44	1B/NC	**1B/NC**
1,2-Dibromo-2,4-dicyanobutane	35691-65-7	0.90	1A	1A	-0.55	1A	**1A**
4-Nitrobenzyl bromide	100-11-8	0.05	1A		-0.01	1A	**1A**
4-Methylaminophenol sulphate	55-55-0	0.80	1A	1B	-0.25	1A	**1A**
4-Ethoxymethylene-2-phenyl-2-oxazolin-5-one (Oxazolone)	15646-46-5	0.00	1A		-0.14	1A	**1A**
α-Hexyl cinnamic aldehyde	101-86-0	10.60	1B	NC	not reactive	1B/NC	**1B/NC**
Methylmethacrylate	80-62-6	75.00	1B	1B	-2.00	1B/NC	**1B/NC**
Sodium lauryl sulfate (SDS)	151-21-3	3.55	1B	NC	not reactive	1B/NC	**1B/NC**
Xylene	1330-20-7	95.80	1B	NC	not reactive	1B/NC	**1B/NC**

Table A11 (cont.)

Name	CAS	LLNA EC3 [%]	GHS Cat LLNA	GHS Cat Human	log k_{max} Numerical	Potency classification kDPRA published cut-off	Potency Classification kDPRA reviseed cut-off
Additional chemicals CosEu database excluding metals and multicomponent mixtures							
Lauryl gallate	1166-52-5	0.30	1A	1A	-0.98	1A	**1A**
Diphenylcyclopropenone	886-38-4	0.05	1A	1A	-0.96	1A	**1A**
Dimethyl fumarate	624-49-7	0.35	1A	1A	-0.27	1A	**1A**
1,4-Dihydroquinone	123-31-9	0.12	1A	1B	-1.62	1B/NC	**1A**
Propyl gallate	121-79-9	0.32	1A	1A	-1.96	1B/NC	**1A**
2-aminophenol	95-55-6	0.45	1A	1A	-2.46	1B/NC	**1B/NC**
Iodopropynyl butylcarbamate	55406-53-6	0.90	1A	1B	0.75	1A	**1A**
Toluene diamine sulphate	615-50-9	0.40	1A	1A	-1.96	1B/NC	**1A**
2-Hydroxyethyl acrylate	818-61-1	1.56	1A	1B	-0.39	1A	**1A**
3-Dimethylaminopropylamine	109-55-7	2.20	1B	1A	not reactive	1B/NC	**1B/NC**
Resorcinol	108-46-3	6.30	1B	1B	not reactive	1B/NC	**1B/NC**
Amyl cinnamic aldehyde	122-40-7	11.20	1B	1B	-2.48	1B/NC	**1B/NC**
Abietic acid	514-10-3	11.00	1B	1B	-0.55	1A	**1A**
Linalool	78-70-6	30.40	1B	1B	not reactive	1B/NC	**1B/NC**
Amylcinnamyl alcohol	101-85-9	>25	NC	1B	not reactive	1B/NC	**1B/NC**
Hexyl salicylate	6259-76-3	0.18	1A	1B	not reactive	1B/NC	**1B/NC**
Neomycin sulphate	1405-10-3	Ns	NC	1B	not reactive	1B/NC	**1B/NC**
1-Butanol	71-36-3	>20	NC	NC	not reactive	1B/NC	**1B/NC**
Dextran	3371-50-4	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Diethyl phthalate	84-66-2	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Octanoic acid	124-07-2	>50	NC	NC	not reactive	1B/NC	**1B/NC**
Propylene glycol	57-55-6	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Propyl paraben	94-13-3	>50	NC	NC	not reactive	1B/NC	**1B/NC**
Tween 80	9005-65-6	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Anethole	104-46-1	2.30	1B	NC	not reactive	1B/NC	**1B/NC**
4-Aminobenzoic acid	150-13-0	>10	NC	NC	not reactive	1B/NC	**1B/NC**
Anisyl alcohol	105-13-5	5.91	1B	NC	not reactive	1B/NC	**1B/NC**
Benzyl benzoate	120-51-4	17	1B	NC	not reactive	1B/NC	**1B/NC**
Benzyl salicylate	118-58-1	2.85	1B	NC	not reactive	1B/NC	**1B/NC**
Diethanolamine	111-42-2	40.00	1B	NC	not reactive	1B/NC	**1B/NC**
Hydrocortisone	50-23-7	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
(R)-(+)-Limonene	5989-27-5	52.50	1B	NC	not reactive	1B/NC	**1B/NC**
Phenoxyethanol	122-99-6	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Triethanolamine	102-71-6	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Diethyl toluamide	134-62-3	>60	NC	NC	not reactive	1B/NC	**1B/NC**
Phenol	108-95-2	Ns	NC	NC	not reactive	1B/NC	**1B/NC**
Tocopherol	59-02-9	8.10	1B	NC	not reactive	1B/NC	**1B/NC**
DMSO	67-68-5	72.00	1B	NC	not reactive	1B/NC	**1B/NC**
Benzaldehyde	100-52-7	>25	NC	NC	not reactive	1B/NC	**1B/NC**

Table A11 (cont.)

