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 プログラムにおいて TG と DA を開発するための AOP に関する研究

令和元年度 分担研究報告書

AOP、TG、DAの開発、AOP国内マニュアルの作成

研究分担者　小島　肇
国立医薬品食品衛生研究所　安全性生物試験研究センター　安全性予測評価部　室長

研究要旨

動物実験 3Rs の浸透に加えて、実験動物とヒトとの種差等の克服のために、既存の毒性試験法の見直しが世界的に進んでいる。経済協力開発機構（OECD: Organisation for Economic Co-operation and Development）においても、全身毒性試験（免疫毒性、生殖発生毒性、発がん性及び光安全性等）の有害性発現経路（AOP: Adverse Outcome Pathway）を開発し、その情報を活用して動物実験代替法を念頭においた試験法ガイドライン（TG: Test Guideline）の公定化や *in silico* 法を確立する一方で、AOP 等の毒性情報を網羅した IATA（Integrated Approaches to Testing and Assessment）を開発し、それに基づき、TG と同格の扱いになる DA（Defined Approach）による化学物質の安全性評価を推進している。このような国際的な潮流に乗り、日本が得意とする分野で主導権を握って、AOP や TG を公定化し、さらには IATA や DA の開発に協力することが本研究班の目的である。

昨年度からの継続した活動の中、本年度に以下に示す 3 試験法が OECD の TG として採択された。本年度、AOP の採択は叶わなかったが、AOP154 案“Inhibition of Calcineurin Activity Leading to impaired T-Cell Dependent Antibody Response ”に関しては、来年度採択に目途が立った。

1）光安全性　活性酸素種（ROS）アッセイ（TG495）
2）皮膚感作性試験代替法 アミノ酸誘導体反応試験（ADRA）（TG442）
3）LabCyte EPI-MODEL24 を用いる腐食性試験代替法（TG431）

研究分担者氏名・所属研究機関名及び
所属研究機関における職名

相場節也　東北大学医学系研究科・医学部・皮膚科学分野教授

足利太可雄　国立医薬品食品衛生研究所 安全性生物試験研究センター 安全性予測評価部 主任研究官
大石　巧　日本免疫毒性学会試験法委員会 AOP 検討小委員会（株式会社ボゾリサーチセンター）
尾上誠良　静岡県立大学　薬学部・薬剤学分野　教授

尾上誠良　静岡県立大学　薬学部・薬剤学分野　教授
笠原利彦　富士フイルム株式会社　安全性評価センター技術マネージャー
加藤雅一　株式会社ジャパン・ティッシュ・エンジニアリング(J-TEC)主任研究員
木村　裕　東北大学医学系研究科・医学部・皮膚科学分野准教授
久田　茂　日本免疫毒性学会試験法委員会(あすか製薬株式会社)

A. 研究目的

本研究班では、OECD の AOP 開発プロジェクトの中で、化学物質の毒性情報等を集積しながら、免疫毒性、生殖発生毒性、発がん性及び光安全性等に関する日本発の AOP 開発を進める。既存の AOP 情報をもとに開発された皮膚感作性試験代替法 ADRA（Amino acid Derivative Reactivity Assay)、免疫毒性試験 MITA(Multi-Immuno Toxicity Assay)、発生毒性試験スクリーニング Hand1-Luc EST(Embryonic Stem cell Test)、光安全性試験スクリーニング ROS（Reactive Oxygen Species）アッセイ、LabCyte EPI-MODEL24 を用いる腐食性試験代替法については、試験法毎に独立した国内外の専門家による第三者評価(peer review)を受けた後、TG を開発する。一方で、皮膚感作性 DA(DASS: Defined Approach for Skin Sensitisation)の開発に関与することを通じて、IATA や DA の国内での普及に務める。

B. 研究方法

B.1. AOP、TG、DA の開発、AOP 国内マニュアルの作成

B.1.1. AOP 国内マニュアルの作成

OECD の AOP 開発プロジェクト EAGMST（Extended Advisory Group on Molecular Screening and Toxicogenomics)及び、TG の開発プロジェクト WNT（Working Group of the National Coordinators of the Test Guidelines Programme)の進捗に合わせ、班員を支援した。

OECDの定める「AOPの開発・評価に関するガイダンス　No.184(以下、AOPガイダンスと記す)」を翻訳し、マニュアル作成に向けた資料整備を進めた。

B.1.2. AOP開発

AOP に関しては、日本免疫毒性学会会員をメンバーとする同学会試験法委員会 AOP 検討小委員会に免疫毒性 AOP の開発を委託している。

文献調査の結果に基づいて、カルシニューリン阻害を分子初動(MIE: Molecular initiating event)とし、T 細胞依存性抗体産生抑制(TDAR)を有害性発現(AO：Adverse Outcome)とする AOP154 案“Inhibition of Calcineurin Activity Leading to impaired T-Cell Dependent Antibody Response”を作成し、peer review に対応した。

B.1.3. TG 開発

日本から提案している試験法である皮膚感作性試験代替法 ADRA、光安全性試験 ROS

アッセイ(尾上分担研究者との協同研究)、LabCyte EPI-MODEL24 を用いる腐食性試験代替法の TG 採択のために、WNT や専門家会議にて交渉した。また、ヒトアンドロゲン(AR)受容体安定トランスフェクト転写活性試験(AR STTA: The Stably Transfected Transactivation method using the AR-EcoScreen™ cell line)TG458 の改訂に向け、尽力した。

分担研究者の相場が開発し、他の研究班でバリデーションを終了させた IL-2 を指標とした免疫毒性試験の TG を目指し、海外の専門家を招聘したバリデーション報告書の peer review を進めた。

また、*in vitro* 免疫毒性試験に関する Detailed Review Paper(DRP)の作成を国際専門家とともに進めた。

B.1.4. DA の開発協力

足利分担研究者とともに、OECD 専門家会議において、DASS の開発に協力した。

C. 研究結果

C.1. AOP、TG、DA の開発、AOP 国内マニュアルの作成

C.1.1. AOP国内マニュアルの作成

AOP ガイダンスを翻訳して、班員に普及した(添付資料 1)。今後、HP などで公開するとともに、来年度のマニュアル作成に反映させる予定である。

C.1.2. AOP開発

AOP154 案“Inhibition of Calcineurin Activity Leading to impaired T-Cell Dependent Antibody Response”に関しては、2018 年 6 月に EAGMST の内部 peer reviewer が終了したものの、外部 peer review の開始まで 1 年もまたされた。この理由として、OECD 事務局が外部専門家との連携関係を模索している最中であることが挙げられる。何とか、免疫毒性の専門家グループも見つかり、2019 年末より、外部 peer reviewer による評価が進んでいる。2020 年 6 月の EAGMST 会議で採択されるよう免疫毒性学会の皆様に尽力して頂いている。

C.1.2. TG の開発

1) 皮膚感作性試験

皮膚感作性試験代替法 ADRA は、2019 年 6 月 18 日の OECD TG442C(添付資料 2)として採択された。

ただし、その後、記載ミスが見つかったこと、新たな項目を追加したいとの富士フイルムの意向を受けて、11 月に TG の改定を求める SPSF を提出し、2020 年 4 月の WNT で作業計画に採択された。

2) 光安全性試験

ROS アッセイの TG は、2019 年 6 月 18 日の OECD TG495(添付資料 3)として採択された。

引き続き、ROS アッセイを含む IATA 開発に関する SPSF を 11 月に提出し、2020 年 4 月の WNT で作業計画に採択された。

3)腐食性試験

LabCyte EPI-MODEL24 を用いる腐食性試験代替法は、2019 年 6 月 18 日の OECD TG431 培養表皮モデルを用いた腐食性試験(添付資料 4)の中の一つに採択された。

4) 内分泌かく乱試験

AR-STTA TG458の改訂に向け、追加される2試験法 The AR-CALUX® method using the AR-CALUX® cell line 及び The ARTA method using the 22Rv1/MMTV_GR-KO cell lineのpeer reviewerを務めた。

結果として、2020 年 4 月の WNT で本改定 TG458 が採択された。

5) 免疫毒性試験

表 1 に示す海外の専門家を招聘し、IL-2 Luc アッセイバリデーション報告書の peer review を進め、評価書案(添付資料 5)が最終段階にある。2019 年 2 月の対面会議後、バリデーション報告書の改訂を受けて、2019 年 10 月および 11 月の電話会議を経て(添付資料 6 および 7)作成された。ただし、まだ最終合意には至っていない。

本件に関しては、2019年11月にIL-2 Lucアッセイの TG開発のSPSFを提出した。ただ、2020年4月のWNTにおいてプロジェクトとしての採択は認められなかった。まずは以下に示すDRPの提出を急ぐべきとの忠告を受けた。来年の再提出を目指し、DRPを開発するとともに、評価報告書とバリデーション報告書を確実に作成することを目指していく。

一方、相場らの開発した*in vitro*免疫毒性試験を円滑にTGに導くための準備として、OECDに了承されたDRPの開発を表2に示す国際的な専門家の協力を得て、進めている。2019年10月の電話会議を経て（添付資料8）、2020年1月には専門家を招聘して、対面会議を行い（添付資料9）、表3に示す目次（Table of Content）に合意を得るとともに、項目毎の担当部分について意見交換した。

来年度早々、DRPをOECDに提出する予定である。

C.1.3. DA の開発協力

OECD 専門家会議（電話会議）で DASS の開発に寄与した。2019 年度だけでも、10 回の電話会議が開催されているが、議事録の一部を添付とした（添付資料 10）。

ヒト及び動物実験結果の再評価、適用範囲および不確定因子について、それぞれサブワーキンググループの討議が一年に渡って続いており、その提案を待って DASS の最終的な議論が 2020 年度になされる予定である。

D. 考察

D.1. AOP の開発

本研究班から提案している免疫抑制の AOP154 案 ”Inhibition of Calcineurin Activity Leading to impaired T-Cell Dependent Antibody Response”は OECD における外部 peer review が進行中である。ただ、現在の状況から来年度採択に目途が立ったと考えている。

D.2. TG の開発

皮膚感作性試験代替法 ADRA、光安全性 ROS アッセイ及び LabCyte EPI-MODEL24 を用いる腐食性試験代替法に関しては、2019 年 6 月 18 日に TG として採択された。

一方、新たに 3 件の SPSF を OECD に 2019 年 11 月に提出し、2020 年 4 月の WNT で OECD 作業計画に加えられることになった。継続して、TG の開発および修正に関与していきたい。

D.3. IATA 及び DA の公定化

DASS に関する OECD 活動に対し、引き続き協力していく予定である。来年度は光安全性 IATA の開発を国内外の専門家とともに進め、動物実験を用いない光安全性評価の体系化を日本からも提案していく予定である。

E. 結論

昨年度からの継続した活動の中、本年度 3 試験法が OECD の TG として採択された。本年度、AOP の採択は叶わなかったが、AOP154 案“Inhibition of Calcineurin Activity Leading to impaired T-Cell Dependent Antibody Response ”に関しては、来年度採択に目途が立った。

引き続き、OECD の活動の中で、日本が得意とする分野で主導権を握って、AOP や TG を公定化し、さらには IATA や DA の開発に協力していく予定である。

F. 添付資料

1. AOPの開発・評価に関するガイダンス No.184

2. OECD Test Guideline 442C for the Testing Chemicals on in chemico skin sensitisation assays addressing the Adverse Outcome Pathway Key Event on covalent binding to proteins
3. OECD Test Guideline 495 for the Testing Chemicals on ROS (Reactive Oxygen Species) assay for phototoxicity
4. OECD Test Guideline 431 for the Testing Chemicals on *in vitro* skin corrosion, Reconstructed human epidermis test methods
5. IL-2 Luciferase (Il-2 Luc) Assay Report of the Peer Review Panel.
6. Teleconference for IL-2 PRP October 1, 2019
7. Teleconference for IL-2 PRP, November 11, 2019
8. Teleconference for DRP on in vitro immunotoxicity, October 28, 2019
9. 2020年1月DRP開発のための対面会議議事録
10. Expert Group on DASS, November 18& 19

表 1. IL-2 Luc アッセイ第三者評価委員会メンバーリスト

No.	Name	Affiliation	Country
1	Henk van Loveren	Maastricht University	Netherlands
2	Haley Neff-LaFord	Seattle Genetics, Inc.	USA
3	Barbara Kaplan	Mississippi State University	USA
4	Fujio Kayama	Jichi Medical University	Japan
5	Xingchao Geng	National Center for Safety Evaluation of Drugs (NCSED)	China
6	Sang-Hyun Kim	Kyungpook National University	Korea
7	Takao Ashikaga	National Institute of Health Sciences	Japan

表 2. In vitro 免疫毒性試験　総説（Detailed Review Paper）作成メンバーリスト

No.	Name	Affiliation	Country
1	Emanuela Corsini	Università degli Studi di Milano	Italy
2	Dori Germolec	NTP/NIEHS	USA
3	Barbara Kaplan	Mississippi State University	USA
4	Henk van Loveren	Maastricht University	Netherlands
5	Haley Neff-LaFord	Seattle Genetics, Inc.	USA
6	Erwin L. Roggen	3RsMC ApS	Denmark
7	Setsuya Aiba	Tohoku University	Japan
8	Yutaka Kimura	Tohoku University	Japan
9	Takayuki Yoshimoto	Tokyo Medical University	Japan
10	Hajime Kojima	JaCVAM, National Institute of Health Sciences	Japan

表 3.　Table of Content (ToC) for DRP

Item
ABOUT THE OECD FOREWORD
LIST OF ABBREVIATIONS
EXECUTIVE SUMMARY
I. Introduction
II. Background
III. Basic concept of immunotoxicity
IV. State-of-the-art of AOP on immunotoxicity
V. State-of-the-art in the field of *in vitro or non-animal* assay
VI. Performance factors of *in vitro* assay
VII. Assay qualification information of *in vitro* assay
VIII. Selection factors for the reference compound developing in vitro assay
IX. Reference compound list
X.　Battery of *in vitro or non-animal* assays
XI. Discussion and Recommendation
XII. References
XIII. Appendix

G. 研究発表

G-1.学会誌・雑誌等における論文一覧

（国内誌　1　件、国際誌　7　件）

1. Kobayashi-Tsukumo H, Oiji K, Xie D, Sawada Y, Yamashita K, Ogata S, Kojima H, Itagaki H: Eliminating the contribution of lipopolysaccharide to protein allergenicity in the human cell-line activation test (h-CLAT), J Toxicol Sci. 2019;44(4):283-297. doi: 10.2131/jts.44.283.
2. 荻原 琢男, 細野 麻友, 小島 肇: ヒト肝細胞の 3 次元培養スフェロイドモデルの新展開, 日本薬理学雑誌 2019;153(5):235-241. doi: 10.1254/fpj.153.235.
3. 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, Nicole K, Kim BH, Kojima H, Kasahara T, Ono A: The within- and between-laboratory reproducibility and predictive capacity of the in chemico amino acid derivative reactivity assay: Results of validation study implemented in four participating laboratories. J Appl Toxicol. 2019 Nov;39(11):1492-1505. doi: 10.1002/jat.3834.
4. Kojima H, Sakai Y, Tanaka N: Japanese Contributions to the Development of Alternative Test Methods, The History of Alternative Test Methods in Toxicology, Elsevier, Netherlands, 2019, 79-85.
5. Mizoi K, Hosono M, Kojima H, Ogihara T: Establishment of a primary human hepatocyte spheroid system for evaluating metabolic toxicity using dacarbazine under conditions of CYP1A2 induction. Drug Metab Pharmacokinet. 2019 Dec 24. pii: S1347-4367(19)30233-2. doi: 10.1016/j.dmpk.2019.11.002.
6. 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, Fujita M: Oxidation of a cysteine-derived nucleophilic reagent by dimethyl sulfoxide in the amino acid derivative reactivity assay. J Appl Toxicol. 2020 Feb 12. doi: 10.1002/jat.3948.
7. Marx U, Akabane T, Andersson TB, Baker E, Beilmann M, Beken S, Brendler-Schwaab S, Cirit M, David R, Dehne EM, Durieux I, Ewart L, Fitzpatrick SC, Frey O, Fuchs F, Griffith LG, Hamilton GA, Hartung T, Hoeng J, Hogberg H, Hughes DJ, Ingber DE, Iskandar A, Kanamori T, Kojima H, Kuehnl J, Leist M, Li B, Loskill P, Mendrick DL, Neumann T, Pallocca G, Rusyn I, Smirnova L, Steger-Hartmann T, Tagle DA, Tonevitsky A, Tsyb S, Trapecar M, Van de Water B, Van den Eijnden-van Raaij J, Vulto P, Watanabe K, Wolf A, Zhou X, Roth A: Biology-inspired microphysiological systems to advance patient benefit and animal welfare in drug development. ALTEX. 2020 Feb 28. doi: 10.14573/altex.2001241.
8. 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: 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. 2020 Mar 18;66:104832. doi:

10.1016/j.tiv.2020.104832.

G-2.学会・シンポジウム等における口頭・ポスター発表

1. Kojima H:The Japanese Strategy on Chemical Risk Assessment with New Approaches, International Symposium for EDCs Testing & Assessment（2019.5.31, Seoul, Korea）
2. Kojima H: Use of new approach methods (NAM) in next generation risk assessment (NGRA), International Symposium for EDCs Testing & Assessment (2019.6.4, Munich, Germany)
3. 小島 肇: In vitro から in vivo の予測、ヒト外挿性向上への期待,第 46 回日本毒性学会学術年会（2019.6.26, 徳島）
4. 小島 肇, 小川 久美子, 西川 秋佳, 若林 敬二, 鰐渕 英機, 林 真, 福島 昭治, 遠山 千春: 実験動物を用いた安全性・リスク評価に携わる人材育成の必要性, 第 46 回日本毒性学会学術年会（2019.6.27, 徳島）
5. 小島 肇: 皮膚感作性試験代替法を行政的に受け入れるための国際動向, 第 46 回日本毒性学会学術年会（2019.6.27, 徳島）
6. 小島 肇: OECD AOP プロジェクトにおける日本の対応, 第 46 回日本毒性学会学術年会（2019.6.28, 徳島）
7. 溝井 健太, 細野 麻友, 松本 映子, 矢野 健太郎, 下井 昭仁, 小島 肇, 荻原 琢男: 肝スフェロイドを用いた薬物の経口急性毒性試験の実験動物代替法の検討,第 46 回日本毒性学会学術年会（2019.6.27, 徳島）
8. Kojima H: 21st Century Toxicology and Regulatory Testing: An Update from East Asia, , The 15th International Congress of Toxicology (ICTXV) (2019.7.16, Honolulu, U.S.A)
9. Kojima H: Safety evaluation of cosmetic ingredients using 3-D models, Conference on Analytical Techniques for Cosmetics (2019.7.25, Chiayi, Taiwan)
10. 小島 肇: OECD AOP プロジェクト,第 26 回日本免疫毒性学会学術年会（2019.9.10, 北九州）
11. 小島 肇: 培養組織モデルの国際標準化の状況, LbL-3D 組織シンポジウム（2019.9.12, 吹田）
12. Kojima H: Establishment of the Asian Consortium for Three Rs supported by ASCCT, 8th Annual Meeting of the American Society for Cellular and Computational Toxicology (2019.9.25, Gaithersburg, MD, USA)
13. Kojima H: Asian Consortium for Three Rs, European Society for Alternatives to Animal Testing (EUSAAT) 2019 (2019.10.12, Linz, Austria)
14. 木村 裕, 安野理恵, 渡辺美香, 小林美和子, 岩城知子, 藤村千鶴, 近江谷克裕, 山影康次, 中島芳浩, 真下奈々, 高木佑実, 大森 崇, 小島 肇, 相場節也: Multi-ImmunoTox Assay（MITA）の予測性評価に必要な文献に基づく化学物質免疫毒性分類の試み,日本動物実験代替法学会第 32 回大会（2019.11.20, つくば）
15. 秋元美由紀, 吉田浩介, 渡辺真一, 山鹿宏彰, 若林晃次, 田原 宥, 堀江宣行, 藤本恵一, 草苅 啓, 神谷孝平, 河上強志, 小島幸一, 寒水孝司, 小野 敦, 小島 肇, 藤田正晴, 山本裕介, 笠原利彦: ADRA における DMSO 溶媒中での NAC の酸化と感作性予測精度に与える影響,日本動物実験代替法学会第32回大会（2019.11.21, つくば）
16. 諫田泰成, 大戸茂弘, 鈴木 真, 武吉正博, 竹内小苗, 佐久間めぐみ, 中村 牧, 小島

肇:日本動物実験代替法学会　国際交流委員会報告,日本動物実験代替法学会第32回大会(2019.11.22, つくば)

17.小島 肇: 安全性評価試験法のOECD 等における国際動向と課題,日本動物実験代替法学会第32回大会(2019.11.22, つくば)

H. 知的所有権の取得状況

H-1. 特許取得

特になし

H-2. 実用新案登録

特になし

H-3. その他

特になし

非機密文書

ENV/JM/MONO(2013)6
27-Jul-2017

経済開発協力機構

English, French

環境局
化学品委員会と化学品・農薬・バイオテクノロジーに関する調査委員会の合同会合

ENV/JM/MONO(2013)6
Unclassified

2013年4月17日の同文書を削除・差し替え

AOPの開発・評価に関するガイダンス文書（改訂）

リスク管理シリーズ
No. 184

ガイダンス文書第2版、2013年初版の差し替え

JT03417683

OECD 環境健康安全出版物

リスク管理シリーズ

No. 184

有害性発現経路（AOP）の開発・評価に関するガイダンス文書（改訂）

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

環境局
経済協力開発機構
2017年　パリ

OECDについて

経済協力開発機構（OECD）は、南北アメリカ、ヨーロッパ、アジア太平洋地域の先進工業国35ヶ国と欧州委員会の代表者が政策を調整し、共通の問題を討議し、一致して国際問題に対応するための国際機関である。 OECDの業務の大半は、加盟諸国の代表で構成される200以上の専門委員会やワーキンググループが行っている。OECDで非加盟国から一定の条件を満たした国々や関連する国際機関などが、オブザーバーとして頻繁に出席している。委員会および作業部会は、フランスのパリにあるOECD事務局が支援している。事務局は複数の部局で編成されている。

環境健康安全局は、12の異なるシリーズ、試験・評価; 優良試験所基準（GLP）および遵守状況の監視; 農薬; 殺生物剤; リスク管理; バイオテクノロジーの規制監督の調和; 新規食品・飼料の安全性; 化学事故; 環境汚染物質排出・移動登録; 排出シナリオ文書; 工業ナノ材料の安全性; 有害性発現経路で無料の文書を公開している。環境健康安全プログラム、およびEHSの出版物に関する詳細情報は、OECDのWorld Wide Webサイト（www.oecd.org/chemicalsafety/）を参照されたい。

本書は化学物質の適正管理のための国際機関間プログラム（IOMC）の枠組みで作成された。内容は、必ずしも個々のIOMC参加機関の見解または規定されたポリシーを反映するものではない。

化学物質の適正管理のための国際機関間プログラム（IOMC）は、化学物質の安全という分野での協力を強め、国際調整を進めるべく、国連環境開発会議が1992年に行った勧告を受けて、FAO、ILO、UNDP、UNEP、UNIDO、UNITAR、WHO、世界銀行、OECD（参加機関）によって1995年に設置された。IOMCの目的は、人の健康と環境に関する化学物質の適正な管理のため、参加機関が共同または個別に実行する政策や取り組みの協調を進めることである。

本書はインターネットを通じて無償で入手することができます。

本書または他の環境保健安全関連の各種出版物の全文については、
OECD のウェブサイト **(http://www.oecd.org/ehs)**をご覧いただくか、

または下記にご連絡下さい。

OECD 環境局
環境健康安全部

2 rue André-Pascal
75775 Paris Cedex 16
France

Fax: (33) 01 45 24 16 75

E-mail: ehscont@oecd.org

OECD 環境健康安全出版物

リスク管理シリーズ

No. 184

有害性発現経路（AOP）の開発・評価に関するガイダンス文書

（第二版）

緒言

有害性発現経路（AOP）方法論は、化学物質の生物学的および毒物学的影響に関する関連情報を収集、整理、評価するための枠組みを提供する手法である。より具体的には、AOP手法は、生物学的組織の分子レベルの摂動と規制上の懸念のある生物学的組織のレベルでの有害性発現との間で、生物学的に妥当で経験に支えられる関連する既存の知識を体系づける。この手法は、化学物質や他のストレス因子の悪影響を理解するため、作用機序（MOAおよび/または作用機構）により支えられている。このガイダンス文書は、関連する科学データと結果として得られる知識の特定と使用を含む、語彙、概念、およびAOPの開発に関する洞察を紹介する。また、本書では、AOPの潜在的な規制上の使用についても簡単に説明している。文書の最後には、AOPの概念と構成物、およびその最終的な適用理解の促進を目的とした用語集がある。

補足文書であるOECD AOPユーザーハンドブックは、AOP開発に関する詳細な情報を提供し、AOPの構築方法の詳細を探している人にとってより適切な文書である（OECD, 2016a）。 AOPユーザーハンドブックには、基礎となる情報の信頼性を評価するためのガイダンスも含まれている。試験法と評価への統合アプローチ（IATA）の開発におけるAOPの使用に関するOECDガイダンス文書（OECD, 2016b）は、AOPの概念をさまざまな目的でIATAを開発するための枠組みとして適用する方法を説明している。 文書化されたAOPの数が増えるにつれて、IATAでの適用とさまざまな規制状況での使用について、さらにデモを行い、ガイダンスを作成できる。 経験が増えるにつれて、AOPと調和したIATAに基づくIATA開発のためのガイダンス文書が開発されることが期待される。

このガイダンスは、2012年12月にOECDのExtended Advisory Group on Molecular Screening and Toxicogenomics（EAGMST）と共同で事務局によって作成された。当時、専門家グループと加盟国がAOPの開発と評価の経験を積み、ガイダンスを改訂する必要性が認められた。それ以来、AOPの開発である程度の経験が得られ、AOPの開発と評価のためのガイダンス文書のユーザーハンドブック補足が公開された。ユーザーハンドブックには、AOP雛形（元のガイダンスのパートII）と、データの要約、AOP評価、および元のガイダンスからのAOPの信頼性を扱う項目が記載されている。ユーザーハンドブックは、AOP-Wikiと証拠の重み付け（WoE）の考慮事項に関連する資料を含むように開発されたものであり、必要に応じて定期的な更新が必要になる。これとは対照的に、現在のガイダンスは、AOP枠組みにおいて変更される可能性が低い領域を扱っている。このガイダンス文書の第2版は、AOP開発プログラムの歴史的背景を提供し、AOPの構築に必要な要素とAOP枠組みの原則の概要を示している。

ガイダンス文書の第2版は、2017年4月の第29回WNT会議で承認された。本書は、化学物質委員会とOECDの化学物質、農薬、バイオテクノロジーに関する作業部会の合同会議の責任で発行される。

目次

背景

1. 化学物質の悪影響からヒトと環境を保護するための歴史的な枠組みは、主に懸念のある単一の化学物質による動物個体を用いた毒性試験に集中している。ただし、コストと時間がかかるため、ヒトと生態系に悪影響を与える可能性のあるすべての化学物質を網羅的に試験することは実用的でも実行可能でもない。現実的には、化学物質の悪影響をより効率的に予測するための科学的に防御可能なモデルとツールが必要性とされている。現在まで、生物学的システムに関する私たちの限られた知識は、効果の外挿の基本とした作用機構情報を使用する努力を妨げてきた。それにもかかわらず、トキシコゲノミクス、バイオインフォマティクス、システム生物学、計算毒物学の進歩は、本書にてその可能性を提供している（NRC, 2007; Krewski et al, 2010）。有害性発現経路（AOP）は、関連する化学的、生物学的および毒物学的情報を収集して評価し、より効率的で予測的な評価および試験戦略の推進を支援する枠組みとして提案されている。この枠組みは、有害作用への経路をよりよく理解するために、作用機序（MOAおよび/または作用機構）の利用により支えられる基礎情報で編成されている。既存の方法と新しい方法をシステム生物学にリンクする。簡単に言えば、AOPの証拠の重み付け（WoE）の検討は、Meek et al, 2014aで報告されたように、国際的に多数の科学者が関与する作用機序（MOA）/人間関連分析の既存の進化する枠組みに組み込まれた概念と原則に基づいている。

2. このガイダンスの主な目的は、AOPの開発・評価、およびAOP固有の用語の定義を含む一貫した情報収集と組織化の枠組みを紹介することである。 これに関連して、AOPは、規制上の決定に関連する組織の生物学的レベルでの分子初動事象（MIE）と最終的な有害性発現（AO）との間の因果関係の経路に関する既存の知識を記載した概念的な構成体である（Ankley et al, 2010）。

はじめに

3. 毒物学的評価のための現在の試験手法の限界と、新しい生化学および細胞試験システムと計算予測方法の急速な発展を認識し、規制当局およびその他の利害関係者は、効率と予測可能性を高めるための試験法と評価の戦略を進化させるための基礎として、代替法を含む*in vivo*試験からの既存の知識とその他の情報源を統合する方法を模索してきた。

4. 過去20年にわたり、さまざまなグループが、毒性物質が公衆衛生に関与するAOを誘発するプロセスを定義するためのシステムおよび経路による手法を提唱してきた。経路手法の初期では、曝露用量反応モデルまたは生物学に基づいた用量反応モデルと呼ばれることが多かった（Clewell et al, 1995; Shuey et al, 1995）。 2001年、MOA情報を使用して動物データとヒトとの関連性を判断するための枠組みが、代謝から影響への一連の主要事象による経路の分類に基づき、国際化学安全プログラム（IPCS）（Sonich-Mullin et al, 2001）によって発表された。後者は2002年にOECDによって採択された（OECD, 2002）。 2007年、米国科学アカデミー（NAS）は、同様の原理に基づく「毒性経路」の概念が提案された21世紀の毒性試験の展望と戦略に関する報告書を発表した（NRC、2007）。報告書には、好ましくはヒト由来の細胞を使用し、適切に設計された*in vitro*法での化学物質曝露による生物学的経路の摂動を評価する毒性試験の方向を変える展望が含まれていた。

5. 2006年、2007年、2008年に予測毒性に関するMcKimカンファレンス（http://mckim.qsari.org）以来、IPCS MOA枠組みの改良（Boobis et al, 2006; 2008）と並行して、「有害性発現経路（AOP）」が進化した（http://mckim.qsari.org）。もともとはAnkleyとその同僚によって導入されたものであり（Ankley et al, 2010）、OECDプログラムの状況下で、規制の決定に関連する組織の生物学的レベルでのMIEと最終AOの間の因果関係の経路に関する既存の知識を構成する概念として発展した。AOPでは、さまざまな出展から既知の情報を統合することが重要である。この手法は、化学物質が最初に到達し、生物体の最初の標的と相互作用することで毒性が生じるという概念に基づいている。そのため、AOPは、MIEから対象のin vivo転帰への事象の連続的に進行する（図1）。一般的に、次のような広範な経路のセットを指す：1）ストレッサーと生物学的目標（例えば、DNA結合、タンパク質酸化など）との相互作用から生じる摂動を表すMIEから始まる。2）毒性の進行に不可欠な一連の生物学的活動（例えば、遺伝子活性化、または組織発達の変化など）に続く; 3）最終的に、人間または生態系のリスク評価者に関連する最終的なAO（例：死亡率、生殖障害、がん、または絶滅など）に到達する(OECD 2011、ENV / JM / MONO(2011)8)。

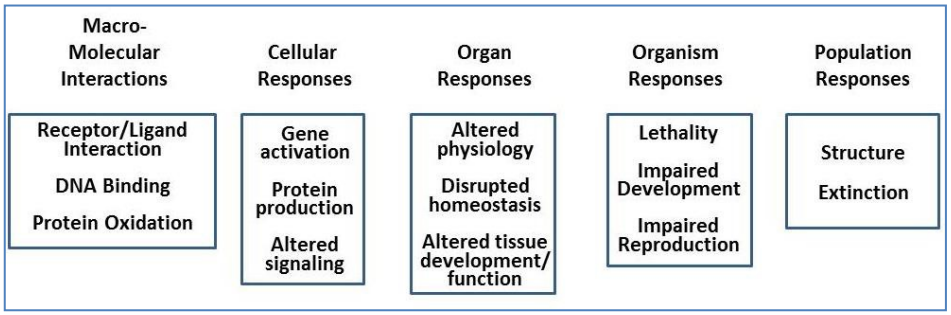


Figure 1. 有害性発現経路（AOP）の概略図は、生物学的組織のさまざまなレベルでのいくつかの例を参照して説明されている。

6. AOPの概念は、2009年に環境毒性および化学協学会によって開催されたPellston会議など、近年の多くの議論に基づいて進化している（Villeneuve and Garcia-Reyero、2011; Watanabe et al.,2011; Perkins et al, 2011; Nichols et al., 2011; Celander et al., 2011; Kramer et al., 2011; ENV / JM / MONO（2012）10 / PART1; ENV / JM / MONO（2012）10 / PART2 ; Enoch and Cronin 2010; ENV / JM / MONO（2011）6; Schultz et al., 2011; Hill,1965; US EPA, 2005; US EPA, 2011）。 これらの議論から作成された論文は、既存のデータからの導出や、ゲノミクスデータからのAOPの逆行工学の手法など、AOPのさまざまな側面を扱っている。 OECDが主催する「化学物質分類における機械的情報の使用」という題名のワークショップが2010年12月にワシントンDCで開催され、短期的に（その後2年間）多くの勧告と結論が出された。 これらの推奨事項は次のとおり：

1) 確立された有害性作用（皮膚感作性など）のAOPを開発する際には、相互作用を促進するために、毒物学者や他の科学者をAOPの議論に参加させる
2) 2010年12月のワークショップで始まった概念実証を、いくつかの異なる長期的なヒトの健康および生態毒性指標のAOPを開発することにより完成させる
3) 以下の開発を含む、AOPを特定、評価、および推進し、OECD QSAR Toolboxに統合するための戦略計画を開発する
    a) AOPの開発と評価に使用できる情報雛形
    b) AOPの完全性と受け入れを評価するための一連の指針
    c) AOPの相互承認を達成するための書式
4) AOPに関連する用語を調和させる (OECD 2011, ENV/JM/MONO(2011)8).

