厚生労働行政推進調査事業費補助金（化学物質リスク研究事業）
OECDプロジェクトでの成果物を厚生労働行政に反映させるための研究

令和4年度 分担研究報告書

OECD における AOP，TG 及び GD の開発

研究分担者　小島肇
国立医薬品食品衛生研究所　食品添加物部

研究要旨

本研究は、化学物質やその混合物の安全性を評価するための国際的な合意を推進する経済協力開発機構（OECD: Organisation for Economic Co-operation and Development）の試験法ガイドライン（TG: Test Guideline）プログラム各国調整官作業グループ（WNT: Working Group of National Co-ordinators of the TGs programme）において、日本で開発された種々の TG やガイダンス文書（GD: Guidance Document）、有害性発現経路（AOP : Adverse Outcome Pathway）などの世界各国が必要とする成果物を公定化させることを目的とする。

これまでの先行研究の成果として、我が国で開発された腐食性試験代替法、皮膚感作性試験代替法、光毒性試験代替法、内分泌かく乱性スクリーニング法などに関する TG や免疫毒性の AOP の公定化に寄与し、皮膚感作性試験の確定方式（DASS: Defined Approach for Skin Sensitisation）ガイドライン 497 の開発に関与してきた。

本年度は、これらの成果を生かし、TG に関しては、既存の TG である皮膚感作性試験代替法 ADRA（Amino acid Derivative Reactivity Assay）を含む TG442C の再改定をなすことができた。GD として、*In vitro* 免疫毒性試験の総説（DRP: Detailed Review Paper）が OECD に採択されたが、*in vitro* 生殖毒性試験の総説は論文投稿に留まった。

研究協力者

相場節也
　元東北大学医学部大学院大学
笠原利彦
　富士フィルム株式会社
小島幸一
　一般財団法人食品薬品安全センター
　秦野研究所

A. 研究目的

本研究は、化学物質やその混合物の安全性を評価するための国際的な合意を推進する経済協力開発機構（OECD: Organisation for Economic Co-operation and Development）の試験法ガイドライン（TG : Test Guideline）プログラム各国調整官作業グループ（WNT: Working Group of National Co-ordinators of the TGs programme）において、1）日本で開発された種々の TG やガイダンス文書（GD : Guidance Document）、有害性発現経路（AOP : Adverse Outcome

Pathway）や評価のための戦略的統合方式（IATA: Integrated Approaches to Testing and Assessment）などの世界各国が必要とする成果物を公定化させることを目的とする。

B. 研究方法

B-1. AOP の開発

EAGMST（Extended Advisory Group on Molecular Screening and Toxicogenomics）で行われているOECDのAOP開発プロジェクトの進捗に合わせ、班員を支援した。

B-2. TG、DRP 及びIATA の開発

OECDのTGの開発プロジェクトWNTの進捗に合わせ、班員を支援した。

B-2-1. 皮膚感作性試験

研究協力者の笠原とともに、皮膚感作性試験代替法 *In Chemico* Skin Sensitisation、ADRA（Amino acid Derivative Reactivity Assay）に混合物が評価できる重量法を加えた TG442C の再改定に向け、尽力した。また、DPRA（Directive Peptide Reactivity Assay）の重量法も同 TG に追加するため、研究協力者の笠原及び小島に加え、他国の機関（P&G 及び Givaudan）とともに共同研究を主導した。具体的には、コード化した 10 物質を 4 施設に配布し、合計 20 物質（分子量が大きく、バラツキが生じる可能性の高い感作性物質）を用い、重量法の施設間再現性を確認するとともに、既知法であるモル濃度法との比較研究を実施した。

また、研究協力者の相場とともに、IL-8 Luc assay TG442E の改定案を作成し、各国からの改定要望に対処した。

さらに、研究分担者の足利とともに、皮膚感作性試験の確定方式（Defined Approach for Skin Sensitisation：DASS）ガイドライン 497 の改定プロジェクトに参加し、他国の専門家と議論した。

B-2-2. 免疫毒性試験

相場及び国際的な専門家とともに、*in vitro* 免疫毒性に関する DRP (Detailed Review Paper)を作成した。

DRP の承認を待って IL-2 を指標とした免疫毒性試験 IL-2 Luc assay の TG 案を提出し、各国からの改定要望に対処した。

B-2-3. 生殖毒性試験の DRP

本分野の国内外の専門家とともに、*in vitro* 生殖毒性毒性に関する総説を作成し、Current Research Toxicology に投稿した。

B-3. 光毒性 IATA 開発

研究分担者の尾上とともに OECD expert group からのコメントに従って光毒性 IATA 案を修正した。

B-4. OECD に提出する資料の事前確認と OECD からの意見募集への対応

B-4-1. SPSF

昨秋、日本から提出した SPSF（Standard Project Submission Form）の内容を検討した。

B-4-2. Emerging technologies in the Test Guidelines Programme に関するワークショップ

代表研究者の平林とともに、emerging technologies in the Test Guidelines Programme のワークショップに関与し、今後の TG の在り方について議論した。

（倫理面への配慮）

特になし。

C. 研究結果

C-1. AOP の開発

免疫毒性、発がん性および光毒性のAOP成立に向け、EAGMSTのメンバー及び開発者の代理として、OECD事務局と交渉した。「IL-1 receptor結合阻害：AOP277」の採択に向け、足利研究分担者を支援した。

C-2. TG 及びDRP の開発

C-2-1. 皮膚感作性試験

一昨年度から検討を続けてきた *in Chemico* Skin Sensitisation、 ADRA の中に混合物が評価できる重量法を加える TG442C の再改定案を作成し、WNT で議論された結果、重量法を加えた TG442C の再改定が 2022 年 9 月に公表された（添付資料１）。引き続き、OECD から要請を受け、TG442C に追加する DPRA 重量法に関する共同研究を主導した。その結果、表 1 に示すように、20 物質すべてでモル濃度法と重量法が一致した結果となることを確認した。

また、IL-8 Luc assay TG442E の改定案を作成し、7 月に OECD に提出し、改定に関する議論を各国の専門家と行った。

一昨年 TG497 として公表された DASS の改定に引き続き参画し、日本の方法である ADRA と IL-8 Luc assay をガイドライン 497 に加えるべく、協力した。

C-2-2. 免疫毒性試験

In vitro 免疫毒性試験の DRP の採択に向けて尽力した結果、本年 9 月に公表された(添付資料 2)。また、IL-2 を指標とした免疫毒性試験 IL-2 Luc assay の TG 案を OECD に提出し、WNT 意見募集を受けて改定した。

C-2-3. 生殖毒性試験の DRP

国際的な専門家とともに、*in vitro* 生殖毒性毒性に関する DRP の作成を継続して実施してきた。成果の一つとして論文が、本年 5 月に Current Research Toxicology に受理された。ただし、この論文をもとに OECD で DRP を作成することは断念した。

C-3. 光毒性 IATA の開発

作成した AOP を基盤として、その枠組みのなかで information sources をマッピングした。光毒性に関与する elements としては、(i) Exposure consideration、 (ii) Chemical descriptors 、 (iii) Skin penetration 、 (iv) Photoexcitation、 (v) Oxidative stress、 (vi) Cell injury を定義し、それらに関わる information sources をリスト化した。この中には *in vitro* 試験、*in vivo* 試験のみならず、*in silico* や QSAR モデルも含めた。また、OECD 専門家会議における有識者の助言に従い、information source に対して詳細な記述を加えることとし、具体的には (1) Regulatory use、 (2) Validation & regulatory acceptance status 、 (3) Potential role in the IATA、 (4) Description、 (5) Scientific basis including MoA、 (6) Protocol available、 (7) Strengths and weakness 、 (8) Applicability domain and limitations、 (9) Predictive capacity、(10) Reliability を各種文献情報やガイドラインを交えつつ追記した。また、これらの information sources を組み合わせた包括的光安全性評価に関して decision tree を新たに提案し、IATA にあくまでも一例として記述した。

アップデートした IATA 案はすでに 各国からの意見募集期間に入ったので今後はコメントや指摘事項に対応して修正作業を行った。

C-4. OECD に提出する資料の事前確認と OECD からの意見募集への対応

C-4-1. SPSF

昨年 11 月に日本から以下の SPSF を提出した。提出にあたり、厚生労働省とも内容を調整した。

1) Proposal for α-Sens® as FBS-free test system for detecting Key Event 2 (ARE-Nrf2 activation) of skin sensitization
2) Proposal for TG 493 ((Performance-Based Test Guideline for Human Recombinant Estrogen Receptor (hrER) *In Vitro* Assays to Detect Chemicals with ER Binding Affinity) performance and acceptability criteria to make it realistic

C-4-2. Emerging technologies in the Test Guidelines Programme に関するワークショップ

Lesson and Learned for a Validation Studyという演題で8月31日に講演した。発表内容を添付資料3に示した。この演題を含め20以上の発表会が数か月に渡り事前に開催された。

D. 考察

免疫毒性や生殖毒性試験などの全身毒性に関する *in vitro* TG の開発は前例がなく、これまで以上に時間を要しており、費用も嵩んでいる。OECD は、こうした前例のない TG を開発するために、まずは DRP の作成を求めており、数年掛かりで免疫毒性と生殖毒性試験の GD 作成を進めてきた。本年、in vitro 免疫毒性試験の DRP を開発できたものの、生殖毒性試験の DRP 開発を断念した。その理由として、DRP の開発は *in vitro* 生殖毒性試験 Hand1-Luc EST の TG 開発を目指したものであったが、開発者の住友化学株式会社がこれ以上の開発を望まないと表明したことによる。新規試験法の導入は、行政的には慎重であるべきとは思うが、時間的なロスを解消しない限り、Emerging technologies の導入は難しいと予想している。

一方、OECD で今年から始まった Emerging technologies in the Test Guidelines Programme への対応は、まさしく日本が直面している問題を解決するプロジェクトである。研究代表者の平林および厚生労働省の担当者とも連携を図り、引き続き、日本として適切な対応を心掛けていく。

E. 結論

TG に関しては、日本主導で取り組んできた *in Chemico* Skin Sensitisation、 ADRA TG442C の再改定が公表された。GD に関しては、*in vitro* 免疫毒性試験の DRP が公表された。引き続き、各国の専門家ともに、TG、GD や IATA 開発を進めていく。

F. 研究発表

F-1. 論文発表

1. Kimura Y, Yasuno R, Iwaki T, Fujimura C, Ohmiya Y, Nakajima Y, Omori T, Corsini E, Inoue T, Rogen EL, Kojima H, Aiba S. An international validation study of the interleukin-2 luciferase leukocyte toxicity test (IL-2 Luc LTT) to evaluate potential immunesuppressive chemicals and its performance after use with the interleukin-2 luciferase assay (IL-2 Luc assay), *Toxicol In Vitro*. 2022;88:105535.
2. Imamura M, Yamamoto Y, Fujita M, Wanibuchi S, Nakashima N, Kojima H, Ono A, Kasahara T. Applicability of ADRA (4 mM) for the prediction of skin sensitization by combining multiple alternative methods to

evaluate key events, *J Appl Toxicol*.2022;42(7): 1159-1167.
3. Yamamoto Y, Fujita M, Watanabe S, Yamaga H, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kawakami T, Kojima K, Sozu T, Kojima H, Kasahara T, Ono A. Within- and between-laboratory reproducibility and predictive capacity of amino acid derivative reactivity assay (ADRA) using a 0.5 mg/mL test chemical solution: Results of the study for reproducibility confirmation implemented in five participating laboratories, *J Appl Toxicol*.2022;42(6): 1078-1090.
4. Fujita M, Yamamoto Y, Wanibuchi S , Watanabe S, Yamaga H, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kawakami T, Kojima K, Sozu T, Kojima H, Kasahara T, Ono A. The within- and between-laboratories reproducibility and predictive capacity of Amino acid Derivative Reactivity Assay using 4 mM test chemical solution: Results of ring study implemented at five participating laboratories, *J Appl Toxicol*. 2022;42(2):318-333.
5. Fujita M, Nakashima N, Wanibuchi S , Yamamoto Y, Kojima H, Ono A, Kasahara T. Assessment of commercial polymers with and without reactive groups using amino acid derivative reactivity assay based on both molar concentration approach and gravimetric approach, *J Appl Toxicol.* 2023;43(3);446-457.
6. Piersma AH, Baker NC, Daston GP, Flick B, Fujiwara M, Knudsen TB, Spielmann H, Suzuki N, Tsaioun K, Kojima H. Pluripotent stem cell assays: Modalities and applications for predictive developmental toxicity, *Current Research in Toxicology*, 2022;3, 100074.
7. Anklam E, Bahl MI, Ball R, Beger RD, Cohen J, Fitzpatrick S, Kojima H, et al. Emerging technologies and their impact on regulatory science. *Exp Biol Med (Maywood).* 2022;247(1):1-75.
8. 小島肇, 渡辺美香．一般財団法人食品薬品安全センターにおける代替法研究, *秦野研究所年報*, 2022;45:8-17.

F-2．学会発表

1. Kojima H. Approach to New Approach Methods developed by Japan in OECD WNT, ICCA-LRI Workshop 2022（2022.6.20-21, 横浜）
2. 小島肇．３Rs を取り巻く国際動向と課題, 第 49 回日本毒性学会学術年会（2022.7.2, 札幌）
3. 小島肇．医薬品等の薬効及び安全性評価を支える in vitro 試験の未来, 日本組織培養学会第 94 回大会（2022.7.8 豊中,大阪）
4. 小島肇．代替法に関する国内外の状況について代替法全般の最新動向, 日本動物実験代替法学会企画委員会主催講習会（2022.8.25, Virtual）
5. 竹内和也，神谷孝平，渡辺真一，山鹿宏彰，若林晃次，田原宥，堀江宣行，藤本恵一，河上強志，小島幸一，寒水孝司，小島肇，山本裕介，藤田正晴，笠原利彦，小野敦．*in chemico* 皮膚感作性試験 ADRA 法のガイドライン改訂に向けたリング試験, 日本動物実験代替法学会 第 35 回大会（2022.11.19, 静岡）
6. 中嶋菜都美，山本裕介，鰐渕彩花，小島肇，小野敦，松本一彦，藤田正晴，笠原利彦．OECD の DASS 化合物セットを対象とした複数の感作性代替法を組み合わせた皮膚感作予測における ADRA の適用

性について, 日本動物実験代替法学会 第35回大会（2022.11.19, 静岡）

7. 鰐渕彩花，藤田正晴，中嶋菜都美，山本裕介，小島肇，小野敦，笠原利彦. ADRA による反応性基を有するポリマーの感作性評価, 日本動物実験代替法学会 第 35 回大会（2022.11.19, 静岡）
8. 小島肇. OECD テストガイダンス作成の経験から見たツールガイダンス整備の課題, 日本学術会議公開シンポジウム（2022.11.19, 東京）
9. 小島肇. Computational Toxicology の利用の実際と将来展望, 第96回日本薬理学会年会（2022.11.30, 横浜）

F-3. OECD 成果物

1.OECD TG 442C, in chemico skin sensitisation assays addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins
2. OECD Series on Testing and Assessment No. 360, Detailed Review Paper on In Vitro Test Addressing Immunotoxicity With a Focus on Immunosuppression
3. Emerging technologies in the Test Guidelines Programme に関するワークショップにおける小島発表資料

表１．DRPA モル濃度法および重量法の比較結果要約

No.	Curated.name	CASRN	Supplier	MW	Purity	Approach	DRPA prediction	DRPA prediction	LLNA call
							Lab 1	Lab2	
1	Bis-GMA	1565-94-2	Sigma-Aldrich	512.6	<=100%	Molar approach	Positive	Positive	1B
						Gravimetric approach	Positive	Positive	
2	Imidazolidinyl urea	39236-46-9	Sigma-Aldrich	388.3	<=100%	Molar approach	Positive	Positive	1B
						Gravimetric approach	Positive	Positive	
3	Tetrachlorosalicylanilide	1154-59-2	ThermoFisher Scientific（ACROS）	351	<100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
4	BADGE	1675-54-3	TCI	340.4	<=100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
5	Lauryl gallate	1166-52-5	TCI	338.4	>98%	Molar approach	Negative	Negative	1A
						Gravimetric approach	Negative	Negative	
6	Abietic acid	514-10-3	TCI	302.5	>80%	Molar approach	Positive	Positive	1B
						Gravimetric approach	Positive	Positive	
7	Dibenzoyl peroxide	94-36-0	Kanto Chemical	302.5	>80%	Molar approach	n.a	n.a	1A
						Gravimetric approach	n.a	n.a	
8	Iodocarb	55406-53-6	TCI	281.09	>97.0%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
9	Isopropyl myristate	110-27-0	Fujifilm Wako	270.5	95%	Molar approach	Negative	Negative	1B
						Gravimetric approach	Negative	Negative	
10	Bromothalonil	35691-65-7	TCI	265.9	>98%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
11	Benzyl cinnamate	103-41-3	Fujifilm Wako	238.3	98%	Molar approach	Negative	Negative	1B
						Gravimetric approach	Negative	Negative	
12	Benzyl salicylate	118-58-1	TCI	228.3	>99%	Molar approach	Negative	Negative	1B
						Gravimetric approach	Negative	Negative	
13	Farnesal	19317-11-4	Sigma-Aldrich	220.4	<=100%	Molar approach	Positive	Positive	1B
						Gravimetric approach	Positive	Positive	
14	Oxazolone	15646-46-5	Sigma-Aldrich	217.2	<=100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
15	4-Nitrobenzyl bromide	100-11-8	Sigma-Aldrich	216	<=100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
16	Propyl gallate	121-79-9	Sigma-Aldrich	212.2	<=100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
17	α-iso-Methylionone	127-51-5	TCI	206.3	>70%	Molar approach	Positive	Positive	1B
						Gravimetric approach	Positive	Positive	
18	Diphenylcyclopropenone	886-38-4	Sigma-Aldrich	206.2	<=100%	Molar approach	Positive	Positive	1A
						Gravimetric approach	Positive	Positive	
19	N,N-Dibutylaniline	613-29-6	Fujifilm Wako	205.3	98%	Molar approach	Negative	Negative	1B
						Gravimetric approach	Negative	Negative	
20	Lilial	80-54-6	TCI	204.3	>96%	Molar approach	Negative	Negative	1B
						Gravimetric approach	Negative	Negative	

Section 4
Health effects

Test Guideline No. 442C
In Chemico Skin Sensitisation

Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins

30 June 2022

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

OECD/OCDE

442C

Adopted:
30 June 2022

OECD KEY EVENT BASED GUIDELINE FOR THE TESTING OF CHEMICALS

In chemico skin sensitisation assays addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins

INTRODUCTION

Covalent binding to proteins Key Event based Test Guideline.

1. A skin sensitiser refers to a substance that will lead to an allergic response following repeated skin contact as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (1). There is general agreement on the key biological events underlying skin sensitisation. The current knowledge of the chemical and biological mechanisms associated with skin sensitisation has been summarised as an Adverse Outcome Pathway (AOP) (2) starting with a molecular initiating event through intermediate events to the adverse effect, namely allergic contact dermatitis. This AOP focuses on chemicals that react with amino-acid residues (i.e. cysteine or lysine) such as organic chemicals. In this instance, the molecular initiating event (i.e. the first key event), is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event in this AOP takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signalling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation.
2. The assessment of skin sensitisation has typically involved the use of laboratory animals. The classical methods that use guinea-pigs, the Guinea Pig Maximisation Test (GPMT) of Magnusson and Kligman and the Buehler Test (OECD TG 406) (11) assess both the induction and elicitation phases of skin sensitisation. The murine tests, such as the LLNA (OECD TG 429) (12) and its three non-radioactive modifications — LLNA:DA (OECD TG 442A) (13), LLNA:BrdU-ELISA, and BrdU-FCM (OECD TG 442B) (14) — all assess the induction response exclusively and have gained acceptance, since they provide an

advantage over the guinea pig tests in terms of animal welfare together with an objective measurement of the induction phase of skin sensitisation.

3. Mechanistically-based *in chemico* and *in vitro* test methods addressing the first three key events of the skin sensitisation AOP have been adopted for contributing to the evaluation of the skin sensitisation hazard potential of chemicals: the present Test Guideline assesses covalent binding to proteins, addressing the first key event; the OECD TG 442D assesses keratinocyte activation (15), the second key event and the OECD TG 442E addresses the activation of dendritic cells (16), the third key event of the skin sensitisation AOP. Finally, the fourth key event representing T-cell proliferation is indirectly assessed in the murine Local Lymph Node Assay (LLNA) (12).

Background and principles of the test methods included in the Key Event based Test Guideline

4. This Test Guideline (TG) describes *in chemico* assays that address mechanisms described under the first key event of the AOP for skin sensitisation, namely covalent binding to proteins (2). The test methods currently included in this Test Guideline are:

   - The Direct Peptide Reactivity Assay (DPRA) (Appendix I),
   - The Amino Acid Derivative Reactivity Assay (ADRA) (Appendix II), and
   - The kinetic Direct Peptide Reactivity Assay (kDPRA) (Appendix III).

5. The test methods are based on *in chemico* covalent binding to proteins and are considered to be scientifically valid. The DPRA has been evaluated in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)-lead validation study and subsequent independent peer review by the EURL ECVAM Scientific Advisory Committee (ESAC) (3) (4) (5). The ADRA underwent a validation study coordinated by the Japanese Center for the Validation of Alternative Methods (JaCVAM) (6) (7) (8) (9) followed by an independent peer-review (10). The kDPRA underwent an industry-coordinated validation study followed by an independent peer-review (17).

6. The test methods included in this Test Guideline might differ with regard to the procedures used to generate the data but can each be used to address countries' requirements for test results on protein reactivity, while benefiting from the Mutual Acceptance of Data.

7. The correlation of protein reactivity with skin sensitisation potential is well established (18) (19) (20). Nevertheless, since protein reactivity represents only one key event of the skin sensitisation AOP (2) (21), information generated with test methods developed to address this specific key event may not be sufficient as stand-alone methods to conclude on the presence or absence of skin sensitisation potential of chemicals. Therefore, data generated with the test methods described in this Test Guideline are proposed to be used within Integrated Approaches to Testing and Assessment (IATA), together with other relevant complementary information from *in vitro* assays addressing other key events of the skin sensitisation AOP as well as non-testing methods, including *in silico* modelling and read-across from chemical analogues (21). Examples on the use of data generated with these methods within Defined Approaches (DAs), i.e. approaches standardised both in relation to the set of information sources used and in the procedure applied to derive predictions, have been published (21) and are implemented in an OECD TG on defined approaches for skin sensitisation (22).

8. The DPRA and ADRA described in Appendixes I and II to this Test Guideline, respectively, support the discrimination of skin sensitisers (Category 1) from non-sensitisers. Depending on the regulatory framework, positive results generated with these methods may be used on their own to classify a chemical into UN GHS Category 1. However, these test methods do not allow on their own, the sub-categorisation of skin sensitisers into subcategories 1A

and 1B (23), as defined by UN GHS (1) for authorities implementing these two optional subcategories, or potency prediction for safety assessment decisions.

9. In contrast, the kDPRA described in Appendix III of this Test Guideline, allows discrimination of UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory 1A) i.e., subcategory 1B or no category (1) but does not allow to distinguish sensitisers (Category 1) from non-sensitisers. Depending on the regulatory framework, positive results generated with the kDPRA may be used on their own to classify a chemical into UN GHS subcategory 1A.

10. Definitions are provided in the Annex. Performance Standards for the assessment of proposed similar or modified *in vitro* skin sensitisation DPRA and ADRA test methods have been developed (24).

Literature for introduction

(1) United Nations (UN) (2021), Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth revised edition, UN New York and Geneva, 2021. Available at: [https://unece.org/sites/default/files/2021-09/GHS_Rev9E_0.pdf]

(2) OECD (2012), The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence, Series on Testing and Assessment No. 168, Organisation for Economic Cooperation and Development, Paris. Available at:http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2012)10/PART1&docLanguage=En

(3) GF Gerberick, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP (2004), Development of a peptide reactivity assay for screening contact allergens. Toxicol Sci. 81, 332-343.

(4) GF Gerberick, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP (2007), Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. Toxicol Sci. 97, 417-427. .

(5) EC EURL-ECVAM (2013), Recommendation on the Direct Peptide Reactivity Assay (DPRA) for the skin sensitisation testing Available at: https://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/eurl-ecvam-recommendaion-on-the-directpeptide-reactivity-assay-dpra.

(6) M Fujita, Yamamoto Y, Tahara H, Kasahara T, Jimbo Y, Hioki T (2014), Development of a prediction method for skin sensitisation using novel cysteine and lysine derivatives. J Pharmacol Toxicol Methods. 70, 94-105.

(7) Y Yamamoto, Tahara H, Usami R, Kasahara T, Jimbo Y, Hioki T, Fujita M.(2015) A novel *in chemico* method to detect skin sensitisers in highly diluted reaction conditions. J Appl Toxicol. 35, 1348-1360.

(8) M Fujita, Yamamoto Y, Watanabe S, Sugawara T, Wakabayashi K, Tahara K, Horie N, Fujimoto K, Kusakari K, Kurokawa Y, Kawakami T, Kojima K, Kojima H, Ono A, Katsuoka Y, Tanabe H, Yokoyama H and Kasahara T (2019), Cause of and Countermeasures for Oxidation of the Cysteine-Derived Reagent Used in the Amino acid Derivative Reactivity Assay, J. Appl. Toxicology, Feb;39(2):191-208 (doi: 10.1002/jat.3707).

(9) OECD (2019), Validation study report: Amino acid Derivative Reactivity Assay (ADRA), Series on testing and Assessment n° 304. Organisation for Economic Cooperation and Development, Paris.

(10) OECD (2019), Report of the Peer Review Panel on the Validation of the Amino acid Derivative Reactivity Assay (ADRA), Series on testing and Assessment n° 305. Organisation for Economic Cooperation and Development, Paris.

(11) OECD (2021), Test No. 406: Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264070660-en.

(12) OECD (2010), Test No. 429: Skin Sensitisation: Local Lymph Node Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264071100-en.

(13) OECD (2010), Test No. 442A: Skin Sensitization: Local Lymph Node Assay: DA, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264090972-en.

(14) OECD (2018), Test No. 442B: Skin Sensitization: Local Lymph Node Assay: BrdU-ELISA or –FCM, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264090996-en.

(15) OECD (2018), Test No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, OECD

Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264229822-en.

(16) OECD (2018), Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/9789264264359-en.

(17) Wareing, B., Kolle, S.N., Birk, B., Alépée, N., Haupt, T., Kathawala, R., Kern, P., Nardelli, L., Raabe, H., Rucki, M., Ryan, C., Verkaart, S., Westerink, W., Landsiedel, R., Natsch, A. (2020), 'The kinetic Direct Peptide Reactivity Assay (kDPRA): Intra- and inter-laboratory reproducibility in a seven-laboratory ring trial', ALTEX, preprint, DOI: 10.14573/altex.2004291.

(18) Landsteiner and Jacobs (1936), Studies on the sensitisation of animals with simple chemical compounds. Journal of Experimental Medicine 64:625-639.

(19) Dupuis and Benezra (1982), Allergic contact dermatitis to simple chemicals: a molecular approach. New York & Basel: Marcel Dekker Inc.

(20) JP Lepoittevin, Basketter DA, Goossens A, Karlberg AT (1998), Allergic contact dermatitis: the molecular basis, Springer, Berlin (doi: 10.1007/978-3-642-80331-4).

(21) OECD (2016), Guidance Document On The Reporting Of Defined Approaches And Individual Information Sources To Be Used Within Integrated Approaches To Testing And Assessment (IATA) For Skin Sensitisation, Annex 1 and Annex 2. ENV/JM/HA(2016)29, Series on Testing & Assessment No. 256, Organisation for Economic Cooperation and Development, Paris. Available at: [Series on Testing and Assessment: publications by number - OECD].

(22) OECD (2021), Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, https://doi.org/10.1787/b92879a4-en.

(23) B Wareing, Urbisch D, Kolle SN, Honarvar N, Sauer UG, Mehling A, Landsiedel R(2017) Prediction of skin sensitization potency sub-categories using peptide reactivity data, Toxicol In Vitro Dec;45(Pt 1):134-145 (doi: 10.1016/j.tiv.2017.08.015).

(24) OECD (2022), Draft Updated Performance Standards for the assessment of proposed similar or modified in vitro skin sensitisation DPRA and ADRA test methods, Series on Testing & Assessment No. 303, Organisation for Economic Cooperation and Development, Paris.

Annex 1.A. DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of relevance. The term is often used interchangeably with concordance to mean the proportion of correct outcomes of a test method (1). The formula used to derive accuracy is shown under "Calculation" of predictive capacity.

ADRA: Amino acid Derivative Reactivity Assay.

AOP (Adverse Outcome Pathway): sequence of events from the chemical structure of a target chemical or group of similar chemicals through the molecular initiating event to an *in vivo* outcome of interest (2).

Balanced accuracy: The average of sensitivity and specificity. This metric is particularly useful when a different number of in vivo positive and in vivo negative chemicals were tested. It is an important consideration in assessing the relevance of a test method. The formula used to derive balanced accuracy is shown under "Calculation" of predictive capacity.

Calculation

Calculating predictive capacity

Sensitivity, specificity, accuracy, and balanced accuracy are calculated based on the true positive (TP), true negative (TN), false negative (FN), and false positive (FP) values as follows:

$$\text{Sensitivity} = \frac{\text{Number of true positives (TP)}}{\text{Number of all positive chemicals (TP+FN)}} \text{ x } 100$$

$$\text{Specificity} = \frac{\text{Number of true negatives (TN)}}{\text{Number of all negative chemicals (TN+FP)}} \text{ x } 100$$

$$\text{Accuracy} = \frac{\text{Number of correct predictions (TP+TN))}}{\text{Number of all chemicals (TP+FN+TN+FP)}} \text{ x } 100$$

$$\text{Balanced accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

Calibration curve: The relationship between the experimental response value and the analytical concentration (also called standard curve) of a known substance.

Coefficient of variation: a measure of variability that is calculated for a group of replicate data by dividing the standard deviation by the mean. It can be multiplied by 100 for expression as a percentage.

Defined Approach (DA): a DA consists of a fixed data interpretation procedure (e.g. statistical, mathematical models) applied to data (e.g. *in silico* predictions, *in chemico*, *in vitro* data) generated with a defined set of information sources to derive a prediction.

DPRA: Direct Peptide Reactivity Assay.

EDTA: Ethylenediaminetetraacetic acid.

EURL ECVAM: the European Union Reference Laboratory for Alternatives to Animal Testing.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

IATA (Integrated Approach to Testing and Assessment): A structured approach used for hazard identification (potential), hazard characterisation (potency), and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazards, risks, and the need for further targeted and therefore minimal testing.

JaCVAM: Japanese Center for the Validation of Alternative Methods.

kDPRA: kinetic Direct Peptide Reactivity Assay.

$\mathbf{k_{max}}$: is the maximum rate constant (in $s^{-1}M^{-1}$) determined from the reaction kinetics for a tested substance in the kDPRA (see Appendix III, paragraph 24).

LLNA: murine Local Lymph Node Assay issued as OECD TG 429 in 2010.

Molecular Initiating Event: Chemical-induced perturbation of a biological system at the molecular level identified to be the starting event in the adverse outcome pathway.

Mixture: A solid or liquid comprising two or more substances which do not react chemically (3).

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent comprises at least 80% (w/w) of the whole.

Multi-constituent substance: A substance, defined by its quantitative composition, in which two or more main constituents are present in concentrations ≥ 10% (w/w) and < 80% (w/w). Multi-constituent substances are the result of a manufacturing process. The difference between a mixture and a multi-constituent substance is that a mixture comprises two or more substances which do not react chemically, whereas a multi-constituent substance comprises two or more substances that do react chemically.

NAC: N-(2-(1-naphthyl) acetyl)-L-cysteine (4) (5) (6).

NAL: α-N-(2-(1-naphthyl) acetyl)-L-lysine (4) (5) (6).

Positive control: A replicate containing all components of a test system and treated with a substance known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be excessive.

Pre-haptens: chemicals which become sensitisers through abiotic transformation.

Pro-haptens: chemicals requiring enzymatic activation to exert skin sensitisation potential.

Reference control: An untreated sample containing all components of a test system, including the solvent or vehicle that is processed with the test chemical treated and other control samples to establish the baseline response for the samples treated with the test chemical dissolved in the same solvent or vehicle. When tested with a concurrent negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (1).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability (1).

Reproducibility: The concordance of results obtained from testing the same substance using the same test protocol (see reliability). (1)

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (1). The formula used to derive sensitivity is shown under ”Calculation” of predictive capacity.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method (1). The formula used to derive specificity is shown under “Calculation” of predictive capacity.

Substance: Chemical elements and their compounds in the natural state or resulting from a manufacturing process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process, but excluding solvents that may be separated without affecting the stability of the substance or changing its composition (3).