Name	CAS	LLNA EC3 {%]	GHS Cat LLNA	GHS Cat Human	log k_{max} Numerical	Potency classification kDPRA published cut-off	Potency Classification kDPRA reviseed cut-off
Isopropyl myristate	110-27-0	44.00	1B	NC	not reactive	1B/NC	**1B/NC**
Bisphenol A-diglycidyl ether	1675-54-3	1.50	1A	1B	-2.53	1B/NC	**1B/NC**
Lyral	31906-04-4	17.10	1B	1A	-3.31	1B/NC	**1B/NC**
Allyl phenoxyacetate	7493-74-5	3.10	1B	1B	not reactive	1B/NC	**1B/NC**
Cinnamyl nitrile	1885-38-7	>25	NC	1B	-3.49	1B/NC	**1B/NC**
Dibenzyl ether	103-50-4	6.3	1B	1B	not reactive	1B/NC	**1B/NC**
Isocyclocitral	1335-66-6	7.35	1B	1B	-2.80	1B/NC	**1B/NC**
4-Methoxy-α-methyl benzenpropanal	5462-06-6	23.63	1B	1B	-2.45	1B/NC	**1B/NC**
α-Methyl-1,3-benzodioxole- 5-propionaldehyde	1205-17-0	16.40	1B	NC	-3.14	1B/NC	**1B/NC**
6-Methyl-3,5-heptadien-2-one	1604-28-0	>5	NC	1A	-3.29	1B/NC	**1B/NC**
α-iso-Methylionone	127-51-5	21.80	1B	NC	not reactive	1B/NC	**1B/NC**
OTNE	54464-57-2	25.14	1B	NC	not reactive	1B/NC	**1B/NC**
Benzyl Alcohol	100-51-6	>50	NC	1B	not reactive	1B/NC	**1B/NC**
Benzyl Cinnamate	103-41-3	18.4	1B	1B	not reactive	1B/NC	**1B/NC**
p-Isobutyl-α-methyl hydrocinnamaldehdye	6658-48-6	9	1B	1B	not reactive	1B/NC	**1B/NC**
α-Methyl cinnamic aldehyde	101-39-3	4.50	1B	1B	not reactive	1B/NC	**1B/NC**
Methyl 2-nonynoate	111-80-8	2.50	1B	1A	-1.66	1B/NC	**1A**
2-Methoxy-4-methylphenol	93-51-6	5.60	1B	1A	not reactive	1B/NC	**1B/NC**
n-hexane	110-54-3	NC	NC	NC	not reactive	1B/NC	1B/NC
pyridine	110-86-1	72.00	1B	1B	not reactive	1B/NC	1B/NC
kanamycin sulfate	25389-94-0	NC	NC		not reactive	1B/NC	1B/NC
beta,beta-3-trimethylbenzenepropanol	103694-68-4	NC	NC	NC	not reactive	1B/NC	1B/NC
Additional chemicals tested kDPRA (Wareing et al.)							
p-Benzoquinone	106-51-4	0.01	1A		0.55	1A	**1A**
Farnesal	502-76-0	11.70	1B		-2.60	1B/NC	**1B/NC**
2,3-Butanedione	431-03-8	11.30	1B		-2.62	1B/NC	**1B/NC**
2,4-Dinitrobenzenesulfonic acid	885-62-1	1.90	1A		-2.30	1B/NC	**1B/NC**
Tosylchloramide sodium	127-65-1	0.40	1A		0.06	1A	**1A**
Phthalic anhydride	85-44-9	0.16	1A		-0.07	1A	**1A**
Dinitrofluorobenzene	70-34-8	0.03	1A		1.02	1A	**1A**
Additional chemicals tested kDPRA (Natsch et al.; Roberts and Natsch)							
2,4-Dinitrobromobenzene	584-48-5	0.085	1A		-0.47	1A	**1A**
2,4-Dinitroiodobenzene	709-49-9	0.17	1A		-0.44	1A	**1A**
2,4-Dichloronitrobenzene	611-06-3	20	1B		-3.47	1B/NC	**1B/NC**
Benzyl bromide	100-39-0	0.2	1A		-0.43	1A	**1A**
Benzene, (2-bromoethyl)-	103-63-9	6.2	1B		-3.03	1B/NC	**1B/NC**
2,4,6-Trinitrobenzene sulfonic acid	2508-19-2	0.3	1A		0.53	1A	**1A**
Chlorothalonil	1897-45-6	0.004	1A		0.65	1A	**1A**
2,4,6-Trinitrochlorobenzene	88-88-0	0.05	1A		2.00	1A	**1A**
2,4-Dinitrothiocyanatobenzene	1594-56-5	0.047	1A		2.20	1A	**1A**
α-Damascone	24720-09-0	3.3	1B		-1.64	1B/NC	**1A**