7.（1）と（2）の提言に対応して、OECDは、皮膚感作を引き起こすタンパク質結合のAOPを開発した。 図2は、タンパク質への共有結合によって開始される皮膚感作に関連する経路の流れ図を示している（OECD 2012, ENV / JM / MONO(2012)10 / PART1; ENV / JM / MONO(2012）10 / PART2）。 このAOPの開発で得られた経験と提言への取り組みに基づいて（3）と（4）、現在のガイダンスの初期版が開発された。 この初期のガイダンスは、AOPの開発と評価に不可欠な要素と原則を示すだけだったが、本書にて改良された。 詳細はユーザーズハンドブックに記載されている。このハンドブックは、AOP開発者と評価者から受け取ったフィードバックに基づいて、このガイダンスよりも簡単に更新できるように作成されている。

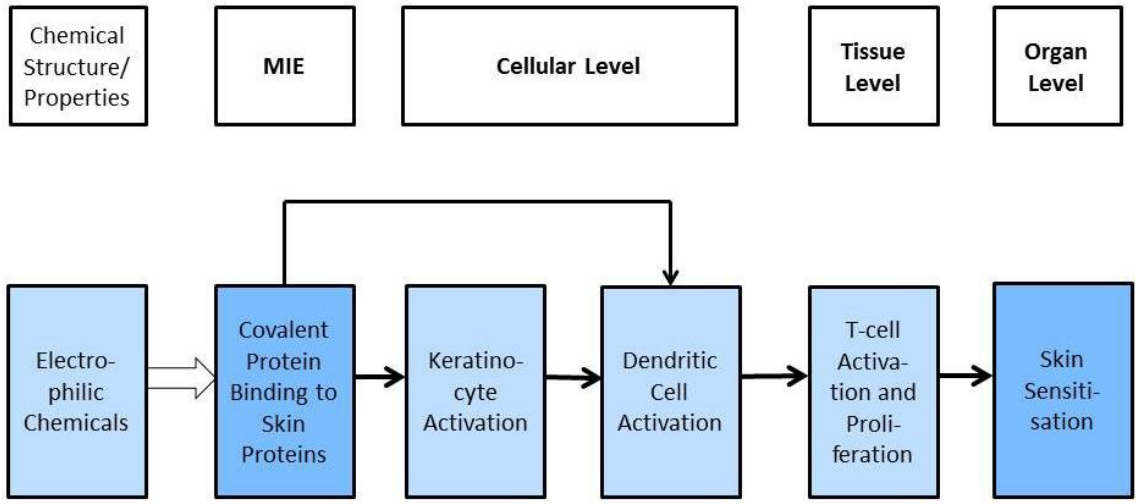


Figure 2. 皮膚感作に関するAOPの例 (出典 OECD 2012, ENV/JM/MONO(2012)10/PART1).

8. AOPの分野での国際的な共同作業を支援するOECD Extended Advisory Group on Molecular Screening and Toxicogenomics（EAGMST）のメンバーは、2014年の国際ワークショップ（http：//www.saaop.org / workshops / somma.html）に参加した。 その結果である論文は、AOPの開発と評価の概念的側面の改良に貢献した（Villeneuve et al,2014a,b; Perkins et al,2015; Garcia-Reyero,2015; Becker et al.,2015; Groh et al,2015a,b; Tollefsen et al.,2014）。 論文の一つ（Villeneuve et al.,2014a）には、AOP開発を導く5つの基本原則が説明されている。

(1) AOPは化学的特異性ではない

(2) AOPは組み換え可能な構成方式であり、主要事象（KE）および主要事象の関係（KER）という名前の再利用可能な要素で構成されている

(3) KEとKERの単一軸で構成される個々のAOPは、AOPの開発と評価の実用的な単位である

(4) 共通のKEとKERを共有する複数のAOPで構成される連絡網は、ほとんどの実際のシナリオの予測機能単位となる可能性がある

(5) AOPは、新しい知識が生成されるにつれて時間とともに進化する生きた文書である

9. 別の論文では、ユーザーズハンドブックに含まれる修正されたBradford-Hillの考慮事項（Becker et al, 2015）に基づいて、AOPの成熟度と信頼レベルの評価の基礎としてWoE分析の事例を示した。 これらの経験から得られた主要な教訓も文書化され、事例と合わせて、AOPの適用に対する信頼性を高めるために必要な文書の性質と形式の共通理解を深めることに寄与している。

10．AOPは単一の軸（たとえば、生物組織のレベル; 図1参照）で表されるが、毒性は多次元（たとえば、性別、種、年齢）であるため、MIEと最終AOの間の経路はかなり異なる。これは、複数の臓器相互作用（例：皮膚感作）、複数の事象（例：反復投与毒性）、経時的蓄積（例：神経毒性）の結果である、より「複雑な」長期指標に特に当てはまるか、または生物の特定の生命段階に関連している（例：発生毒性）。それにもかかわらず、毒性反応を実現するには多くの生化学的段階が必要であるが、MIEはその後のすべての段階の前提条件である（Enoch and Cronin, 2010）。そうは言っても、単一のMIEがいくつかの多段階のシグナル伝達に影響を与える可能性があると理解されており（例えば、Casperse-3のタンパク質発現の減少、Casperse-6の同時活性化）、そのうちの1つ以上が特定のAOに寄与する可能性がある。さらに、AOPは、生物レベル全体ではなく分子レベルで化学的相互作用が始まるという事実に基づいている。したがって、*in vivo*で観察されるAOは、化学物質と内因性生体分子との相互作用によって引き起こされる多段階の生物学的な結果である。

11. AOPは生物学の実用的な単純化であるが、AOPネットワークで表すことができるより広範な状況を発生させると認識できる。 特定のMIEはいくつかの最終結果につながる可能性があり、逆に、いくつかのMIEは同じ最終結果につながる可能性がある。 ただし、AOPは、一つのMIEと一つの最終AOのみに焦点を当てた評価するように設計する必要がある。

12. AOPの各構成物自体は、モデル化されている生物学的システム内で進行している他の経路の影響を受ける場合がある。 さらに、AOPユニットを構成するKEの非分岐の連続は、他のAOPと共有されるKEまたはKERである可能性が高く、AOPネットワークの作成につながる相互作用をつくりうる。 AOPネットワークとして知られるこれらの相互に作用するAOPのシステムは、ほとんどの実際のシナリオをより代弁していると考えられている（Villeneuve et al, 2014a; Knapen et al, 2015）。 AOPは、AOにつながるKEの単一の線形連続として視覚化できるが、AOPネットワークはより広範な画像を提供し、同じAOの発現につながる可能性のある摂動の可能性があるさまざまな利用可能なKEおよびKERを含む。 これらのより複雑なネットワークは、AOP知識根拠（KB）（後述）にある他のツールを使用して表示される。 これらのネットワークは、たとえば、ある器官（脳）のMIEが別の器官（生殖腺）のAOとしてどのように現れるかを示す。

13. OECDは、AOPの開発と適用を可能にするITツールの構築に取り組みんでいる。 AOP-KBは2014年9月25日に開始され、いくつかの基本単位で構成されている。 AOP-Wikiは、KEの共有とAOP開発者間の協力を促進する、開発済みまたは構築中のAOPの中央貯蔵スペースとして機能するオープンソースインターフェイスである。 AOP XPlorerは、AOPネットワークの視覚化を支援するAOP-KBの追加基本単位である。 Effectopediaは、KE間の定量的反応-反応関係、KEの測定に利用できる試験法および生体指標などの追加の構造化情報を把握できる。意思決定段階を促進するために、グラフィカルインターフェイスを介してこの情報を表示する。 4番目の基本単位である中間効果データベース（IEDB）は、AOP-KBを、典型的な化学物質、つまり特定のAOPにつながる科学的推論を裏付ける化合物を使用した*in vitro*アッセイの実際の試験結果につながる。 IEDBはIUCLID事例であり、ほとんどがOECD Harmonized Template（OHT）201を使用する。これは、非古典的な試験方法を対象としている。第三極のIUCLIDシステムからのOHT 201データは、IEDBに簡単に転送できる。 e.AOP.Portalは、AOP-KBの主要な導入点であり、AOP-WikiおよびEffectopediaで現在利用可能なAOPを見つけるための検索メカニズムを提供している。 e.AOP.Portalは、OECD作業計画にあるすべてのAOPの状態を提供する。公開されたすべてのAOPは、この同じサイトからアクセスできる。

14. AOPおよびAOP KBの開発と広報により、さまざまな規制活用が可能になると予想される。たとえば、上記の2014年のワークショップの一環として、Groh et al.（2015a）はAOの予測能を向上させることを目的とした研究を進めるために、AOP概念をどのように使用できるかを調査

した。 さらに、魚の成長を事例研究として使用し、AOP概念を使用して、既存の知識と潜在的な代替試験のギャップを確認するため、特定の慢性毒性事例に利用可能な知識について高い評価を得た（Groh et al, 2015b）。 また、ワークショップグループは、さまざまな規制目的でIATAの開発を通知する際のAOPの役割の解明に焦点を当てており（Tollefsen et al, 2014）、AOPの開発の相対的な程度と科学的信頼度に応じて、AOPのさらなる規制適用の可能性を調査した（Perkins et al, 2015）

15. 最初から、目的の十分性を判断するために定義とチェックリストおよび/または評価枠組みが必要である。開発のあらゆる段階のすべてのAOPが役立つ。ただし、特定の規制状況でAOPを信頼できる範囲は、その開発レベル、許容できる不確実性の段階、および証拠の段階（たとえば、情報とデータの詳細、品質、および量）に関連しており、利用可能である（ユーザーズハンドブックを参照）。 OECDは、AOP-Wikiに入力されたAOPのガイダンス原則、完全性、および科学的堅牢性の順守を評価するために、特定のAOPに関して提出された提案を作業計画に承認する主な責任を持つEAGMSTと連携して機能し、加盟国の承認を求めている。 EAGMSTの承認後、Working Group of National Co-ordinators of the TGs programme（WNT）のワーキンググループとWorking Party on Hazard Assessment（WPHA）は、AOPの承認とその後のJoint Meeting（JM）からの機密解除について責任を負う。

16. この文書の目的は、AOP固有の用語の定義の用語集を含む、一貫した情報収集とAOPへの編成のための枠組みを提供することである。 この文書は、AOPと使用する必要のある用語を開発するために必要な情報の断片についての見通しを示すことを目的としている。 最後に、この文書では、規制状況におけるAOPの潜在的な使用についても簡単に概説している。 IATAにAOPを使用する方法に関するガイダンスが利用可能（OECD, 2016b）だが、リスク評価へのAOPの使用に関するさらなるガイダンスは将来開発される予定である。 AOP WikiでのAOPの開発と実装に関する詳細なガイダンスは、ユーザーズハンドブックに記載されている。これは、得られた経験に基づいて時間とともに進化する現在の文書の補足である。 AOPを開発または評価する場合、現在のガイダンスとユーザーズハンドブックを意味する両方の文書の内容を考慮する必要がある。

有害性発現経路(AOP)の開発

主要情報部分の特定

17. AOPの基本的な構成要素は、通常、生物学的組織の異なるレベルにあるKEsである。これらのKEは、検討中のAOに因果関係があり、不可欠であり、測定可能である。MIE と AO は、単一のAOPの状況下において、特殊なKEsを必要とする。AOPは生物学的標的との化学物質の直接的な相互作用を表すMIEによって一方の端に固定され、もう一方の端には規制上の決定に関連する生物学的レベルの組織であるAOで固定されている。AOP の 2 番目の基本構成物は、主要事象の 関係 (KER) です。KER はAOP の上流の KE と下流の KE の間をつなぐ。これらの情報は、AOP開発時に明確に識別し、記述する必要がある。

18. AOPを開発するために、経路(すなわち経路を開始する分子標的と相互作用する可能性を持つ)を開始できる化学物質の種類を反映する構造アラート、相対反応性または化学的-生物学的相互作用を測定する*in chemico*法、一連の細胞応答(例えば遺伝子発現)を与える*in virto*法、*ex vivo*および*in vivo*機構試験、直接指標(複数可)を測定する*in vivo*試験において、規制上の意思決定を推進するAOに関連する指標を測定する*in vivo*試験を含むさまざまな種類のデータを利用することができる(OECD 2011,ENV/JM(2011)6)。AOPは化学的に特異的ではなく、記述経路は特定の化学初動からの独立を繰り返していることに価値がある。それにもかかわらず、特定のAOの状況下において、化学物質への曝露から得られた実験データは、生物学的応答の様式を理解するために有用である。この情報は、AOP 内の KE および KE を識別し、AOP を裏付ける科学的証拠を提供するために使用される。したがって、AOPは、異なる次元(例えば、異なるレベルの生物学的組織)における効果をAOPの最終指標に結びつける科学的基礎を提供する(図1)。

分子初動事象の定義(作用部位)

19.化学的に誘導された生物学的システムの摂動は、分子レベルに始まる。ほとんどの化学物質は、複数の分子標的と相互作用する。MIEは、AOPの主要なアンカーまたは「基盤」である。したがって、特に、MIE/AOPが*in silico*または*in chemico*予測を支援するために使用されている場合は、評価に関連する特定の最終AOにつながる始まりを明確に識別することが重要である。多くのMIEは、タンパク質および/またはDNAに対する共有結合の形で定義されている。これらのタイプのMIEは、有機化学（すなわち求電気基球反応性）の原理に基づいている。対照的に、「受容体結合」または酵素への結合は、多くの場合、自然界でより選択的である非共有結合相互作用に基づいている。化学物質は、異なるターゲットに対して異なる親和性を持っている。内部曝露が受容体または酵素上の結合部位を飽和するのに十分であるならば、活性の活性化または阻害の効力は毒性を引き起こすかもしれない。MIEの理解は、生物学的利用可能性、構造要件（特に受容体結合）および代謝変化などの摂動を誘発する可能性が最も高い化学物質の特性の同定と定義の理解を容易にする。潜在的な化学的誘導物質を理解することは、同様の方法で作用する化学分類物質の分子構造の限界を定義するのに役立つ。

20.理想的なシナリオでは、MIEが明確に定義されている場合、その事象を引き出す化学物質の可能性が認識されるだけでなく、摂動の活性部位にも注目すべきである。例えば、求電子性の種に対する物質の代謝変換は、皮膚感作および肝線維症に関して同じであってもよいが、作用部位は異なることになる(ケラチノサイト対肝細胞)。いくつかの指標、特に受容体結合機構に基づいての作用部位の同定は、受容体の「立体構造」および他の特性が構造的にそれに結合できる分子の種類を定義するので非常に重要である。

しかし、MIEの作用の部位の同定が非常に困難であるかもしれない（例えば、反復投与）または正確に定義されていない（例えば、魚の単純ナルコシス）AOが多数存在する。 定義されてい

ない機能部位を持つAOPは有用ではなく、必要ではない。

有害発現の特定

21. AOは、一般的に、承認された規制ガイドライン毒性試験での確立された保護目標または先端指標に対応する規制的に重要性であるとして承認されている専門的なタイプの主要事象である(OECD、2016a)。AO は、さまざまな次元 (例えば、曝露の　持続時間、性別、種など)に基づいて定義できる。最終的なAOは、個人レベルまたは集団レベルにおける特定のAOPのアンカーである。規制上の意思決定に関連する最終的なAOを明確かつ正しく定義することが不可欠である（すなわち、それは確立された規制ガイダンスの承認された保護目標または共通の先端指標に対応する)。これは、ヒト健康の場合、個人または集団全体または特定された集団のいずれかに、特定の臓器または器官系における病状のリスクの増加を促すAOにつながる事象の機械学的連続を説明するのに役立つ。

一方、生態系では、野生動物の集団持続可能性の推定値の観点から意味を持つ集団統計学的意義の結果であることが最も多い。生存、胎児性または成長に関連した結果は、容易に集団の持続可能性に関係する。しかし、生物の多くの構造的（例えば、重大な異常）および機能的（例えば、行動異常）な変化は、他の要因と組み合わせると調節的意義を有し、ヒトおよび生態学的リスク評価における貴重な証拠を付加する可能性がある。少なくとも、AOPは、組織の器官レベル以上で規制上重要な少なくとも一つのAOに関連する必要がある。可能な場合は、ヒトの健康と生態学的リスク評価のためのAOPの有用性を最大化するために、KERの連続を集団レベルまで拡張する必要がある（適用範囲内にて）。

有害経路につながる主要事象の特定

22. KEは、通常、さまざまなレベルの生物学的組織で、中間事象経路に沿った段階として定義される。KEになるためには、AOPの中間段階にあることが不可欠であり、実験的に測定可能でなければならない。 ユーザーハンドブックには、AOPにKEを含めるために使用できるデータの種類に関する追加の考慮事項と事例が記載されている。

23. AOPには、評価に関連する最終AOとMIEの間にあるKEの収集が含まれる。隣接する事象間の関係は、多くの場合、すべての中間事象を記述するのではなく、経路に沿って発生する主な効果をとらえ、重要で測定可能なKEを含めることができる方法と定義できる。 理想的なシナリオでは、AOPには、MIE（アンカー1）と最終AO（アンカー2）の因果関係/接続を確立するために必要な比較的少数または最小限のKEを含める必要がある。 AOPを評価するための支援とデータを提供するため、KEおよびKERに関するさまざまな*in vivo*および*in vitro*の情報、およびハイスループットスクリーニング（HTS）アッセイからの情報、ハイコンテンツスクリーニング（HCS）からの指標、オミクスアプローチさらに、*in silico*法も使用できる。 KEの数が増えると、毒物学的な複雑さが明らかになる。

24. AOにつながるKEを特定する前に、正常な生理学的経路の理解が不可欠である（例：生殖段階、肝機能）。これは、生物学的組織のさまざまなレベルでかく乱する可能性のある過程の複雑なネットワークの認識に役立つ。 KEの特定中に、既存の文献のレビューは、最終的な悪影響につながる妥当な機構と中間段階について可能な限り多くの情報を評価するために必要である。入手可能な文献のKEデータの信頼性と関連性を判断するには、最終AOに関する比較と解釈のために、研究計画の重要な因子（曝露摂生、曝露期間、サンプリング時間）の評価が含まれる。系統的**レビュー**を含む自動化された文献を掘り起こし、評価ツールは、AOPの開発を加速し、データ評価の客観性と透明性を高めるのに役立つが、必須ではない。重要な点は、AOPの開発は科学文献によって支援されるべきであり、それがどのように達成されるかは、ユーザーズハンドブック

で議論されている主要な考慮事項が扱われる限り、AOPの開発者次第である。通常、複数の中間事象は、特定のAOPの構築中に同定される。 したがって、集められた知識は、関心のある特定のAOPに適切であるというフィルターをかける必要がある。 KEが複数のAOPに存在する場合、情報はAOP間で共有できる。

主要事象関係の特定

25. AOPで識別されたすべてのKEは関連し、KERで定義される。 AOPでのKERの記述には、上流のKEの既知の状態または測定された状態から下流のKEの推定変化または状態を推定するための科学的根拠を定義する情報と証拠の種類を集めて整理することが含まれる。

26. 慣例により、KERは2つの形式のいずれかをとることができる（OECD、2016a）。 KERを介してリンクされたKEのペアは特定のAOPを定義するKEの連続で互いに隣接している場合がある。あるいは、KERは、関係が別のKE（つまり、AOP内の非隣接KE）を通過すると考えられるKEのペアを指す場合がある。隣接しない可能性のあるKEの可能な２つ毎にKERを記述する必要はない。ただし、隣接していないKEのKER記述を提供するオプションは、特にAOP Wiki内で役立つ。隣接するKE間の関連を支援する経験的証拠が利用できないか、直接隣接していないKEのみが利用できるためである。 たとえば、一部のKEの測定は、通常の研究ではめったに行われないように、行うのがかなり難しい場合がある。 AOPの一部としてKEを確立するのに十分なデータがあるかもしれないが、利用可能なWoEの多くは、その特定のKEを無視するか、「飛び越える」ことができる。隣接しないKEの記述を含めると、これらの関係のWoEを記述し、他のAOPに連結できる。

27. AOP開発の重要な要素は、KERのWoEの評価である。これには、関係の生物学的妥当性の評価、および最初のKEが後続のKE以下の用量で、下流のKEよりも早い時点で発生することを裏付ける経験的証拠の検査が含まれる。事例などの詳細情報は、以下およびユーザーズハンドブックの補遺にある。

AOP評価

28. 規制目的での使用を意図して開発された特定のAOPの基礎となる証拠は、MOA背景でのWoEの比較分析に使用される進化したBradford-Hillの考慮事項に基づいて評価される（Meek et al., 2014a; Meek et al., 2014b）。 ただし、必要に応じて、化学的にとらわれないAOPの状況に対応するように変更されており、ユーザーズハンドブック（OECD、2016a）に記載されている。

29. 3つの主要な考慮事項は次の通り

1) KERの生物学的妥当性。これは、関与する基本的な生物学的段階と、それらがAOPで提案されている因果関係と一致しているかどうかの理解に依存している。
2) KEの本質。AOP全体の状況下で考慮され、上流のKEがブロックされた場合に下流のKEまたはAOが防止または変更されるかの実験データを指す（ノックアウトモデルでの試験または 可逆性）
3) KERの経験的支援。これは、多くの場合、KEペアの用量反応と時間的一致を評価できる一津つ以上の参照化学物質から得られた毒性データに基づいている。

30. AOPを評価する場合、生物学的妥当性と経験的支援の両方がKER毎に個別に評価されるが、KEのそれぞれの支援情報に基づいて、AOPの状況下では本質性が考慮される。 その段階は、前述のようにAOP間の構成物の共有を支援するだけでなく、AOP内の重要なデータのギャップと不確実性を明確に特定し、規制活用に目標を絞った研究と評価を促進するのに役立つ。 WoE（生物学的妥当性と経験的支援で構成される）が各KERで評価され、各KEで本質性が評価されると、

AOP全体を支援する証拠をユーザーハンドブック（OECD,2016a）の指示に従って表にまとめることができる。

AOPの利用

31. リスク評価の基礎として有害性の定量化を支援するには、数学的にモデル化できる記述的な定量的KERを備えた、よく記述説明されたAOPが通常必要である。 しかし、明確に定義された定性AOPでさえ、さまざまなレベルの生物学的組織を通じて正確に記述された一連の事象で、多くの目的に使用できる貴重な機構情報を提供できる（OECD 2011、ENV / JM（2011）6）。

32. AOPのさまざまな潜在的な用途が説明されている。 特定のAOPの決定を支援できる範囲は、不確実性のレベルとKERの定量的理解に依存する。たとえば、KEを特定して説明することにより、AOPは特定の方法における使用の理論的根拠を説明し、潜在的により予測的な方法の開発により、OECD試験法ガイドラインプログラムの作業に通じる（以下でさらに説明）。 AOPは、IATAまたは統合的試験戦略（ITS）を開発するための基礎としても使用できる。 また、区分を最初に形成し、区分内でデータギャップを埋めるRead Acrosssなど、代替アプローチのさらなる開発と適用にも使用できる。これにより、従来の生体内試験の改良、削減、および/または置換が可能になる。

33. AOPは、以下に限定されない多くの規制状況に貢献するために使用できる：そのため、（1）から（4）に進むと、許容できる不確実性のレベルが低下し、AOP開発を支援する際に提示される証拠のレベル（詳細、品質、および情報とデータの量など）が増加する。

34. 部分的に開発されたAOP（つまり、すべてのKEが知られているわけではないもの）は、さらなる試験と開発の優先順位設定に役立つ場合がある。 同様に、現在OECD QSAR Toolboxで実行されているように、部分的に開発されたAOPが有害性識別に使用される場合がある。 生理学に基づいた薬物動態（PBPK）モデリングおよび吸収、分布、代謝、排泄（ADME）に関するトキシコキネティクス情報は、AOP開発では考慮されないが、上記の規制状況のいずれかでAOPの適用に対処する必要がある。 AOPは、特定の化学物質のMOA分析の開始点としても機能し、化合物空間とADMEの考慮を取り入れることができる。

35. 定性的AOPは、AOPを支援するWoE全体の定性的評価、妥当性または統計的推論に加えて、KEの測定方法の説明およびKEが経験的証拠によって支援される記述によってKEが支援される（Villeneuve et al., 2014a）。 対照的に、定量的AOPは、KEの測定方法の説明によって支援されるKEの機械言語と、変化の大きさおよび/または持続時間の定量的理解によって支援されるKERとともに測定が行われる精度と精度に基づいている 下流のKEにある程度の変化を引き起こすには、上流のKEが必要である（Villeneuve et al., 2014a）。

36. AOPはその開発中に、推定AOPとして始まり、より多くの情報が蓄積されるにつれて定量的AOPに進化する可能性がある。一般に、AOPは推定、完全定性、完全定量の間の連続体に沿って存在する。 AOPは決して完全ではないことに注意することが重要である。AOPは進化し続けることができるが、意思決定を支援するためにその使用を妨げるものではない。 OECD内でのAOPの潜在的な用途を以下に説明する。

化学品区分の開発とOECD QSARツールボックスのさらなる開発

37. AOPの主な対象活用の一つは、Read acrossおよび化学物質区分形成の分野である。 AOPは、一つまたは複数のAOP内のMIEとKEの類似性に対処することにより、OECD QSAR Toolboxを使用する場合、グループ化アプローチにおける物質の生物学的類似性に関する作用機構情報を提供

できる。 さらに、AOPは、特定のKEの試験に焦点を当てることにより、化学物質区分内の洗練された試験法戦略の識別を支援し、構造的に類似した物質の共通の作用モードに対処できる（試験と評価のための統合アプローチの状況下でのAOPの使用も参照）。

38. 実証されているように、OECD QSAR Toolboxのバージョン3.0で皮膚感作を引き起こすタンパク質結合については、AOPを使用して化学物質区分を開発および改良できる。 この例では、3つの情報セットが照合および統合される。（1）評価で通常使用される生体内効果のライブラリ（例、局所リンパ節試験のEC3値）、（2）MIEのライブラリ（例、タンパク質結合反応）、および（3）中間事象のライブラリ、通常は*in vitro*法を使用して作製されたデータ（樹状細胞表面バイオマーカーなど）（OECD 2012, ENV / JM / MONO（2012）10 / PART1; ENV / JM / MONO（2012） 10 / PART2）。 理論的には、各指標は単一または複数の化学物質の範囲内に関連付けることができる。 化学物質の区分に関しては、対象となる化学構造空間、または適用領域は、AOP内のMIEおよびKEに対して評価される化学物質に依存する。

試験法ガイドラインプログラム

39. KEを特定して説明することにより、AOPは試験法ガイドラインプログラムの作業に通じる。 実際、KEが特定されると、細胞またはより高いレベルの生物学的組織で直接的な化学効果または応答を検出する*in vitro*および*ex vivo*試験法の開発、および特定されたMIEに関連する目的のスクリーニング試験の開発を提案できる（ OECD 2011, ENV / JM / MONO（2011）8）。 逆に、*in vitro*試験法の開発に関する提案をAOPのKEに連結することにより、規制目的に関連する有害性指標との関係を確立できる。

40. たとえば、皮膚感作をもたらすタンパク質結合についてAOPで特定された2つの方法が、試験法ガイドラインの開発のためにOECDに提案されている。ケラチノセンスアッセイ（ヒトケラチノサイトでの遺伝子発現）（OECD、TG No.442CおよびD）および h-CLATアッセイ（ヒト単球細胞における細胞表面マーカー(CD86) 発現）（OECD、TG No. 442E）である。

41. しかし、単一のAOPにより規制に関連する可能性のあるすべての事象を捉えることはほとんどない。 少なくとも一つの共通要素を共有するAOPに基づくAOPネットワークは、潜在的にAOに至る経路とネットワークの変更より現実的になる。 これらのAOPネットワークの分析は、複数の結果の予測実用性を備えた単一の試験法を開発することであろうと、規制上の懸念に関する特定の指標を予測するために非常に特異的な一連の試験法を開発することであろうと、試験法開発の優先順位付けに役立つ。 たとえば、魚の生殖および発生毒性に関連する5つのAOPを使用して、AOPネットワークを記述し、AOPネットワークを試験法の開発と改良に使用する方法を説明している(Knapen et al., 2015)。

試験法と評価のための統合アプローチ（IATA）の状況下でのAOPの使用

42. AOPは、特定の指標のIATAまたはITSを開発するための基盤として機能する。 AOPは、どの追加情報（および、もしあれば、どの試験）が初動事象と悪影響に関連する確実性を高めるかを決定するのに役立つ。さらに、確立されたAOPを使用して、種から種への外挿に使用できる。 AOPの基礎となる信頼性と信頼性、およびIATAでのそれらの適用が提案されている(Tollefsen et al., 2014; Patlewicz et al., 2015; Perkins et al., 2015)。 IATAの開発と使用のための枠組みに関する2014年に開催されたワークショップ（OECD, 2015年）に基づいて、IATAの開発におけるAOP概念の使用のための手法の概要を記したガイダンスが利用可能である。

要約

43. リスク評価の予測戦略を実施するには、MIEまたはMIEに対する細胞応答に焦点を合わせた*in vitro*毒性試験の結果を、生物および最終的には集団への影響に外挿する必要がある。 これは、MIEとAOを因果的に関係するAOPの開発により実現できる。 これらのAOPは規制当局による使用を目的としているため、このガイダンスでは、AOPの開発、評価、およびAOPの開発、文書化、レビューの方法の標準化を目的としている。 承認されたAOPは、その開発とレビューに続いて、AOPに関するOECDシリーズで公開されている。 ただし、科学的知識が進歩するため、このシリーズのAOPの公開は、そのAOPへのさらなる更新または新しい貢献を妨げるものではない。

44. AOPは、化学的有害性分類を開発または改良するための、透明性があり、作用機構に基づいた枠組みを提供し、目的となる*in vitro*および*in vivo*試験の提案および優先順位付けする必要がある。 構造活性相関（SAR）および*in chemico*および*in vitro*試験から、より低いレベルの生物学的組織における影響の可能性を理解することにより、より高いレベルの生物学的組織（たとえば*in vivo*試験）で追加の試験が必要かどうかを効率的に判断できる（ Meek et al., 2011）。 ここで提供されるガイダンスは、MOA枠組み（2012年にWHO / IPCSにより更新; Meek et al., 2014a）で提示され、ユーザーズハンドブックに記載された進化するMOA分析の組み込みとともに、意思決定プロセスにおける機械的データと計算モデルの組み込みを支援する。

45. Bauchらにより示されているように、AOP内のすべてのKEが、評価で使用されるために完全に記述されている必要はない（Bauch et al., 2011)。特定の目的のためのAOP またはAOPネットワークの使用には、 各KERおよび 全体的なAOPと同様に、各KERの評価の基礎である最終AOにつながるMIEおよびKEに関する情報の結果を含んでいる。AOPの十分な知識と考えられるものは用途に依存する可能性があり、潜在的な影響が大きな応用により知識や信頼度が高くなる（Meek et al.,2011）

参考文献

Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrano, J.A., Tietge, J.E., Villeneuve, D.L. (2010). Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment. Environ. Toxicol. Chem. 29: 730-741.

Bauch, C., Kolle, S.N., Fabian, E., Pachel, C., Ramirez, T., Wiench, B., Wruck, C.J., van Ravenzwaay, B., Landsiedel, R. (2011). Intralaboratory validation of four in vitro assays for prediction of the skin sensitizing potential of chemicals. Toxicol. In Vitro 25: 1162-1168.

Becker, R.A., Ankley, G.T., Edwards, S.W., Kennedy, S., Linkov, I., Meek, B., Sachana, M., Segner, H., Van Der Burg, B., Villeneuve, D.L., Watanabe, H., Barton-Maclaren, T.S. (2015) Increasing Scientific Confidence in Adverse Outcome Pathways: Application of Tailored Bradford-Hill Considerations for Evaluating Weight of Evidence. Regul. Toxicol. Pharmacol. 72: 514-537.

Boobis, A.R., Cohen, S.M., Dellarco, V., McGregor, D., Meek, M.E., Vickers, C., Willcocks, D., Farland, W. (2006). IPCS Framework for Analyzing the Relevance of a Cancer Mode of Action for Humans. Crit. Rev. Toxicol. 36(10): 781-792.

Boobis, A.R., Doe, J.E., Heinrich-Hirsch, B., Meek, M.E., Munn, S., Ruchirawat, M., Schlatter, J., Seed, J., Vickers, C. (2008). IPCS Framework for Analyzing the Relevance of a Noncancer Mode of Action for Humans. Crit. Rev. Toxicol. 38(2): 87-96.