System suitability: Determination of instrument performance (e.g., sensitivity) by analysis of reference standards prior to running the analytical run (7).

Test chemical: The term test chemical is used to refer to the substance being tested.

TFA: Trifluoroacetic acid.

United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS): A system proposing the classification of chemicals (substances and mixtures) according to standardised types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (3).

UVCB: substances of unknown or variable composition, complex reaction products or biological materials.

Valid test method: A test method considered to have sufficient relevance and reliability for a specific purpose and which is based on scientifically sound principles. A test method is never valid in an absolute sense, but only in relation to a defined purpose (1).

Literature for definitions

(1) OECD (2005), Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Series on Testing and Assessment, No. 34. Organisation for Economic Cooperation and Development, Paris, France.

(2) OECD (2012), The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Series on Testing and Assessment No. 168, OECD, Paris.

(3) United Nations (UN) (2021), Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth revised edition, UN New York and Geneva, 2021. Available at: https://unece.org/sites/default/files/2021-09/GHS_Rev9E_0.pdf

(4) M Fujita, Yamamoto Y, Tahara H, Kasahara T, Jimbo Y and Hioki T (2014), Development of a prediction method for skin sensitisation using novel cysteine and lysine derivatives, Journal of pharmacological and toxicological methods, 70:94-105.

(5) Y Yamamoto, Tahara H, Usami R, Kasahara T, Jimbo Y, Hioki T and Fujita M (2015), A novel *in chemico* method to detect skin sensitisers in highly diluted reaction conditions, Journal of Applied Toxicology, 35(11):1348-60, (doi: 10.1002/jat.3139).

(6) M Fujita, Yamamoto Y, Watanabe S, Sugawara T, Wakabayashi K, Tahara K, Horie N, Fujimoto K, Kusakari K, Kurokawa Y, Kawakami T, Kojima K, Kojima H, Ono A, Katsuoka Y, Tanabe H, Yokoyama H and Kasahara T (2019), Cause of and Countermeasures for Oxidation of the Cysteine-Derived Reagent Used in the Amino acid Derivative Reactivity Assay, J. Appl. Toxicology, Feb;39(2):191-208 (doi: 10.1002/jat.3707).

(7) FDA (Food and Drug Administration) (2018), Guidance for Industry: Bioanalytical Method Validation 41 pp. Accessible at: https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf

APPENDIX I

In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA)

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

1. The DPRA is proposed to address the molecular initiating event of the skin sensitisation AOP, namely protein reactivity, by quantifying the reactivity of test chemicals towards model synthetic peptides containing either lysine or cysteine (1). Cysteine and lysine percent peptide depletion values are then used to categorise a substance in one of four classes of reactivity for supporting the discrimination between skin sensitisers and non-sensitisers (2).

2. The DPRA test method proved to be transferable to laboratories experienced in high-performance liquid chromatography (HPLC) analysis. The level of reproducibility in predictions that can be expected from the test method is in the order of 85% within laboratories and 80% between laboratories (3). Results generated in the validation study (4) and published studies (5) overall indicate that the accuracy of the DPRA in discriminating sensitisers (i.e. UN GHS Category 1) from non-sensitisers is 80% (N=157) with a sensitivity of 80% (88/109) and specificity of 77% (37/48) when compared to LLNA results. The DPRA is more likely to under predict chemicals showing a low to moderate skin sensitisation potency (i.e. UN GHS subcategory 1B) than chemicals showing a high skin sensitisation potency (i.e. UN GHS subcategory 1A) (4) (5). However, the accuracy values given here for the DPRA as a stand-alone test method are only indicative since the test method should be considered in combination with other sources of information in the context of an IATA or a DA and in accordance with the provisions of paragraphs 7 and 8 in the General introduction. Furthermore when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the LLNA test as well as other animal tests may not fully reflect the situation in the species of interest, i.e. humans. On the basis of the overall data available, the DPRA was shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, skin sensitisation potency (as determined in *in vivo* studies) and physico-chemical properties (1) (2) (3) (5). Taken together, this information indicates the usefulness of the DPRA to contribute to the identification of skin sensitisation hazard.

3. The term "test chemical" is used in this Test Guideline to refer to what is being tested[1] and is not related to the applicability of the DPRA to the testing of substances and/or mixtures (see a summary of the known limitations of the DPRA in Annex 1 of this Appendix). This test method is not applicable for the testing of metal compounds since they are known to react with proteins with mechanisms other than covalent binding. A test chemical should be soluble in an appropriate solvent at a final concentration of 100 mM (see paragraph 10).

[1] In June 2013, the Joint Meeting agreed that where possible, a more consistent use of the term “test chemical” describing what is being tested should now be applied in new and updated Test Guidelines.

However, test chemicals that are not soluble at this concentration may still be tested at lower soluble concentrations. In such a case, a positive result could still be used to support the identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result. Limited information is currently available on the applicability of the DPRA to mixtures of known composition (4) (5). The DPRA is nevertheless considered to be technically applicable to the testing of multi-constituent substances and mixtures of known composition (see paragraph 4 and 10). When considering testing of mixtures, difficult-to-test chemicals (e.g. unstable), or test chemicals not clearly within the applicability domain described in this Appendix of the Test Guideline, upfront consideration should be given to whether the results of such testing will yield results that are meaningful scientifically. The current prediction model cannot be used for complex mixtures of unknown composition or for substances of unknown or variable composition, complex reaction products or biological materials (i.e. UVCB substances) due to the defined molar ratio of test chemical and peptide. For this purpose a new prediction model based on a gravimetric approach will need to be developed. In cases where evidence can be demonstrated on the non-applicability of the test method to other specific categories of chemicals, the test method should not be used for those specific categories of chemicals.

4. The test method described in this Appendix of the Test Guideline is an *in chemico* method that does not encompass a metabolic system. Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential (i.e. pro-haptens) cannot be detected by the test method. Chemicals that become sensitisers after abiotic transformation (i.e. pre-haptens) are reported to be in most cases correctly detected by the test method (4) (9) (10). In the light of the above, negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA or a DA. Test chemicals that do not covalently bind to the peptide but promote its oxidation (i.e. cysteine dimerisation) could lead to a potential over estimation of peptide depletion, resulting in possible false positive predictions and/or assignment to a higher reactivity class (see paragraphs 21 and 22).

5. As described, the DPRA assay supports the discrimination between skin sensitisers and non-sensitisers. However, it may also potentially contribute to the assessment of sensitising potency (6) (11) when used in integrated approaches such as IATA or DA (12). However further work, preferably based on human data, is required to determine how DPRA results may possibly inform potency assessment.

PRINCIPLE OF THE TEST

6. The DPRA is an *in chemico* method which quantifies the remaining concentration of cysteine- or lysine-containing peptide following 24 hours incubation with the test chemical at 22.5-30°C. The synthetic peptides contain phenylalanine to aid in the detection. Relative peptide concentration is measured by high-performance liquid chromatography (HPLC) with gradient elution and UV detection at 220 nm. Cysteine- and lysine peptide percent depletion values are then calculated and used in a prediction model (see paragraph 21) which allows assigning the test chemical to one of four reactivity classes used to support the discrimination between sensitisers and non-sensitisers.

7. Prior to routine use of the method described in this Appendix, laboratories should demonstrate technical proficiency, using the ten proficiency substances listed in Annex 2.

PROCEDURE

8. This test method is based on the DPRA DB-ALM protocol n° 154 (7) which represents the protocol used for the EURL ECVAM-coordinated validation study. It is recommended that this protocol is used when implementing and using the method in the laboratory. The following is a description of the main components and procedures for the DPRA. If an alternative HPLC set-up is used, its equivalence to the validated set-up described in the DB-ALM protocol should be demonstrated (e.g. by testing the proficiency substances in Annex 2).

Preparation of the cysteine or lysine-containing peptides

9. Stock solutions of cysteine (Ac-RFAA**C**AA-COOH) and lysine (Ac-RFAA**K**AA-COOH) containing synthetic peptides of purity higher than 85% and preferably > 90%, should be freshly prepared just before their incubation with the test chemical. The final concentration of the cysteine peptide should be 0.667 mM in pH 7.5 phosphate buffer whereas the final concentration of the lysine peptide should be 0.667 mM in pH 10.2 ammonium acetate buffer. The HPLC run sequence should be set up in order to keep the HPLC analysis time less than 30 hours. For the HPLC set up used in the validation study and described in this test method, up to 26 analysis samples (which include the test chemical, the positive control and the appropriate number of solvent controls based on the number of individual solvents used in the test, each tested in triplicate), can be accommodated in a single HPLC run. All of the replicates analysed in the same run should use the identical cysteine and lysine peptide stock solutions. It is recommended to prove individual peptide batches for proper solubility prior to their use.

Preparation of the test chemical

10. Solubility of the test chemical in an appropriate solvent should be assessed before performing the assay following the solubilisation procedure described in the DPRA DB-ALM protocol (7). An appropriate solvent will dissolve the test chemical completely. Since in the DPRA the test chemical is incubated in large excess with either the cysteine or the lysine peptides, visual inspection of the forming of a clear solution is considered sufficient to ascertain that the test chemical (and all of its components in the case of testing a multi-constituent substance or a mixture) is dissolved. Suitable solvents are, acetonitrile, water, 1:1 mixture water:acetonitrile, isopropanol, acetone or 1:1 mixture acetone:acetonitrile. Other solvents can be used as long as they do not have an impact on the stability of the peptide as monitored with reference controls C (i.e. samples constituted by the peptide alone dissolved in the appropriate solvent; see Annex 3). If the test chemical is not soluble in any of the solvents mentioned above, DMSO can be used as a last resort and in minimal amounts. It is important to note that DMSO may lead to peptide dimerisation and as a result, it may be more difficult to meet the acceptance criteria. If DMSO is chosen, attempts should be made to first solubilise the test chemical in 300 µL of DMSO and dilute the resulting solution with 2700 µL of acetonitrile. If the test chemical is not soluble in this mixture, attempts should be made to solubilise the same amount of test chemicals in 1500 µL of DMSO and dilute the resulting solution with 1500 µL of acetonitrile. The test chemical should be pre-weighed into glass vials and dissolved immediately before testing in an appropriate solvent to prepare a 100 mM solution. For mixtures and multi-constituent substances of known composition, a single purity should be determined by the sum of the proportion of its constituents (excluding water), and a single apparent molecular weight should be determined by considering the individual molecular weights of each component in the mixture (excluding water) and their individual proportions. The resulting purity and apparent molecular weight should then be used to calculate the weight of test chemical necessary to prepare a 100 mM solution. For polymers for which a predominant molecular weight cannot be determined, the molecular weight of the monomer (or the apparent molecular weight of the various monomers constituting the polymer) may be considered to

prepare a 100 mM solution. However, when testing mixtures, multi-constituent substances or polymers of known composition, it should be considered to also test the neat chemical. For liquids, the neat chemical should be tested as such without any prior dilution by incubating it at 1:10 and 1:50 ratio with the cysteine and lysine peptides, respectively. For solids, the test chemical should be dissolved to its maximum soluble concentration in the same solvent used to prepare the apparent 100 mM solution. It should then be tested as such without any further dilution by incubating it at 1:10 and 1:50 ratio with the cysteine and lysine peptides, respectively. Concordant results (reactive or non-reactive) between the apparent 100 mM solution and the neat chemical should allow for a firm conclusion on the result.

Preparation of the positive control, reference controls and coelution controls

11. Cinnamic aldehyde (CAS 104-55-2; ≥95% food-grade purity) should be used as positive control (PC) at a concentration of 100 mM in acetonitrile. Other suitable positive controls providing mid-range depletion values may be used if historical data are available to derive comparable run acceptance criteria. In addition reference controls (i.e. samples containing only the peptide dissolved in the appropriate solvent) should also be included in the HPLC run sequence and these are used to verify the HPLC system suitability prior to the analysis (reference controls A), the stability of the reference controls over time (reference control B) and to verify that the solvent used to dissolve the test chemical does not impact the percent peptide depletion (reference control C) (see Annex 3). The appropriate reference control for each substance is used to calculate the percent peptide depletion for that substance (see paragraph 18). In addition, a co-elution control constituted by the test chemical alone for each of the test chemicals analysed should be included in the run sequence to detect possible co-elution of the test chemical with either the lysine or the cysteine peptide.

Incubation of the test chemical with the cysteine and lysine peptide solutions

12. Cysteine and lysine peptide solutions should be incubated in glass autosampler vials with the test chemical at 1:10 and 1:50 ratio respectively. If a precipitate is observed immediately upon addition of the test chemical solution to the peptide solution, due to low aqueous solubility of the test chemical, one cannot be sure how much test chemical remained in the solution to react with the peptide. Therefore, in such a case, a positive result could still be used, but a negative result is uncertain and should be interpreted with due care (see also provisions in paragraph 10 for the testing of chemicals not soluble up to a concentration of 100 mM). The reaction solution should be left in the dark at 22.5-30°C for 24±2 hours before running the HPLC analysis. Each test chemical should be analysed in triplicate for both peptides. Samples have to be visually inspected prior to HPLC analysis. If a precipitate or phase separation is observed, samples may be centrifuged at low speed (100-400xg) to force precipitate to the bottom of the vial as a precaution since large amounts of precipitate may clog the HPLC tubing or columns. If a precipitation or phase separation is observed after the incubation period, peptide depletion may be underestimated and a conclusion on the lack of reactivity cannot be drawn with sufficient confidence in case of a negative result.

Preparation of the HPLC standard calibration curve

13. A standard calibration curve should be generated for both the cysteine and the lysine peptides. Peptide standards should be prepared in a solution of 20% or 25% acetonitrile:buffer using phosphate buffer (pH 7.5) for the cysteine peptide and ammonium acetate buffer (pH 10.2) for the lysine peptide. Using serial dilution standards of the peptide stock solution (0.667 mM), 6 calibration solutions should be prepared to cover the range from 0.534 to 0.0167 mM. A blank of the dilution buffer should also be included in the standard calibration curve. Suitable calibration curves should have an $r^2>0.99$.

HPLC preparation and analysis

14. The suitability of the HPLC system should be verified before conducting the analysis. Peptide depletion is monitored by HPLC coupled with an UV detector (photodiode array detector or fixed wavelength absorbance detector with 220 nm signal). The appropriate column is installed in the HPLC system. The HPLC set-up described in the validated protocol uses a Zorbax SB-C-18 2.1 mm x 100 mm x 3.5 micron as preferred column. With this reversed-phase HPLC column, the entire system should be equilibrated at 30°C with 50% phase A (0.1% (v/v) trifluoroacetic acid in water) and 50% phase B (0.085% (v/v) trifluoroacetic acid in acetonitrile) for at least 2 hours before running. The HPLC analysis should be performed using a flow rate of 0.35 mL/min and a linear gradient from 10% to 25% acetonitrile over 10 minutes, followed by a rapid increase to 90% acetonitrile to remove other materials. Equal volumes of each standard, sample and control should be injected. The column should be re-equilibrated under initial conditions for 7 minutes between injections. If a different reversed-phase HPLC column is used, the set-up parameters described above may need to be adjusted to guarantee an appropriate elution and integration of the cysteine and lysine peptides, including the injection volume, which may vary according to the system used (typically in the range from 3-10 µL). Importantly, if an alternative HPLC set-up is used, its equivalence to the validated set-up described above should be demonstrated (e.g. by testing the proficiency substances in Annex 2). Absorbance is monitored at 220 nm. If a photodiode array detector is used, absorbance at 258 nm should also be recorded. It should be noted that some supplies of acetonitrile could have a negative impact on peptide stability and this has to be assessed when a new batch of acetonitrile is used. The ratio of the 220 peak area and the 258 peak area can be used as an indicator of co-elution. For each sample a ratio in the range of 90%<mean[2] area ratio of control samples<100% would give a good indication that co-elution has not occurred.

15. There may be test chemicals which could promote the oxidation of the cysteine peptide. The peak of the dimerised cysteine peptide may be visually monitored. If dimerisation appears to have occurred, this should be noted as percent peptide depletion may be over-estimated leading to false positive predictions and/or assignment to a higher reactivity class (see paragraphs 21 and 22).

16. The HPLC analysis should be timed to assure that the injection of the first sample starts 22 to 26 hours after the test chemical was mixed with the peptide solution. The HPLC run sequence should be set up in order to keep the HPLC analysis time less than 30 hours. For the HPLC set up used in the validation study and described in this test method, up to 26 analysis samples can be accommodated in a single HPLC run (see also paragraph 9). An example of HPLC analysis sequence is provided in Annex 3.

DATA AND REPORTING

Data evaluation

17. The concentration of cysteine or lysine peptide is photometrically determined at 220 nm in each sample by measuring the peak area (area under the curve, AUC) of the appropriate peaks and by calculating the concentration of peptide using the linear calibration curve derived from the standards.

18. The percent peptide depletion is determined in each sample by measuring the peak area and dividing it by the mean peak area of the relevant reference controls C (see Annex 3) according to the formula described below.

[2] For mean it is meant arithmetic mean throughout the document.

$$\text{Percent peptide depletion} = \left[1 - \left(\frac{\text{Peptide peak area in replicate injection}}{\text{Mean peptide peak area in reference controls } C}\right)\right] \times 100$$

Acceptance criteria

19. The following criteria should be met for a run to be considered valid:

a) the standard calibration curve should have an $r^2 > 0.99$,

b) the mean percent peptide depletion value of the three replicates for the positive control cinnamic aldehyde should be between 60.8% and 100% for the cysteine peptide and between 40.2% and 69.0% for the lysine peptide (for other positive controls a reference range needs to be established) and the maximum standard deviation (SD) for the positive control replicates should be <14.9% for the percent cysteine depletion and <11.6% for the percent lysine depletion and

c) the mean peptide concentration of reference controls A should be 0.50±0.05 mM and the coefficient of variation (CV) of peptide peak areas for the nine reference controls B and C in acetonitrile should be <15.0%.

If one or more of these criteria is not met the run should be repeated.

20. The following criteria should be met for a test chemical's results to be considered valid:

a) the maximum standard deviation for the test chemical replicates should be <14.9% for the percent cysteine depletion and <11.6% for the percent lysine depletion,

b) the mean peptide concentration of the three reference controls C in the appropriate solvent should be 0.50±0.05 mM.

If these criteria are not met the data should be rejected and the run should be repeated for that specific test chemical.

Prediction model

21. The mean percent cysteine and percent lysine depletion value is calculated for each test chemical. Negative depletion is considered as "0" when calculating the mean. By using the cysteine 1:10/lysine 1:50 prediction model shown in Table 1, the threshold of 6.38% average peptide depletion should be used to support the discrimination between skin sensitisers and non-sensitisers in the framework of an IATA or DA. Application of the prediction model for assigning a test chemical to a reactivity class (i.e. low, moderate and high reactivity) may perhaps prove useful to inform potency assessment within the framework of an IATA or DA.

Table 1: Cysteine 1:10/lysine 1:50 prediction model[1]

Mean of cysteine and lysine % depletion	Reactivity Class	DPRA Prediction[2]
0% ≤ mean % depletion ≤ 6.38%	No or minimal reactivity	Negative
6.38% < mean % depletion ≤ 22.62%	Low reactivity	Positive
22.62% < mean % depletion ≤ 42.47%	Moderate reactivity	
42.47% < mean % depletion ≤ 100%	High reactivity	

[1] The numbers refer to statistically generated threshold values and are not related to the precision of the measurement (2).
[2] A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 2 and 4.

22. There might be cases where the test chemical (the substance or one or several of the components of a multi-constituent substance or a mixture) absorbs significantly at 220 nm and has the same retention time of the peptide (co-elution). Co-elution may be resolved by slightly adjusting the HPLC set-up in order to further separate the elution time of the test chemical and the peptide. If an alternative HPLC set-up is used to try to resolve co-elution, its equivalence to the validated set-up should be demonstrated (e.g. by testing the proficiency substances in Annex 2). When co-elution occurs the peak of the peptide cannot be integrated and the calculation of the percent peptide depletion is not possible. If co-elution of such test chemicals occurs with both the cysteine and the lysine peptides, or with the cysteine peptide only, then the analysis should be reported as “inconclusive”. In cases where co-elution occurs only with the lysine peptide, then the cysteine 1:10 prediction model reported in Table 2 can be used.

Table 2: Cysteine 1:10 prediction model[1]

Cysteine (Cys) % depletion	Reactivity class	DPRA prediction[2]
0% ≤ Cys % depletion ≤ 13.89%	No or minimal reactivity	Negative
13.89% < Cys % depletion ≤ 23.09%	Low reactivity	Positive
23.09% < Cys % depletion ≤ 98.24%	Moderate reactivity	
98.24% < Cys % depletion ≤ 100%	High reactivity	

[1] The numbers refer to statistically generated threshold values and are not related to the precision of the measurement.
[2] A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 2 and 4.

23. There might be other cases where the overlap in retention time between the test chemical and either of the peptides is incomplete. In such cases percent peptide depletion values can be estimated and used in the cysteine 1:10/lysine 1:50 prediction model, however assignment of the test chemical to a reactivity class cannot be made with accuracy.

24. A single HPLC analysis for both the cysteine and the lysine peptide should be sufficient for a test chemical when the result is unequivocal. However, in cases of results

close to the threshold used to discriminate between positive and negative results (i.e. mean percent depletion falls in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10 prediction model), additional testing is recommended. In particular, in case of negative results in these ranges (i.e. 3% to 6.38% for the cysteine 1:10/lysine 1:50 prediction model or 9% to 13.89% for the cysteine 1:10 prediction model), a second run should be conducted, as well as a third one in case of discordant results between the first two runs.

Test report

25. The test report should include the following information

Test chemical and Controls (positive control and solvent/vehicle)

- Mono-constituent substance (test and control chemicals)
    - Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
    - Physicochemical properties such as physical state, appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;
    - Purity, chemical identity of impurities as appropriate and practically feasible, etc;
    - Treatment prior to testing, if applicable (e.g. warming, grinding);
    - Concentration(s) tested;
    - Storage conditions and stability to the extent available.
- Multi-constituent substance, UVCB and mixture:
    - Characterisation as far as possible by e.g. chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available;
    - Physical appearance, water solubility and additional relevant physicochemical properties, to the extent available;
    - Molecular weight or apparent molecular weight in case of mixtures/polymers of known compositions or other information relevant for the conduct of the study;
    - Treatment prior to testing, if applicable (e.g. warming, grinding);
    - Concentration(s) tested;
    - Storage conditions and stability to the extent available.
- Additional information for positive control
    - Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.
- Additional information for solvent/vehicle control
    - Solvent/vehicle used and ratio of its constituents, if applicable;
    - Justification for choice of solvent for each test chemical;
    - For acetonitrile, results of test of impact on peptide stability.

Peptides

- Supplier, lot, purity

HPLC instrument setting and analysis

- Type of HPLC instrument, HPLC and guard columns, detector, autosampler;
- Parameters relevant for the HPLC analysis such as column temperature, injection volumes, flow rate and gradient.

System suitability

- Peptide peak area at 220 nm of each standard and reference control A replicate;
- Linear calibration curve graphically represented and the r^2 reported;
- Peptide concentration of each reference control A replicate;
- Mean peptide concentration (mM) of the three reference controls A, SD and CV;
- Peptide concentration of reference controls A and C.

Analysis sequence

- For reference controls:
    - Peptide peak area at 220 nm of each B and C replicate;
    - Mean peptide peak area at 220 nm of the nine reference controls B and C in acetonitrile, SD an CV (for stability of reference controls over analysis time);
    - For each solvent used, the mean peptide peak area at 220 nm of the three appropriate reference controls C (for the calculation of percent peptide depletion);
    - For each solvent used, the peptide concentration (mM) of the three appropriate reference controls C;
    - For each solvent used, the mean peptide concentration (mM) of the three appropriate reference controls C, SD and CV.
- For positive control:
    - Peptide peak area at 220 nm of each replicate;
    - Percent peptide depletion of each replicate;
    - Mean percent peptide depletion of the three replicates, SD and CV.
- For each test chemical:
    - Appearance of precipitate in the reaction mixture at the end of the incubation time, if observed. If precipitate was re-solubilised or centrifuged;
    - Presence of co-elution;
    - Description of any other relevant observations, if applicable;
    - Peptide peak area at 220 nm of each replicate;
    - Percent peptide depletion of each replicate;
    - Mean of percent peptide depletion of the three replicate, SD and CV;
    - Mean of percent cysteine and percent lysine depletion values;
    - Prediction model used and DPRA prediction.

Proficiency testing

- Statement that the testing facility has demonstrated proficiency in the use of the test method before routine use by testing of the proficiency chemicals.

Discussion of the results

- Description of any unintended modifications to the test procedure.
- Discussion of the results obtained with the DPRA test method and if it is within the ranges described in paragraph 24.

Conclusion

LITERATURE FOR APPENDIX I

(1) Gerberick *et al.* (2004), Development of a peptide reactivity assay for screening contact allergens. Toxicological Sciences 81:332-343.

(2) Gerberick *et al.* (2007), Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicological Sciences 97:417-427.

(3) EC EURL-ECVAM (2013), Recommendation on the Direct Peptide Reactivity Assay (DPRA) for skin sensitisation testing. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra.

(4) EC EURL ECVAM (2012), Direct Peptide Reactivity Assay (DPRA) Validation Study Report 74pp. Accessible at: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/eurl-ecvam-recommendations/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra

(5) Natsch *et al.* (2013), A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. Journal of Applied Toxicology, published online, 9 April 2013, DOI:10.1002/jat.2868.

(6) Jaworska *et al.* (2013), Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice Journal of Applied Toxicology, published online, 14 May 2013, DOI: 10.1002/jat.2869.

(7) DB-ALM (INVITTOX) Protocol 154: Direct Peptide Reactivity assay (DPRA) for skin sensitisation testing 17pp. Accessible at: http://ecvam-dbalm.jrc.ec.europa.eu/

(8) ECETOC (2003), Contact sensitization: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No. 87).

(9) Urbisch et al. (2016), Assessment of Pre- and Pro-haptens Using Nonanimal Test Methods for Skin Sensitization, Chem Res Toxicol. 29(5):901-13 doi: 10.1021/acs.chemrestox.6b00055

(10) Pattlewicz et al. (2016), Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization? Regul Toxicol Pharmacol.; 82:147-155. doi: 10.1016/j.yrtph.2016.08.007

(11) Jaworska et al. (2015), Arch Toxicol. 2015 Dec;89(12):2355-83. doi: 10.1007/s00204-015-1634-2)

(12) OECD (2016). Series on Testing & Assessment No. 256: Guidance Document On The Reporting Of Defined Approaches And Individual Information Sources To Be Used Within Integrated Approaches To Testing And Assessment (IATA) For Skin Sensitisation, Annex 1 and Annex 2. ENV/JM/HA(2016)29. Organisation for Economic Cooperation and Development, Paris. Available at: [https://community.oecd.org/community/iatass].

APPENDIX I, ANNEX 1

KNOWN LIMITATIONS OF THE DIRECT PEPTIDE REACTIVITY ASSAY

The table below provides a summary of the known limitations of the DPRA.

Substance class / interference	Reason for potential underprediction or interference	Data interpretation	Example substance
Metals and inorganic compounds	Known to react with proteins via mechanisms other than covalent binding	Should not be tested	Nickel sulphate; 7786-81-4
Pro-haptens	Test Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential; cannot be detected by the test method unless activation is caused by auto-oxidation to a similar degree as in vivo /in humans. It will however normally not be known whether this will be the case	May lead to false negatives. Negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA or a DA	Diethylenetriamine; 111-40-0 (1A chez l'homme, LLNA n/a)
Pre-haptens	Chemicals that become sensitisers after abiotic transformation are reported to be in most cases correctly detected by the test method		Linalool: 78-70-6
Test chemicals absorbing significantly at 220 nm and having the same retention time of the peptides (co-elution)	When co-elution occurs the peak of the peptide cannot be integrated and the calculation of the percent peptide depletion is not possible	If co-elution of such test chemicals occurs with both the cysteine and the lysine peptides, or with the cysteine peptide only, then the analysis should be reported as "inconclusive" and alternative HPLC set up should be considered (see paragraph 22). In cases where co-elution occurs only with the lysine peptide, then the cysteine 1:10 prediction model reported in Table 2 can be used.	Salicylic acid: 69-72-7
Complex mixtures of unknown composition, substances of unknown or variable composition, complex reaction products or biological materials	This is due to the need for defined molar ratio of test chemical and peptide	n/a	UVCBs, chemical emissions, products or formulations with variable or not fully known composition
Test chemicals which cannot be dissolved in an appropriate solvent at a final concentration of 100 mM	Not sure if sufficient exposure can be achieved	Test chemicals that are not soluble at this concentration may still be tested at lower soluble concentrations. In such a case, a positive result could be used to support the identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result.	n/a
Chemicals which precipitate in reaction solution	Not sure if sufficient exposure can be achieved	A conclusion on the lack of reactivity cannot be drawn with sufficient confidence in case of a negative result	Isopropyl myristate CAS: 110-27-0
Test chemicals that do not covalently bind to the cysteine-peptide but promote	Could lead to a potential over-estimation of cysteine-peptide depletion, resulting in possible false positive predictions.		DMSO Oxidant

its oxidation (i.e. cysteine dimerisation)			
Test chemicals that are only soluble in DMSO	DMSO causes excessive peptide depletion due to cysteine dimerization resulting in high background cysteine depletion.	May lead to false negative results	n/a

APPENDIX I, ANNEX 2

PROFICIENCY SUBSTANCES

In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay

Prior to routine use of the test method described in this test method, laboratories should demonstrate technical proficiency by correctly obtaining the expected DPRA prediction for the 10 proficiency substances recommended in Table 1 and by obtaining cysteine and lysine depletion values that fall within the respective reference range for 8 out of the 10 proficiency substances for each peptide. These proficiency substances were selected to represent the range of responses for skin sensitisation hazards. Other selection criteria were that they are commercially available, that high quality *in vivo* reference data and high quality *in vitro* data generated with the DPRA are available, and that they were used in the EURL ECVAM-coordinated validation study to demonstrate successful implementation of the test method in the laboratories participating in the study.

Table 1: Recommended proficiency substances for demonstrating technical proficiency with the Direct Peptide Reactivity Assay

Proficiency substances	CASRN	Physical state	*In vivo* prediction[1]	DPRA prediction[2]	Range[3] of % cysteine peptide depletion	Range[3] of % lysine peptide depletion
2,4-Dinitrochlorobenzene	97-00-7	Solid	Sensitiser (extreme)	Positive	90-100	15-45
Oxazolone	15646-46-5	Solid	Sensitiser (extreme)	Positive	60-80	10-55
Formaldehyde	50-00-0	Liquid	Sensitiser (strong)	Positive	30-60	≤ 24
Benzylideneacetone	122-57-6	Solid	Sensitiser (moderate)	Positive	80-100	≤ 7
Farnesal	19317-11-4	Liquid	Sensitiser (weak)	Positive	15-55	≤ 25
2,3-Butanedione	431-03-8	Liquid	Sensitiser (weak)	Positive	60-100	10-45
1-Butanol	71-36-3	Liquid	Non-sensitiser	Negative	≤ 7	≤ 5.5
6-Methylcoumarin	92-48-8	Solid	Non-sensitiser	Negative	≤ 7	≤ 5.5
Lactic Acid	50-21-5	Liquid	Non-sensitiser	Negative	≤ 7	≤ 5.5
4-Methoxyacetophenone	100-06-1	Solid	Non-sensitiser	Negative	≤ 7	≤ 5.5

[1]The *in vivo* hazard and (potency) predictions are based on LLNA data (5). The *in vivo* potency is derived using the criteria proposed by ECETOC (8).
[2] A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 2 and 4.
[3] Ranges determined on the basis of at least 10 depletion values generated by 6 independent laboratories.

APPENDIX I, ANNEX 3

EXAMPLES OF ANALYSIS SEQUENCE

Calibration standards and reference controls	STD1 STD2 STD3 STD4 STD5 STD6 Dilution buffer Reference control A, rep 1 Reference control A, rep 2 Reference control A, rep 3
Co-elution controls	Co-elution control 1 for test chemical 1 Co-elution control 2 for test chemical 2
Reference controls	Reference control B, rep 1 Reference control B, rep 2 Reference control B, rep 3
First set of replicates	Reference control C, rep 1 Cinnamic aldehyde, rep 1 Sample 1, rep 1 Sample 2, rep 1
Second set of replicates	Reference control C, rep 2 Cinnamic aldehyde, rep 2 Sample 1, rep 2 Sample 2, rep 2
Third set of replicates	Reference control C, rep 3 Cinnamic aldehyde, rep 3 Sample 1, rep 3 Sample 2, rep 3
Reference controls	Reference control B, rep 4 Reference control B, rep 5 Reference control B, rep 6

Three sets of reference controls (i.e. samples constituted only by the peptide dissolved in the appropriate solvent) should be included in the analysis sequence:

Reference control A: used to verify the suitability of the HPLC system.