Name	CAS	LLNA EC3 [%]	GHS Cat LLNA	GHS Cat Human	log k_{max} Numerical	Potency classification published cut-off	Potency Classification reviseed cut-off
(2E)-5,6,7-trimethyl-2,5-octa-dien-4-one	357650-26-1	1.6	1A		-1.09	1A	**1A**
4-Vinyl-pyridine	100-43-6	1.6	1A		-1.04	1A	**1A**
2-Decenal	3913-71-1	2.5	1B		-1.03	1A	**1A**
Galbanone	56973-85-4	3	1B		not reactive	1B/NC	**1B/NC**
Spirogalbanone	224031-70-3	2.2	1B		-2.28	1B/NC	**1B/NC**
Safranal	116-26-7	7.5	1B	1A	-1.74	1B/NC	**1A**
Damascenone	23696-85-7	1.22	1A	1A	-1.28	1B/NC	**1A**
Damascone beta	23726-91-2	4.01	1B		-2.35	1B/NC	**1B/NC**
4-vinylcyclohex-1-ene-carbal-dehyde	1049017-68-6	3.4	1B		-1.77	1B/NC	**1A**
Bergamal	22418-66-2	4.5	1B		-1.64	1B/NC	**1A**
γ-Damascone	35087-49-1	4.5	1B		-2.50	1B/NC	**1B/NC**
Additional chemicals in Urbisch et al database with 2 out of 3 data							
Streptomycin sulfate	3810-74-0	Ns	NC		-2.95	1B/NC	**1B/NC**
6-Methylcoumarin	92-48-8	>25	NC		not reactive	1B/NC	**1B/NC**
Clofibrate	637-07-0	NC	NC		not reactive	1B/NC	**1B/NC**
Dimethyl formamide	68-12-2	Ns	NC		not reactive	1B/NC	**1B/NC**
Dibutyl phthalate	84-74-2	Ns	NC		not reactive	1B/NC	**1B/NC**
Butylbenzylphthalate	85-68-7	Ns	NC		not reactive	1B/NC	**1B/NC**
2,2-Bis-[4-(2-hydroxy-3-me-thacryloxypropoxy)phenyl)]-propane (Bis-GMA)	1565-94-2	45.00	1B		not reactive	1B/NC	**1B/NC**
Cyclamen aldehyde	103-95-7	22.30	1B		-3.53	1B/NC	**1B/NC**
N,N-Dibutylaniline	613-29-6	19.60	1B		not reactive	1B/NC	**1B/NC**
4-Allylanisole	140-67-0	20.20	1B		not reactive	1B/NC	**1B/NC**
Oxalic acid anhydrous	144-62-7	15.00	1B		-1.01	1A	**1A**
Butyl acrylate	141-32-2	20.00	1B		-0.83	1A	**1A**
2-Ethylhexyl acrylate	103-11-7	19.18	1B		-2.13	1B/NC	**1B/NC**
Trimellitic anhydride	552-30-7	9.20	1B		-0.13	1A	**1A**
Methyl methanesulphonate	66-27-3	8.10	1B		-2.15	1B/NC	**1B/NC**
Undec-10-enal	112-45-8	6.80	1B		not reactive	1B/NC	**1B/NC**
4-Chloroaniline	106-47-8	6.50	1B		not reactive	1B/NC	**1B/NC**
2,4-Heptadienal	5910-85-0	4.00	1B		-1.52	1B/NC	**1A**
Diethyl sulfate	64-67-5	3.30	1B		-2.60	1B/NC	**1B/NC**
Methyl pyruvate	600-22-6	2.40	1B		not reactive	1B/NC	**1B/NC**
1-Bromohexane	111-25-1	10.00	1B		not reactive	1B/NC	**1B/NC**
1-Phenyl-1,2-propanedione	579-07-7	1.30	1A		-1.18	1B/NC	**1A**
CD3	25646-71-3	0.60	1A		-0.34	1A	**1A**
Maleic anhydride	108-31-6	0.16	1A		0.94	1A	**1A**
Palmitoyl Chloride	112-67-4	8.80	1B		-3.38	1B/NC	**1B/NC**
Chemicals which could not be evaluated due to fluorescence interference							
Tetrachlorsalicylanilide	1154-59-2	0.04	1A		-0.458, Autofluo.	1A	**1A**
2-Nitro-1,4-phenylenediamine	5307-14-2	3.95	1A		Quenching	1A	**1A**
Vanillin	121-33-5	>50	NC		Quenching	1A	**1A**
Tropolone	533-75-5	4.30	1B		Quenching	1A	**1A**
Bandrowski's Base	20048-27-5	0.02	1A		-1.86 (quenching)	1A	**1A**

Thioglycerol	96-27-5	3.6	1A		Reaction with mBrB		

Appendix 4: Statistical analysis for the contribution of continuous parameters from validated in vitro assays to predict LLNA EC3: kDPRA vs. existing validated in vitro tests (KeratinoSens and h-CLAT) as of October 11, 2019

Goal

The primary goal of the kDPRA is the identification of GHS 1A sensitizers. The second potential use is for the prediction of sensitizer potency on a more granular, continuous scale in an integrated approach (ITS) or defined approach (DA). The goal of this anaylsis is to evaluate the correlative contribution of the log k_{max} obtained from the kDPRA in single or multiple linear regressions along with other *in vitro* parameters from existing, validated tests when compared to LLNA EC3 values. This analysis also evaluates, whether a combination of multiple tests is superior to the kDPRA alone for GHS class 1A assignment, and whether the kDPRA can be proposed as stand-alone assay for 1A identification / subclassification.

This analysis is not made here in order to directly promote a multiple linear regression as a final defined approach (DA) – other mathematical ways such as neuronal nets, Bayesian nets or other non-linear models may be better for a final ITS or DA and this will be a separate discussion. Yet simple correlation between the *in vitro* data and the *in vivo* data may provide the most transparent indication on:

a) Predictive power of single parameters
b) Potential more predictive linear combinations
c) Data redundancy when adding additional parameters

In principle such correlations can be made for all data including sensitizers and non-sensitizers [1]. If this is done, the result is influenced by the potential of an assay to both predict hazard and by its potential to predict potency. In parallel, it may be more interesting to analyze only chemicals with an LLNA EC3 below a given threshold. Here we chose to focus specifically on chemicals with an **LLNA EC3 < 30%**: These are the chemicals with, in general, a *bona fide* positive LLNA result and sensitization potential as revealed by the LLNA. Chemicals with EC3 of 30 – 100 % are in a grey zone (most chemicals were only tested up to 25 or 50% when the LLNA was validated and many non-sensitizers were only tested up to such intermediary concentrations in original databases). Hence we cannot firmly conclude on the relevance of LLNA results above EC3 of 30%, else that they indicate a very weak/non-sensitizer status (some regulators may see that differently, but scientifically it does make little sense to try to model LLNA potency responses of an EC3 of 30 – 100% based on these limitations in the dataset and in the way the LLNA was validated).

A framework to predict potency therefore ideally focuses on chemicals with EC3 < 30%. Here we thus specifically look at the chemicals with LLNA EC3 < 30%, but we also make the same analysis on all chemicals including LLNA negatives and those in the grey zone.