Celander, M.C., Goldstone, J.V., Denslow, N.D., Iguchi, T., Kille, P., Meyerhoff, R.D., Smith, B.A., Hutchinson, T.H., Wheeler, J.R. (2011). Species extrapolation for the 21st century. Environ. Toxicol. Chem. 30: 52-63.

Clewell, H.J., Gentry, P.R., Gearhart, J.M., Allen, B.C., Andersen, M.E. (1995). Considering pharmacokinetic and mechanistic information in cancer risk assessments for environmental contaminants: examples with vinyl chloride and trichloroethylene. Chemosphere 31: 2561-2578.

Enoch, S.J., and Cronin, M.T.D. (2010). A review of the electrophilic reaction chemistry involved in covalent DNA binding. Crit. Rev. Toxicol. 40: 728-748.

Groh, K J., Carvalho, R., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K., Watanabe, K. H. (2015a). Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology Chemosphere 120: 764-777.

Groh, K.J., Carvalho, R., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K., Watanabe, K.H. (2015b). Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: II. A focus on growth impairment in fish Chemosphere 120: 778-792.

Hill, A.B. (1965). The environment and disease: Association or causation? Proc. R. Soc. Med. 58: 295-300.

Knapen, D., Vergauwen, L., Villeneuve, D.L., Ankley, G.T. (2015) The potential of AOP networks for reproductive and developmental toxicity assay development. Reprod. Toxicol. 56: 52-55.

Knudsen, T.B. and Kleinstreuer, N.C. (2011). Disruption of embryonic vascular development in predictive toxicology. Birth Defects Res. 93: 312-323.

Kramer, V.J., Etterson, M.A., Hecker, M., Murphy, C.A., Roesijadi, G., Spade, D.J., Spromberg, J., Wang, M., Ankley, G.T. (2011). Adverse outcome pathways and ecological risk assessment: bridging to population-level effects. Environ. Toxicol. Chem. 30: 64-76. Krewski, D., Acosta, D.Jr., Andersen, M., Anderson, H., Bailar, J.C. 3rd., Boekelheide, K., Brent, R., Charnley, G., Cheung, V.G., Green, S. Jr., Kelsey, K.T., Kerkvliet, N.I., Li, A.A., McCray, L., Meyer, O., Patterson, R.D., Pennie, W., Scala, R.A., Solomon, G.M., Stephens, M., Yager, J., Zeise, L. (2010) Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health B Crit Rev. 13: 51-138.

Meek, M.E., Boobis, A.R., Crofton, K.M., Heinemeyer, G., Raaij, M.V., Vickers, C. (2011). Risk assessment of combined exposure to multiple chemicals: A WHO/IPCS framework. Regul. Toxicol. Pharmacol. 60: S1-S14.

Meek, M.E., Boobis, A., Cote, I., Dellarco, V., Fotakis, G., Munn, S., Seed, J., Vickers, C. (2014a). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 34: 1-18.

Meek, M.E., Palermo, C.M., Bachman, A.N., North, C.M., Jeffrey Lewis R (2014b). Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J. Appl. Toxicol. 34: 595-606.

Nichols, J.W., Breen, M., Denver, R.J., DiStefano, J.J. III, Edwards, J.S., Hoke, R.A., Volz, D.C., Zhang, X. (2011). Predicting impacts on vertebrate endocrine systems. Environ. Toxicol. Chem. 30: 39-51.

NRC (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

OECD (2002) Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies. OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 35 and Series on Pesticides No. 14. ENV/JM/MONO(2002)19.

OECD (2011). Provision of knowledge and information: hazard assessment: follow-up from the OECD workshop on using mechanistic information in forming chemical categories: a long-term vision for the (Q)SAR project. ENV/JM(2011)6.

OECD (2011). Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 138. ENV/JM/MONO(2011)8.

OECD (2012). The adverse outcome pathway for skin sensitization initiated by covalent binding to proteins. Part 1 scientific evidence. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 168. ENV/JM/MONO(2012)10/PART1.

OECD (2012). The adverse outcome pathway for skin sensitization initiated by covalent binding to proteins. Part 2 Use of AOP to develop chemical categories and integrated assessment and testing approaches. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 169. ENV/JM/MONO(2012)10/PART2.

OECD (2015). Report of the workshop on a framework for the development and use of integrated approaches to testing and assessment. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 215. ENV/JM/MONO(2015)22.

OECD (2016a) Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. OECD Series on Adverse Outcome Pathways, No. 1, OECD Publishing, Paris.

OECD (2016b) Guidance document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 257. ENV/JM/MONO(2016)64.

Patlewicz, G., Simon, T. W., Rowlands, J. C., Budinsky, R. A., Becker, R. A. (2015). Proposing a scientific confidence framework to help support the application of adverse outcome pathways for regulatory purposes. Regul. Toxicol. Pharmacol. 71: 463-477.

Perkins, E.J., Chipman, J.K., Edwards, S., Habib, T., Falciani, F., Taylor, R., Van Aggelen, G., Vulpe, C., Antczak, P., Loguinov, A. (2011). Reverse engineering adverse outcome pathways. Environ. Toxicol. Chem. 30: 22-38.

Perkins, E.J., Antczak, P., Burgoon, L. D., Falciani, F., Garcia-Reyero, N., Gutsell, S., Hodges, G., Kienzler, A., Knapen, D., McBride, M., Willett, C. (2015). Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. Toxicol. Sci. 148: 14-25.

Schultz, T.W. (2010). Adverse outcome pathways: A way of linking chemical structure to in vivo toxicological hazards. In: Cronin, M.T.D. and Madden, J.C. eds., In Silico Toxicology: Principles and Applications, The Royal Society of Chemistry pp. 346-371.

Schultz, T.W., Diderich, B., Enoch, S. (2011). The OECD Adverse Outcome Pathway Approach a case study for skin sensitisation. In: Seidle, T. and Spielmann, H. Eds., Alterative Testing Strategies Progress Report 2011 & AXLR8-2 Workshop Report on a Roadmap to Innovative Toxicity Testing. pp. 288-300.

Shuey, D.L., Setzer, R.W., Lau, C., Zucker, R.M., Elstein, K.H., Narotsky, M.G., Kavlock, R.J., Rogers, J.M. (1995). Biological modeling of 5-fluorouracil developmental toxicity. Toxicology 102: 207-213.

Sonich-Mullin, C., Fielder, R., Wiltse, J., Baetcke, K., Dempsey, J., Fenner-Crisp, P., Grant, D., Hartley, M., Knaap, A., Kroese, D., Mangelsdorf, I., Meek, E., Rice, J.M., Younes, M. (2001). International Programme on Chemical Safety. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regul. Toxicol. Pharmacol. 34: 146-152.

Tollefsen, K.E., Scholz, S., Cronin, M.T., Edwards, S.W., de Knecht, J., Crofton, K., Garcia-Reyero, N., Hartung, T., Worth, A., Patlewicz, G. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). Regul. Toxicol. Pharmacol. 70: 629-640.

U.S. EPA (2005). Guidelines for Carcinogen Risk Assessment. Washington, DC: Risk Assessment Forum. March. EPA/630/P-03/001B. http://www.epa.gov/raf/publications/pdfs/CANCER_GUIDELINES_FINAL_3-25-05.pdf

US EPA (2011) Integrated Approaches to Testing and Assessment Strategies: Use of New Computational and Molecular Tools. SAP Minutes No. 2011-04. http://www.epa.gov/scipoly/sap/meetings/2011/may/052411minutes.pdf.

Villeneuve, D.L. and Garcia-Reyero, N. (2011). Vision & strategy: Predictive ecotoxicology in the 21st century. Environ. Toxicol. Chem. 30: 1-8.

Villeneuve, D., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landersmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M. (2014a). Adverse outcome pathway (AOP) development I: Strategies and principles Toxicol. Sci. 142: 312-320.

Villeneuve, D., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landersmann, B., Lettieri, T., Munn, S., Nepelska, M., Ottinger, M.A., Vergauwen, L., Whelan, M. (2014b). Adverse outcome pathway development II: Best practices. Toxicol. Sci. 142: 321-330.

Volz, D.C., Belanger, S., Embry, M., Padilla, S., Sanderson, H., Schirmer, K., Scholz, S., Villeneuve, D. (2011). Adverse outcome pathways during early fish development: a conceptual framework for identification of chemical screening and prioritization strategies. Toxicol. Sci. 123: 349-358.

Watanabe, K.H., Andersen, M.E., Basu, N., Carvan, M.J. III, Crofton, K.M., King, K.A., Sunol, C., Tiffany-Castiglioni, E., Schultz, I.R. (2011). Defining and modeling known adverse outcome pathways: domoic acid and neuronal signaling as a case study. Environ. Toxicol. Chem. 30: 9-21.

補遺 I: 有害性発現経路に関する用語集

このガイダンス文書に記載されている用語は、アルファベット順に整理されている。 以下の用語のいくつかは、定義が大きく重複しているさまざまなソースで説明されている。 ただし、これらの用語の最も進化した完全な定義は、考慮のためにここに含まれている。

ADME
薬物動態/毒物動態学および薬理学/毒物学で用いられる吸収（Absorption）、分布（Distribution）、代謝（Metabolism）、排泄（Excretion）の頭字語。生物内の医薬品/化学化合物の性質を示す。 4つの段階はすべて、薬物/化学物質のレベルと組織への薬物曝露の動態に影響を与えるため、化合物の性能と薬理/毒物学的活性に影響する（Pharmacology Study Guide、2007）。

Adverse outcome
有害作用は、確立された保護目標への対応または承認された規制ガイドライン毒性試験（OECD、2016a）の先端指標との同等性に基づいて規制上の重要性があるとして一般に受け入れられている特殊なタイプの主要事象である。

注：保護の目標が人間の健康か環境の健康かによって、考慮される指標は異なる場合がある。

Adverse Outcome Pathway (AOP)

概念的には、AOPは、ストレッサーと標的細胞または組織内の生体分子との最初の相互作用から始まる一連の事象（つまり、分子開始事象）として見ることができ、一連の依存事象を経て進行し、有害な結果に至る代償作用とフィードバックループが克復されると、AOPは通常、1つの重要な事象から別の事象に移動して順番に表される (OECD, 2016a)。

Apical endpoint
先端指標は、死亡、発達異常、繁殖行動、生殖障害、身体的変化、および疾患状態を示す臨床徴候または病理学的状態を含む臓器の大きさおよび組織病理学の変化など、曝露の経験的に検証可能な結果である (Krewski et al., 2011; Villeneuve and Garcia-Reyero, 2011)。

注：先端とみなされる指標（結果）は、人間の健康と生態学的な健康の代理として使用される場合、異なる場合がある。

Cellular response

化学物質の対応する受容体への結合は、細胞内で最終的にその攪乱を変化させる事象を誘発する。これらの細胞内事象の性質は、受容体の種類によって異なる。 また、同じ化学シグナルが異なる細胞タイプで異なる反応を引き起こす可能性がある (http://global.britannica.com/EBchecked/topic/101396/cell/37445/Cellular-response).

Chemical category

物理化学的および人間の健康および/または環境毒物学的特性および/または環境運命特性が類似している可能性が高い、または構造的類似性（または他の類似性特性）の結果として規則的なパターンに従う化学物質のグループ (OECD, 2007).

Effectopedia

Effectopediaは、AOP開発、モデリング、使用のための共同研究プラットフォームである。知識（定性的および定量的）は入れ子になった層で構成される。経路表現の最も抽象的な層は、経路構造の視覚的な図である。次のレベルの詳細は、個々の要素（化学物質、効果、リンク、試験法、in silicoモデルなど）のカスタムビルドインターフェイスで取り込まれ、構造化された要約情報の統一表現を提供する。Effectopediaの全体的な目標は、定量的AOPの開発と規制上の意思決定の状況（Aladjov、パーソナルコミュニケーション）での使用を可能にするために必要なすべての情報を単一源に集約することである。

Endpoint

*in chemico*法、*in vitro*試験法または*in vivo*試験から得られ、記録された結果(OECD, 2011).

Integrated Approaches to Testing and Assessment (IATA)

試験法と評価への統合アプローチは、化学物質の区分識別、有害性評価、および/または安全性評価に使用される複数の情報源に基づく手法である。IATAは、関連するすべての既存の証拠を統合および重み付けし、必要に応じて、潜在的な有害性および/またはリスクに関する規制上の意思決定を通知するために、新しいデータの目的形成を説明する。 IATA内では、さまざまな情報源（物理化学的特性、*in siloco*モデル、グループ化およびRead Across アプローチ、*in vitro*試験法、*in vivo*試験、およびヒトデータ）からのデータが評価および統合され、有害性および/またはリスクに関する結論が導き出される。この段階内で、非動物試験および非試験方法で生成されたデータの組み込みは、動物での試験の削減に大きく貢献すると予想される（OECD、2016b）。 IATAの出力は、他の考慮事項とともに、規制上の意思決定に役立つ結論である (OECD, 2016b)。

Integrated Testing Strategy (ITS)

試験および評価に対する定義済みの手法はさまざまな方法で設計でき、たとえば、連続試験戦略（STS）または統合的試験戦略（ITS）の形式をとることができる。ITSはさまざまな特定の方法論を適用して、異なる情報源からの入力を予測に変換することにより、複数のデータまたは情報源を同時に評価する手法である。この目的のために、統計モデルや数学モデルなど、さまざまな特定の方法論を適用できる (OECD, 2016b)。

Key Event (KE)

主要事象は、特定の有害な結果につながる定義済みの生物学的摂動の進行に測定可能で不可欠な生物学的状態の変化(OECD, 2016a)。

Key event relationship (KER)

主要事象の関係は、一つの主要事象を別の主要事象に接続し、二つの間の有用な関係を定義し（つまり、一方を上流として、もう一方を下流として識別）、上流に主要事象の既知、測定、あるいは予測状態から下流の状態の推論または外挿を容易にする科学的根拠の関係（OECD, 2016a)。

Levels of biological organisation

原子、分子、細胞、組織、器官、器官系、生物（個体）、集団、コミュニティ(図１参照)

(Villeneuve and Garcia-Reyero, 2011)。

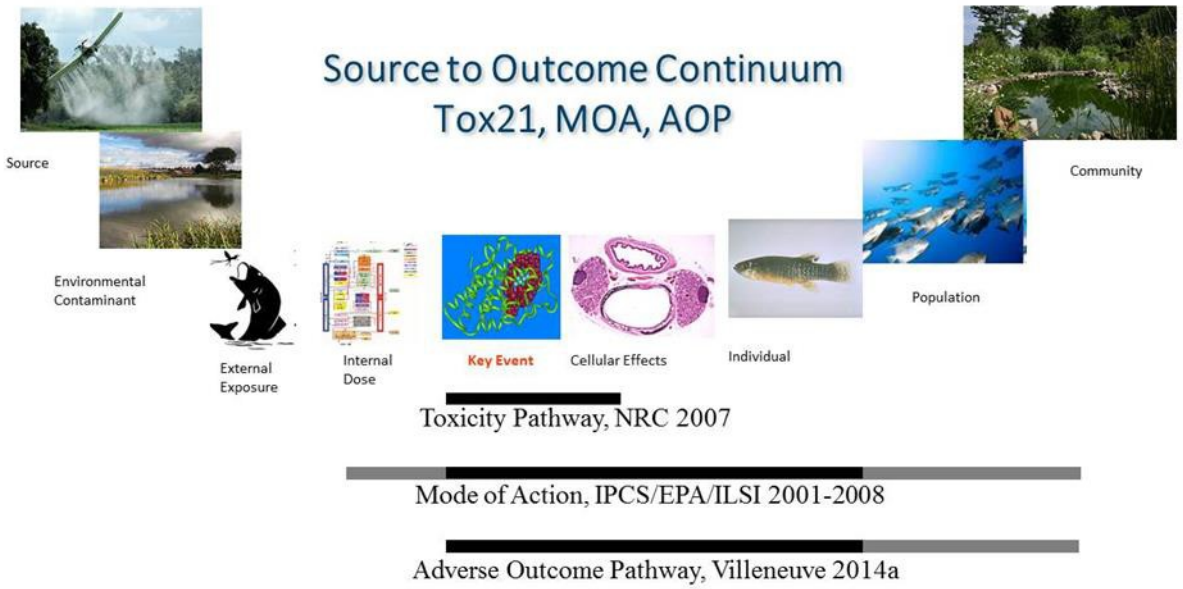


図1. 毒性経路、作用機序経路および有害転帰経路の間の関係の表現。黒いバーはこれらの概念に共通する幅広い研究を表す。 灰色のバーは、概念の理論的な範囲を表す (adapted from OECD 2011).

Mechanism of action

毒性の作用機序は、がんまたはその他の健康指標の誘発における重要な事象の詳細な分子記述である。作用機構は、作用形態が意味するものよりも詳細な事象の理解と説明を表す(North American Free Trade Agreement NAFTA, 2011)。

Mode of action (MOA)

作用機序は、WHOによって「生物学的にもっともらしいKEの連続として定義されており、粗い実験的観測と機械的データによって裏付けられた観測された効果をもたらす。作用形態は、重要な細胞学的および生化学的事象、つまり、測定可能であり、観測された効果に必要な事象を論理的な枠組みで説明する。 World Health Organization (2009) Environmental Health Criteria 240: Principles and Methods for the Risk Assessment of Chemicals in Food. WHO, Geneva, (Definitions page A-25) http://www.who.int/foodsafety/publications/chemical-food/en/.

Molecular Initiating Event (MIE)

分子開始事象は、AOPを開始する攪乱をもたらす生物内の分子レベルでの化学的相互作用の初期点を表す特殊なタイプの主要事象である (OECD, 2016a)。

Molecular screening

分子スクリーニングは、迅速なスクリーニング法とトキシコゲノミクスを、生化学および細胞ゲノム法をカテゴリー分析に適用する目的と組み合わせている。分子毒性スクリーニングの前提は、化学物質と何らかの形の細胞標的との相互作用によって引き起こされる。最初に毒性を評価するには、適切な懸念対象を特定する必要があり、懸念の化学物質と相互作用の可能性を評価するには適切な試験が必要である。 (OECD, 2008).

Pathway perturbation

有害な健康影響が発生する程度まで正常な生物学的機能を損なう可能性のある環境因子またはその代謝物による毒性経路の重大な変化(Krewski et al., 2011).

Site of action

作用部位は、化学物質と相互作用する生体分子であるか、受容体のリガンド結合ドメインなど、目的の高分子上のより特異的な部位を指す。作用部位は、分子開始事象が発生する特定の細胞または組織タイプの状況と見ることもできる（シュルツ、個人的なコミュニケーション）。 作用部位は種固有である場合があることに注意。

Structural alerts

構造アラートは、原子ベースのフラグメントであり、分子内に存在する場合、化合物を特定のカテゴリーに配置できることを示す(Schultz, 2010)。

Systems biology

システム生物学は、動的に相互作用するネットワークの生物学として定義される。それは、生の全体を形成するネットワークはその部分の合計以上のものであるという理解から始めて、生物システムの複雑さを解読することを目的とした生物医学の全体論的アプローチである。Systems biologyは、生物学、コンピューターサイエンス、エンジニアリング、バイオインフォマティクス、物理学、およびその他の分野のアプローチを統合し、これらのシステムとネットワークが化学的曝露を含むさまざまな条件下でどのように変化するかを予測する。バイオインフォマティクス、データ統合、モデリングに加えて、さまざまな分析プラットフォームの使用が必要である(Jain, 2010)。

注：システム生物学には、（1）大量の実験データの収集（ハイスループット技術および/または還元主義分子生物学および生化学の文献のマイニングによる）、（2）少なくともいくつかの重要な側面を説明する可能性のある数学モデルの提案が含まれる。このデータセットの（3）数値予測を得るための数学方程式の正確なコンピューターソリューション、および（4）数値シミュレーションと実験データの比較によるモデルの品質の評価 (Duffus et al., 2007).

Toxicity Pathway

十分に摂動すると、健康への悪影響をもたらすと予想される細胞応答経路は、毒性経路と呼ばれる（NRC、2007）（図1内で定義）。 毒性経路は、分子開始事象から細胞効果への正常な生化学経路の摂動に関連している。 MOAとAOP概念の中心にあるが、先端効果に直接リンクしていない。

Weight of evidence (WoE)

WoEは、状況に応じてアプローチとツールが異なる仮説を裏付ける情報の範囲と質の、包括的で統合された、しばしば定性的な判断である（Weed、2005; WHO-UNEP、2012）。AOPの場合、疫学研究（Hill、1965）で因果関係を評価するためにBradford Hill（B / H）によって提案されたものから修正された特定の考慮事項のサブセットに基づいて、WoEが対処される。関連する各考慮事項の質問と補助データの性質の定義は、AOPに関するOECDガイダンスのユーザーハン度ブックの補遺 1 に含まれる (OECD, 2016a)。

参考文献

Duffus, J.H., Nordberg, M., Templeton, D.M. (2007). Glossary of terms used in toxicology, 2nd edition. Pure Appl. Chem. 79: 1153-1344.

Hill, A.B. (1965). The environment and disease: An association of causation. Proc. R. Soc. Med. 58: 295-300.

Jain, K.K. (2010). The Handbook of Biomarkers. London, Humana Press.

Keller, D.A., Juberg, D.R., Catlin, N., Farland, W.H., Hess, F.G., Wolf, D.C., Doerrer, N.G. (2012). Identification and Characterization of Adverse Effects in 21st Century Toxicology. Tox. Sci. 126: 291-297.

Krewski, D., Westphal, M., Al-Zoughool, M., Croteau, M.C., Andersen, M.E. (2011). New directions in toxicity testing. Annu. Rev. Public Health 32: 161-178.

North American Free Trade Agreement (NAFTA), Technical Working Group on Pesticides (TWG). (2011). (Quantitative) Structure Activity Relationship ((Q)SAR) Guidance Document.

NRC (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: The National Academies Press.

OECD (2007). Guidance on grouping of chemicals ENV/JM/MONO(2007)28.

OECD (2008). Report of the Second Survey on Available Omics Tools. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 100. ENV/JM/MONO(2008)35.

OECD (2011). Report of the Workshop on Using Mechanistic Information in Forming Chemical Categories. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 138. ENV/JM/MONO(2011)8.

OECD (2016a). Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. OECD Series on Adverse Outcome Pathways, No. 1, OECD Publishing, Paris.

OECD (2016b). Guidance Document on the Reporting of Defined Approaches to be used within Integrated Approaches to Testing and Assessment. ENV/JM/MONO(2016)28.

Schultz, T.W. (2010). Adverse outcome pathways: A way of linking chemical structure to in vivo toxicological hazards. In: Cronin, M.T.D. and Madden, J.C. eds., *In Silico* Toxicology: Principles and Applications, The Royal Society of Chemistry, Cambridge, UK, pp. 346-371.

Villeneuve, D.L. and Garcia-Reyero, N. (2011). A vision and strategy for predictive ecotoxicology testing in the 21st century. Environ. Toxicol. Chem. 30: 1-8.

Weed D.L. (2005). Weight of evidence: a review of concept and methods. Risk. Anal. 25: 1545-1557.

World Health Organization-United Nations Environment Programme (WHO-UNEP) (2012). Bergman, A., Heindel, J.J., Jobling, S., Kidd, K.A., Zoeller, R.T. (Eds.), State of the Science of Endocrine

Disrupting Chemicals 2012. Summary for Decision-Makers.
http://www.who.int/ceh/publications/endocrine/en/index.html Accessed January 9, 2017.

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

18 June 2019

OECD Guidelines for the Testing of Chemicals

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Key–Event-Based Test Guideline For In Chemico Skin Sensitisation Assays Addressing The Adverse Outcome Pathway Key Event On Covalent Binding To Proteins

GENERAL 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 signaling 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 Guideline comprises test methods to be used for supporting the discrimination between skin sensitisers and non-sensitisers in accordance with the UN GHS (1). The test methods currently described in this Test Guideline are:

- The Direct Peptide Reactivity Assay (DPRA) (Appendix I), and
- The Amino acid Derivative Reactivity Assay (ADRA) (Appendix II).

5. These two 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).

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 (17) (18) (19). Nevertheless, since protein reactivity represents only one key event of the skin sensitisation AOP (2) (20), 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 support the discrimination between skin sensitisers (i.e. UN GHS Category 1) and non-sensitisers when 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 modeling and read-across from chemical analogues (20). 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 (20) and can be employed as useful elements within IATA.

8. The test methods described in this Test Guideline do not allow either sub-categorisation of skin sensitisers into subcategories 1A and 1B (21), as defined by UN GHS (1) for authorities implementing these two optional subcategories, or potency prediction for safety assessment decisions. However, 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.

9. 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 (22).

Literature for introduction

(1) United Nations (UN) (2017), Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Seventh revised edition, New York and Geneva, United Nations Publications. Available at: [https://www.unece.org/trans/danger/publi/ghs/ghs_rev07/07files_e0.html]

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

(10) 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.

(11) OECD (1992), OECD Guidelines for the Testing of Chemicals No. 406. Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

(12) OECD (2010), OECD Guidelines for Chemical Testing No. 429. Skin sensitisation: Local Lymph Node assay. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

(13) OECD (2010), OECD Guidelines for Chemical Testing No. 442A.Skin sensitisation: Local Lymph Node assay: DA. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

(14) OECD (2018), OECD Guidelines for Chemical Testing No. 442B. Skin sensitisation: Local Lymph Node assay: BrdU-ELISA or –FCM. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

(15) OECD (2018), OECD Key Event based test Guideline 442D: In vitro Skin Sensitisation Assays Addressing AOP Key Event on Keratinocyte Activation. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

(16) OECD (2018), OECD Key event based test Guideline 442E: In Vitro Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [http://www.oecd.org/env/testguidelines].

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

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

(19) 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).

(20) 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].

(21) 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).

(22) OECD (2019), 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 - 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).

(Formula shown below.)

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).

Calculation

Calculating depletion of either NAC or NAL

Depletion is calculated as follows:

Percent depletion of either NAC or NAL = {1- (NAC or NAL peak area in replicate injection ÷ mean NAC or NAL peak area in reference control C)} × 100

Calculating predictive capacity

There are several terms that are commonly used along with the description of sensitivity, specificity and accuracy. They are true positive (TP), true negative (TN), false negative (FN), and false positive (FP).

Sensitivity, specificity and accuracy are described in terms of TP, TN, FN, and FP.

Sensitivity: Number of true positives ÷ Number of all positive chemicals, TP ÷ (TP + FN)

Specificity: Number of true negatives ÷ Number of all negative chemicals, TN ÷ (TN + FP)

Accuracy: Number of correct predictions ÷ Number of all predictions, (TN + TP) ÷ (TN+TP+FN+FP)

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

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) (Formula shown below.)

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) (Formula shown below.)

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 a reference standard prior to running the analytical batch (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) (2013), Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Fifth revised edition, UN New York and Geneva, 2013. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html

(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) (2001), Guidance for Industry: Bioanalytical Method Validation 22pp. Accessible at:www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidance/ucm070 107.pdf - 138 (23)

APPENDIX I

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

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

10. 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).

11. 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 Cat. 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.

12. 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. 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). 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

[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.

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.

13. 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).

14. 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

15. 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.

16. Prior to routine use of the method described in this Appenix, laboratories should demonstrate technical proficiency, using the ten proficiency substances listed in Annex 1.

PROCEDURE

17. This test method is based on the DPRA DB-ALM protocol nº 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 1).

Preparation of the cysteine or lysine-containing peptides

18. 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

19. 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 2). 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

20. 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 2). 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

21. 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

22. 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

23. 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 1). 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.

24. 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).

25. 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 2.

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

DATA AND REPORTING

Data evaluation

26. 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.

27. 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 2) according to the formula described below.

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

Acceptance criteria

28. 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.

29. 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

30. 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] A DPRA prediction should be considered in the framework of an IATA and in accordance with the provisions of paragraphs 2 and 4.

31. 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 1). 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 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.

32. 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.

33. 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. borderline results), additional testing may be necessary. If situations where the mean percent depletion falls in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or the cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10 prediction model, a second run may be considered, as well as a third one in case of discordant results between the first two runs.

Test report

34. The test report should include the following information

Test chemical

- Mono-constituent substance
    - Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
    - Physical 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.

Controls

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

    - Physical 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;
    - Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.
- Solvent/vehicle
    - Solvent/vehicle used and ratio of its constituents, if applicable;
    - Chemical identification(s), such as IUPAC or CAS name(s), CAS number(s), and/or other identifiers;
    - Purity, chemical identity of impurities as appropriate and practically feasible, etc;
    - Physical appearance, molecular weight, and additional relevant physicochemical properties in the case other solvents/vehicles than those mentioned in the test method are used and to the extent available;
    - Storage conditions and stability to the extent available;
    - Justification for choice of solvent for each test chemical;
    - For acetonitrile, results of test of impact on peptide stability.

Preparation of peptides, positive control and test chemical

- Characterisation of peptide solutions (supplier, lot, exact weight of peptide, volume added for the stock solution);
- Characterisation of positive control solution (exact weight of positive control substance, volume added for the test solution);
- Characterisation of test chemical solutions (exact weight of test chemical, volume added for the test solution).

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

- If applicable, the procedure used to demonstrate proficiency of the laboratory in performing the test method (e.g. by testing of proficiency substances) or to demonstrate reproducible performance of the test method over time.

Discussion of the results

- Discussion of the results obtained with the DPRA test method;

- Discussion of the test method results in the context of an IATA if other relevant information is available.

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

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 2

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

35. 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 cysteine and lysine derivatives are then used to support the discrimination between skin sensitisers and non-sensitisers (1) (2) (3).

36. The ADRA proved to be transferable to laboratories experienced in high-performance liquid chromatography (HPLC) analysis. ADRA's WLR was 100% (10/10), 100% (7/7), 90% (9/10), and 100% (10/10) in four participating laboratories. BLR for 40 test chemicals calculated based the results from three participating laboratories was 91.9% (4). For the 40 chemicals tested in the validation study in four laboratories, the cumulative accuracy was 86.9% (139/160), sensitivity was 81.5% (88/108), and specificity was 98.1% (51/52) (4) (5). Results from the validation study (4) (5) as well as from other published studies (3) indicate that ADRA identified sensitisers and non-sensitisers with an accuracy of 79% (98/124) (124 compounds that fall within ADRA's applicability domain), a sensitivity of 74% (65/88), and a specificity of 92% (33/36) relative to LLNA results (6). In addition, the prediction of human skin sensitisation for 73 compounds that fall within ADRA's applicability domain has an accuracy of 86% (63/73), a sensitivity of 85% (44/52), and a specificity of 90% (19/21) (6). 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 test as well as other animal tests may not fully reflect the situation in the species of interest, which is 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 study was considered to demonstrate that this method should be acceptable as part of an integrated testing strategy for the predictive identification of skin sensitisation hazard (7).

37. 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 ADRA 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. 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). 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 paragraphs 27 and 28); it may be possible to detect and quantify any NAC dimer formed by HPLC, 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).

38. The ADRA test method allows testing of poorly soluble chemicals. To be tested, a test chemical should be soluble in an appropriate solvent at a final concentration of 1 mM (see paragraph 15). 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.

39. In general, many organic compounds absorb UV in the range of 220 nm. In the case of co-elution of the nucleophilic reagent and the test chemical, this might result in false negative prediction. This may happen with the DPRA which specifies that quantification of the peptide-based nucleophilic reagents has to be performed at 220 nm. In contrast to this, the nucleophilic reagents used in ADRA are quantified at 281 nm. 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 (8).

40. 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 (UVCB substances) due to the need for defined molar ratio of test chemical and nucleophilic reagents. Limited information is currently available on the applicability of the ADRA to mixtures (9) (10). A new protocol has to be developed for multi-constituent substances and mixtures to be used with test methods like ADRA, which utilise HPLC analysis to quantify the depletion of nucleophilic reagents (9) (10). Thus, although it is impossible to define fixed methods in this guideline, which can evaluate multi-constituent substances and mixtures, paragraph 16 describes an evaluation method that is considered to be applicable at the present time for multi-constituent substances or mixtures of known composition (9). Such substances were however not tested during the validation studies. 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.

41. 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.

PRINCIPLE OF THE TEST

42. ADRA is an *in chemico* test method that quantifies residual concentrations 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, 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. The relative concentrations of NAC and NAL are measured by high-performance liquid chromatography (HPLC) with gradient elution and UV detection at 281 nm. Percent depletion values are then calculated for both NAC and NAL and compared to a prediction model (see paragraph 26).

43. Prior to routine use of the method described in this test method, laboratories should demonstrate technical proficiency, using the ten proficiency substances listed in Annex 1 of this Appendix.

PROCEDURE

44. This test method is based on the protocol (11) 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 Annex 1 of this Appendix.

Quality of NAC and NAL

45. The Nucleophilic Reagents can be obtained as an ADRA Kit for Skin Sensitisation Test, from FUJIFILM Wako (FFWK) Pure Chemical Corporation, Catalog No. 296-80901. Manufacturing NAC/NAL is patented in Japan only, by Fujifilm Corporation. Therefore, manufacturers in other countries can produce NAC/NAL without permission. In case other NAC/NAL are used, these should satisfy three quality criteria described below. Quality checks can be obviated and ADRA testing 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. Residual levels of NAC and NAL are to be a minimum of 90% in either case (11). The residual level of NAC is calculated as a percentage of the sum of NAC and the residual level of NAC dimers.