Reference control B: included at the beginning and at the end of the analysis sequence to verify stability of reference controls over the analysis time.

Reference control C: included in the analysis sequence to verify that the solvent used to dissolve the test chemical does not impact the percent peptide depletion.

APPENDIX II

In Chemico Skin Sensitisation: Amino acid Derivative Reactivity Assay (ADRA)

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

1. The ADRA is proposed to address the molecular initiating event of the skin sensitisation AOP - namely, protein reactivity - by quantifying the reactivity of test chemicals towards model synthetic amino acid derivatives containing either lysine or cysteine (1) (2) (3). Depletion values of the cysteine derivative N-(2-(1-naphthyl)acetyl)-L-cysteine (CAS. 32668-00-1), which is known as NAC, and the lysine derivative α-N-(2-(1-naphthyl)acetyl)-L-lysine (CAS. 397841-92-8), known as NAL are then used to support the discrimination between skin sensitisers and non-sensitisers (1) (2) (3).

2. The reproducibility and transferability of the ADRA protocol were confirmed using validation studies coordinated by the Japanese Center for validation of alternative methods (JaCVAM) (4) (5) (6) (7) (8) (9) (10). There are two detection types of ADRA: ultraviolet (UV) detection and fluorescence (FL) detection (11) (12). Within-laboratory reproducibility (WLR) and between-laboratory reproducibility (BLR) of ADRA were 100% each determined using both the UV detection and fluorescence detection (9) (10). Prediction of skin sensitisation potential based on local lymph node assay (LLNA) data indicated that ADRA with UV-detection identified sensitisers and non-sensitisers with an accuracy of 76 % (104/136), a sensitivity of 76% (74/98), a specificity of 79% (30/38) and a balanced accuracy of 77% (8). In addition, the prediction of the skin sensitisation potential based on human data indicated that ADRA with UV detection has an accuracy of 84% (67/80), a sensitivity of 83% (48/58), a specificity of 86% (19/22) and a balanced accuracy of 84% (8). However, the accuracy values given here for ADRA as a stand-alone test method are for reference only, since it is recommended that the test method be used in combination with other sources of information in the context of an IATA and in accordance with the provisions of paragraphs 7 and 8 in the General Introduction. Furthermore, when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the LLNA as well as other animal tests may not fully reflect the situation in humans. On the basis of the overall data available, ADRA's applicability domain was shown to include a variety of organic functional groups, reaction mechanisms, skin sensitisation potencies (as determined in in vivo studies), and physicochemical properties (1) (2) (3) (4). Following an independent peer review, the ADRA validation studies were considered to demonstrate that this method should be acceptable as part of an integrated testing strategy for the predictive identification of skin sensitisation hazard (6) (13) (14).

3. Co-elution occurs when the test chemical (the substance or one or several of the constituents of a multi-constituent substance or a mixture) was detected significantly at an OD of 281 nm (UV detector) or Ex/Em 284/333 nm (FL detector) and has the same retention time as NAC or NAL (15). Co-elution of UV absorbing-compounds using with the nucleophiles NAC and NAL can lead to inconclusive results when using conventional ultraviolet (UV) detection (11) (12). This problem can be prevented by an alternative or parallel measurement using a fluorescence (FL) detector; thus, the depletion values

obtained by simultaneous measurement using both detectors were also collected in the validation studies (9) (10) and equivalent results to those obtained with UV-detection were obtained, indicating that both detection methods are valid, but FL-detection may lead to fewer inconclusive results. Known limitations of the ADRA are tabulated in Appendix II, Annex 1.

4. The term "test chemical" is used in this Test Guideline to refer to what is being tested[3]. This test method is not applicable to the testing of metal compounds, which are known to react with proteins via mechanisms other than covalent binding. The test method described in this Appendix of the Test Guideline is an in chemico method that does not encompass a metabolic system. Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential (i.e., pro-haptens) cannot be detected by the test method. Chemicals that become sensitisers after abiotic transformation (i.e., pre-haptens) are reported to be in some cases correctly detected by the test method (1) (2) (3) (4) (7) (8). In the light of the above, negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA. Test chemicals that promote the oxidation of the N-(2-(1-naphthyl)acetyl)-L-cysteine (NAC) reagent (i.e. cysteine dimerisation) could lead to a potential over-estimation of NAC depletion, resulting in possible false positive predictions (see paragraph 22 and Appendix II, Annex 1); it may be possible to detect and quantify any NAC dimer formed by high-performance liquid chromatography (HPLC) using a UV detector, thus confirming or ruling out that the NAC reagent has been depleted via oxidative dimerisation as opposed to reaction and covalent bonding to the test item substance(s).

5. The ADRA test method allows testing of poorly soluble chemicals (16). To be tested, a test chemical should be soluble in an appropriate solvent at a final concentration of 4 mM (see paragraph 14). Test chemicals that are not soluble at this concentration may still be tested at lower concentrations. In such cases, a positive result could still be used to support identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result.

6. The nucleophilic reagents used in ADRA are quantified at 281 nm (1) (2). In the case of co-elution of the nucleophilic reagent and the UV-absorbing test chemical, this might result in inconclusive predictions. However, substances that absorb UV in this range of the spectrum are generally limited to those having conjugated double bonds, which significantly lowers the potential for inconclusive results due to co-elution of UV-absorbing components (15). Furthermore, NAC and NAL are fluorescent and thus, they can be detected using a FL detector (11) (12). Since test chemicals rarely have fluorescence at the specific excitation/emission wavelengths, it is possible to further reduce frequency of inconclusive results by using a FL detector. This is particularly useful in the case of multi-constituent substances with UV absorbance.

7. When assessing the sensitisation potential of a test chemical by using ADRA, there are two options for the preparation of the stock solution (see Figure 1 and paragraphs 15-16): a) If the test chemical is a mono-constituent substance with a known molecular weight or a mixture or multi-constituent substance of known composition, ADRA should be performed using a stock solution prepared at a concentration of 4 mM (8); b) If the test chemical is a mono-constituent substance of unknown molecular weight or a mixture and there is no defined molecular weight (mixtures of unknown or variable composition, complex reaction products, or biological materials (UVCB)), ADRA should be performed using a gravimetric approach based on a stock solution prepared at 0.5 mg/mL. In addition, the gravimetric approach with ADRA (0.5 mg/mL) can also be used for polymers. Assessment of the predictive capacity of ADRA conducted with this gravimetric approach indicated that ADRA (0.5 mg/mL) identified sensitisers and non-sensitisers with an accuracy of 76 % (103/136), a sensitivity of 74% (73/98), a specificity of 79% (30/38) and

[3] In June 2013, the Joint Meeting agreed that where possible, a more consistent use of the term “test chemical” describing what is being tested should now be applied in new and updated Test Guidelines.

a balanced accuracy of 77% when compared to LLNA data (8). In addition, the predictive capacity for human data indicated that the gravimetric ADRA (0.5 mg/mL) has an accuracy of 83% (66/80), a sensitivity of 81% (47/58), and a specificity of 86% (19/22) (8). The molecular weight range of the test chemicals used in the validation study of ADRA (0.5 mg/mL) was 60.10 - 388.29, and the ratio of nucleophilic reagent to test chemical in the reaction solution at that time was 1:416 - 1:64 (9).

8. ADRA can be used to support the discrimination between skin sensitisers and non-sensitisers. Further work, preferably based on human data, is necessary to determine whether ADRA results can contribute to potency assessment when considered in combination with other information sources (13) (14).

PRINCIPLE OF THE TEST

9. ADRA is an in chemico test method that quantifies residual concentrations of the NAC and NAL, following a 24±1 hour incubation at 25±1ºC in the presence of a test chemical. Both these derivatives include a naphthalene ring that is introduced to their N-terminal in order to facilitate UV detection and FL detection. The relative concentrations of NAC and NAL are measured by HPLC using UV detection (optical density, 281 nm), optionally in combination with FL detection (excitation/emission [Ex/Em], 284/333 nm) and with gradient elution (see paragraph 19). To ultimately support the discrimination between skin sensitisers and non-sensitisers, percent depletion values are then calculated for both NAC and NAL and compared to a prediction model (see paragraph 27).

10. Prior to routine use of the method described in this test method, laboratories should demonstrate technical proficiency, using the ten proficiency substances listed in Appendix II, Annex 2.

PROCEDURE

11. This test method is based on the protocol (17) used for the JaCVAM-coordinated ADRA validation study and is recommended for use when implementing ADRA at a laboratory. The main components and procedures for the ADRA are described below. Before using an alternative HPLC set-up, its equivalence to the validated set-up described in the protocol should be demonstrated, preferably by testing the proficiency substances in Appendix II, Annex 2.

Quality of NAC and NAL

12. The Nucleophilic Reagents can be obtained as an ADRA Kit for Skin Sensitisation Test, from FUJIFILM Wako Pure Chemical Corporation, Catalogue No. 296-80901. The use of NAC/NAL as reagent for detecting sensitisation is patented in Japan only, by Fujifilm Corporation. Therefore, in other countries, NAC/NAL can be used without permission. In case other manufacturer's NAC/NAL are used, these should satisfy three quality criteria described below. Quality checks can be obviated and ADRA testing can be performed without delay by purchasing NAC and NAL that have been manufactured specifically to satisfy these quality criteria.

Quality required for NAC and NAL:

1) Purity: Both NAC and NAL are to be at least 98% pure.

2) Stability: Using NAC and NAL stock solution, prepare a reference control free of any test chemical and quantify the residual levels of NAC and NAL both immediately after preparation (0 hours) and after a 24 hour incubation. The residual level of NAC and NAL is calculated as follows:

$$\text{Residual levels of NAC} = \frac{\text{Peak area of NAC}}{\text{Total peak area of NAC and NAC dimer}} \times 100$$

$$\text{Residual levels of NAL} = \frac{\text{Peak area of NAL at 24 hour}}{\text{Peak area of NAL at 0 hour}} \times 100$$

The main cause of NAC stability degradation is dimerisation, which may affect reactivity with the test chemical and test reproducibility (3). Therefore, the residual level of NAC should be calculated with respect to the total amount of NAC and dimer. Since the dimers may be formed over time or may have already been formed during the preparation of the stock solution, residual level of NAC is calculated at the time of stock solution preparation and after 24 hours. Residual levels of NAC (both of 0 hour and 24 hour) and NAL (24 hour) should be a minimum of 90% in either case (17).

3) Reactivity: NAC and NAL are to be evaluated for reactivity with the ten proficiency substances given in Appendix II, Annex 2 and should satisfy the requirement given therein.

Preparation of the NAC and NAL stock solution

13. The solubility of individual NAC and NAL batches should be verified prior to use. NAC stock solution should be prepared to a concentration of 2 mM in 100 mM of pH 8.0 phosphate buffer, including 0.333 µM of EDTA, as well as NAL stock solution to a concentration of 2 mM in 100 mM of pH 10.2 phosphate buffer. These two stock solutions are then diluted in buffer to prepare 6.667 µM stock solutions. Both NAC and NAL stock solutions should be used as soon as possible after preparation (3). In the event that they are to be stored, these stock solutions may be frozen and stored for up to twelve months time at less than -75°C prior to use. The final concentration of the NAC in the incubation mixture is 5 µM in pH 8.0 phosphate buffer, and the final concentration of the NAL in the incubation mixture is 5 µM in pH 10.2 phosphate buffer.

Preparation of the test chemical solution

14. Solubility of the test chemical in an appropriate solvent should be assessed before performing the assay in accordance with the solubilisation procedure described in the ADRA JaCVAM protocol (17). An appropriate solvent should dissolve the test chemical completely. Since the ADRA protocol stipulates that either NAC or NAL are incubated in an excess volume of the test chemical, visual inspection of the clear test chemical solution is considered sufficient to confirm that the test chemical (and all its constituents, if testing a multi-constituent substance or a mixture) is dissolved (17). Suitable solvents are distilled water, acetonitrile and acetone. If the test chemical is not soluble in any of the solvents mentioned above, DMSO can be used as a last resort and in minimal amounts (19). It is important to note that DMSO may lead to dimerisation of the nucleophilic reagent NAC (18) (19) and as a result, it may be more difficult to meet the acceptance criteria. If a DMSO-acetonitrile solvent is chosen (5% DMSO in acetonitrile), the test chemical should be dissolved at 80 mM in DMSO, and then this solution should be diluted 20-fold with

acetonitrile to prepare a 4 mM test chemical solution. In case the use of DMSO leads to increased dimerisation of the NAC reagent, this can be checked analytically as the NAC dimer can be detected by HPLC. If a solvent other than those already considered appropriate for the ADRA is used for the test chemical, it is necessary to confirm that the solvent itself does not lead to NAC or NAL depletion (e.g., dimerisation, oxidation) and does not degrade or disrupt the integrity of the test subtances or mixture components. The test chemical should be pre-weighed into a disposable polypropylene tube and dissolved immediately before testing in an appropriate solvent to prepare a 4 mM stock solution (See paragraph 5).

15. This molecular weight approach should apply if the test chemical is a mono-constituent substance with a known molecular weight or a mixture or multi-constituent substance of known composition (See Figure 1). For mixtures and multi-constituent substances of known composition, a single aggregated purity value should be determined by the sum of the proportion of its constituents (excluding water), and a single aggregated molecular weight should be determined by considering the individual molecular weights of each component in the mixture (excluding water) and their individual proportions. The resulting purity and aggregated molecular weight should then be used to calculate the weight of test chemical necessary to prepare a 4 mM solution.

16. Mono-constituent substances of unknown molecular weight should be tested based on a test chemical stock solution at a concentration of 0.5 mg/mL rather than 4 mM (7) (See Figure1 and paragraph 7). Polymers can also be tested at a concentration of 0.5 mg/mL. For mixtures and multi-constituent substances of unknown composition (i.e. UVCB substances of unknown or variable composition, complex reaction products or biological materials), the test solution can be prepared with a gravimetric approach. The substance should then be dissolved in the stock solution at 0.5 mg/mL on the basis of the weight of the total components (excluding solvent) in an appropriate solvent (See paragraph 14 and Figure 1). This 0.5 mg/mL of test chemical concentration corresponds to a molecular weight of 125 g/mol when ADRA (4 mM) is performed. The ADRA gravimetric approach with ADRA (0.5 mg/mL) has been shown to be almost as accurate in prediction as ADRA (4 mM) for 136 chemicals in a wide molecular weight range (30.03 - 512.60) (8) (see paragraph 7). This assessment of the predictive capacity of the gravimetric approach is based on testing chemicals with defined molecular weight and not based on the testing of mixtures, as no reference data for mixtures are available. Therefore, if the mixture to be investigated is known to contain a chemical class with a typical molecular weight which is significantly higher, this default molecular weight and the test solution concentration should be adjusted accordingly [see e.g. approach for agrochemical formulations in (24)]. The gravimetric approach should only be applied as a last resort if no aggregated molecular weight can be calculated. As for any testing with mixtures, as much as possible, information should be gathered on the sensitization potential and reactivity of individual constituents.

Preparation of the positive control, reference controls and co-elution controls

17. Either phenylacetaldehyde (CAS 122-78-1, purity ≥ 90%) or squaric acid diethyl ester (CAS 5231-87-8, purity > 95%) should be used as the positive control (PC) at a concentration of 4 mM in acetonitrile (10). Phenylacetaldehyde is prone to oxidation and polymerisation and integrity of the sample has to be assured by proper storage or by using fresh samples. Squaric acid diethyl ester should be stored protected from high temperature or humidity, since it is prone to hydrolysis. Other suitable positive controls that provide mid-range depletion values may be used if historical data are available to derive comparable run acceptance criteria. In addition, reference controls comprising only NAC or only NAL dissolved in the appropriate solvent should also be included in the HPLC run sequence, and these are used to verify the HPLC system suitability prior to analysis (Reference Control A), the stability of the reference controls over time (Reference Control B), and any effects of the solvent used on depletion of NAC or NAL (Reference Control C) (See Appendix II, Annex 3). The percent NAC and NAL depletion for a test chemical is calculated using an appropriate reference control for that test chemical (see paragraph 23). Also, a

co-elution control comprising only the test chemical should be included in the run sequence to detect possible co-elution of the test chemical with either the NAC or NAL.

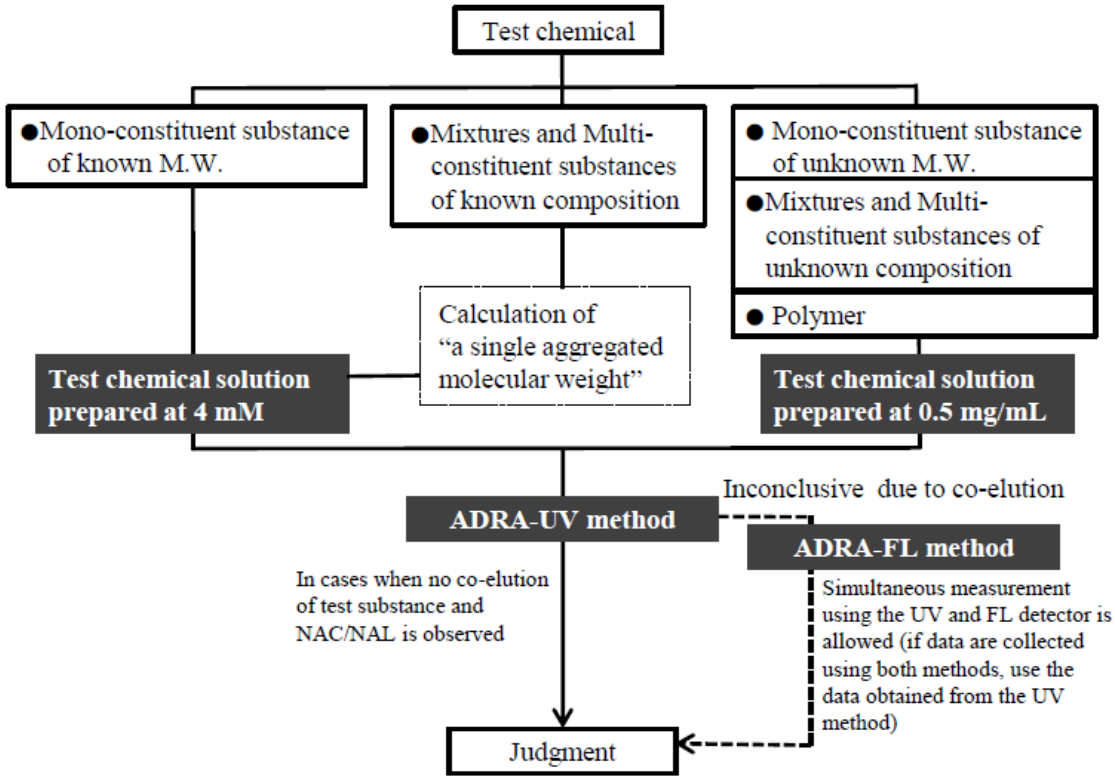
Incubation of the test chemical with the NAC and NAL solutions

18. Both the NAC and the NAL stock solutions are incubated with the test chemical stock solution in a 3:1 ratio in a 96-well microplate. For the 4 mM test chemical stock solution this gives a final concentration of 1 mM test chemical and 5 µM NAC/NAL (17). For the 0.5 mg/ml test chemical stock solution, the final level of the test chemical is 0.125 mg/ml. The observation of precipitate immediately upon addition of the test chemical solution to the NAC and the NAL solutions is an indication of poor solubility, which means that there is no way to know exactly how much test chemical is contained in the solution. Thus, although positive results can be used with confidence, negative results are uncertain and no firm conclusion on the lack of reactivity should be drawn from a negative result (see also paragraph 5 regarding the testing of chemicals not soluble at concentrations as high as 4 mM). The reaction solution should be incubated in the dark at 25±1ºC for 24±1 hours before performing HPLC analysis. After incubation, trifluoroacetic acid (TFA) (≥ 98%) should be added to reaction solution as a fixing solution to stop the reaction (3). 2.5% (v/v) TFA aqueous solution is added to the reaction solution in a 1:4 ratio. Thus, final concentration of NAC/NAL and TFA are 4 µM and 0.5%, respectively.

HPLC preparation and analysis

19. NAC/NAL depletion is monitored by HPLC coupled with an UV-detector. In case of co-elution of NAC/NAL with an UV-absorbing component in the test chemical solution, a fluorescence detector is used (11) (12). There are two options for NAC/NAL detection: Successive measurement should be started with UV-detection and fluorescent detection is used only if inconclusive results due to co-elution are obtained. Alternatively, simultaneous measurement is performed by connecting both the UV and FL detector to the HPLC system for parallel detection. If no co-elution of UV-absorbing components is observed, only the UV data are used. If inconclusive results due to co-elution are observed, FL data will be used (see Figure 1). In the unlikely event that a co-elution also appears in ADRA-FL, the operation should be performed according to paragraph 28. Each test chemical should be analysed in triplicate to determine percent depletion for both NAC and NAL. Although adding the fixing solution does stop the reaction, measurement of the reaction solution is to be performed as soon as possible and in any case within three days after adding the fixing solution. For example, when HPLC analysis of NAC and NAL are performed separately using two 96-well microplates, up to 34 samples may be analysed at one time, including the test chemical, the positive control, and the appropriate number of solvent controls based on the number of individual solvents used in the test, each in triplicate. All of the replicates analysed in a single run should use identical batches of NAC and NAL stock solution. Test chemical and control solutions are to be visually inspected prior to HPLC analysis and may be centrifuged at low speed (100–400 × g) to force any precipitate to the bottom of the vial as a precaution against large amounts of precipitate clogging the HPLC tubing or columns. Observation of precipitation or phase separation after the incubation period is an indication that NAC and NAL depletion could be misleading, and negative results in that case are uncertain and should be interpreted with due care, as well as for any precipitate observed at the beginning of the incubation period (see above).

Figure 1: Procedure to assess NAC/NAL depletion in ADRA including a gravimetric approach for mixtures and alternative fluorescent detection in case of co-elution with UV-absorbing components.



MW, molecular weight; ADRA, amino acid derivative reactivity assay; UV, ultraviolet; FL, fluorescence

20. A standard calibration curve should be generated for both NAC and NAL. Standard solutions of both NAC and NAL should be prepared in 20% acetonitrile in buffer and containing 0.5% trifluoroacetic acid. For NAC, a phosphate buffer at pH 8.0, and for NAL, a phosphate buffer at pH 10.2 should be used. Using the NAC and NAL stock solutions (6.667 µM), six calibration solutions should be prepared in concentrations from 5.0 to 0.156 µM. A blank of the dilution buffer should also be included in the standard calibration curve. Suitable calibration curves should have an R2 > 0.990.

21. The suitability of the HPLC system should be verified before conducting the analysis. Both NAC and NAL depletion is monitored by HPLC coupled with an UV-detector (photodiode array detector or fixed wavelength absorbance detector with 281 nm signal) and a FL detector (Ex, 284 nm and Em, 333 nm) (see paragraph 19). The appropriate column is installed in the HPLC system. The recommended HPLC set-up described in the validated protocol uses a column with the following specifications. Base particle: core-shell type silica gel, Particle size: 2.5~2.7 µm, column size: 3.0 × 150 mm as preferred column. With this reversed-phase HPLC column, the entire system should be equilibrated for at least 30 minutes at 40ºC with 50% phase A (0.1% (v/v) trifluoroacetic acid in water), 50% phase B (0.1% (v/v) trifluoroacetic acid in acetonitrile) before use. Then, the column is conditioned by running the gradient at least twice before actual use. The HPLC analysis should be performed using a flow rate of 0.30 mL/min and a linear gradient from 30% to 55% acetonitrile for NAC and from 25% to 45% acetonitrile for NAL within 10 minutes, followed by a rapid increase to 100% acetonitrile to remove other materials. Equal volumes of the standard solutions, test chemical solutions, and control solutions should be injected. The column should be re-equilibrated under initial conditions for 6.5 minutes between

injections. If a different reversed-phase HPLC column is used, the set-up parameters described above may need to be adjusted to guarantee an appropriate elution and integration of the NAC and NAL, including the injection volume, which may vary according to the system used (typically in the range from 10–20 µL). Importantly, if an alternative HPLC set-up is used, its equivalence to the validated set-up described above should be demonstrated, preferably by testing the proficiency substances in Appendix II, Annex 2. Using the UV detection method, absorbance is monitored at 281 nm. If a photodiode array detector is used, absorbance at 291 nm should also be recorded. It should be noted that some batches of acetonitrile could have a negative impact on NAC and NAL stability and this has to be assessed when a new batch of acetonitrile is used. The ratio of the 281 nm peak area and the 291 nm peak area can be used as an indicator of co-elution. For each sample a ratio in the range of 90% < mean area ratio of control samples < 100% would give a good indication that co-elution has not occurred. An example of HPLC analysis sequence is provided in Appendix II, Annex 3.

22. There are some test chemicals that could potentially promote oxidation of NAC. The peak of the dimerised NAC may be monitored visually in the case of ADRA-UV. However, since the NAC dimer does not exhibit fluorescence, it cannot be detected in the fluorescent detection mode. Any apparent dimerisation should be noted, since overestimation of NAC depletion could result in false-positive predictions (See paragraphs 4, 14 and Appendix II, Annex 1).

DATA AND REPORTING

Data evaluation

23. The concentration of both NAC and NAL is photometrically determined at 281 nm (UV detector) and if needed by fluorescence detection with Ex/Em, 284/333 nm (FL detector) (see paragraph 21) in each sample by measuring the peak area (area under the curve, AUC) of the appropriate peaks and by calculating the concentration of both NAC and NAL using the linear calibration curve derived from the standards.

24. The percent depletion for both NAC and NAL is determined in each sample by measuring the peak area and dividing it by the mean peak area of the relevant Reference Controls C (See Appendix II, Annex 3) according to the formula described below.

$$\text{Percent NAC or NAL depletion} = \left[1-\left[\frac{\text{NAC or NAL peak area in replicate injection}}{\text{Mean NAC or NAL peak area in reference controls C}}\right]\right]\text{x}100$$

Acceptance criteria

25. The following criteria should be met for a run to be considered valid:

a) the standard calibration curve should have an $R^2 > 0.990$,

b) the mean percent NAC and NAL depletion value and the maximum standard deviation (SD) of the three replicates for the positive control (phenylacetaldehyde or squaric acid diethyl ester) should meet the following criteria:

- NAC depletion:

Phenylacetaldehyde: 30 - 80%; Squaric acid diethyl ester: 30 - 80 %

- NAL depletion:

Phenylacetaldehyde: 70 - 100%; Squaric acid diethyl ester: 70 - 100 %

- Maximum standard deviation (SD) for NAC and NAL depletion for both phenylacetaldehyde and squaric acid diethyl ester: < 10%,

c) the mean NAC and NAL concentration of both Reference Controls A and C should be 3.2–4.4 µM and the coefficient of variation (CV) of NAC and NAL peak areas for the nine Reference Controls B and C in acetonitrile should be < 10%.

If one or more of these criteria is not satisfied, the data should be rejected and the run should be repeated for that specific test chemical.

26. The following criteria should be satisfied for a test chemical's results to be accepted as valid:

a) the maximum standard deviation for the test chemical replicates should be < 10% for the percent depletion of both NAC and NAL,

b) the mean NAC and NAL concentration of the three Reference Controls C in the appropriate solvent should be 3.2–4.4 µM. The permissible range of the mean NAC concentration of Reference Control C when 5% DMSO in acetonitrile is used as a solvent is 2.8 to 4.0 µM (19).

If one or more of these criteria is not satisfied, the data should be rejected and the run should be repeated for that specific test chemical.

Prediction model

27. The mean percent depletion of NAC and NAL is calculated for each test chemical. Negative depletion is considered to be "0" when calculating the mean. By using the NAC/NAL prediction model shown in Table 1, the threshold of 4.9% mean depletion should be used to support the discrimination between skin sensitisers and non-sensitiser in the framework of an IATA or a DA. The 4.9% of cut-off value for the mean percent depletion of NAC and NAL was set by using 2 class classification model so that the sensitizer and non-sensitizer could be predicted most appropriately.

Table 1: NAC/NAL prediction model[1]

Mean NAC and NAL percent depletion	ADRA prediction[2]
Less than 4.9%	Negative
4.9% or higher	Positive

[1] The numbers refer to statistically generated threshold values and are not related to the precision of the measurement.

[2] An ADRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 13 and 14.

28. If co-elution is observed using either the UV or the FL detector, the depletion value measured using the detector in which co-elution is not observed should be used (See Figure 1). If co-elution is observed with both detectors, co-elution may be resolved by slightly adjusting the HPLC set-up in order to further separate the elution time of the test chemical and NAC or NAL. If an alternative HPLC set-up is used to try to resolve co-elution, its equivalence to the validated set-up should be demonstrated, preferably by testing the proficiency substances in Appendix II, Annex 2. When co-elution occurs, it is not possible to integrate the peak of the NAC or NAL, thereby preventing calculation of the percent depletion of NAC or NAL. If co-elution of test chemicals occurs with both the NAC and NAL and separation of elution time is not feasible, then the analysis should be reported to be inconclusive. In cases where co-elution occurs only with NAL and separation of elution time is not feasible, the NAC-only prediction model (See Table 2) can be used to make a prediction. In this case, the NAC data of ADRA-UV should still be preferentially adopted

than that of ADRA-FL. The 5.6% cut-off value for the percent depletion of NAC was set by using 2 class classification model so that the sensitizer and non-sensitizer could be predicted most appropriately.

Table 2: NAC-only prediction model[1]

Mean NAC percent depletion	ADRA prediction[2]
Less than 5.6%	Negative
5.6% or higher	Positive

[1] The numbers refer to statistically generated threshold values and are not related to the precision of the measurement.
[2] An ADRA prediction should be considered in the framework of an IATA (13) (14).

29. When a result is unequivocal, a single HPLC analysis for both NAC and NAL should be sufficient for a test chemical. However, in case of results close to the threshold used to discriminate between positive and negative results (i.e. in the range of 3% to 10% for NAC/NAL prediction model or NAC percent depletion falls in the range of 4% to 11% for NAC-only prediction model), additional testing is recommended. In particular, in case of negative results in these ranges (i.e. 3% to 4.9% for NAC/NAL prediction model or 4 % to 5.6% for NAC-only prediction model), a second run should be conducted, as well as a third one in case of discordant results between the first two runs. In the above cases, the majority of the three test results is adopted.

Test report

30. The test report should include the following information:

Test chemical and Controls (positive control and solvent/vehicle)

- For all mono-constituent substance (test and control chemicals)
    - Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers
    - Physicochemical properties such as physical state, appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available
    - Purity, chemical identity of impurities as appropriate and practically feasible, etc.
    - Treatment prior to testing, if applicable (warming, grinding)
    - Concentration(s) tested
    - Storage conditions and stability to the extent available
- Multi-constituent substance, UVCB, and mixtures
    - Characterisation by chemical identity (see above), purity, quantitative occurrence and relevant physicochemical properties (see above) of the constituents, to the extent available
    - Physical appearance, water solubility, and additional relevant physicochemical properties, to the extent available
    - Molecular weight (or apparent molecular weight) for mixtures or polymers of known composition, or other information relevant to the study
    - Treatment prior to testing, if applicable (warming, grinding)
    - Concentration(s) tested

- Storage conditions and stability, to the extent available.

- Additional information for positive control
  - Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.
- Additional information for solvent/vehicle control
  - Solvent used and ratio of its constituents, if applicable
  - Justification for choice of solvent for each test chemical
  - Impact on NAC and NAL stability when using acetonitrile

Preparation of NAC and NAL, positive control and test chemical solution

- Characterisation of NAC and NAL solutions (supplier, lot, exact weight of NAC and NAL, volume added for the stock solution)
- Characterisation of positive control solutions (exact weight of positive control reagent, volume added for the control solution)
- Characterisation of test chemical solutions (exact weight of test chemical, volume added for the test chemical solution)

HPLC instrument setting and analysis

- Type of HPLC instrument, HPLC and guard columns, UV or FL detector, autosampler
- Parameters relevant for the HPLC analysis such as column temperature, injection volumes, flow rate and gradient

System suitability

- NAC and NAL peak area at OD 281 nm (UV detector) or Ex/Em 284/333 nm (FL detector) of each standard and reference control A replicate
- Linear calibration curve graphically represented and the R2 reported
- NAC and NAL concentration of each Reference Control A replicate
- Mean NAC and NAL concentration (µM) of the three reference controls A, SD and CV
- NAC and NAL concentration of Reference Controls A and C.