We look here at two data-sets:

Set I Chemicals with log k_{max} from kDPRA **AND** KeratinoSens data (n=173; EC3<30% n=121)
Set II Chemicals with log k_{max} from kDPRA **AND** KeratinoSens data **AND** h-CLAT data (n=154; EC3<30% n=107)

When only evaluating predictivity of kDPRA and KeratinoSens, the most comprehensive Set I can be used. To evaluate predictivity in combination with h-CLAT and DPRA the smaller set II is used. In both cases we look at (i) all chemicals and (ii) those with LLNA < 30%.

Data source and approach

Data on **kDPRA** are from the testing of the extended dataset in the kDPRA as summarized in Appendix 3 of the kDPRA validation report, i.e. the data used to set a refined cut-off to identify GHS 1A sensitizers as described in the validation report. KeratinoSens, h-CLAT and DPRA data were compiled from the three databases compiled by Urbisch et al.[2], Hoffmann et al.[3] and Jaworska et al.[4].

LLNA data were sourced from the same references[2]. For numerical analysis, EC3 of non-sensitizers (i.e. SI of 3 not reached at maximum tested concentrations) were set to 100, although some may have only been tested up to 20%. LLNA EC3 values were transformed to pEC3 values:

$$pEC3 = Log(\frac{Mol.weight}{EC3})$$

The numerical data from **KeratinoSens** (EC1.5 and EC3 for luciferase induction, IC50 for 50% cytotoxicity, all in µM) were transformed to logarithmic values. If no induction was reached and in case of no toxicity up to maximal test concentration of 2000 µM observed, a default value of 4000 µM was used.

The numerical data from **h-CLAT** (EC150 for CD86, EC200 CD54, MIT, minimum induction threshold of EC150 and EC200, and CV75 for 25% cytotoxicty, all in µg/mL) were transformed to logarithmic values. If no induction was reached and in case of no toxicity up to maximal test concentration of 5000 µg/mL observed, a default value of 5000 mg/mL was used for numerical analysis.

For the DPRA, the depletion values are transformed to rate constants according the equation:

$$\log k = \log \frac{\ln \frac{100}{100 - dp}}{[E]_0 t}$$

where dp is the peptide depletion, [E] the test substance concentration in mM and t the incubation time (1440 min). These logarithmic data were then used for linear regression analysis in the MiniTab software (Minitab® 18.1; © 2017 Minitab).

[2] Note: For this analysis the reference data from the historical databases were used, not adjusting for the corrections introduced by the OECD 2019 review in the group on DA SS guideline, as these may be subject to further discussions and as all the validations done here for kDPRA relied on these historical databases also used for validation of other methods.

A) Contribution of individual parameters

First, single linear regression of each *in vitro* parameter vs. the pEC3 form the LLNA was calculated. As shown in Table A12, the strongest correlation is observed for the full **dataset Set I** for the k_{max} from the kDPRA (r^2 = 0.51). On the same set, the parameters from KeratinoSens have a range of r^2 = 0.29 - 0.35. Limiting the evaluation to the clear positives in the LLNA (EC3 < 30%), the correlation is reduced: r^2 = 0.40 for the kDPRA and r^2 = 0.13 - 0.17 for KeratinoSens.

For the smaller **dataset Set II** also including h-CLAT data, the correlation to potency in general is weaker: r^2 = 0.45 and 0.32 for k_{max} compared to all LLNA values or the chemicals with EC3 < 30%, respectively. For the parameters from the cellular assays, a range of r^2 = 0.16 - 0.43 is observed for all chemicals and 0.11 – 0.20 for chemical with EC3< 30%. Based on this analysis, k_{max} as a single parameter has the strongest correlation to potency among all parameters investigated, and this is in particular true when focusing on chemicals with EC3 < 30%, i.e. those where the potential for hazard ID of the *in vitro* tests is no longer affecting the overall predictivity.

All the correlations shown in Table A12 are statistically highly significant at $p \leq 0.0005$ (with the exception of EC200 / Set II /EC<30%, where p = 0.039). Table A13 additionally lists all the F values for these statistical comparisons.

It has to be emphasized, that these are very 'crude' comparisons, as they integrate chemicals form all chemical domains, and they focus on predictivity of the LLNA with all its limitations of variability and regarding its potential to predict human sensitization potential. Also, *in vitro* assays may have a stronger correlation to potency when used in specific mechanistic domains [1], and these benefits are not taken into account by these overall statistical comparisons.

Table A12: R^2 coefficient for linear regression of logarithmic *in vitro* parameters vs. pEC3

		Set I: Wwth KeratinoSens (n = 173)	Set II: with KeratinoSens and h-CLAT and DPRA (n = 154)	Set I: with KeratinoSens, EC3 <30% (n = 121)	Set II: with KeratinoSens and h-CLAT and DPRA, EC3 <30% (n = 107)
kDPRA	k_{max}	0.51	0.45	0.40	0.32
Keratino-Sens	EC1.5	0.29	0.27	0.13	0.11
	EC3	0.35	0.35	0.17	0.16
	IC50	0.34	0.34	0.14	0.14
h-CLAT	EC150		0.28		0.17
	EC200		0.16		0.04
	MIT[1)]		0.36		0.20
	CV75		0.43		0.21
DPRA	kCys		0.33		0.19
	kLys		0.16		0.17

[1)] The MIT is not a directly measured parameter, but already a mathematical combination of two parameters, i.e. the minimal value of EC150 and EC200; as this is widely used in publications on h-CLAT it is provided here, too.