3) Reactivity: NAC and NAL are to be evaluated for reactivity with the ten proficiency substances given in Annex 1 and should satisfy the requirement given therein.

Preparation of the NAC and NAL stock solution

46. 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 solution is 5 μM in pH 8.0 phosphate buffer, and the final concentration of the NAL solution is 5 μM in pH 10.2 phosphate buffer.

Preparation of the test chemical solution

47. 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 (11). An appropriate solvent should dissolve the test chemical completely. Since the ADRA protocol stipulates that the test chemical be incubated in an excess volume of both NAC and NAL, 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. 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. It is important to note that DMSO may lead to dimerisation of the nucleophilic reagent NAC (12) and as a result, it may be more difficult to meet the acceptance criteria. 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). When using a DMSO-acetonitrile solvent, the test chemical should be dissolved in DMSO, and then this solution should be diluted 20-fold with acetonitrile to prepare a 1 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. The test chemical should be pre-weighed into a disposable polypropylene tube and dissolved immediately before testing in an appropriate solvent to prepare a 1 mM solution.

48. Mono-constituent substances of unknown molecular weight may be tested in a test chemical solution at a concentration of 0.5 mg/mL rather than 1 mM (9). Polymers which are well characterised should also be tested at a concentration of 1 mM based on the mean number average molecular weight, in a manner analogous to the procedure for mono-constituent compounds.

49. Mixtures and multi constituent substances, of known composition are to be tested as follows:

> 1) Liquids: Generally, tested as an undiluted mixture. In cases where low solubility of the test item prevents formation of reaction solution, i.e. undissolved material, clouding, and/or precipitation is observed, a positive result may still be used in the assessment, whereas a negative result is uncertain and should be interpreted with due care. Insofar as results could be false positives, however, predictions should be interpreted with due care.
>
> 2) Solids: The test chemical should be dissolved to maximum soluble concentration in the same solvent used to prepare the 1 mM test chemical solution. The test

chemical solution of the highest concentration possible is then tested as an undiluted mixture. In cases where low solubility of the test item prevents formation of reaction solution, i.e. undissolved material, clouding, and/or precipitation is observed, a positive result may still be used in the assessment, whereas a negative result is uncertain and should be interpreted with due care. Insofar as results could be false positives, however, predictions should be interpreted with due care.

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

50. Phenylacetaldehyde (CAS 122-78-1, ≥90% purity) should be used as positive control (PC) at a concentration of 1 mM in acetonitrile. 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, so they can be 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 Annex 2). 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

51. Both the NAC and the NAL solutions should be incubated with the test chemical at 1:50 ratio in a 96-well microplate. 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 should be interpreted with due care (see also paragraph 4 regarding the testing of chemicals not soluble at concentrations as high as 1 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 as a fixing solution to stop the reaction (3).

HPLC preparation and analysis

52. 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).

53. 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. Serial dilution of the NAC and NAL stock solutions (5.0 μM) will be used to prepare six calibration solutions in concentrations from 5.0 to 0.156 μM as well as a blank of the dilution buffer. Suitable calibration curves should have an $R^2 > 0.990$.

54. 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). The appropriate column is installed in the HPLC system. The recommended HPLC set-up described in the validated protocol uses a column (Base particle: core-shell type silica gel, Particle size: 2.5~2.7 μm, colomn 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 Annex 1. 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 Annex 2.

55. There are some test chemicals that could potentially promote oxidation of NAC. The peak of the dimerised NAC may be monitored visually. Any apparent dimerisation should be noted, since overestimation of NAC depletion could result in false-positive predictions (See paragraphs 26 and 27).

DATA AND REPORTING

Data evaluation

56. The concentration of both NAC and NAL is photometrically determined at 281 nm 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.

57. 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 Annex 2) 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

58. The following criteria should be met:

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

b) the mean percent NAC and NAL depletion value of the three replicates for the positive control phenylacetaldehyde should be between 6% and 30% for NAC and between 75% and 100% for NAL, while the maximum standard deviation (SD) for the positive control replicates should be < 10% for both NAC and NAL depletion, and

c) the mean NAC and NAL concentration of both Reference Controls A and Reference Control 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.

59. 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.

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

Prediction model

60. 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 percent depletion should be used to support the discrimination between skin sensitisers and non-sensitiser in the framework of an IATA or a DA.

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 2 and 3.

61. Co-elution occurs when the test chemical (the substance or one or several of the constituents of a multi-constituent substance or a mixture) absorbs significantly at 281 nm and has the same retention time as NAC or NAL. 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 Annex 1. 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.

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 and in accordance with the provisions of paragraphs 2 and 3.

62. When a result is unequivocal, a single HPLC analysis for both NAC and NAL should be sufficient for a test chemical. Additional testing is sometimes necessary, however, when the results lie close to the threshold value used to discriminate between positive and negative results (borderline results). If the mean percent depletion falls between 3.0% and 10.0% when using the NAC/NAL prediction model or the NAC percent depletion falls between 4.0% and 11.0% when using the NAC-only prediction model, a second run is advisable, as is a third run in the event of discordant results between the first two runs.

Test report

63. The test report should include the following information:

Test chemical

- Mono-constituent substance
    - Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers
    - Physical 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.

Controls

- Positive control
    - Chemical identification, such as IUPAC or CAS name(s), CAS number(s), SMILES or InChI code, structural formula, and/or other identifiers;
    - Physical appearance, water solubility, molecular weight, and additional relevant physicochemical properties, to the extent available;
    - Purity, chemical identity of impurities, as appropriate or feasible
    - Treatment prior to testing, if applicable (warming, grinding)
    - Concentration(s) tested
    - Storage conditions and stability, to the extent available;
    - Reference to historical positive control results demonstrating suitable run acceptance criteria, if applicable.

- Solvent
    - Solvent used and ratio of its constituents, if applicable
    - Chemical identification(s), such as IUPAC or CAS name(s), CAS number(s), and/or other identifiers
    - Purity, chemical identity of impurities, as appropriate and feasible
    - Physical appearance, molecular weight, and additional relevant physicochemical properties when solvents other than those mentioned in the test method are used
    - Storage conditions and stability, to the extent available
    - 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, 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 281 nm 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 281 nm of each replicate of Reference Controls B and C
    - Mean NAC and NAL peak area at 281 nm 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 281 nm 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 281 nm 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 281 nm 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

- If applicable, the procedure used to demonstrate proficiency of the laboratory in performing the test method (testing of proficiency substances, etc.) or to demonstrate reproducible performance of the test method over time.

Discussion of the results

- Discussion of the results obtained with the ADRA test method
- Discussion of the test method results in the context of an IATA if other relevant information is available

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, Jon R, Kleinstreuer N, Bae-Hwa K, Kojima H, Kasahara T, Ono A, 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, accepted on May 10, 2019

(5) OECD (2019), Draft 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) Yamamoto Y, Fujita M, Wanibuchi S, Sato A, Katsuoka Y, Ono A, Kasahara T. (2019), Availability of ADRA for prediction of skin sensitization by combining multiple alternative methods. Journal of Toxicological Sciences, accepted on May 9, 2019

(7) 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.

(8) Fujira M, Yamamoto Y, Wanibuchi S, Katsuoka Y, 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

(9) YamamotoY, Fujita M, Manibuchi S, Katsuoka Y, Ono A, 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

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

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

(12) James P. Tam JP, Cui Rong Wu CR, Wen Liu W, Jing Wen Zhang JW. (1991), Disulfide bond formation in peptides by dimethyl sulfoxide. Scope and applications. Journal of the American Chemical Society, 113, 6657–6662. DOI: 10.1021/ja00017a044

(13) Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F, Kern P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitisation undergoing prevalidation. *Journal of Applied Toxicology*, **33**(11):1337-52, DOI:10.1002/jat.2868.

(14) Gerberick GF, Vassalo JD, Foertsch LM, Price BB, Chaney JG, 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.

(15) 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.

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

APPENDIX II, ANNEX 1

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. 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

No.	Test chemicals	CAS No.	Physical state	Molecular weight	*In vivo* Prediction[1]	ADRA prediction[2]	Range of % depletion	
							NAC[3]	NAL[3]
1	*p*-Benzoquinone	106-51-4	Solid	108.09	Sensitiser (extreme)	Positive	90-100	40-70
2	Chloramine T trihydrate	7080-50-4	Solid	281.69	Sensitiser (strong)	Positive	90-100	90-100
3	Trans-Cinnamaldehyde	14371-10-9	Liquid	132.16	Sensitiser (moderate)	Positive	40-100	≤20
4	Palmitoyl Chloride	112-67-4	Liquid	274.87	Sensitiser (moderate)	Positive	≤10	50-100
5	Imidazolidinyl urea	39236-46-9	Solid	388.29	Sensitiser (weak)	Positive	10-45	≤10
6	Farnesal	19317-11-4	Liquid	220.35	Sensitiser (weak)	Positive	20-40	≤15
7	Glycerol	56-81-5	Liquid	92.09	Non-sensitiser	Negative	≤7	≤7
8	Benzyl alcohol	100-51-6	Liquid	108.14	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	110.11	Non-sensitiser	Negative	≤7	≤7

[1]The *in vivo* hazard (and potency) predictions are based on LLNA data. (13) (14) (15). The *in vivo* potency is derived using the criteria proposed by ECETOC (16).

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

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

APPENDIX II, ANNEX 2

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 Phenylacetaldehyde, rep 1 Sample 1, rep 1 Sample 2, rep 1
Second set of replicates	Reference control C, rep 2 Phenylacetaldehyde, rep 2 Sample 1, rep 2 Sample 2, rep 2
Third set of replicates	Reference control C, rep 3 Phenylacetaldehyde, 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.

Section 4
Health effects

Test Guideline No. 495

Reactive Oxygen Species (ROS) Assay for Photoreactivity

18 June 2019

OECD Guidelines for the Testing of Chemicals

OECD GUIDELINE FOR TESTING OF CHEMICALS

Ros (Reactive Oxygen Species) Assay For Photoreactivity

INTRODUCTION

1. Phototoxicity is defined as a toxic response is elicited by topically or systemically administered photoreactive chemicals after the exposure of the body to environmental light. Several classes of photoreactive chemicals could cause phototoxic reactions when activated by light at otherwise non-toxic doses. Phototoxicity can be categorized as photoirritation, photoallergy, and photogenotoxicity (1). Photoirritation is characterized as an acute light-induced skin response to a photoreactive chemical. Photoallergy is an immune-mediated reaction in which light may cause a structural change in a drug so that it acts as a hapten, possibly by binding to proteins in the skin (2). Photogenotoxicity is a genotoxic response after exposure to a chemical by two mechanisms: either directly by photoexcitation of DNA or indirectly by excitation of photoreactive chemicals.

2. In 2002, regulatory agencies in the US (US Food and Drug Administration, FDA) and EU (European Medicines Agency, EMA) published guidelines for photosafety assessments of drug candidates (3)(4). In 2004, the Organisation for Economic Co-operation and Development (OECD) adopted Test Guideline 432: *In vitro* 3T3 Neutral Red Uptake (NRU) Phototoxicity Test as a validated methodology for evaluating the phototoxic potential of chemicals (5). The EMA also published a concept paper in 2008 (6), which proposes a testing strategy that merges the testing proposals recommended by FDA and EMA. Considering these documents, the International Council of Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) published ICH S10 guideline, "Photosafety Evaluation of Pharmaceuticals" in 2014 (7).

3. According to above referenced guidelines, chemicals or drug candidates need to be examined for their phototoxic potential. Since light must be absorbed by a compound in order for photochemical reactions to take place (8), the phototoxic potential of chemicals is related to the photochemical properties of compounds, especially light absorption properties within 290–700 nm. The guidelines suggested the need for measurement of the light absorption properties of chemicals as a first round of screening (3)(4). The ICH S10 guideline recommends UV-visible light absorption spectral analysis as a criterion for evaluating the phototoxic potentials of drugs (7); however, UV-visible light absorption of chemicals would not always correlate directly with their phototoxic potential, so a

combination of UV data (molar extinction coefficient, MEC) with other appropriate screening systems might be advantageous in avoiding false predictions.

4. In addition to light absorption and distribution to light-exposed tissue, the generation of a reactive species from chemicals following absorption of UV-visible light is described as a key determinant of chemicals for causing direct phototoxic reactions in an older guidance document (7)(9). Thus, the Reactive Oxygen Species (ROS) assay (10)(11) has been also included by the ICH S10 guideline as an optional initial *in chemico* screening tool for evaluating the photoreactivity of pharmaceuticals (7).

5. As an alternative method for *in vivo* phototoxicity testing, the OECD TG432 (5) describes an *in vitro* 3T3 NRU Phototoxicity Test and sets specific criteria for evaluating phototoxic hazard. The 3T3 NRU Phototoxicity Test evaluates photo-cytotoxicity by the relative reduction in viability of cells exposed to the chemical in the presence versus absence of light. Chemicals identified by this test are likely to be photoreactive, following systemic application and distribution to the skin, or after topical application. Although most of the photoirritant chemicals were correctly identified by the 3T3 NRU Phototoxicity Test, it provided false predictions for almost half of the chemicals in the photoallergens group. However, the 3T3 NRU Phototoxicity Test was not originally designed for specific prediction of chemical photoallergenicity (2). The 3T3 NRU Phototoxicity Test seems to be less reliable for photoallergenicity prediction. The photochemical assays such as ROS assay and UV/VIS spectral analysis can predict photoallergenic potential of tested chemicals, although there is still a substantial risk of false positive predictions (12).

6. Definitions used are provided in Annex A.

INITIAL CONSIDERATION AND LIMITATIONS

7. Before photosafety assessments are considered, a UV-visible light absorption spectrum of the test chemical should be determined according to OECD Test Guideline 101(13). Based on an analysis of data, the ICH S10 guideline has suggested that no further photosafety testing is needed if the MEC of a chemical is less than 1,000 $L \cdot mol^{-1} \cdot cm^{-1}$ (7). Few phototoxic chemicals showed a MEC less than 1,000 $L \cdot mol^{-1} \cdot cm^{-1}$ and these chemicals may not need to be tested in the ROS assay or any other photosafety assessments (9) (14) (15). Data collected for the limits of photoreactivity are discussed in Henry *et al.* (16) and Bauer *et al.* (17). It should be noted that phototoxicity by indirect mechanisms (e.g., pseudoporphyria or porphyria), although rare, could still occur. For compounds with MEC values of 1000 $L \cdot mol^{-1} \cdot cm^{-1}$ or higher, if the drug developer chooses to conduct a test for photoreactivity a negative result could support a decision that no further photosafety assessment is warranted.

8. The reliability and relevance of the ROS assay was recently evaluated in a multi-laboratory validation study using two different solar simulators (18)(19)(20)(21). In both solar simulators, the intra- and inter-day precisions for quinine, a positive control, were found to be above 90%, and the data suggested high inter-laboratory reproducibility (19). In a multi-laboratory validation study, the ROS assay on 2 standards and 42 coded chemicals, including 23 phototoxins and 19 non-phototoxic drugs/chemicals, provided no false negative predictions upon defined criteria as compared with the *in vitro/in vivo* phototoxicity. The, sensitivity, individual specificity, positive and negative predictivities of the ROS assay on the 42 tested chemical were calculated to be 100%, 42–82%, 75–92% and 100%, respectively. The ROS assay was designed for qualitative photoreactivity assessment of chemicals, the principle of which is

monitoring of type I (an electron or hydrogen transfer, resulting in the formation of free radical species) and type II (an energy transfer from excited triplet photosensitizer to the oxygen) photochemical reactions in test chemicals exposed to simulated sunlight (10), possibly leading to photodegradation and various phototoxic reactions, including photoirritation, photoallergy, and photogenotoxicity. Further, this assay has been optimised for detecting positive test chemicals. Test chemicals found to be negative in the ROS assay are likely to be negative in *in vivo* test systems; however, additional data may be required to determine if chemicals that are photoreactive in the ROS assay are likely to be positive *in vivo*. The test has not been designed to address indirect mechanisms of phototoxicity, such as effects of metabolites of a test chemical.

9. The applicability domain of the ROS assay is currently restricted to only those chemicals that meet the solubility criteria outlined in the protocol (see paragraph 22). Insoluble chemicals in the reaction mixtures are not suitable for testing with the ROS assay using this protocol (DMSO or NaPB solvent) but might be tested in the ROS assay with addition of solubility enhancers in the reaction mixtures (22)(23)(24). However, further characterization and standardization of procedures using these alternative vehicles should be performed by testing proficiency chemicals before incorporation into routine use. In the ROS assay, superoxide anion (SA) can be measured upon the reduction of nitroblue tetrazolium, and the determination of singlet oxygen (SO) can be made on the basis of bleaching of *p*-nitrosodimethylaniline by oxidized imidazole (11). Test chemicals that interfere with these reactions are sometimes best considered outside of the applicability domain of the ROS assay. For example, ascorbic acid and other reducing chemicals reduce the tetrazolium salt to formazan directly (25). Some skin-lightening cosmetics may also have potent reducing properties that interfere with ROS determinations. Ascorbic acid also accelerates the oxidation of imidazole derivatives (26), providing false positive prediction in the ROS assay.

10. 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 ROS assay to the testing of mono-constituent chemicals, multi-constituent chemicals and/or mixtures. Based on the data currently available, the ROS assay was shown to be applicable to test chemicals covering a variety of organic functional groups, reaction mechanisms, phototoxic potency (as determined in *in vivo* studies) and physicochemical properties. Limited information is currently available on the applicability of the ROS assay to multi-constituent chemicals/mixtures (27). When considering testing of mixtures, difficult to test chemical (e.g. unstable) or 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.

PRINCIPLE OF THE TEST

11. Chemical phototoxicity can be caused by topical and systemic application of chemicals in combination with exposure to environmental light. There are several classes of chemicals that are nontoxic by themselves but could become reactive in the skin or eyes when exposed to environmental light and thereby result in toxicity. The primary event in any phototoxic reaction is the absorption of photons of a wavelength that induces excitation of the chromophore. The excitation energy is often transferred to oxygen molecules, followed by generation of ROS, including SA through type I photochemical reactions and SO through type II photochemical reactions by photo-excited molecules. These appear to be the principal intermediate species in many phototoxic responses. Direct reaction of excited chromophores with cellular constituents may also lead to phototoxicity. Therefore, while the ROS assay may not detect all ultimate mechanisms of phototoxicity, the

determination of ROS generation from chemicals irradiated with simulated sunlight is indicative of phototoxic potential.

12. In the ROS assay, SO generation is detected by spectrophotometric measurement of *p*-nitrosodimethylaniline (RNO) bleaching, followed by decreased absorbance of RNO at 440 nm (28). Although SO does not react chemically with RNO, the RNO bleaching is a consequence of SO capture by the imidazole ring, which results in the formation of a trans-annular peroxide intermediate capable of inducing the bleaching of RNO, as follows:

SO + Imidazole → [Peroxide intermediate] → Oxidized imidazole

[Peroxide intermediate] + RNO → RNO + Products

13. SA generation is detected by observing the reduction of nitroblue tetrazolium (NBT). As indicated below; NBT can be reduced by SA via a one-electron transfer reaction, yielding partially reduced (2 e^-) monoformazan (NBT^+) as a stable intermediate (29). Thus, SA can reduce NBT to NBT^+, the formation of which can be monitored spectrophotometrically at 560 nm.

$SA + NBT \rightarrow O_2 + NBT^+$

Demonstration of Proficiency

14. Prior to routine use of the test method described in this Test Guideline, laboratories should demonstrate technical proficiency, using the proficiency chemicals listed and described in Annex C. The 9 proficiency chemicals (Nos. 1–9) for the two recommended solar simulators (Suntest CPS+ or CPS and SXL-2500V2) or the 17 proficiency chemicals (Nos. 1–17) for a solar simulator other than the two recommended models are to be tested to ensure that measured values of SO and SA on all proficiency chemicals are within the range described in Annex C.

PROCEDURE

Solar simulator

15. Typically calibrated solar simulators are used because photoreactivity in the presence of natural sunlight is of concern, due to the spectral differences of global positioning and the time of day. For other circumstances where photoreactivity in response to artificial light is of interest, other sources of light may be considered. An appropriate solar simulator is to be used for irradiation of UV and visible light. The irradiation power distribution is to be kept as close to that of outdoor daylight as possible by using an appropriate filter to reduce UVC wavelengths. Recommended test conditions are as follows:

Solar simulator with filter to reduce UV wavelengths <290 nm (See Annex B)

- 1.8 to 2.2 mW/cm^2 (e.g. the indicator setting value of 250 W/m^2 for CPS+) for 1 hour,
- 6.5 to 7.9 J/cm^2 of UVA intensity (Annex B).

SXL-2500V2 (Seric) with UV filter (to reduce wavelengths <300 nm)

- 3.0 to 5.0 mW/cm^2 for 1 hour,
- 11 to 18 J/cm^2 of UVA intensity (Annex B).

16. The solar simulator is to be equipped with an appropriate temperature control or fan to stabilize the temperature during irradiation, because ROS production is affected by

temperature. Standard temperature for a solar simulator with temperature control is 25°C. The acceptable temperature range during irradiation is 20 to 29°C (20)(21).

Quartz reaction container

17. A quartz reaction container is used to avoid loss of UV due to passing through a plastic lid and vaporization of the reaction mixture (20)(21)(30). Specifications for the recommended container are provided in Annex D. If a different container is used, a lid or seal with high UV transmittance should be used. In this case, a feasibility study using the reference chemicals (Nos. 1–17) is to be conducted to determine an appropriate level of exposure to UV and visible light.

Reagents

18. All reagents should be used within 1 month after preparation and should be sonicated immediately prior to use (20)(21). Representative preparation methods are shown as follows:

20 mM sodium phosphate buffer (NaPB), pH 7.4

- Weigh 593 mg of $NaH_2PO_4 \cdot 2H_2O$ (CAS No. 13472-35-0) and 5.8 g of $Na_2HPO_4 \cdot 12H_2O$ (CAS No. 10039-32-4), add 900 mL of purified water, adjust with HCl to a pH of 7.4, dilute with purified water up to 1 L, and mix.
- Store in a refrigerator or at room temperature.

0.2 mM *p*-nitrosodimethylaniline (RNO, CAS No. 138-89-6)

- Dissolve 3 mg of RNO in 100 mL of 20 mM NaPB.
- Store in a refrigerator and protect from light.

0.2 mM imidazole (CAS No. 288-32-4)

- Dissolve 13.6 mg of imidazole in 10 mL of 20 mM NaPB.
- Dilute the 20 mM imidazole solution 100 times with 20 mM NaPB.
- Store in a refrigerator and protect from light.

0.4 mM nitroblue tetrazolium chloride (NBT, CAS No. 298-83-9)

- Dissolve 32.7 mg of NBT in 100 mL of 20 mM NaPB.
- Store in a refrigerator and protect from light.

Solvents

19. Use analytical grade DMSO at first. For chemicals that are not soluble in DMSO, 20 mM NaPB is to be used as a solvent. Some chemicals react with DMSO and test chemical stability in DMSO should be determined. If the test chemical is not soluble or stable in DMSO or NaPB, other solvents may be used. However the test chemical must be demonstrated to be stable in the selected solvent, and SO and SA ranges for proficiency chemicals must fall within the ranges defined in Annex C.

Test chemicals

20. Test chemicals must be prepared fresh, immediately prior to use unless data demonstrate their stability in storage. It is recommended that all chemical handling and the initial treatment of cells be performed under light conditions that would avoid photoactivation or degradation of the test chemical prior to irradiation. Chemicals should be tested at 200 µM (final concentration). A 20-µM concentration can be used if precipitation occurs before light exposure, coloration, or other interference is observed in the reaction mixture at 200 µM. A positive result at 20 µM can be used to indicate

photoreactivity; however, a negative result at the lower 20 µM concentration is not indicative of absence of photoreactivity. The molecular weight of the test chemical must be available.

21. The test chemical solutions are to be prepared immediately before use in a solvent as described in paragraph 19. Each test chemical is to be weighed in a tube, and solvent added to achieve a 10 mM concentration of the test chemical (20)(21). The tube is to be mixed with a vortex mixer and sonicated for 5 to 10 minutes. All preparations are to be protected from strong UV and intense visible light (e.g. direct overhead light, working near windows exposed to natural light) at all times during preparation. When precipitation before light exposure or other interference is observed in the reaction mixture at 200 µM, a 1-mM solution (20 µM as the final concentration) is to be prepared by dilution of the stock solution of chemicals at 10 mM using DMSO. For chemicals that are not soluble in DMSO, 20 µL of DMSO (2 v/v%) is to be contained in the reaction mixture.

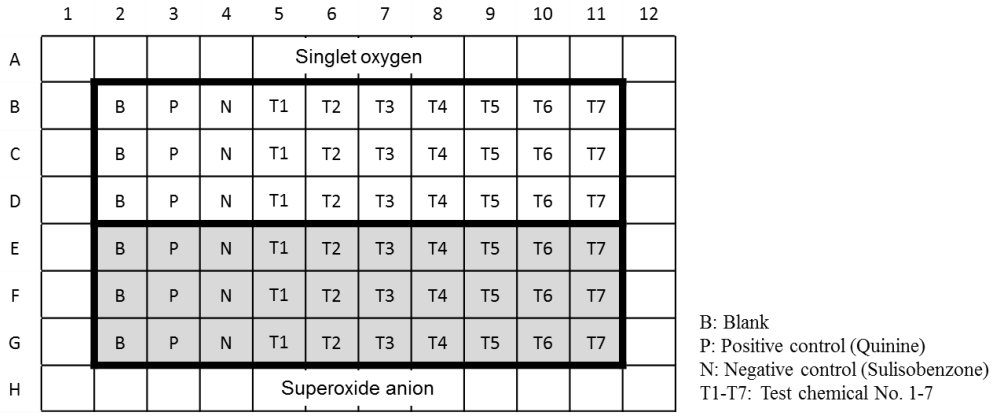
Positive and negative controls

22. Stock solutions of quinine hydrochloride (a positive control, CAS No. 6119-47-7) and sulisobenzone (a negative control, CAS No. 4065-45-6) are to be prepared at 10 mM each in DMSO (final concentration of 200 µM) according to the above procedure, divided into tubes, and stored in a freezer (generally below -20°C) for up to 1 month. The stock solution is to be thawed just before the experiment and used within the day.

Test procedure

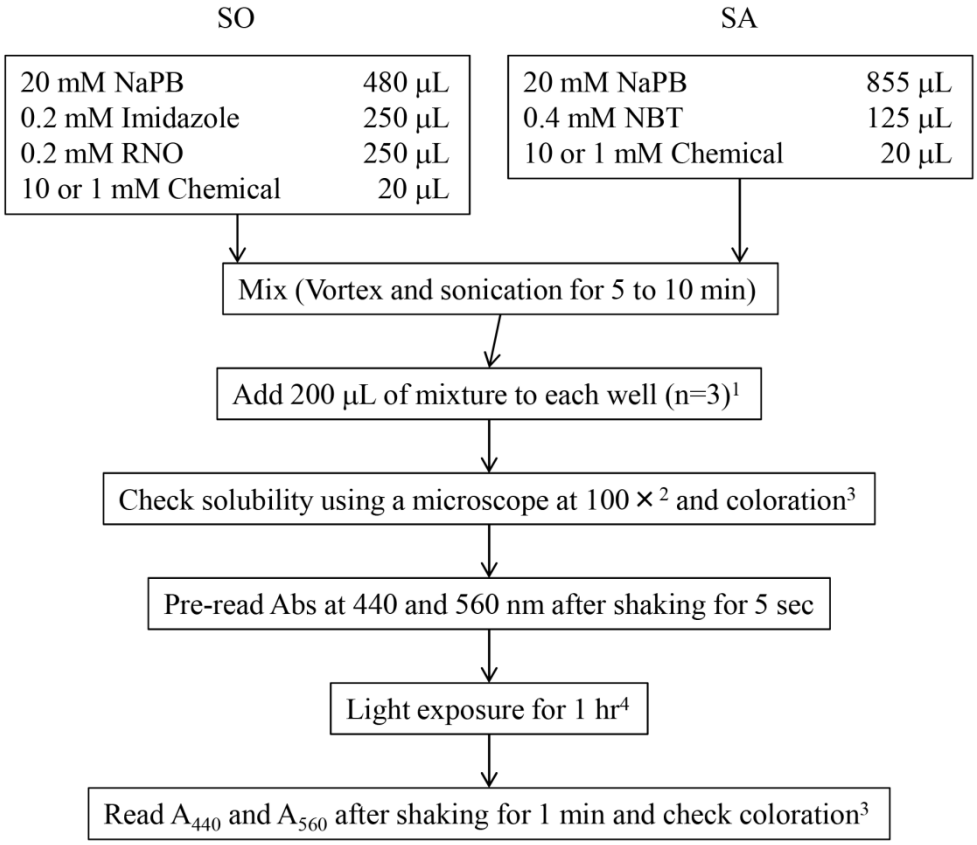
23. A typical 96-well plate configuration is as follows, but other configurations are also acceptable:

Figure 1. An example of a typical plate configuration



24. A tube (e.g. 1.5 mL micro tube) and a plastic clear flat bottomed 96-well microplate are to be used. The reaction mixture is to be prepared by vortex mixing and/or sonication under UV-cut illumination or shade. The same volume of DMSO, 20 µL, is to be added in a vehicle control instead of test chemical solution.

Figure 2. Workflow diagram if the stock solution of the test chemical is prepared in DMSO.



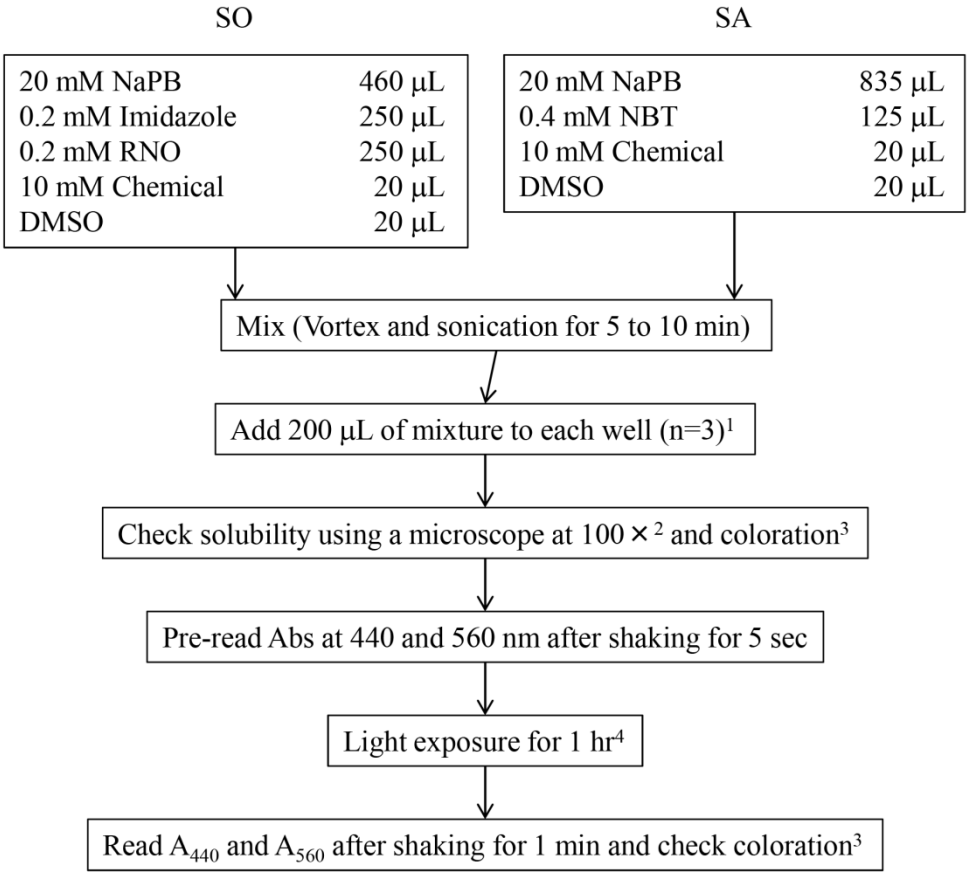
[1] Avoid using peripheral wells. More than one test chemical can be tested on a plate.

[2] Some chemicals might precipitate in the reaction mixture. It is therefore important to check solubility prior to irradiation. Solubility of each reaction mixture in its well is to be observed with a microscope prior to irradiation. Test chemical concentrations are to be selected so as to avoid precipitation or cloudy solutions.

[3] The reaction mixture is to be checked for coloration with the naked eye.

[4] The 96-well plate is to be placed in the quartz reaction container. A quartz cover is to be set on the plate and fastened with bolts. Ensure that temperature and other ambient conditions are stable when using the solar simulator. Measure UVA intensity and temperature at the plate position using a UVA detector and thermometer both before and after irradiation.

Figure 3. Workflow diagram if the stock solution of the test chemical is prepared in 20 mM NaPB.



[1] Avoid using peripheral wells. More than one test chemical can be tested on a plate.

[2] Some chemicals might precipitate in the reaction mixture. It is therefore important to check solubility prior to irradiation. Solubility of each reaction mixture in its well is to be observed with a microscope prior to irradiation. Test chemical concentrations are to be selected so as to avoid precipitation or cloudy solutions.

[3] The reaction mixture is to be checked for coloration with the naked eye.