Analysis sequence

- For Reference Controls
  - NAC and NAL peak area at an OD of 281 nm (UV detector) or an Ex/Em of 284/333 nm (FL detector)of each replicate of Reference Controls B and C
  - Mean NAC and NAL peak area at an OD of 281 nm (UV detector) or an Ex/Em of 284/333 nm (FL detector) of the nine Reference Controls B and C in acetonitrile, SD and CV (for stability of reference controls over analysis time)
  - For each solvent used, the mean NAC and NAL peak area at an OD of 281 nm (UV detector) or an Ex/Em of 284/333 nm (FL detector) of the three appropriate Reference Controls C (for the calculation of percent NAC and NAL depletion)
  - For each solvent used, the NAC and NAL concentration (µM) of the three appropriate Reference Controls C

- For each solvent used, the mean NAC and NAL concentration (µM) of the three appropriate Reference Controls C, SD and CV.

- For positive controls
  - NAC and NAL peak area at an OD of 281 nm (UV detector) or an Ex/Em of 284/333 nm (FL detector) of each replicate
  - Percent NAC and NAL depletion of each replicate
  - Mean percent NAC and NAL depletion of the three replicates, SD and CV.
- For each test chemical
  - Appearance of precipitate in the reaction mixture at the end of the incubation time, if observed. If precipitate was re-solubilised or centrifuged;
  - Presence of co-elution
  - Description of any other relevant observations, if applicable
  - NAC and NAL peak area at an OD of 281 nm (UV detector) or an Ex/Em of 284/333 nm (FL detector) of each replicate
  - Percent NAC and NAL depletion of each replicate
  - Mean of percent NAC and NAL depletion of the three replicate, SD and CV
  - Mean of percent NAC and percent NAL depletion values
  - Prediction model used and ADRA prediction

Proficiency testing

- Statement that the testing facility has demonstrated proficiency in the use of the test method before routine use by testing of the proficiency chemicals

Discussion of the results

- Description of any unintended modifications to the test procedure.
- Discussion of the results obtained with the ADRA test method and if it is within the ranges described in paragraph 29.

Conclusion

Literature for Appendix II

(1) Fujita M, Yamamoto Y, Tahara H, Kasahara T, Jimbo Y and Hioki T (2014), Development of a prediction method for skin sensitisation using novel cysteine and lysine derivatives, Journal of Pharmacological and Toxicological Methods, 70:94-105. DOI: 10.1016/j.vascn.2014.06.001.

(2) Yamamoto Y, Tahara H, Usami R, Kasahara T, Jimbo Y, Hioki T and Fujita M (2015), A novel *in chemico* method to detect skin sensitisers in highly diluted reaction conditions, Journal of Applied Toxicology, 35:1348-60. DOI: 10.1002/jat.3139.

(3) Fujita M, Yamamoto Y, Watanabe S, Sugawara T, Wakabayashi K, Tahara K, Horie N, Fujimoto K, Kusakari K, Kurokawa Y, Kawakami T, Kojima K, Kojima H, Ono A, Katsuoka Y, Tanabe H, Yokoyama H and Kasahara T (2019), Cause of and Countermeasures for Oxidation of the Cysteine-Derived Reagent Used in the Amino acid Derivative Reactivity Assay, Journal of Applied Toxicology, 39:191-208. DOI: 10.1002/jat.3707.

(4) Fujita M, Yamamoto Y, Watanabe S, Sugawara T, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Kusakari K, Kurokawa Y, Kawakami T, Kojima K, Sozu T, Nakayama T, Kusao T, Richmond J, Kleinstreuer N, Kim BH, Kojima H, Kasahara T and Ono A (2019), The within- and between-laboratory reproducibility and predictive capacity of the *in chemico* Amino acid Derivative Reactivity Assay (ADRA): Results of validation study implemented in four participating laboratories, Journal of Applied Toxicology, 39: 1492-1505. DOI: 10.1002/jat.3834.

(5) OECD (2022), Updated Validation report: Amino acid Derivative Reactivity Assay (ADRA) – JaCVAM Validation Study Report. Series on testing and Assessment n° 304. Organisation for Economic Cooperation and Development, Paris.

(6) OECD (2019), Amino acid Derivative Reactivity Assay (ADRA) – Report of the Peer Review Panel. Series on testing and Assessment n° 305. Organisation for Economic Cooperation and Development, Paris.

(7) Yamamoto Y, Fujita M, Wanibuchi S, Katsuoka Y, Ono A and Kasahara T (2019), Expanding the applicability of the Amino acid Derivative Reactivity Assay (ADRA): Determining a weight concentration for preparation of test chemical solutions that yields a predictive capacity identical to the conventional method using molar concentration and demonstrating the capacity to detect sensitizers in liquid mixtures. Journal of Pharmacological and Toxicological Methods. 97: 67-79. DOI: 10.1016/j.vascn.2019.01.001

(8) Imamura M, Wanibuchi S, Yamamoto Y, Kojima H, Ono A, Kasahara T and Fujita M (2021), Improving predictive capacity of the Amino acid Derivative Reactivity Assay test method for skin sensitisation potential with an optimal molar concentration of test chemical solution. Journal of Applied Toxicology, 41: 303-329. DOI: 10.1002/jat.4082.

(9) Yamamoto Y, Watanabe S, Yamaga H, Yoshida K, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kojima K, Kawakami T, Sozu T, Wanibuchi S, Fujita M. Kasahara T, Ono, A and Kojima H (2022), Within- and between-laboratory reproducibility and predictive capacity of amino acid derivative reactivity assay (ADRA) using a 0.5 mg/ml test chemical solution: Results of the study for reproducibility confirmation implemented in five participating laboratories. Journal of Applied Toxicology, DOI: 10.1002/jat.4279. Epub ahead of print.

(10) Fujita M, Yamamoto Y, Watanabe S, Yamaga H, Yoshida K, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kojima K, Kawakami T, Sozu T, Wanibuchi S, Kasahara T, Ono, A and Kojima H (2022), The within- and between-laboratories reproducibility and predictive capacity of Amino acid Derivative Reactivity Assay using 4 mM test chemical solution: Results of ring study implemented at five participating laboratories. Journal of Applied Toxicology, 42: 318-333. DOI:

10.1002/jat.4268.

(11) Fujita M, Yamamoto Y, Wanibuchi S, Katsuoka Y and Kasahara T (2019), A Newly Developed Means of HPLC-Fluorescence Analysis for Predicting the Skin Sensitisation Potential of Multi-Constituent Substances Using ADRA. Toxicology In Vitro, 59, 161-178. DOI: 10.1016/j.tiv.2019.04.014.

(12) Wanibuchi S, Yamamoto Y, Sato A, Kasahara T and Fujita M (2019), The amino acid derivative reactivity assay with fluorescence detection and its application to multi-constituent substances. Journal of Toxicological Sciences, 44: 821-832. DOI: 10.2131/jts.44.821.

(13) Yamamoto Y, Fujita M, Wanibuchi S, Sato A, Akimoto M, Katsuoka Y, Ono A and Kasahara T (2019), Applicability of amino acid derivative reactivity assay for prediction of skin sensitisation by combining multiple alternative methods to evaluate key events. Journal of Toxicological Sciences, 44: 585-600. DOI: 10.2131/jts.44.585.

(14) Imamura M, Yamamoto Y, Fujita M, Wanibuchi S, Nakashima N, Kojima H, Ono A and Kasahara T (2022), Applicability of ADRA (4mM) for prediction of skin sensitisation by combining multiple alternative methods to evaluate Key events. Journal of Applied Toxicology, DOI: 10.1002/jat.4283. Epub ahead of print.

(15) Fujita M, Yamamoto Y, Wanibuchi S, Katsuoka Y and Kasahara T (2019), The underlying factors that explain why nucleophilic reagents rarely co-elute with test chemicals in the ADRA. Journal of Pharmacological and Toxicological Methods. 96: 95-105. DOI: 10.1016/j.vascn.2019.02.004.

(16) Yamamoto Y, Wanibuchi S, Sato A, Kasahara T and Fujita M (2019), Precipitation of test chemicals in reaction solutions used in the amino acid derivative reactivity assay and the direct peptide reactivity assay. Journal of Pharmacological and Toxicological Methods, 100: 106624. DOI: 10.1016/j.vascn.2019.106624.

(17) ADRA protocol: JaCVAM Statements. Available at: http://www.jacvam.jp/en_effort/effort02.html

(18) Tam JP, Wu CR, Liu W and Zhang JW (1991), Disulfide bond formation in peptides by dimethyl sulfoxide. Scope and applications. Journal of the American Chemical Society, 113, 6657–6662. https://pubs.acs.org/doi/pdf/10.1021/ja00017a044

(19) Akimoto M, Yamamoto Y, Watanabe S, Yamaga H, Yoshida K, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Kusakari K, Kamiya K, Kojima K, Kawakami T, Kojima H, Ono A, Kasahara T and Fujita M (2020), Oxidation of a cysteine-derived nucleophilic reagent by dimethyl sulfoxide in the amino acid derivative reactivity assay. J Appl Toxicol., 40: 843-854. DOI: 10.1002/jat.3948.

(20) Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F and Kern P (2013), A dataset on 145 chemicals tested in alternative assays for skin sensitisation undergoing prevalidation. Journal of Applied Toxicology, 33: 1337-52, DOI:10.1002/jat.2868.

(21) Gerberick GF, Vassalo JD, Foertsch LM, Price BB, Chaney JG and Lepoittevin J-P (2007), Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach, Toxicological Sciences, 97: 417-427. DOI: 10.1093/toxsci/kfm064.

(22) Basketter DA, Scholes EW (1992), Comparison of the local lymph node assay with the guinea-pig maximization test for the detection of a range of contact allergens, Food and Chemical Toxicology, 30: 65-69.

(23) ECETOC (2003), Contact sensitisation: Classification according to potency. European Centre for Ecotoxicology & Toxicology of Chemicals (Technical Report No. 87).

(24) Raja S. Settivari, Sean C. Gehen, Ricardo Acosta Amado, Nicolo R. Visconti, Darrell R. Boverhof, Edward W. Carney, (2015) Application of the KeratinoSens™ assay for assessing the skin sensitization potential of agrochemical active ingredients and formulations, Regulatory Toxicology and

Pharmacology, 72, 350-360.

APPENDIX II, ANNEX 1

Known limitations of the Amino acid Derivative Reactivity Assay (ADRA)

The table below provides a summary of the known limitations of the ADRA.

Substance class / interference	Reason for potential underprediction or interference	Data interpretation	Example substance
Metals and inorganic compounds	Known to react with proteins via mechanisms other than covalent binding	Should not be tested	Nickel sulphate; 7786-81-4
Pro-haptens	Test Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential cannot be detected by the test method unless activation is caused by auto-oxidation to a similar degree as in vivo /in humans. It will however normally not be known whether this will be the case	May lead to false negatives. Negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA	Diethylenetriamine; 111-40-0 (human 1A, LLNA n/a)
Pre-haptens	Chemicals that become sensitisers after abiotic transformation are reported to be in some cases correctly detected by the test method		Linalool: 78-70-6
Test chemicals that have a UV absorption (OD, 281 nm) or FL (Ex/Em, 284/333 nm) and have the same retention time than NAC or NAL (co-elution)	When co-elution occurs the peak of the NAC or NAL cannot be integrated and the calculation of the percent NAC or NAL depletion is not possible.	The substances that absorb UV in this range of the spectrum are generally limited to those having conjugated double bonds, which significantly lowers the potential for co-elution. The substances that have a FL in this range are generally limited to polyaromatic or polyheterocyclic compounds, including naphthalene derivatives. If co-elution of such test chemicals occurs with both the NAC and the NAL or with the NAC only, then the analysis should be reported as "inconclusive" and alternative HPLC set up should be considered (see paragraph 28). In cases where co-elution occurs only with the NAL, then the NAC-only prediction model reported in Table 2 can be used."	Safranal; 116-26-7
Complex mixtures of unknown composition, substances of unknown or variable composition, complex reaction products or biological materials	ADRA using a 4 mM chemical solution needs for defined molar ratio of test chemical and nucleophilic reagent, but ADRA using a 0.5 mg/mL solution does not need the defined molar ratio of a test chemical and can predict sensitisation for test chemicals, which are prepared at a weight concentration of 0.5 mg/mL. When the mixture is a liquid, the evaluation of sensitisation using ADRA cannot be performed if the total weight of the mixture components dissolved in solvent (water, dissolving solution, extraction solvent, etc) is not known, since it is then impossible to prepare a 0.5 mg/mL test chemical solution.	Since plant extract contains various polyphenols, which react with NAC, it may be judged as a sensitiser when a solution containing a high concentration of the plant extract is evaluated using ADRA. Therefore, these results should be considered with reference to results obtained using alternative methods for other key events and *in vivo* results of similar substances.	n/a
Test chemicals which cannot be dissolved in an appropriate solvent	Not sure if sufficient exposure can be achieved	The ADRA test method allows testing of poorly soluble chemicals. Test chemicals that are not soluble at this	n/a

at a final concentration of 4 mM	If the mixture is liquid and the total weight of the mixed components dissolved in a solvent (e.g., water, dissolving solution, extraction solvent) is not known, it is not possible to prepare a 0.5 mg/mL test substance solution, and thus the sensitisation potential cannot be evaluated by ADRA.	concentration though may still be tested at lower soluble concentrations. In such a case, a positive result could be used to support the identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result.	
Chemicals which precipitate in reaction solution	Not sure if sufficient exposure can be achieved	Test chemicals that precipitate in the reaction solution even if dissolved in the solvent may still be tested at lower soluble concentrations. In such a case, a positive result could still be used to support the identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result.	Isopropyl myristate CAS: 110-27-0
Test chemicals that do not covalently bind to the NAC but promote its -oxidation (i.e. NAC dimerisation)	Could lead to a potential over-estimation of NAC depletion, resulting in possible false positive predictions.	It may be possible to detect and quantify any NAC dimer formed by HPLC (UV detector), thus confirming or ruling out that the NAC reagent has been depleted via oxidative dimerisation as opposed to reaction and covalent bonding to the test item substance(s) Therefore, ADRA may prevent erroneous judgement due to the oxidizing action of the test chemical. However, since the NAC dimer does not have fluorescence, it can only be detected by ADRA-UV.	DMSO Oxidant
Test chemicals that are only soluble in DMSO	DMSO causes excessive NAC depletion due to NAC dimerization resulting in high background NAC depletion.	DMSO is allowed to be contained in the test chemical solution up to 5%. If DMSO is chosen, attempts should be made to solubilise the test chemical in a 1:20 mixture of DMSO and acetonitrile (5% DMSO in acetonitrile).	n/a

APPENDIX II, ANNEX 2

Proficiency Substances

In Chemico Skin Sensitisation: Amino acid Derivative Reactivity Assay (ADRA)

Prior to routine use of the test method, laboratories should demonstrate technical proficiency by correctly obtaining the expected ADRA prediction for the 10 proficiency substances recommended in Table 1 and by obtaining NAC and NAL depletion values that fall within the respective reference ranges for 8 out of the 10 proficiency substances. The test to demonstrate technical proficiency in ADRA is basically ADRA with 4 mM (10). If ADRA with 4 mM has been proven to be mastered by performing proficiency substances, ADRA with 0.5 mg/mL can be exempt from demonstrating the technical proficiency (9). These proficiency substances were selected to represent the full range of responses for skin sensitisation hazards. Other selection criteria were that they are commercially available, that high quality *in vivo* reference data and high quality ADRA data are available, and that they were used during the JaCVAM-coordinated validation study to demonstrate successful implementation.

Table 1. Recommended chemicals for demonstrating technical proficiency with ADRA_4 mM

No.	Test chemicals	CAS No.	Physical state	Molecular weight	*In vivo* Prediction[1]	ADRA 4 mM prediction[2]	Range of % depletion	
							NAC[3]	NAL[3]
1	*p*-Benzoquinone	106-51-4	Solid	108.09	Sensitiser (extreme)	Positive	90-100	70-100
2	Diphenylcyclopropenone	4886-38-4	Solid	206.24	Sensitiser (strong)	Positive	50-90	≤ 10
3	2-Methyl-2H-isothiazol-3-one	2682-20-4	Solid	115.15	Sensitiser (moderate)	Positive	80-100	≤10
4	Palmitoyl Chloride	112-67-4	Liquid	274.87	Sensitiser (moderate)	Positive	≤ 40	70-100
5	Imidazolidinyl urea	39236-46-9	Solid	388.29	Sensitiser (weak)	Positive	40-70	≤ 20
6	Farnesal	19317-11-4	Liquid	220.35	Sensitiser (weak)	Positive	60-100	5-40
7	Glycerol	56-81-5	Liquid	92.09	Non-sensitiser	Negative	≤ 7	≤ 7
8	Isopropanol	67-63-0	Liquid	60.10	Non-sensitiser	Negative	≤ 7	≤ 7
9	Dimethyl isophthalate	1459-93-4	Solid	194.19	Non-sensitiser	Negative	≤ 7	≤ 7
10	Propyl paraben	94-13-3	Solid	180.20	Non-sensitiser	Negative	≤ 7	≤ 7

[1]The *in vivo* hazard (and potency) predictions are based on LLNA data (20) (21) (22). The *in vivo* potency is derived using the criteria proposed by ECETOC (23).

[2] An ADRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 3 and 5.

[3] Ranges determined on the basis of at least 10 depletion values generated by 5 independent laboratories.

APPENDIX II, ANNEX 3

EXAMPLES OF ANALYSIS SEQUENCE

Each sample of HPLC analysis should be analysed in number order below. Refer to the table showing Examples of HPLC Sample Analysis Sequences for more practical sequences about HPLC analysis.

1. Start to analyse calibration standards and Reference Control A (N = 3).

2. The co-elution Control does not need to be analysed by turns if it is analysed after analysis of standard solution and Reference Control A.

3. Reference Control B should be analysed three times (total six times) before and after the analysis of sample, Reference Control C and Positive Control.

4. The Reference Control C, Positive Control and Test chemical solutions are analysed. (After the first set of replicates of each sample is analysed, the second set of replicates of each should be analysed).

Calibration standards and reference controls	STD1 STD2 STD3 STD4 STD5 STD6 Dilution buffer Reference control A, rep 1 Reference control A, rep 2 Reference control A, rep 3
Co-elution controls	Co-elution control 1 for test chemical 1 Co-elution control 2 for test chemical 2
Reference controls	Reference control B, rep 1 Reference control B, rep 2 Reference control B, rep 3
First set of replicates	Reference control C, rep 1 Positive control, rep 1 Sample 1, rep 1 Sample 2, rep 1
Second set of replicates	Reference control C, rep 2 Positive control, rep 2 Sample 1, rep 2 Sample 2, rep 2
Third set of replicates	Reference control C, rep 3 Positive control, rep 3 Sample 1, rep 3 Sample 2, rep 3
Reference controls	Reference control B, rep 4 Reference control B, rep 5 Reference control B, rep 6

Three sets of reference controls (NAC or NAL dissolved in the appropriate solvent) should be included in the analysis sequence:

Reference control A: Control for verifying validity of the HPLC system. Reference Control A is used to verify concentration of NAC and NAL from each calibration curve after addition of acetonitrile rather than test chemical.

Reference control B: Control for verifying stability of reaction solution under analysis. Reference Control B is used to verify variability (CV) of each three NAC/NAL peak areas in the solution after addition of acetonitrile rather than test chemical at the start of analysis and at the end of analysis.

Reference control C: Control for calculating NAC/NAL depletion of each test chemical solution. To calculate depletion of NAC/NAL, measure three Reference Controls C after addition of solvent instead of test chemical. Prepare reference Control C for all solvents used to dissolve the test chemicals.

APPENDIX III

In Chemico Skin Sensitisation: kinetic Direct Peptide Reactivity Assay (kDPRA)

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

1. The kDPRA is proposed to address the molecular initiating event of the skin sensitisation AOP - namely, protein reactivity - by quantifying the reactivity of test chemicals towards a synthetic model peptide containing cysteine in a time- and concentration dependent manner (1) (2). Kinetic rate constants are calculated and the logarithm of the maximum rate constant (log k_{max} value in $s^{-1}M^{-1}$) for a tested substance is then used to support the discrimination of UN GHS subcategory 1A skin sensitisers (subcategory 1A) from those not categorised as subcategory 1A (non-subcategory 1A) i.e., subcategory 1B or no category according to UN GHS (3). Based on theoretical consideration, the rate constant of the reaction between a test chemical and skin proteins will determine the amount of epitope formed from a given amount of chemical or, vice-versa, determine the dose needed to form the amount of epitope needed for induction of sensitization to occur and it is thus a rate limiting and potency determining step. Based on empirical evidence when evaluating 180 chemicals, the rate constant was shown to be the strongest determinant of potency among all evaluated parameters measured in OECD 442C, 442D and 442E (3).

2. The kDPRA proved to be transferable to laboratories without hands-on training (4). For the 24 test chemicals tested during the validation study, the overall within-laboratory reproducibility of kDPRA for assigning UN GHS subcategory 1A was 96% and the average between-laboratory reproducibility was 88% (4). Results from the validation study (4) as well as from other published studies (3) encompassing 180 test chemicals that fall within kDPRA's applicability domain indicate that kDPRA allows to discriminate UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory 1A) according to UN GHS with a balanced accuracy of 85%, a sensitivity of 84% (38/45), and a specificity of 86% (116/135) relative to LLNA results (3). Similar performances were obtained when comparing kDPRA outcomes with the OECD LLNA database compiled within the context of the Test Guideline on Defined Approaches for Skin Sensitization (15)[4]. In addition, the prediction for 123 test chemicals (out of the 180) having human skin sensitisation data (5) (6) has a balanced accuracy of 76%, a sensitivity of 64% (21/33), and a specificity of 89% (80/90) (3), although the human reference data are subject to a significant uncertainty[5]. Furthermore, when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the

[4] A balanced accuracy of 85%, a sensitivity of 82% (31/38), and a specificity of 88% (102/116) were found relative to LLNA dataset compiled within the context of the Test Guideline on Defined Approaches for Skin Sensitization (15).

[5] A balanced accuracy of 67%, a sensitivity of 53% (9/17), and a specificity of 81% (25/31) were found relative to human skin sensitisation dataset compiled within the context of the Test Guideline on Defined Approaches for Skin Sensitization (15).

LLNA test as well as other animal tests may not fully reflect the situation in the species of interest, which is humans. For comparison, based on a data set of 123 chemicals used to evaluate the kDPRA vs. human sensitising potential, the LLNA showed a 73% balanced accuracy, a 55% (18/33) sensitivity and a 91% (82/90) specificity for the identification of UN GHS subcategory 1A. On the basis of the overall data available, kDPRA's applicability domain was shown to include a variety of organic functional groups, reaction mechanisms, skin sensitisation potencies (as determined in *in vivo* studies), and physicochemical properties (3). Following an independent peer review (16), the kDPRA was considered to be scientifically valid to discriminate UN GHS subcategory 1A skin sensitisers from those not categorised as 1A (non-subcategory 1A) according to UN GHS (7). The kDPRA can therefore be used (i) as a follow-up test method for sub-categorisation of chemicals identified as UN GHS Category 1 skin sensitisers, or (ii) on its own by using positive results for direct classification of a chemical into UN GHS subcategory 1A, depending on the regulatory framework.

3. The term "test chemical" is used in this Test Guideline to refer to what is being tested and is not related to the applicability of the kDPRA to the testing of substances and/or mixtures. This test method is not applicable to the testing of metal compounds, which are known to react with proteins via mechanisms other than covalent binding. Furthermore, kDPRA only measures reactivity with the cysteine peptide, so that strong sensitisers having an exclusive lysine-reactivity, such as some acyl-halides, phenol-esters or aldehydes are outside of the applicability domain of kDPRA. However, only few UN GHS subcategory 1A skin sensitisers are known currently to react exclusively with lysine residues. In addition, considering exclusive strong Lysine-reactivity from the DPRA or ADRA in a tiered strategy may reduce this uncertainty. Test chemicals that do not covalently bind to the peptide but promote its oxidation (i.e. cysteine dimerisation) could lead to a potential over estimation of peptide depletion, resulting in possible false positive predictions and/or assignment to a higher reactivity class. The test method described in this Appendix of the Test Guideline is an *in chemico* method that does not encompass a metabolic system. Reactivity of chemicals that require enzymatic bioactivation to exert their skin sensitisation potential (i.e. pro-haptens) cannot be reliably detected by the test method. However, the limitation for detecting pro-haptens was found to be less pronounced when identifying strong sensitisers as compared to the identification of weak sensitisers (3). The majority of chemicals that become sensitisers after abiotic transformation (i.e. pre-haptens) were reported to be correctly detected by *in chemico* test methods (8) (9). However, spontaneously rapidly oxidizing pre-haptens may be under-predicted by kDPRA (as in any *in vitro* skin sensitisation assay) due to a lag-phase for oxidation which reduces the overall reaction rate. In the light of the above, results obtained with the test method that do not lead to subcategory1A categorisation should be interpreted in the context of the currently known limitations (see also Annex 1 of this Appendix), i.e.:

   - aromatic amines, catechols or hydroquinones may require further data to confirm their weak reactivity even under oxidizing conditions, and
   - acyl-halides, phenol-esters or aldehydes specifically reacting with Lysine-residue according to e.g. the DPRA or ADRA, may require further data to confirm their weak reactivity.

4. To be tested, a test chemical should be soluble in an appropriate solvent at a final concentration of 20 mM (see paragraphs 12-13). Test chemicals that are not soluble at this concentration may still be tested at lower concentrations as long as a k_{max} value (i.e., the maximum rate constant (in $s^{-1}M^{-1}$) determined from the reaction kinetics for a tested substance in the kDPRA (see paragraph 24)), can be derived. In such a case, a positive result leading to a UN GHS subcategory 1A skin sensitization prediction (i.e. log $k_{max} \geq -2.0$) could still be used, but no firm conclusion should be drawn from a negative result (i.e., non-reactive or log $k_{max} < -2.0$ outcome).

5. The kDPRA uses a fluorescence readout which requires attention for potential test chemical autofluorescence, fluorescence quenching or interaction with the reagent (monobromobimane). In particular, it is important to include the respective test chemical controls as described in paragraph 16 and to assess the incubation time dependence of the determined peptide depletion. Furthermore, test chemicals with primary SH-group (thiols) cannot be tested with the kDPRA as the thiol group can interact with the monobromobimane (see paragraph 8) leading to enhanced fluorescence. Finally, chemicals decomposing under the conditions of the assay (neutral, aqueous conditions) and releasing a free SH-group will be prone to the same limitations.

6. The kDPRA is considered to be technically applicable to the testing of multi-constituent substances and mixtures of known composition, although such substances were not tested during the validation studies. In this case, a single purity may be determined by the sum of the proportion of its constituents (excluding water), and a single apparent molecular weight may be determined by considering the individual molecular weights of each component in the mixture (excluding water) and their individual proportions. The resulting purity and apparent molecular weight can then be used to calculate the weight of test chemical necessary to prepare a 20 mM solution. Results obtained with mixtures and multi-constituent substances of known composition can lead to a non-linear behaviour, so that the provisions described in paragraph 27(ii) should be used. Regarding mixtures and substances of unknown or variable composition, complex reaction products or biological materials (i.e. UVCB substances), the current model cannot be used due to the need for defined molar ratios. In any case, when considering testing of mixtures, difficult-to-test chemicals (e.g. unstable), or test chemicals not clearly within the applicability domain described in this Guideline, upfront consideration should be given to whether the results of such testing will yield results that are meaningful scientifically. Finally, in cases where evidence can be demonstrated on the non-applicability of the test method to specific categories of chemicals, the test method should not be used for those specific categories of chemicals.

7. The kDPRA can be used for the discrimination of UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory1A) according to UN GHS (3). As for any key-event based test method, the performance of kDPRA will have to be further assessed when used in combination with other assays such as DPRA or ADRA, and within integrated approaches such as IATA or DA for a more comprehensive analysis of skin sensitisation (3) (10).

PRINCIPLE OF THE TEST

8. The kDPRA is a modification of the *in chemico* test method DPRA (described in Appendix I of this Test Guideline). The kDPRA uses the cysteine peptide (Ac-RFAA**C**AA-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 and in the DPRA. While the DPRA measures only at one concentration of the test chemical (5 mM for the cysteine peptide) and at one time point (≥ 24 h), the kDPRA performs parallel reactions at five concentrations (5, 2.5, 1.25, 0.625 and 0.3125 mM) and at six time-points (10, 30, 90, 150, 210 and 1440 min) at 25±2.5ºC. Residual concentration of the cysteine peptide after the respective reaction time is measured after stopping the reaction by the addition of monobromobimane (mBrB; CAS 74235-78-2). The highly reactive and non-fluorescent mBrB rapidly reacts with unbound cysteine moieties of the model peptide to form a fluorescent complex which is measured in order to quantify the non-depleted peptide concentration. If the depletion of the highest concentration surpasses the threshold of 13.89% (cut-off used in the DPRA for positivity in the cysteine only prediction model) and this depletion is statistically significant vs. controls with peptide

only, further calculations are performed (otherwise the test chemical is considered to be non-reactive according to the prediction model shown in paragraph 28). 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 coefficient > 0.90), 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 calculated. The maximum value observed at any time point is taken as the log k_{max}, and this maximum rate constant is the primary read-out of the test. It gives a quantification of the maximum kinetic rate of the reaction of the test chemical with the test peptide. Kinetic reaction rates of the cysteine peptide depletion are then used to discriminate UN GHS subcategory 1A skin sensitisers from those not categorised as 1A (non-subcategory 1A) according to UN GHS. Chemicals with a log $k_{max} \geq -2.0$ are predicted as UN GHS subcategory 1A. The kinetic rate constant may be further used in integrated approaches such as IATA or DA to assess the skin sensitisation potency of a test chemical in a continuous scale as needed for risk assessment (3) (10).

9. Prior to routine use of this test method, laboratories should demonstrate technical proficiency, using the nine proficiency substances listed in Annex 2 of this Appendix.

PROCEDURE

10. This test method is based on the kDPRA DB-ALM protocol no 217 (11) which represents the protocol used for the industry-coordinated validation study. It is recommended that this protocol is used when implementing and using the method in a laboratory. The main components and procedures for the kDPRA are described below.

Preparation of the cysteine-peptide

11. The stock solution of the cysteine containing synthetic peptide (Ac-RFAA**C**AA-COOH) of purity equal to or higher than 95% should be freshly prepared just before the incubation with the test chemical. The final concentration of the cysteine peptide should be 0.667 mM in pH 7.5 phosphate buffer for test chemical soluble in acetonitrile and 1.0 mM for chemicals soluble in pH 7.5 phosphate buffer.

Preparation of the test chemical

12. Solubility of the test chemical in an appropriate vehicle should be assessed before performing the assay. A non-reactive, water-miscible vehicle able to completely dissolve the test chemical should be used. Solubility is checked by visual inspection where the forming of a clear solution is considered sufficient to ascertain that the test chemical is dissolved. The preferred vehicle is acetonitrile. When a substance is not soluble in acetonitrile, solubilisation in pH 7.5 phosphate buffer should be assessed. Further vehicles have not been tested yet but may be used if it is demonstrated that the vehicle does not interfere with the assay, e.g. all controls should be prepared using the same vehicle, and the reaction rates obtained for the positive control and for the proficiency chemicals should fall within the ranges described in paragraph 26 and Annex 2 of this Appendix, respectively. It is important to note that use of DMSO as a vehicle should be avoided as it may lead to peptide dimerisation.

13. The test chemical should be pre-weighed into glass vials and dissolved immediately before testing to prepare a 20 mM solution using the appropriate vehicle as described in paragraph 12. Test chemical dilutions are prepared by serial dilution to obtain concentrations of 20, 10, 5, 2.5 and 1.25 mM.

Preparation of controls

14. Cinnamic aldehyde (CAS 104-55-2; ≥95% food-grade purity) should be used as positive control (PC). It is dissolved at a concentration of 20 mM in acetonitrile immediately before testing. Serial dilutions are then prepared to obtain PC concentrations of 20, 10, 5, 2.5 and 1.25 mM. Use of other positive controls is not recommended since in this assay an exact reaction rate is measured and consistent use of the positive control allows quantitative comparison between laboratories, with validation study data and as intra-laboratory historical control.

15. A vehicle control (VC), considered as the negative control, includes the peptide dissolved in buffer and vehicle respectively but no test chemical nor PC. The peptide-depletion of test chemical or PC incubated samples is calculated relative to the respective VC.

16. The assay also includes test chemical controls at the respective test chemical concentration in the vehicle and buffer but without peptide. This set of controls is used for the identification of interference of the test chemical with the fluorescence measurement (autofluorescence and quenching) to assess e.g., interference with monobromobimane and as a background measurement.

17. A blank control (BC) is used as a background measurement and is prepared with vehicle and buffer but without test chemical, PC, or peptide.