Table A13: F-Values for linear regression of logarithmic *in vitro* parameters vs. pEC3

		Set I: With KeratinoSens	Set II: with KeratinoSens and h-CLAT and DPRA	Set I: With KeratinoSens, EC3 <30%	Set II: with KeratinoSens and h-CLAT and DPRA, EC3 <30%
kDPRA	k_{max}	191.14	126.05	84.34	50.55
Keratino-Sens	EC1.5	77.84	57.41	18.58	13.28
	EC3	95.91	80.24	25.78	20.47
	IC50	95.48	78.72	20.45	17.24
h-CLAT	EC150		59.00		20.99
	EC200		29.71		4.35
	MIT		85.23		27.04
	CV75		115.24		28.66
DPRA	kCys		71.9		23.57
	kLys		28.12		20.87

B) Chemicals with LLNA EC3 < 30%: Continuous models with multiple regressions

Next, we performed multiple regressions of several *in vitro* parameters vs. the pEC3 form the LLNA. This allows estimating, whether a model incorporating several parameters will significantly improve potency prediction and it gives a first indication of data redundancy. Results of the key linear combinations are shown in Table 3.

For the largest **Set I**, adding KeratinoSens parameters (EC3 and IC50) to k_{max} improves r^2 from 0.51 to 0.61 for all chemicals and from 0.40 to 0.45 for those with EC3 < 30%.

On the **Set II**, this same improvement is from 0.45 to 0.57 for all chemicals and from 0.32 to 0.38 for those with EC3<30%. Similarly, when adding h-CLAT (MIT and CV75) to k_{max} , the r^2 is improved from 0.45 to 0.59 for all chemicals and from 0.32 to 0.40 for those with EC3 < 30%. Thus adding one cellular assay improves predictivity, but especially if integrating all chemicals and less so when only those with EC3 < 30% are considered. If both cell-based assays are added to k_{max}, the further increase in correlation is only very marginal (row four in the Table) as compared to using only one cellular assay, which indicates a strong data redundancy as had already been observed before [1].

When combining all *in vitro* parameters from DPRA, h-CLAT and KeratinoSens, the quantitative correlation to potency for the chemicals with EC3 < 30% is weaker ($r^2 = 0.27$) as compared to k_{max} alone ($r^2 = 0.32$). This is a further indication that k_{max} is the strongest contributing parameter form potency of the validated *in vitro* assays, which is in line with the view that protein modification as molecular initiating event is an important rate-limiting step for the acquisition of skin sensitization.

Table A14: R^2 coefficient for linear multiple regression of logarithmic *in vitro* parameters vs pEC3

	Set I: With KeratinoSens (n = 173)	Set II: with KeratinoSens and h-CLAT and DPRA (n = 154)	Set I: With KeratinoSens, EC3 <30% (n = 121)	Set II: with KeratinoSens and h-CLAT and DPRA, EC3 <30% (n = 107)
k_{max}	0.51	0.45	0.40	0.32
KS+k_{max}	0.61	0.57	0.45	0.38
h-CLAT+k_{max}		0.59		0.40
h-CLAT +KS+k_{max}		0.60		0.41
h-CLAT +KS+DPRA		0.54		0.27
h-CLAT +KS		0.51		0.27

KS, KeratinoSens

While the quantitative models integrating multiple parameters will be interesting for quantitative risk assessment and especially also when evaluating specific chemical domains, in the short term a crucial question is whether a model with multiple parameters could better predict and separate GHS 1A chemicals form GHS 1B/NC chemicals as compared to kDPRA alone. We thus used the models in Table A14 (i.e. the regression equations underlying the reported r^2 values) and calculated for each chemical the predicted LLNA EC3. Chemicals with a predicted EC3 < 2% where then assigned a GHS 1A subcategory, which was compared to the *in vivo* value.

In Table A15 we show the results, here only for the chemicals with LLNA EC3 < 30%. The same analysis can be done with all chemicals, in that case specificity and balanced accuracy will be higher as all the chemicals with weak/no response in LLNA and *in vitro* tests are included boosting specificity, while sensitivity, by definition, will remain the same.

In the first analysis we show the predictivity of k_{max} when applying the simple prediction model with a threshold of log k_{max} > -2 for GHS class 1A attribution ("k_{max} threshold -2"). For comparison with the other models, which are based on regression, we also calculated a regression model with k_{max} as single predictor and used this for GHS class 1A attribution[3].

As shown in Table A15, sensitivity for identifying GHS class 1A sensitizers is at 86.4% with the threshold prediction model and slightly lower (84.1%) with the regression model using k_{max} alone. Adding to the regression model with the single parameter further parameters from KeratinoSens and /or h-CLAT, sensitivity and balanced accuracy are not enhanced. A model with h-CLAT, KeratinoSens and DPRA can reach higher sensitivity but not surprisingly at significant costs regarding specificity.

This analysis thus further indicates, that, overall, a decision model for GHS class 1A attribution based on log k_{max} alone has a high predictivity and can be used as stand-alone model for the identification of GHS 1A sensitizers.

---

[3] Note: If the average molecular weight of MW=185 for the chemicals with an EC3 < 30% and available k_{max} values is used and entered along an EC3 = 2% into the regression equation with k_{max} as single predictor (pEC3 = 2.652 + 0.3491 × Log k_{max}), which is then solved for log k_{max}, the resulting log k_{max} =-1.96 is obtained as decision threshold to predict chemicals with EC3 ≤2%, indicating that with regression analysis we obtain a very similar result as with the ROC analysis.