[4] The 96-well plate is to be placed in the quartz reaction container. A quartz cover is to be set on the plate and fastened with bolts. Ensure that temperature and other ambient conditions are stable when using the solar simulator. Measure UVA intensity and temperature at the plate position using a UVA detector and thermometer both before and after irradiation.

DATA AND REPORTING

Data analysis

25. Data from three wells for each chemical concentration is used to calculate mean and standard deviation.

SO

Decrease of A_{440} ×1000 = [A_{440} (–) – A_{440} (+) – (a – b)] ×1000
A_{440} (–): Absorbance before light exposure at 440 nm
A_{440} (+): Absorbance after light exposure at 440 nm
a: Vehicle control before light exposure (mean)
b: Vehicle control after exposure (mean)

SA

Increase of A_{560} ×1000 = [A_{560} (+) – A_{560} (–) – (b – a)] ×1000
A_{560} (–): Absorbance before light exposure at 560 nm
A_{560} (+): Absorbance after light exposure at 560 nm
a: Vehicle control before light exposure (mean)
b: Vehicle control after exposure (mean)

Criteria for data acceptance

26. The following criteria are to be satisfied in each experiment.

- No precipitation of test chemical in the reaction mixture before light exposure.
- No color interference by test chemical in the reaction mixture before or after light exposure.
- No technical problems, including temperature range (20–29°C), when collecting data set.
- The ranges of raw A_{440} and A_{560} values: 0.02 to 1.5.
- Historical positive and negative control values are to be developed by each laboratory based on a mean +/-2 SD. The following range was defined based on the 95% confidence interval (mean +/- 1.96SD) obtained from the validation data. When a solar simulator other than a recommended model is used, establish modified criteria based on 95% confidence interval.
-

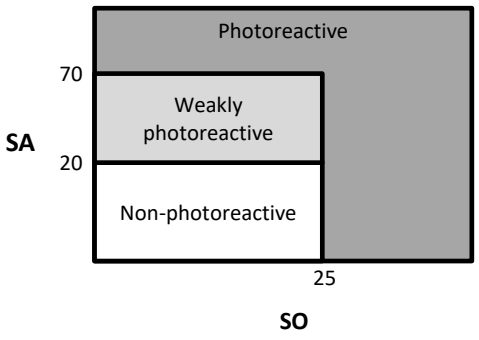
  Positive control (quinine hydrochloride) value at 200 µM (mean of 3 wells)
  SO: 319 to 583
  SA: 193 to 385
  Negative control (sulisobenzone) value at 200 µM (mean of 3 wells)
  SO: -9 to 11
  SA: -20 to 2
- Laboratories should demonstrate technical proficiency, as described in Annex C, prior to routine use of the test method described in this Test Guideline.

Criteria for judgment

27. Each test chemical is to be judged as follows:

ROS assay prediction model

Judgment[1, 2]	Concentration[3]	SO (mean of 3 wells)[6]		SA (mean of 3 wells)[6]
Photoreactive	200 μM	≥25	and	≥70
		<25 and/or I[4]	and	≥70
		≥25	and	<70 and/or I[4]
Weakly photoreactive	200 μM	<25	and	≥20, <70
Photoreactive	20 μM	≥25	and	≥20
Non-photoreactive	200 μM	<25	and	<20
Inconclusive	The results do not meet any of the above-mentioned criteria.[5]			



[1] A single experiment is sufficient for judging results, because the ROS assay shows good intra- and inter-laboratory reproducibility in the validation studies.

[2] If precipitation, coloration, or other interference is observed at both 20 and 200 μM, the chemical is considered incompatible with the ROS assay and judged as inconclusive.

[3] 20 μM can be used for judgment when precipitation or coloration is observed at 200 μM. A positive results at 20 μM can be used to indicate photoreactivity; however, a negative result at the lower 20 μM concentration is not indicative of absence of photoreactivity.

[4] Interference such as precipitation or coloration.

[5] Positive prediction can be made on the basis of SO only, SA only, or both; however, both SO and SA values should be obtained for reliable negative prediction.

[6] Classification criteria defined in published manuscripts. (11)(20)(21)

Data quality

28. Studies for regulatory purposes are to be conducted to the highest of quality standards, with data collection records readily available, in compliance with GLP regulations whenever possible, and all documents checked by the Quality Assurance Unit of the laboratory.

Test report

29. The test report should include the following information:

Test chemical:

- identification data, common generic names and IUPAC and CAS number, if known;
- physical nature and purity;
- physicochemical properties relevant to conduct of the study;
- UV/vis absorption spectrum;
- stability and photostability, if known.

Control chemicals:

- name, manufacturer, and lot No.;
- physical nature and purity;
- storage condition;
- preparation of control chemical solutions;
- final concentrations tested.

Solvent:

- name, manufacturer, and lot No.;
- justification for choice of solvent;
- solubility of the test chemical in solvent.

Irradiation condition:

- manufacturer and type of the solar simulator used;
- rationale for selection of the solar simulator used;
- UVA detector used;
- UVA irradiance, expressed in mW/cm^2
- UVA dose, expressed in J/cm^2;
- temperature before and after irradiation.

ROS assay procedure.

Acceptance and decision criteria.

Results.

Discussion.

Conclusions.

LITERATURE

(1) Moore, D. E. (2002). Drug-induced cutaneous photosensitivity: incidence, mechanism, prevention and management. *Drug Saf*, **25**, 345-72

(2) Onoue, S., Seto, Y., Sato, H., Nishida, H., Hirota, M., Ashikaga, T., Api, A. M., Basketter, D. and Tokura, Y. (2017). Chemical photoallergy: photobiochemical mechanisms, classification, and risk assessments. *J Dermatol Sci*, **85**, 4-11

(3) The European Agency for the Evaluation of Medicinal Products, Evaluation of Medicines for Human Use, Committee for Proprietary Medicinal Products (EMEA/CPMP) (2002). Note for Guidance on Photosafety Testing, CPMP/SWP/398/01.

(4) The Food and Drug Administration, Center for Drug Evaluation and Research (FDA/CDER) (2002). Guidance for Industry, Photosafety Testing.

(5) The Organisation for Economic Co-operation and Development (OECD) (2004). OECD guideline for testing of chemicals, 432, *In vitro* 3T3 NRU phototoxicity test.

(6) The European Agency for the Evaluation of Medicinal Products, Evaluation of Medicines for Human Use, Committee for Proprietary Medicinal Products (EMEA/CPMP) (2008). Concept Paper on the Need for Revision of the Note for Guidance on Photosafety testing, CPMP/SWP/398/01.

(7) International Council on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2014). ICH guideline S10 Guidance on photosafety evaluation of pharmaceuticals.

(8) *Radiation. Encyclopedia Britannica Online. Retrieved 2009-11-09.*

(9) The Organisation for Economic Co-operation and Development (OECD) (1997). Environmnental Health and Safety Publications, Series on Testing and Assessment No.7 "Guidance Document on Direct Phototransformation of Chemicals in Water" Environment Directorate.

(10) Onoue, S. and Tsuda, Y. (2006). Analytical studies on the prediction of photosensitive/phototoxic potential of pharmaceutical substances. *Pharm Res*, **23**, 156-64

(11) Onoue, S., Kawamura, K., Igarashi, N., Zhou, Y., Fujikawa, M., Yamada, H., Tsuda, Y., Seto, Y. and Yamada, S. (2008). Reactive oxygen species assay-based risk assessment of drug-induced phototoxicity: classification criteria and application to drug candidates. *J Pharm Biomed Anal*, **47**, 967-72

(12) Onoue, S., Ohtake, H., Suzuki, G., Seto, Y., Nishida, H., Hirota, M., Ashikaga, T. and Kouzuki, H. (2016). Comparative study on prediction performance of photosafety testing tools on photoallergens. *Toxicol In Vitro*, **33**, 147-52

(13) The Organisation for Economic Co-operation and Development (OECD) (1981). OECD guideline for testing of chemicals, 101, UV-VIS absorption spectra.

(14) Onoue, S., Suzuki, G., Kato, M., Hirota, M., Nishida, H., Kitagaki, M., Kouzuki, H. and Yamada, S. (2013). Non-animal photosafety assessment approaches for cosmetics based on the photochemical and photobiochemical properties. *Toxicol In Vitro*, **27**, 2316-24

(15) Lovell, W. W. (1993). A scheme for in vitro screening of substances for photoallergenic potential. *Toxicol In Vitro*, **7**, 95-102

(16) Henry, B., Foti, C. and Alsante, K. (2009). Can light absorption and photostability data be used to assess the photosafety risks in patients for a new drug molecule? *J Photochem Photobiol B*, **96**, 57-62

(17) Bauer, D., Averett, L. A., De Smedt, A., Kleinman, M. H., Muster, W., Pettersen, B. A. and Robles, C. (2014). Standardized UV-vis spectra as the foundation for a threshold-based, integrated photosafety evaluation. *Regul Toxicol Pharmacol*, **68**, 70-5

(18) Onoue, S., Hosoi, K., Wakuri, S., Iwase, Y., Yamamoto, T., Matsuoka, N., Nakamura, K., Toda, T., Takagi, H., Osaki, N., Matsumoto, Y., Kawakami, S., Seto, Y., Kato, M., Yamada, S., Ohno, Y. and Kojima, H. (2013). Establishment and intra-/inter-laboratory validation of a standard

protocol of reactive oxygen species assay for chemical photosafety evaluation. *J Appl Toxicol*, **33**, 1241-1250

(19) Onoue, S., Hosoi, K., Toda, T., Takagi, H., Osaki, N., Matsumoto, Y., Kawakami, S., Wakuri, S., Iwase, Y., Yamamoto, T., Nakamura, K., Ohno, Y. and Kojima, H. (2014). Intra-/inter-laboratory validation study on reactive oxygen species assay for chemical photosafety evaluation using two different solar simulators. *Toxicol In Vitro*, **28**, 515-23

(20) ROS assay validation management team (2013). Validation report for the international validation study on ROS (Reactive Oxygen Species) assay as a test evaluating phototoxic potential of chemicals (Atlas Suntest version), http://www.jacvam.jp/files/news/ROS_Assay_full%2020130920%20atlas_fourth%20data.pdf (accessed May, 2015).

(21) ROS assay validation management team (2013). Validation report for the international validation study on ROS (Reactive Oxygen Species) assay as a test evaluating phototoxic potential of chemicals (Seric version), http://www.jacvam.jp/files/news/ROS_Assay_full%2020131009Seric%20fourth.pdf (accessed May, 2015).

(22) Onoue, S., Yamauchi, Y., Kojima, T., Igarashi, N. and Tsuda, Y. (2008). Analytical studies on photochemical behavior of phototoxic substances; effect of detergent additives on singlet oxygen generation. *Pharm Res*, **25**, 861-8

(23) Seto, Y., Kato, M., Yamada, S. and Onoue, S. (2013). Development of micellar reactive oxygen species assay for photosafety evaluation of poorly water-soluble chemicals. *Toxicol In Vitro*, **27**, 1838-46

(24) Onoue, S., Kato, M. and Yamada, S. (2014). Development of an albuminous reactive oxygen species assay for photosafety evaluation under experimental biomimetic conditions. *J Appl Toxicol*, **34**, 158-65

(25) Bruggisser, R., von Daeniken, K., Jundt, G., Schaffner, W. and Tullberg-Reinert, H. (2002). Interference of plant extracts, phytoestrogens and antioxidants with the MTT tetrazolium assay. *Planta Med*, **68**, 445-8

(26) Imanaga, Y. (1955). Autooxidation of L-ascorbic acid and imidazole nucleus. *J Biochem*, **42**, 657-67

(27) Nishida, H., Hirota, M., Seto, Y., Suzuki, G., Kato, M., Kitagaki, M., Sugiyama, M., Kouzuki, H. and Onoue, S. (2015). Non-animal photosafety screening for complex cosmetic ingredients with photochemical and photobiochemical assessment tools. *Regul Toxicol Pharmacol*, **72**, 578-85

(28) Kraljic, I. and Mohsni, S. E. (1978). A new method for the detection of singlet oxygen in aqueous solutions. *Photochem Photobiol*, **28**, 577-81

(29) Pathak, M. A. and Joshi, P. C. (1984). Production of active oxygen species ($^1O_2$ and O_2^-.) by psoralens and ultraviolet radiation (320-400 nm). *Biochim Biophys Acta*, **798**, 115-26

(30) Onoue, S., Igarashi, N., Yamada, S. and Tsuda, Y. (2008). High-throughput reactive oxygen species (ROS) assay: an enabling technology for screening the phototoxic potential of pharmaceutical substances. *J Pharm Biomed Anal*, **46**, 187-93

Annex A. Definitions

3T3 NRU Phototoxicity Test: *In vitro* 3T3 neutral red uptake phototoxicity test.

Irradiance: The intensity of UV or visible light incident on a surface, measured in W/m^2 or mW/cm^2.

Dose of light: The quantity [= intensity × time (seconds)] of UV or visible light incident on a surface, expressed in J/m^2 or J/cm^2.

MEC: Molar Extinction Coefficient (also called molar absorptivity) is a constant for any given molecule under a specific set of conditions (e.g. solvent, temperature, and wavelength) and reflects the efficiency with which a molecule can absorb a photon (typically expressed as $L \cdot mol^{-1} \cdot cm^{-1}$).

Photoreactivity: The property of chemicals that react with another molecule as a consequence of absorption of photons.

Phototoxicity: Toxic responses that can be elicited after the exposure of skin to certain chemicals and subsequent exposure to light, or that is induced similarly by skin irradiation after systemic administration of a chemical.

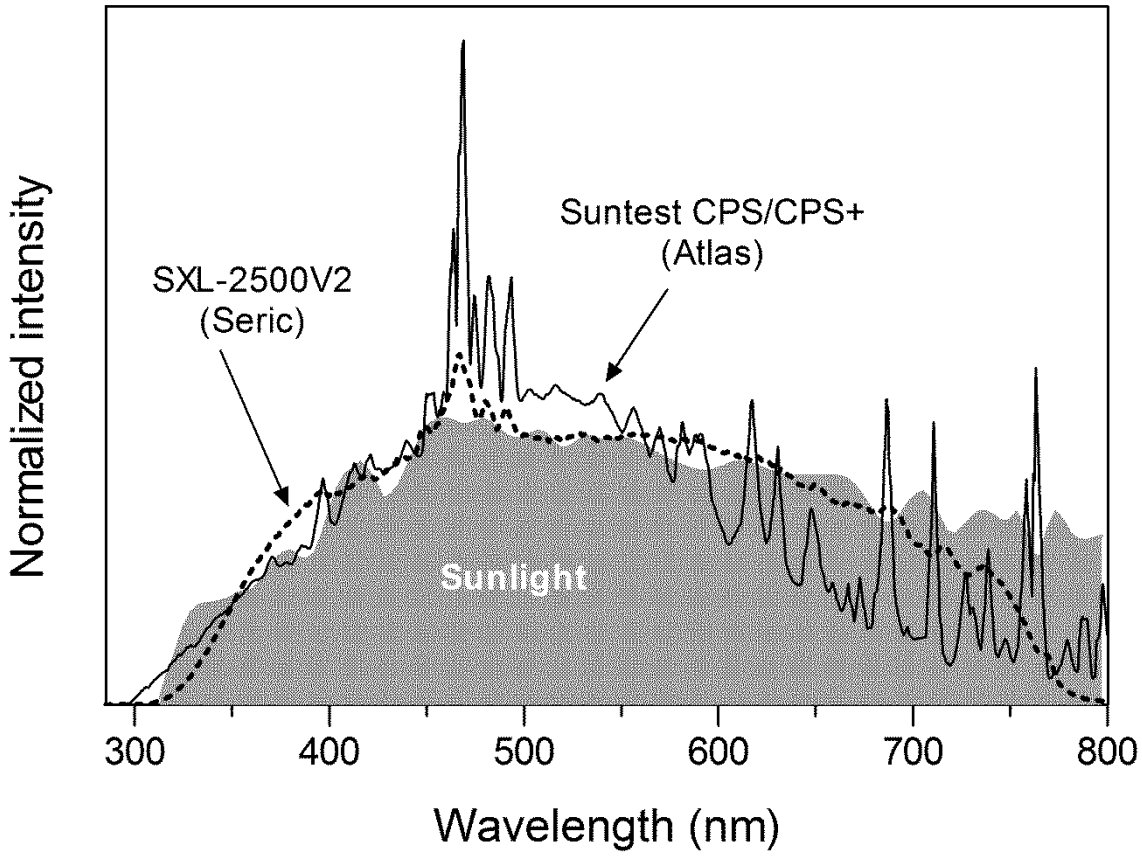
ROS: Reactive Oxygen Species, including superoxide anion (SA) and singlet oxygen (SO).

SA: Superoxide anion is one of radical species, generated from photo-irradiated chemicals through type I photochemical reaction.

SO: Singlet oxygen is one of radical species, generated from photo-irradiated chemicals through type II photochemical reaction.

UV light wavebands: The designations recommended by the CIE (Commission Internationale de L'Eclairage) are: UVA (315–400 nm) UVB (280–315 nm) and UVC (100–280 nm). Other designations are also used; the division between UVB and UVA is often placed at 320 nm, and the UVA may be divided into UV-A1 and UV-A2 with a division made at about 340 nm.

Annex B. Spectrum of solar stimulators used in the validation studies.



Annex C. Proficiency Chemicals

Prior to routine use of the test method described in this Test Guideline, laboratories should demonstrate technical proficiency by correctly obtaining the expected ROS prediction for proficiency chemicals recommended in the Table. For Suntest CPS/CPS+ (Atlas) or SXL-2500V2 (Seric) solar simulators, nine chemicals (Nos. 1–9) are to be tested. For other solar simulators, all 17 chemicals (Nos. 1–17) are to be tested. These proficiency chemicals were selected to represent the range of responses for phototoxic potential. 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 ROS assay are available, and that they were used in the JaCVAM-coordinated validation study to demonstrate successful implementation of the test method in the laboratories participating in the study (20)(21).

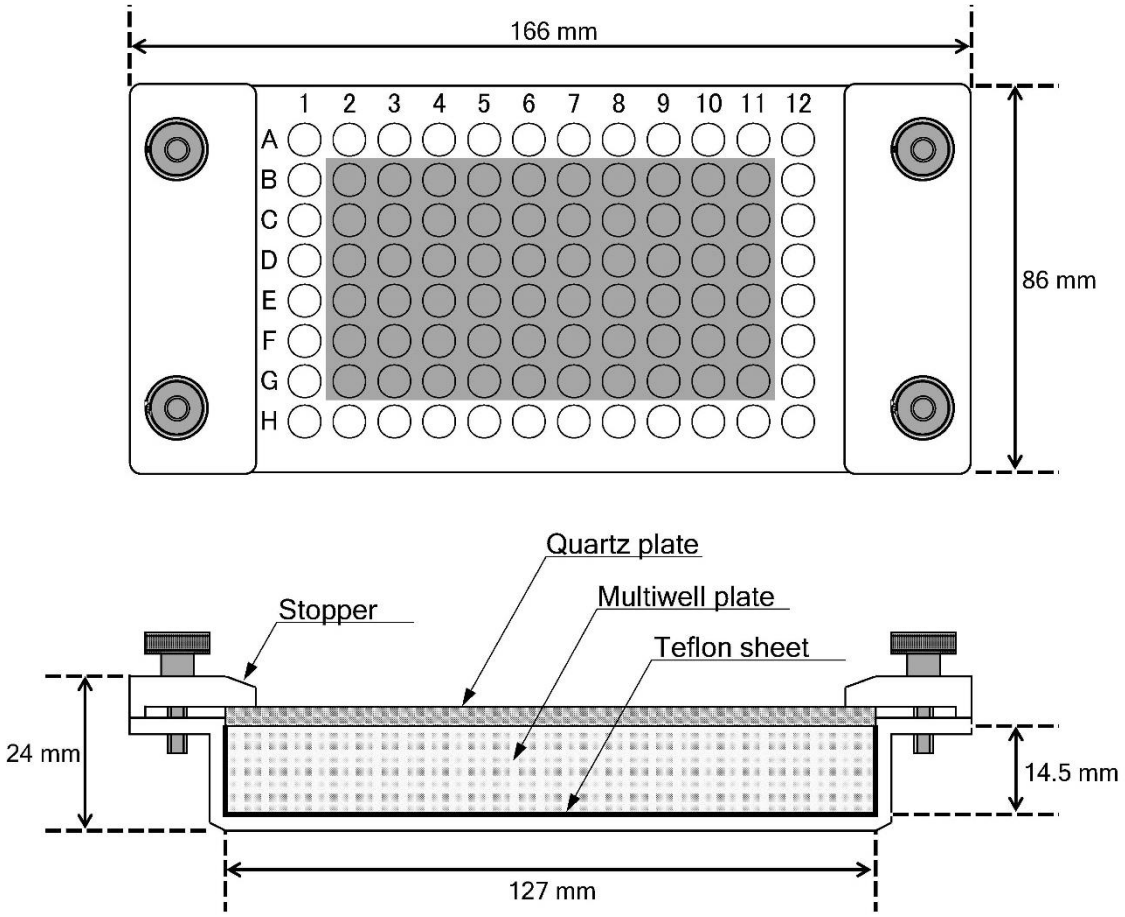
Table A C.1. Table of proficiency chemicals.

The expected ROS prediction for proficiency chemicals and the acceptable range..

No.	Chemical[1]	CAS No.	SO [2]	SA [2]	Solvent	Concentration
1	*p*-Aminobenzoic acid	150-13-0	-8 to 12	-11 to 7	DMSO	200 µM
2	Benzocaine	94-09-7	-7 to 9	-7 to 17	DMSO	200 µM
3	Doxycycline hydrochloride	10592-13-9	115 to 429	230 to 468	DMSO	200 µM
4	Erythromycin	114-07-8	-15 to 11	-9 to 21	DMSO	200 µM
5	Fenofibrate	49562-28-9	77 to 203	-31 to 11	DMSO	20 µM
6	L-Histidine	71-00-1	-8 to 12	8 to 120	NaPB	200 µM
7	Norfloxacin	70458-96-7	131 to 271	57 to 161	DMSO	200 µM
8	8-Methoxy psoralen	298-81-7	31 to 137	0 to 126	DMSO	200 µM
9	Octyl salicylate	118-60-5	-5 to 11	-8 to 20	DMSO	20 µM
10	Acridine	260-94-6	182 to 328	121 to 243	DMSO	200 µM
11	Chlorpromazine hydrochloride	69-09-0	-56 to 70	66 to 106	DMSO	200 µM
12	Diclofenac	15307-79-6	34 to 416	47 to 437	DMSO	200 µM
13	Furosemide	54-31-9	31 to 225	-7 to 109	DMSO	200 µM
14	Ketoprofen	22071-15-4	120 to 346	77 to 151	DMSO	200 µM
15	Nalidixic acid	389-08-2	54 to 246	88 to 470	DMSO	200 µM
16	Omeprazole	73590-58-6	-221 to 103	30 to 216	DMSO	200 µM
17	Promethazine hydrochloride	58-33-3	20 to 168	-3 to 77	DMSO	200 µM

[1]All chemicals are solid
[2]The values were calculated as means +/- 1.96 SD from the validation data..

Annex D. Quartz reaction container used in the validation studies.

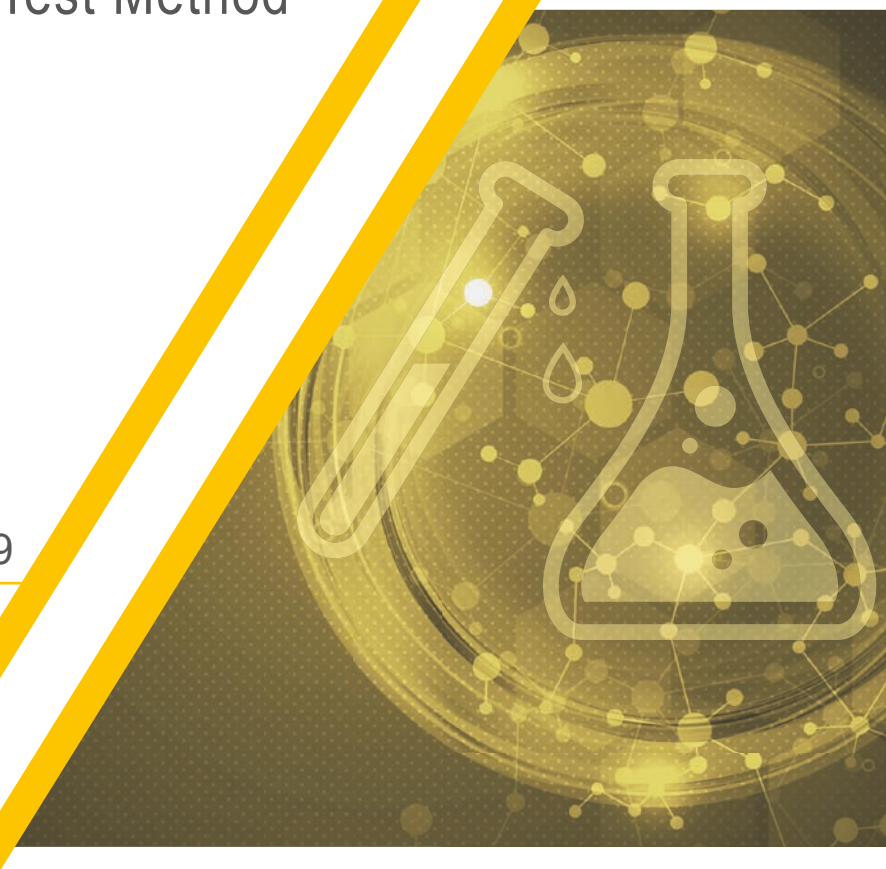


Recommended thickness of quartz plate: ca. 3 mm.

Section 4
Health effects

Test Guideline No. 431

In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE)Test Method



18 June 2019

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

OECD GUIDELINE FOR TESTING OF CHEMICALS

In Vitro Skin Corrosion: Reconstructed Human Epidermis (RhE) Test Method

INTRODUCTION

1. Skin corrosion refers to the production of irreversible damage to the skin manifested as visible necrosis through the epidermis and into the dermis, following the application of a test chemical [as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS)] (1). This updated Test Guideline 431 provides an in vitro procedure allowing the identification of non-corrosive and corrosive substances and mixtures in accordance with UN GHS (1). It also allows a partial sub-categorisation of corrosives.

2. The assessment of skin corrosion potential of chemicals has typically involved the use of laboratory animals (OECD Test Guideline 404 (TG 404); originally adopted in 1981 and revised in 1992, 2002 and 2015) (2). In addition to the present TG 431, two other in vitro test methods for testing corrosion potential of chemicals have been validated and adopted as OECD Test Guidelines 430 (3) and 435 (4). Furthermore the in vitro OECD TG 439 (5) has been adopted for testing skin irritation potential. A document on Integrated Approaches to Testing and Assessment (IATA) for Skin Corrosion and Irritation describes several modules which group information sources and analysis tools, and provides guidance on (i) how to integrate and use existing testing and non-testing data for the assessment of skin irritation and skin corrosion potentials of chemicals and (ii) proposes an approach when further testing is needed (6).

3. This Test Guideline addresses the human health endpoint skin corrosion. It makes use of reconstructed human epidermis (RhE) (obtained from human derived non-transformed epidermal keratinocytes) which closely mimics the histological, morphological, biochemical and physiological properties of the upper parts of the human skin, i.e. the epidermis. This Test Guideline was originally adopted in 2004 and updated in 2013, 2016 and 2019 to include additional test methods using the RhE models. The Test Guideline was also updated in 2015 to introduce the possibility to use the methods to support the sub-categorisation of corrosive chemicals, and to refer to the IATA guidance document, and introduce the use of an alternative procedure to measure viability.

4. Five validated test methods using commercially available RhE models are included in this Test Guideline, as described below. Prevalidation studies (7), followed by a formal validation study for assessing skin corrosion (8) (9) (10) have been conducted (11) (12) for two of these commercially available test methods, EpiSkin™ Standard Model (SM), and EpiDerm™ Skin Corrosivity Test (SCT) (EPI-200) (referred to in the following text as the Validated Reference Methods – VRMs, EpiSkin^{TM}=VRM1, EpiDerm^{TM}= VRM2). The outcome of these studies led to the recommendation that the two VRMs mentioned above could be used for regulatory purposes for distinguishing corrosive (C) from non-corrosive (NC) substances, and that the EpiSkin™ could moreover be used to support sub-categorisation of corrosive substances (13) (14) (15). Two other commercially available in vitro skin corrosion RhE test methods have subsequently shown similar results to the EpiDerm™ SCT according to PS-based Validation (16) (17) (18). These are the SkinEthicTM RHE1 and epiCS® (previously named EST-1000) that can also be used for regulatory purposes for distinguishing corrosive from non-corrosive substances (19) (20). Post validation studies performed by the RhE model producers in the years 2012 to 2014 with a refined protocol correcting interferences of unspecific MTT reduction by the test chemicals improved the performance of both discrimination of C/NC as well as supporting sub-categorization of corrosives (21) (22). Further statistical analyses of the post-validation data generated with Epiderm™ SCT, SkinEthic™ RHE and epiCS® have been performed to identify alternative predictions models that improved the predictive capacity for sub-categorisation (23). Finally, the LabCyte EPI-MODEL24 is another commercially available in vitro skin corrosion RhE test that was shown to be scientific similar to the VRMs and can therefore be used for regulatory purposes to distinguish corrosive from non-corrosive substances as well as support sub-categorization of corrosives (40) (41) (42)(43).

5. Before a proposed similar or modified in vitro RhE test method for skin corrosion other than the VRMs can be used for regulatory purposes, its reliability, relevance (accuracy), and limitations for its proposed use should be determined to ensure its similarity to the VRMs, in accordance with the requirements of the Performance Standards (PS) (24) set out in accordance with the principles of Guidance Document No.34 (25). The Mutual Acceptance of Data will only be guaranteed after any proposed new or updated test method following the PS have been reviewed and included in this Test Guideline. The test methods included in this Test Guideline can be used to address countries' requirements for test results on in vitro test method for skin corrosion, while benefiting from the Mutual Acceptance of Data.

DEFINITIONS

6. Definitions used are provided in Annex I.

INITIAL CONSIDERATIONS

7. This Test Guideline allows the identification of non-corrosive and corrosive substances and mixtures in accordance with the UN GHS (1). This Test Guideline further supports the sub-categorisation of corrosive substances and mixtures into optional Sub-category 1A, in accordance with the UN GHS (1), as well as a combination of Sub-categories 1B and 1C (21) (22) (23). A limitation of this Test Guideline is that it does not allow discriminating between skin corrosive Sub-category 1B and Sub-category 1C in accordance with the UN GHS (1) due to the limited set of well-known in vivo corrosive Sub-category 1C chemicals. The five test methods under this test guideline are able to discriminate sub-categories 1A versus 1B-and-1C versus NC.

8. A wide range of chemicals representing mainly individual substances has been tested in the validation studies supporting the test methods included in this Test Guideline. The original database of the validation study conducted for identification of non-corrosives versus corrosives amounted to 60 chemicals covering a wide range of chemical classes (8) (9) (10). Testing to demonstrate sensitivity, specificity, accuracy and within-laboratory-reproducibility of the assay for sub-categorisation was further performed by the test method developers using 79 to 80 chemicals also covering a wide range of chemical classes, and results were reviewed by the OECD (21) (22) (23). On the basis of the overall data available, the Test Guideline is applicable to a wide range of chemical classes and physical states including liquids, semi-solids, solids and waxes. The liquids may be aqueous or non-aqueous; solids may be soluble or insoluble in water. Whenever possible, solids should be ground to a fine powder before application; no other prior treatment of the sample is required. In cases where evidence can be demonstrated on the non-applicability of test methods included in the Test Guideline to a specific category of test chemicals, these test methods should not be used for that specific category of test chemicals. In addition, this Test Guideline is assumed to be applicable to mixtures as an extension of its applicability to substances. However, due to the fact that mixtures cover a wide spectrum of categories and composition, and that only limited information is currently available on the testing of mixtures, in cases where evidence can be demonstrated on the non-applicability of the Test Guideline to a specific category of mixtures (e.g. following a strategy as proposed in (26)), the Test Guideline should not be used for that specific category of mixtures. 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. Such considerations are not needed, when there is a regulatory requirement for testing of the mixture. Gases and aerosols have not been assessed yet in validation studies (8) (9) (10). While it is conceivable that these can be tested using RhE technology, the current Test Guideline does not allow testing of gases and aerosols.

9. Test chemicals absorbing light in the same range as MTT formazan and test chemicals able to directly reduce the vital dye MTT (to MTT formazan) may interfere with the tissue viability measurements and need the use of adapted controls for corrections. The type of adapted controls that may be required will vary depending on the type of interference produced by the test chemical and the procedure used to measure MTT formazan (see paragraphs 25-31).

10. While this Test Guideline does not provide adequate information on skin irritation, it should be noted that OECD TG 439 specifically addresses the health effect skin irritation in vitro and is based on the same RhE test system, though using another protocol (5). For a full evaluation of local skin effects after a single dermal exposure, the Guidance Document No. 203 on Integrated Approaches for Testing Assessment should be consulted (6). This IATA approach includes the conduct of in vitro tests for skin corrosion (such as described in this Test Guideline) and skin irritation before considering testing in living animals. It is recognized that the use of human skin is subject to national and international ethical considerations and conditions.