Incubation of the test chemical with the cysteine peptide solution

18. Serial dilutions of the test chemical and PC are prepared in a 96-well microtiter plate referred to as the application plate. Further, a 96-well black assay plate for each exposure time is prepared, referred to as the assay plates, by adding the relevant reagents (i.e., peptide stock solution, vehicle and buffer solution) according to a predefined plate layout such as recommended within the kDPRA protocol (11). Each test chemical concentration should be analysed in triplicate. The reaction is started by adding the test chemical and PC dilutions from the application plates to the assay plates. If a precipitate is observed immediately upon addition of the test chemical solution to the peptide solution, due to low aqueous solubility of the test chemical, one cannot be sure how much test chemical remained in the solution to react with the peptide. In such a case, a positive result (i.e. log k_{max} ≥ -2.0) could still be used, but a negative result (i.e., non-reactive or log k_{max} < -2.0 outcome) should be interpreted with due care (see also provisions in paragraph 4 for the testing of chemicals not soluble up to a concentration of 20 mM in the kDPRA). After adding the test chemical and PC, plates are sealed with gas-tight adhesive foil and shaken at least 200 rpm for 5 min. Assay plates solution should be incubated in the dark at 25 ± 2.5° C for several incubation (exposure) times, i.e. 10, 30, 90, 150, 210, and 1440 min before addition of mBrB solution. Incubation times may be adapted to investigate the most relevant time points for a specific chemical (e.g., shorter incubation times might be more suitable for fast reacting chemicals). However, 1440 min should always be tested, as it corresponds to the incubation time of the DPRA. The incubation (exposure) time is the time interval from the application of the test chemical and PC dilutions to the assay plate until the addition of mBrB.

Fluorescence measurement

19. When the desired incubation (exposure) time is reached, freshly prepared mBrB solution (3 mM in acetonitrile) is added rapidly to the wells of the assay plates (one per exposure time) in the dark. Plates are sealed with gas-tight adhesive foil and shaken at least 200 rpm for 5 min. Fluorescence intensity is then determined using an excitation filter of 390 nm and an emission filter of 480 nm.

DATA AND REPORTING

Data evaluation

20. An automated Excel-evaluation spreadsheet is available with the DB-ALM protocol and should be used for data evaluation. Detailed instructions are provided in the DB-ALM protocol no. 217 (11).

21. For each incubation (exposure) time 't' the following parameters are calculated:

- The arithmetic mean and standard deviation of the fluorescence intensity of the 12 blank controls (BC);
- The arithmetic mean and standard deviation of the fluorescence intensity of the 12 vehicle controls (VC);
- The mean BC value is subtracted from the VCs to obtain corrected VC values.
- For each test chemical and PC concentration, the respective test chemical control value is subtracted from their obtained values to calculate corrected test chemical or PC values.

22. To determine the relative peptide depletion in % for each test chemical concentration per exposure time, the following calculation is performed:

$$relative\ peptide\ depletion\ [\%] \quad = \left[1 - \left(\frac{corrected\ test\ chemical\ or\ PC\ value}{mean\ of\ corrected\ VC}\right)\right] x100\%$$

23. For each test chemical concentration, the arithmetic mean and standard deviation of the three replicates is calculated (per exposure time). A student's t-test is performed to test whether the peptide concentrations measured in the three replicates is statistically significantly lower as compared to the concentration in the 12 VC wells.

24. In the kDPRA, reaction kinetic rate constants are determined as explained below if (i) a peptide depletion of ≥ 13.89% is observed at the highest test chemical concentration (final test chemical concentration 5 mM) at a given time and if (ii) the difference is statistically different from the VC. This 'positivity criterion' is based on the 'positive' criterion for peptide reactivity in the cysteine only prediction model of the DPRA described in Appendix I of this test guideline. If the positive criterion is not met, the test chemical is considered to be non-reactive according to the prediction model shown in paragraph 28.

    The natural logarithm of the non-depleted peptide concentrations (100-relative peptide depletion (%)) is plotted vs. the concentration of the test chemical at each time point. If a linear relationship is observed (correlation coefficient > 0.90), the slope of this curve is determined. The absolute value of this negative slope corresponds to the observed reaction kinetic constant (pseudo first order rate constants $k_{observed}$ in mM^{-1}). From the $k_{observed}$ value for each exposure time, the reaction kinetic constant (k_t) per concentration and incubation (exposure) time 't' is calculated as follows:

$$k_t\ [M^{-1}s^{-1}] = k_{observed} \cdot \frac{1000}{60 \cdot t}$$

    with 't' being the exposure time in minutes. If no linear relationship is observed (i.e., correlation coefficient < 0.90), the recommendations within paragraph 27.ii should be followed.

25. For each exposure time 't' with a correlation > 0.90, the decimal logarithm (log k_t) is calculated and the highest value is determined as log k_{max}.

Acceptance criteria

26. The following criteria should be met for a run to be considered valid. If one or more of these criteria is not met the run should be repeated.

    a. PC: the log k of the PC at 90 min (log $k_{90\ min}$) should be within the following range: -1.75 to -1.40 $M^{-1}s^{-1}$. If no log $k_{90\ min}$ is obtained in case of e.g., reactivity is not yet statistically significant, the value at 150 min (log $k_{150\ min}$) can be taken into account and should lie in the following range: -1.90 to -1.45 $M^{-1}s^{-1}$.

    b. VC: The coefficient of variance of the 12 VC values of a plate should be < 12.5% for at least 5 of the 6 exposure times.

27. The data obtained for the test chemical are further assessed to check for possible conditions which may affect results:

    (i) Interrupted time-course: If significant peptide depletion is observed at early time-points but not at following time points, there is either an intrinsic non-linear reaction for the test chemical or an experimental variation. In such cases the run is repeated. If the same pattern is reproducible, a non-linear kinetic is proven and the rate-constant observed at early time points is accepted.

    (ii) Non-linear concentration-response: There are few cases where the concentration-response is not linear, but clear depletion is noted. In such cases no rate constant is calculated by the slope method, as regression coefficient is R^2 < 0.90. Alternatively, rate constants can also be calculated based on individual depletion values according to the formula:

    $$k = [\ln(100/(100 - dp))]/(E \times t)$$

    Where 'dp' is depletion in %, 'E' is the concentration of test chemical and 't' is the incubation (exposure) time. Rate constants according to this formula are calculated at each time point 't' and at each concentration 'E' with depletion values above the threshold of 13.89%. For each time point 't' the average of the values for the different concentrations is taken, and then again the log k_{max} for the highest rate at any given time point is reported.

    In such a case a repetition should be performed to check whether this non-linear behaviour is intrinsic to the test chemical, or whether an experimental variation is the cause. If the non-linearity is reproducible, this alternative rate calculation based on the individual depletion values is used for the final rating.

    (iii) Fluorescence interference, namely autofluorescence or fluorescence quenching: Based on the control wells with test chemical only in absence of the test peptide, incidences of autofluorescence and fluorescence quenching by the test chemical can be detected. As the values are corrected for the autofluorescence recorded in the test chemical control wells, this shall not be a problem for low autofluorescence, but with a high autofluorescence, the fluorescence of the peptide-adduct and the autofluorescence may not be fully additive, and subtraction of autofluorescence may lead to apparent depletion, which is not due to loss of peptide signal but to this non-additivity. Thus, one should check whether the observed depletion is time dependent. If this is not the case and autofluorescence is observed, then depletion from autofluorescence is assumed to occur. Fluorescence quenching can also lead to 'pseudo-depletion', but this would happen immediately and resulting depletion would not increase with time. If both conditions are met, it is assumed that depletion from quenching occurs. These cases are rare. If this is not clear from the results a run may be repeated, but if the effect is clear-cut no repetition is needed. In such a case, the test chemical cannot be assessed in the kDPRA (technical limitation)

unless the reaction can be measured with an alternative fluorescent probe not leading to autofluorescence or quenching (see Section II of the Annex 1 to DB-ALM protocol (11)).

(iv) All above cases are detailed in the DB-ALM protocol and automatic alerts appear in the Excel template provided with the DB-ALM protocol when evaluating the data.

Prediction model

28. The kDPRA uses kinetic rates of cysteine peptide depletion for discrimination of UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory 1A) according to UN GHS (3). Results obtained with the test method that do not lead to subcategory 1A categorisation should be interpreted in the context of the limitations stated in paragraph 3 and Annex 1 of this appendix.

Table 1: kDPRA prediction model

Reaction rate	kDPRA Prediction
log k_{max} ≥ -2.0	UN GHS subcategory 1A
Non-reactive or log k_{max} < -2.0	Not categorised as UN GHS subcategory 1A* (non-subcategory 1A)

* Further information is needed to discriminate UN GHS subcategory 1B from UN GHS No Category. Depending on the context (e.g. IATA, DA) this information can be generated prior to or after performing the kDPRA.

29. In cases of a log k_{max} result close to the -2.0 threshold falling in the borderline range calculated for kDPRA (i.e., between -1.93 and -2.06 (12)), no conclusive prediction can be made. In this case, re-testing and/or additional data/information is needed before a conclusive prediction can be made.

30. The kinetic rate constant may be further used in integrated approaches such as IATA or DA to assess the skin sensitisation potency of a test chemical in a continuous scale as needed for risk assessment (3) (10).

Test report

31. The test report should include the following information

Test chemical and Controls (positive control and solvent/vehicle)

For all mono-constituent substance (test and control chemicals)

Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;

Physicochemical properties such as physical state, appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;

Purity, chemical identity of impurities as appropriate and practically feasible, etc;

Treatment prior to testing, if applicable (e.g. warming, grinding);

Concentration(s) tested;

Storage conditions and stability to the extent available.

Additional information for positive control

Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.

Additional information for solvent/vehicle control

Solvent/vehicle used and ratio of its constituents, if applicable;

Justification for choice of other solvent than acetonitrile and experimental assessment of the solvent effect on peptide stability.

Peptide

Supplier, lot, purity

Fluorescence analysis

Fluorimeter used (e.g., model and type), including wavelengths settings

Proficiency testing

Statement that the testing facility has demonstrated proficiency in the use of the test method before routine use by testing of the proficiency chemicals.

Discussion of the results

Description of any unintended modifications to the test procedure.

Discussion of the results obtained with the kDPRA test method and if it is within the ranges described in paragraph 29.

Description of any relevant observations made, such as appearance of precipitate in the reaction mixture at the end of the incubation time, if precipitate was resolubilised or centrifuged.

Conclusion

LITERATURE FOR APPENDIX III

(1) Wareing, B., Urbisch, D., Kolle, S. N., Honarvar, N., Sauer, U. G., Mehling, A.and Landsiedel, R., (2017) Prediction of skin sensitization potency sub-categories using peptide reactivity data. Toxicol In Vitro 45, 134-145.

(2) Roberts, D. W.and Natsch, A., (2009) High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles. Chem. Res. Toxicol. 22, 592-603.

(3) Natsch, A., Haupt, T., Wareing, B., Landsiedel, R., and Kolle, S.N., (2020) Predictivity of the kinetic direct peptide reactivity assay (kDPRA) for sensitizer potency as-sessment and subclassification. ALTEX, 37(4), 652-664

(4) Wareing, B., Kolle, S.N., Birk, B., Alépée, N., Haupt, T., Kathawala, R., Kern, P., Nardelli, L., Raabe, H., Rucki, M., Ryan, C., Verkaart, S., Westerink, W., Landsiedel, R., Natsch, A., (2020) The kinetic Direct Peptide Reactivity Assay (kDPRA): Intra- and inter-laboratory reproducibility in a seven-laboratory ring trial. ALTEX, 37(4), 639-651, DOI: 10.14573/altex.2004291.

(5) Hoffmann S., Kleinstreuer N., Alépée N., Allen D., Api A. M., Ashikaga T., Clouet E., Cluzel M., Desprez B., Gellatly N., Goebel C., Kern P. S., Klaric M., Kühnl J., Lalko J. F., Martinozzi-Teissier S., Mewes K., Miyazawa M., Parakhia R., van Vliet E., Zang Q., Petersohn D.., Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. Crit Rev Toxicol, 2018. 48(5): p. 344-358.

(6) Basketter, D. A., Alepee, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Teissier, S., Maxwell, G., Reisinger, K., Sakaguchi, H., Schepky, A., Tailhardat, M.and Templier, M., (2014) Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis 25, 11-21.

(7) International ad hoc Expert Panel, (2020) Independent Peer-Review Panel Report on the scientific validity of the kinetic Direct Peptide Reactivity Assay (kDPRA) as a modified version of the DPRA assay according to OECD TG 442C, to extend its regulatory applicability to identify UN GHS Subcategory 1A.

(8) S. Casati, K. A., B.D. Asturiol, D. Basketter, S. Dimitrov, C. Dumont, A.T. Karlberg, J. P. Lepoittevin, G. Patlewicz, D. Roberts, A. Worth, (2016) Ability of non-animal methods for skin sensitisation to detect pre- and pro-haptens: Report and recommendations of an EURL ECVAM expert meeting. Available at: http://publications.jrc.ec.europa.eu/repository/handle/JRC100479http://publications.jrc.ec.europa.eu/repository/handle/JRC100479, last accessed 27.05.206.

(9) Urbisch, D., Becker, M., Honarvar, N., Kolle, S. N., Mehling, A., Teubner, W., Wareing, B.and Landsiedel, R., (2016) Assessment of Pre- and Pro-haptens Using Nonanimal Test Methods for Skin Sensitization. Chem. Res. Toxicol. 29, 901-13.

(10) Natsch, A., Emter, R., Gfeller, H., Haupt, T.and Ellis, G., (2015) Predicting Skin Sensitizer Potency Based on In Vitro Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment. Toxicol. Sci. 143, 319-32.

(11) DB-ALM Protocol 217: The kinetic Direct Peptide Reactivity assay (kDPRA). Accessible at http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/DBALM/LATEST/online/DBALM_docs/217_P_kDPRA_final_27Oct20.pdf.

(12) Kolle, S., Mathea, M., Natsch, A., Landsiedel, R., (2021) Borderline ranges for in chemico and in vitro skin sensitization methods determined from ring trial data - acknowledging experimental

uncertainty in defined approaches, Toxicology Letters, 350.

(13) 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. 2011. NIH Publication 11-7709.

(14) ECETOC, ECETOC Document No. 46: Potency Values from the Local Lymph Node Assay: Application to Classification, Labelling and Risk Assessment. 2008.

(15) OECD (2021), Validation report: Kinetic Direct Peptide Reactivity Assay (kDPRA). Series on testing and Assessment n° 337 Organisation for Economic Cooperation and Development, Paris.

(16) 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 Cooperation and Development, Paris.

APPENDIX III, ANNEX 1

KNOWN LIMITATIONS OF THE KINETIC DIRECT PEPTIDE REACTIVITY ASSAY

The table below provides a summary of the known limitations of the kDPRA.

Substance class / interference	Reason for potential underprediction or interference	Data interpretation	Example substance
Metals and inorganic compounds	Known to react with proteins via mechanisms other than covalent binding	Should not be tested	Nickel sulphate; 7786-81-4
Hydroquinones, catechols and aromatic amines	Lag time of oxidation may reduce apparent reaction rate	Results with log k_{max} < - 2.0 can only be accepted if low reactivity can be confirmed after oxidation	Para-phenylenediamine; 106-50-3; Human and LLNA 1A
Thiols or thiol-releasers	Test chemicals with primary SH-groups and those decomposing under the conditions of the assay can react with the detection probe	Test chemical cannot be tested in the kDPRA with derivatisation by thiol reactive probes: other kinetic data with the test peptide e.g. by HPLC may need to be generated (not part of this guideline)	Thioglycerol; 96-27-5; LLNA UN GHS category 1B; Human n/a
Test chemicals having an exclusive lysine-reactivity as observed in DPRA or ADRA	kDPRA only measures reactivity with the cysteine peptide	Results with log k_{max} < - 2.0 for chemicals which specifically deplete NH_2-groups, but not SH-groups in DPRA or ADRA are not conclusive	Some acyl-halides, phenol-esters or aldehydes, Dihydrocoumarin, 119-84-6; LLNA UN GHS category 1B; Human n/a, Glutaric aldehyde; 111-30-8; Human and LLNA UN GHS category 1A
Pro-haptens	Test chemicals for which there is evidence that they strictly require enzymatic bioactivation to exert their skin sensitizing potential	Strict pro-haptens may be underestimated. However chemicals which are i) strict pro-haptens (i.e. test chemicals not also acting as direct haptens or prehaptens, too) and ii) strong allergens were found to be rare	Diethylenetriamine; 111-40-0 (human 1A, LLNA UN GHS category 1)
Fluorescent chemicals with excitation in the range of the fluorescent probe	If fluorescence of test chemicals and of the mBrB-peptide adduct is not additive, pseudo-depletion is observed	Follow the considerations in the DB-ALM Protocol n° 217 to evaluate assay interference	Tetrachlorosalicylanilide; 1154-59-; Human and LLNA UN GHS category 1A
Test chemicals absorbing in the emission range of the probe	If test chemical quenches fluorescence emission of the mBrB-peptide adduct, pseudo-depletion is observed	Follow the considerations in the DB-ALM Protocol n° 217 to evaluate assay interference	Vanillin, 121-33-5; LLNA NC; Human n/a
Mixtures of unknown composition, substances of unknown or variable composition, complex reaction products or biological materials	no information on applicability of kDPRA is available in the published literature	n/a	UVCBs, chemical emissions, products or formulations with variable or not fully known composition
Test chemicals which cannot be dissolved in water or	Not sure if sufficient exposure can be achieved	In such cases, a log k_{max} > -2.0 could still be used to support the identification of the test chemical as a UN GHS subcategory	n/a

acetonitrile or a compatible water-miscible solvent		1A skin sensitiser but no firm conclusion should be drawn in case log k_{max} is < -2.0. Alternative vehicle may be used according to the prescriptions given in paragraph 12.	
Test chemicals which precipitate in reaction solution	Not sure if sufficient exposure can be achieved: If a precipitate is observed immediately upon addition of the test chemical solution to the peptide solution, due to low aqueous solubility of the test chemical, one cannot be sure how much test chemical remained in the solution to react with the peptide.	In such a case, a positive result (i.e. log kmax ≥ -2.0) could still be used, but a negative result (i.e., non-reactive or log kmax < -2.0 outcome) should be interpreted with due care (see also provisions in paragraph 4 for the testing of chemicals not soluble up to a concentration of 20 mM in the kDPRA).	Methyl-2-nonynoate[6]; 111-80-8; LLNA NC
Test chemicals promoting cysteine-peptide oxidation		May lead to a potential over estimation of peptide reactivity.	DMSO

[6] 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

APPENDIX III, ANNEX 2

PROFICIENCY SUBSTANCES

In Chemico Skin Sensitisation: kinetic Direct Peptide Reactivity Assay (kDPRA)

Prior to routine use of the test method described in this appendix, laboratories should demonstrate technical proficiency by correctly obtaining the expected kDPRA prediction for at least 8 of the 9 proficiency substances recommended in Table 1 and by obtaining cysteine rate constants log k_{max} that fall within the respective reference range for 7 out of the 9 proficiency substances. These proficiency substances were selected to represent the range of responses for skin sensitisation hazard and potency. Other selection criteria were that they are commercially available, that high quality *in vivo* reference data and high quality *in vitro* data generated with the kDPRA are available, and that they were used in the industry-coordinated validation study to demonstrate successful implementation of the test method in the laboratories participating in the study.

Table 1: Recommended proficiency substances for demonstrating technical proficiency with the kinetic Direct Peptide Reactivity Assay

Proficiency substances	CASRN	Physical state	*In vivo* prediction[1]	UN GHS Category LLNA	UN GHS Category human	kDPRA prediction [2]	Range of log k_{max} [2]
2,4-Dinitrochlorobenzene	97-00-7	Solid	Sensitiser (extreme)	1A	1A	1A	(-0.8) – (-0.4)
Methylisothiazolinone	2682-20-4	Solid	Sensitiser (extreme)	1A	1A	1A	(-0.5) – (-0.1)
Oxazolone	15646-46-5	Solid	Sensitiser (extreme)	1A	No data	1A	(-0.3) – (0.0)
Methyl-2-octynoate	111-12-6	Liquid	Sensitiser (strong)	1A	1A	1A	(-1.6) – (-1.2)
Isoeugenol	97-54-1	Liquid	Sensitiser (moderate)	1A	1A	1A	(-1.4) - (-1.1)
2,3-Butanedione	431-03-8	Liquid	Sensitiser (weak)	1B	No data	non-1A (1B or NC)	(-3.2) – (-2.1)
Ethylene glycol dimethacrylate (EGDMA)	97-90-5	Liquid	Sensitiser (weak)	1B	1B	non-1A (1B or NC)	(-2.8) – (-2.1)
4-Methoxyacetophenone	100-06-1	Solid	Non-sensitiser	No Cat.[3]	No Cat.[3]	non-1A (1B or NC)	Not reactive
Chlorobenzene	108-90-7	Liquid	Non-sensitiser	No Cat.[3]	No Cat.[3]	non-1A (1B or NC)	Not reactive

[1]The *in vivo* hazard and (potency) predictions are based on LLNA data (13). The *in vivo* potency is derived using the criteria proposed by ECETOC (14).
[2] Rounded ranges determined on the basis of at least 14 log k_{max} determinations generated by 7 independent laboratories.
[3] Non sensitisers according to the UN GHS.

Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2022)16 | 1

Unclassified

English - Or. English

19 September 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Detailed Review Paper on In Vitro Test Addressing Immunotoxicity With a Focus on Immunosuppression

Series on Testing and Assessment No. 360

JT03502631

OECD Environment, Health and Safety Publications
Series on Testing & Assessment
No. 360

Detailed Review Paper on In Vitro Test Addressing Immunotoxicity With a Focus on Immunosuppression

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 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 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 Series on 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

FOREWORD

The project to develop a Detailed Review Paper (DRP) on non-animal approaches that could be used to test chemicals for their potential immunotoxic effects was initiated by the Japanese Center for the Validation of Alternative Methods (JaCVAM), who submitted a Standard Project Submission Form (SPSF) proposing the preparation of this DRP to OECD. Japan had already investigated the potential relevance and reproducibility of certain *in vitro* assays used in specific combinations to inform the assessment of immunotoxicity. Given the absence of standardized *in vitro* methods in this particular area of chemical safety testing, the Working Party of the National Coordinators of the Test Guidelines Programme recommended to start with a DRP to enable an overview of the state-of-the-science, techniques, and methods available.

The Detailed Review Paper was prepared by an international team of subject matter experts. A dedicated OECD Expert Group was formed to provide input into the draft DRP and two WNT commenting rounds were organized in 2020 and 2021 to subject the document to broad review and comments from the regulatory science community.

The Working Party of the National Coordinators of the Test Guidelines Programme approved this Detailed Review Paper at its 34th meeting in April 2022. This document is published under the responsibility of the Chemicals and Biotechnology Committee.

LIST OF ABBREVIATIONS

AOP: Adverse Outcome Pathway
CFU-GM: Colony Forming Unit-Granulocyte-Macrophage
CTL: Cytotoxic T lymphocyte
DC: Dendritic Cell
DHR: Delayed Hypersensitivity Response
DTH: Delayed Type Hypersensitivity
EAGMST: Extended Advisory Group on Molecular Screening and Toxicogenomics
ECVAM: European Centre for the Validation of Alternative Methods
FCC: Fluorescent Cell chip
HWBCRA: Human Whole-Blood Cytokine Release Assay
ICH: International Conference on Harmonisation
IL: interleukin
IL2-LA: IL-2 Luciferase activity
IL8-LA: IL-8 Luciferase activity
iPSC: induced Pluripotent Stem Cells
LOEL: Lowest Observed Effect Levels
LPS: Lipopolysaccharide
MCP-1: Monocyte chemoattractant protein-1
M-CFS: Macrophage Colony-Stimulating Factor
MIG: Macrophage-Induced Gene
MITA: Multi-ImmunoTox Assay
MLR: Mixed Leukocyte Reaction
mMITA: modified Multi-ImmunoTox Assay
NK: Natural Killer
OECD: Organisation for Economic Co-operation and Development
PGE_2: Prostaglandin E_2
SLG: Stable Luciferase Green
SLR: Stable Luciferase Red
SLO: Stable Luciferase Orange
SRB: Sulforhodamine B
TDAR: T-cell-Dependent Antibody Response
TF: Tissue Factor
TM: Thrombomodulin
TNF: Tumor Necrosis Factor
TPA: Phorbol Myristate Acetate
VCAM-1: Vascular Cell Adhesion Molecule 1

Table of contents

EXECUTIVE SUMMARY

This Detailed Review Paper (DRP) aims to present and discuss the application and interpretation of *in vitro* immunotoxicity assays, mainly covering immunosuppression, and to define an *in vitro* tiered approach to testing and assessment. This project was led by Japan, as Japan has developed three Adverse Outcome Pathways (AOPs) for immunotoxicity in the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST). Japan also coordinated a validation study of the Multi-ImmunoTox Assay (MITA), based on one of the proposed AOPs.

A well-functioning immune system is essential for maintaining the integrity of an organism. Immune cells are an integral part of other systems including the respiratory, dermal, gastrointestinal, neurological, cardiovascular, reproductive, hepatobiliary, musculoskeletal system, and endocrine systems. As such, exposure to immunotoxic compounds can have serious adverse health consequences affecting responses to both communicable and non-communicable diseases. It is therefore important to understand the immunotoxic potential of xenobiotics and the risk(s) they pose to humans.

In contrast to the *in vivo* testing batteries traditionally used to investigate systemic chemical toxicity, *in vitro* methods have historically been used to generate focused mechanistic information. When using *in vitro* assays for screening purposes it is likely that several assays will be required to identify immunotoxicants because of the different components of the immune system and their influences on other systems. A tiered testing strategy is proposed to assess immunotoxicity *in vitro.* In the proposed tiered approach, pre-screening for direct immunotoxicity *in vitro* begins by evaluating myelotoxicity (Tier 1). Compounds capable of damaging or destroying bone marrow cells will most likely have immunotoxic effects, as the majority of immune cells are derived from a common precursor located in the adult bone marrow. If compounds are not potentially myelotoxic, they should be tested for direct leukotoxicity, as defined as toxicity to any cell of the lymphoid or myeloid lineages (Tier 2). Compounds should then be tested for immunotoxicity at non-cytotoxic concentrations using various approaches, such as cytokine production, T cell–dependent antibody response, lymphocyte proliferation assay, mixed leukocyte reaction, and natural killer cell assay (Tier 3).

Despite the need for an *in vitro* tiered system to evaluate immunotoxicity, at present there is no consensus on which assays to use, nor how, and there are no Organisation for Economic Co-operation and Development (OECD) Test Guidelines to detect chemical immunosuppression *in vitro*. It is clear that one assay alone will not be able to cover all of the potential adverse effects of chemicals on the immune system and that a larger set of assays that will cover the spectrum of immunotoxicity is needed. The MITA is one example of such an integrated testing strategy which may be used to predict the immunotoxicity of chemicals, and may be even more powerful when combined with complementary assays. More research and investigation are needed to develop candidate assays amenable to detect immunotoxic substances without the use of animals, but available tools can already be used in an integrated fashion for that purpose.

. Introduction

1. Since the early 1980s, there has been increasing recognition that some natural and synthetic substances to which humans may be exposed are able to interfere with the function of the immune system (e.g. Vos, 1977; Dean et al 1982). As an adequately functioning immune system is essential for maintaining the integrity of an organism, immune dysregulation can have serious adverse health consequences, ranging from reduced resistance to infection and neoplasia to allergic and autoimmune conditions. For example, it is well-known in clinical practice that treatment of patients with immunosuppressive drugs, such as those intended to prevent rejection of organ transplants, is often associated with increased numbers of infections or tumors. In addition to drugs, environmental contaminants and food additives can also target the immune system, resulting in immune dysregulation. Originally the emphasis of immunotoxicology, which is defined as the study of toxicological effects of xenobiotics on the immune system, was on immunosuppression. Later, more attention was given to chemical-induced allergies, inadvertent immunostimulation, and chemical-induced autoimmunity. The immune system comprises a complex network of different cell types located in various organs and their mediators, which operate to maintain homeostasis. Because of its complex nature, influences of chemical exposure can occur on different components of the immune system, with different mechanisms, eventually leading to adverse health outcomes. For this reason, testing has been done most often in the intact animal if not in humans themselves, and *in vitro* testing was predominantly used to unravel specific mechanisms of immunotoxicity.

2. Current practices in immunotoxicity testing are still varied and employ either or both unchallenged and challenged immune systems. Although useful information can be obtained by the histopathology of immune organs and enumeration of immune cells obtained from regular 28-day general toxicity tests, most immunotoxicity testing historically has been organized into tiers (Hinton 2000; Luster et al. 1988). Functional immune tests, which may be used in various tiers, enable the generation of data of increased quality and specificity.

3. There are several regulatory guiding principles in immunotoxicology published as pharmaceutical industry guidances (eg. ICH S8 guidance) or chemical industry guidelines for immunotoxicity. In addition to the specific OECD Test Guidelines for skin sensitizing activity (*in vivo* OECD TG 406, *in vivo* TG 429, *in vivo* TG 442A-B, *in vitro* TG 442C-E, *in silico/in vitro* GL 497), other OECD guidelines for toxicity testing include assays for assessing immunotoxic potential in the context of more general toxicity testing, such as the 28-Day Repeated Dose Toxicity Study (OECD 407), the 90-Day Repeated Dose Toxicity Study (OECD 408), and the Extended One Generation Reproduction Toxicity Study (OECD 443). The World Health Organisation/International Programme on Chemical Safety (WHO/IPCS) has published the Guidance for Immunotoxicity Risk Assessment for Chemicals (IPCS Harmonization project No. 10) oriented at immunosuppression, inadvertent immunostimulation, and autoimmunity caused by chemical exposure.

4. A workshop hosted by the International Life Sciences Institute-Health and Environmental Sciences Institute (ILSI-HESI) was held to share perspectives on immunotoxicity testing, developmental immunotoxicity, and integrated approaches to testing and assessment (IATA[1]) of immunotoxicity. The

[1] IATA Are pragmatic, science-based approaches for chemical hazard characterization that rely on an integrated analysis of existing information coupled with the generation of new information using testing strategies [See OECD

workshop summarized that standard toxicity studies, combined with trigger-based functional immune testing approaches, represent effective approaches to evaluate immunotoxic potential (Boverhof et al. 2014). To date, the chemical risk assessment practice in OECD member countries has relied to a large extent on animal data. However, these animal models and assays have many drawbacks: they are resource intensive (time, costs, and animal numbers), pose ethical problems, and have varying ability for predicting human health outcomes. For these reasons, there is a clear societal desire to minimize the use of experimental animals for toxicity testing, while ensuring adequate protection to human health and the environment. Efforts have been and are being made to replace, reduce, or refine animal-based assays as much as possible. The EU Directive (2010/63/EU) was sanctioned to achieve these goals, and some US regulatory agencies aim to phase out animal testing by 2035 (EPA, 2019).

5. In the regulatory context, while animal models for hypersensitivity (respiratory and skin sensitization for type 1 and type 4 hypersensitivity, respectively) and immunosuppression show an overall good correlation with human data, currently available animal models and assays are not valid to assess the potential for systemic hypersensitivity (type 2 and type 3 hypersensitivity) and autoimmunity. Because we understand the mechanism of dermal sensitization to a large degree, there has been success with the development of *in vitro* methods for hypersensitivity. Additional efforts for the development of *in vitro* assays to detect other forms of immunotoxicity (e.g. immunosuppression) are needed. One aim is that this DRP, which is focused on immunosuppression, will trigger further development of methods and approaches in those immunotoxicological aspects not covered by the current Test Guidelines. Several non-animal–based testing methods for immunotoxicity have been published, although few have reached the stage of validation and acceptance by international regulatory bodies. A workshop hosted by the European Centre for the Validation of Alternative Methods (ECVAM2) in 2003 focused on what was at that time the current status of *in vitro* systems for evaluating immunotoxicity (Gennari et al. 2005). In this workshop, a tiered approach for *in vitro* immunotoxicity testing (similar to that used for *in vivo* immunotoxicity testing) was proposed. The proposed tiered approach would begin with pre-screening for direct immunotoxicity by evaluating myelotoxicity (Tier 1). Compounds capable of damaging or destroying bone marrow will most likely have immunotoxic effects. If compounds are not potentially myelotoxic, they should be tested for leukotoxicity (Tier 2). Compounds should then be tested for immunotoxicity using various approaches, such as T cell–dependent antibody response (TDAR), lymphocyte proliferation assay, mixed leukocyte reaction (MLR), natural killer (NK) cell assay, dendritic cell (DC) maturation assay, human whole-blood cytokine release assay (HWBCRA), and fluorescent cell chip (FCC) assay (Tier 3).