Table A15: Predictivity of linear regression models for sub-classification of chemicals with EC3 <30% into GHS 1A and 1B chemicals

	n	Sensitivity [%] [1)]	Specificity [%]	Balanced accuracy [%]
Set I				
k_{max} threshold -2	125	86.4	77.8	82.0
k_{max} regression model	125	84.1	75.3	79.7
KS+k_{max}	125	84.1	71.6	77.8
SetII				
k_{max} threshold -2	106	83.3	77.1	80.2
KS+k_{max}	106	80.5	75.7	78.1
h-CLAT +k_{max}	106	83.3	77.1	80.2
h-CLAT +KS+k_{max}	106	77.8	80.0	78.9
h-CLAT +KS	106	66.7	78.6	72.6
h-CLAT +KS+DPRA	106	91.6	62.9	77.3

KS, KeratinoSens

[1)] Here prediction statistics are given for identifying GHS 1A sensitizers within the chemicals with an LLNA EC3 < 30%.

[2)] For each chemical, the predicted EC3 was calculated with the regression models in Table 3, column 4 and 5, and chemicals with predicted EC3 < 2% were assigned to GHS 1A.

Discussion

The analysis presented here shows that all the continuous parameters from the validated *in vitro* assays have a statistically highly significant correlation to sensitizer potency as expressed by the EC3 or pEC3 values form the LLNA.

This quantitative correlation is strongest for the parameter log k_{max} measured in the kDPRA. This is alignment with an earlier analysis, in which LLNA potency was correlated to multiple parameters from *in vitro* tests and in which k_{max} values measured with a different peptide (Cor1-C420) and with a protocol not yet undergone full standardization as for the kDPRA was identified as the strongest predictor, too [1].

The superior performance of log k_{max} for potency assessment is particularly clear when focusing on those chemicals with a clear positive call in the LLNA (EC3 < 30%), i.e. excluding those which are negative or in the grey zone with EC3 values between 30 and 100% (the zone where very little validation against human and guinea pig data had been done for the LLNA).

From a mechanistic point of view, this result is in alignment with the view, that the reaction with a protein is the rate limiting step for acquisition of skin sensitization and with the view that a rate constant is the best single predictor of chemical reactivity [5]. Of course other factors such as (i) the nature of the target nucleophile (ii) ADME factors and (iii) danger signal / activation of cellular pathways will have an influence on potency, and thus an *in vitro* rate constant can be only an approximation, but maybe the best single predictor to be measured *in vitro* based on this analysis.

The analysis of the improvement of models adding other predictors in addition to the k_{max} indicates that, although these predictors have a highly significant correlation to potency themselves, they do improve correlation for chemicals with EC3 < 30% to a limited extend and do not improve identification of GHS 1A sensitizers, which is a strong argument for data redundancy. Of course, the hope behind the concept of using multiple tests addressing different key events in the adverse outcome pathway was always that these tests and events are orthogonal and thus would lead to a cumulatively improved predictive capacity. While this is partly true, especially when including all chemicals, data redundancy limits the benefit of combinations of assays especially for the chemicals with EC3 < 30%.

From a ***statistical point***, the current analysis thus indicates that:

- For identification of GHS class 1A sensitzers among chemicals rated positive by hazard models, **proposing the kDPRA as a stand-alone assay is justified** based on analysis of a large dataset, for which evidence from multiple *in vitro* assay is available
- The balanced accuracy and especially **the sensitivity are sufficient**, esp. in the light of the variability of the LLNA itself to predict 1A sensitizers. Thus sensitivity > 90% cannot be expected for any test or defined approach, if the figure of LLNA reproducibility for 1A identification given in [6] is correct ($p = 0.9^3 = 0.73$ for three LLNA EC3 < 2% results for a bona fide 1A)
- **Combinations of assays would not further improve the model to identify 1A sensitizers** in general, statistical terms.

From a ***mechanistic point*** there are four clear limitations to this proposal:

a) **Pro-haptens** requiring metabolic activation will not always be correctly classified. However, the number of known GHS 1A prohaptens is rather limited. This may be mechanistically explained by the fact that metabolism may often be a rate limiting step in the skin and thus the requirement for metabolic activation will reduce the potential of a chemical to act as strong skin sensitizer. Among the false-negative GHS 1A sensitizers according to the LLNA, we identified not a single pro-hapten, while against human data two false-negative (i.e. 1B/NC instead of 1A) chemicals were identified: 3-dimethylaminopropylamine and diethylenetriamine, although the assignment as 1A for the latter chemical was now removed from the OECD reference list due to insufficient human data.

b) **Pre-haptens** requiring rapid oxidation (< days) for hapten formation may be underpredicted. This was the case for 4-phenylenediamine and 2-aminophenol, which reacts more slowly in the kinetic assay as the observed kinetic rate is the sum of the oxidation rate and the reaction rate.

c) **Chemicals reacting predominantly with other nucleophiles.** The ability to covalently modify lysine residues has been proposed as a contributor to particularly high sensitization potential esp. for some extreme sensitizers [7]. This potential is not covered by the kDPRA. However, among the kDPRA false negatives (compared to the LLNA), only glutaraldehyde falls into this class. The low number of such false-negatives can be explained by the fact that most chemicals with high Lys-reactivity also are highly reactive towards Cys. Indeed, in the published databases on the DPRA, anhydrides are the key group with selective Lys reactivity, however they are positive in the kDPRA as they form rapid, semistable adducts with Cys at early timepoints. Besides anhydrides and glutaraldehyde, the published databases report only hydrocortisone and bromo-hexane as LLNA positives with selective Lys-reactivity (> 50% Lys depletion, < 50% Cys depletion). These chemicals are weak sensitizers, and the strong Lys reactivity remains to be verified in an independent assay.

d) **Metals** form coordination bonds with proteins and cannot be detected by peptide depletion with Cys- or Lys-model peptides.

This discussion indicates, that ***by mechanistic reasons***, there can be chemicals which would not be recognized by the kDPRA, but it shows that they are a very small group which explains the ***statistically satisfying performance*** of the kDPRA.

References

1. Natsch, A., et al., Predicting Skin Sensitizer Potency Based on In Vitro Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment. Toxicol. Sci., 2015. 143(2): p. 319-32.