PRINCIPLE OF THE TEST

11. The test chemical is applied topically to a three-dimensional RhE model, comprised of non-transformed, human-derived epidermal keratinocytes, which have been cultured to

form a multi-layered, highly differentiated model of the human epidermis. It consists of organized basal, spinous and granular layers, and a multi-layered stratum corneum containing intercellular lamellar lipid layers representing main lipid classes analogous to those found in vivo.

12. The RhE test method is based on the premise that corrosive chemicals are able to penetrate the stratum corneum by diffusion or erosion, and are cytotoxic to the cells in the underlying layers. Cell viability is measured by enzymatic conversion of the vital dye MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide; CAS number 298-93-1], into a blue formazan salt that is quantitatively measured after extraction from tissues (27). Corrosive chemicals are identified by their ability to decrease cell viability below defined threshold levels (see paragraphs 35 and 36). The RhE-based skin corrosion test methods have shown to be predictive of in vivo skin corrosion effects assessed in rabbits according to the OECD guideline 404 (2).

DEMONSTRATION OF PROFICIENCY

13. Prior to routine use of any of the five validated RhE test methods that adhere to this Test Guideline, laboratories should demonstrate technical proficiency by correctly classifying the twelve Proficiency Substances listed in Table 1. In case of the use of a method for sub-classification, also the correct sub-categorisation should be demonstrated. In situations where a listed substance is unavailable or where justifiable, another substance for which adequate in vivo and in vitro reference data are available may be used (e.g. from the list of reference chemicals (24)) provided that the same selection criteria as described in Table 1 are applied.

Table 1. List of Proficiency Substances[1]

Substance	CASRN	Chemical Class[2]	UN GHS Cat. Based on *In Vivo* results [3]	Cat. Based on *In Vitro* results[4]	Mean cell viability for VRMs				Physical State
					VRM1		VRM2		
					3 min	60 min.	3 min.	60 min	
Sub-category 1A *In Vivo* Corrosives									
Bromoacetic acid	79-08-3	Organic acid	1A	(3) 1A	3	2.8	3.2	2.8	S
Boron trifluoride dihydrate	13319-75-01	Inorganic acid	1A	(3) 1A	2.4	4.2	4.4	10.1	L
Phenol	108-95-2	Phenol	1A	(3) 1A	29.8	21.8	22.6	13.5	S
Dichloroacetyl chloride	79-36-7	Electrophile	1A	(3) 1A	5.6	6.3	1.3	1.4	L
Combination of sub-categories 1B-and-1C *In Vivo* Corrosives									
Glyoxylic acid monohydrate	563-96-2	Organic acid	1B-and-1C	(3) 1B-and-1C	110.4	22.5	90.4	3.1	S
Lactic acid	598-82-3	Organic acid	1B-and-1C	(3) 1B-and-1C	80.2	9.4	90	3.5	L
Ethanolamine	141-43-5	Organic base	1B	(3) 1B-and-1C	66.2	40.3	69.7	9.3	Viscous
Hydrochloric acid (14.4%)	7647-01-0	Inorganic acid	1B-and-1C	(3) 1B-and-1C	69.3	5.7	80.8	9	L

***In Vivo* Non Corrosives**									
Phenethyl bromide	103-63-9	Electrophile	NC	(3) NC	141	117.2	112.5	71.2	N
4-Amino-1,2,4-triazole	584-13-4	Organic base	NC	(3) NC	116.8	120.6	105.7	88.2	N
4-(methylthio)-benzaldehyde	3446-89-7	Electrophile	NC	(3) NC	136.7	150.4	85.4	81.6	N
Lauric acid	143-07-7	Organic acid	NC	(3) NC	102	117.4	90.7	64.4	N

Abbreviations: CASRN = Chemical Abstracts Service Registry Number; UN GHS = United Nations Globally Harmonized System (1); VRM = Validated Reference Method, EpiSkin™=VRM1, EpiDerm™= VRM2; NC = Not Corrosive

[1]The proficiency substances, sorted first by corrosives versus non-corrosives, then by corrosive sub-category and then by chemical class, were selected from the substances used in the ECVAM validation studies EpiSkin™ and EpiDerm™ (8) (9) (10) and from post-validation studies based on data provided by EpiSkin™ (22), EpiDerm™, SkinEthic™ and epiCS® developers (23). Unless otherwise indicated, the substances were tested at the purity level obtained when purchased from a commercial source (8) (10). The selection includes, to the extent possible, substances that: (i) are representative of the range of corrosivity responses (e.g. non-corrosives; weak to strong corrosives) that the VRMs are capable of measuring or predicting; (ii) are representative of the chemical classes used in the validation studies; (iii) have chemical structures that are well-defined; (iv) induce reproducible results in the VRM; (v) induce definitive results in the in vivo reference test method; (vi) are commercially available; and (vii) are not associated with prohibitive disposal costs.

[2]Chemical class assigned by Barratt et al. (8).

[3]The corresponding UN Packing groups are I, II and III, respectively, for the UN GHS 1A, 1B and 1C.

[4]The in vitro predictions reported in this table were obtained with all five test methods covered in TG 431; for phenol though the LabCyte EPI-MODEL24 had slightly discordant results across runs, i.e. 1A-1BC-1BC; other methods achieved these classifications in validation or post-validation testing performed by the test method developers.

[5]The viability values obtained in the ECVAM Skin Corrosion Validation Studies were not corrected for direct MTT reduction (killed controls were not performed in the validation studies). However, the post-validation data generated by the test method developers that are presented in this table were acquired with adapted controls (23).

14. As part of the proficiency exercise, it is recommended that the user verifies the barrier properties of the tissues after receipt as specified by the RhE model manufacturer. This is particularly important if tissues are shipped over long distance/time periods. Once a test method has been successfully established and proficiency in its use has been demonstrated, such verification will not be necessary on a routine basis. However, when using a test method routinely, it is recommended to continue to assess the barrier properties in regular intervals.

PROCEDURE

15. The following is a generic description of the components and procedures of the RhE test methods for skin corrosion assessment covered by this Test Guideline. The RhE models endorsed as scientifically valid for use within this Test Guideline, i.e. the EpiSkin™ (SM), EpiDerm™ (EPI-200), SkinEthic™ RHE, epiCS® and LabCyte EPI-MODEL24 (16) (17) (19) (28) (29) (30) (31) (32) (33) (40) (41), can be obtained from commercial sources. Standard Operating Procedures (SOPs) for these five RhE models are available (34) (35) (36) (37) (42), and their main test method components are summarised in Annex 2. It is recommended that the relevant SOP be consulted when implementing and using one of these methods in the laboratory. Testing with the five RhE test methods covered by this Test Guideline should comply with the following:

RHE TEST METHOD COMPONENTS

General conditions

16. Non-transformed human keratinocytes should be used to reconstruct the epithelium. Multiple layers of viable epithelial cells (basal layer, stratum spinosum, stratum granulosum) should be present under a functional stratum corneum. The stratum corneum should be multi-layered containing the essential lipid profile to produce a functional barrier with robustness to resist rapid penetration of cytotoxic benchmark chemicals, e.g. sodium dodecyl sulphate (SDS) or Triton X-100. The barrier function should be demonstrated and may be assessed either by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50% (IC50) after a fixed exposure time, or by determination of the exposure time required to reduce cell viability by 50% (ET50) upon application of the benchmark chemical at a specified, fixed concentration (see paragraph 18). The containment properties of the RhE model should prevent the passage of material around the stratum corneum to the viable tissue, which would lead to poor modelling of skin exposure. The RhE model should be free of contamination by bacteria, viruses, mycoplasma, or fungi.

Functional conditions

Viability

17. The assay used for quantifying tissue viability is the MTT-assay (27). The viable cells of the RhE tissue construct reduce the vital dye MTT into a blue MTT formazan precipitate, which is then extracted from the tissue using isopropanol (or a similar solvent). The OD of the extraction solvent alone should be sufficiently small, i.e., OD < 0.1. The extracted MTT formazan may be quantified using either a standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure (38). The RhE model users should ensure that each batch of the RhE model used meets defined criteria for the negative control. An acceptability range (upper and lower limit) for the negative control OD values should be established by the RhE model developer/supplier. Acceptability ranges for the negative control OD values for the five validated RhE test methods included in this Test Guideline are given in Table 2. An HPLC/UPLC-Spectrophotometry user should use the negative control OD ranges provided in Table 2 as the acceptance criterion for the negative control. It should be documented that the tissues treated with negative control are stable in culture (provide similar OD measurements) for the duration of the exposure period.

Table 2. Acceptability ranges for negative control OD values to control batch quality

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM)	≥ 0.6	≤ 1.5
EpiDerm™ SCT (EPI-200)	≥ 0.8	≤ 2.8
SkinEthic™ RHE	≥ 0.8	≤ 3.0
epiCS	≥ 0.8	≤ 2.8
LabCyte EPI-MODEL24 SCT	≥ 0.7	≤ 2.5

Barrier function

18. The stratum corneum and its lipid composition should be sufficient to resist the rapid penetration of certain cytotoxic benchmark chemicals (e.g. SDS or Triton X-100), as estimated by IC50 or ET50 (Table 3). The barrier function of each batch of the RhE model used should be demonstrated by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Morphology

19. Histological examination of the RhE model should be performed demonstrating multi-layered human epidermis-like structure containing stratum basale, stratum spinosum, stratum granulosum and stratum corneum and exhibits lipid profile similar to lipid profile of human epidermis. Histological examination of each batch of the RhE model used demonstrating appropriate morphology of the tissues should be provided by the RhE model developer/vendor upon supply of the tissues to the end user (see paragraph 21).

Reproducibility

20. Test method users should demonstrate reproducibility of the test methods over time with the positive and negative controls. Furthermore, the test method should only be used if the RhE model developer/supplier provides data demonstrating reproducibility over time with corrosive and non-corrosive chemicals from e.g. the list of Proficiency Substances (Table 1). In case of the use of a test method for sub-categorisation, the reproducibility with respect to sub-categorisation should also be demonstrated.

Quality control (QC)

21. The RhE model should only be used if the developer/supplier demonstrates that each batch of the RhE model used meets defined production release criteria, among which those for viability (paragraph 17), barrier function (paragraph 18) and morphology (paragraph 19) are the most relevant. These data are provided to the test method users, so that they are able to include this information in the test report. Only results produced with QC accepted tissue batches can be accepted for reliable prediction of corrosive classification. An acceptability range (upper and lower limit) for the IC50 or the ET50 is established by the RhE model developer/supplier. The acceptability ranges for the five validated test methods are given in Table 3.

Table 3. **QC batch release criterion**

	Lower acceptance limit	Upper acceptance limit
EpiSkin™ (SM) (18 hours treatment with SDS)(33)	IC_{50} = 1.0 mg/mL	IC_{50} = 3.0 mg/mL
EpiDerm™SCT (EPI-200) (1% Triton X-100)(34)	ET_{50} = 4.0 hours	ET_{50} = 8.7 hours
SkinEthic™ RHE (1% Triton X-100)(35)	ET_{50} = 4.0 hours	ET_{50} = 10.0 hours
epiCS (1% Triton X-100)(36)	ET_{50} = 2.0 hours	ET_{50} = 7.0 hours
LabCyte EPI-MODEL24 SCT (18 hours treatment with SDS) (42)	IC_{50} = 1.4 mg/mL	IC_{50} = 4.0 mg/mL

Application of the Test Chemical and Control Substances

22. At least two tissue replicates should be used for each test chemical and controls for each exposure time. For liquid as well as solid chemicals, sufficient amount of test chemical should be applied to uniformly cover the epidermis surface while avoiding an infinite dose, i.e. a minimum of 70 µL/cm2 or 30 mg/cm2 should be used. Depending on the methods, the epidermis surface should be moistened with deionized or distilled water before application of solid chemicals, to improve contact between the test chemical and the epidermis surface (34) (35) (36) (37) (42). Whenever possible, solids should be tested as a fine powder. The application method should be appropriate for the test chemical (see e.g. references (34-37). At the end of the exposure period, the test chemical should be carefully washed from the epidermis with an aqueous buffer, or 0.9% NaCl. Depending on which of the five validated RhE test methods is used, two or three exposure periods are used per test chemical (for all five valid RhE models: 3 min and 1 hour; for EpiSkin™ an additional exposure time of 4 hours). Depending on the RhE test method used and the exposure period assessed, the incubation temperature during exposure may vary between room temperature and 37ºC.

23. Concurrent negative and positive controls (PC) should be used in each run to demonstrate that viability (with negative controls), barrier function and resulting tissue sensitivity (with the PC) of the tissues are within a defined historical acceptance range. The suggested PC chemicals are glacial acetic acid or 8N KOH depending upon the RhE model used (see Annex 2 and relevant SOP for details). It should be noted that 8N KOH is a direct MTT reducer that might require adapted controls as described in paragraphs 25 and 26. The suggested negative controls are 0.9% (w/v) NaCl or water.

Cell Viability Measurements

24. The MTT assay, which is a quantitative assay, should be used to measure cell viability under this Test Guideline (27). The tissue sample is placed in MTT solution of appropriate concentration (0.3, 0.5 or 1 mg/mL, see Annex 2 and relevant SOP for details) for 3 hours. The precipitated blue formazan product is then extracted from the tissue using a solvent (e.g. isopropanol, acidic isopropanol), and the concentration of formazan is

measured by determining the OD at 570 nm using a filter band pass of maximum ± 30 nm, or by an HPLC/UPLC spectrophotometry procedure (see paragraphs 30 and 31) (38).

25. Test chemicals may interfere with the MTT assay, either by direct reduction of the MTT into blue formazan, and/or by colour interference if the test chemical absorbs, naturally or due to treatment procedures, in the same OD range of formazan (570 ± 30 nm, mainly blue and purple chemicals). Additional controls should be used to detect and correct for a potential interference from these test chemicals such as the non-specific MTT reduction (NSMTT) control and the non-specific colour (NSC) control (see paragraphs 26 to 30). This is especially important when a specific test chemical is not completely removed from the tissue by rinsing or when it penetrates the epidermis, and is therefore present in the tissues when the MTT viability test is performed. Detailed description of how to correct direct MTT reduction and interferences by colouring agents is available in the SOPs for the test methods (34) (35) (36) (37) (42).

26. To identify direct MTT reducers, each test chemical should be added to freshly prepared MTT medium (34) (35) (36) (37) (42). If the MTT mixture containing the test chemical turns blue/purple, the test chemical is presumed to directly reduce the MTT, and further functional check on non-viable epidermis should be performed, independently of using the standard absorbance (OD) measurement or an HPLC/UPLC-spectrophotometry procedure. This additional functional check employs killed tissues that possess only residual metabolic activity but absorb the test chemical in similar amount as viable tissues. Each MTT reducing chemical is applied on at least two killed tissue replicates per exposure time, which undergo the whole skin corrosion test. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the MTT reducer minus the percent non-specific MTT reduction obtained with the killed tissues exposed to the same MTT reducer, calculated relative to the negative control run concurrently to the test being corrected (%NSMTT).

27. To identify potential interference by coloured test chemicals or test chemicals that become coloured when in contact with water or isopropanol and decide on the need for additional controls, spectral analysis of the test chemical in water (environment during exposure) and/or isopropanol (extracting solution) should be performed. If the test chemical in water and/or isopropanol absorbs light in the range of 570 ± 30 nm, further colorant controls should be performed or, alternatively, an HPLC/UPLC-spectrophotometry procedure should be used in which case these controls are not required (see paragraphs 30 and 31). When performing the standard absorbance (OD) measurement, each interfering coloured test chemical is applied on at least two viable tissue replicates per exposure time, which undergo the entire skin corrosion test but are incubated with medium instead of MTT solution during the MTT incubation step to generate a non-specific colour (NSCliving) control. The NSCliving control needs to be performed concurrently per exposure time per coloured test chemical (in each run) due to the inherent biological variability of living tissues. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the interfering test chemical and incubated with MTT solution minus the percent non-specific colour obtained with living tissues exposed to the interfering test chemical and incubated with medium without MTT, run concurrently to the test being corrected (%NSCliving).

28. Test chemicals that are identified as producing both direct MTT reduction (see paragraph 26) and colour interference (see paragraph 27) will also require a third set of controls, apart from the NSMTT and NSCliving controls described in the previous

paragraphs, when performing the standard absorbance (OD) measurement. This is usually the case with darkly coloured test chemicals interfering with the MTT assay (e.g., blue, purple, black) because their intrinsic colour impedes the assessment of their capacity to directly reduce MTT as described in paragraph 26. These test chemicals may bind to both living and killed tissues and therefore the NSMTT control may not only correct for potential direct MTT reduction by the test chemical, but also for colour interference arising from the binding of the test chemical to killed tissues. This could lead to a double correction for colour interference since the NSCliving control already corrects for colour interference arising from the binding of the test chemical to living tissues. To avoid a possible double correction for colour interference, a third control for non-specific colour in killed tissues (NSCkilled) needs to be performed. In this additional control, the test chemical is applied on at least two killed tissue replicates per exposure time, which undergo the entire testing procedure but are incubated with medium instead of MTT solution during the MTT incubation step. A single NSCkilled control is sufficient per test chemical regardless of the number of independent tests/runs performed, but should be performed concurrently to the NSMTT control and, where possible, with the same tissue batch. The true tissue viability is then calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT minus %NSCliving plus the percent non-specific colour obtained with killed tissues exposed to the interfering test chemical and incubated with medium without MTT, calculated relative to the negative control run concurrently to the test being corrected (%NSCkilled).

29. It is important to note that non-specific MTT reduction and non-specific colour interferences may increase the readouts of the tissue extract above the linearity range of the spectrophotometer. On this basis, each laboratory should determine the linearity range of their spectrophotometer with MTT formazan (CAS # 57360-69-7) from a commercial source before initiating the testing of test chemicals for regulatory purposes. In particular, the standard absorbance (OD) measurement using a spectrophotometer is appropriate to assess direct MTT-reducers and colour interfering test chemicals when the ODs of the tissue extracts obtained with the test chemical without any correction for direct MTT reduction and/or colour interference are within the linear range of the spectrophotometer or when the uncorrected percent viability obtained with the test chemical already defined it as a corrosive (see paragraphs 35 and 36). Nevertheless, results for test chemicals producing %NSMTT and/or %NSCliving ≥ 50% of the negative control should be taken with caution.

30. For coloured test chemicals which are not compatible with the standard absorbance (OD) measurement due to too strong interference with the MTT assay, the alternative HPLC/UPLC-spectrophotometry procedure to measure MTT formazan may be employed (see paragraph 31) (37). The HPLC/UPLC-spectrophotometry system allows for the separation of the MTT formazan from the test chemical before its quantification (38). For this reason, NSCliving or NSCkilled controls are never required when using HPLC/UPLC-spectrophotometry, independently of the chemical being tested. NSMTT controls should nevertheless be used if the test chemical is suspected to directly reduce MTT or has a colour that impedes the assessment of the capacity to directly reduce MTT (as described in paragraph 26). When using HPLC/UPLC-spectrophotometry to measure MTT formazan, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. For test chemicals able to directly reduce MTT, true tissue viability is calculated as the percent tissue viability obtained with living tissues exposed to the test chemical minus %NSMTT. Finally, it should be noted that direct MTT-reducers

that may also be colour interfering, which are retained in the tissues after treatment and reduce MTT so strongly that they lead to ODs (using standard OD measurement) or peak areas (using UPLC/HPLC-spectrophotometry) of the tested tissue extracts that fall outside of the linearity range of the spectrophotometer cannot be assessed, although these are expected to occur in only very rare situations.

31. HPLC/UPLC-spectrophotometry may be used also with all types of test chemicals (coloured, non-coloured, MTT-reducers and non-MTT reducers) for measurement of MTT formazan (38). Due to the diversity of HPLC/UPLC-spectrophotometry systems, qualification of the HPLC/UPLC-spectrophotometry system should be demonstrated before its use to quantify MTT formazan from tissue extracts by meeting the acceptance criteria for a set of standard qualification parameters based on those described in the U.S. Food and Drug Administration guidance for industry on bio-analytical method validation (38) (39). These key parameters and their acceptance criteria are shown in Annex 4. Once the acceptance criteria defined in Annex 4 have been met, the HPLC/UPLC-spectrophotometry system is considered qualified and ready to measure MTT formazan under the experimental conditions described in this Test Guideline.

Acceptance Criteria

32. For each test method using valid RhE models, tissues treated with the negative control should exhibit OD reflecting the quality of the tissues as described in table 2 and should not be below historically established boundaries. Tissues treated with the PC, i.e. glacial acetic acid or 8N KOH, should reflect the ability of the tissues to respond to a corrosive chemical under the conditions of the test method (see Annex 2 and relevant SOP for details). The variability between tissue replicates of test chemical and/or control substances should fall within the accepted limits for each valid RhE model requirements (see Annex 2 and relevant SOP for details) (e.g. the difference of viability between the two tissue replicates should not exceed 30%). If either the negative control or PC included in a run fall out of the accepted ranges, the run is considered as not qualified and should be repeated. If the variability of test chemicals falls outside of the defined range, its testing should be repeated.

Interpretation of Results and Prediction Model

33. The OD values obtained for each test chemical should be used to calculate percentage of viability relative to the negative control, which is set at 100%. In case HPLC/UPLC-spectrophotometry is used, the percent tissue viability is calculated as percent MTT formazan peak area obtained with living tissues exposed to the test chemical relative to the MTT formazan peak obtained with the concurrent negative control. The cut-off percentage cell viability values distinguishing corrosive from non-corrosive test chemical (or discriminating between different corrosive sub-categories) are defined below in paragraphs 35 and 36 for each of the test methods covered by this Test Guideline and should be used for interpreting the results.

34. A single testing run composed of at least two tissue replicates should be sufficient for a test chemical when the resulting classification is unequivocal. However, in cases of borderline results, such as non-concordant replicate measurements, a second run may be considered, as well as a third one in case of discordant results between the first two runs.

35. The prediction model for the EpiSkin™ skin corrosion test method (9) (34) (22), associated with the UN GHS (1) classification system, is shown in Table 4:

Table 4. EpiSkin™ prediction model

Viability measured after exposure time points (t='3,' 60 and 240 minutes)	Prediction to consider
< 35% after 3 min exposure	**Corrosive:** Optional Sub-category 1A *
≥ 35% after 3 min exposure **AND** < 35% after 60 min exposure **OR** ≥ 35% after 60 min exposure **AND** < 35% after 240 min exposure	**Corrosive:** A combination of optional Sub-categories 1B-and-1C
≥ 35% after 240 min exposure	**Non-corrosive**

*) According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 22 % of the Sub-category 1A results of the EpiSkin™ test method may actually constitute Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over classifications) (see Annex 3).

36. The prediction models for the EpiDerm™ SCT (10) (23) (35), the SkinEthic™ RHE (17) (18) (23) (36), the epiCS® (16) (23) (37) and LabCyte EPI-MODEL24 (41) (42) skin corrosion test methods, associated with the UN GHS (1) classification system, are shown in Table 5:

Table 5. EpiDerm™ SCT, SkinEthic™ RHE epiCS® and LabCyte EPI-MODEL24 SCT

Viability measured after exposure time points (t=3 and 60 minutes)	Prediction to be considered
STEP 1 for EpiDerm™ SCT, SkinEthic™ RHE, epiCS® and LabCyte EPI-MODEL24 SCT	
< 50% after 3 min exposure	Corrosive
≥ 50% after 3 min exposure AND < 15% after 60 min exposure	Corrosive
≥ 50% after 3 min exposure AND ≥ 15% after 60 min exposure	Non-corrosive
STEP 2 for EpiDerm™ SCT - for substances/mixtures identified as Corrosive in step 1	
< 25% after 3 min exposure	Optional Sub-category 1A *
≥ 25% after 3 min exposure	A combination of optional Sub-categories 1B-and-1C
STEP 2 for SkinEthic™ RHE - for substances/mixtures identified as Corrosive in step 1	
< 18% after 3 min exposure	Optional Sub-category 1A *
≥ 18% after 3 min exposure	A combination of optional Sub-categories 1B-and-1C
STEP 2 for epiCS® - for substances/mixtures identified as Corrosive in step 1	
< 15% after 3 min exposure	Optional Sub-category 1A *
≥ 15% after 3 min exposure	A combination of optional Sub-categories 1B-and-1C
STEP 2 for LabCyte EPI-MODEL24 SCT - for substances/mixtures identified as Corrosive in step 1	
< 15% after 3 min exposure	Optional Sub-category 1A *
≥ 15% after 3 min exposure	A combination of optional Sub-categories 1B-and-1C

* According to the data generated in view of assessing the usefulness of the RhE test methods for supporting sub-categorisation, it was shown that around 29%, 31%, 33% and 30% of the Sub-category 1A results of the EpiDerm™ SCT, SkinEthic™ RHE epiCS® and LabCyte EPI-MODEL24 SCT, respectively, may actually constitute Sub-category 1B or Sub-category 1C substances/mixtures (i.e. over-classifications) (see Annex 3).

DATA AND REPORTING

Data

37. For each test, data from individual tissue replicates (e.g. OD values and calculated percentage cell viability for each test chemical, including classification) should be reported in tabular form, including data from repeat experiments as appropriate. In addition, means and ranges of viability and CVs between tissue replicates for each test should be reported. Observed interactions with MTT reagent by direct MTT reducers or coloured test chemicals should be reported for each tested chemical.

Test Report

38. The test report should include the following information:

Test Chemical and Control Substances:

- Mono-constituent substance: chemical identification, such as IUPAC or CAS name, CAS number, SMILES or InChI code, structural formula, purity, chemical identity of impurities as appropriate and practically feasible, etc;
- Multi-constituent substance, UVCB and mixture: characterised as far as possible by chemical identity (see above), quantitative occurrence and relevant physicochemical properties of the constituents;
- Physical appearance, water solubility, and any additional relevant physicochemical properties;
- Source, lot number if available;
- Treatment of the test chemical/control substance prior to testing, if applicable (e.g. warming, grinding);
- Stability of the test chemical, limit date for use, or date for re-analysis if known;
- Storage conditions.

RhE model and protocol used and rationale for it (if applicable)

Test Conditions:

- RhE model used (including batch number);
- Calibration information for measuring device (e.g. spectrophotometer), wavelength and band
- pass (if applicable) used for quantifying MTT formazan, and linearity range of measuring device;
- Description of the method used to quantify MTT formazan;
- Description of the qualification of the HPLC/UPLC-spectrophotometry system, if applicable;
- Complete supporting information for the specific RhE model used including its performance. This should include, but is not limited to:
    - i) Viability;
    - ii) Barrier function;
    - iii) Morphology;
    - iv) Quality controls (QC) of the model;
- Reference to historical data of the model. This should include, but is not limited to acceptability of the QC data with reference to historical batch data;
- Demonstration of proficiency in performing the test method before routine use by testing of the proficiency substances.

Test Procedure:

- Details of the test procedure used (including washing procedures used after exposure period);
- Doses of test chemical and control substances used;
- Duration of exposure period(s) and temperature(s) of exposure;
- Indication of controls used for direct MTT-reducers and/or colouring test chemicals, if applicable;
- Number of tissue replicates used per test chemical and controls (PC, negative control, and NSMTT, NSCliving and NSCkilled, if applicable), per exposure time;
- Description of decision criteria/prediction model applied based on the RhE model used;
- Description of any modifications of the test procedure (including washing procedures).
- Run and Test Acceptance Criteria:
- Positive and negative control mean values and acceptance ranges based on historical data;
- Acceptable variability between tissue replicates for positive and negative controls;
- Acceptable variability between tissue replicates for test chemical.

Results:

- Tabulation of data for individual test chemicals and controls, for each exposure period, each run and each replicate measurement including OD or MTT formazan peak area, percent tissue viability, mean percent tissue viability, differences between replicates, SDs and/or CVs if applicable;
- If applicable, results of controls used for direct MTT-reducers and/or colouring test chemicals including OD or MTT formazan peak area, %NSMTT, %NSCliving, %NSCkilled, differences between tissue replicates, SDs and/or CVs (if applicable), and final correct percent tissue viability;
- Results obtained with the test chemical(s) and control substances in relation to the defined run and test acceptance criteria;
- Description of other effects observed;
- The derived classification with reference to the prediction model/decision criteria used.

Discussion of the results:

Conclusions:

LITERATURE

1. UN. (2013). United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Fifth Revised Edition, UN New York and Geneva. Available at: [http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html]
2. OECD. (2015). Guideline for Testing of Chemicals. (No. 404.): Acute Dermal Irritation, Corrosion, Organisation for Economic Cooperation and Development, Paris.
3. OECD. (2015). Guideline for the Testing of Chemicals (No. 430.): In Vitro Skin Corrosion: Transcutaneous Electrical Resistance (TER). Organisation for Economic Cooperation and Development, Paris.
4. OECD. (2015). Guideline for the Testing of Chemicals (No. 435.): In Vitro Membrane Barrier Test Method. Organisation for Economic Cooperation and Development, Paris. (5) OECD. (2015). Guideline for the Testing of Chemicals (No. 439.): In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Cooperation and Development, Paris.
5. OECD. (2015). Guideline for the Testing of Chemicals (No. 439.): In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Cooperation and Development, Paris.
6. OECD. (2014). Guidance Document on Integrated Approaches to Testing and Assessment of Skin Irritation/Corrosion. Environment, Health and Safety Publications, Series on Testing and Assessment, (No. 203.) Organisation for Economic Cooperation and Development, Paris.
7. Botham P.A., Chamberlain M., Barratt M.D., Curren R.D., Esdaile D.J., Gardner J.R., Gordon V.C., Hildebrand B., Lewis R.W., Liebsch M., Logemann P., Osborne R., Ponec M., Regnier J.F., Steiling W., Walker A.P., and Balls M. (1995). A Prevalidation Study on In Vitro Skin Corrosivity Testing. The report and Recommendations of ECVAM Workshop 6.ATLA 23, 219-255.
8. Barratt M.D., Brantom P.G., Fentem J.H., Gerner I., Walker A.P., and Worth A.P. (1998). The ECVAM International Validation Study on In Vitro Tests for Skin Corrosivity. 1. Selection and distribution of the Test Chemicals. Toxicol.In Vitro 12, 471-482.
9. Fentem J.H., Archer G.E.B., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J., Holzhutter H.- G., and Liebsch M. (1998). The ECVAM International Validation Study on In Vitro Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. Toxicol.in Vitro 12, 483-524.
10. Liebsch M., Traue D., Barrabas C., Spielmann H., Uphill, P., Wilkins S., Wiemann C., Kaufmann T., Remmele M. and Holzhütter H. G. (2000). The ECVAM Prevalidation Study on the Use of EpiDerm for Skin Corrosivity Testing, ATLA 28, pp. 371-401.
11. Balls M., Blaauboer B.J., Fentem J.H., Bruner L., Combes R.D., Ekwall B., Fielder R.J., Guillouzo A., Lewis R.W., Lovell D.P., Reinhardt C.A., Repetto G., Sladowski D., Spielmann H. et Zucco F. (1995). Practical Aspects of the Validation

of Toxicity Test Procedures. The Report and Recommendations of ECVAM Workshops, ATLA 23, 129-147.

12. ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (1997). Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No. 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA. Available at: [http://www.iccvam.niehs.nih.gov/docs/guidelines/validate.pdf].

13. ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2002). ICCVAM evaluation of EpiDerm™ (EPI-200), EPISKIN™ (SM), and the Rat Skin Transcutaneous Electrical Resistance (TER) Assay: In Vitro Test Methods for Assessing Dermal Corrosivity Potential of Chemicals. NIH Publication No. 02-4502. National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.Available at:[http://www.iccvam.niehs.nih.gov/methods/epiddocs/epis_brd.pdf]

14. EC-ECVAM. (1998). Statement on the Scientific Validity of the EpiSkin™ Test (an In Vitro Test for Skin Corrosivity), Issued by the ECVAM Scientific Advisory Committee (ESAC10), 3 April 1998. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

15. EC-ECVAM. (2000). Statement on the Application of the EpiDerm™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC14), 21 March 2000. Available at: [http://ecvam.jrc.ec.europa.eu].

16. Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr H.J., Vohr H.W. and Fuchs H.W. (2005). Epidermal-Skin-Test 1000 (EST-1000)-A New Reconstructed Epidermis for In Vitro Skin Corrosivity Testing. Toxicol.In Vitro 19, 925-929.

17. Kandárová H., Liebsch M., Spielmann,H., Genschow E., Schmidt E., Traue D., Guest R., Whittingham A., Warren N, Gamer A.O., Remmele M., Kaufmann T., Wittmer E., De Wever B., and Rosdy M. (2006). Assessment of the Human Epidermis Model SkinEthic RHE for In Vitro Skin Corrosion Testing of Chemicals According to New OECD TG 431. Toxicol.In Vitro 20, 547 559.

18. Tornier C., Roquet M. and Fraissinette A.B. (2010). Adaptation of the Validated SkinEthic™ Reconstructed Human Epidermis (RHE) Skin Corrosion Test Method to 0.5 cm2 Tissue Sample. Toxicol. In Vitro 24, 1379-1385.