6. The T lymphocyte (T cell), being a crucial cell type in function and regulation of many aspects of the immune system, has been a prime target in *in vitro* assays for immunotoxicity testing. The IL-2 Luc assay developed by Dr. S. Aiba from the Department of Dermatology at Tohoku University School of Medicine in Japan is one such assay that has reached a level of validation (Kimura, 2020). Yet, it is clear that one assay alone will not cover the entire spectrum of potential adverse effects of chemicals on the immune system. Any validated test will therefore be part of a larger set of different assays that could potentially assess all types of immunotoxicity including immunosuppression, sensitization, and autoimmunity, which together will be able to adequately predict immunotoxic action of chemicals. An integral battery of tests to achieve this goal, designated as Multi-ImmunoTox Assay (MITA), is being developed by the same institute, and is described in this document as an example of how to integrate multiple tests.

(2016) Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA), Series on Testing and Assessment No. 260].

7. This DRP reviews developments made in the field of non-animal-based immunotoxicity testing. Immunotoxicity by drugs or chemicals can be manifested in various ways, including dysregulation of the immune response, which could lead to immunosuppression or inappropriate immunostimulation. The latter can include unintended immune stimulation, sustained inflammation, hypersensitivity reactions and autoimmune disease. With reference to chemical-induced immunotoxicity, the effect may not be exclusively in one direction and the same substance can produce immunosuppression or immune stimulation, depending on the dose and the cellular target. Thus, it may be more appropriate to define an immunotoxic substance as any agent that can alter one or more immune functions resulting in an adverse effect for the host. In this way we focus not on the direction of the effect, but on its consequence. For this reason, we prefer to use the term immunotoxicant/immunotoxicity throughout the document, although the primary focus of this DRP will be on immunosuppression. This document is not meant to be an extensive review of immunology but does include a brief overview of the immune system, a description of commonly available assays and data that have been used to determine suppressive effects of chemicals on immune responses, and considerations for establishing assays for immunotoxicity testing. Thus, the purpose of this DRP is to provide a brief overview of the complicated nature of the immune system and assessment of immunotoxicity using cell-based methods from a regulatory standpoint. The focus of this DRP is thus on in vitro test methods that are considered ready for standardisation as OECD Test Guidelines. Although there is on-going research to further develop computational approaches, organ-on-a-chip and other technologies, these are not covered in this DRP since the level of readiness for regulatory application in the area of immunotoxicity may still be limited.

. Basic concept of immunotoxicity

8. The immune system is responsible for providing protection against foreign invaders while not reacting to self-entities. The responses required to eliminate a threat occur at the same time that responses to a non-threat must be quelled. For these reasons, immunity is dynamic, with constant surveillance needed to determine whether to initiate an immune response or to not respond. Immunity can therefore be considered a continuum along which the actions of initiating a response, resolving a response, or not responding at all are carefully balanced to achieve immune homeostasis. A tip of the balance can lead to morbidity or mortality; suppression of immune responses renders an individual susceptible to infections or cancer, while enhancement can result in hypersensitivity or autoimmune disease.

9. The immune system is comprised not only of specific immune organs, but also specialized immune cells present in most tissues. It is therefore an integral part of other systems including the respiratory, dermal, gastrointestinal, neurological, cardiovascular, reproductive, hepatobiliary, musculoskeletal, and endocrine systems. An immune response occurs through the coordination of many different cell types and can involve several tissues. The thymus and bone marrow are critical for immune cell development, while the lymph nodes and spleen are organs in which many immune responses occur.

10. Initially, the cells that are involved in detecting a threat are those belonging to the innate arm of the immune response. Innate cells, such as macrophages and neutrophils, express receptors that specifically recognize pathogen-specific patterns of proteins or lipids on foreign invaders. These innate cells can release directly cytotoxic proteins, or produce cytokines or chemokines to recruit other immune cells to the area of insult. Macrophages and dendritic cells (DCs) also serve as antigen presenting cells (APCs), which provide a bridge between the innate and adaptive arms of the immune response. The most critical APC is the DC. DCs are capable of antigen uptake, allowing for removal and destruction of pathogens. The DCs also process the antigen and present antigenic epitopes to T cells, allowing for activation of the adaptive arm of the immune response.

11. The adaptive immune response includes actions by T cells and B cells, which express receptors that recognize antigenic epitopes. Pathogen-specific T cells and B cells then undergo robust proliferation, known as clonal expansion, to ensure that a large population of cells is present to react to the current threat. Specialized subsets of T cells aid in the immune response by recruiting or activating other immune cells (TH1 or TH2 cells) or by directly killing infected cells (cytotoxic T cells, CTL). Other specialized T cells include TH17 cells, which produce high levels of pro-inflammatory cytokines that recruit innate cells, and regulatory T cells (Tregs), which help to regulate the immune response and prevent autoimmune responses. The primary role of the B cell in an adaptive response is to produce antibodies, which can neutralize foreign invaders, initiate cytolysis of infected cells, or enhance the actions of innate cells, such as phagocytosis. There are also specialized cells called innate lymphoid cells (iLCs) that play a critical role in early responses in part through robust cytokine production. Both the innate and adaptive arms of an immune response can contribute to inflammation, and even though inflammation is a normal process of pathogen destruction, it can also produce tissue damage.

12. It is clear from the above that the immune system has the functional mechanisms to eradicate threats but must also be tightly regulated to avoid inappropriate reactions. Thus, the immune system is

susceptible to toxic insults in part because of: 1) the need to maintain the delicate balance between activation, regulation, and silencing; 2) its dependence on regeneration of cells from hematopoietic stem cells in the bone marrow; 3) its requirement of clonal expansion of T cells and B cells by cellular proliferation during the adaptive response; 4) the required maintenance of appropriate levels of lymphocyte subsets, including effector, memory and regulatory subsets; and 5) its interaction with other physiological systems (i.e., gut microbiota) to maintain immune homeostasis.

13. In considering how a drug or chemical exhibits immunosuppression, the agent might alter the number of cells (innate or adaptive), the ability of the cells to produce cytokines, chemokines, antibodies or growth factors, the composition of the subpopulations of cells present at the site of the response, or the function of the cells (i.e., kill infected cells or proliferate). Signs of immunotoxic potential of agents in standard animal toxicology studies can be defined by hematological changes (i.e., leukocytopenia/leukocytosis, granulocytopenia/granulocytosis, or lymphopenia/lymphocytosis), alterations in immune system organ weights or histology, changes in serum antibodies, or changes in incidence of infections or tumors (Galbiati, et al., 2010).

14. Once it has been determined that an agent possesses immunosuppressive potential, the mechanism(s) by which an agent acts can be investigated. First, given the extensive involvement of different cell types and organs in an immune response, the cellular target(s) must be defined, which could identify that the innate or adaptive (or both) arms are sensitive to alteration by a drug or chemical. Second, it is important to define whether the parent compound or a metabolite is mediating the immunotoxic effects. Immune cells have limited capacity to metabolize chemicals, but immune cells may be targeted by metabolites generated in other organs, such as the liver. Third, it is important to determine whether the agent is directly or indirectly producing immune system toxicity. For example, there are critical interactions between the immune system and endocrine systems such that immunity is regulated by various neurotransmitters and hormones (Karmaus et al 2015*)*. Thus, the mechanisms by which an immunotoxicant acts might be different in males versus females, or the mechanism of immune suppression might involve induction of stress, as high glucocorticoid levels suppress immunity. Fourth, an immunotoxicant might alter the gut microbiome, subsequently alerting immune homeostasis. Finally, the intracellular components altered by the agent that led to immune alteration should be defined. For instance, identification of whether a drug or chemical alters specific cytokines could dictate if the agent will affect all T cells, T cells and B cells, or subpopulations of one or both cells (i.e., TH1, TH2, TH17, and/or Tregs).

15. In summary, a robust immune response requires the careful coordination of cellular interactions, subsequent recruitment and/or activation of various cells, and mechanisms for regulation at all steps. There are several cell types, immune cellular functions, and/or changes in distinct physiological systems that influence immune homeostasis and might be disrupted by an immunotoxicant. Thus, *in vitro* immune toxicity tests are critical tools for deciphering whether a drug or chemical suppresses the immune response, but it is just as important to use a battery of tests to fully characterize how an agent exhibits immunotoxicity.

. Current status of AOPs on immunotoxicity testing

~~16.~~ An Adverse Outcome Pathway (AOP) describes a logical sequence of causally linked events at different levels of biological organization, which follows exposure to a chemical and leads to an adverse health effect in humans or wildlife. AOPs are the central element of a toxicological knowledge framework, promoted by member countries through OECD, built to support chemical risk assessment based on mechanistic reasoning (OECD, 2020a). These AOPs are available in the AOP Wiki (OECD, 2020b), an interactive and virtual encyclopedia for AOP development.

17. All AOPs on immunosuppression currently available in the OECD work plan are on-going and shown in Table 1. Project 1.74: Inhibition of JAK3 leading to impairment of TDAR is under development and will not be discussed. However, two of the proposed AOPs, Project 1.38 " No. 154: Inhibition of Calcineurin Activity Leading to Impaired T-Cell Dependent Antibody Response" and Project 1.48 "No. 277: Inhibition of IL-1 binding to IL-1 receptor leading to increased susceptibility to infection" are undergoing peer review. No. 154 shows calcineurin (CN) activity is inhibited when CN inhibitors bind to CN with their respective immunophilins, which interferes with the nuclear localization of nuclear factor of activated T cells (NFAT), a substrate of CN. As a result, the formation of functional NFAT complexes with activator protein-1 (AP-1) that bind at the site of IL-2, IL-4 and other T cell-derived cytokine promoters is reduced, thereby suppressing production of these cytokines. Among the affected cytokines from each of the helper T cell subsets, reduced production of IL-2 and IL-4 affects the proliferation and differentiation of B cells to suppress the TDAR. AOP 277 addresses one Molecular Initiating Event (MIE), impaired IL-1 receptor signaling. The biological plausibility of the signaling cascade from the activation of IL-1 receptor to the activation of nuclear factor κB (NF-κB) is already confirmed (Verstrepen et al., 2008). In addition, the biological plausibility that suppressed NF-κB activation leads to impaired T cell activation and antibody production leading to increased susceptibility to infection is supported by several published works (OECD, 2020b). To recapitulate some aspects of the *in vivo* immunotoxic responses by using *in vitro* methods, it will be very important to more closely mimic respective *in vivo* situations based on individual AOPs, although this may be complicated and laborious.

Table 1. Ongoing AOPs for Immunosuppression in the OECD work plan

Project 1.38: The Adverse Outcome Pathway on Binding of FK506-binding protein (FKBP12) by calcineurin inhibitors leading to immunosuppression	
Lead: Inclusion in work plan: Current situation:	Japan 2015 No. 154: Inhibition of Calcineurin Activity Leading to Impaired T-Cell Dependent Antibody Response , External review completed as presented in EAGMST meeting 2020.
Project 1.48: The Adverse Outcome Pathway on Dysregulation of IL-1 transcription leading to immunotoxicity	
Lead: Inclusion in work plan: Current situation:	Japan 2016 No. 277: Inhibition of IL-1 binding to IL-1 receptor leading to increased susceptibility to infection, External review completed as presented in EAGMST meeting 2020.
Project 1.74: Inhibition of JAK3 leading to impairment of TDAR	
Lead: Inclusion in work plan: Current situation:	Japan 2018 No. 315: Inhibition of JAK3 leading to impairment of T-Cell Dependent Antibody Response, Under Development

. State-of-the-art knowledge in the field of *in vitro* or non-animal assays

18. As mentioned in the previous sections, an immunotoxic compound is a compound that can alter one or more immune functions resulting in an adverse effect for the host (Luster et al., 1992). Any alteration in immune functions (e.g., antigen presentation, cytokine production, cell proliferation) that significantly deviates from control values and that can be linked to a downstream immunotoxic effect, should be considered as adverse if the immunomodulation is unintended. Considering this definition and that many functional immune tests following *in vivo* exposure are de facto *ex vivo* tests, attempts are being made to recapitulate the immune response following chemical exposure *in vitro* considering as much as possible the complexity of immune function and integration with multiple cells and soluble mediators. When interpreting the *in vitro* results, as discussed in section V, any limitations of the assay(s) (e.g. metabolic capacity) should be taken into consideration.

19. Although outside the scope of this DRP, the success in replacing animal testing for assessment of skin sensitization is a noteworthy accomplishment in the field of immunotoxicology. In the past two decades, thanks to the mechanistic understanding of the skin sensitization process that made it possible to define the first AOP (OECD (2014), *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins*, OECD Series on Testing and Assessment, No. 168, OECD Publishing, Paris, https://doi.org/10.1787/9789264221444-en) incredible progress has been made into the development and validation of non-animal models to detect skin sensitizers, and several OECD test guidelines have been published (TG 442C-E and GL 497). In addition, several reviews have been recently published on this topic (de Avila, 2019; Casati et al., 2018; Corsini et al., 2018).

20. With the successful development of new approach methodologies for the assessment of skin sensitization, the next step must be the development of a strategy to address immunosuppression without the use of animals. We recognize that the level of complexity and our understanding of the mechanistic pathways that lead to immunosuppression are less clear than those for hypersensitivity, and additional studies are needed to prove the possibility and feasibility to address immunotoxicity using *in vitro* approaches (Corsini and Roggen, 2009; Lankveld et al., 2010; Galbiati et al., 2010; Luster and Gerberick, 2010; Hartung and Corsini, 2013). Nevertheless, we must acknowledge that important progress has been made in the development of *in vitro* assays for the assessment of immunotoxicity, with the HWBCRA and the MITA representing significant achievements (Langezaal et al., 2001; Kimura et al., 2018).

In vitro opportunities to identify immunosuppressive agents

21. As noted above, factors such as age at onset, gender, dose, duration, and route of exposure may result in differing effects on the immune system and skew the adverse response in the direction of immunosuppression or immunostimulation. Thus, while this DRP focuses on immunosuppression, all

the assays described in this section could lead to results demonstrating either no effect, immunosuppression and/or immune enhancement when compared to a control group, with the last two indicating an immunotoxic effect. Before beginning any evaluation, it would be useful to collect all available information in the literature, including information from sources such as the CompTox Dashboard, and the Integrated Chemical Environment data integrator, which include data from multiple endpoints, several of which may be relevant for immunotoxicity (Bell et al., 2017; Naidenko et al., 2021).

22. Due to the complexity and diversity of the immune responses, it was generally assumed that it would be very difficult to reproduce all the key events and processes *in vitro*. To a large extent, *in vitro* systems do not consider the interactions of the different cellular and soluble components involved in the immune response, nor the potential for neuro-immuno-endocrine interactions. Therefore, the assessment of *in vitro* immunotoxicity has often been valuable only in the cases of a direct immunotoxicant (Gennari et al., 2005). Several isolated processes can be studied *in vitro* including antigen presentation, lymphocyte proliferation, cytokine production, phagocytosis, lytic functions, and even primary antibody production, offering the possibility to assess immunotoxicity *in vitro.* Recently there has been incredible progress in 3D models with engineered immune tissues and organs, such as bone marrow, thymus, lymph nodes and spleen being described (see review by Gosselin et al., 2018), and microfluidic body-on-a-chip, and in the future it may be possible to identify both direct and indirect immunotoxicants using an integrated model of the whole human immune system (Shanti et al., 2018).

23. Primary human immune cells such as monocyte-derived DCs, T cells, and B cells obtained from human peripheral blood may be useful materials for *in vitro* testing and are highly clinically relevant. However, the use of human primary cells for developing a testing assay may have several issues regarding ethics, donor-to-donor variability, versatility, and reproducibility. Variability reflects diversity in individual immune capability that requires consideration, and it is important to understand and ensure that it is reflected in *in vitro* systems developed using non-primary cells. Variability and predictive capacity are important considerations for establishing scientific confidence for individual or combinations of *in vitro* methods. Use of induced pluripotent stem cells (iPSC) technology may further improve the versatility of *in vitro* assays. Several human immune cells including T cells, B cells, DCs, and NK cells have been generated from iPSC (Vizcardo et al., 2013; French et al., 2015; Senju et al., 2011; Kitayama et al., 2016). In the future, iPSC technology might be used to provide different populations of bone marrow cells such as iPSC-derived hematopoietic stem cells and mesenchymal stem/stromal cells to obtain a more complete picture of myelotoxicity *in vitro*.

24. When assessing the potential immunotoxicity of xenobiotics, bioavailability should also be considered as part of *in vitro* testing. If a compound is not systemically available, a direct adverse effect on the immune system should not be expected as the compound would not reach immune cells or tissues. However, local effects at the site of exposure would still be possible. For example, it is important to consider that the immune system is closely linked and influenced by the microbiota. A substance taken orally could influence the microflora and mucosal DCs, and even if it is not absorbed into the systemic circulation, it may in turn influence the immune response (Belkaid and Hand, 2014). *In vivo* toxicokinetic studies, if available, or physiologically based pharmacokinetic (PBPK) models should be used to assess or predict absorption. For consistency, any alternative means to obtain information on systemic bioavailability without *in vivo* animal data should be preferred. At this regard, the ECHA Guidance in information requirements and chemical safety assessment Chapter R.7c (available at: https://echa.europa.eu/documents/10162/13632/information_requirements_r7c_en.pdf/e2e23a98-adb2-4573-b450-cc0dfa7988e5) contains a section on how information on systemic bioavailability can be gathered.

25. As a general strategy, *in vitro* testing for direct immunotoxicity should be done in a tiered approach (adapted from Gennari et al., 2005; Corsini and Roggen, 2009), with the first tier measuring myelotoxicity (Tier 1). Myelotoxicity or bone marrow toxicity represents the decrease in production of cells responsible for providing immunity (leukocytes), carrying oxygen (erythrocytes), and/or those

responsible for normal blood clotting (thrombocytes). In the context of immunotoxicity, myelotoxicity would refer to toxicity to precursors of immune cells. Compounds that are capable of damaging or destroying the bone marrow will have a profound immunotoxic effect, since the effectors of the immune system itself will no longer be available. Therefore, if a compound is myelotoxic, according to the specific assay performed, the chemical will *de facto* be an immunotoxicant. *Ex vivo* colony forming assays are used to assess bone marrow toxicity in animal models. The methodologies for evaluating myelotoxicity *in vitro* using bone marrow culture systems are well-characterized and scientifically validated for reproducibility and predictive capacity (Pessina et al., 2003; 2005; Rich and Hall, 2005; Haglund et al., 2010), but they are not required for regulatory testing or widely accepted as a standard screening tool due to technical challenges. Results of a pre-validation study showed that the *in vitro* colony forming unit-granulocyte-macrophage assay (CFU-GM) is linear and highly reproducible within and between laboratories (Pessina et al., 2001; Pessina et al., 2010). In an international blind trial (Pessina et al., 2002), the model correctly predicted the human maximum tolerated dose (MTD) for 20 drugs out of the 23 tested (87%).

26. Compounds that are not directly myelotoxic may still selectively damage leukocytes (defined as cells of lymphoid or myeloid lineage), which are the primary effectors and regulators of immunity, so the next step in evaluating potential immune toxicity *in vitro* is testing for leukotoxicity (Tier 2). If the agent is cytotoxic to immune cells at concentrations relevant for human exposure, the agent should be considered an immunotoxicant. If not, data on leukotoxicity will be used for the selection of the concentration range to be used in Tier 3, in which only non-cytotoxic concentrations should be used (cell viability > 80%). The choice of the cellular model to be used in Tier 3 will depend on the target identified in Tier 2 and the functional test to be performed (e.g., B or T cells, DCs, NK cells). There are several methods that can be used to assess cytotoxicity, among which the fluorometric microculture cytotoxicity assay to screen for leukotoxicity can be mentioned. This assay is based on measurement of fluorescence generated from hydrolysis of fluorescein diacetate to fluorescein by cells with intact plasma membranes. 20This method has been used in primary peripheral blood mononuclear cells (from different species) by Hassan et al. (2007) to measure cell survival following exposure to cytotoxic drugs and shown to correlate well with CFU-GM data. Cytotoxicity may result from the destruction of rapidly dividing cells by necrosis or apoptosis. Alternatively, chemicals may cause cell death by interfering with cell activation by affecting signal transduction pathways. A variety of well-established and accepted methods are available for assessing cell viability (e.g., colorimetric, flow cytometric assays), and several of these assays are integral parts of currently accepted OECD TGs. If leukotoxicity occurs at concentrations relevant to the expected *in vivo* human exposure, then cytotoxicity remains a relevant effect associated with immunosuppressive potential. If expected *in vivo* concentrations are unknown, the use of PBPK models should be considered to predict pharmacokinetic parameters. After consideration of myelotoxicity and leukotoxicity, which if positive are sufficient to classify the compound as immunotoxic, basic immune cell functionality may then be assessed by performing specific functional assays that identify targeted cells and processes (Tier 3). These assays (i.e., proliferative responses, lytic activity, cytokine production), should be conducted using concentrations of the test chemical that are not cytotoxic and provide acceptable viability for the specific assay.

27. Alternative *in vitro* methods have the potential to reduce animal use and testing cost, to facilitate immunotoxicity screening, and prioritization efforts (Luebke, 2012). Several *in vitro* assays that evaluate specific functions or functional correlates of the immune system (e.g., CTL activity, NK cell activity, antibody production, cytokine production, cell proliferation) have been used to assess immunotoxicity. A significant response in any of these assays should be interpreted as the chemical possessing the potential for immunotoxicity and should warrant further investigation. In Table 2, relevant immune components, and opportunities for *in vitro* assessment of immunotoxicity (immunosuppression) are reported, and readers are referred to the cited works for further details. The tests shown in Table 2 refer to what has been published with the specific purpose of identifying immunotoxic substances *in vitro*. Other aspects of immune function may be evaluated (e.g,. phagocytosis, production of lysozyme,

microbicidal activity) to identify the immunotoxic potential of a substance, but have been used less frequently for screening purposes. In addition, while there are numerous immune cell subtypes, (e.g., Th subpopulations or DC subsets), involved in the different immune responses, there is a need to establish validated methods to study how alterations in their functional capabilities contribute to immunotoxicity. Considering the complexity of the immune response, more than one *in vitro* test will likely be needed to define the immunotoxic potential of a xenobiotic. Table 2 also includes the source of cell used (i.e. human vs animal, or primary vs cell line). While some of the methods reported involve the use of primary cultures of animal origin, the partial replacement still allows for a reduction and refinement in the use of animals. Currently, the main issue for most of the *in vitro* models mentioned in Table 2 is the limited number of chemicals tested. Some of the most promising tests will be described in more detail in the following paragraphs, either because they have been validated or are in the process of validation or because they measure the production of antibodies, which in animal models is considered the most predictive parameter.

Table 2. Key targets in chemical-induced immunosuppression and *in vitro* test opportunities

KEY IMMUNOLOGICAL TARGETS (TIER)	*IN VITRO* OPPORTUNITIES	CELL MODEL	REFERENCES
Bone marrow (Tier 1)	Human lympho-hematopoietic colony-forming assay for myelotoxicity (e.g. CFU-GM)	Human bone marrow and umbilical cord blood; rodent bone marrow	Pessina et al., 2003; 2005; 2010; Rich and Hall, 2005; Haglund et al., 2010
Leukotoxicity (Tier 2)	Cell viability (e.g., MTT, LDH release assay, flow cytometry)	Rodent splenocytes; human peripheral blood mononuclear cells	Hassan et al 2007; GIVIMP, 2018
Innate immunity (Tier 3)	NK cell activity	Rodent splenocytes; human peripheral blood mononuclear cells	Lebrec at al., 1995
	Monocytes/macrophages cytokines	Human peripheral blood mononuclear cells (e.g. whole blood assay); rodent splenocytes; cell lines (e.g. THP-1)	Langezaal et al., 2001; Langezaal et al., 2002; Carfì et al., 2007; Vessillier et al., 2015; Kimura et al., 2018
	Mast Cells/Basophils	Human basophils	McGowan et al., 2013
Cell mediated immunity (Tier 3)	T cell proliferation	Rodent splenocytes; human peripheral blood mononuclear cells	Lebrec at al., 1995; Carfì et al., 2007
	Mixed leukocyte response (MLR)	Rodent splenocytes; human peripheral blood mononuclear cells	Lebrec at al., 1995
	Cytotoxic T lymphocyte (CTL)	Rodent splenocytes; human peripheral blood mononuclear cells	Lebrec at al., 1995
	Cytokine production	Rodent splenocytes; human peripheral blood mononuclear cells (e.g. HWBCRA); human cell lines (e.g. Jurkat T cells)	Langezaal et al., 2001; Langezaal et al., 2002; Ullerås et al., 2005; Carfì et al., 2007; Ringerike et al., 2005; Stølevik et al., 2010; Kimura et al., 2018
	Transcriptomic profiles	Human peripheral blood mononuclear cells; human cell lines (e.g. Jurkat T cells)	Hochstenbach et al., 2010; Shao et al., 2014; Schmeits et al., 2015
	In vitro antigen presentation to T cells	Mouse cell lines (e.g. 3A9; Ch27B)	Lehmann and Williams, 2018

Humoral immunity (Tier 3)	B cell proliferation	Rodent splenocytes; human peripheral blood mononuclear cells	Carfì et al., 2007
	In vitro antibody production	Rodent splenocytes; human peripheral blood mononuclear cells	Keoper and Vohr, 2009; Lu et al., 2009; Collinge et al., 2010; Fischer et al., 2011

The table reports methods that have been proposed as alternatives to animals for the identification of immunotoxicants.

28. Among these assays, the HWBCRA has the advantage of comprising multiple cell types in their natural proportion and environment, allowing the evaluation of both monocyte and lymphocyte functions by using selective stimuli (Langezaal et al., 2001 and 2002), while “omics” techniques can provide additional mechanistic understanding and hold promise for the characterization of classes of compounds and prediction of specific toxic effects (Hochstenbach et al., 2010; Shao et al., 2014; Schmeits et al., 2015). The IL-2 Luc assay also allows high-throughput analysis (Kimura et al., 2018), which will greatly expand the opportunities for *in vitro* testing. The CFU-GM assay, the HWBCRA as a pyrogen test, and the IL-2 Luc assay have undergone validation for reproducibility and predictive capacity.

29. The whole blood assay provides a more physiological environment, as compared to isolated peripheral mononuclear cells, which may allow for a broader assessment of immune functions. In addition to cytokine production as in the HWBCRA, the whole blood assay can be used to address many other relevant immunological endpoints, including NK cell activity, lymphocyte proliferation, and antibody production. The cost of performing these assays depends on the endpoints, but it is overall relatively inexpensive compared to *in vivo* studies, and feasibility is high due to extensive use of this methodology (Hartung and Corsini, 2013). In the *in vitro* pyrogen test (i.e., HWBCRA), which is used in the same way as the Limulus test to analyze the possible presence of contamination of Gram negative bacteria or their remnants in drugs, and in the analysis of water and industrial raw materials, samples are incubated with fresh or cryopreserved human whole blood for the detection of the production of the proinflammatory cytokine IL-1β by enzyme-linked immunosorbent assay (Hartung and Wendel, 1995). In addition, the HWBCRA has also been adapted for immunotoxicity testing, to permit the potency testing of immunostimulants and immunosuppressants (Langezaal et al., 2001 and 2002). In this case, tested compounds are incubated in the presence of lipopolysaccharide (LPS) to activate monocytes or staphylococcal enterotoxin B to activate lymphocytes (mainly CD4+), and the release of IL-1β and IL-4, respectively, are subsequently measured by ELISA. Results are then expressed as IC50 values for immunosuppression, or SC(4) (stimulatory concentration resulting in a four-fold increase) values for immunostimulation, depending on the results observed. Thirty-one pharmaceutical compounds were used to 2222optimize and standardize the method. The *in vitro* results correlated well with *in vivo* data, and the test appears to reflect immunomodulation, meaning that both immunosuppression and immunostimulation can be detected. Results were reproducible (CV = 20 +/- 5%), and the method could be successfully transferred to another laboratory. A sensitivity of 67% and a specificity of 100% for the combined endpoints were calculated, where “sensitivity” refers to correctly identifying positive immunotoxicants, and “specificity” refers to correctly identifying negative immunotoxicants (reviewed in Hartung and Corsini, 2013).

30. Progress in *in vitro* testing for direct immunotoxicity includes validation of existing assays and selection of the assay (or combination of assays) that performs best, as described in the last section of this document (section VIII). In particular, the two luciferase assays that comprise the MITA are undergoing official validation studies; the IL-2 Luc assay that evaluates the effects of chemicals on the IL-2 promoter activity in response to stimulation with phorbol 12-myristate 13-acetate (PMA) and ionomycin (Io), and the IL-1 Luc assay that evaluates the effects of chemicals on the IL-1β promoter activity in response to stimulation with LPS. During the validation studies, the lead laboratory evaluated

the predictivity of the IL-2 Luc assay by examining 25 chemicals in the validation studies and 60 chemicals in the test data set. The predictivity of the IL-2 Luc assay was 75% in 25 chemicals and 82.5% in 60 chemicals, respectively. This predictivity is not optimal to predict immunotoxicity of chemicals as a stand-alone test method. Combination with other immune function tests, in particular myelotoxicity tests or leukotoxicity tests, will increase thepredictivity of the IL-2 Luc assay, as it is unable to detect myelotoxicity or antiproliferative effects (Kimura et al., 2014; Kimura et al., 2018).

31. In experimental animals, the TDAR is considered the “gold standard” to identify immunotoxic compounds (Luster et al., 1992; Lebrec et al., 2014). The TDAR has been the consensus choice for a functional endpoint to identify immunotoxicity hazard in most, if not all, regulatory guidelines, because the TDAR requires many of the cellular components of an immune response and thus, is a sensitive indicator of the overall immunotoxic potential of chemicals. Koeper and Vohr (2009), Lu et al. (2009), Collinge et al. (2010, 2020), and Fischer et al. (2011) reported the possibility to assess *in vitro* antibody production in the context of immunotoxicity. Antibody production provides a holistic summation of antigen processing, presentation and recognition, gene transcription and rearrangement, cell proliferation and differentiation, and ultimately, the production of antibodies, the effector molecules (Luebke, 2012). Koeper and Vohr (2009) and Fischer et al. (2011) used the *in vitro* antibody response (Mishell-Dutton culture) as an alternative to the existing animal tests to predict different immunosuppressants. Using this model, they were able to show that cell sources from both rats and mice were able to correctly predict all of 11 tested compounds and to clearly distinguish immunosuppressants from negative control substances. In another model proposed by Lu et al. (2009) a polyclonal immunoglobulin M (IgM) antibody-forming cell (AFC) response model to directly characterize immunotoxicity in primary mouse or human B cells was developed. CD40 ligand (CD40L) is used to activate B cells and to mimic T cell-dependent antibody responses *in vivo*. Antibody production, proliferation, and phenotypic changes characteristic of B cell activation as well as the plasma cell phenotype are measured. Two well-characterized immunotoxicants, arsenic and benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide, were tested. The novel model proposed by Lehmann and Williams (2018) to evaluate effects on antigen presentation, which is a key step in successful immunization is also of interest. Even if it is based on the use of two mouse T and B cell lines (3A9 hen egg lysozyme-specific I-A^k restricted T cell hybridoma cell line and the mouse Ch27 B lymphoma cell line), the method allows the evaluation of the effect of chemical exposure on several integrated events critical for immunization, including uptake, processing and presentation of antigen by antigen presenting cells, and antigen recognition and IL-2 production and secretion by T cells (Lehmann and Williams, 2018). However, these assays do not address all aspects of the humoral immune response, and it will be necessary to develop methods which assess the ability of B cells to undergo somatic hypermutation, affinity maturation, and class-switch. In the future, the gold standard TDAR might be replaced with *in vitro* coculture systems using iPSC-derived DCs, T cells, and B cells, or ultimately a microphysiological system of human immune system-on-a-chip consisting of these cell types to capture all aspects of humoral immunity (Miller at al., 2020). Further explorations of these models are recommended.

32. Finally, the BioMapTM Diversity Plus platform can be mentioned (Singer et al., 2019). The BioMap® Diversity Plus platform consists of several human primary cell-based assays modeling complex tissue and disease biology of organs (vasculature, immune system, skin, lung) and general tissue biology. Among the 12 systems, monocyte activation (readouts: MCP-1, VCAM-1, TM, TF, CD40, E-selectin, CD69, IL-8, IL-1α, M-CSF, sPGE2, SRB, sTNFα), T cell activation (readouts: MCP-1, CD38, CD40, E-selectin, CD69, IL-8, MIG, PBMC Cytotoxicity, Proliferation, SRB), and B and T cell autoimmunity (readouts: B cell Proliferation, PBMC Cytotoxicity, Secreted IgG, sIL-17A, sIL-17F, sIL-2, sIL-6, sTNFα) are likely to be relevant to identify chemical-induced immunotoxicity. However, its predictivity and reproducibility for chemicals inducing immunotoxicity, has not been demonstrated with a diverse set of environmental contaminants. Further explorations are recommended.