2. Urbisch, D., et al., Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul. Toxicol. Pharmacol., 2015. 71(2): p. 337-51.

3. Hoffmann, S., et al., Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. Crit. Rev. Toxicol., 2018. 48(5): p. 344-358.

4. Jaworska, J.S., et al., Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Arch. Toxicol., 2015. 89(12): p. 2355-83.

5. Roberts, D.W. and A.O. Aptula, Determinants of skin sensitisation potential. J. Appl. Toxicol., 2008. 28(3): p. 377-387.

6. OECD, Supporting document for evaluation and review of draft Guideline (GL) for Defined Approaches (DAs) for Skin Sensitisationhttps://www.oecd.org/env/ehs/testing/DAGL%20supporting%20document_23%20Sep2019.pdf OECD Publishing, Paris, 2019. https://www.oecd.org/env/ehs/testing/DAGL%20supporting%20document_23%20Sep2019.pdf.

7. Natsch, A. and R. Emter, Reaction Chemistry to Characterize the Molecular Initiating Event in Skin Sensitization: A Journey to Be Continued. Chem. Res. Toxicol., 2017. 30(1): p. 315-331.

Appendix 5: Follow-up to the public consultation on the validation report (circulated for comments on 7 July 2020).

Following completion of the validation study, conducted between spring 2018 and summer 2019, the kDPRA test method and its validation study underwent an independent peer-review between August 2019 and February 2020 by an international panel (1).

A public consultation of the draft validation report and the draft Test Guideline 442C was organised in July 2020. The comments received on the validation report and responses are provided in this Appendix. This option was chosen in order to keep the validation report as it was when submitted to peer review, while not losing the information gained from the commenting round.

This Appendix has been developed and added to the validation report after completion of the peer review.

Member country	Comments	Responses
	General	
US	A general question is if values close to the cutoff should be assigned a category or listed as not classifiable.	Guidance is provided in paragraph 29 of the kDPRA Appendix.
	The authors calculate standard deviation values in multiple places, but it is unclear how they are calculated. Was the propagation of errors approach used since multiple components in the log k calculation have variability? This needs to be more clearly described.	No – no error propagation is done. From the triplicate depletion values the mean is taken, then the slope is derived from this mean, which then leads to the reported individual k_{max} value of a given experiment. For intra and inter-lab comparison, there is then the standard deviation within or between Lab calculated for these Log k_{max} values
	Section 3.3 – 3rd Paragraph	

Member country	Comments	Responses
US	Line 3: Should the statistical analysis be one-sided instead of two-side? The main focus is to find compounds that have greater depletion than the negative control.	Good point – the standard evaluation sheet does calculate a two-tailed t-test. Actually we would be hesitant to change that as it was used as such in the validation study, although there is a very good argument to do it one sided as you correctly note!
	Figure 1 (and other similar figures)	
US	Where are the error bars?	These are always calculated in the standard evaluation template, but have not been added to these graphs.
	Section 5.5.2 – 2nd paragraph	
US	While standard deviation values are higher for less reactive chemicals, it seems that the coefficient of variation values are similar.	A coefficient of variation does not make a lot of sense for logarithmic averages with their associated logarithmic standard deviations (which are actually factors of variation). (i.e. If actual not log transformed values would be normalized by e.g. a factor of 1000, the same would happen for the SD, and the CV would remain the same, while in the log scale the values would move by 3, the SD remain the same and the CV be completely different…). So we would not refer here to CV.

Member country	Comments	Responses
	Tables 7, 8	
US	Should “SDSD” be “SD”?	YES, typo
	Section 5.2.4, page 21	
US	If there is significant fluorescence or quenching, it may be helpful to test a larger number of interference controls to have a more precise quantification of this effect. Testing only one well (as described in the papers published on this assay) may lead to large errors.	See comment above and revised ANNEX to the SOP which gives more detail to deal with the (relatively few) cases of fluorescence interference.
	Page 37, first paragraph	
US	Line 2: It is unclear why COV is meaningless if the mean and standard deviation are meaningful. It would still provide information about the relative size of the standard deviation relative to the mean.	See above, logarithmic standard deviations are factors in the normal scale and are thus scale-independent. COV calculated as SD in Log scale divided by AVG in Log scale mathematically does not make sense.
	Figure 12	
US	Why does the line not cross the origin? That is unexpected. Does the 95 % confidence interval for the slope overlap with 1?	Actually in this graph only the few chemicals from the ring study are plotted, which are biased towards sensitizers and significant reactivity, thus the low reactive chemicals have low leverage, and thus the regression line at x = 0 has low statistical power.

Member country	Comments	Responses
		When analyzing the 158 chemicals from the database with DPRA data, the regression equation of kDPRA 24 h/5mM to Cys depletion DPRA is: Cys-Dep = 5.4% + 0.95 x kDPRA24h/5mM depletion. Thus with a y-intercept of 5.4% it is much closer to zero as you would expect.
	Paragraph 5.4.18	
US	3rd bullet point: Is a factor of 1.75 actually that small? It is nearly a factor of 2.	More precision would always be optimal, however, Hoffmann, 2015[4], found an average logarithmic SD of 0.252 for repeated LLNA testing, which is very close to our value of 0.244 for inter-laboratory testing of the rate constant k_{max}. Thus variability is similar to the reference method.

(1) OECD (2021), Independent Peer-Review Panel Report on the scientific validity of the kinetic Direct Peptide Reactivity Assay (kDPRA). Series on testing and Assessment n°338. Organisation for Economic Co-operation and Development, Paris.

[4] Hoffmann, S., LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. ALTEX, 2015. 32(4): p. 379-83.