19. EC-ECVAM. (2006). Statement on the Application of the SkinEthic™ Human Skin Model for Skin Corrosivity Testing, Issued by the ECVAM Scientific Advisory Committee (ESAC25), 17 November 2006. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

20. EC-ECVAM. (2009). ESAC Statement on the Scientific Validity of an In-Vitro Test Method for Skin Corrosivity Testing: the EST-1000, Issued by the ECVAM Scientific Advisory Committee (ESAC30), 12 June 2009. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

21. OECD. (2013). Summary Document on the Statistical Performance of Methods in OECD Test Guideline 431 for Sub-categorisation. Environment, Health, and Safety

Publications, Series on Testing and Assessment (No. 190.). Organisation for Economic Cooperation and Development, Paris.

22. Alépée N., Grandidier M.H., and Cotovio J. (2014). Sub-Categorisation of Skin Corrosive Chemicals by the EpiSkin™ Reconstructed Human Epidermis Skin Corrosion Test Method According to UN GHS: Revision of OECD Test Guideline 431. Toxicol.In Vitro 28, 131-145.
23. Desprez B., Barroso J., Griesinger C., Kandárová H., Alépée N., and Fuchs, H. (2015). Two Novel Prediction Models Improve Predictions of Skin Corrosive Sub-categories by Test Methods of OECD Test Guideline No. 431. Toxicol. In Vitro 29, 2055 2080.
24. OECD. (2015). Performance Standards for the Assessment of Proposed Similar or Modified In Vitro Reconstructed Human Epidermis (RHE) Test Methods For Skin Corrosion in Relation to OECD TG 431. Environmental Health and Safety Publications, Series on Testing and Assessment (No. 219). Organisation for Economic Cooperation and Development, Paris
25. OECD. (2005). Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. .Environment, Health and Safety Publications, Series on Testing and Assessment (No. 34.), Organisation for Economic Cooperation and Development, Paris.
26. Eskes C. et al. (2012). Regulatory Assessment of In Vitro Skin Corrosion and Irritation Data Within the European Framework: Workshop Recommendations. Regul.Toxicol.Pharmacol. 62, 393-403.
27. Mosmann T. (1983). Rapid Colorimetric Assay for Cellular Growth and Survival: Application to Proliferation and Cytotoxicity Assays. J. Immunol. Methods 65, 55-63.
28. Tinois E., et al. (1994). The Episkin Model: Successful Reconstruction of Human Epidermis In Vitro. In: In Vitro Skin Toxicology. Rougier A.,. Goldberg A.M and Maibach H.I. (Eds): 133-140.
29. Cannon C. L., Neal P.J., Southee J.A., Kubilus J. and Klausner M. (1994), New Epidermal Model for Dermal Irritancy Testing. Toxicol.in Vitro 8, 889 - 891.
30. Ponec M., Boelsma E, Weerheim A, Mulder A, Bouwstra J and Mommaas M. (2000). Lipid and Ultrastructural Characterization of Reconstructed Skin Models. Inter. J. Pharmaceu. 203, 211 - 225.
31. Tinois E., Tillier, J., Gaucherand, M., Dumas, H., Tardy, M. and Thivolet J. (1991). In Vitro and Post Transplantation Differentiation of Human Keratinocytes Grown on the Human Type IV Collagen Film of a Bilayered Dermal Substitute. Exp. Cell Res. 193: 310-319.
32. Parenteau N.L., Bilbo P, Nolte CJ, Mason VS and Rosenberg M. (1992). The Organotypic Culture of Human Skin Keratinocytes and Fibroblasts to Achieve Form and Function. Cytotech. 9, 163-171.
33. Wilkins L.M., Watson SR, Prosky SJ, Meunier SF and Parenteau N.L. (1994). Development of a Bilayered Living Skin Construct for Clinical Applications. Biotech.Bioeng.43/8, 747-756.

34. EpiSkin™ (December 2011)SOP,. INVITTOX Protocol (No. 118.). EpiSkin™ Skin Corrosivity Test.. Available at: [http://www.ecvam.jrc.ec.europa.eu.hm;].

35. EpiDremTM SOP. (February 2012). Version MK-24-007-0024 Protocol for : In vitro EpiDerm™ Skin Corrosion Test (EPI-200-SCT), for Use with MatTek Corporation's Reconstructed Human Epidermal Model EpiDerm. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

36. SkinEthic™ RHE SOP,INVITTOX Protocol (January 2012). SkinEthic™ Skin Corrosivity Test. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

37. EpiCS® SOP, Version 4.1 (January 2012). In Vitro Skin Corrosion: Human Skin Model Test Epidermal Skin Test 1000 (epiCS®) CellSystems. Available at: [http://www.ecvam.jrc.ec.europa.eu.html].

38. Alépée N., Barroso J., De Smedt A., De Wever B., Hibatallah J., Klaric M., Mewes K.R., Millet M., Pfannenbecker U., Tailhardat M., Templier M., and McNamee P. Use of HPLC/UPLCspectrophotometry for Detection of MTT Formazan in In Vitro Reconstructed Human Tissue (RhT)- based Test Methods Employing the MTT Assay to Expand their Applicability to Strongly Coloured Test Chemicals. Manuscript in Preparation.

39. US FDA. (2001). Guidance for Industry: Bioanalytical Method Validation. U.S. Department of Health and Human Services, Food and Drug Administration. (May 2001). Available at: [http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf].

40. Katoh M., Hamajima F., Ogasawara T., and Ken-ichiro Hata. (2010). Assessment of the human epidermal model LabCyte EPI-MODEL for in vitro skin corrosion testing according to the OECD test Guideline 431. J Toxicol. Sci. 35, 411-417.

41. LabCyte Validation Management Team (2018). Validation Study for in vitro skin corrosion test method using reconstructed human epidermal tissue LabCyte EPI-MODEL24. Available at: [http://www.jacvam.jp/files/doc/01_05/01_05_Z1.pdf].

42. LabCyte EPI-MODEL24 SCT SOP, Version 1.6. (May, 2017). Skin corrosion test using the reconstructed human model “LabCyte EPI-MODEL24”. Available at: [http://www.jacvam.jp/files/doc/01_05/01_05_Z2.pdf].

43. Report of the Peer-review of the validation study for LabCyte EPI-MODEL24 In Vitro Skin Corrosion Test Method. Available at [http://www.jacvam.jp/files/doc/01_05/01_05_Z3.pdf].

ANNEX 1- 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 (25).

Cell viability: Parameter measuring total activity of a cell population e.g. as ability of cellular mitochondrial dehydrogenases to reduce the vital dye MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide, Thiazolyl blue), which depending on the endpoint measured and the test design used, correlates with the total number and/or vitality of living cells.

Chemical: means a substance or a mixture.

Concordance: This is a measure of test method performance for test methods that give a categorical result, and is one aspect of relevance. The term is sometimes used interchangeably with accuracy, and is defined as the proportion of all chemicals tested that are correctly classified as positive or negative. Concordance is highly dependent on the prevalence of positives in the types of test chemical being examined (25).

ET50: Can be estimated by determination of the exposure time required to reduce cell viability by 50% upon application of the benchmark chemical at a specified, fixed concentration, see also IC50.

GHS (Globally Harmonized System of Classification and Labelling of Chemicals): A system proposing the classification of chemicals (substances and mixtures) according to standardized 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 (1).

HPLC: High Performance Liquid Chromatography.

IATA: Integrated Approach on Testing and Assessment.

IC50: Can be estimated by determination of the concentration at which a benchmark chemical reduces the viability of the tissues by 50% (IC50) after a fixed exposure time, see also ET50.

ET50. Infinite dose: Amount of test chemical applied to the epidermis exceeding the amount required to completely and uniformly cover the epidermis surface.

Mixture: means a mixture or solution composed of two or more substances in which they do not react.

Mono-constituent substance: A substance, defined by its quantitative composition, in which one main constituent is present to at least 80% (w/w).

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; Thiazolyl blue tetrazolium bromide.

Multi-constituent substance: A substance, defined by its quantitative composition, in which more than one main constituent is present in a concentration ≥ 10% (w/w) and <

80% (w/w). A multi-constituent substance is the result of a manufacturing process. The difference between mixture and multi-constituent substance is that a mixture is obtained by blending of two or more substances without chemical reaction. A multi-constituent substance is the result of a chemical reaction.

NC: Non corrosive.

NSCkilled control: Non-Specific Colour control in killed tissues.

NSCliving control : Non-Specific Colour control in living tissues.

NSMTT: Non-Specific MTT reduction.

OD: Optical Density

PC: 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.

Performance standards (PS): Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are; (i) essential test method components; (ii) a minimum list of Reference Chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (iii) the similar levels of reliability and accuracy, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of Reference Chemicals (25).

Relevance: Description of relationship of the test method to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test method correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (25).

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 (25).

Run: A run consists of one or more test chemicals tested concurrently with a negative control and with a PC.

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 (25).

Skin corrosion in vivo: The production of irreversible damage of the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test chemical for up to four hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars. Histopathology should be considered to evaluate questionable lesions.

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 (25).

Substance: means chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the

product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Test chemical: means what is being tested.

UPLC: Ultra-High Performance Liquid Chromatography.

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

ANNEX 2 - MAIN TEST METHOD COMPONENTS OF THE RhE TEST METHODS VALIDATED FOR SKIN CORROSION TESTING

Nr.	1	2	3	4	5
Test Method Component	**EpiSkin™**	**EpiDerm™ SCT**	**SkinEthic™ RHE**	**epiCS®**	**LabCyte EPI-MODEL24 SCT**
Model surface	0.38 cm^2	0.63 cm^2	0.5 cm^2	0.6 cm^2	0.3 cm^2
Number of tissue replicates	At least 2 per exposure time	2-3 per exposure time	At least 2 per exposure time	At least 2 per exposure time	At least 2 per exposure time
Treatment doses and application	Liquids and viscous: 50 ± 3 µL (131.6 µL/cm^2) Solids: 20± 2 mg (52.6 mg/cm^2) +100 µL±5µL NaCl solution (9 g/L) Waxy/sticky: 50 ± 2 mg (131.6 mg/cm^2) with a nylon mesh	Liquids: 50 µL (79.4 µL/cm^2) with or without a nylon mesh *Pre-test compatibility of test chemical with nylon mesh* Semi solids: 50 µL (79.4 µL/cm^2) Solids: 25 µL H_2O (or necessary) + 25 mg (39.7 mg/cm^2) Waxes: flat "disc like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15µL H_2O.	Liquids and viscous:40 ± 3 µL (80µL/cm^2) using nylonmesh *Pre-test compatibility of test chemical with nylon mesh* Solids: 20 µL ± 2µl H_2O + 20± 3 mg (40 mg/cm2) Waxy/sticky: 20 ± 3 mg (40 mg/cm^2) with a nylon mesh	Liquids and viscous:50 µL (83.3µL/cm^2) using nylonmesh *Pre-test compatibility of test chemical with nylon mesh* Semi solids: 50 µL (83.3 µL/cm^2) Solids: 25 mg (41.7 mg/cm^2) + 25 µL H_2O (or more if necessary) Waxy/sticky: flat "cookie like" piece of ca. 8 mm diameter placed atop the tissue wetted with 15µL H_2O	Liquids and viscous:50 µL (166.7µL/cm^2) Solids: 50± 2 mg (166.7 mg/cm^2) + 50 µL H_2O Waxy: Use a positive displacement pipette and tip as liquid and viscous substance.

Nr.	1	2	3	4	5
Test Method Component	**EpiSkin™**	**EpiDerm™ SCT**	**SkinEthic™ RHE**	**epiCS®**	**LabCyte EPI-MODEL24 SCT**
Pre-check for direct MTT reduction	50 µL (liquid) or 20 mg (solid) + 2 mL MTT 0.3 mg/mL solution for 180±5 min at 37°C, 5% CO_2, 95% RH ➔ if solution turns blue/purple, water-killed adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 1 mL MTT 1 mg/mL solution for 60 min at 37°C, 5% CO_2, 95% RH ➔ if solution turns blue/purple, freeze-killed adapted controls should be performed	40 µL (liquid) or 20 mg (solid) + 1 mL MTT 1 mg/mL solution for 180±15 min at 37°C, 5% CO_2, 95% RH ➔ if solution turns blue/purple, freeze-killed adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 1 mL MTT 1 mg/mL solution for 60 min at 37°C, 5% CO_2, 95% RH ➔ if solution turns blue/purple, freeze-killed adapted controls should be performed	50 µL (liquid) or 50 mg (solid) + 500 µL MTT 0.5 mg/mL solution for 60 min at 37°C, 5% CO_2, 95% RH ➔ if solution turns blue/purple, freeze-killed adapted controls should be performed
Pre-check for colour interference	10 µL (liquid) or 10 mg (solid) + 90µL H2O mixed for 15 min at RT ➔ if solution becomes coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H2O mixed for 60 min at 37oC, 5% CO2, 95% RH ➔ if solution becomes coloured, living adapted controls should be performed	40 µL (liquid) or 20 mg (solid) + 300 µL H2O mixed for 60 min at RT ➔ if solution becomes coloured, living adapted controls should be performed	50 µL (liquid) or 25 mg (solid) + 300 µL H2O mixed for 60 min at 37oC, 5% CO2, 95% RH ➔ if solution becomes coloured, living adapted controls should be performed	50 µL (liquid) or 50 mg (solid) + 500 µL H2O mixed for 60 min at 37oC, 5% CO2, 95% RH ➔ if solution becomes coloured, living adapted controls should be performed
Exposure time and temperature	3 min, 60 min (±5 min) and 240 min (±10 min) In ventilated cabinet Room Temperature (RT, 18-28oC)	3 min at RT, and 60 min at 37oC, 5% CO2, 95% RH	3 min at RT, and 60 min at 37oC, 5% CO2, 95% RH	3 min at RT, and 60 min at 37oC, 5% CO2, 95% RH	3 min at RT, and 60 min at 37oC, 5% CO2, 95% RH
Rinsing	25 mL 1x PBS (2 mL/throwing)	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS	20 times with a constant soft stream of 1x PBS	10 times or more with a constant strong stream of 1x PBS
Negative control	50 µL NaCl solution (9 g/L) Tested with every exposure time	50 µL H2O Tested with every exposure time	40 µL H2O Tested with every exposure time	50 µL H2O Tested with every exposure time	50 µL H2O Tested with every exposure time
Positive control	50 µL Glacial acetic acid Tested only for 4 hours	50 µL 8N KOH Tested with every exposure time	40 µL 8N KOH Tested only for 1 hour	50 µL 8N KOH Tested with every exposure time	50 µL 8N KOH Tested only for 1 hour
MTT solution	2 mL 0.3 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL	300 µL 1 mg/mL	500 µL 0.5 mg/mL
MTT incubation	180 min (±15 min) at 37oC, 5%	180 min at 37oC, 5% CO2, 95%	180 min (±15 min) at 37oC, 5%	180 min at 37oC, 5% CO2, 95%	180 min (±5 min) at 37oC, 5%

Nr.	1	2	3	4	5
Test Method Component	**EpiSkin™**	**EpiDerm™ SCT**	**SkinEthic™ RHE**	**epiCS®**	**LabCyte EPI-MODEL24 SCT**
time and temperature	CO2, 95% RH	RH	CO2, 95% RH	RH	CO2, 95% RH
Test Method Component	EpiSkin™ EIT	EpiDerm™ SCT	SkinEthic™ RHE EIT	epiCS®	LabCyte EPI-MODEL24 SCT
Extraction solvent	500 µL acidified isopropanol (0.04 N HCl in isopropanol) (isolated tissue fully immersed)	2 mL isopropanol (extraction from top and bottom of insert)	1.5 mL isopropanol (extraction from top and bottom of insert)	2 mL isopropanol (extraction from top and bottom of insert)	300 µL isopropanol (isolated tissue fully immersed)
Extraction time And temperature	Overnight at RT, protected from light	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight without shaking at RT or for 120 min with shaking (~120 rpm) at RT	Overnight at RT, protected from light
OD reading	570 nm (545 - 595 nm) without reference filter	570 nm (or 540 nm) without reference filter	570 nm (540 - 600 nm) without reference filter	540 - 570 nm without reference filter	570 nm with reference filter 650 nm
Tissue Quality Control	18 hours treatment with SDS 1.0mg/mL ≤ IC_{50} ≤ 3.0mg/mL	Treatment with 1% Triton X-100 4.08 hours ≤ ET_{50} ≤ 8.7 hours	Treatment with 1% Triton X-100 4.0 hours ≤ ET_{50} ≤ 10.0 hours	Treatment with 1% Triton X-100 2.0 hours ≤ ET_{50} ≤ 7.0 hours	18 hours treatment with SDS 1.4mg/mL ≤ IC_{50} ≤ 4.0 mg/mL
Acceptability Criteria	1. Mean OD of the tissue replicates treated with the negative control (NaCl) should be ≥ 0.6 and ≤ 1.5 for every exposure time 2. Mean viability of the tissue replicates exposed for 4 hours with the positive control (glacial acetic acid), expressed as % of the negative control, should be ≤ 20% 3. In the range 20-100% viability and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be ≤ 15% 3. In the range 20 - 100% viability, the Coefficient of Variation (CV) between tissue replicates should be ≤ 30%	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 3.0 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour (and 4 hours, if applicable) with the positive control (8N KOH), expressed as % of the negative control, should be ≤ 15% 3. In the range 20-100% viability and for ODs ≥ 0.3, difference of viability between the two	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.8 and ≤ 2.8 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be ≤ 15%. 3. In the range 20-100% viability and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not	1. Mean OD of the tissue replicates treated with the negative control (H_2O) should be ≥ 0.7 and ≤ 2.5 for every exposure time 2. Mean viability of the tissue replicates exposed for 1 hour with the positive control (8N KOH), expressed as % of the negative control, should be ≤ 15%. 3. In the range 20-100% viability and for ODs ≥ 0.3, difference of viability between the two tissue replicates should not

Nr.	1	2	3	4	5
Test Method Component	EpiSkin™	EpiDerm™ SCT	SkinEthic™ RHE	epiCS®	LabCyte EPI-MODEL24 SCT
	exceed 30%.		tissue replicates should not exceed 30%.	exceed 30%.	exceed 30%.

ANNEX 3 - PERFORMANCE OF TEST METHODS FOR SUB-CATEGORISATION

The table below provides the performances of the five test methods calculated based on a set of 79 or 80 chemicals tested by the five test developers. Calculations of four test methods (EpiSkin™, EpiDerm™ SCT, SkinEthic™ RHE and epiCS®) were performed by the OECD Secretariat, reviewed and agreed by an expert subgroup (21) (23). Calculation of LabCyte EPI-MODEL24 SCT was performed by the test developer, reviewed and agreed by the validation management group and a peer review panel (41) (43).

STATISTICS ON PREDICTIONS OBTAINED ON THE ENTIRE SET OF CHEMICALS (n= 80 chemicals tested over 2 independent runs for epiCS® or 3 independent runs for EpiDerm™ SCT, EpiSkin™ and SkinEthic™RHE *i.e.* respectively 159* or 240 classifications.
n= 79 chemicals tested over 3 independent runs for LabCyte EPI-MODEL24 SCT, *i.e.* 237 classification.)**
*one chemical was tested once in epiCS® because of no availability (23).
** one chemical was not tested in LabCyte EPI-MODEL24 SCT because of no availability.

	EpiSkin	**EpiDerm**	**SkinEthic**	**epiCS**	**LabCyte EPI-MODEL24**
Overclassifications:					
1B-and-1C overclassified 1A	21.5%	29.0%	31.2%	32.8%	30.0%
NC overclassified 1B-and-1C	20.7%	23.4%	27.0%	28.4%	18.9%
NC overclassified 1A	0.0%	2.7%	0.0%	0.0%	2.7%
Overclassified as Corrosive	20.7%	26.1%	27.0%	28.4%	21.6%
Global overclassification rate (all categories)	17.9%	23.3%	24.5%	25.8%	21.5%
Underclassifications:					
1A underclassified 1B-and-1C	16.7%	16.7%	16.7%	12.5%	13.9%
1A underclassified NC	0.0%	0.0%	0.0%	0.0%	0.0%
1B-and-1C underclassified NC	2.2%	0.0%	7.5%	6.6%	0.0%
Global underclassification rate (all categories)	3.3%	2.5%	5.4%	4.4%	2.1%
Correct Classifications:					
1A correctly classified	83.3%	83.3%	83.3%	87.5%	86.1%
1B-and-/1C correctly classified	76.3%	71.0%	61.3%	60.7%	70.0%
NC correctly classified	79.3%	73.9%	73.0%	71.62%	78.4%
Overall Accuracy	78.8%	74.2%	70.0%	69.8%	76.4%

ANNEX 4 - Key parameters and acceptance criteria for qualification of an HPLC/UPLC-spectrophotometry system for measurement of MTT formazan extracted from RhE tissues

Parameter	Protocol Derived from FDA Guidance (36)(38)	Acceptance Criteria
Selectivity	Analysis of isopropanol, living blank (isopropanol extract from living RhE tissues without any treatment), dead blank (isopropanol extract from killed RhE tissues without any treatment)	$Area_{interference}$ = 20% of $Area_{LLOQ}$[1]
Precision	Quality Controls (i.e., MTT formazan at 1.6 g/mL, 16 g/mL and 160 g/mL) in isopropanol (n=5)	CV = 15% or = 20% for the LLOQ
Accuracy	Quality Controls in isopropanol (n=5)	%Dev = 15% or = 20% for LLOQ
Matrix Effect	Quality Controls in living blank (n=5)	85% = %Matrix Effect= 115%
Carryover	Analysis of isopropanol after an ULOQ[2] standard	$Area_{interference}$ = 20% of $Area_{LLOQ}$
Reproducibility (intra-day)	3 independent calibration curves (based on 6 consecutive 1/3 dilutions of MTT formazan in isopropanol starting at ULOQ, i.e., 200 g/mL); Quality Controls in isopropanol (n=5)	Calibration Curves:%Dev = 15% or = 20% for LLOQ Quality Controls:
Reproducibility (inter-day)	Day 1: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 2: 1 calibration curve and Quality Controls in isopropanol (n=3) Day 3: 1 calibration curve and Quality Controls in isopropanol (n=3)	%Dev= 15% and CV = 15%
Short Term Stability of MTT Formazan in RhE Tissue Extract	Quality Controls in living blank (n='3)' analysed the day of the preparation and after 24 hours of storage at room temperature	%Dev = 15%
Long Term Stability of MTT Formazan in RhE Tissue Extract, if required	Quality Controls in living blank (n='3)' analysed theday of the preparation and after several days of storageat a specified temperature (e.g., 4ºC, -20ºC, -80ºC)	%Dev = 15%

Note:
[1]LLOQ: Lower Limit of Quantification, defined to cover 1-2% tissue viability, i.e., 0.8 µg/mL.
[2]ULOQ: Upper Limit of Quantification, defined to be at least two times higher than the highest expected MTT formazan concentration in isopropanol extracts from negative controls i.e., 200 µg/mL.

添付資料5

IL-2 Luciferase (IL-2 Luc) Assay Report of the Peer Review Panel

on

a JaCVAM co-ordinated study programme addressing the validation status of the IL-2 Luc assay for the prospective identification of immunotoxicic substances on T-cells

Report completed by the Peer review Panel on May XX, 2020

Table of Contents

Executive Summary

The IL-2 Luc assay has been proposed as an in vitro alternative, providing information on in the adverse outcome pathway (AOP) for immunotoxicity, especially on T-Cells. This assay identifies the effects of chemicals on the IL-2 luciferase (Luc) activity.

The peer review panel (PRP) found the Validation Management Team's report presented the necessary information for an independent review.

Consequently, the PRP were able to conclude that the IL-2 Luc assay was well defined, with a clear protocol and criteria for data interpretation. Both within and between laboratory reproducibility information were satisfactory. On the other, the predictive capacity was not satisfactory as a stand alone method. All necessary information including performance standards were detailed.

The PRP should stress this assay is that it is not intended to be used as a sole indicator of immunotoxicity and the reliability of the criteria about the immunotoxicological chemicals is to be discussed further.

Accordingly, the PRP concluded that the IL-2 Luc assay validation has demonstrated that the method would be acceptable as part of an integrated testing strategy for the predictive screening of T-cell targeted immunotoxicity.

Peer Review Panel Composition

Fujio Kayama (chair)	Jichi Medical University, Japan
Henk van Loveren	Maastricht University, Netherland
Haley LaNef Ford	Seattle Genetics, Inc., USA
Barbara Kaplan	Mississippi State University, USA
Xingchao Geng	National Center for Safety Evaluation of Drugs (NCSED), China
Takao Ashikaga (Vice chair)	JaCVAM, Kawasaki, Japan
Sang-Hyun Kim	Kyungpook National University, Korea

Background

Immune dysregulation can have serious adverse health consequences and it could be caused by many types of chemicals, such as environmental contaminants, food additives, and drugs. It ranges from reduced resistance to infection and neoplasia to allergic and autoimmune conditions. For many years, to identify such immunotoxici chemicals depends on animal models. For animal welfare, ethical and scientific reasons there has been a desire to replace in vivo methods with non-animal alternatives (1). The immune system comprises innate and adaptive immunity. Both arms of the immune response function differently and are driven by different population of cells. A variety of intracellular signaling pathways also play roles in innate and adaptive immune responses. Given the complexity of the immune system, it is unlikely that a single in vitro method will cover all immunotoxicants. Therefore, Integrated Approaches to Testing and Assessment (IATA) has been taken for this field. For example, target compounds are tested using various approaches such as the human whole blood cytokine release assay (HWBCRA), lymphocyte proliferation assay, mixed lymphocyte reaction and fluorescent cell chip assay after the evaluation of myelotoxicity (2). There are many in vitro methods aiming at evaluating various aspects of molecular and cellular events in the adverse outcome pathway (AOP) for immunotoxicity. However, no in vitro method is validated formally.

The IL-2 Luc assay was developed as a part of high-throughput screening system which enables to evaluate chemical immunotoxicity. This screening system was named Multi-ImmunoTox assay or MITA (3) (4) (5). The IL-2 Luc assay, using human cell line transfected with luciferase genes under control of the IL-2 promoter, identifies the effect of chemicals on the IL-2 activity in the 2H4 cells in the presence of stimulants (3).

The PRP was assembled and met in February 2019 to review a progress report on the IL-2 Luc assay prepared by the Validation Management Team (VMT). Following the commentary on this work by the PRP, the VMT refined the validation report.

The PRP engaged in follow-up telephone conferences in October and December 2019. With the provision of all of the amended, updated and additional material, including the final VMT report, this PRP Validation Report was prepared.

IL-2 Luc Test Method Definition

The PRP confirmed that the IL-2 Luc assay test method has been fully described in the report of the Validation Management Team (VMT) and in the associated detailed test protocol. During the validation study, the test developer changed their prediction model. A clear definition of the 35% threshold and its reason was explained. The VMT report describes the need for the assay in the current regulatory context (6). Furthermore, a clear rationale for the assay has been given (the rationale for the test method is that drugs and chemicals, environmental contaminants, food additives, and drugs

can target the immune system, resulting in immune dysregulation). It is known that IL-2 exerts pleiotropic actions on CD4+ T cell differentiation via its modulation of cytokine receptor expression. It promotes Th1 differentiation by inducing IL-12Rb2 (and IL-12Rb1), promotes Th2 differentiation by inducing IL-4Ra, inhibits Th17 differentiation by inhibiting gp130 (and IL-6Ra), and drives Treg differentiation by inducing IL-2Ra. IL-2 also potently represses IL-7Ra, which decreases survival signals that normally promote cell survival and memory cell development (7). Therefore, it is reasonable that the test developer focused on the regulation of IL-2 transcription and attempted to construct an AOP of immunotoxicity with transcriptional dysregulation of IL-2 as a central key event. The VMT report mentioned that IL-2 Luc assay as part of development of a broader tier approach to eventually include IL-8 and IL-1β for corresponding to the AOP.

The PRP agreed that the mechanistic basis of the method and how it related to the T-cell specific endpoint also was well described in the VMT report.

Within Laboratory Reproducibility

The PRP agreed that the results which emerged have demonstrated a sufficient degree of within laboratory reproducibility. For achieving such conclusion, the PRP focused on results obtained with the final protocol and prediction model.

A total of 5 coded chemicals (4 T-cell targeting and 1 non T-cell targeting) were evaluated by 3 experimental sets. Based on such assumptions, the success criterion of >80% within laboratory reproducibility was achieved in each of the three participating laboratories (Lab. A: 80.0% (4/5), Lab. B: 100% (5/5), Lab. C: 80.0% (4/5)).

The PRP notes that extensive documentation of within laboratory reproducibility data for the final and all the development phases of the IL-2 Luc assay has been displayed in the VMT report and its appendices. Taken into account, the data lends support to the view that the assay has a sufficient level of reproducibility within laboratories.

Interlaboratory Transferability

The PRP noted that the technical transfer of the IL-2 Luc assay involved training and successful assessment of 3 experiments of 5 test substances (not blinded) by each of the participating laboratories. That work was prior to their approval to participate in the subsequent validation work.

Between Laboratory Reproducibility

With regard to between-laboratory reproducibility, the PRP recognized that the test results gave 80% (20/25) and met the success criterion of >80% between laboratory reproducibility. The number of test chemicals is combined of the Phase I (5) and Phase II (20) studies.

Again, the PRP notes that extensive and transparent documentation of between laboratory reproducibility data for all phases of the IL-2 Luc assay has been displayed in the VMT report and its appendices.

The PRP concluded that the assay demonstrated successful between laboratory reproducibility.

Predictive Capacity

To determine the predictivity of the IL-2 Luc assay, it is crucial to understand the immunotoxic characteristics of chemicals used in the study. The PRP agreed that classification chemicals into those that affect T cell function, i.e., T cell-targeting chemical (TTC) and those that do not directly affect T cell function, i.e., non-T cell-targeting chemicals (NTTC) . The PRP confirmed the rationale for classifying immunotoxic chemicals are clearly discribed in the VMT report.

Demonstration of a test method’s performance should be based on the testing of representative, preferably coded, reference chemicals. The PRP concluded that the validation study used an appropriate level of test chemical coding to ensure fully blinded evaluation. With respect to chemical selection, the PRP confirmed that the criteria for chemical selection were clearly outlined. On the other, it should be noted that there is a question if the number of true negatives (8/25) in the set was big enough or not.

The immunotoxic characteristics of each chemical used in the Phase I and Phase II studies are shown in the VMT report and based on the criteria total 25 chemicals were classified into 16 positives, 8 negatives and 1 unclassified. According to the classification, accuracy is 75% (18/24), specificity is 75% (6/8) and sensitivity is 75% (12/16). The PRP concluded the predictive capacity of the test was not sufficient, if one would see this test as a stand-alone to detect immunotoxicity.

The PRP basically agreed with the test developer's opinion that there are at least 2 reasons for this poor predictivity. First, the reliability of the criteria about the immunotoxicological chemicals is sometimes uncertain, because the information available was very limited. Second, the IL-2 Luc assay does not cover every aspect of the effects of chemicals on T cell function.

Regarding the second point, the PRP noted again this assay should be used in the context of IATA (that means combination with other assays targeting T-cell functions should be mandatory). With respect to the first point, the PRP noted the use of compounds with clear immunotoxicity mechanism will help to ultimately improve the accuracy of the validation study and, also recommends a Detailed Review Paper (DRP) on the immunotoxicity should stress the importance of reliability of in vivo data .

Following phases I and II, the assay was then applied to over 60 chemicals that were previously evaluated by the test developer (8). The accuracy was calculated as 76%, which is in line with results obtained in phases I and II. As a result, the PRP concluded the predictive capacity of this assay is reasonable if at least the two issues pointed above would be considered.

Applicability Domain

The PRP shared the applicability domain of this assay communication with VMT. Especially because of the use of a cell line, the method can't detect immunotoxicity associated with the inhibition of DNA synthesis and cell division. Another limitation is that the assay might not detect compounds that require metabolic activation to a toxic intermediate. In addition, the use of PMA/Io as stimulants bypasses signaling through the T cell receptor and therefore this stimulation could affect the results. Inevitably, the IL-2 Luc assay shares limitations common to many suspension cell-based techniques, not least in dealing with highly hydrophobic substances.

Performance Standards

The PRP was of the view that the list of performance standard (PS) substances placed in the appendix 15 to the VMT report was satisfactory. Mutual Acceptance of Data (MAD) will only be guaranteed for test methods validated according to the PS, if these test methods have been reviewed and included in this Test Guideline by the OECD. The PS is supplemented by a list of proficiency chemicals, listed in the appendix 14, to be used as a routine check on performance of the assay.

Additional Comments

The PRP concluded that the validation study management and conduct met the criteria set out in OECD GD 34 (2005). The PRP concluded also that the study was conducted not under GLP certification but in the spirit of GLP.

The PRP appreciated the transparency with which all the IL-2 Luc assay material was presented. The PRP notes that during the conduct of the review, it was possible access to the full raw data files associated with the IL-2 Luc assay development/validation work.

The PRP also noted that AOP networks and DRP in this field must be essential in order to construct the IATA.

Conclusions and Recommendations

The PRP concluded that the IL-2 Luc assay validation has demonstrated that the method should be acceptable as part of IATA for the predictive screening of T-cell targeted immunotoxicity.

Acknowledgements

The PRP is grateful to the members of the VMT for their hard work and patience and to JaCVAM for their support in setting up and hosting meetings in Japan, as well as for the setting up of several telephone conferences.