. Performance factors of *in vitro* assay(s)

33. To be appropriate for regulatory use, alternative *in vitro* assay(s) examining immunotoxicity should be characterized using the Reference Compound List (Table 3). The list is not exhaustive but is a compilation of environmental contaminants and drugs that have been shown to induce immunotoxicity in non-clinical studies and/or humans. When evaluating assay performance, to be in line with other validation studies, typically at least 20 compounds in total should be considered, but the required number of chemicals will depend on the specific test being performed and the regulatory guidance for the assay type. Compounds other than those in the reference list may be included, but their use should be justified according to the selection factors listed in section VII. The compounds in Table 3 have been selected from multiple classes, covering a wide range of biological and chemical modalities that have multiple immune targets. While it is beneficial to use compounds from multiple classes with different immune-related targets, some consideration of the cell type or response evaluated in the assay is necessary to interpret the sensitivity and specificity of each test. An approximate 2:1 ratio of positive to negative compounds, selected from different chemical classes, should be tested to ensure selectivity with the limited number of reference materials available.

34. A range of concentrations should be tested in each assay, and clearly described, as the same compound may be immuno-stimulatory or immunosuppressive depending on the concentration level. A certain degree of toxicity is expected at the highest concentration tested, which indicates that the chemical is doing something to the cells (similarly to the maximum tolerated dose in animal studies). However, significant toxicity should be avoided as it is difficult to ascertain whether the compound is toxic to a particular cell type, or if it decreases the proliferative capacity of those cells in the assay. When available, information on internal dose from *in vivo* exposure should be used to guide dose selection. Ideally, sensitivity and specificity would be 100%. However, these levels cannot be standardized and are dependent on the assay and chemicals being used. In general, the sensitivity to detect a positive compound in an assay(s), when applied on chemicals from the Reference Compound List (Table 3), should be at least 75%, with evidence of sufficient specificity (i.e., differentiating between true positives and true negatives, OECD GD286, 2018).

35. Inter-laboratory reproducibility and transferability between laboratories is required for the purpose of validation to establish an OECD test guideline if a particular assay is to be used in more than one laboratory. The minimum and maximum number of laboratories needed for a comprehensive assessment of the validity of the test method will depend on the type of test, the questions being addressed, and/or the overall amount of testing required of each laboratory. In many cases, three or four laboratories per test method may be an adequate number for an assessment of the inter-laboratory reproducibility (OECD GD34, 2005). Evaluation should also include assessments of accuracy and reproducibility over time. The performance characteristics of each assay, as well as the performance of the combined battery (if used) should be specified.

36. When interpreting results, the applicability domain and any limitations of the assay(s) should be taken into consideration (i.e., solubility, stability in culture media, metabolism). If the compound requires metabolism to exert its effects, but the system lacks metabolic enzymes, the results need to be considered in this context. Immune cells are generally considered to have low metabolic capacity, therefore, the use of S9 or other alternative metabolic activation systems should be considered by test

developers (Ooka et al, 2020). If not, it must be clearly stated that compounds that require metabolic activation fall outside the applicability domain of the test, and in case of negative results, one must be sure before classifying the compound as non-immunotoxic that the compound does not undergo bioactivation.

. Performance information of *in vitro* assay(s) to be provided to health authorities

37. To enable evaluation of an alternative assay(s) for use in immunotoxicity risk assessment for regulatory purposes, the following information should be provided: a detailed description of the predictive model including the *in vivo* endpoints for which it is trying to predict an outcome, and its use in the context of a tiered approach and integrated testing strategy. The *in vitro* model can consist of a single assay or a battery of assays together (a battery of tests measuring different immune endpoints is more likely to be predictive). If a battery of tests is used, each individual endpoint should be fully described with how the assessment of validity is made, including how the endpoints were selected.

38. The details of the prediction model used for determining positive and negative outcomes from the assay, including the borderline results and their interpretation should be presented for each assay. The model should correlate concentrations tested in the *in vitro* assays to the *in vivo* internal dose required to result in immunotoxicity in the species being predicted. For example, concentrations associated with immunotoxic effects should be interpreted in the context of expected *in vivo* exposure parameters such as C_{max} or AUC. If available, PBPK models can also inform concentration ranges.

39. The compound list used to qualify the assay performance should be presented. For purposes of establishing the predictive capacity (i.e., sensitivity, specificity and accuracy) of proposed test methods to be used by several laboratories, typically at least 20 compounds encompassing multiple chemical classes (examples listed in Table 3) in total should be tested by multiple laboratories, recognizing that the number of chemicals and testing laboratories will depend on the variability of the specific test being performed. In each laboratory, the chemicals should be tested in three independent runs performed with different cell batches on multiple days. Each run should consist of at least three concurrent replicates for each test chemical, negative, and positive control.

40. The calculation of the accuracy, sensitivity, and specificity values of the proposed *in vitro* test method for a single assay or battery of assays, should be equal to or better than the target values derived from the validated *in vivo* reference method(s). If a battery of assays is being proposed, the above information for each individual test method should be included. The combined accuracy, sensitivity, and specificity of the battery of assays should be greater than that of each of the individual assays. Typically, the sensitivity should be ≥75% (OECD 286). Any participating laboratory may misclassify an *in vivo* non-immunotoxic chemical as long as the final specificity of the test method is within the acceptable range.

41. The source of all reagents, biologic materials, and test compounds should be included. Test compound purity, stability and CAS number should be documented if available. The source/reference of all *in vivo* exposure data used for comparison should also be provided. Assays should be developed with the understanding that regulatory studies should generally be conducted in compliance with current

Good Laboratory Practice (GLP). Consultation with the relevant health authority(ies) is highly recommended to determine the level of validation needed.

. Selection factors for the reference compounds to be used in development of the in vitro assay(s)

42. The Reference Compound List (Table 3) contains environmental contaminants and drugs that have been shown to induce immunotoxicity in nonclinical studies and/or humans. The list includes representative chemicals from a number of classes of compounds that have been demonstrated to be immunotoxic (i.e. perfluoroalkyl substances, polycyclic aromatic hydrocarbons, organotins), known immunosuppressive therapeutics and other substances, but is not a comprehensive listing of all immunotoxicants. These compounds, as well as others, can be used to support qualification of an alternative assay or battery of assays. Many of the suggested reference compounds have been evaluated using tiered testing panels of *in vivo* assays in rodents (Luster et al., 1988; Vos and Van Loveren, 1989) and have been reported to suppress functional immune responses or modulate disease resistance. Modulation of observational measures, such as organ weights and cell subpopulations have also been described. It should be noted that while these compounds generally have immunosuppressive effects, depending on the exposure concentration and experimental design, some of these compounds may exhibit immunostimulatory effects, particularly when they act upon regulatory cells. In addition, positive control compounds may have different potency or target different cell types *in vitro* and an understanding of their mode of action and *in vivo* exposure parameters such as internal dose and metabolism may inform effects on specific cell populations. For a limited number, there is evidence from human epidemiology (i.e. ethanol, lead, and 2929 perfluorooctanoic acid) or clinical studies (cyclosporine A, dexamethasone, and diethylstilbestrol). In addition, there is a growing body of *in vitro* evidence supporting the immunotoxicity of these reference materials, and for many there are data with human cells or cell lines. The selected compounds target a variety of cell types and processes and will thus be useful to identify defined approaches for the *in vitro* assessment of immunotoxicity. The proposed negative compounds have been tested in a full battery of immune function assays in rodents as described in Luster et al (1988) and were negative under the conditions of those studies.

Table 3. Reference Compound List

Positive Controls	CAS Number	Reported Immune Targets
Aflatoxin B1*	1162-65-8	DTH, Cell Proliferation, Innate Immunity
2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine*	105650-23-5	Cell Proliferation, Antibody Response, Cytokine Production
Atrazine	1912-24-9	Cell Proliferation, Lymphoid Organ Weights, Clinical Pathology
Azathioprine*	446-86-6	Antibody Response, CTL
Benzidine*	92-87-5	Antibody Response, NK Cell Activity, Cell Proliferation, CTL
Benzo(a)pyrene*	50-32-8	Antibody Response, Cell Proliferation, Host Resistance, NK Cell Activity, Lymphoid Organ Weights
Cadmium Chloride	10108-64-2	Antibody Response, Cell Proliferation, Cytokine Production
Chloroquine*	54-05-7	Innate Immunity
Chrysene*	218-01-9	Antibody Response
Cyclophosphamide*	50-18-0	Antibody Response, Cell Proliferation, DTH, Lymphoid Organ Weights
Cyclosporine A	59865-13-3	Antibody Response
Deoxynivalenol	51481-10-8	Antibody Response, Cytokines
Dexamethasone	50-02-2	Macrophage Function, NK Cell Activity, Clinical Pathology, Lymphoid Organ Weights, Cytokines
2,4 Diaminotoluene*	95-80-7	Antibody Response, DTH
Dibromoacetic acid	631-64-1	Antibody Response, NK Cell Activity, CTL
Dichlorodiphenyltrichloroethane (DDT)	50-29-3	Antibody Response, Cell Proliferation, Cytokine Production, Lymphoid Organ Weights
Dideoxyadenosine	4097-22-7	Antibody Response, Cell Proliferation
Diethanolamine	111-42-2	Antibody Response, CTL, Clinical Pathology
Di(2-Ethylhexyl) Phthalate	117-81-7	Antibody Response, Innate Immunity, Cytokines
Diethylstilbestrol*	56-53-1	Antibody Response, Cell Proliferation, Cytokine Production
Dimethylbenz(a)-anthracene*	57-97-6	Antibody Response, Cell Proliferation, Lymphoid Organ Weights
Diphenylhydantoin*	630-93-3	NK Cell Activity, Lymphoid Organ Weights
Ethanol*	64-17-5	Antibody Response, NK Cell Activity, Cytokine Production
Ethyl carbamate*	51-79-6	Antibody Response, Cell Proliferation
Ginseng	50647-08-0	NK Cell Activity, Lymphoid Organ Weights

Glycidol	556-52-5	Antibody Response, NK Cell Activity, Clinical Pathology
Hexachlorobenzene*	118-74-1	Antibody Response
Hexachlorobiphenyl 153	35065-27-1	Antibody Response, CTL, DTH
γ-Hexachlorocyclohexane (Lindane)*	58-89-9	Cell Proliferation, CTL, NK Cell Activity
1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin	57653-85-7	Antibody Response, Cell Proliferation, DTH, Lymphoid Organ Weights
Indomethacin	95-83-0	Cell Proliferation
Lead acetate	6080-56-4	Antibody Response, DTH
Methadone Hydrochloride	1095-90-5	NK Cell Activity, Lymphoid Organ Weights
2-Methyoxyacetic Acid	625-45-6	Antibody Response, Cell Proliferation, Lymphoid Organ Weights
Morphine Sulfate	64-31-3	Antibody Response, NK Cell Activity, Cell Proliferation, Lymphoid Organ Weights
Mycophenolic Acid	24280-93-1	Antibody Response, Cell Proliferation
Nitrobenzene*	98-95-3	Antibody Response, NK Cell Activity, Cell Proliferation
n-Nitrosodimethylamine*	62-75-9	Antibody Response, NK Cell Activity, Cell Proliferation, DTH
m-Nitrotoluene*	99-08-1	Antibody Response, Cell Proliferation, DTH
Ochratoxin A	303-47-9	NK Cell Activity, Lymphoid Organ Weights
Parathion	56-38-2	Host Resistance, Antibody Response
Pefluorooctanoic Acid	335-67-1	Antibody Response, NK Cell Activity, Cytokine Production
Perfluorooctanesulfonic acid	1763-23-1	Antibody Response, NK Cell Activity, Cell Proliferation, Cytokine Production
3,3',4,4',5-Pentachlorobiphenyl*	57465-28-8	Antibody Response, NK Cell Activity, Cell Proliferation, Lymphoid Organ Weights
Pentachlorophenol*	87-86-5	Antibody Response, NK Cell Activity, Cell Proliferation
Prednisolone	50-24-8	Macrophage Function, NK Cell Activity, Clinical Pathology, Lymphoid Organ Weights, Cytokines
Propanil*	709-98-8	Antibody Response, NK Cell Activity, CTL, Cell Proliferation, Cytokine Production
Tetrabromobisphenol A	79-94-7	Cytokines, Innate Immunity
2,3,7,8-Tetrachlorodibenzo-p-dioxin	1746-01-6	Antibody Response, CTL, Thymus Weight,
δ 9- Tetrahydrocannabinol	1972/8/3	Antibody Response, Cell Proliferation, Cytokine Production
Thalidomide	50-35-1	Antibody Response, CTL, Clinical Pathology
Tributyltin	56-24-6	Antibody Response, Lymphoid Organ Weights, Cytokine Production
Tributyltin Chloride	1461-22-9	Antibody Response, NK Cell Activity, CTL, Cell Proliferation, Lymphoid Organ Weights
Tributyltin Oxide	56-35-9	Antibody Response, NK Cell Activity, CTL, Cell Proliferation, DTH, Lymphoid Organ Weights

		Negative Controls
Chloramine	10599-90-3	
4-Chloro-o-phenylenediamine	95-83-0	
Dichloroacetic acid	79-43-6	
Methyl carbamate	598-55-0	
Nitrofurazone	59-87-0	
Oxymethalone	434-07-1	
Patulin	149-29-1	
Sodium Bromate	7789-38-0	
Sodium Chlorite	7758-19-2	

*Denotes compounds which require metabolic activation for immunotoxicity to manifest. Because there is a large body of literature for each of these individual compounds the reader is referred to compilations of information or data on immuntoxicity that review these effects including: Cohen et al 2000; Corsini and van Loveren, 2015; Descotes 2004; Dewhurst et al 2015; House et al 2007; Kaplan et al 2019; Kimura et al 2020; Luster et al 1988; 1992; Tryphonas et al 2005; Vohr 2005, WHO 1996

. *In vitro* immunotoxicological assessments using combinations of cell types or cell lines

8.1. Lessons from immunotoxicological assessments using rodents

43. The use of tiered testing panels of *in vivo* assays in rodents has been the most common methodology for assessment of immunotoxicity since the inception of the discipline (Luster et al., 1988; Van Loveren and Vos, 1989). The *in vivo* tiered approach proposed by the National Toxicology Program at the NIH contains both screening assays to detect immunologic effects (Tier I) and a comprehensive suite of assays to provide an in-depth assessment of immune function and host resistance endpoints (Tier II), as listed in Table 4 (modified from (Luster 1998)). Chemicals are judged as immunotoxicants based on whether they produced a significant dose-response effect ($p<0.05$) in a measure of a functional immune response (rather than an observational measure such as a change in body or organ weights), or if they significantly altered two or more test results at the highest dose of chemical tested ($p<0.05$). Based on the ability of various immune tests to predict increased susceptibility in disease resistance assays, Luster et al. (1992), demonstrated that: 1) a number of the immune tests provided a relatively high association with changes in host resistance (i.e., > 70%), such as the TDAR, delayed hypersensitivity response (DHR), cell surface immunophenotyping markers, and CTL assay. In contrast, several of the tests, such as leukocyte counts and lymphoproliferative response to LPS were poor predictors, with concordance values of approximately 50%; and 2) the combination of two immune tests significantly increased the predictive value from that obtained using individual tests. Pair-wise combinations which included either the plaque forming cell (PFC) response (a TDAR endpoint), cell surface immunophenotyping markers, or DHR gave consistently higher concordances and combinations of two or three immune tests involving these measures could give more than 90% concordance with effects on disease resistance.

44. Several regulatory guidelines or guidance documents have since been developed for the assessment of immunotoxicity of pharmaceuticals or industrial products (e.g. ICH S8). The majority of these suggest that standard toxicity studies, combined with trigger-based functional immune testing, represent an effective approach to evaluate immunotoxic potential (Boverhof et al., 2014). Among the various functional immune tests, the TDAR has been the consensus choice for a functional endpoint to identify immunotoxicity hazard in most, if not all, regulatory guidelines (Fischer et al., 2011; Koeper and Vohr, 2009). These approaches affirm the need to evaluate multiple aspects of the immune response to accurately predict immunotoxicity.

Table 4. *In vivo* tiered approach for detecting immune alterations in rodents

	Procedures
Tier I	
	Hematology
	Organ Weights – Spleen, thymus
	Cellularity – Spleen and bone marrow
	Histology of lymphoid organs
	IgM antibody plaque-forming cells (PFCs)
	Lymphocyte blastogenesis
	Natural killer cell activity Surface markers (peripheral or tissue immunophenotyping)
Tier II	
	IgG antibody PFC response
	Cytotoxic T lymphocyte cytolysis
	Delayed hypersensitivity response (DHR) Macrophage/neutrophil functional assays
	Host resistance (syngeneic tumors, bacterial, viral, and parasite models)

2. *In vitro* immunotoxicological assessments using primary cells

45. To maximize relevance to human immunotoxicology and to avoid inter-species extrapolation, it is recommended to use human cells for all *in vitro* tests. Although the use of primary human cells, which are available from peripheral blood or from buffy coats, are of highest clinical relevance, consideration can be given to the use of sufficiently well-characterized and largely used cell lines for certain aspects of the test systems (Gennari et al., 2005). Although continuous cell lines are not physiologically equivalent, many have proven to be valid surrogates but require appropriate characterization to ensure they are accurately recapitulating the normal immunological responses and functions (Boverhof et al., 2014). However, as cell lines are subject to genetic drift over excessive passages, it is important to control the stability of the cell line used. Another aspect that should not be disregarded, is that some compounds' immunotoxicity, especially drug-induced hypersensitivity, has been associated with specific HLA-types (Fan et al. 2017). Thus, the haplotype of the cell lines used should also be considered. If human cells are not available or human cell lines cannot be used, the use of non-human cells may be considered if the method proves to have acceptable predictive capacity or if the same response is expected both qualitatively and/or quantitatively, similar to what is expected using animals in toxicology (Corsini and Roggen, 2017; Lankveld et al., 2010).

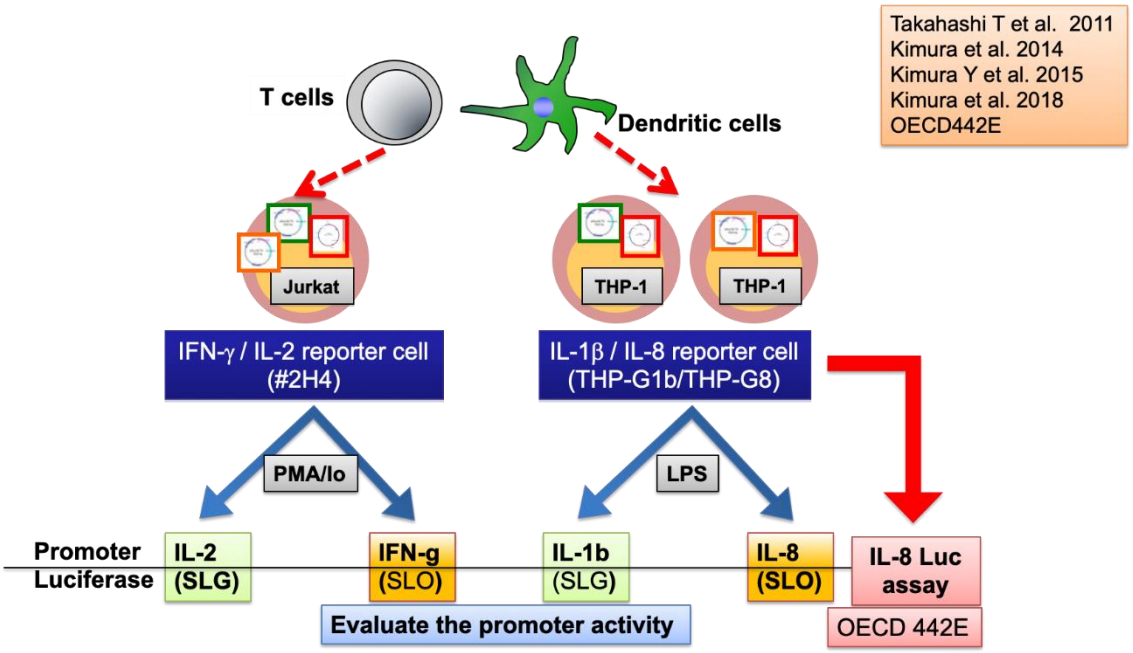
3. *In vitro* immunotoxicological assessments using cell lines

46. Most of the *in vitro* immune tests described in section IV use animals or human samples to obtain immune cells. The use of primary human cells for *in vitro* tests can entail challenges and

inconsistencies due to constraints around securing human samples, large variability among donors, poor reproducibility, inability to control donors' disease state and environmental factors, or the need to differentiate progenitor cells. Although continuous cell lines are not the physiological equivalent of primary cells or animals, many have proven to be valid surrogates if appropriately characterized to ensure they are accurately recapitulating the normal immunological responses and functions (Boverhof et al., 2014).

47. Almost a decade ago, a luciferase reporter assay system, MITA, was established to evaluate the effects of chemicals on the human immune system using stable reporter cell lines instead of primary human cells (Figure 1). The MITA is composed of 3 stable reporter cell lines: 1) 2H4 derived from Jurkat cells containing luciferase genes regulated by the IL-2, IFNγ, and G3PDH promoters (Saito et al. 2011), 2) THP-G8 cells derived from THP-1 cells containing luciferase genes regulated by the IL-8 and G3PDH promoters (Takahashi et al., 2011), and 3) THP-G1b cells derived from THP-1 cells containing luciferase genes regulated by the IL-1β and G3PDH promoters (Kimura et al. 2018). Using these cell lines, it has been demonstrated that MITA can reflect the effects of immunosuppressive drugs on cytokine expression by T cells or macrophages, and that the evaluation of drugs using the MITA was consistent with those obtained using the mother cell lines (Jurkat and THP-1 cells) or by using stimulated human whole blood cells ((Kimura et al. 2014) and Table 5).

Figure 1. Multi-ImmunoTox Assay (MITA)



MITA is an approach to detect immunotoxic chemicals using a combination of 3 luciferase reporter cell lines. 2H4 cells are used to evaluate the effects of chemicals on IL-2 and IFNγ promoter activity under stimulation with PMA/ionomycin. THP-G1b cells evaluate IL-1β promoter activity under stimulation with LPS, while THP-G8 cells evaluate IL-8 promoter activity under stimulation with LPS. The IL-8 Luc assay (OECD 442E) is an *in vitro* skin s35ensitization test that determines the induction of IL-8 promoter activity by chemical-treated THP-G8 cells. A modified version of the MITA (mMITA) is composed of the MITA and the IL-8 Luc assay (OECD 442E). In support of the initiative to increase the number of available *in vitro* immunotoxicity assays, two of the four reporter assays comprising MITA (IL-2 Luc and IL-1 Luc) are currently undergoing validation studies (Kimura et al. 2020).

48. When the performance of the IL-2 reporter assay was evaluated by examining immunosuppressive drugs whose effects in humans have been well-established (Table 5, reviewed by Allison 2000), the results demonstrated that the majority of the known agents caused reductions in IL-2 transcription including tacrolimus, cyclosporine A, dexamethasone, chloroquine, minocycline, sulfasalazine, ruxolitinib, tofacitinib, and baricitinib. However, decreased IL-2 transcription was not observed with several immunosuppressants whose mechanism of action is dependent on the inhibition of DNA synthesis or anti-proliferative effects on T cells, such as rapamycin, cyclophosphamide, azathioprine, mycophenolic acid, mizoribine, and methotrexate. Thus, it is critical to define the applicability domain for any proposed *in vitro* testing strategy to understand if the assay can appropriately assess the effects of certain compounds (e.g., those that require metabolism or target particular pathways).

Table 5. Evaluation of immunomodulatory drugs by MITA (modified from the original report by Kimura et al 2014 and Kimura et al 2018)

Principal Mechanism of Action	Drug	Effects of Transcriptional Activity			
		IL-2	IFNγ	IL-1β	IL-8
					IMMUNOSUPPRESSIVE DRUGS
Regulation of gene expression	Dexamethasone	↓	—	↓	↓
Kinase and phosphatase inhibitors	Cyclosporin A	↓	↓	—	—
	Tacrolimus	↓	↓	—	—
	Rapamycin	↑	—	—	—
JAK inhibitors	Ruxolitinib	↓	N	N	N
	Tofacitinib	↓	N	N	N
	Baricitinib	↓	N	N	N
Nrf-2 inhibitor	Dimethyl fumarate	↓	N	N	N
PDE4 inhibitor	Apremilast	↑	N	N	N
Alkylation	Cyclophosphamide	—	—	—	—
Inhibition of de novo purine synthesis	Azathioprine	—	—	—	—
	Mycophenolic acid	↑	↑	—	—
	Mizoribine	—	—	↑	↑
Inhibition of pyrimidine and purine synthesis	Methotrexate	↑	↑	—	—
					OFF-LABEL IMMUNOSUPPRESSIVE DRUGS
	Sulfasalazine	↓	↓	↓	↓
	Colchicine	↑	—	↑	↑
	Chloroquine	↓	—	—	↓
	Minocycline	↓	↓	—	—
	Nicotinamide	↑	—	↓	↓
					NON-IMMUNOMODULATORY DRUGS
	Acetaminophen	↑	—	—	↑
	Digoxin	↓	↓	—	—
	Warfarin	↑	—	↓	↓

JAK=Janus kinase, Nrf-2=nuclear factor erythroid 2-related factor 2, PDE4=phosphodiesterase 4

↑=stimulation, ↓=suppression, —=no effect, N=not tested

4. Combination of *in vitro* assays and clustering analysis has the potential to increase predictivity

49. As stated throughout this document it is likely that several assays will need to be used in combination to increase the ability to predict immunotoxicity using in vitro tests. For example, while the predictivity of the IL-2 Luc assay is high when testing T cell-targeting chemicals, combination with other immune function tests, in particular myelotoxicity tests or leukotoxicity tests, will be essential to cover global immunotoxicity. Methods for evaluating myelotoxicity *in vitro* using bone marrow culture systems (Haglund et al., 2010; Pessina et al., 2003) could be combined with the MITA to provide a more complete assessment. In addition, a novel luciferase assay designated as the IL-2 Luc LTT has recently been developed, that can detect the antimitotic effects of chemicals and can correctly identify immunosuppressive chemicals that were negative in the IL-2 assay (Kimura et al., 2021). Thus, the combination of the IL-2 Luc assay and the IL-2 Luc LTT can overcome some of the limitations of the IL-2 Luc assay alone (Kimura et al., 2021) and may be even more predictive when combined with a myelotoxicity assay.

50. Another example of this combinatorial approach is the addition of the MITA (IL-2 reporter assay in response to PMA/ionomycin [IL-2LA] plus the IL-8 reporter assay in response to LPS [IL-8+LPS]) with the IL-8 reporter assay for skin s37ensitization test (IL-8-LPS, OECD 442E), denoted as the modified MITA (mMITA). Sixty chemicals with well-known immunotoxic profiles were examined by the mMITA and were classified based on multiple approaches into a final group of six clusters with distinct characteristics. **Cluster 1:** chemicals that preferentially suppressed IL-8+LPS and showed a negative IL-8-LPS (preferential IL-8+LPS suppression); **Cluster 2:** those that suppressed IL-2LA and showed a positive IL-8-LPS, but did not affect IL-8+LPS (IL-2LA suppression and IL-8-LPS(+)); **Cluster 3:** those that suppressed both IL-2LA and IL-8+LPS and showed a positive IL-8-LPS (IL-2LA and IL-8+LPS suppression and IL-8-LPS(+)); **Cluster 4:** those that did not suppress either IL-2LA or IL-8+LPS and showed a negative IL-8-LPS (all negative); **Cluster 5:** those that suppressed both IL-2LA and IL-8+LPS but showed a negative IL-8-LPS (IL-2LA and IL-8+LPS suppression); and **Cluster 6:** those that preferentially suppressed IL-2LA and showed a negative IL-8-LPS (preferential IL-2LA suppression). The power of this approach is highlighted by the fact that although there were less well characterized chemicals tested, their potential *in vivo* effects could be inferred by comparing to well understood medicinal drugs that landed in the same cluster, such as sulfasalazine for Cluster 1, chloroquine for Cluster 2, colchicine for Cluster 3, acetaminophen for Cluster 4, dexamethasone for Cluster 5, and cyclosporine A and FK506 for Cluster 6 (Kimura et al., 2014; Kimura et al., 2018).

51. Overall, it is crucial to examine the correlation between *in vitro* assays and their *in vivo* effects. As demonstrated using the example of the mMITA, the clustering of chemicals using only three parameters may be inadequate to detect every aspect of their immunotoxic effects. However, such clustering in the context of a battery of complementary assays can be a first step to profile the immunotoxicity of chemicals.

. Discussion and Conclusion

52. Besides the presence of specific immune organs, immune cells are an integral part of other systems including the respiratory, dermal, gastrointestinal, neurological, cardiovascular, reproductive, hepatic, and endocrine systems. Consequently, exposure to immunotoxic compounds can have detrimental effects on the response to both communicable and non-communicable diseases. It is therefore important to understand the immunotoxic potential of xenobiotics and the risk they pose to humans.

53. Current practices in overt immunotoxicity testing are still varied and employ either experimental animals or humans themselves, while *in vitro* testing has predominantly been used to study specific mechanisms of immunotoxicity. There are several regulatory guiding principles in immunotoxicology, including IPCS/WHO Guidance for Immunotoxicity Risk Assessment for Chemicals (Harmonization project No. 10). All of these guidelines involve animal testing, and the current practice of risk assessment of chemical exposure is based to a large extent on animal testing.

54. While a number of non-animal-based testing methods for immunotoxicity have been published, not many have reached the stage of validation and acceptance. In accordance with the promotion of alternative testing methods and the global desire to reduce the use of laboratory animals, the purpose of this document is to provide the state-of-the-art knowledge for non-animal testing in immunotoxicology, and a way forward. Currently, a tiered approach is the most appropriate means to assess immunotoxicity, as described above. In the proposed tiered approach pre-screening for direct immunotoxicity *in vitro* begins by evaluating myelotoxicity (Tier 1). Compounds capable of damaging or destroying bone marrow cells will most likely have immunotoxic effects, as all immune cells derive from a common precursor located in the adult bone marrow. If compounds are not myelotoxic, they should be tested for direct leukotoxicity (Tier 2). Compounds should then be tested for immunotoxicity using various approaches, such as TDAR, lymphocyte proliferation assay, MLR, and NK cell assay (Tier 3).

55. It is likely that multiple assays will be required to define immunotoxicants because of the complexity and varied components of the immune system (e.g,. innate or adaptive immune responses). The combined use of several in vitro assays in IATAs or defined approaches, such as those used for skin sensitization, should increase the ability to predict immunotoxicity over an individual assay. It is critical to define whether the parent compound or a metabolite is mediating the immunotoxic effects, as immune cells have limited capacity to metabolize drugs. Even upon determining that a compound is immunotoxic, it is important to remember that there are interactions between the immune, nervous, and endocrine systems such that immunity is regulated by various neurotransmitters and hormones, which could also result in sex differences in the sensitivity to immunotoxicity.

56. At present there is no consensus on which assays to use, or how, and there are no OECD test guidelines to detect chemical immunotoxicity *in vitro*. Any validated test should therefore be part of a larger set of different assays that preferably covers all types of immunotoxicity including immunosuppression, sensitization, and autoimmunity. The MITA described in Section VIII, is one example of a combination of in vitro assays which may be used to predict the immunotoxicity of chemicals (and could be considered a potential future Tier 3 approach), with the potential for including additional endpoints which address gaps in the assessment of innate and humoral immunity. Comparing unknown chemicals to known immunotoxicants through clustered analyses such as those used in the

mMITA, information can be obtained on potential cellular targets and mode of action. Considering the incredible progress in 3D models, with engineered immune tissues and organs, such as bone marrow, thymus, lymph nodes and spleen being described, and microfluidic body-on-a-chip, it is reasonable to think that in the future it will be possible to assess immunotoxicity in an integrated model of the whole human immune system. While additional studies are certainly needed to define the possibility of identifying immunotoxic substances without the use of animals, the road is being paved for the use of integrated testing strategies that together hold the promise of being able to adequately predict the immunotoxic action of chemicals *in vitro*.

Acknowledgments: We thank the International Cooperation on Alternative Test Methods (ICATM) members, especially EURL ECVAM and NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) for their support and recommendations. This work was partly supported by the Health and Labor Science Research Grant, Japan.

. References

Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Leite, S.B., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tähti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A., Zaldivar, J.M., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85(5), 367-485.

Allison, A.C., 2000. Immunosuppressive drugs: the first 50 years and a glance forward. Immunopharmacology 47(2-3), 63-83.