Appendix 6: The predictivity of the kDPRA to identify 1A sensitizers vs. the consolidated OECD database

Author and Date: Andreas Natsch and Susanne Kolle, 2.10.2020

This Appendix has been developed and added to the validation report after completion of the peer review (1).

The database on 180 chemicals published in ALTEX and included in the validation report of the kDPRA is significantly, but not completely overlapping with the DASS reference database. In order to fully assess predictivity of the kDPRA also on the DASS database, a data gap filling exercise was conducted and all missing chemicals were tested in the kDPRA.

This analysis was conducted based on the DASS database, status June 21, 2020. Table 1 indicates how many chemicals in the different classes are in the two databases, to which extend the databases overlap, and to which extend the same chemicals are rated differently in the different databases.

For all the analysis, the same approach as in the validation report was used: Chemicals are classified into two classes: GHS 1 A and not – GHS 1A (i.e. GHS 1B/NC), as the goal of the evaluation is to rate the predictivity for identifying GHS 1A chemicals.

As shown in Table 1, the LLNA data review only led to different class attribution for 3 chemicals as compared to the ALTEX publication, while for five 1A chemicals, the OECD DASS group came to the conclusion that they cannot be rated as 1A but rather as inconclusive based on the LLNA result (mostly because the EC3 value was extrapolated from a higher test concentration). Overall the ALTEX database contained 6 more 1A chemicals, but in general a similar number of chemicals is evaluated in both databases and the two assessments thus have similar statistical power.

Table 1.Overview of the database used in the validation report and in ALTEX vs. the OECD database

	Validation / ALTEX Database	OECD Database
Number of LLNA 1A	45	39
Number of LLNA 1B/NC	135	116
Number of LLNA 1A unique to the database	7	6
Number of LLNA 1A in both databases	33	33
Number of LLNA 1A rated 1B/NC in the other database	1	2
Number of LLNA 1A rated NA (1B/1A attribution cannot be made) in the other database	5	0

Table 2 lists the results for the two databases vs. LLNA results. Sensitivity is 81.6% and specificity is 87.9%. A very similar balanced accuracy is observed as in the previous analysis, with 2.8% lower sensitivity and 2% higher specificity for the DASS database compared to the ALTEX / validation database.

Thus the conclusion from the validation report and supported by the validation peer review, that the kDPRA has an adequate predictivity for identifying GHS 1A sensitizers can be confirmed when analyzing the curated LLNA data in the DASS database. The conclusion is also supported when comparing to ITSv2.6 as a benchmark, which is currently being discussed as a DA for identifying 1A chemicals based on *in vitro/ in silico* data to be included in the DA guideline: ITSv2.6 has 68.4% sensitivity for 1 A sensitizers with a specificity of 90.4% for the LLNA assessment in the curated DASS database, using the same two class analysis as made here, this sensitivity of the kDPRA is clearly superior.

Table 2. Comparison of the predictivity for LLNA data

	Sensitivity [%]	Specificity [%]	Balanced accuracy [%]	n tested
kDPRA vs LLNA OECD DASS DB	81.6% (31/38)[1)]	87.9% (102/116)	84.8%	154 [2)]
kDPRA vs LLNA ALTEX DB	84.4% (38/45)[1)]	85.9% (116/135)	85.2%	180

[1)] given are predictivity data in percent and in brackets n correct predictions / n chemicals in the group

[2)] 12 of 167 chemicals in OECD database are NA, i.e. cannot be attributed to class 1A or 1B

Overall there are seven LLNA 1A chemicals underpredicted as 1B, as shown in Table 3, five of those are common to the ALTEX publication, one could now be tested by using an alternative fluorescent probe thus avoiding autofluorescence of Tetrachlorosalicylanilide. Only one chemical (TCSA) has a cLogP of > 3.5, thus the concern that chemicals in this physicochemical range would be specifically mispredicted due to limited solubility appears not to be relevant when looking at this specific dataset, although the 1A chemicals in general contain very few chemicals of high cLogP.

Table 3. Chemicals underpredicted by the kDPRA

Chemical	**CASRN**	**LogP**	**Discussion**
Glutaraldehyde (act. 50%)	111-30-8	-0.14	Amine reactive, see ALTEX paper
1,4-Phenylenediamine	106-50-3	-0.30	Pre-hapten, see ALTEX paper
2-aminophenol	95-55-6	0.62	Pre-hapten, see ALTEX paper
Tetrachlorosalicylanilide	1154-59-2	5.11	Autofluorescence, could now be tested with alternative fluorescent probe
Bisphenol A-diglycidyl ether	1675-54-3	3.36	See ALTEX paper, Close to threshold (LLNA EC3 1.5%)
1-Naphthol	90-15-3	2.85	Close to threshold (LLNA EC3 1.3%)
2,4-Dinitrobenzenesulfonic acid, sodium salt	885-62-1	-2.40	See ALTEX paper, Close to / at threshold (LLNA EC3 2%)

Comparison to the human data in the DASS database

The DASS Database contains only 17 chemicals rated as human 1A chemicals. The sensitivity here is clearly lower, with 9 of 17 chemicals rated 1A (sensitivity 53%) by the kDPRA. However, also the LLNA has limited predictivity for this set of human data and the attribution made by the human data review, with 9 of 17 chemicals rated 1A by the LLNA and 3 rated NA, thus while the predictivity of the kDPRA for this dataset is not satisfying, the same holds true for the current reference method, the LLNA.

An Excel document containing all data is available from the OECD Secretariat upon request (nathalie.delrue@oecd.org)

(1) OECD (2021), Independent Peer-Review Panel Report on the scientific validity of the kinetic Direct Peptide Reactivity Assay (kDPRA). Series on testing and Assessment n°338. Organisation for Economic Co-operation and Development, Paris.