References

1) Adler et al. (2010) Alternative (non-animal) methods for cosmetics testing: current status and future prospects. Arch Toxicol. 2011 85(5):367-485
2) Corsini E and Roggen EL., 2017. Overview of in vitro assessment of immunotoxicity. Curr Opin Toxicol. 5: 13-18.
3) Saito, R., Hirakawa, S., Ohara, H., et al., 2011. Nickel differentially regulates NFAT and NF-kappaB activation in T cell signaling. Toxicology and applied pharmacology, 254, 245-255.
4) Takahashi, T., Kimura, Y., Saito, R., et al., 2011. An in vitro test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. Toxicological sciences: an official journal of the Society of Toxicology 124, 359-369.
5) Kimura, Y., Fujimura, C., Ito, Y., et al., 2014. Evaluation of the Multi-ImmunoTox Assay composed of 3 human cytokine reporter cells by examining immunological effects of drugs. Toxicol In Vitro 28, 759-768.6) (WHO)/, W.H.O., Meeting, I.P.o.C.S.I.S., 2012. Guidance for immunotoxicity risk assessment for chemicals. WHO press.
7) Liao, W., Lin, J.X., Wang, L., et al., 2011, Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. Nat Immunol. 12: 551-559.
8) Kimura, Y., Fujimura, C., Ito, Y., et al., 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, 2043-2054.

Teleconference for IL-2 PRP

October 1, 2019

Peer Review Panel: Henk van Loveren, Haley Neff-LaFord, Barbara Kaplan, Fujio Kayama, Takao Ashikaga
VMT: Hajime Kojima
Observers: Steve Venti (meeting minutes)

Kojima:	In this meeting, we will discuss the revised validation report and the schedule going forward. I will explain the changes in the report, which are shown in red. One important point is Appendix 7. It has 290 pages and discusses the data available on immunotoxic effects of chemicals. Mainly, the figures for predictivity and the summary were revised. I heard Dr. Aiba is on-going to revise minorly. After the meeting, I ill share the newest Validation report.
Kaplan:	This summary is in line with what we discussed at the FTF meeting.
Kojima:	Does everyone accept this summary?
Everyone:	Yes.
Kojima:	Section 9-1-3 addresses predictivity and describes the effects of chemicals on T-cells. And there is a definition of T-cell targeting chemicals (TTCs).
Kaplan:	Criterion 3 says “#2 or #3 on two or more cytokines.” Does that refer only to the three cytokines mentioned in #2 and #3? For example, is IL-17 excluded? This is not clear. If there is a report for other cytokines, would they be considered TTCs?
Kojima:	I can’t answer at the moment, but I will ask Dr. Aiba.
Kaplan:	This is an improvement over the original report. Once we have some clarification on Criterion 3, I think that these criteria are acceptable.
van Loveren:	Although I think it would be good to extend this to other cytokines, not just the ones listed.
Kojima:	(Brief review of other changes in red. Please see revised Validation Study Report.) If you are happy with this report, then we can move on to reviewing the PRP Evaluation Criteria and creating the PRP report.
Kaplan:	Do we need to read this and provide comments? What do you need from the PRP to submit to the OECD?
Kojima:	If you feel that the Validation Study Report satisfies the 14 PRP Evaluation Criteria, then you can prepare a Peer Review Report of about 12 pages with a comment about each criterion. And then the Validation Study Report and the Peer Review Report will be reviewed by an OECD expert working group.
van Loveren:	Are there specific places we should comment on?
Kojima:	We revised the Validation Report based on the comments from the PRP.
Kaplan:	So we have already covered the critical issues. But if there is anything specific you want us to look at, please tell us now.
van Loveren:	Is there any issue we need to address now?
Kojima:	I will share these documents with you, and after we have your comments, Dr. Kayama will write the final PRP report.
Neff-LaFord:	Once you see the documents, it is pretty easy to follow what has been changed, so we should be able to follow it.
Kojima:	The deadline for comments if possible, would be by the end of October and then we can have another teleconference in early or mid-November. OK, I will send you meeting minutes, the newest validation Study Report, and the evaluation

	criteria.

Teleconference for IL-2 PRP

November 11, 2019

Peer Review Panel: Henk van Loveren, Barbara Kaplan, Haley Neff-LaFord, Fujio Kayama, Takao Ashikaga, Lin Shi, Xingchao Geng
VMT: Hajime Kojima, Setsuya Aiba, Takuya Kimura
Observers: Steve Venti (meeting minutes)

Kojima:	In this meeting, we will discuss the revised validation report prior to discuss the peer review items. We revised the report based on your comments. After the previous teleconference, we received it in accordance with the comments from Barbara, and you have some other comments that have not been reflected yet, so I think we need to discuss this report more.
Kaplan:	I think these revisions are fine as long as things are separated into a table on criteria and clearly intelligible.
Aiba:	I don't know who made this table, but it presents what I wanted to say, so I think we can use this if the PRP agrees.
Kojima:	Dr. Aiba will calculate predictive capacity based on this table, so the most important thing is that the PRP finds this table acceptable.
Kayama:	I think these criteria are easier to understand as presented in the table.
van Loveren:	I am still concerned that the introduction is confusing to a naïve reader. We of the PRP understand that MITA is the context, *not* the aim, of this study. But the introduction needs a clear statement at the start of the introduction that the aim of this validation study is the IL-2, not MITA in general. Mentioning MITA in the introduction is fine, but you cannot have MITA at the start of the introduction. The introduction must begin with the aim of the study, which is IL-2.
Kaplan:	The first time I read this introduction, I thought that you were validating the entire MITA, but later I realized that is not the case. The goal is to validate the IL-2 assay. I agree with Henk and Haley that the goal of the validation needs to be stated clearly at the start of the introduction. Even just one sentence is enough. Just clearly state that the goal is to validate the IL-2 assay.
Neff-LaFord:	Yes, just more section 3 up higher.
van Loveren:	We need to say "proposed AOP" because this AOP has not yet been accepted.
Neff-LaFord:	The expression "IL-2 LA" appears to mean the same thing as "IL-2 Luc Assay." If IL-2 LA is intended to mean something different, then this needs to be spelled out more clearly.
Aiba:	Yes, I will clarify that.
van Loveren:	On page seven in introduction, I have suggested a revision, but perhaps the information about the applicability range that I deleted needs to be added back.
Kaplan:	I think that in context, the meaning of "applicability domain" is clear enough to be left in. But the word "however" should be removed for clarity.
van Loveren:	The applicability domain is discussed in the preceding paragraph, so maybe we can just use Haley's suggestion as is.
Kojima:	In section 9-5, I will inform you the detailed records collected in the principles of GLP.
Neff-LaFord:	I don't understand what "almost comparable" means in section 10-3-1.
Kaplan:	Given the emphasis on comparing IL-2 results with the results of other tests, I think that this section needs to be expressed more clearly. I think this information is important, which is why it should be described more clearly.
Ashikaga:	I couldn't find any description about regulatory application in the report.

Aiba:	Do I need to respond to each of these comments one by one?
Ashikaga:	Why is SFO-luciferase activity measured in this assay?
Aiba:	It is automatically measured but it is not necessary for this assay.
Kaplan:	This is related to what we were talking about before. This report contains a lot of information that is only incidentally related to IL-2, which confuses the reader.
Ashikaga:	I could not find a list of proficiency chemicals. Shouldn't the developer submit a list?
Aiba:	Yes. Appendix 14 and 15 have a list of proficiency chemicals.
Kojima:	Are there any other comments?
Xingchao:	I agree with the comments and I think the report is improved.
Lin:	(inaudible)
Aiba:	(inaudible)
van Loveren:	The applicability domain does not seem to be defined anywhere. Where is the definition of the applicability domain? All the information is there, but there is no single clear definition. You could rename 10-6 and start with a simple explanation of the applicability domain.
Kaplan:	This is a good point. We have defined a T-cell target, so we need to say that is what the applicability domain.
Aiba:	OK, I will provide a clear definition of what the applicability domain is.
Kojima:	I will share the minutes of this meeting, and then Dr. Aiba and the VMT will revise the validation report to share with the PRP. Perhaps you can then submit your comments for Dr. Kayama within one month and to be created the PRP report by Dr. Kayama.
Kayama:	The most important comment today is Henk's last comment.
Aiba:	I'd like to ask Dr. Kayama to summarize the PRP comments, because I already answered the original comments. I would like to know what I should respond to.
Kayama:	Will the PRP report be incorporated into the validation report or separately attached?
Kojima:	Separately attached.

Teleconference for DRP on in vitro immunotoxicity

October 28, 2019

Emanuela Corsini, Erwin Roggen, Dori Germolec, Henk van Loveren, Barbara Kaplan, Setsuya Aiba, Yutaka Kimura, Takayuki Yoshimoto, Hajime Kojima, Steve Venti (meeting minutes)

Kojima:	Today's agenda is as follows: Opening comments What is DRP in the OECD TG programme? Japanese SPSF adopted at WNT Supporters Discussion of Table of Contents We will allocate responsibilities for the table of contents. I would like to follow Emanuela's suggestion. Do you all agree with this?
Kaplan:	Yes, I will do III.
Germolec:	I can do IV.
Kojima:	I talked with Dr. Aiba and he will do X.
Loveren:	I am fine with II, although I will need some guidance about what to do.
Kojima:	I will do IV, but I have not set I, VI, VII, VIII, or XI. So do we have any ideas for people who could do these sections. The deadline is the end of this year, and then I will share everything by mid-January to prepare for our meeting at the end of January.
Corsini:	Can you explain why these two documents you sent us will help us?
Kojima:	Here is the draft guideline for detection of reproductive and developmental toxicity for human pharmaceuticals. This contains a list of reference compounds based on MOA. I think we need to discuss and select which chemicals we want to discuss.
Corsini:	So you feel these are relevant for immunotoxicity?
Kojima:	We will have to recommend positive and negative compounds for immunotoxicity, so this is a good example for us to follow. But in addition to pharmaceuticals, we have to also address industrial and other chemicals.
Corsini:	So we can use this as a possible template for developing our approach to immunotoxicity.
Aiba:	It's very difficult to determine immunotoxicity just by in vitro tests, because they do not correlate strongly with in vivo test results. So how can we approach this issue? Even chemicals that are classified as immunotoxic do not necessarily increase susceptibility to infection.
Germolec:	I think the goal should be to achieve similar results to in vivo test results for compounds that we know are immunotoxic.
Kojima:	The goal of this document is to encourage the development of in vitro assays for immunotoxicity.
Corsini:	There is no single test that can predict immunotoxicity. So we must always correlate the results of several tests. And the definition will be very broad, because it will encompass any change in functionality of the immune system.
Kojima:	This paper on the Curated Database of Rodent Uterotropic Bioactivity is also an example of an approach we can take.
Corsini:	Yes, we must think of a way to transpose this approach to fit immuotox purposes, but if I understand correctly, these documents are examples that we can follow in preparing our DRP.

Kojima:	Yes.
Germolec:	The information in the literature is rather spotty for trying to see if a chemical was targeting T-cells. So it is difficult to classify what the immunotoxicity would be in vivo. We can try to develop a table like this, but I am not sure it will be useful for all chemicals.
Kaplan:	A lot of work has already been done, so I hope we don't try to reinvent the wheel. We should try to build on what has already been done.
Corsini:	There are many compounds for which we don't have a clear picture of whether they actually target T-cells or not.
Kojima:	I feel that VIII and IX are rather related, so perhaps Dori could make a draft of both these sections.
Germolec:	Yes, I can do VIII and IX. Maybe I will come back to Drs. Kojima and Aiba to ask about how chemicals were selected for the first round of the validation.
Kojima:	Yes, in XI we will be recommending Dr. Aiba's assays. So the validation reports describes much information that can be used as a reference.
Germolec:	Before I send my draft to anyone else, I will send it to Dr. Aiba together with any questions I might have, so that he can confirm that I have covered everything that needs to be covered.
Corsini:	The draft of the introduction can be very brief, because we really all are working independently until we all share our drafts and begin to share comments. So if Hajime can circulate these documents to us then we can all be on the same page about what the introduction should say.
Kaplan:	I agree. Once we see sections I to X, then we will know how to introduce the subject.
Corsini:	Yes, it will be easier to do this at the face to face meeting. So we will submit our drafts by December?
Kojima:	Yes, and then I will distribute them by mid-January, so we can discuss them face to face at the end of January. I will submit to the OECD this May, and if it is approved, then we can discuss our future schedule at the Expert meeting in OECD.

F2F Meeting for the OECD DRP on in vitro immunotoxicity

January 28 & 29, 2020

H. Loveren, B. Kaplan, H. Neff-LaFord, T. Yoshimoto, E. Corsini, D. Germolec, S. Aiba, Y. Kimura, H. Kojima, S. Venti

	January 28
Kojima:	(OECD follow-up activities) SPSF was submitted a year and half ago and was approved in Feb. 2019. Japan is to coordinate creation of a draft DRP. The OECD Secretariat wants to coordinate an expert workshop during 2020 but after the DRP is available.
Corsini:	Does this mean that we have more time than just the end of March to complete the DRP?
Kojima:	We need to identify our action items by the end of this meeting, but we are all busy so maybe the end of March is not possible. Perhaps by the beginning of summer break is a good target.
Corsini:	We have most of the contributions, but we still lack the performance factors. So, we need to ask Erwin to provide this.
Kaplan:	We need to define a difference between IL2-Luc and IL2-LA and when to use.
Germolec:	Is it fair to expect Erwin to contribute? I think his focus is on his family, so we need a contingency in case he cannot write the section on performance factors.
Corsini:	What should be described in Chapter VI?
Germolec:	And how does Chapter VI differ from Chapter VII?
Kojima:	Chapter VII is assay qualification information, which refers to within- and between-laboratory replicability. But the performance factors of Chapter VI refer to test validation criteria.
Aiba:	One issue is that we often don't have a consensus about whether or not a particular chemical is immunotoxic. To create an immunotoxicity test, we need some kind of definition of what is an immunotoxic chemical.
Corsini:	If a chemical presents the possibility of immuno-augmentation or suppression, then we should consider it immunotoxic.
Germolec:	I think that we have pretty strong evidence for many chemicals that they are immunotoxic, so I cannot agree with Dr. Aiba.
Aiba:	This handout is one example of the kind of data we use to determine if a chemical is immunotoxic.
Germolec:	The WHO and the EPA have issued documents that define what makes a compound immunotoxic.
Kaplan:	Are these performance factors specific to IL-2 or are they for all in vitro assays? For all in vitro assays.

Corsini:	We should use the WHO/EPA criteria for classifying compounds immunotoxic. We will adapt the text from the ICHS5 for sections VI and VII.
Loveren:	My understanding is that we need to describe these things at a very general level. VI about is how to arrive at a valid test result, VII is about how to validate the test method itself.
Kaplan:	Do we also need to address issues related to applicability domain? For example, chemicals that require metabolism cannot be identified by an in vitro test.
Corsini:	An in vitro test method should be able to distinguish whether a chemical positive or negative.
Germolec:	It is true that we do not have as extensive a database of chemicals that affect the immune system as we do for chemicals that are sensitizers.
Corsini:	Not all chemicals need to be tested for immunotoxicity, which is another reason that we don't have as much data as for sensitizers.
Germolec:	There are many chemicals I left off my list of immunotoxic chemicals, because I focused on chemicals that would be useful for developing an in vitro test method. Maybe this document is too focused and needs to be made more general.
Corsini:	I received a suggestion to move the background to the start, so that it comes before the introduction.
Loveren:	Yes, and since there is a lot of overlap, I can make the introduction a bit shorter.
Germolec:	Perhaps after having a general explanation of in vitro assays or immunotoxicity, we should have a section at the end of the document using the MITA as an example of how a battery of tests for immunotoxicity is a good approach for in vitro assays.
Aiba:	I don't mind if you delete my introduction and just have your background, because they do describe similar things.
Loveren:	But we do need both sections, don't we? I'm still not 100% sure of how to revise my section. How broad is this? Apparently, we want to be very broad.
Kaplan:	What we can do is provide a broad introduction but then say that we will present the MITA as a possible implementation at the end of the document.
Germolec:	There will be some duplication across sections that we will either have to eliminate or make sure is consistent.
Neff-LaFord:	We are all working on our own sections, but could we share our revisions in real time so that we can each see what has changed?
Kojima:	Maybe we can use the JaCVAM website to share the latest version of the document. I can give everyone a username and password.
Corsini:	What was the source of the text in Chapter IV?
Kojima:	From the AOP wiki
Kaplan:	The OECD project numbers need to be coordinated with AOP numbers so that we know how they relate to the numbers that the OECD secretariat uses. The figures that shows the

	structure of the AOP is much more informative than the current text. Also, the references to “Pending” or “under development” are difficult to understand.
Loveren:	Do we have diagrams for all these AOP?
Kojima:	No, some are under development and there is no diagram.
Germolec:	We should provide links to information that comes from the AOP wiki.
Loveren:	Why is there information on the sensitization AOP? Is it relevant to our DRP? Doesn’t it create confusion to talk about sensitization in the middle of a DRP on immunotoxicity?
Corsini:	Maybe we can just refer to it rather than providing so much information. Immunotoxicity does encompasses skin sensitization.
Kaplan:	Since this is the only AOP that is currently approved by the OECD, perhaps this is a good example of the state of the art.
Germolec:	It might be thought of the type of event that might be found in an immune response AOP. Hypersensitivity is one of the clusters that we refer to when assessing immunotoxicity, so it might be relevant in that sense.
Aiba:	The IL1 AOP is now under EAGMAST review.
Germolec:	The IL1 AOP is a good example of what an immunosuppression AOP should look like, so I would like to see a diagram of it included.
Corsini:	Let’s use the same layout as in Dr. Kimura’s diagram, so that we know the OECD Project No. and status, and include a diagram if available. The AOP chapter needs to be linked with MITA as an example. The next section on the state of the art of in vitro or nonanimal testing, there is still much that needs to be added, which I will do.
Loveren:	But the information on predictive capacity can also be misleading. Because it depends on what chemicals were chosen.
Germolec:	But I think it is valuable to expand this section to discuss these and various other considerations.
Aiba:	There are few drugs that cause immunosuppression by affecting B-cell function.
Neff-LaFord:	There are pharmaceuticals but they are very targeted. There are few environmental chemicals that do that, however.
Aiba:	Our goal is to develop an immunotoxicity test that does not need primary animal cells.
Corsini:	The message that we want to put here is that there are systems that we want to develop into an in vitro test that incorporates T-cell dependent antibody development.
Kaplan:	Should we mention the IL-2 Luc at the end of this chapter or wait until the end? Maybe we can mention that we will talk about the IL-2 Luc in detail at the end of the paper. What are the two assays you mentioned?
Germolec:	The human whole blood and IL-2 Luc assays.

Kaplan:	These two in vitro TDAR assays and the antigen presentation assay are the wave of the future in that they are moving away from using animals, so maybe we can expand on that a little.
Aiba:	Are there any studies that examine cytotoxicity in chemicals rather than drugs?
Corsini:	There are some, but many examine drugs.
Aiba:	The IL-2 test needs to be combined with a test for cytotoxicity. So, what I want to know is if there is a study that shows immunosuppression by suppressing T-cell proliferation.
Kaplan:	Can this discussion of cytotoxicity be incorporated into the performance factors of this paper? I think that this is critical to determining if something is immunotoxic.
Kojima:	This ICHS5 document provides information on what is meant by performance factors and qualification information. So, we can use this document as a reference.
Corsini:	You are suggesting that we adopt this information for our paper? Maybe we can have a flow chart to propose an integrated testing strategy.
Aiba:	We should comment that the IL1-Luc test covers only a narrow range of immunotoxicity, so it cannot achieve a predictivity as high as 80%. We need multiples tests to do that.
Corsini:	Why did you include this sentence saying that "if the proposed test method is to be used in a single laboratory only, multi-laboratory testing will not be required for validation?"
Kojima:	There is need to promote the development of new test methods, so this was included to make it easier for companies to develop in-house assays.
Corsini:	But then the assay cannot be used in a regulatory context, so there is no need for the OECD to review it.
Kojima:	This comes from TG 439 for the development of me-too assays.
Aiba:	The context is different. Our DRP is about the development of a new test method, not a derivation of an already validated one.
Loveren:	Well, we could include that information and keep this sentence.
Corsini:	I think it would be better to delete this sentence.
Kojima:	My idea is to test at least 10 chemicals and achieve a within-laboratory reproducibility of 80%.
Aiba:	My experience is that even just 5 chemicals is enough. 10 chemicals require 90 runs.
Corsini:	It is a statistical issue. It could be 5, it could be 10. Let's leave it at ten. We will get comments from member countries.
Aiba:	Let's make it 5 and get comments.
Corsini:	Let's say "at least 5" but leave it to the agencies and statisticians to create a consensus.
Kojima:	And predictive capacity?
Corsini:	We need to add a phrase to provide context, predictivity might be lower than 80% for a particular assay, but within the context of using multiple assays, that will be acceptable.

Germolec & Loveren:	At the end of this paragraph, let's add this: "It is understood that due to the complexity of the immune system, a highly specific individual test might not achieve the desired levels of specificity, sensitivity, and accuracy, because of the different modes of action of the test chemicals. This can be overcome with a battery of tests, but the battery of tests must achieve the specified predictive capacity.
Corsini:	We also need to provide an explanation as to why some chemicals are excluded from the applicability domain.
Germolec:	We don't have data yet to support an assertion that using a battery of tests will improve predictive capacity. That is our hope, and theoretically we expect this to happen, but we don't have the data to assert this. Perhaps we have to say "We expect that increasing the number of tests will increase predictivity."
Kaplan:	But can we really say that yet?
Germolec:	Yes, it is ok for only one test to give positive results because they all test different things.
Kaplan:	Which means we cannot classify compounds as negative. Only as positive.
Germolec:	The Luster data shows that you need multiple in vivo tests, so based on that we can say that you would need multiple in vitro tests.
Germolec:	I will expand our list of chemicals by adding the ones that I consider relevant and will have to add some negatives, too.
Corsini:	Please add the CAS No. in the table, since this will be a reference.
Germolec:	And I can add some reference works as a footnote to the table.
Loveren:	Do we need to provide chemical selection criteria? Or does it not make so much difference? But anyway, we do need to include negative chemicals.
Germolec:	Please send me any suggestions you have over the next couple of weeks, and I will circulate the new list when it is ready.
Aiba:	Should we suggest that in vitro tests not be performed at concentrations beyond normal human exposure?
Corsini:	We should comment that if an effect is found at a very high concentration, it needs to correlate with in vivo exposure to be considered relevant.

	January 29
Corsini:	Do we want to move some of the discussion in IX to State of the Art and focus here on MITA and other tiered or battery approaches? For example, Lessons from Rodents could be moved.
Kaplan:	I think we can move 2 and 3 up to State of the Art.
Germolec:	We should add an introduction here to in vitro tiered approaches, giving some history. And do we want to move that from where it is now or simply repeat it for emphasis.

Corsini:	I will integrate 2 and 3 into State of the Art, and then add an introduction here to in vitro tiered approach.
Neff-LaFord:	But we will also need to add a conclusion.
Corsini:	Yes, but we will write a discussion and conclusion after we have finalized the content we have now.
Loveren:	Is there really a need to discuss the in vivo tiered approach in this paper? I don't think it has a real bearing on what we are doing here.
Germolec:	I think having that background here establishes a rationale for using the tiered approach for the immune system.
Corsini:	The benefits of a tiered approach are obvious to some, but there are other who are less familiar.
Kaplan:	We don't what to give too much detail about in vivo, just use it to provide a rationale for why a tiered approach is also valid for in vitro testing.
Germolec:	Should the reference chemical section come after Section IX? That way we introduce the MITA and then provide a list of chemicals that can be used to validate any in vitro assay.
Germolec:	We need to remove the "non-immunomodulatory" from headings, because that is misleading.
Kaplan:	We need to be clear in this section as well that the MITA is IL-2 and interferons, but the modified MITA includes IL-8. So, we need to explain clearly what is MITA, what is modified MITA, and where the clusters come from.
Germolec:	What I am struggling with is that we have these clusters, but we don't have a clear concept of what each cluster means for predicting immunotoxicity. Does it mean potency, does it mean mechanism?
Aiba:	They are grouped per their response in the assays.
Germolec:	OK, but if I test an unknown compound and its response in the assays resembles formaldehyde, what does that mean? That is the piece we are missing. We need to include some characterization of the immunosuppressive effect of each cluster.
Corsini:	What we have here is a way of showing that a chemical is, for example, both an immunosuppressive and a skin sensitizer. But we need to spell out for the reader what each cluster represents.
Neff-LaFord:	Yes, use the figure showing the hierarchical clusters but add a table that explains the characteristics of each cluster. But the problem is that we also need to clarify that we are now introducing the element of sensitization, whereas the focus until now has been on immunomodulation.
Aiba:	I will make that table.
Germolec:	We should link the IL-8 Luc assay to the sensitization AOP so that that we can tie them together in this section.

Kaplan:	There seems to be a focus on MITA being IL-2 when actually it includes three cell line and four endpoints.
Neff-LaFord:	If LA refers luciferase activity, then that should be spelled out so that it is less confusing. And we are talking about both MITA and modified MITA in this paper, so we need to clarify that throughout.
Corsini:	Are there any knowledge gaps here that we need to address?
Germolec:	Do we need to propose activities leading to the establishment of a defined approach?
Neff-LaFord:	Do we need to address chemicals that present immunoaugmentation in these assays?
Germolec:	What about expanding the number of chemicals with published test results from a tiered approach? How about the applicability domain and what the limitations are?
Aiba:	We might have to discuss the administration or concentrations, which can also affect the interpretation of results.
Loveren:	That is a consideration but not a data gap.
Germolec:	We might need to distinguish here between key considerations and data gaps. How does a change in cytokine expression translate into health risk assessment? So perhaps we need to call out in-vivo–in-vitro extrapolation, and also mention that it is for all in vitro studies, not just immunotoxicity. And then we need a conclusion, that might be a place to discuss the use of multiple assays to test different aspects of immunotoxicity.
Corsini:	And say that much progress has been made in the field of immunosuppression using non-animal approaches.
	Schedule
	Revisions by end of February
	Collated document distributed before SOT
	Additional comments by end of March
	Conference call in April before WNT
	Present document to WNT at end of April
	Conference call in May
	Final document by end of June

Expert Group on DASS

Teleconferences 18/19 November, 2019 14h30-17h (Central Europe)

Attendees
18 November: Susanne Kolle, Donna Macmillan, Chantra Eskes, Knud Ladegaard Pedersen, Amaia Irizar, Nathalie Printemps, Michele Regimbald-Krnel, Silvia Casati, Erika Witasp Henriksson, Andrea Gissi, Tim Singer, Grace Patlewicz, Andre Muller, Raja Settivari, Henrik Tyle, Elena, Eva Bay Wedebye, Martin Paparella, Sebastian Dunst, Judy Strickland, Kristie Sullivan, Nicole Kleinstreuer, Petra Kern, Roman Liska, LMC, Martina Klaric, Nathalie Alepee, Matthias Herzler, Andreas Natsch, Paul Brown, Janine Ezendam, Andrew Williams, Emma Grange, Hermann-Josef Thierse, Laura Rossi, Gavin Maxwell, Anne Gourmelon, Patience Browne

November 19: Kristie Sullivan, Michele Regimbald-Krnel, Takao Ashikaga, Andrea Gissi, Silvia Casati, Chantra Eskes, Donna Macmillan, Janine Ezendam, Judy Strickland, Tim Singer, Grace Patlewicz, Petra Kern, Nicole Kleinstreuer, Nathalie Alepee, Andreas Natsch, Betty Hakkert, Gavin Maxwell, Matthias Herzler, Ovanes Mekenyan, Sebastian Dunst, Emma Grange, Martin Paparella, Laura Rossi, Hermann-Josef Thierse, David Asturiol, Erika Witasp Henriksson, Andrew Williams, Amaia Irizar, Jong Kwon Lee, Eva Bay Wedebye, LMC, Susanne Kolle, JaCVAM, Andre Muller, Nathalie Printemps, Andrea Gissi, Gavin Maxwell, Anne Gourmelon, Patience Browne

DAY 1

1. Discussion of LLNA and Human reference data
    a. Matthias Herzler (BfR) presented an evaluation of uncertainty in LLNA data
        i. Table 4 summarised reference chemical classifications for
            1. binary calls (GHS_{BIN}),
            2. GHS subcategory classification of potency (GHS_{SUB}), and
            3. using the new proposed approach, two additional classifications for borderline 1A/1B chemicals =1 and 1B/NC chemicals (GHS_{BORDER}).
        ii. In most cases (86/108 chemicals), the analysis from the LLNA data subgroup (using the Median-Like Location Parameter (Hoffmann et al. 2018)) and this new approach produced consistent results for LLNA reference chemical GHS classifications.
        iii. For 17 chemicals, previous classifications were proposed to be changed to “ambiguous” calls (1 or NC/1B).
        iv. The EG discussed the change to calls, specifically
            1. Changes from NC to 1B/NC would reduce the number of negative reference chemicals to 11
            2. The change in potency calls for some reference chemicals was based on rejecting negative results with maximum test chemical concentrations of $\leq$25%
            3. Feedback was provided, that while few chemicals may be tested at higher concentrations, it is rare.
        v. A calculation of the LLNA reproducibility based on the number of studies available per individual substance and removing the ambiguous results increases the reported value considerably (~95%).
    b. The JRC presented an overview of ambiguous LLNA reference chemical classifications in the LLNA data subgroup database supporting the revised GL and SD
        i. The presentation included some feedback on specific chemicals received following the 1 November commenting round
        ii. As with Matthias’s presentation, for many chemicals, understanding the data required a deeper examination of the LLNA studies

c. The US presented an overview of ambiguous human reference chemical classification from the human data reference subgroup report and database
    i. The presentation also included feedback on specific chemicals received following the 1 November commenting round
    ii. Several experts commented that human data could not be reliably used for potency but a comparison for hazard assessment should be included in the SD (and summarised in GL)
d. Following some discussion of individual chemicals, **a decision was taken to reflect on the analyses and provide input on a final reference chemical list by 26 November** to:
    i. address questions regarding specific chemical data, and
    ii. propose an approach for dealing with borderline chemicals

DAY 2 – addition topics for discussion in presentation from US

2. The group discussed the status of the 2 out of 3 DA in the draft GL
    a. As stated in the email from Anna Lowit (18November 2019), the US is accepting the DA in lieu of LLNA data for hazard classification
    b. Several members of the DASS EG noted the high FN rate and the questionable relevance of the 2 out of 3 DA for negative calls
        i. While it was noted that the DA predictions may change a bit with updated reference chemical data, the FN rate is not expected to change dramatically
        ii. It was also noted that the mispredictions often involve chemicals with borderline LLNA results
    c. The US presented several options for the 2 out of 3 DA moving forward
    d. DK remarked that there is support for moving the ITS forward at this time but the 2 out of 3 may benefit from additional analyses
3. The versions of the ITS DA were discussed
    a. Several comments remarked that it was unclear if the ITS v1 (using Derek) and v2 (using OECD QSAR TB) should be one or two DAs.
    b. There was general agreement that these should be two versions of one DA
        i. Donna Macmillan/UK expressed a dissenting opinion
    c. The text proposed by DK to describe the change in the DIP for the ITS was supported with a modification to explain the rationale of the revised cutoff to 6 (due to the change between ECETOC and GHS potency classification systems and resulting in better prediction of 1A chemicals, etc.)
4. Substitute in vitro and in silico data in DAs
    a. The EG supported including more details on assumptions made and protocol to produce in silico predictions in GL (e.g. in an annex)
    b. More details are needed in the text proposed by the UK/LMC regarding “sufficiently reliable” commercial in silico tools (in addition to Derek and OECD TB)
        i. The group supported reversion to previous version with more details regarding publications, QMRF, etc.
        ii. The following text is proposed “The ITS may use other reliable in silico tools which are well supported by publications, have QMRF with complete information for the model, such as training set, applicability domain, performance, etc. with documentation equivalent to the in silico tools included in the GL.”
            1. This will be further refined between the leads/Secretariat
        iii. DK suggests that similar language be crafted for other in vitro information sources.
5. Comments on applicability domain
    a. More guidance may be needed on how to use metabolism/reactivity domain profilers – these are not intended to be used as QSARs and may need to be validated

b. Comments received regarding applicability domain were discussed regarding more/less prescriptive information on the applicability domain (Figure 1 from GL)
    i. EG suggested framing the analysis as a description of the current dataset rather than an interpreting the figure as prescriptive
    ii. Suggestion to remove statements recommending methods/DAs as "better" for some chemicals than others, as this only reflects the limited number of chemicals tested in each domain
    iii. Details should be only in the SD, including how reactivity domains were defined for this analysis
    iv. Leads will revise figure to clarify workflow and associated text
c. Add text to GL specifying that mixtures/UVBCs, (and in this case) metal, and inorganic chemicals are outside the applicability domain

6. Responses to comments on me-too in vitro information sources
    a. DK suggested "parking" the issue for now and having a broader discussion at the WNT
        i. Support for adding the topic to the 2020 WNT meeting
    b. Topic is probably easier to address in the hazard identification mode than for potency
        i. For ITS with quantitative cutoffs, they are assay specific and cannot be extended to a drop-in replacement
    c. As noted in the discussion of in silico, language could be included to suggest other options may be acceptable
7. Next steps
    a. EG will have an additional chance to provide feedback on the LLNA and human reference chemicals (circulated in September) and the analysis from Matthias and provide additional chemical-specific consideration until 26 Nov
    b. TC scheduled for 1st week in December to resolve final reference chemicals