Belkaid Y, Hand T. Role of the Microbiota in Immunity and Inflammation. Cell. 2014 Mar 27; 157(1): 121–141. doi: 10.1016/j.cell.2014.03.011

Bell SM, Phillips J, Sedykh A, Tandon A, Sprankle C, Morefield SQ, Shapiro A, Allen D, Shah R, Maull EA, Casey WM, Kleinstreuer NC. An Integrated Chemical Environment to Support 21st-Century Toxicology. Environ Health Perspect. 2017 May 25;125(5):054501.

Boverhof, D.R, Ladics, G., Luebke, B., Botham, J., Corsini, E., Evans, E., Germolec, D., Holsapple, M., Loveless, SE., Lu, H., van der Laan, JW., White, K.L. Jr., Yang, Y., 2014. Approaches and considerations for the assessment of immunotoxicity for environmental chemicals: a workshop summary. Regul. Toxicol. Pharmacol. 68(1), 96-107Carfi', M., Gennari, A., Malerba, I., Corsini, E., Pallardy, M., Pieters, R., Van Loveren, H., Vohr, H.W., Hartung, T., Gribaldo, L., 2007. In vitro tests to evaluate immunotoxicity: a preliminary study. Toxicology 229(1-2), 11-22.

Casati, S., Aschberger, K., Barroso, J., Casey, W., Delgado, I., Kim, T.S., Kleinstreuer, N., Kojima, H., Lee, J.K., Lowit, A., Park, H.K., Régimbald-Krnel, M.J., Strickland, J., Whelan, M., Yang, Y., Zuang, V., 2018. Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods. Arch. Toxicol. 92(2), 611-617.

Chary, A., Serchi, T., Moschini, E., Hennen, J., Cambier, S., Ezendam, J., Blomeke, B., Gutleb, A.C., 2019. An in vitro coculture system for the detection of sensitization following aerosol exposure. ALTEX 36(3), 403-418.

Cohen, MD., Zelikoff, JT., Schlesinger, RB. eds.,2000 Pulmonary Immunotoxicology, Springer, Boston, pp 465.

Collinge, M., Cole, S.H., Schneider, P.A., Donovan, C.B., Kamperschroer, C., Kawabata, T.T., 2010. Human lymphocyte activation assay: an in vitro method for predictive immunotoxicity testing. J. Immunotoxicol. 7(4), 357-66.

Collinge, M., Schneider, P., Li, D., Parish, S., Dumont, C., Freebern, W., Ghanime, J., Guimont-Derochers, F., Langenkamp, A., Lebron, J., Li, N., Marban, C., Plitnick, L., Wheeler, J. 2020 Cross-company evaluation of the human lymphocyte activation assay. J Immunotoxicol 17(1):51-58.

Corsini, E., Roggen, E.L., 2009. Immunotoxicology: opportunities for non-animal test

development. Altern. Lab. Anim. 37(4), 387-97.
Corsini, E. and van Loveren, H. eds.,2015 Molecular Immunotoxicology, Wiley-VCH, Weinhelm, pp 334.
Corsini, E., Roggen, E., 2017. Overview of in vitro assessment of immunotoxicity. Curr. Opin. Toxicol. 5, 13-18.
Corsini, E., Engin, A.B., Neagu, M., Galbiati, V., Nikitovic, D., Tzanakakis, G., Tsatsakis, A.M., 2018. Chemical-induced contact allergy: from mechanistic understanding to risk prevention. Arch. Toxicol. 92(10), 3031-3050.
de Ávila, R.I., Lindstedt, M., Valadares, M.C., 2019. The 21st Century movement within the area of skin sensitization assessment: From the animal context towards current human-relevant in vitro solutions. Regul. Toxicol. Pharmacol. 108. doi: 10.1016/j.yrtph.2019.104445.
Dean JH, Luster MI, Boorman GA. Methods and approaches for assessing immunotoxicity: an overview. Environ Health Perspect. 1982 Feb;43:27-9.
Descotes, J. ed. 2004 Immunotoxicology of Drugs and Chemicals: An Experimental and Clinical Approach, Elsevier, pp 398.
Dewhurst, I., Koshy, L., Samuels, S., Shillaker, D., 2015. Retrospective analysis of the immunotoxic effects of plant protection products as reported in the Draft Assessment Reports for their peer review at EU level. EFSA supporting publication 2015:EN-782, 104 pp.
Fan, Wen-Lang, Shiao, Meng-Shin, Hui, Rosaline Chung-Yee, Su, Shih-Chi, Wang, Chuang-Wei, Chang,Ya-Ching, Chung, Wen-Hung, 2017. HLA Association with Drug-Induced Adverse Reactions
J Immunol Res. 3186328. doi: 10.1155/2017/3186328
Fischer, A., Koeper, L.M., Vohr, H.W., 2011. Specific antibody responses of primary cells from different cell sources are able to predict immunotoxicity in vitro. Toxicol. in Vitro 25, 1966-1973.
French, A., Yang, C.T., Taylor, S., Watt, S.M., Carpenter, L., 2015. Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. Stem Cells Dev. 24(9), 1082-1095.
Galbiati, V., Mitjans, M. & Corsini, E., 2010. Present and future of in vitro immunotoxicology in drug development. J. Immunotoxicol. 7(4), 255-267.
Gennari, A., Ban, M., Braun, A., Casati, S., Corsini, E., Dastych, J., Descotes, J., Hartung, T., Hooghe-Peters, R., House, R., Pallardy, M., Pieters, R., Reid, L., Tryphonas, H., Tschirhart, E., Tuschl, H., Vandebriel, R., Gribaldo, L., 2005. The Use of In Vitro Systems for Evaluating Immunotoxicity: The Report and Recommendations of an ECVAM Workshop. J. Immunotoxicol. 2(2), 61-83.
Germolec, D., Luebke, R., Rooney, A., Shipkowski, K., Vandebriel, R., van Loveren, H., 2017. Immunotoxicology: A brief history, current status and strategies for future immunotoxicity assessment. Curr. Opin. Toxicol. 5, 55-59.
Gosselin, E.A., Eppler, H.B., Bromberg, J.S., Jewell, C.M., 2018. Designing natural and synthetic immune tissues. Nat. Mater. 17(6), 484-498.
Haglund, C., Aleskog, A., Håkansson, L.D., Höglund, M., Jacobsson, S., Larsson, R., Lindhagen. E., 2010. The FMCA-GM assays, high throughput non-clonogenic alternatives to CFU-GM in preclinical hematotoxicity testing. Toxicol. Lett. 194(3), 102-107.
Hartung, T., Corsini, E., 2013. Immunotoxicology: challenges in the 21st century and in vitro opportunities. ALTEX 30(4), 411-426.
Hartung, T., Wendel, A., 1995. Detection of Pyrogens using human whole blood.ALTEX 12(2):70-75.
Hassan, S.B., Haglund, C., Aleskog, A., Larsson, R., Lindhagen, E., 2007. Primary lymphocytes as predictors for species differences in cytotoxic drug sensitivity. Toxicol.

in Vitro 21(6):1174-1181.
Hinton, D.M., 2000. US FDA "Redbook II" immunotoxicity testing guidelines and research in immunotoxicity evaluations of food chemicals and new food proteins. Toxicol. Pathol. 28(3), 467-478.
Hochstenbach, K., van Leeuwen, D.M., Gmuender, H., Stølevik, S.B., Nygaard, U.C., Løvik, M., Granum, B., Namork, E., van Delft, J.H., van Loveren, H., 2010. Transcriptomic profile indicative of immunotoxic exposure: in vitro studies in peripheral blood mononuclear cells. Toxicol. Sci. 118(1), 19-30.
House, RV., Luebke, R. and Kimber, I., eds., 2007 Immunotoxicology and Immunopharmacology, CRC Press Boca Raton, pp 672.
International Conference on Harmonisation (ICH) Tripartite Guideline S8, Immunotoxicity Studies for Human Pharmaceuticals, 2005.
Kaplan, B., Sulentic, C., Haggerty, .H, Holsaplle, M., Kaminski, N., Toxic Responses of the Immune System, Casarett and Doull's Toxicology: The Basic Science of Poisons 9th Edition, Klaassen D, ed. McGraw Hill, New York, 2019. pp 633-718.
Karmaus, A.L., Zacharewski, T.R., 2015. Atrazine-Mediated Disruption of Steroidogenesis in BLTK1 Murine Leydig Cells.Toxicol. Sci. 148(2), 544-554.
Kimura, Y., Fujimura, C., Ito, Y., Takahashi, T., Aiba, S., 2014. Evaluation of the Multi-ImmunoTox Assay composed of 3 human cytokine reporter cells by examining immunological effects of drugs. Toxicol. in Vitro 28(5), 759-768.
Kimura, Y., Fujimura, C., Ito, Y., Takahashi, T., Terui, H., Aiba, S., 2018. Profiling the immunotoxicity of chemicals based on in vitro evaluation by a combination of the Multi-ImmunoTox assay and the IL-8 Luc assay. Arch. Toxicol. 92(6), 2043-2054.
Kimura, Y., Yasuno, R., Watanabe, M., Kobayashi, M., Iwaki, T., Fujimura, C., Ohmiya, Y., Yamakage, K., Nakajima, Y., Kobayashi, M., Mashimo, N., Takagi, Y., Omori, T., Corsini, E., Germolec, D., Inoue, T., Rogen, EL., Kojima, H., Aiba, S., 2020. An international validation study of the IL-2 Luc assay for evaluating the potential immunotoxic effects of chemicals on T cells and a proposal for reference data for immunotoxic chemicals. Toxicol. in Vitro 66. doi: 10.1016/j.tiv.2020.104832.
Kitayama, S., Zhang, R., Liu, T.Y., Ueda, N., Iriguchi, S., Yasui, Y., Kawai, Y., Tatsumi, M., Hirai, N., Mizoro, Y., Iwama, T., Watanabe, A., Nakanishi, M., Kuzushima, K., Uemura, Y., Kaneko, S., 2016. Cellular adjuvant properties, direct cytotoxicity of re-differentiated valpha24 invariant NKT-like cells from human induced pluripotent stem cells. Stem Cell Reports 6(2), 213-227.
Koeper, L. M., Vohr, H. W., 2009. Functional assays are mandatory for correct prediction of immunotoxic properties of compounds in vitro. Food Chem. Toxicol. 47(1), 110-118.
Langezaal, I., Coecke, S., Hartung, T., 2001. Whole blood cytokine response as a measure of immunotoxicity. Toxicol. in Vitro 15(4-5), 313-318.
Langezaal, I., Hoffmann, S., Hartung, T., Coecke, S., 2002. Evaluation and prevalidation of an immunotoxicity test based on human whole-blood cytokine release. Altern. Lab. Anim. 30(6), 581-595.
Lankveld, D.P., Van Loveren, H., Baken, K.A., Vandebriel, R.J., 2010. In vitro testing for direct immunotoxicity: state of the art. Methods Mol. Biol. 598, 401-423.
Lebrec, H., Roger, R., Blot, C., Burleson, G. R., Bohuon, C., Pallardy, M., 1995. Immunotoxicological investigation using pharmaceutical drugs. In vitro evaluation of immune effects using rodent or human immune cells. Toxicology 96(2), 147-156.
Lebrec, H., Molinier, B., Boverhof, D., Collinge, M., Freebern, W., Henson, K., Mytych, DT., Ochs, HD., Wange, R., Yang, Y., Zhou, L., Arrington, J., Christin-Piché, MS., Shenton, J., 2014. The T-cell-dependent antibody response assay in nonclinical studies of pharmaceuticals and chemicals: study design, data analysis, interpretation.Regul. Toxicol. Pharmacol. 69(1), 7-21.
Lehmann, D.M., Williams, W.C., 2018. Development and utilization of a unique in vitro antigen

presentation co-culture model for detection of immunomodulating substances. Toxicol. in Vitro 53, 20-28.
Luebke, R., 2012. Immunotoxicant screening and prioritization in the twenty-first century. Toxicol Pathol. 40(2), 294-299.
Luster, M.I., Munson, A.E., Thomas, P.T., Holsapple, M.P., Fenters, J.D., White, K.L. Jr., Lauer, L.D., Germolec, D.R., Rosenthal, G.J., Dean, J.H., 1988. Development of a testing battery to assess chemical-induced immunotoxicity: National Toxicology Program's guidelines for immunotoxicity evaluation in mice. Fundam. Appl. Toxicol. 10(1), 2-19.
Luster, M.I., Portier, C., Pait, D.G., White, K.L. Jr., Gennings, C., Munson, A.E., Rosenthal, G.J., 1992. Risk assessment in immunotoxicology. I. Sensitivity and predictability of immune tests. Fundam. Appl. Toxicol. 18(2), 200-210.
Luster, M.I., Portier, C., Pait, D.G., Rosenthal, G.J., Germolec, D.R., Corsini, E., Blaylock, B.L., Pollock, P., Kouchi, Y., Craig, W., White, K.L., Munson, A.E., Comment, C.E., 1993.Risk assessment in immunotoxicology. II. Relationships between immune and host resistance tests. Fundam. Appl. Toxicol. 21(1), 71-82.
Luster, M.I., Gerberick, GF., 2010. Immunotoxicology testing: past and future. Methods Mol. Biol. 598, 3-13.
Lu, H., Crawford, RB., North, CM., Kaplan, BL., Kaminski, NE., 2009. Establishment of an immunoglobulin m antibody-forming cell response model for characterizing immunotoxicity in primary human B cells.Toxicol. Sci. 112(2), 363-373.
McGowan EC, Saini S., 2013. Update on the performance and application of basophil activation tests. Curr Allergy Asthma Rep. 13:101
Miller et al. 2020: Engineering Microphysiological Immune System Responses on Chips. Trends in Biotechnology Volume 38, Issue 8, 2020, pp 857-872.
Mizoguchi, I., Ohashi, M., Chiba, Y., Hasegawa, H., Xu, M., Owaki, T., Yoshimoto, T.. 2017. Prediction of chemical respiratory and contact sensitizers by OX40L expression in dendritic cells using a novel 3D coculture system. Front. Immunol. 8. doi: 10.3389/fimmu.2017.00929.
Naidenko OV, Andrews DQ, Temkin AM, Stoiber T, Uche UI, Evans S, Perrone-Gray S. Investigating Molecular Mechanisms of Immunotoxicity and the Utility of ToxCast for Immunotoxicity Screening of Chemicals Added to Food. International Journal of Environmental Research and Public Health. 2021; 18(7):3332. https://doi.org/10.3390/ijerph18073332
OECD Guidance document No. 286, 2018, Guidance Document on Good In Vitro Method Practices (GIVIMP) Available at: https://www.oecd-ilibrary.org/environment/guidance-document-on-good-in-vitro-method-practices-givimp_9789264304796-en
OECD Test No. 406, 1992. Skin sensitisation, Available at: https://www.oecd-ilibrary.org/environment/test-no-406-skin-sensitisation_9789264070660-en
OECD, Test No.407, 2008. Repeated Dose 28-day Oral Toxicity Study in Rodents, Available at: https://www.oecd-ilibrary.org/environment/test-no-407-repeated-dose-28-day-oral-toxicity-study-in-rodents_9789264070684-en
OECD, Test No.408, 2018. Repeated Dose 90-Day Oral Toxicity Study in Rodents, Available at: https://www.oecd-ilibrary.org/environment/test-no-408-repeated-dose-90-day-oral-toxicity-study-in-rodents_9789264070707-en
OECD, Test No. 443, 2018. Extended One-Generation Reproductive Toxicity Study, Available at: https://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en
OECD, Test No. 442E, 2018. In Vitro Skin Sensitisation, In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation, Available at: https://www.oecd-ilibrary.org/environment/test-no-442e-in-vitro-skin-sensitisation_9789264264359-en

OECD, Adverse Outcome Pathways, Molecular Screening and Toxicogenomics, 2020a. Available at: http://www.oecd.org/chemicalsafety/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm

OECD, AOP wiki, 2020b. Available at: https://aopwiki.org/

Ooka, M., Lynch, C., Xia, M., 2020 Application of *In Vitro* Metabolism Activation in High-Throughput Screening. *Int. J. Mol. Sci*. 21(21):E8182.

Peiser, M., Tralau, T., Heidler, J., Api, AM., Arts, JHE., Basketter, DA., English, J., Diepgen TL., Fuhlbrigge, RC, Gaspari, AA., Johansen, JD., Karlberg, AT., Kimber, I., Lepoittevin, JP., Liebsch, M., Maibach, HI., Martin, SF., Merk, HF., Platzek, T., Rustemeyer, T., Schnuch, A., Vandebriel, RJ., White, IR., Luch, A. 2012 Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Cell. Mol. Life Sci. 69:763-781.

Pessina, A., Albella, B., Bueren, J., Brantom, P., Casati, S., Gribaldo, L., Croera, C., Gagliardi, G., Foti, P., Parchment, R., Parent-Massin, D., Sibiril, Y., Van Den Heuvel, R., 2001. Prevalidation of a Model for Predicting Acute Neutropenia by Colony Forming Unit Granulocyte/Macrophage (CFU-GM) Assay, Toxicol. in Vitro 15(6), 729-740.

Pessina, A., Albella, B., Bayo, M., Bueren, J., Brantom,P., Casati, S., Croera, C., Parchment, R., Parent-Massin, D., Schoeters, G., Sibiri, Y., Van Den Heuvel, R., Gribaldo, L., 2002. In Vitro Tests for Haematotoxicity: Prediction of Drug-Induced Myelosuppression by the CFU-GM Assay, Altern. Lab. Anim. 30 Suppl 2, 75-79.

Pessina, A., Albella, B., Bayo, M., Bueren, J., Brantom, P., Casati, S., Croera, C., Gagliardi, G., Foti, P., Parchment, R., Parent-Massin, D., Schoeters, G., Sibiril, Y., Van Den Heuvel, R., Gribaldo, L., 2003. Application of the CFU-GM assay to predict acute drug-induced neutropenia: an international blind trial to validate a prediction model for the maximum tolerated dose (MTD) of myelosuppressive xenobiotics. Toxicol. Sci. 75(2), 355-367.

Pessina, A., Malerba, I., Gribaldo, L., 2005. Hematotoxicity testing by cell clonogenic assay in drug development and preclinical trials. Curr. Pharm. Des. 11(8), 1055-1065.

Pessina, A., Bonomi, A., Cavicchini, L., Albella, B., Cerrato, L., Parent-Massin, D., Sibiril, Y., Parchment,R., Behrsing, H., Verderio, P., Pizzamiglio, S., Giangreco, M., Baglio, C., Coccè, V., Sisto, F., Gribaldo, L., 2010. Prevalidation of the Rat CFU-GM Assay for in Vitro Toxicology Applications, Altern. Lab. Anim. 38(2), 105-117.

Rich, I.N., Hall, K.M., 2005. Validation and development of a predictive paradigm for hemotoxicology using a multifunctional bioluminescence colony-forming proliferation assay. Toxicol. Sci. 87(2), 427-441.

Ringerike, T., Ullerås, E., Völker, R., Verlaan, B., Eikeset, A., Trzaska, D., Adamczewska, V., Olszewski, M., Walczak-Drzewiecka, A., Arkusz, J., van Loveren, H., Nilsson, G., Lovik, M., Dastych, J., Vandebriel, R. J., 2005. Detection of immunotoxicity using T-cell-based cytokine reporter cell lines (“Cell Chip”). Toxicology 206(2), 257-272.

Saito, R., Hirakawa, S., Ohara, H., Yasuda, M., **Yamazaki, T**., Nishii, S., Aiba, S., 2011. Nickel differentially regulates NFAT and NF-kappaB activation in T cell signaling. Toxicol. Appl. Pharmacol. 254(3), 245-255.

Schmeits, P.C., Shao, J., van der Krieken, D.A., Volger,O.L., van Loveren, H., Peijnenburg, A.A., Hendriksen, P.J., 2015. Successful validation of genomic biomarkers for human immunotoxicity in Jurkat T cells in vitro. J. Appl. Toxicol. 35(7), 831-841.

Senju, S., Haruta, M., Matsumura, K., Matsunaga, Y., Fukushima, S., Ikeda, T., Takamatsu, K., Irie, A., Nishimura, Y., 2011. Generation of dendritic cells and macrophages from human induced pluripotent stem cells aiming at cell therapy. Gene. Ther. 18(9), 874-883.

Shao, J., Berger, L.F., Hendriksen, P.J., Peijnenburg, A.A., van Loveren, H., Volger, O.L., 2014. Transcriptome-based functional classifiers for direct immunotoxicity. Arch Toxicol., 88(3), 673-689.

Shanti, A., Teo, J., Stefanini, C., 2018. In Vitro Immune Organs-on-Chip for Drug Development: A Review. Pharmaceutics 10(4). doi: 10.3390/pharmaceutics10040278.

Stølevik, S.B., Nygaard, U.C., Namork, E., Granum, B., Pellerud, A., van Leeuwen, D.M., Gmuender, H., van Delft, J.H., van Loveren, H. & Løvik, M., 2011. In vitro cytokine release from human peripheral blood mononuclear cells in the assessment of the immunotoxic potential of chemicals. Toxicol. in Vitro 25(2), 555-562.

Takahashi, T., Kimura, Y., Saito, R., Nakajima, Y., Ohmiya, Y., Yamasaki, K., Aiba, S., 2011. An in vitro test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. Toxicol Sci. 124(2), 359-369.

Investigative Immunotoxicology, Tryphonas, H., Fournier, M., Blakley, B., Smits, J. and Brousseau, P. eds. CRC Press Boca Raton 2005, pp 480.

The European Parliament and the Council, Retrieved 17 April 2016, EU Directive (2010/63/EU). Official Journal of the European Union.

Ullerås, E., Trzaska, D., Arkusz, J., Ringerike, T., Adamczewska, V., Olszewski, M., Wyczółkowska, J., Walczak-Drzewiecka, A., Al-Nedawi, K., Nilsson, G., Białek-Wyrzykowska, U., Stepnik, M., Loveren, H.V., Vandebriel, R.J., Løvik, M., Rydzyński, K., Dastych, J., 2005. Development of the "Cell Chip": a new in vitro alternative technique for immunotoxicity testing. Toxicology 15, 206(2), 245-256.

USEPA. 2019. Directive to Prioritize Efforts to Reduce Animal Testing. https://www.epa.gov/research/administrator-memo-prioritizing-efforts-reduce-animal-testing-september-10-2019. Accessed January 28,2021

Verstrepen L, Bekaert T, Chau TL, Tavernier J, Chariot A, Beyaert R. TLR-4, IL-1R and TNF-R signaling to NF-κB: variations on a common theme. Cell Mol Life Sci 2008, 65:2964-2978

Vessillier, S., Eastwood, D., Fox, B., Sathish, J., Sethu, S., Dougall, T., Thorpe, SJ., Thorpe, R., Stebbings, R., 2015. Cytokine release assays for the prediction of therapeutic mAb safety in first-in man trials--Whole blood cytokine release assays are poorly predictive for TGN1412 cytokine storm. J Immunol Methods 424:43-52.

Encyclopedic Reference of Immunotoxicology, Vohr, H-W. ed. Springer Berlin 2005 pp 730.

Vos JG. Immune suppression as related to toxicology. CRC Crit Rev Toxicol. 1977 May;5(1):67-101.

Vos, J.G., Van Loveren, H.,1998. Experimental studies on immunosuppression: how do they predict for man? Toxicology 129(1), 13-26.

Vizcardo, R., Masuda, K., Yamada, D., Ikawa, T., Shimizu, K., Fujii, S., Koseki, H., Kawamoto, H., 2013. Regeneration of human tumor antigen-specific T cells from iPSCs derived from mature CD8(+) T cells. Cell Stem Cell 12(1), 31-36.

WHO, Guidance for Immunotoxicity Risk Assessment for Chemicals. IPCS Harmonization Project Document No. 10. 2012. ISBN 978 92 4 150330.

Principles and Methods for Assessing Direct Immunotoxicity Associated with Exposure to Chemicals. International Programme on Chemical Safety: Environmental Health Criteria 180 1996. World Health Organization, Geneva

. Appendix

Antibody Response: Alterations in antigen-specific T cell-dependent antibody responses. Antigens most commonly used include sheep erythrocytes and keyhole limpet hemocyanin.
Cell Proliferation: Alterations in the proliferative response following stimulation with mitogenic compounds such as concanavalin A, phytohemagluttinin, lipopolysaccharide or with allogenic leukocytes in a mixed lymphocyte reaction.
Clinical pathology: Indicates alterations in the leukogram or hematology parameters, such as leukocytopenia/leukocytosis, granulocytopenia/granulocytosis or lymphopenia/lymphocytosis.
CTL: The cytotoxic T lymphocyte assay measures cytotoxity against tumor or virally infected cells. For the reference compounds above, the majority of studies used either *ex vivo* or *in vitro* assays.
DTH: Alterations in delayed type hypersensitivity were measured by assessing the cell-mediated immune response to a soluble antigen such as keyhole limpet hemocyanin following sensitization and secondary challenge.
Lymphoid Organ weights: Decreases in the weight of spleen, thymus or relevant lymph nodes.
NK cytotoxicity: Natural killer cell cytotoxicity was measured using either *ex vivo* or *in vitro* assays which quantitate cytotoxicity against tumor cell lines or virally infected cells.

OECD 3rd webinar
August 31th, 2022

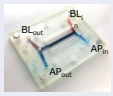
Lesson and Learned for a Validation Study

Hajime Kojima, Ph. D.,
JaCVAM, CBSR, NIHS, Japan

1

MPS Devices for Commercialization in AMED-MPS Project

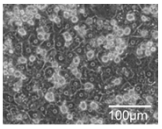
http://www.scetra.or.jp/business/

	Overview http://www.scetra.or.jp/business/#section-6							
Name (tentative name)		Design/Manufacturing Related device manufacturing	Structure	Target organs				Characteristics
				Liver	Intestine	BBB	Kidney	
Matsunaga Device		Prof. Matsunaga's Laboratory, Nagoya City University/ Shinko Chemical Co., Ltd.	Two-organs connected culture	◎	◎			• In addition to two-organ coupling assays, it is also applicable to membranous tissue/organ assays. • Applicable to 5 commercially available inserts. • Adopt pump drive. • Size complies with SBS standards.
Kimura Device Fluid3D-X		Prof. Kimura's Laboratory, Tokai University/ TOKYO OHKA KOGYO Co., Ltd.	Multi-layered Microchannel with Porous Membrane Experimental example • Shear stress load • Coculture of Epithelial and endothelial cells • AP-BL material transfer observation	○	○	○	◎	• Plastic is used as a chip material to low drug adsorption. • A unique manufacturing process enables flexible microchannel design change. • Low cytotoxicity and good observability (fluorescence/phase-contrast)
PD-MPS (Pressure Driven-MPS)		National Institute of Advanced Industrial Science and Technology (AIST)/ Sumitomo Bakelite Co., Ltd. Control and sensing unit: SHIMADZU CORPORATION・SCREEN Holdings Co., Ltd.	Two-organs connected culture	○	○			• A platform device that can accommodate a variety of organ models by flexible fluid channel design and on-board inserts. • A closed system fluid flow is possible by pressure driving.
			Flow culture (Medium flow under a membrane)		○	○		
Sakai/Kimura Device (On-chip pump type MPS device)		Prof. Sakai's Laboratory, Tokyo University; Prof. Kimura's Laboratory, Tokai University/ Sumitomo Bakelite Co., Ltd.	Two-organs connected culture	○	○			• An organ block type platform device that can accommodate various organ assays through recombination of various culture equipment (inserts and cell desk). • Perfusion culture is possible by driving the device built-in stirrer type pump. • Oxygen permeable film, honeycomb structure membrane can be used for cell culture bottom in addition to conventional polystyrene.

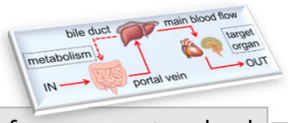
Column for the target organs: It shows the combination of experimental systems in which cells derived from various organs are cultured on each device to evaluate the function of cells and the device. Available in all combinations. The cells not shown are also available. Liver: Liver cells ; Gut: Gut cells; BBB: Blood-brain barrier cells; Kidney: Kidney cells. The current development status is shown as follows; ◎: Analysis data using the applicable cells have been obtained; ○: An experiment using the applicable cells is desired based on the user needs.

2

Preparing Operating Procedure for Round-Robin Test

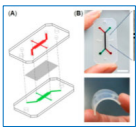


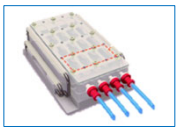
Cell Criteria



MPS performance standard

Standard for material and equipment





Standard Operating Procedures

Test developers have many concerns for a validation study

Complicated process

Excessive demands?

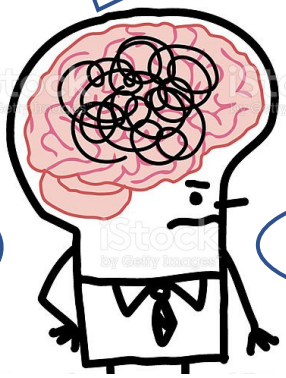
My protocol has already published.

No budget

long-term restraint?

Cannot discussion in English

Why not a peer review after the validation ?



Simple and easy study is only acceptable.

Evaluation criteria for the validated novel test methods

1. Study objective and test method purpose
2. Biological and mechanistic relevance
3. Test method protocol
4. Appropriateness of the validation study management and conduct
5. Adequacy of chemicals selected for validation of the study objective
6. Quality of the reference data used for evaluation of the relevance
7. Training of naïve laboratories and transferability
8. Use of quality assurance system(s) during data generation
9. Within- and between-laboratory reproducibility
10. Predictive capacity
11. Applicability domain and limitations
12. Completeness of data and documentation

OECD GD34 (2005)

5

Simple validation study - Why now?

- A validation study for a novel test method is crucial for the regulatory acceptance of new approach methods.
- Limitations:
  - Numerous evaluation criteria before acceptance
  - Difficulty in performing an international study owing to limited budget
  - Members involved become exhausted while performing several phases owing to revising immature protocols
  - Availability of a poor database concerning several fields of toxicology

6

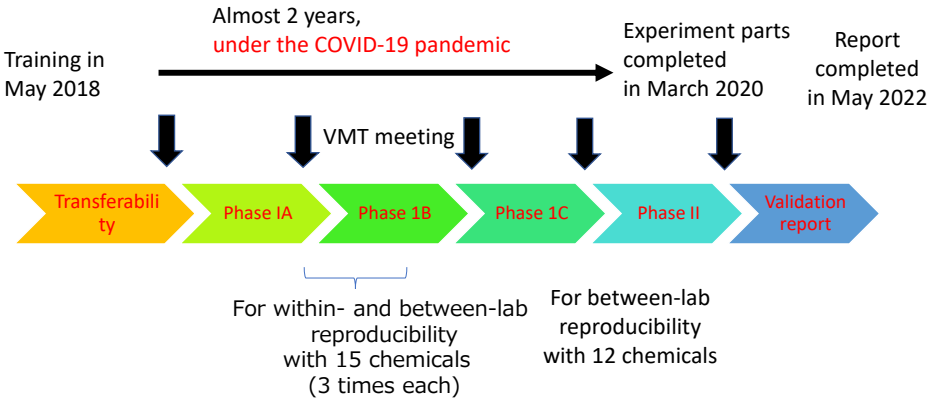
My proposal is based on Lessons and Learned

- A detailed review paper including a performance standard should be developed for each systemic toxicity by OECD.
- International experts who are familiar with TG development are mandatory on the validation management team and peer review panel.
- The purpose of the study is limited to checking the robustness and reliability of the protocol based on the performance standard.
- The chemical number is up to 30.

The bottleneck of the validation study is resolved simply and easily

7

Case study: EpisensA for skin sensitization validation study



#This validation study is a performance standard based on the ECVAM validation study.
#The predictive capacity and applicability domain were evaluated based on the data by test developers in the validation report.

8

Short cut for evaluation criteria

Checkpoints prior to a validation study

1. Study the objective and purpose of test method
2. Need and benefits in comparison to those of the existing test methods
3. Comparison of the test method with the essential test method components in the Performance Standards
4. Biological and mechanistic relevance
5. Test method protocol
6. Appropriateness of the validation study management and conduct

Mandatory:
High Quality of VMT

7. Adequacy of chemicals selected for validation of the study objective
8. Quality of the reference data used for the evaluation of relevance
9. Training of naive laboratories and transferability
10. Use of quality assurance system(s) during data generation
11. Within- and between-laboratory reproducibility
12. Predictive capacity
13. Applicability domain and limitations
14. Completeness of data and documentation

Mission during the validation study

After the validation study, the predictive capacity and applicable domain is evaluated by VMT members and peer review panel with the test developers.

9

For NAM validation

- Test method is an important element for NAM.
- The purpose is limited to checking the robustness and reliability of the test protocol.
- Before starting the validation study, the VMT members should discuss the objective, benefit, and biological and mechanistic relevance with the test developers and the study plan, including acceptance criteria, with a sponsor.
- The VMT members should discuss protocol development and chemical selection during the ongoing validation study.

10